Review 1: Author responses

Panis & Ramsey

2025-07-14

Editor

> 15-Mar-2025
>
> Dear Dr. Panis:
>
> Thank you for submitting your Tutorial entitled "Event History Analysis for psychological time-to-event data: A tutorial in R with examples in Bayesian and frequentist workflows" . I have now received input from three experts in the field, and you will find the reviews below this message and attached. As you will see, all three reviewers believe your paper can make a strong contribution to the literature. Prior to considering the reviewers' input, I read your paper to form an independent opinion. I share the positive sentiments about this work and invite you to revise and resubmit your manuscript for further consideration.

> The reviews are through and thoughtful, and I will not reiterate all the details here. For my part, I would like your revision to focus on accessibility. Your paper is a bit challenging– it's long and a little tough to get through. What can make it easier for less familiar readers? Perhaps more of a basic set up earlier? A glossary defining specific terms (e.g., what is an event? how is it defines? censoring? hazards vs. survival curves, conditional probabilities, etc.)? Also, clarify when readers need to shift from frequentist to Bayesian model... and why? The main consideration here is that wish to make the paper much more accessible– from start to finish– for readers who are novices and entirely unfamiliar with this area. This can be done any number of ways, and R3 also includes some suggestions for making these reviews. I encourage you to have the revised paper read and pre-reviewed by colleagues who are novices in this area with the explicit goal of making the paper accessible to them– this will help you hit the more for our broad readership.

> If you choose to submit a revision, please include include a letter detailing your point-by-point responses to each reviewer comment and indicating how you changed the manuscript to address them. The revised manuscript may undergo further peer review. We ask that you submit your revision within three months. Please let us know if you will not be able to meet this deadline.

> Sincerely,

Response: Dear Editor,

thank you for providing constructive feedback and allowing us to submit a revision of our manuscript (AMPPS-24-0232).

We tried to focus on accessibility in the revision of our Tutorial in several ways. First, we reduced the number of Figures: we replaced Figure 1 with a simpler Figure that directly conveys our message and deleted Figures 2 and 3, and we replaced the 6 old figures on modeling results with 2 new figures. Second, we shortened the Introduction and Discussion sections. Third, we added a section in the supplementary material (section A) in which we visualize the different types of time-to-event data that are obtained in typical RT tasks

(detection, discrimination, bistable perception). We also mention the unique contribution of our work, we removed redundancies, and tried to improve the clarity of the paper. As such, we have substantially revised the manuscript in accordance with feedback from you as well as the reviewers.

Reviewer: 1

> Comments to the Author It was a pleasure to review this manuscript as someone who has spent years applying survival analysis methods to psychological data. This is an ambitious manuscript and I believe it achieves its goals. The key contributions of this manuscript from my perspective are the creation of custom functions for data wrangling, and the section on planning studies/simulations for power analysis (of course, in addition to the introduction to this category of methods and the practical tutorials provided. I have some comments on this work that I believe could strengthen this manuscript's contribution.

- General comments:

> The Introduction is really nicely framed in terms of providing compelling information for why this class of statistical models would be useful in some specific areas of psychology (i.e., reaction time data). Given that there are already a couple of tutorial-types of articles introducing specific applications of survival analysis in developmental psychology (references below, both regarding the use of these methods for behavioural observations of emotional expressions in a multilevel framework for recurring events), it would be helpful to further frame the unique contribution that this work adds. The authors may also consider further emphasizing the specific theoretical contributions that using these methods may lead to in the areas of reaction time data etc.

Response: In the revised ms. we frame the unique contribution that this works adds on page 6: "... we are not aware of any tutorials that are aimed specifically at psychological RT (+ accuracy) data, and which provide worked examples of the key data processing and Bayesian multilevel regression modelling steps."

> Lougheed, J. P., Benson, L., Ram, N., & Cole, P. M. (2019). Multilevel survival analysis: Studying the timing of children's recurring behaviors. Developmental Psychology, 55(1), 53–65. https://doi.org/10.1037/dev0000619

> Stoolmiller, M. (2016). An introduction to using multivariate multilevel survival analysis to study coercive family process. In T. J. Dishion & J. J. Snyder (Eds.), The Oxford handbook of coercive relationship dynamics (pp. 363–378). Oxford University Press.

> Stoolmiller, M., & Snyder, J. (2006). Modeling heterogeneity in social interaction processes using multilevel survival analysis. Psychological Methods, 11(2), 164–177. https://doi.org/10.1037/1082-989X.11.2.164

Response: We added these references to the revised ms. on page 6.

> Similarly, it would be helpful to make it clear that this manuscript pertains to estimating single-event occurrences rather than recurring events (which are described in the Lougheed and Stoolmiller tutorials mentioned above)

Response: We explicitly mention this on page 6 of the revised ms. Also, we added a section in the Supplemental Material (section A in the revised ms.) which explains the types of time-to-event data typically obtained in RT tasks (detection, discrimination, and bistable perception tasks). This section makes it clear that our tutorials pertain only to estimating single-event occurrences. It also visualizes recurrent events and includes references to the Lougheed and Stoolmiller tutorials.

Specific comments:

The authors use several different terms to refer to time-to-event models throughout the manuscript. The title refers to EHA, and the running head refers to Hazard analysis. Survival analysis is also a common term in Psychology (see tutorials by Lougheed et al., 2019, and Stoolmiller & Snyder, 2006). Disambiguating the various terms early in the manuscript and then constantly using one term throughout the manuscript would be helpful.

Response: We disambiguate the various terms in the revised ms. on pages 7-8, and we now only use one term throughout.

Page 5: it would be helpful to refer to the same time unit in reference to Figure 1, which shows time bins along the x-axis but the examples discussed in the manuscript refer to units of milliseconds.

Response: Figure 1 has been replaced.

Page 7: The authors state that they are not aware of EHA tutorials for psychological time to event data. Previoulsy mentioned tutorial articles by Lougheed and Stoolmiller could be mentioned here— these are demonstrating applications to behavioural observation data within psychology.

Response: We added these 3 references to the revised ms. on page 6.

Page 8: I'm curious about the decision to focus on discrete time rather than continuous time models. It would help the reader to explain the decision either way, or refer the reader to Lougheed et al., 2019 who discuss this distinction and choices that researchers may make regarding how to handle decisions around time given data formats. I don't work with reaction time data, but it seems to me that many experiments in this area would collect data in continuous time intervals, and it's not clear to me what the benefit of discretizing into time bins is. Now that I've read the entire manuscript, I see that some information is presented on this on Page 50, but it would be helpful to make this information known to the reader earlier in the manuscript.

Response: We address our choice for discrete-time methods on pages 8-9 of the revised ms.:

"... the definition of hazard and the type of models employed depend on whether one is using continuous or discrete time units. As a lab, and mainly for practical reasons, we have much more experience using discrete-time EHA, and that is the approach that we describe and focus on in this paper. This choice may seem counter-intuitive, given that RT is typically treated as a continuous variable. However, continuous forms of EHA require much more data to estimate the continuous-time hazard (rate) function well (Bloxom, 1984; Luce, 1991; Van Zandt, 2000). Thus, by trading a bit of temporal resolution for a lower number of trials, discrete-time methods seem ideal for dealing with typical psychological RT data sets for which there are less than ~200 trials per condition per participant (Panis, Schmidt, et al., 2020). Moreover, as indicated by Allison (2010), learning discrete-time EHA methods first will help in learning continuous-time methods, so it seems like a good starting point."

Page 8-9, excellent and accessible explanations of the survival function in the discrete time framework.

Response: Thank you for the positive feedback. To make the manuscript more accessible and shorter and to respond to some other comments from the editor and other reviewers, we moved the explanation of the survival function to the Supplementary Material in the revised ms.

Page 11: In my experience, researchers in psychology are generally not familiar with the concept of right censored data. This term is first used on page 11 with no explanation— a brief, even parenthetical, definition would be helpful. For example, the parenthetical definition on page 48 could be provided on page 11.

Response: We introduce the concept of right-censoring on page 11 of the revised ms.

> Taken together, this is an excellent manuscript with the potential to further bring survival analysis methods into Psychology, where time-to-event and right-censored cases abound, yet very few people seem to consider these issues.

Thank you very much for your helpful and positive feedback.

Reviewer: 2

- Comments to the Author Please see attached file for comments.

> Reviewer report for AMPPS-24-0232: "Event History Analysis for psychological time-to- event data: A tutorial in R with examples in Bayesian and frequentist workflows" Summary: The authors present a tutorial paper on the analysis of discrete time-to-event outcomes. They motivate their tutorials with a description of psychological experiments that produce time-to-event data and explain how simply summarizing event-time outcomes using means results in a loss of potentially important information. They then explain how more information can be gleaned from event-time data if they are summarized using event history analysis methods. They focus specifically on the analysis of discrete time-to-event outcomes. They provide code for conducting the analyses in the tutorials on their Github. Overall, I like the idea of this paper and strongly agree that using methods appropriate for analyzing time-to-event data, rather than simply summarizing event-time outcomes using means, is very important and currently underused. While I like the premise of the paper and think that such a paper could be a valuable contribution to the literature, I do think the paper has room for improvement in a number of areas (see comments below). The writing is approachable (which is advantageous for a tutorial paper) but some text in the paper was repetitive and could be a bit more polished; additional editing with a focus on clarity and brevity would be beneficial.

Response: Thanks for your constructive and encouraging feedback. As we also noted in our response to the editor, we have made substantial revisions to the manuscript that include removing repetitive text, and placing a stronger focus on clarity and brevity in the revised ms. For example, we describe the modeling results in the paper using only two instead of six figures.

> Major comments: 1. Some discussion on the choice of bin width is needed. Since all of the example time-to-event data is collected in continuous time, the analyst must discretize it by defining bins. Defining bins is a key step in the analysis pipeline and so should not be overlooked. It seems like results—and potentially conclusions drawn—may depend on the choice of bin width. What factors should the analyst consider when defining the width of the bins? Does the number of bins matter? Or is the number of events per bin important? What about other factors? Should analysts (or potentially these authors too) consider multiple bin widths when conducting an analysis? An example comparing results when bin width varies could be both interesting and informative.

Response: We discuss the issue of defining bins on page 12 of the revised ms.:

"Third, the width of each time bin will need to be determined. For instance, in Figure 1B we chose 100 ms in an arbitrary manner. In reality, however, bin width will need to be set by considering a number of factors simultaneously. The optimal bin width will depend on (a) the length of the observation period in each trial, (b) the rarity of event occurrence, (c) the number of repeated measures (or trials) per condition per participant, and (d) the shape of the hazard function. Finding an appropriate bin width in a given user case before fitting models will require testing a number of options, when calculating and plotting the descriptive statistics (see section 3.1). The goal is to find the smallest bin width that is supported by the amount of data available. Based on our experience, a bin width of 50 ms is a good starting value when the number of repeated measures is 100 or less. Too small bin widths will result in erratic hazard functions as many bins will have no events, and thus hazard estimates of zero. Interestingly, the time bins do not need to have the same width. For example, Panis (2020) used larger bins towards the end of the observation period, as fewer events occurred there."

2. In the discussion, the authors mention that they choose to present a tutorial on discrete time analysis methods because they believe these are easier to understand. My personal impression is that awareness of continuous time analysis methods is slightly higher and so researchers may be more familiar with methods such as Cox models, even if they have not used them themselves; however, this could vary by specific sub-field. I agree with the authors that the interpretation of the hazard is easier in the discrete time setting than in the continuous time setting, but I do think that the distinction between the interpretation of the hazard in continuous vs discrete time should be discussed in the introduction. When I first read the introduction, I was confused as to why the focus was on discrete time methods, when the outcome measures collected in the motivating data were collected in continuous time. I would suggest either of the following: (a) provide concrete examples of when truly discrete timeoutcomes are collected to strengthen the motivation of the paper and justification for use of discrete time analysis methods or (b) keep the motivating examples as- is and include a short comparison of discrete vs. continuous time analysis methods to help explain your reasoning for choosing to focus on discrete time analysis methods.

Response: We address our choice for discrete-time methods on pages 8-9 of the revised ms.:

"... the definition of hazard and the type of models employed depend on whether one is using continuous or discrete time units. As a lab, and mainly for practical reasons, we have much more experience using discrete-time EHA, and that is the approach that we describe and focus on in this paper. This choice may seem counter-intuitive, given that RT is typically treated as a continuous variable. However, continuous forms of EHA require much more data to estimate the continuous-time hazard (rate) function well (Bloxom, 1984; Luce, 1991; Van Zandt, 2000). Thus, by trading a bit of temporal resolution for a lower number of trials, discrete-time methods seem ideal for dealing with typical psychological RT data sets for which there are less than ~200 trials per condition per participant (Panis, Schmidt, et al., 2020). Moreover, as indicated by Allison (2010), learning discrete-time EHA methods first will help in learning continuous-time methods, so it seems like a good starting point."

3. Censoring is an important part of time-to-event data and so more description of censoring—and how it is (not) accounted for in "orthodox" analysis methods vs. discrete time-to-event analysis methods would be useful, beyond the already- included mention of censoring in Section 5.3.

Response: We introduce the concept of right-censoring on page 11 of the revised ms.

4. Section 1.2 is a bit long. Please shorten this section to improve the clarity of the text and reduce redundancy. There is also overlap between text in Section 1.2 and the start of Section 4.

Response: We shortened this section and removed the overlap.

5. Section 2.3.2. I like the idea of making a connection with experimental design, but am having trouble seeing how the discussion in this section is specific to EHA and how these comments do not apply also to the "orthodox method". I think this section could be improved by providing more specific distinctions between experimental design implications for EHA vs. the "orthodox method".

Response: We created a new Figure 1 that makes the intended point, and removed section 2.3.2.

6. Line 517: What are model weights? Please explain briefly for the reader, including how the interpretation differs for loo and waic. Also, in the provided reference (Kruz 2023a), I was only able to find a discussion of dWAIC, rather than WAIC, so could you please include a definition of or reference for dWAIC in your context? Lastly, when reporting the model weights in the example in the text, I'd suggest including more than 2 digits so that it is clear if the reported weights are actually 0 and 1 or if rounding has just made them appear so.

Response: We briefly explain model weights for loo and waic on page 27 of the revised ms. A discussion of WAIC in Kurz (2023a) can be found in section 4.6.4., so we did not include a reference for dWAIC in the revised ms. Lastly, we print more digits when reporting the model weights.

7. Tutorials 3a and 3b: Rather than simply stating that the frequentist models did not converge and resulted in singular fits, it would be helpful to provide more discussion. Do you have an idea why the models did not converge? Since this is a tutorial paper, more explanation is needed. These frequentist examples don't add much to the paper in their current form, so I'd either suggest either (a) providing some suggestions to the reader of alternative approaches or specific strategies that could be used to define a model that will converge or (b) removing these examples from the main text and directing the reader to the supplement instead.

Response: We removed the frequentist examples from the main text and direct the reader to the Tutorials on page 32 of the revised ms. We mention in Tutorials 3a and 3b that it is common for models to fail to converge in lme4 when they have a reasonably complex random effects structure, and we provide a reference.

Minor comments: 1. Is "mean-average comparison" standard terminology? This phrase seems redundant to me, since means and averages are the same thing.

Response: We removed this term.

2. Figure 1: In the caption, please state what the small error bars in the inset plots are showing.

Response: The old Figure 1 has been replaced, and we state what error bars are showing in the new Figure 1.

3. Please consider updating all plots so that they are understandable when viewed in black and white. In addition to using different colors, I'd suggest distinguishing the conditions using shapes (e.g., circles vs. triangles) and different line type (e.g., solid vs. dashed lines).

Response: We updated all plots by distinguishing the conditions using different shapes and/or line types, so that they are understandable when viewed in black and white.

4. Figure 2 caption:a. Please provide a reference for the statement beginning, "Because the survival function. . . ".

Response: Figure 2 has been removed. We only discuss the survival function in the Supplemental Material (section B) of the revised ms.

b. I am confused by the sentence beginning: "For example, the high hazard of . . . ". Could you please provide some additional explanation or a reference for more details on this interpretation?

Response: We removed this section.

5. Lines 156-166: Generally, survival probabilities are described as probabilities of remaining event-free up until a certain point, rather than the probability of an event occurring in the future. Some clarity in wording could be used here.

Response: We follow this definition in section B of the Supplemental Material.

6. Please confirm that you have defined all acronyms, abbreviations, and functions at their first use in the text. E.g., please define WAIC, LOO, P(t), ca(t), SOA, CrI, among others.

Response: We define all acronyms on their first mention in the revised ms.

7. Section 3.1 could be condensed in the main paper (additional details could be included in the supplementary material if important). Also, please double check your definition of functional programming.

Response: Section 3 has been removed in the revised ms.

8. Table 2: It could be helpful to show an example of a trial where an event never occurs, to help the reader connect what is written in the text (line 325-327) to what appears in the table.

Response: We added data visualizations in section A of the Supplemental Material.

9. Figure 4: Why is there no bin 40?

Response: Because no responses occurred in the first bin with endpoint 40 ms. However, we added this bin back in Figure 2 (previously 4) of the revised ms to avoid confusion.

10. Line 423: The phrase "early responses" is confusing as it implies temporal ordering of the responses (i.e., early vs. late trials). I'd suggest using "fast response" or "rapid responses", or some other phrase along those lines.

Response: We replaced early with fast on page 21 of the revised ms.

11. Is the discussion of reference and index coding (lines 457 – 464) needed? In the introduction, the authors state that knowledge of regression is assumed so I am not sure this is necessary to include, but I defer to the authors' judgment.

Response: Index coding is not used a lot in experimental psychology so we included a short description.

12. Figure 6: I only see one width of credible interval for each line on this plot— please clarify if you are plotting 80% or 95% credible intervals.

Response: We corrected this (Figure 4 in the revised ms.).

13. Line 536: To me, "subject-specific" and "marginal" mean the opposite things. That is, subject-specific is an effect that is conditional on an individual while a "marginal" effect has been marginalized/averaged over individuals. Could you please clarify your use of these terms here?

Response: We feel there is no real consensus in the literature about the use of the term marginal. We thus removed this term in the revised ms. But we kept the term in tutorials 2a and 2b where we base our use of the terms on this cited reference: Heiss (2021).

14. Lines 584-587: Since this is a tutorial paper, I'd suggest including explanations for why you make changes (a), (b), and (c).

Response: We now write "The general process is similar to Tutorial 2a, except that (a) we use the person-trial data, (b) we use the symmetric logit link function, and (c) we change the priors (our prior belief is that conditional accuracy values between 0 and 1 are equally likely)." on page 30 of the revised ms. The logit is typically used for logistic regression by experimental psychologists, so we feel that no extra explanation is needed.

15. Line 659: What is a "fully varying effects structure"? Please define/explain.

Response: We explain this on pages 22-23 of the revised ms.: "We also use a"keep it maximal" approach by specifying a full varying (or random) effects structure. This means that wherever possible we include varying intercepts and slopes per participant."

16. Line 832-833: Please provide a reference for the statement, "However, they require much more data. . . ".

Response: References are provided on page 8 of the revised ms.

17. Github code: Instruction 2 of the README states "open the reproducible_workflow.Rproj" – I can't find this R project, should it be the "Tutorial_EHA.Rproj" instead?

Response: Yes, this has been corrected now.

Reviewer: 3

Comments to the Author This tutorial guides readers through how to wrangle cognitive reaction time data and address research questions pertaining to temporal dynamics of cognition using event history analysis (EHA) and speed/accuracy tradeoff analysis (SAT). Models are illustrated for hazard functions, and conditional accuracy functions. A nice feature of this study is that each model is presented in both the Bayesian and frequentist frameworks. Having both frameworks provides side-by-side comparisons such as when the Bayesian framework might be preferred (e.g., when models do not converge in the Frequentist framework, which is common when modeling multilevel data). The tutorial also includes a section on simulation and power analysis for planning experiments. I believe this tutorial could eventually be a useful contribution to fellow psychological scientists. However, in its current state, it was difficult to follow/ make my way through (despite existing expertise in EHA), and also difficult to understand what the main contributions were above and beyond existing works. Additionally, it was not clear why the discrete time framework was chosen. The temporal granularity of reaction time data seems like it would be better suited for a continuous time framework. Altogether, the tutorial covers important concepts, but it would benefit from greater accuracy and clearer definitions of key terms. Carefully removing redundancies, getting to the tutorial section much quicker (currently not until pg. 17), and being more precise with explanations of technical jargon would enhance its clarity and accuracy.

Response: Thank you for your constructive feedback. We discuss the main contributions above and beyond existing works on page 6 of the revised ms. We address the issue of treating time as discrete or continuous on pages 8-9 of the revised ms. We removed redundancies, get to the tutorial section quicker (page 13) and included more precise explanations of technical jargon.

Code for the tutorials: I went to the github page and found a folder for each of the 4 tutorials, which included sub-folders for data, figures, and tables. It took a minute for me to realize that the actual R code for each tutorial was in the main folder, not within each sub-folder. This may also be confusing to readers. Additionally, each of the 4 tutorials, along with their sub-tutorials are provided in separate .Rmd files. There were several download steps needed to get all the information. To me, this made the user experience tedious, and made the tutorials seem disjointed from one another. The authors might consider a more comprehensive single Rmd file that is neatly organized with an interactive table of contents (in the knitted html file). To me, this would make the experience much more enjoyable and useful for readers. Given that this is a tutorial, it would also be nice if the the RMarkdowns were somewhat more standalone e.g., by including a brief description of what a life table is and how to interpret the plot.

Response: We explain the location of the files and the folder structure in the README file. In addition we make it clear that a single download step is possible by either cloning the project, or downloading the ZIP file. We also included a brief description of a life table and the definitions of the discrete-time functions in the R Markdown file called Tutorial_1a.Rmd. At this point, we also prefer a modular approach to the tutorials rather than combining them. We appreciate that not everyone might share this preference, but we think there are many benefits to a modular workflow in this instance.

Contribution of this tutorial: It wasn't clear to me as a reader what the unique contribution is for this tutorial and what gaps in the literature it fills beyond existing texts and tutorials. For example, is this a tutorial that is meant to specifically focus on experimental psychology cognitive task data – and that this

> type of data has special characteristics/associated research questions that necessitate their own stand-alone tutorial? If so, it may be helpful to give more detail on what these specific characteristics tend to be that may make it different from the many other types of psychology data. Or, is this a tutorial on discrete time survival analysis that is meant to be broadly applicable to psychological scientists, and the cognitive reaction time data is just for illustrative purposes. If so, then much less emphasis should be placed on the specific data type and more emphasis should be placed more broadly on types of psychological data/ research questions this method could be applied to. I think the distinction makes a big difference in the entire framing of the tutorial, and choosing one or the other will make the tutorial content much easier for readers to absorb.

Response: This tutorial is meant to specifically focus on experimental psychology RT data and this type of data has special characteristics that necessitate their own stand-alone tutorial. Therefore, we added visualizations of the types of time-to-event data that are obtained in typical RT tasks (detection, discrimination, bistable perception) in section A of the Supplemental Material.

> On page 7, it states that there are not tutorials aimed at psychological time-to-event data. Yet, there are at least a few existing empirical article examples with accompanying tutorials in psychology (listed below), and arguably some of the popular textbooks are also geared toward psychological scientists (e.g., Singer & Willett). Similarly, many of the textbooks contain code to accompany their text. Regardless, even if those are not counted, I still think it will be beneficial for readers if the authors can make a clear case for what is unique about psychological time-to-event data that necessitates a stand alone tutorial/ what gaps in the existing textbooks/empirical articles/tutorials the present tutorial serves to fill.

Response: In the revised ms. we frame the unique contribution that this works adds on page 6: "...we are not aware of any tutorials that are aimed specifically at psychological RT (+ accuracy) data, and which provide worked examples of the key data processing and Bayesian multilevel regression modelling steps."

> Lougheed, J. P., Benson, L., Cole, P. M., & Ram, N. (2019). Multilevel survival analysis: Studying the timing of children's recurring behaviors. Developmental Psychology, 55(1), 53.

> Elmer, T., van Duijn, M. A., Ram, N., & Bringmann, L. F. (2023). Modeling categorical time-to-event data: The example of social interaction dynamics captured with event-contingent experience sampling methods. Psychological Methods.

> Mills, M. (2011). The fundamentals of survival and event history analysis. Introducing Survival Analysis and Event History Analysis. London: SAGE Publications, 1-17.

Response: We included these references in the revised ms. on page 6.

> Organization of the tutorial:The current version of the tutorial is much longer than a typical AMPPStutorial. There are a few possibilities for dealing with this. First, it may be that the tutorial could be split into 2, such as splitting into tutorials 1-3 and tutorial 4 on its own. Another possibility is that the focus could be primarily on the Bayesian models, and then leave the frequentist models for either a supplement, or to only live in the RMarkdown. This possibility may also be beneficial due to the model convergence issues with at least one of the frequentist models (which is common with multilevel data and why many psychological scientists are moving to the Bayesian framework). A third possibility is that the overall tutorial text could be greatly streamlined, with major cuts to both the introduction and the discussion sections.

Response: We focus primarily on the Bayesian models in the revised ms., and keep Tutorial 4. We also made major cuts to both the introduction and the discussion sections to shorten the ms.

> Clarity and accuracy: -There are several instances where jargon terms and key concepts pertinent to survival analysis need to be defined and/or described in slightly more detail than what is currently provided. For example, clearer descriptions of censoring and risk sets (plus what it means to be at risk) should be given in the introduction.

Response: We introduce the definition of right-censoring on page 11 of the revised ms. We discuss what it means to be at risk in section A of the Supplemental Material.

> Additionally, many key details, such as the EHA equations are relegated to supplementary materials. In my opinion, any key information that makes the tutorial more accessible and precise should not be in supplementary materials.

Response: We placed the equations for h(t) and ca(t) in Figure 1 of the revised ms. Beyond Figure 1, we still choose to keep any further reference of equations to the supplemental material (section B). This is because we feel that one barrier to uptake of these methods is an aversion to equations. And by placing them in the supplemental material we feel that we strike a nice balance between non-technical accessibility and a fullness of mathemathical explanation for key ideas and concepts that remain easily accessible to interested readers.

> There are many abbreviations that were difficult to follow. EHA is arguably standard, but SAT (for speed/accuracy trade-off) is less familiar, and may therefore benefit from not being an abbreviation. Similarly referring to hazard function as h(t), conditional accuracy function as c(t) was less clear than just saying hazard function and conditional accuracy function.

Response: Although SAT is rather standard in cognitive experimental psychology, we tried to limit the use of abbreviations in the main body of the revised ms.

> It would be helpful if the research questions that can be addressed by the methods presented are clearly presented in each section. I think that currently they are getting lost in the tutorial.

Response: We now discuss general research questions on page 10 of the revised ms.

We did not add more information about possible research questions in each section of Tutorial 2a, because we feel that (a) research questions are highly dependent on the field of study (perception, behavioral control, memory), (b) this would make the tutorial even longer, and (c) we mention in the paper that we do not discuss **why** you might perform EHA. In other words, we think that the knowledge of experimental psychologists on regression modeling should be large enough for them to link the model structure to possible research questions, once they have read the paper and worked through the code examples.

> There are several opportunities to describe the method more precisely. For example, the discrete time hazard function gives a 'conditional' probability (pg. 9).

Response: Indeed, this was a big error on our part, which we corrected in the revised ms. Thank you for pointing this out.

> Additionally, it may be helpful to use more standard notations such as h(t_ij) to indicate the conditional probability that the event occurs in bin j for individual i, given that the individual had not already experienced the event in a previous time period (lines 159-160, pg. 9).

Response: We added indexes for individual and trial in the regression model equations in section E of the supplemental material. Furthermore, our new description on pages 8 and 9, together with the new section A in the supplemental material, should make this issue clear for the reader.

> Similarly, the survival probability is the probability that the event occurs after bin j, *given that the individual lhas survived up until bin j.*

Response: We define S(t) = P(RT > t) now as "the probability that the event does not occur before the endpoint of bin t" in section B of the Supplemental Material.

> Throughout the tutorial, reference to individuals in the notation is missing.

Response: We added reference to individuals and trials (repeated measures) in the notation of the model equations in section E of the Supplemental Material. Furthermore, we feel that the information provided by our new description on pages 8 and 9, together with the new section A in the supplemental material, is sufficient for understanding what we mean without including a reference to individuals (and trials) in the main text.

> Another inaccuracy is that in the discrete time framework, it is not an instantaneous likelihood (line 161, pg. 9).

Response: Indeed, only continuous-time hazard can be called an instantaneous likelihood. We replaced the term "likelihood" with "risk".

> Introduction: -Section 1.1 – I had a difficult time understanding this section as the opening to the tutorial, and the last paragraph seemed out of place/abrupt. felt too vague to be understandable as the opening to the tutorial.

Response: We rewrote this section.

> Given that the speed-accuracy tradeoff analysis is described as central to this tutorial, it may benefit from being introduced earlier than pg. 7.

Response: We introduce conditional accuracy functions already in Figure 1 of the revised ms.

> It seems that comparing and contrasting Bayesian and frequentist methods may be outside the scope of a tutorial (pg. 7)

Response: We agree and focus now on the Bayesian models in the revised ms.

> As mentioned above, tutorial 4 – while related – could potentially be its own stand-alone tutorial.

Response: We decided to keep tutorial 4, as we feel that planning future studies is an important and difficult step to take in these kinds of studies, which we wanted to keep close to the rest of the material for ease of understanding a reasonably complete workflow.

> Censoring is a key component of survival analysis that is missing from the introduction

Response: We introduce the concept of right-censoring on page 11 of the revised ms.

> It might be useful to indicate early on the distinctions between multilevel data, single events, and recurring events – and that this particular tutorial is only focused on single events.

Response: We explicitly mention "single event" on page 6 of the revised ms. Also, in section A of the Supplemental Material we mention that this Tutorial does not cover recurrent events, and we visualize the different types of time-to-event data that are obtained in typical RT tasks (detection, discrimination, bistable perception tasks). This should make it clear to the reader that we are dealing with single events that are repeatedly measured within the same individual.

> It's not clear why a discrete time framework is used given the granular temporal resolution of reaction time data – this would also mean that the data would not need to be separated into bins. Most textbooks (e.g., Allison's) often indicate discrete time methods are common for temporal scales such as months, years, or decades (not milliseconds). Clearer justification for the discrete time framework and separating the data into bins would be helpful.

Response: We address our choice for discrete-time methods on pages 8-9 of the revised ms.:

"... the definition of hazard and the type of models employed depend on whether one is using continuous or discrete time units. As a lab, and mainly for practical reasons, we have much more experience using discrete-time EHA, and that is the approach that we describe and focus on in this paper. This choice may seem counter-intuitive, given that RT is typically treated as a continuous variable. However, continuous forms of EHA require much more data to estimate the continuous-time hazard (rate) function well (Bloxom, 1984; Luce, 1991; Van Zandt, 2000). Thus, by trading a bit of temporal resolution for a lower number of trials, discrete-time methods seem ideal for dealing with typical psychological RT data sets for which there are less than ~200 trials per condition per participant (Panis, Schmidt, et al., 2020). Moreover, as indicated by Allison (2010), learning discrete-time EHA methods first will help in learning continuous-time methods, so it seems like a good starting point."

> Clearer description of the data would be helpful for readers (as well as only having this description in one place in the tutorial to be more concise). For example, on pg. 10 it is indicated that 200 trials of 1 experimental condition are simulated. Can the event occur at any time in those 200 trials? If the event occurs in trial 150, do the remaining 50 trials still occur, or does the time series end for that person? How are those extra trials after an event occurred handled in the data wrangling and analysis. Can events occur more than once?

Response: We think that there is a misunderstanding of what "trial" means. We now provide a clear description of the meaning of "trial" and of the various types of time-to-data obtained in typical RT tasks in section A of the Supplemental Material.

> Section 2.2 (pg. 10, lines 168-169), if describing what 'statisticians and mathematical psychologists say', it would be helpful to provide supporting references.

Response: We provide supporting references on page 10 of the revised ms.

> The numbered list on pg. 11 contained jargon statements without explanations. Suggest to simplify and keep the focus on the data you are using and specific research questions you are using survival analysis as a tool to address.

Response: We removed the numbered list in the revised ms.

> In section 2.3, suggest to lead with the data and research questions of interest and then show how features of survival analysis make it a good method to examine the research questions. I also felt that up until now, the tutorial has been very confusing and hard to follow. It could be that starting the tutorial at section 2.3 (with a few extra details filled in), may be an alternative.

Response: Section 2.3 has been rewritten. The new Figure 1 in the tutorial directly focuses on the comparison between analysing means vs. distributional shapes. We also discuss general research questions on page 10 of the revised ms.

> Again, section 2.3.1 (the majority of pg. 12) was unclear and difficult to follow for someone who already has knowledge of survival analysis, but not much familiarity with cognitive process/reaction time data.

Response: Section 2.3.1 has been removed.

> From what is written, I did not have a clear sense of how temporal states of cognitive processes and theory development could occur (pg. 14, lines 231-29). These ideas and the research questions are scattered in many places throughout the tutorial, and are not very clear/straightforward. It may be helpful to locate them in just one place in the tutorial and provide a clearer description to guide readers. Similarly, the statement on pg. 14 (lines 238-239) felt rather strong without any references, and also out of scope for a tutorial.

Response: We explain the idea of temporal states of cognitive processes now on pages 3-4 of the revised ms. and illustrate it in Figure 1.

- It wasn't clear how section 3.1 added value to the tutorial

Response: Section 3 has been removed.

> Section 3.2 – it would be helpful if there was clarification in the text on when it would not be possible to include varying intercepts and slopes (i.e., at least clarifying that random intercepts are a requirement, but sometimes random slopes are removed based on model convergence issues – however, this also begs the question of whether a model should even be presented if it does not converge when it includes random slopes. . .)

Response: We do not discuss frequentist approaches in the main text anymore (to save space) and only mention that the reader can consult Tutorials 3a and 3b. We agree that a model should not be presented if it does not converge when it includes random slopes. However, to save space, and because this is an active area of research, we do not attempt to clarify when it would not be possible to include varying intercepts and/or slopes. We do mention in Tutorials 3a and 3b that it is common for models to fail to converge in lme4 when they have a reasonably complex random effects structure, and we provide a reference.

- Additionally in section 3.2, the text on pg. 16 is vague and does not give much detail to readers.

Response: Section 3 has been removed.

> Overall, the introduction spans across 16 pages, which feels long even for a standard empirical article, let alone a tutorial. As currently written, I think many readers will get lost before even getting to the tutorials section.

Response: We agree, and restructured and shortened the introduction.

> Tutorials section: -Rather than only using h(t), and ca(t), suggest to use hazard function and conditional accuracy function along with their mathematical notation. Without having presented any equations, using terms such as h(t) and ca(t) seems odd.

Response: We try to limit the number of abbreviations used in the main body of the revised ms.

- Similarly, P(t) is never defined for readers.

Response: P(t) was and still is defined only in the Supplementary Material (section B in the revised version).

- Section 4.1.1. – it would be helpful to clarify that the risk set also pertains to individuals (not just bins)

Response: We clarify this in section A of the Supplemental Material.

> The connection between trials and bins could be made clearer (this is in reference to the text on pg. 18, but I think the clarity would need to come earlier on in the tutorial when describing the data)

Response: This connection should be clear now based on section A of the Supplemental Material, and based on the text on pages 8 and 9.

> Comprehension of the text on pg. 18 would be enhanced if both right and left censoring had been clearly introduced previously. Although left censoring is probably unlikely in the reaction time data used in the tutorial – if it is the case that this tutorial is meant for broader data types in psychology, left censoring would certainly be relevant to many other use cases.

Response: We discuss various types of censoring (right, left, random informative, random uninformative) in section A of the Supplemental Material.

> Section 4.1.2 was a clearer data description than had been presented up until that point in the tutorial. Again, it is suggested that the manuscript/tutorial could be dramatically reduced in length if information was more concisely but descriptively written in a single place (e.g., data descriptions).

Response: We provide the information on data descriptions in a single place in the revised ms.

> Suggest to be more precise with the language on pg. 21 (line 355) – which data are nested within participants? The trial data? The binned data?

Response: We specify "person-trial data" in the revised ms on page 17.

> pg. 22 (line 358) the way that censoring is described here is confusing.

Response: We agree and adjusted the description (see pages 10-11 of the revised ms.).

> pg. 22 discusses warning messages and lets readers know these can be ignored but does not give details as to why. Since this is a didactic tutorial, it may be helpful to have at least some explanation as to why the warning can be ignored.

Response: On line 362 it was stated that "the warning messages are generated because some bins have no hazard and ca(t) estimates, and no error bars. They can thus safely be ignored". We now removed these messages from the paper and mention them now only in the tutorial itself.

> pg. 22 – since this is a tutorial, it might also be helpful to unpack in more detail the section about identifying individuals who may be guessing etc – especially since these are very important considerations.

Response: We have no more detail to share on this issue. On page 18 of the revised ms. we now write "In general, it is important to visually inspect the functions first for each participant, in order to identify individuals that may not be following task instructions (e.g., a flat conditional accuracy function at .5 indicates that someone is just guessing), outlying individuals, and/or different groups with qualitatively different behavior.".

> section 4.3.1 – (line 451) it would be helpful if more details is given on how to set the analysis time window. What counts as "enough data"? Given that this is a tutorial, these types of detail are critical for readers to be able to implement the method in their own work

Response: We changed this sentence into: "... because the first few bins typically contain no responses, one has to select an analysis time window, i.e., a contiguous set of bins for which there is data for each participant." on page 23 of the revised ms.

> section 4.3.1 (line 453) – more detail is needed on the cloglog link function and how it differs from the more commonly known about logit link.

Response: A visual comparison between the cloglog and logit link functions is provided in section D of the Supplemental Material.

> section 4.3.1 (line 456) – what goes into the decision about whether to treat time as a categorical or continuous predictor?

Response: We added the following information on page 23 of the revised ms.: "Third, one has to choose whether to treat TIME (i.e., the time bin index t) as a categorical or continuous predictor (see also section E of the Supplemental Material). For example, when you want to know if cloglog-hazard is changing linearly or quadraticly over time, you should treat TIME as a continuous predictor. When you are only interested in the effect of covariates on hazard, you can treat TIME as a categorical predictor (i.e., fit an intercept for each bin), in which case you can choose between reference coding and index coding."

> section 4.3.3. – it might be helpful to give a brief indicator up front to readers about the rationale/purpose of fitting both M0i and M1i, especially given that no interpretation is given for M0i. Similarly, it would be helpful to give a bit more detail on the purpose of the comparison of M0i and M1i (given that this is a didactic tutorial).

Response: We mention now that model M0i is a baseline or reference model, to be compared with model M1i (see page 25 of the revised ms.). We also mention that we only report 2 models in the paper from a larger set of models that are fitted in the file Tutorial_2a.Rmd.

> section 4.3.6. – same thing here – since this is a tutorial, at least brief details on what the purpose of this section is would be helpful. It might also be helpful here to define what is meant by credible interval so that more novice readers don't confuse it with a confidence interval.

Response: We added the purpose ("To make causal inferences...") and define a credible interval on page 27 of the revised ms.

> The term SOA is used on pg. 34, but had never previously been defined or described.

Response: Stimulus-onset-asynchrony (SOA) is a well-known concept in experimental psychology. We now only use the term SOA in the Supplemental Material.

> The text on pg. 35 (lines 567-574) seems too in-depth for a tutorial

Response: We removed this text.

> pg. 36 – it would be helpful to give more detail on how to interpret the logit-ca scale

Response: We decided not to include more detail because logistic regression is very familiar to experimental psychologists.

> section 4.7 – as previously mentioned, it is possible that tutorial 4 could be a stand alone tutorial. If it's not though, more details are needed up front to tell readers what the purpose is and why this is useful.

Response: We decided to keep tutorial 4, and discuss its usefulness on pages 40-41 of the revised ms.

> Discussion section: The Discussion section would benefit from being shortened (AMPPS guidelines indicate discussions should be a brief summary of their contents rather than a general discussion).

Response: We shortened the Discussion section.

> is 'how long' a component of survival analysis – or is it more accurate to say 'whether and when psychological states occur'?

Response: It is more accurate to say 'whether and when' psychological states occur. However, based on those results, we can deduce information about how long a psychological state lasts within a trial. Nevertheless, we replaced this with "how effects evolve with increasing waiting time".

> pg. 47 (lines 753-761) – I had a difficult time understanding this section

Response: We rewrote this section.

> section 5.5 – unclear that the multi-state model is actually an example where a multi-state model would be relevant... (the conceptual example given seems like it would be more linked to either a hidden Markov model or recurring event survival model)

Response: We removed section 5.5.

> Figures: -Figure 1 is difficult to follow. It is unclear how the means are being calculated. For example, it seems like in examples 2 and 3, the means between conditions should be different. I was also confused in example 4 because the difference in probability between the two conditions in time bin 2 seems tiny. I was also confused because it still looked like there was a big difference in probability between the two conditions at bin 5. Was bin 6 by chance what was meant instead of bin 5? It also still looked like there were differences later on such as in bin 7. Example 5 also left me confused. It looked to me like the effect started at bin 4 and ended at bin 8. Clarification (and perhaps simplifying) this figure would be useful.

Response: We created a new, simpler Figure 1.

> The description for figure 2 is unclear, particularly with respect to the contrasting knowledge that can be gained from the survival function and hazard function

Response: Figure 2 has been removed.