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

December 31, 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 and discussion sections.

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 same 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.

## Author's response

We are very grateful for the insightful comments of the reviewer. We use the same CIF specification of Cederkvist et al. (2019) but with a different likelihood formulation (a full likelihood approach instead of a composite one) and parameter estimation routine (a Laplace approximation instead an expensive AGQ), taking advantage of state-of-the-art computational libraries and algorithm implementations. We stress out these points in the paper, as recommended. In terms of motivational applications, it was also added to the introduction.

## Reviewer 1

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?

## Author's response

We are very grateful for the insightful comments of the reviewer. We had access to Cederkvist et al. (2019)'s code through GitHub. Running their code we've not been able to reproduce their results. By adapting it to perform some extra simulation studies like the ones that we did with our model, in the majority of the scenarios convergence has not been reached. Given that and the facts that 1) our likelihood functions, formulations, and estimation procedures are very different and not directly comparable; 2) our model runs in parallel and its running time is quite fast considering its complexity and sample size, we end up choosing not to make any statement about these differences in the results section.

### Reviewer 1

Minor

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

### Author's response

We are very grateful for the insightful comments of the reviewer. The clarification and the correction were made.

## 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)?

### Author's response

We are very grateful for the insightful comments of the reviewer. We use the same CIF specification of Cederkvist et al. (2019) but with a different likelihood formulation (a full likelihood approach instead of a composite one) and parameter estimation routine (a Laplace approximation instead an expensive AGQ), taking advantage of state-of-the-art computational libraries and algorithm implementations. We stress out these points in the paper, as recommended.

### Reviewer 2

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)?

### Author's response

We are very grateful for the insightful literature recommendation. The model proposed by He et al. (2022) is also a valid approach to handling clustered competing risk data. The mentioned approach is similar to ours in the sense that it is not necessary to model the censoring distribution. The differences lie in the model category and parameter estimation and uncertainty quantification. We have a parametric approach based on a proper likelihood function instead a semiparametric one based on a pseudo-likelihood. With a standard likelihood formulation based on GLMM, we can take advantage of all well-established inferential frameworks. With the semiparametric pseudo-likelihood formulation a nonparametric estimator was needed to be proposed together with a sandwich covariance estimator extra computations.

### Reviewer 2

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?

### Author's response

We are very grateful for the insightful comments of the reviewer. EM algorithms are a very interesting maximization routine since they can be applied virtually to any model with latent components, but they present the caveats of being slow, in the majority of the cases, and that each model requests a specific EM formulation. Here, with our model being easily written in the exponential family framework, a very general and immediate Laplace approximation is easily obtained. This very practical formulation together with state-of-the-art Laplace approximation implementation and the automatic-differentiation tool makes it a better deal than the EM algorithm for our model specification.

## 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.

He, Y., Kim, S., Mao, L. and Ahn, K. W. (2022). Marginal semiparametric transformation models for clustered multivariate competing risks data, *Statistics in Medicine* **41**: 5349–5364.