Center for Innovative Design & Analysis
colorado school of **public health**

December 19, 2021

Resubmission of Manuscript 2020-190 “cpsurvsim: An R Package for Simulating Data from Change-Point Hazard Distributions”

Dear Dr. Michael Kane, Executive Editor, and the reviewers,

On behalf of my coauthor, I thank you the thoughtful review of our manuscript. In response to your comments, we added additional examples to the package documentation, clarified the scope of our simulation study, and strengthened the argument for the utility of our R package. Following this letter are detailed responses to each reviewer comment, with changes reflected in our updated manuscript. Your reviews helped us to improve this paper and we hope that you agree.

Thank you for the opportunity to revise and improve our manuscript for *The R Journal*. We greatly appreciate your time and re-consideration of our work.

Sincerely,

Camille J. Hochheimer, PhD

Response to reviewers

> -- Reviewer 1 --

>cpsurvsim: An R Package for Simulating Data from Change-Point Hazard Distributions by Camille
>J. Hochheimer and Roy T. Sabo

>The R package "survival" originally attributed to Terry Therneau is concerned with modelling,
>model fitting but not simulation. This popular package is frequently cited and taught in every
>classroom so there is clearly a need for packages such as the one being proposed here.

We agree. This was the inspiration for publishing this R package.

>Things I like: the ability to generate time-dependent covariates, albeit binary valued. Every
>feature in this package being proposed has several references to motivate its application and
>value to users. I poked around the code on Github and like there was ample parameter checking
>at the beginning of routines.

Thank you for this feedback.

>Something I would like to see added to the wrapper: An example where the whole model is crazy
>wrong. For example, simulate change-point survival times with a decreasing step-wise constant
>hazard and then fit a parametric model that does the opposite. This type of example would be
>instructive to a clueless user who is really fitting an inappropriate model but doesn't know
>better.

We thank the reviewer for this suggestion. We have added this kind of example to the wrapper for each of the simulation functions in our package. This update has been submitted to CRAN.

>Future directions to build a larger package: Use the bootstrap to simulate survival data from
>existing data under such models as Cox, accelerated failure time, and your changepoint models.
>Anything with accelerated failure time models (I don't feel this model gets the credit it deserves).

Thank you for these suggested extensions. We feel there is strength in modular packages and look to other researchers to implement simulation methods for these important models in future R packages.

>Informative censoring. I don't know what this entails but few researchers want to discuss this. It
>might be a good thing to add to encourage the conversation.

This is another interesting suggestion. Unfortunately, it is a non-trivial extension. We have added this to the discussion as a future direction.

> --Reviewer 2--

>Report on the manuscript by Camille J. Hochheimer and Roy T. Sabo titled as "cpsurvsim: An R
>Package for Simulating Data from Change-Point Hazard Distributions"

>This manuscript introduces the cpsurvsim package provides two methods for simulating data
>from a both the exponential and Weibull hazard model with Type I right censoring allowing for
>multiple change-points. It is also demonstrated how to simulate data using cpsurvsim and
>compare the performance of these methods through a simulation study. The manuscript looks
>interesting and may contribute the reader of R Journal, however I have some concern about the
>simulation study such as:

>• Why the intervals are taken as -+5 in Figure 1? Is there any specific reason for that? It needs to
>be clarified.

We thank the reviewer for pointing out the lack of clarity on this decision. We used +/- 5 because it represents a range of 10% within our time range from 0-100 and have added this clarification to the simulation methods.

>• Why the results did not compared with the other packages which generate data for one change-
>point?

Ideally, we would've compared with other existing packages, however, our package is different enough that this would not have made for a meaningful comparison. We are unable to do a direct comparison with CPsurv because it relies on existing data (survival time and censoring) as input whereas our functions generate this information. SimSCRPiecewise requires at least one patient covariate as an input, which was not part of our simulation study. We've added some information on these points to the Introduction.

>Also, some real data examples could be given to show the importance of the what package
>provides us in real life? Why we need to use it?

For packages detailing exclusively methodological functions, a demonstration on existing data is an excellent way to exemplify performance. However, this R package focuses on creating simulated data, not in analyzing existing data. Our work in creating this package was inspired by the simulations needed for our research on developing a test for multiple change-points. We hope that others will be able to use it for similar purposes, to address the appropriateness of a change-point model for data, or to conduct a power analysis. We've added a comment on this to the introduction.

>Another important point is about the sample size. It is indicated that the performance of the
>package is poor when the sample size is 50 and 100, this need to be discussed.

We share the reviewer's concern about the poor performance with sample sizes of 50 and 100. We suspect that the issues in accuracy come from relying on a sample that would be reasonable for estimating a single shape and scale parameter to estimate several additional scale parameters. This note has been added to our discussion section where we discuss other potential simulation methods to address inaccuracies within small sample sizes.

>One more suggestion is that the R-codes must be given in a github repo or etc. for reproducibility >of the simulation.

We agree with the reviewer that it is helpful to share the R code for this simulation. For that reason, it is included in the files for this submission, however, we are also happy to publish to GitHub once this paper is in press.

>I think that the paper needs to be improved. Especially, there are vogues must be eliminated on >simulation studies. I suggest that re-evaluation after this revisions are done.

Thank you for your thoughtful review. We hope that we have successfully addressed your concerns.