

1 Event History Analysis for psychological time-to-event data: A tutorial in R with examples  
2 in Bayesian and frequentist workflows

3 Sven Panis<sup>1</sup> & Richard Ramsey<sup>1</sup>

4 <sup>1</sup> ETH Zürich

5 Author Note

6 Neural Control of Movement lab, Department of Health Sciences and Technology  
7 (D-HEST). Social Brain Sciences lab, Department of Humanities, Social and Political  
8 Sciences (D-GESS).

9 Correspondence concerning this article should be addressed to Sven Panis, ETH  
10 GLC, room G16.2, Gloriastrasse 37/39, 8006 Zürich. E-mail: sven.panis@hest.ethz.ch

11

## Abstract

12 Time-to-event data such as response times and saccade latencies form a cornerstone of  
13 experimental psychology, and have had a widespread impact on our understanding of  
14 human cognition. However, the orthodox method for analyzing such data – comparing  
15 means between conditions – is known to conceal valuable information about the timeline of  
16 psychological effects, such as their onset time and duration. The ability to reveal  
17 finer-grained, “temporal states” of cognitive processes can have important consequences for  
18 theory development by qualitatively changing the key inferences that are drawn from  
19 psychological data. Luckily, well-established analytical approaches, such as event history  
20 analysis (EHA), are able to evaluate the detailed shape of time-to-event distributions, and  
21 thus characterize the time course of psychological states. One barrier to wider use of EHA,  
22 however, is that the analytical workflow is typically more time-consuming and complex  
23 than orthodox approaches. To help achieve broader uptake of EHA, in this paper we  
24 outline a set of tutorials that detail one distributional method known as discrete-time  
25 EHA. We touch upon several key aspects of the workflow, such as how to process raw data  
26 and specify regression models, and we also consider the implications for experimental  
27 design, as well as how to manage inter-individual differences. We finish the article by  
28 considering the benefits of the approach for understanding psychological states, as well as  
29 the limitations and future directions of this work. Finally, the project is written in R and  
30 freely available, which means the approach can easily be adapted to other data sets.

31       *Keywords:* response times, event history analysis, Bayesian multilevel regression  
32 models, experimental psychology, cognitive psychology

33 Word count: 11664 (body) + 1593 (references) + 2394 (supplemental material)

34

## 1. Introduction

### 35 1.1 Motivation and background context: Comparing means versus 36 distributional shapes

37 In experimental psychology, it is standard practice to analyse response times (RTs),  
38 saccade latencies, and fixation durations by calculating average performance across a series  
39 of trials. Such comparisons between means have been the workhorse of experimental  
40 psychology over the last century, and have had a substantial impact on theory development  
41 as well as our understanding of the structure of cognition and brain function. However,  
42 differences in mean RT conceal important pieces of information, such as when an  
43 experimental effect starts, how it evolves with increasing waiting time, and whether its  
44 onset is time-locked to other events (Panis, 2020; Panis, Moran, Wolkersdorfer, & Schmidt,  
45 2020; Panis & Schmidt, 2016, 2022; Panis, Torfs, Gillebert, Wagemans, & Humphreys,  
46 2017; Panis & Wagemans, 2009; Wolkersdorfer, Panis, & Schmidt, 2020). Such information  
47 is useful not only for the interpretation of experimental effects under investigation, but also  
48 for cognitive psychophysiology and computational model selection (Panis, Schmidt,  
49 Wolkersdorfer, & Schmidt, 2020).

50 As a simple illustration, Figure 1 shows how comparing means between two  
51 conditions conceal the shapes of the underlying RT and accuracy distributions. We  
52 simulated a RT + accuracy data set for a single subject who performed 200 trials (i.e.,  
53 repeated measurements) in each of two conditions. For example, while this subject is 71 ms  
54 faster on average in condition 1 (481 ms) compared to condition 2 (552 ms), the  
55 corresponding hazard functions of response occurrence show that the effect starts in time  
56 period (400,500] or bin t=5, and is present in three consecutive time bins (i.e., for 300 ms).  
57 Similarly, while this subject makes less errors in condition 1 (86% accuracy) compared to  
58 condition 2 (64% accuracy), the conditional accuracy functions show that (a) the effect is  
59 present only for responses emitted before 400 ms, (b) erroneous responses in condition 1

60 are confined to a single time bin, and (c) the observed conditional accuracies (0, 1, 0.51,  
61 0.48) are never even close to the mean accuracies.

62 Why does this matter for research in psychology? Compared to the aggregation of  
63 data across trials, a distributional approach offers the possibility to reveal the time course  
64 of psychological states. For example, Figure 1B shows a first state (up to 400 ms after  
65 target onset) for which the early upswing in hazard is equal for both conditions, and the  
66 emitted responses are always correct in condition 1 and always incorrect in condition 2. In  
67 a second state (400 to 500 ms), hazard is higher in condition 1, and conditional accuracies  
68 are close to .5 in both conditions. In a third state (>500 ms), the effect disappears in  
69 hazard, and all conditional accuracies are equal to 1.

70 Note that the distributional shapes are inspired by published results from interference  
71 paradigms such as priming and cueing tasks (refs). For example, if the target stimulus at  
72 time zero is preceded by a prime stimulus that can be congruent (condition 1) or  
73 incongruent (condition 2) to the target, then the distributions show that (a) the fastest  
74 responses (< 400 ms) are triggered exclusively by the prime, (b) between 400 and 500 ms,  
75 the prime-triggered response tendency is actively inhibited (leading to conditional  
76 accuracies close to .5), and (c) the slowest responses (>500 ms) are controlled by the target  
77 stimulus.

78 For many psychological questions, such “temporal states” information can be  
79 theoretically meaningful by leading to more fine-grained understanding of psychological  
80 processes, by adding a relatively under-used dimension – the passage of time – to the theory  
81 building toolkit. Thus, a distributional approach permits different kinds of questions to be  
82 asked, different inferences to be made, and it holds the potential to better discriminate  
83 between different theoretical accounts of psychological and/or brain-based processes.

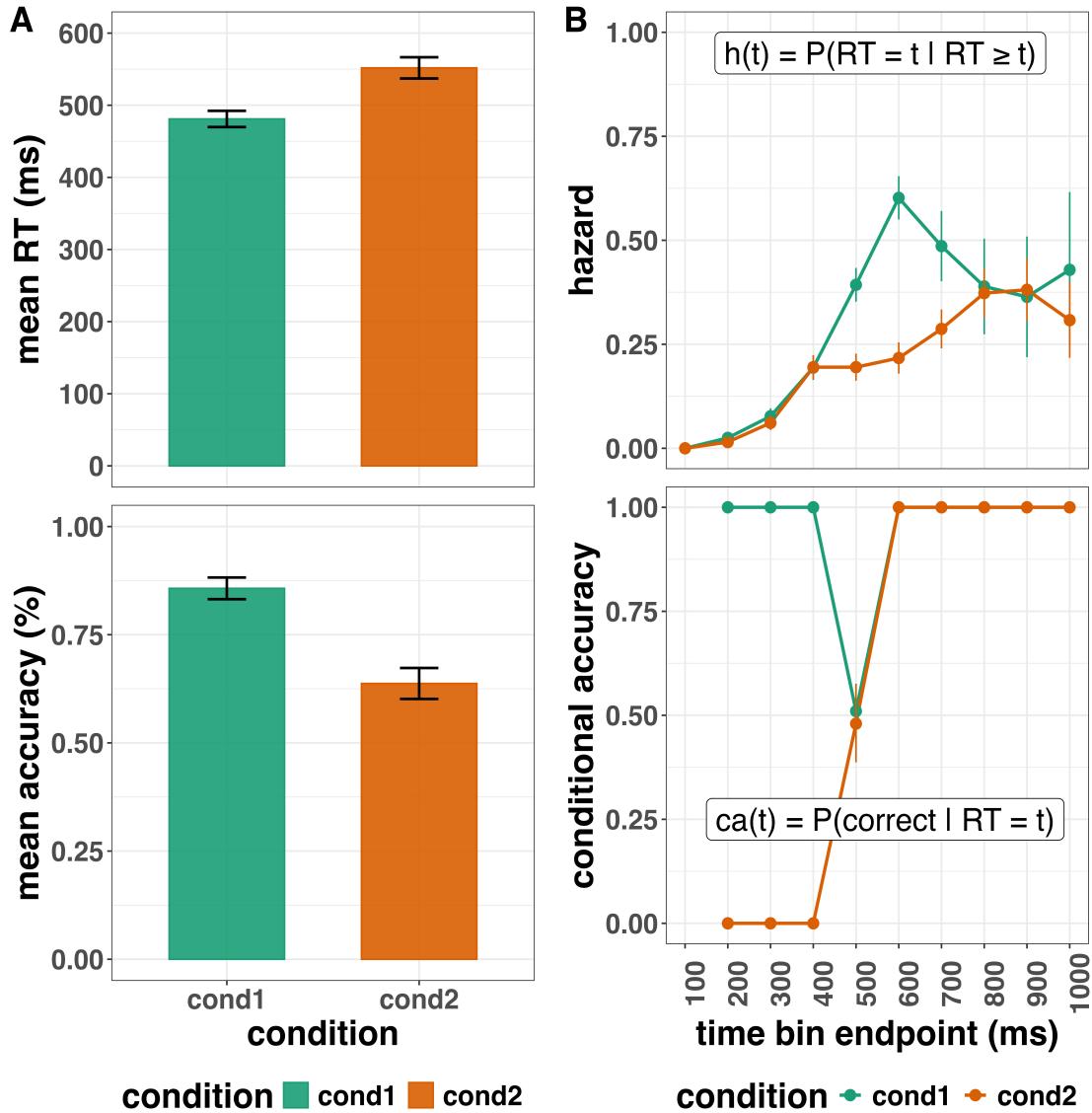


Figure 1. Mean performance versus distributional analyses. (A) The mean RT (top) and overall accuracy (bottom) for two conditions are plotted. (B) The discrete-time hazard functions (top) and conditional accuracy functions (bottom) are plotted for the same data. The first second after target stimulus onset (time zero) is divided in ten bins of 100 ms. The first bin is (0,100], the last bin is (900,1000]. Note that the hazard and conditional accuracy estimates are plotted at the endpoint of each time bin. The definitions of discrete-time hazard and conditional accuracy are further explained in section 2. Error bars represent +/- 1 standard error of the mean (A) or proportion (B).

## 84 1.2 Aims and structure of the paper

85 In this paper, we focus on a distributional method for time-to-event data known as  
86 discrete-time Event History Analysis (EHA), a.k.a. survival analysis, hazard analysis,  
87 duration analysis, failure time analysis, and transition analysis (Singer & Willett, 2003).

88 Our ultimate goal is twofold: first, we want to convince readers of the many benefits of  
89 using EHA when dealing with psychological RT data, and second, we want to provide a set  
90 of practical tutorials, which provide step-by-step instructions on how you actually perform  
91 a discrete-time EHA on RT data, as well as a complementary discrete-time speed-accuracy  
92 tradeoff (SAT) analysis on timed accuracy data in case of choice RT data.

93 Even though EHA is a widely used statistical tool and there already exist many  
94 excellent reviews (e.g., Blossfeld & Rohwer, 2002; Box-Steffensmeier, 2004; Hosmer,  
95 Lemeshow, & May, 2011; Teachman, 1983) and tutorials (e.g., Allison, 2010; Landes,  
96 Engelhardt, & Pelletier, 2020), we are not aware of any tutorials that are aimed specifically  
97 at psychological RT (+ accuracy) data, and which provide worked examples of the key  
98 data processing and Bayesian multilevel regression modelling steps. From a historical  
99 perspective, it is worth noting that the development of analytical tools that can estimate or  
100 predict whether and when events will occur is not a new endeavour. Indeed, hundreds of  
101 years ago, analytical methods were developed to predict the duration of time until people  
102 died (e.g., Halley, 1693; Makeham, 1860). The same logic can be applied to psychological  
103 time-to-event data, as previously demonstrated [Panis, Schmidt, et al. (2020); XXXXX].

104 We first provide a brief overview of EHA to orient the reader to the basic concepts  
105 that we will use throughout the paper. However, this will remain relatively short, as this  
106 has been covered in detail before (Allison, 1982, 2010; Singer & Willett, 2003). Indeed, our  
107 primary aim here is to introduce the set of tutorials, which explain **how** to do such  
108 analyses, rather than repeat in any detail **why** you may do them.

109 We then provide seven different tutorials, which are written in the R programming

language and publicly available on our Github page ([https://github.com/sven-panis/Tutorial\\_Event\\_History\\_Analysis](https://github.com/sven-panis/Tutorial_Event_History_Analysis)), along with all of the other code and material associated with the project. The tutorials provide hands-on, concrete examples of key parts of the analytical process, so that others can apply EHA to their own time-to-event data measured in RT tasks. Each tutorial is provided as an RMarkdown file, so that others can download and adapt the code to fit their own purposes. Additionally, each tutorial is made available as a .html file, so that it can be viewed by any web browser, and thus available to those that do not use R. Finally, the manuscript itself is written in R using the *papaja* package (Aust & Barth, 2024a), which makes it computationally reproducible, in terms of the underlying data and figures.

120 .

## 121 2. A brief introduction to event history analysis

122 EHA is a class of statistical methods to study the occurrence and timing of events, such as disease onset, marriages, arrests, and job terminations (Allison, 2010). To apply EHA, one must be able to:

- 125 1. define an event of interest that represents a qualitative change - a transition from one  
126 discrete state to another - that can be situated in time (e.g., a button press, a  
127 saccade onset, a fixation offset, etc.);
- 128 2. define time point zero (e.g., target stimulus onset, fixation onset, etc.);
- 129 3. measure the passage of time between time point zero and event occurrence in discrete  
130 or continuous time units.

### 131 2.1 Single, repeatable, and recurrent events

132 While people can die only once, in experimental RT tasks the events of interest are  
133 typically repeatable. For example, in the target-present condition of a one-button detection

task the participant is presented in each trial with a faint target stimulus whose presence (s)he has to detect by pressing a button within a certain time window (e.g., the first second after target onset). In EHA parlance, the single event of interest is a button press response, *time zero* is defined as target display onset, the *observation period* is 1 second long in each trial or repeated measurement, in each trial the participant is *at risk* for response occurrence as long as the response has not occurred yet, and the individual always starts in an “idle” state in each trial and *transitions* to a “detected” state when a response occurs.

In a two-button discrimination task, the participant is presented in each trial with a target stimulus that (s)he has to categorize by pressing one of two buttons within a certain time window. In the world of EHA, this is known as a “competing risks” situation, because in each trial the participant can transition from an idle state to either a “correct response” state or an “incorrect response” state.

In a bistable perception task, the participant is looking at an ambiguous stimulus (e.g., the duck-rabbit illusion, the Necker cube) for two minutes, for example, and asked to press a button each time when her/his perception switches from one possible interpretation to the other possible interpretation. In this task, there are two events (percept A switches to percept B, percept B switches to percept A) that can recur within the same observation period of two minutes, so that the individual transitions back and forth between two states.

In section A of the Supplemental Material we visualize the types of time-to-event data that are obtained in these typical RT tasks (detection, discrimination or categorization, bistable perception). Note that we do not analyse recurrent events in this tutorial. More information about recurrent events analysis can be found in REF and REF...

## 2.2 Right censoring versus data trimming

What do you do with trials in which no response occurs during the observation period? EHA treats such trials as *right-censored* observations on the variable RT, because

159 all we know is that RT is greater than some value. Right-censoring is a type of missing  
160 data problem and a nearly universal feature of survival data including RT data. For  
161 example, in the one-button detection task example from above, all trials have a *censoring*  
162 *time* of 1 second, but some trials result in observed event times (those with a RT below 1  
163 second), while the other trials result in response times that are right-censored at 1 second.

164 EHA can deal in a straight-forward fashion with right-censored time-to-event data.  
165 In contrast, experimental psychologists are used to either (a) use a response deadline and  
166 discard all trials without a response, or (b) wait in each trial until a response occurs and  
167 then apply data trimming techniques, i.e., discarding too short or too long RTs before  
168 calculating a mean RT (REF). Discarding data can introduce biases, however.

### 169 **2.3 Discrete vs continuous time units**

170 All man-made measurements of duration are discrete in nature. However, when the  
171 temporal resolution is high relative to the duration of the observation window, researchers  
172 typically treat time as continuous. RT data can thus be analysed using continuous-time  
173 EHA methods which use the exact event times, including parametric models (e.g., an  
174 exponential hazard model, a Weibull hazard model, a lognormal hazard model) and the  
175 popular Cox regression model () .

176 However, in this tutorial we focus on discrete-time methods for three reasons: First,  
177 we are interested in studying the shape of the hazard function (Cox regression ignores this  
178 and only tests the effects of covariates); Second, empirical hazard and conditional accuracy  
179 functions from certain RT tasks (e.g., interference tasks; Figure 1B) can show abrupt  
180 changes in their shape (parametric methods assume smooth distributions), and the shape  
181 of the hazard function in many experimental tasks is still unknown (parametric methods  
182 assume well-defined probability distributions); Third, in discrete time, hazard is simply  
183 defined as a conditional probability (see 2.4) and we can apply logistic regression modeling

<sup>184</sup> with which most experimental psychologists are already familiar.

<sup>185</sup> In sum, due to their simplicity and flexibility, we believe that discrete-time methods  
<sup>186</sup> are a good starting point for experimental psychologists that want to abandon ANOVA  
<sup>187</sup> and learn to apply EHA, even though continuous-time methods might be more suited in  
<sup>188</sup> certain situations.

<sup>189</sup> **2.4 Discrete-time hazard functions and conditional accuracy functions**

<sup>190</sup> After dividing time in discrete, contiguous time bins indexed by  $t$  (e.g.,  $t = 1:10$  time  
<sup>191</sup> bins; Figure 1B), let  $RT$  be a discrete random variable denoting the *rank* of the time bin in  
<sup>192</sup> which a particular person's response occurs in a particular experimental condition. For  
<sup>193</sup> example, the detection response in trial 1 might occur at 546 ms and it would be in time  
<sup>194</sup> bin 6 (any RTs from 501 ms to 600 ms). Thus, the RT data are interval-censored, because  
<sup>195</sup> we only use the information that  $a < RT \leq b$  when the response occurs in time bin  $(a,b]$ .

<sup>196</sup> While experimental psychologists are familiar with the cumulative distribution  
<sup>197</sup> function or  $F(t) = P(RT \leq t)$  and the probability mass function or  $P(t) = P(RT = t)$ ,  
<sup>198</sup> discrete-time EHA focuses on the discrete-time hazard function of event occurrence:

$$\sup_{199} h(t) = P(RT = t | RT \geq t) \quad (1)$$

<sup>200</sup> and the discrete-time survivor function:

$$\sup_{201} S(t) = P(RT > t) = 1 - F(t) = [1-h(t)].[1-h(t-1)].[1-h(t-2)] \dots [1-h(1)] \quad (2)$$

<sup>202</sup> The discrete-time hazard function gives you, for each time bin, the conditional  
<sup>203</sup> probability that the event occurs (sometime) in bin  $t$ , given that the event does not occur  
<sup>204</sup> in previous bins. In other words, it reflects the instantaneous risk that the response occurs  
<sup>205</sup> in bin  $t$ , given that it has not yet occurred in one of the prior bins. In contrast, the  
<sup>206</sup> discrete-time survivor function cumulates the bin-by-bin risks of event *nonoccurrence* to  
<sup>207</sup> obtain the survival probability, the probability that the event does not occur before the  
<sup>208</sup> endpoint of bin  $t$ . As a result, only the hazard function conveys the risk of event

209 occurrence associated with each bin, and . . . suited for online tracking of performance.. cfr  
 210 mouse cursor movements. . . .

211 For two-choice RT data, the discrete-time hazard function can be extended with the  
 212 discrete-time conditional accuracy function

$$213 \quad ca(t) = P(\text{correct} \mid RT = t) \quad (5)$$

214 which gives you for each bin the probability that a response is correct given that it is  
 215 emitted in time bin  $t$  (Allison, 2010; Kantowitz & Pachella, 2021; Wickelgren, 1977). The  
 216  $ca(t)$  function is also known as the micro-level speed-accuracy tradeoff (SAT) function. We  
 217 refer to this extended (hazard + conditional accuracy) analysis for choice RT data as  
 218 EHA/SAT.

219 As we will illustrate in Tutorials 1a and 1b, performing a descriptive EHA/SAT  
 220 analysis by calculating the sample-based estimates of  $h(t)$ ,  $S(t)$  and  $ca(t)$  for each  
 221 combination of participant and condition requires setting up a *life table*. Definition life  
 222 table. . .

## 223 2.5 Bayesian vs. frequentist approaches to regression

224 To study how the risk of a response, and the accuracy of an emitted response,  
 225 depends on covariates (i.e., explanatory predictor variables) we can estimate regression  
 226 models for hazard and for conditional accuracy, i.e., perform inferential EHA/SAT  
 227 analysis. Such covariates can be constant over within-trial time (e.g., gender, race, trial  
 228 number, block number) or vary with within-trial time (e.g., heart rate, eye gaze position,  
 229 eye pupil dilation). Note that time-varying covariates are not covered in this tutorial.

230 Heterogeneity -> Multilevel survival analysis: Methods, Models and Applications  
 231 Austin 2017 !!

232 fitting problems -> Bayesian

233 **2.6 Number of samples, repeated measures, time bins**

234 In a typical RT data set from a within-subject design, there are N individuals and M  
235 repeated measures or trials per experimental condition. To test process models of  
236 cognition, .. advises to use small-N designs, ... eACH SUBJECT REPLICATION UNIT

237 Power IS A COMPLEX FUNCTION OF .... (REF A, REF B ON POWER WITH  
238 EXPONENTIAL)

239 Number of time bins?

240 **bin width**

241 We recommend several excellent textbooks for a comprehensive background context  
242 to EHA (Allison, 2010; Singer & Willett, 2003) and for a more general introduction to  
243 understanding regression equations (Gelman, Hill, & Vehtari, 2020; Winter, 2019). Our  
244 focus here is not on providing a detailed account of the underlying regression equations,  
245 since this topic has been comprehensively covered many times before. Instead, we want to  
246 provide an intuition regarding how EHA works in general, as well as in the context of  
247 experimental psychology. As such, we only supply regression equations in section D of the  
248 Supplemental Material.

249

#### 4. Tutorials

250 We used R (Version 4.4.0; R Core Team, 2024)<sup>1</sup> for all reported analyses. The  
 251 content of the tutorials, in terms of EHA and multilevel regression modelling, is mainly  
 252 based on Allison (2010), Singer and Willett (2003), McElreath (2020), Heiss (2021), Kurz  
 253 (2023a), and Kurz (2023b).

254 Tutorials 1a and 1b show how to calculate and plot the descriptive statistics of  
 255 EHA/SAT when there are one or two independent variables, respectively. Tutorials 2a and  
 256 2b illustrate how to use Bayesian multilevel modeling to fit hazard and conditional  
 257 accuracy models, respectively. Tutorials 3a and 3b show how to implement, respectively,  
 258 multilevel models for hazard and conditional accuracy in the frequentist framework.  
 259 Additionally, to further simplify the process for other users, the first two tutorials rely on a  
 260 set of our own custom functions that make sub-processes easier to automate, such as data  
 261 wrangling and plotting functions (see section B in the Supplemental Material for a list of  
 262 the custom functions).

---

<sup>1</sup> We, furthermore, used the R-packages *bayesplot* (Version 1.11.1; Gabry, Simpson, Vehtari, Betancourt, & Gelman, 2019), *brms* (Version 2.22.0; Bürkner, 2017, 2018, 2021), *citr* (Version 0.3.2; Aust, 2019), *cmdstanr* (Version 0.8.1.9000; Gabry, Češnovar, Johnson, & Brønner, 2024), *dplyr* (Version 1.1.4; Wickham, François, Henry, Müller, & Vaughan, 2023), *forcats* (Version 1.0.0; Wickham, 2023a), *ggplot2* (Version 3.5.1; Wickham, 2016), *lme4* (Version 1.1.35.5; Bates, Mächler, Bolker, & Walker, 2015), *lubridate* (Version 1.9.3; Grolemund & Wickham, 2011), *Matrix* (Version 1.7.1; Bates, Maechler, & Jagan, 2024), *nlme* (Version 3.1.166; Pinheiro & Bates, 2000), *papaja* (Version 0.1.3; Aust & Barth, 2024b), *patchwork* (Version 1.3.0; Pedersen, 2024), *purrr* (Version 1.0.2; Wickham & Henry, 2023), *RColorBrewer* (Version 1.1.3; Neuwirth, 2022), *Rcpp* (Eddelbuettel & Balamuta, 2018; Version 1.0.13.1; Eddelbuettel & François, 2011), *readr* (Version 2.1.5; Wickham, Hester, & Bryan, 2024), *RJ-2021-048* (Bengtsson, 2021), *rstan* (Version 2.32.6; Stan Development Team, 2024), *standist* (Version 0.0.0.9000; Girard, 2024), *StanHeaders* (Version 2.32.10; Stan Development Team, 2020), *stringr* (Version 1.5.1; Wickham, 2023b), *tibble* (Version 3.2.1; Müller & Wickham, 2023), *tidybayes* (Version 3.0.7; Kay, 2024), *tidyR* (Version 1.3.1; Wickham, Vaughan, & Girlich, 2024), *tidyverse* (Version 2.0.0; Wickham et al., 2019) and *tinylabels* (Version 0.2.4; Barth, 2023).

263 Our list of tutorials is as follows:

- 264 • 1a. Wrangle raw data and calculate descriptive stats for one independent variable
- 265 • 1b. Wrangle raw data and calculate descriptive stats for two independent variables
- 266 • 2a. Bayesian multilevel modeling for  $h(t)$
- 267 • 2b. Bayesian multilevel modeling for  $ca(t)$
- 268 • 3a. Frequentist multilevel modeling for  $h(t)$
- 269 • 3b. Frequentist multilevel modeling for  $ca(t)$
- 270 • 4. Simulation and power analysis for planning experiments

271 **4.1 Tutorial 1a: Calculating descriptive statistics using a life table**

272 **4.1.1 Data wrangling aims.** Our data wrangling procedures serve two related

273 purposes. First, we want to summarise and visualise descriptive statistics that relate to our  
274 main research questions about the time course of psychological processes, using a life table.

275 A life table includes for each time bin, the risk set (i.e., the number of trials that are  
276 event-free at the start of the bin), the number of observed events, and the estimates of  
277  $h(t)$ ,  $S(t)$ ,  $P(t)$ , possibly  $ca(t)$ , and their estimated standard errors (se).

278 Second, we want to produce two different data sets that can each be submitted to

279 different types of inferential modelling approaches. The two types of data structure we  
280 label as ‘person-trial’ data and ‘person-trial-bin’ data. The ‘person-trial’ data (Table 1)  
281 will be familiar to most researchers who record behavioural responses from participants, as  
282 it represents the measured RT and accuracy per trial within an experiment. This data set  
283 is used when fitting conditional accuracy models (Tutorials 2b and 3b).

284 `## Warning in attr(x, "align"): 'xfun::attr()' is deprecated.`

285 `## Use 'xfun::attr2()' instead.`

286 `## See help("Deprecated")`

Table 1

*Data structure for ‘person-trial’ data*

pid	trial	condition	rt	accuracy
1	1	congruent	373.49	1
1	2	incongruent	431.31	1
1	3	congruent	455.43	0
1	4	incongruent	622.41	1
1	5	incongruent	535.98	1
1	6	incongruent	540.08	1
1	7	congruent	511.07	1
1	8	incongruent	444.42	1
1	9	congruent	678.69	1
1	10	congruent	549.79	1

*Note.* The first 10 trials for participant 1 are shown. These data are simulated and for illustrative purposes only.

287 In contrast, the ‘person-trial-bin’ data (Table 2) has a different, more extended  
 288 structure, which indicates in which bin a response occurred, if at all, in each trial.  
 289 Therefore, the ‘person-trial-bin’ data generates a 0 in each bin until an event occurs and  
 290 then it generates a 1 to signal an event has occurred in that bin. This data set is used  
 291 when fitting hazard models (Tutorials 2a and 3a). It is worth pointing out that there is no  
 292 requirement for an event to occur at all (in any bin), as maybe there was no response on  
 293 that trial or the event occurred after the time window of interest. Likewise, when the event  
 294 occurs in bin 1 there would only be one row of data for that trial in the person-trial-bin  
 295 data set.

```

296 ## Warning in attr(x, "align"): 'xfun::attr()' is deprecated.
297 ## Use 'xfun::attr2()' instead.
298 ## See help("Deprecated")

```

Table 2

*Data structure for ‘person-trial-bin’ data*

pid	trial	condition	timebin	event
1	1	congruent	1	0
1	1	congruent	2	0
1	1	congruent	3	0
1	1	congruent	4	1
1	2	incongruent	1	0
1	2	incongruent	2	0
1	2	incongruent	3	0
1	2	incongruent	4	0
1	2	incongruent	5	1

*Note.* The first 2 trials for participant 1 from Table 1 are shown. The width of the time bins is 100 ms. These data are simulated and for illustrative purposes only.

299       **4.1.2 A real data wrangling example.** To illustrate how to quickly set up life  
 300 tables for calculating the descriptive statistics (functions of discrete time), we use a  
 301 published data set on masked response priming from Panis and Schmidt (2016). In their  
 302 first experiment, Panis and Schmidt (2016) presented a double arrow for 94 ms that  
 303 pointed left or right as the target stimulus with an onset at time point zero in each trial.

304 Participants had to indicate the direction in which the double arrow pointed using their  
 305 corresponding index finger, within 800 ms after target onset. Response time and accuracy  
 306 were recorded on each trial. Prime type (blank, congruent, incongruent) and mask type  
 307 were manipulated. Here we focus on the subset of trials in which no mask was presented.  
 308 The 13-ms prime stimulus was a double arrow presented 187 ms before target onset in the  
 309 congruent (same direction as target) and incongruent (opposite direction as target) prime  
 310 conditions.

311 There are several data wrangling steps to be taken. First, we need to load the data  
 312 before we (a) supply required column names, and (b) specify the factor condition with the  
 313 correct levels and labels.

314 The required column names are as follows:

- 315 • “pid”, indicating unique participant IDs;
- 316 • “trial”, indicating each unique trial per participant;
- 317 • “condition”, a factor indicating the levels of the independent variable (1, 2, ...) and  
     the corresponding labels;
- 319 • “rt”, indicating the response times in ms;
- 320 • “acc”, indicating the accuracies (1/0).

321 In the code of Tutorial 1a, this is accomplished as follows.

```
data_wr<-read_csv("../Tutorial_1_descriptive_stats/data/DataExp1_6subjects_wrangled.csv")
data_wr <- data_wr %>%
  rename(pid = vp, condition = prime_type, acc = respac, trial = TrialNr) %>%
  mutate(condition = condition + 1, # original levels were 0, 1, 2.
        condition = factor(condition,
                            levels=c(1,2,3),
                            labels=c("blank","congruent","incongruent")))
```

322        Next, we can set up the life tables and plots of the discrete-time functions  $h(t)$ ,  $S(t)$ ,  
 323       $ca(t)$ , and  $P(t)$  – see section A of the Supplemental Material for their definitions. To do so  
 324      using a functional programming approach, one has to nest the data within participants  
 325      using the `group_nest()` function, and supply a user-defined censoring time and bin width  
 326      to our custom function “`censor()`”, as follows.

```
data_nested <- data_wr %>% group_nest(pid)

data_final <- data_nested %>%
  # ! user input: censoring time, and bin width
  mutate(censored = map(data, censor, 600, 40)) %>%
  # create person-trial-bin data set
  mutate(ptb_data = map(censored, ptb)) %>%
  # create life tables without ca(t)
  mutate(lifetable = map(ptb_data, setup_lt)) %>%
  # calculate ca(t)
  mutate(condacc = map(censored, calc_ca)) %>%
  # create life tables with ca(t)
  mutate(lifetable_ca = map2(lifetable, condacc, join_lt_ca)) %>%
  # create plots
  mutate(plot = map2(.x = lifetable_ca, .y = pid, plot_eha,1))
```

327       Note that the censoring time should be a multiple of the bin width (both in ms). The  
 328      censoring time should be a time point after which no informative responses are expected  
 329      anymore. In experiments that implement a response deadline in each trial the censoring  
 330      time can equal that deadline time point. Trials with a RT larger than the censoring time,  
 331      or trials in which no response is emitted during the data collection period, are treated as  
 332      right-censored observations in EHA. In other words, these trials are not discarded, because  
 333      they contain the information that the event did not occur before the censoring time.  
 334      Removing such trials before calculating the mean event time will result in underestimation  
 335      of the true mean.

336        The person-trial-bin oriented data set is created by our custom function `ptb()`, and it

337        has one row for each time bin (of each trial) that is at risk for event occurrence. The

338        variable “event” in the person-trial-bin oriented data set indicates whether a response

339        occurs (1) or not (0) for each bin.

340        The next step is to set up the life table using our custom function `setup_lt()`,

341        calculate the conditional accuracies using our custom function `calc_ca()`, add the `ca(t)`

342        estimates to the life table using our custom function `join_lt_ca()`, and then plot the

343        descriptive statistics using our custom function `plot_eha()`. When creating the plots, some

344        warning messages will likely be generated, like these:

- 345        • Removed 2 rows containing missing values or values outside the scale range

346            (`geom_line()`).

- 347        • Removed 2 rows containing missing values or values outside the scale range

348            (`geom_point()`).

- 349        • Removed 2 rows containing missing values or values outside the scale range

350            (`geom_segment()`).

351        The warning messages are generated because some bins have no hazard and `ca(t)`

352        estimates, and no error bars. They can thus safely be ignored. One can now inspect

353        different aspects, including the life table for a particular condition of a particular subject,

354        and a plot of the different functions for a particular participant. In general, it is important

355        to visually inspect the functions first for each participant, in order to identify individuals

356        that may be guessing (e.g., a flat conditional accuracy function at .5 indicates that

357        someone is just guessing), outlying individuals, and/or different groups with qualitatively

358        different behavior.

359        Table 3 shows the life table for condition “blank” (no prime stimulus presented) for

360        participant 6.

```
361 ## Warning in attr(x, "align"): 'xfun::attr()' is deprecated.  
362 ## Use 'xfun::attr2()' instead.  
363 ## See help("Deprecated")
```

Table 3

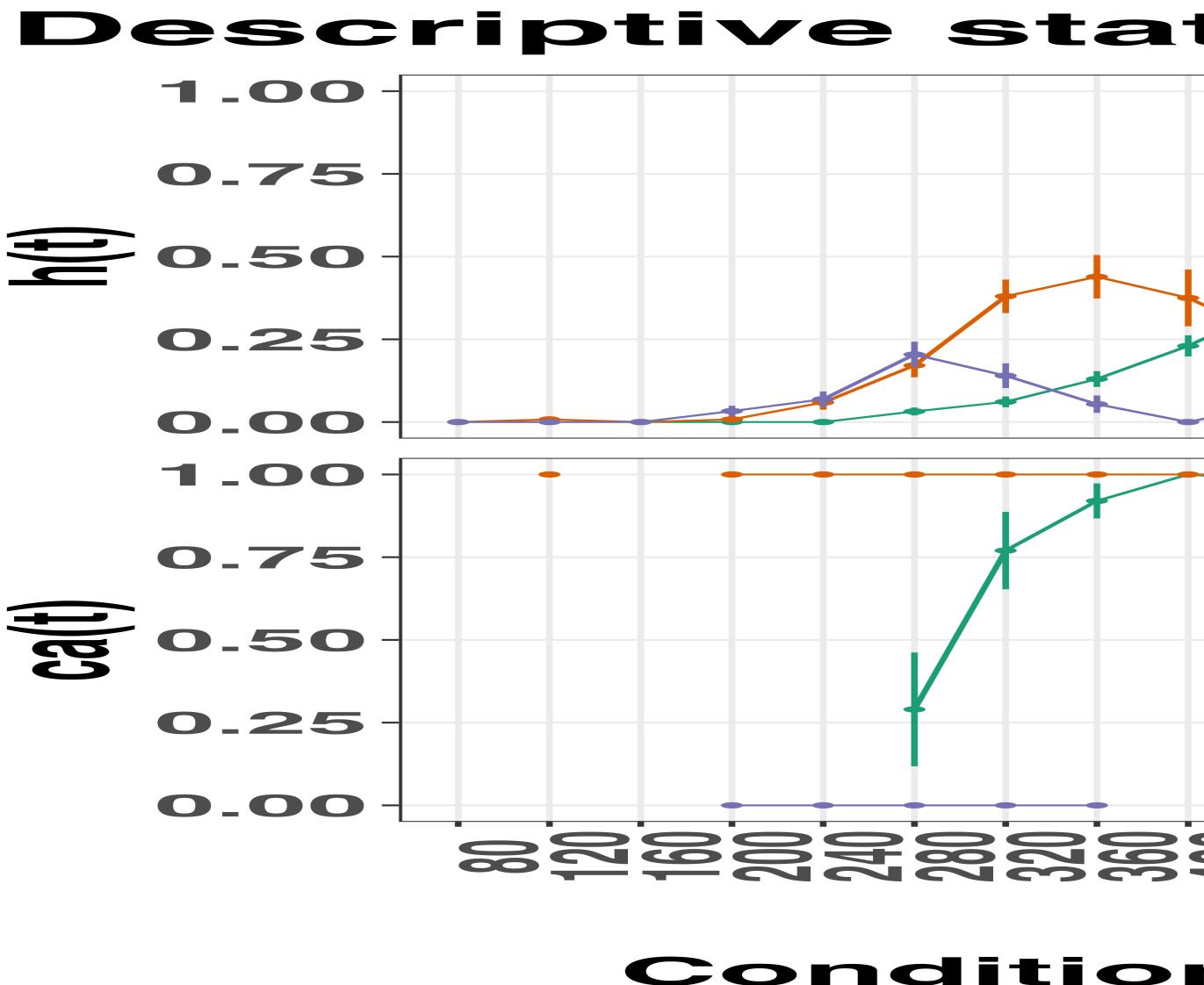
*The life table for the blank prime condition of participant 6.*

bin	risk_set	events	hazard	se_haz	survival	se_surv	ca	se_ca
0	220	NA	NA	NA	1.00	0.00	NA	NA
40	220	0	0.00	0.00	1.00	0.00	NA	NA
80	220	0	0.00	0.00	1.00	0.00	NA	NA
120	220	0	0.00	0.00	1.00	0.00	NA	NA
160	220	0	0.00	0.00	1.00	0.00	NA	NA
200	220	0	0.00	0.00	1.00	0.00	NA	NA
240	220	0	0.00	0.00	1.00	0.00	NA	NA
280	220	7	0.03	0.01	0.97	0.01	0.29	0.17
320	213	13	0.06	0.02	0.91	0.02	0.77	0.12
360	200	26	0.13	0.02	0.79	0.03	0.92	0.05
400	174	40	0.23	0.03	0.61	0.03	1.00	0.00
440	134	48	0.36	0.04	0.39	0.03	0.98	0.02
480	86	37	0.43	0.05	0.22	0.03	1.00	0.00
520	49	32	0.65	0.07	0.08	0.02	1.00	0.00
560	17	9	0.53	0.12	0.04	0.01	1.00	0.00
600	8	4	0.50	0.18	0.02	0.01	1.00	0.00

*Note.* The column named “bin” indicates the endpoint of each time bin (in ms), and includes time point zero. For example the first bin is (0,40] with the starting point excluded and the endpoint included. At time point zero, no events can occur and therefore  $h(t=0)$  and  $ca(t=0)$  are undefined.  $se =$  standard error.  $ca =$  conditional accuracy.  $NA =$  undefined.

Figure 4 displays the discrete-time hazard, survivor, conditional accuracy, and

365 probability mass functions for each prime condition for participant 6. By using  
 366 discrete-time hazard functions of event occurrence – in combination with conditional  
 367 accuracy functions for two-choice tasks – one can provide an unbiased, time-varying, and  
 368 probabilistic description of the latency and accuracy of responses based on all trials of any  
 369 data set.



*Figure 2.* Estimated discrete-time hazard ( $h$ ), survivor ( $S$ ), conditional accuracy ( $ca$ ) and probability mass ( $P$ ) functions for participant 6. Vertical dotted lines indicate the estimated median RTs. Error bars represent  $\pm 1$  standard error of the respective proportion.

370 For example, for participant 6, the estimated hazard values in bin (240,280] are 0.03,

371 0.17, and 0.20 for the blank, congruent, and incongruent prime conditions, respectively. In

372 other words, when the waiting time has increased until *240 ms* after target onset, then the

373 conditional probability of response occurrence in the next 40 ms is more than five times

374 larger for both prime-present conditions, compared to the blank prime condition.

375 Furthermore, the estimated conditional accuracy values in bin (240,280] are 0.29, 1,

376 and 0 for the blank, congruent, and incongruent prime conditions, respectively. In other

377 words, if a response is emitted in bin (240,280], then the probability that it is correct is

378 estimated to be 0.29, 1, and 0 for the blank, congruent, and incongruent prime conditions,

379 respectively.

380 However, when the waiting time has increased until *400 ms* after target onset, then

381 the conditional probability of response occurrence in the next 40 ms is estimated to be

382 0.36, 0.25, and 0.06 for the blank, congruent, and incongruent prime conditions,

383 respectively. And when a response does occur in bin (400,440], then the probability that it

384 is correct is estimated to be 0.98, 1, and 1 for the blank, congruent, and incongruent prime

385 conditions, respectively.

386 These distributional results suggest that participant 6 is initially responding to the

387 prime even though (s)he was instructed to only respond to the target, that response

388 competition emerges in the incongruent prime condition around 300 ms, and that only

389 slower responses are fully controlled by the target stimulus. Qualitatively similar results

390 were obtained for the other five participants. When participants show qualitatively similar

391 distributional patterns, one might consider aggregating their data and plotting the

392 group-average distribution per condition (see Tutorial\_1a.Rmd).

393 In general, these results go against the (often implicit) assumption in research on

394 priming that all observed responses are primed responses to the target stimulus. Instead,

395 the distributional data show that early responses are triggered exclusively by the prime

396 stimulus, while only later responses reflect primed responses to the target stimulus.

397 At this point, we have calculated, summarised and plotted descriptive statistics for  
398 the key variables in EHA/SAT. As we will show in later Tutorials, statistical models for  
399  $h(t)$  and  $ca(t)$  can be implemented as generalized linear mixed regression models predicting  
400 event occurrence (1/0) and conditional accuracy (1/0) in each bin of a selected time  
401 window for analysis. But first we consider calculating the descriptive statistics for two  
402 independent variables.

403 **4.2 Tutorial 1b: Generalising to a more complex design**

404 So far in this paper, we have used a simple experimental design, which involved one  
405 condition with three levels. But psychological experiments are often more complex, with  
406 crossed factorial designs and/or conditions with more than three levels. The purpose of  
407 Tutorial 1b, therefore, is to provide a generalisation of the basic approach, which extends  
408 to a more complicated design. We felt that this might be useful for researchers in  
409 experimental psychology that typically use crossed factorial designs.

410 To this end, Tutorial 1b illustrates how to calculate and plot the descriptive statistics  
411 for the full data set of Experiment 1 of Panis and Schmidt (2016), which includes two  
412 independent variables: mask type and prime type. As we use the same functional  
413 programming approach as in Tutorial 1a, we simply present the sample-based functions for  
414 each participant as part of Tutorial\_1b.Rmd for those that are interested.

415 **4.3 Tutorial 2a: Fitting Bayesian hazard models to discrete time-to-event data**

416 In this third tutorial, we illustrate how to fit Bayesian multilevel regression models to  
417 the RT data of the masked response priming data used in Tutorial 1a. Fitting (Bayesian or  
418 non-Bayesian) regression models to time-to-event data is important when you want to  
419 study how the shape of the hazard function depends on various predictors (Singer &

420 Willett, 2003).

421 **4.3.1 Hazard model considerations.** There are several analytic decisions one  
422 has to make when fitting a discrete-time hazard model. First, one has to select an analysis  
423 time window, i.e., a contiguous set of bins for which there is enough data for each  
424 participant. Second, given that the dependent variable (event occurrence) is binary, one  
425 has to select a link function (see section C in the Supplemental Material). The cloglog link  
426 is preferred over the logit link when events can occur in principle at any time point within  
427 a bin, which is the case for RT data (Singer & Willett, 2003). Third, one has to choose  
428 whether to treat TIME (i.e., the time bin index  $t$ ) as a categorical or continuous predictor.  
429 And when you treat a variable as a categorical predictor, you can choose between reference  
430 coding and index coding. With reference coding, one defines the variable as a factor and  
431 selects one of the  $k$  categories as the reference level. `Brm()` will then construct  $k-1$   
432 indicator variables (see model M1d in Tutorial\_2a.Rmd for an example). With index  
433 coding, one constructs an index variable that contains integers that correspond to different  
434 categories (see models M0i and M1i below). As explained by McElreath (2020), the  
435 advantage of index coding is that the same prior can be assigned to each level of the index  
436 variable, so that each category has the same prior uncertainty.

437 In the case of a large- $N$  design without repeated measurements, the parameters of a  
438 discrete-time hazard model can be estimated using standard logistic regression software  
439 after expanding the typical person-trial data set into a person-trial-bin data set (Allison,  
440 2010). When there is clustering in the data, as in the case of a small- $N$  design with  
441 repeated measurements, the parameters of a discrete-time hazard model can be estimated  
442 using population-averaged methods (e.g., Generalized Estimating Equations), and Bayesian  
443 or frequentist generalized linear mixed models (Allison, 2010).

444 In general, there are three assumptions one can make or relax when adding  
445 experimental predictor variables and other covariates: The linearity assumption for  
446 continuous predictors (the effect of a 1 unit change is the same anywhere on the scale), the

447 additivity assumption (predictors do not interact), and the proportionality assumption  
 448 (predictors do not interact with TIME).

449 In tutorial\_2a.Rmd we fit several Bayesian multilevel models (i.e., generalized linear  
 450 mixed models) that differ in complexity to the person-trial-bin oriented data set that we  
 451 created in Tutorial 1a. We decided to select the analysis time window (200,600] and the  
 452 cloglog link. Below, we shortly discuss two of these models. The person-trial-bin data set is  
 453 prepared as follows.

```
# read in the file we saved in tutorial 1a
ptb_data <- read_csv("Tutorial_1_descriptive_stats/data/inputfile_hazard_modeling.csv")

ptb_data <- ptb_data %>%
  # select analysis time range: (200,600] with 10 bins (time bin ranks 6 to 15)
  filter(period > 5) %>%
    # define categorical predictor TIME as index variable named timebin
  mutate(timebin = factor(period, levels = c(6:15)),
    # factor "condition" using reference coding, with "blank" as the reference level
    condition = factor(condition, labels = c("blank", "congruent", "incongruent")),
    # categorical predictor "prime" with index coding
    prime = ifelse(condition=="blank", 1, ifelse(condition=="congruent", 2, 3)),
    prime = factor(prime, levels = c(1,2,3)))
```

454 **4.3.2 Prior distributions.** To get the posterior distribution of each model  
 455 parameter given the data, we need to specify prior distributions for the model parameters  
 456 which reflect our prior beliefs. In Tutorial\_2a.Rmd we perform a few prior predictive  
 457 checks to make sure our selected prior distributions reflect our prior beliefs (Gelman,  
 458 Vehtari, et al., 2020).

459 The middle column of Supplementary Figure 2 (section E of the Supplemental  
 460 Material) shows six examples of prior distributions for an intercept on the logit and/or  
 461 cloglog scales. While a normal distribution with relatively large variance is often used as a

462 weakly informative prior for continuous dependent variables, rows A and B of  
 463 Supplementary Figure 2 show that specifying such distributions on the logit and cloglog  
 464 scales actually leads to rather informative distributions on the original probability scale, as  
 465 most mass is pushed to probabilities of 0 and 1.

466       **4.3.3 Model M0i: A null model with index coding.** When you do not want to  
 467 make assumptions about the shape of the hazard function, or its shape is not smooth but  
 468 irregular, then you can use a general specification of TIME, i.e., fit one grand intercept per  
 469 time bin. In this first model, we use a general specification of TIME using index coding,  
 470 and do not include experimental predictors. We call this model “M0i”.

471       Before we fit model M0i, we select the necessary columns from the data, and specify  
 472 our priors. In the code of Tutorial 2a, model M0i is specified as follows.

```
model_M0i <-  
  
  brm(data = data_M0i,  
        family = bernoulli(link="cloglog"),  
        formula = event ~ 0 + timebin + (0 + timebin | pid),  
        prior = priors_M0i,  
        chains = 4, cores = 4,  
        iter = 3000, warmup = 1000,  
        control = list(adapt_delta = 0.999,  
                      step_size = 0.04,  
                      max_treedepth = 12),  
        seed = 12, init = "0",  
        file = "Tutorial_2_Bayesian/models/model_M0i")
```

473       After selecting the bernoulli family and the cloglog link, the model formula is  
 474 specified. The specification “0 + ...” removes the default intercept in brm(). The fixed  
 475 effects include an intercept for each level of timebin. Each of these intercepts is allowed to

476 vary across individuals (variable pid). We request 2000 samples from the posterior  
 477 distribution for each of four chains. Estimating model M0i took about 30 minutes on a  
 478 MacBook Pro (Sonoma 14.6.1 OS, 18GB Memory, M3 Pro Chip).

479 **4.3.4 Model M1i: Adding the effects of prime-target congruency.** Previous  
 480 research has shown that psychological effects typically change over time (Panis, 2020;  
 481 Panis, Moran, et al., 2020; Panis & Schmidt, 2022; Panis et al., 2017; Panis & Wagemans,  
 482 2009). In the next model, therefore, we use index coding for both TIME (variable  
 483 “timebin”) and the categorical predictor prime-target-congruency (variable “prime”), so  
 484 that we get 30 grand intercepts, one for each combination of timebin level and prime level.  
 485 Here is the model formula of this model that we call “M1i”.

```
event ~ 0 + timebin:prime + (0 + timebin:prime | pid)
```

486 Estimating model M1i took about 124 minutes.

487 **4.3.5 Compare the models.** We can compare the two models using the Widely  
 488 Applicable Information Criterion (WAIC) and Leave-One-Out (LOO) cross-validation, and  
 489 look at model weights for both criteria (Kurz, 2023a; McElreath, 2020).

```
model_weights(model_M0i, model_M1i, weights = "loo") %>% round(digits = 2)
```

```
490 ## model_M0i model_M1i
491 ## 0 1
```

```
model_weights(model_M0i, model_M1i, weights = "waic") %>% round(digits = 2)
```

```
492 ## model_M0i model_M1i
493 ## 0 1
```

494 Clearly, both the loo and waic weighting schemes assign a weight of 1 to model M1i,  
 495 and a weight of 0 to the other simpler model.

496        **4.3.6 Evaluating parameter estimates in model M1i.** To make inferences

497    from the parameter estimates in model M1i, we first plot the densities of the draws from

498    the posterior distributions of its population-level parameters in Figure 5, together with

499    point (median) and interval estimates (80% and 95% credible intervals).

## Posterior distributions for population-level effects in Model M1i

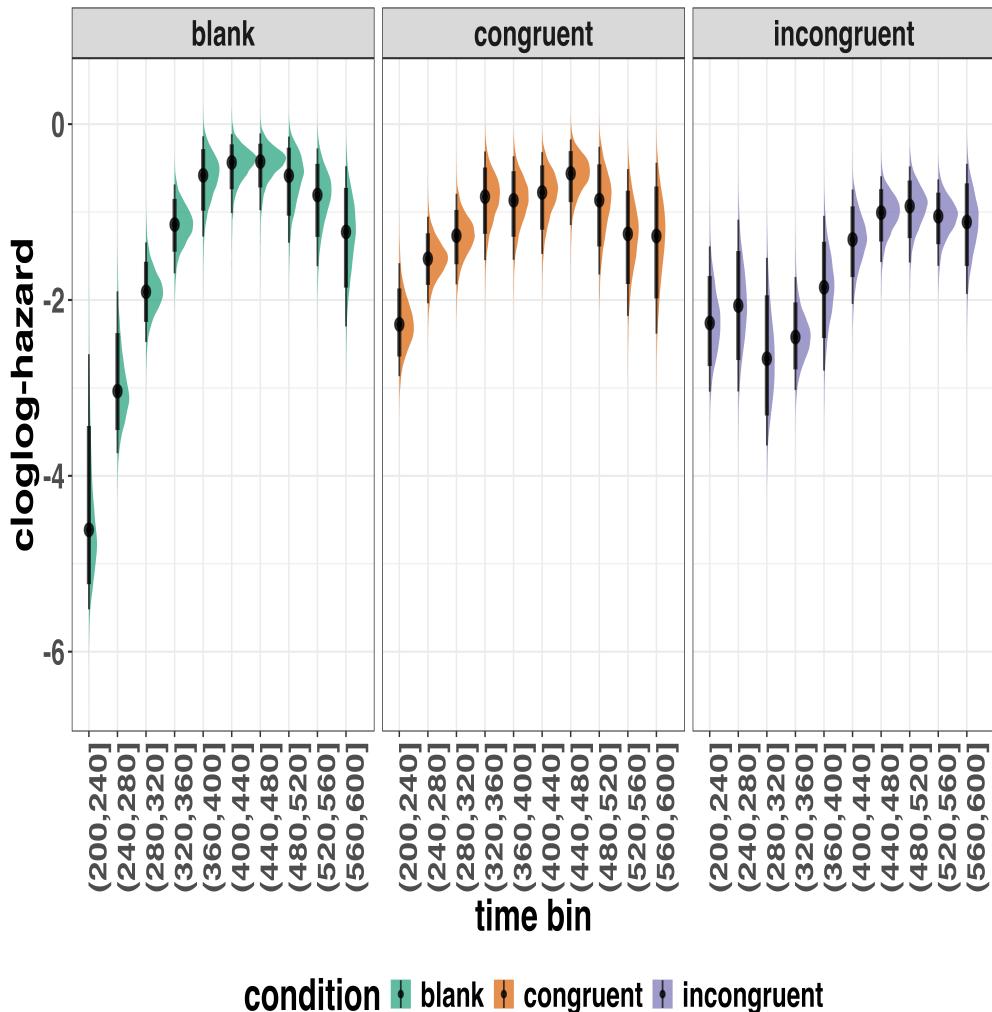


Figure 3. Medians and 80/95% credible intervals of the posterior distributions of the population-level parameters of model M1i.

500    Because the parameter estimates are on the cloglog-hazard scale, we can ease our

501 interpretation by plotting the expected value of the posterior predictive distribution – the  
 502 predicted hazard values – at the population level (Figure 6A), and for each participant in  
 503 the data set (Figure 6B).

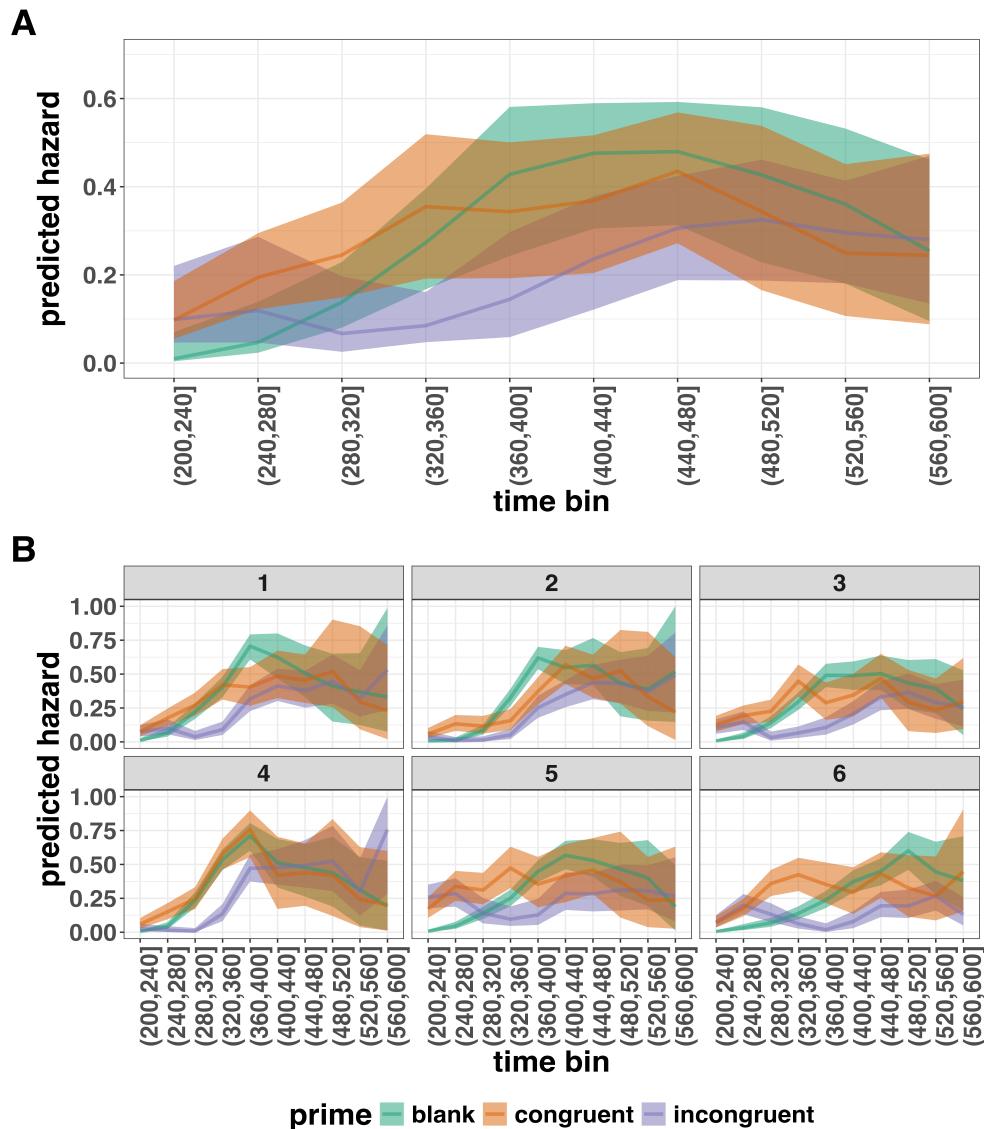


Figure 4. Point (median) and 80/95% credible interval summaries of the hazard estimates (expected values of the draws from the posterior predictive distributions) in each time bin at the population level (A), and for each participant (B).

504 As we are actually interested in the effects of congruent and incongruent primes,

505 relative to the blank prime condition, we can construct two contrasts (congruent-blank,  
 506 incongruent-blank), and plot the posterior distributions of these contrast effects, both at  
 507 the population level (Figure 7A; grand average marginal effect) and at the participant level  
 508 (Figure 7B; subject-specific average marginal effect).

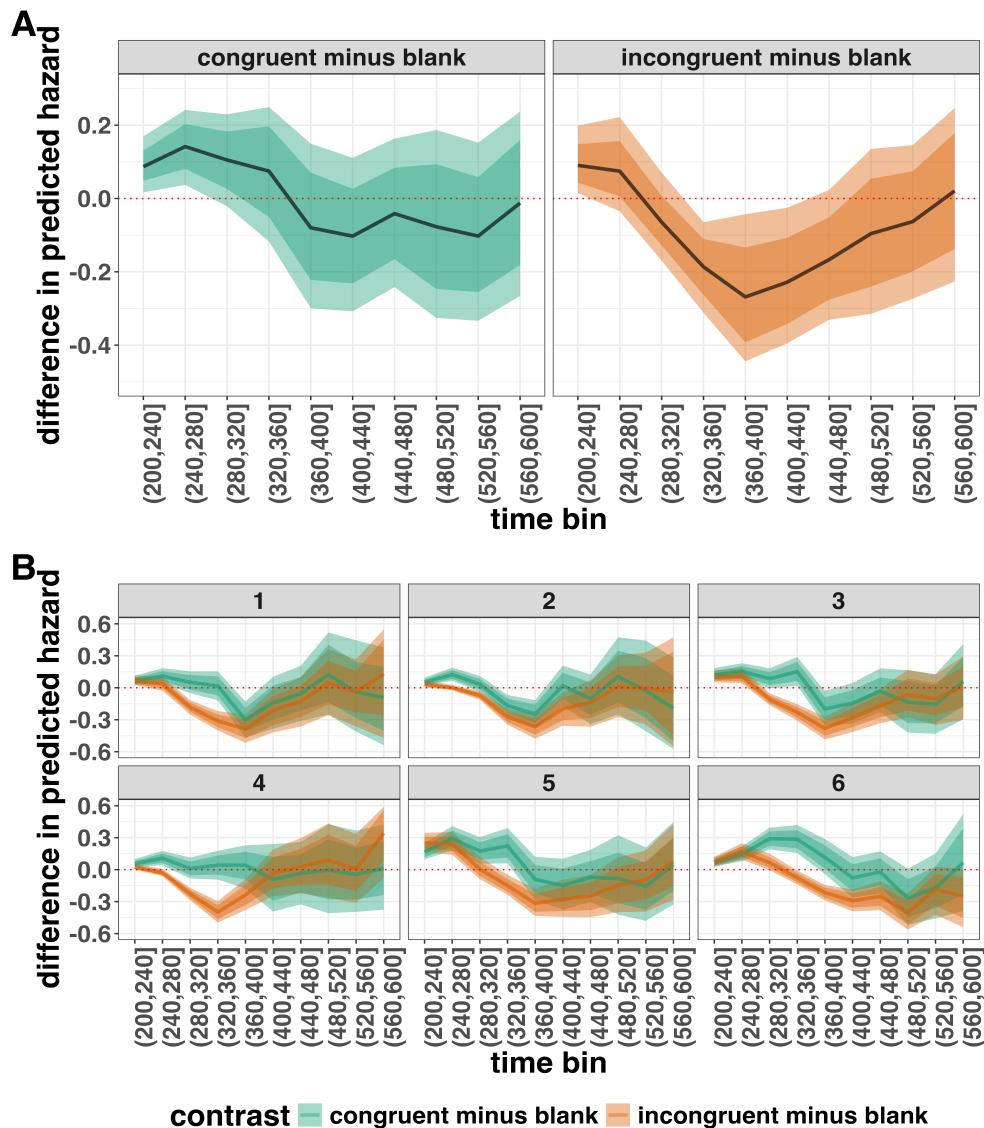


Figure 5. Point (mean) and 80/95% credible interval summaries of estimated differences in hazard in each time bin at the population level (A), and for each participant (B).

509 The point estimates and quantile intervals can be reported in a table (see

510 Tutorial\_2a.Rmd for details).

511 ***Example conclusions for M1i.*** What can we conclude from model M1i about  
512 our research question, i.e., the temporal dynamics of the effect of prime-target congruency  
513 on RT? In other words, in which of the 40-ms time bins between 200 and 600 ms after  
514 target onset does changing the prime from blank to congruent or incongruent affect the  
515 hazard of response occurrence (for a prime-target SOA of 187 ms)?

516 If we want to estimate the population-level effect of prime type on hazard, we can  
517 base our conclusion on Figure 7A. The contrast “congruent minus blank” was estimated to  
518 be 0.09 hazard units in bin (200,240] (95% CrI = [0.02, 0.17]), and 0.14 hazard units in bin  
519 (240,280]) (95% CrI = [0.04, 0.25]). For the other bins, the 95% credible interval contained  
520 zero. The contrast “incongruent minus blank” was estimated to be 0.09 hazard units in bin  
521 (200,240] (95% CrI = [0.01, 0.21]), -0.19 hazard units in bin (320,360] (95% CrI = [-0.31,  
522 -0.06]), -0.27 hazard units in bin (360,400] (95% CrI = [-0.45, -0.04]), and -0.23 hazard  
523 units in bin (400,440] (95% CrI = [-0.40, -0.03]). For the other bins, the 95% credible  
524 interval contained zero.

525 There are thus two phases of performance for the average person between 200 and  
526 600 ms after target onset. In the first phase, the addition of a congruent or incongruent  
527 prime stimulus increases the hazard of response occurrence compared to blank prime trials  
528 in the time period (200, 240]. In the second phase, only the incongruent prime decreases  
529 the hazard of response occurrence compared to blank primes, in the time period (320,440].  
530 The sign of the effect of incongruent primes on the hazard of response occurrence thus  
531 depends on how much waiting time has passed since target onset.

532 If we want to focus more on inter-individual differences, we can study the  
533 subject-specific hazard functions in Figure 7B. Note that three participants (1, 2, and 3)  
534 show a negative difference for the contrast “congruent minus incongruent” in bin (360,400]  
535 – subject 2 also in bin (320,360].

536 Future studies could (a) increase the number of participants to estimate the  
537 proportion of “dippers” in the subject population, and/or (b) try to explain why this dip  
538 occurs. For example, Panis and Schmidt (2016) concluded that active, top-down,  
539 task-guided response inhibition effects emerge around 360 ms after the onset of the stimulus  
540 following the prime (here: the target stimulus). Such a top-down inhibitory effect might  
541 exist in our priming data set, because after some time participants will learn that the first  
542 stimulus is not the one they have to respond to. To prevent a premature overt response to  
543 the prime they thus might gradually increase a global response threshold during the  
544 remainder of the experiment, which could result in a lower hazard in congruent trials  
545 compared to blank trials, for bins after ~360 ms, and towards the end of the experiment.  
546 This effect might be masked for incongruent primes by the response competition effect.

547 Interestingly, all subjects show a tendency in their mean difference (congruent minus  
548 blank) to “dip” around that time (Figure 7B). Therefore, future modeling efforts could  
549 incorporate the trial number into the model formula, in order to also study how the effects  
550 of prime type on hazard change on the long experiment-wide time scale, next to the short  
551 trial-wide time scale. In Tutorial\_2a.Rmd we provide a number of model formulae that  
552 should get you going.

#### 553 4.4 Tutorial 2b: Fitting Bayesian conditional accuracy models

554 In this fourth tutorial, we illustrate how to fit a Bayesian multilevel regression model  
555 to the timed accuracy data from the masked response priming data used in Tutorial 1a.  
556 The general process is similar to Tutorial 2a, except that (a) we use the person-trial data,  
557 (b) we use the logit link function, and (c) we change the priors. To keep the tutorial short,  
558 we only fit one conditional accuracy model, which was based on model M1i from Tutorial  
559 2a and labelled M1i\_ca.

560 To make inferences from the parameter estimates in model M1i\_ca, we first plot the

561 densities of the draws from the posterior distributions of its population-level parameters in  
 562 Figure 8, together with point (median) and interval estimates (80% and 95% credible  
 563 intervals).

## Posterior distributions for population-level effects in Model M1i\_ca

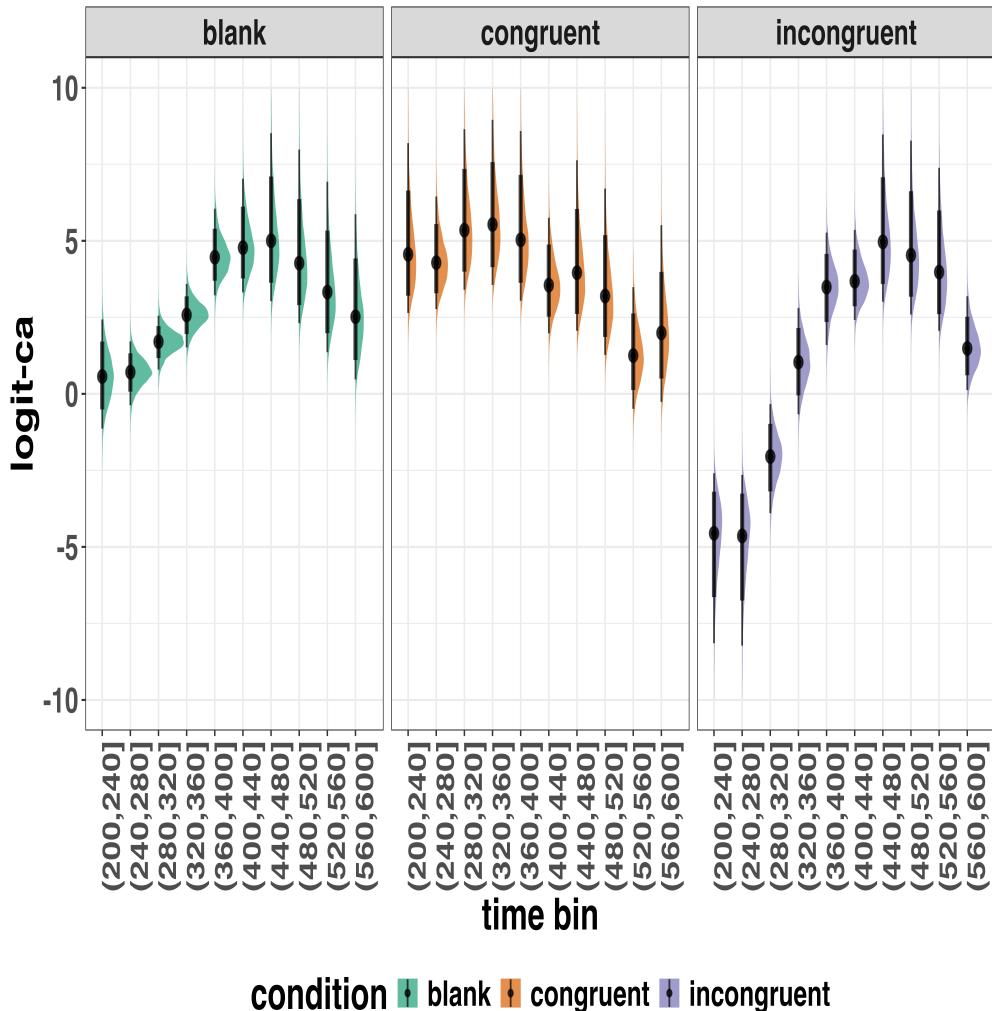


Figure 6. Medians and 80/95% credible intervals of the posterior distributions of the population-level parameters of model M1i\_ca. ca = conditional accuracy.

564 Because the parameter estimates are on the logit-ca scale, we can ease our  
 565 interpretation by plotting the expected value of the posterior predictive distribution – the

<sup>566</sup> predicted conditional accuracies – at the population level (Figure 9A), and for each  
<sup>567</sup> participant in the data set (Figure 9B).

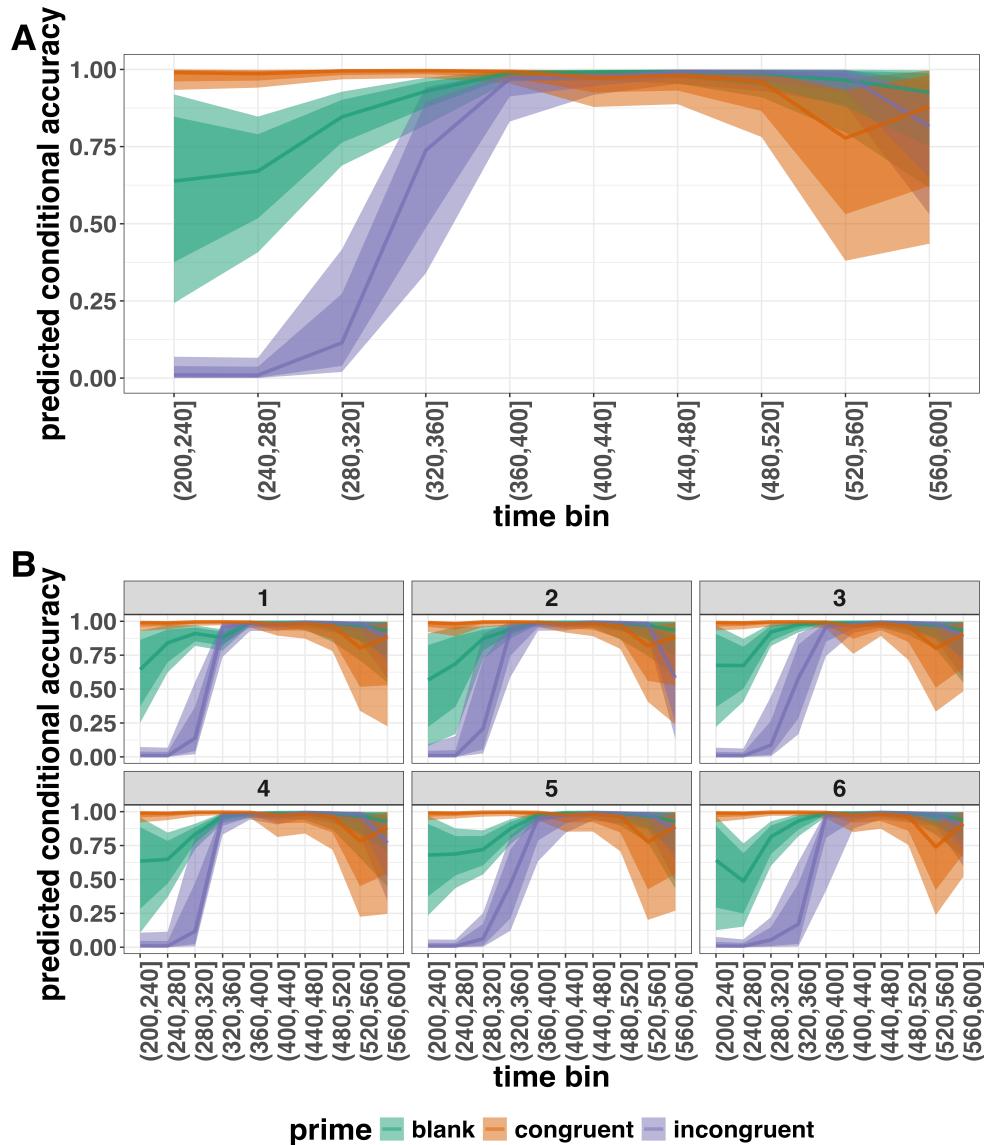


Figure 7. Point (median) and 80/95% credible interval summaries of the conditional accuracy estimates (expected values of the draws from the posterior predictive distributions) in each time bin at the population level (A), and for each participant (B).

<sup>568</sup> As we are actually interested in the effects of congruent and incongruent primes,  
<sup>569</sup> relative to the blank prime condition, we can construct two contrasts (congruent-blank,

570 incongruent-blank), and plot the posterior distributions of these contrast effects at the  
 571 population level (Figure 10A) and for each participant (Figure 10B).

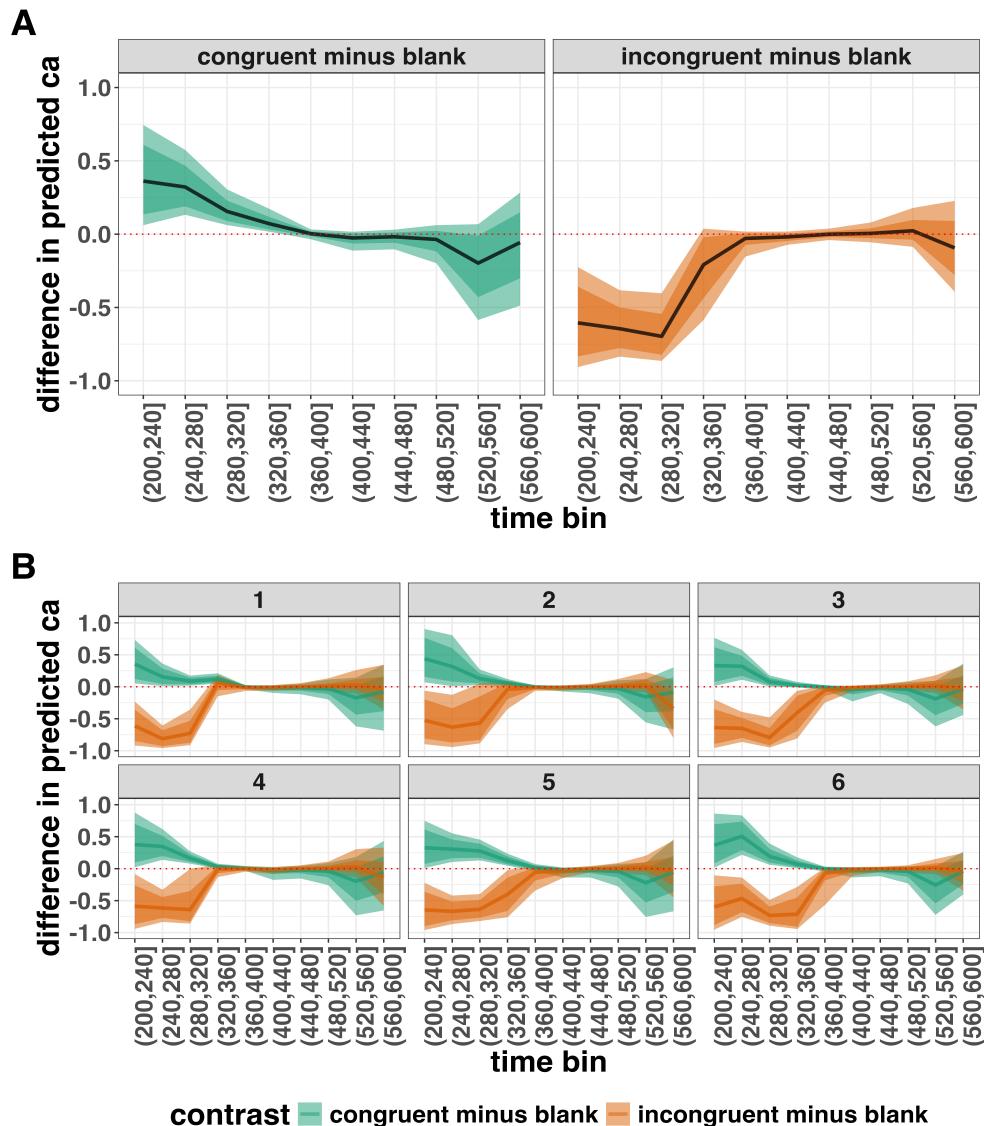


Figure 8. Point (mean) and 80/95% credible interval summaries of estimated differences in conditional accuracy in each time bin at the population level (A), and for each participant (B).

572 Based on Figure 10A we see that on the population level congruent primes have a  
 573 positive effect on the conditional accuracy of emitted responses in time bins (200,240],

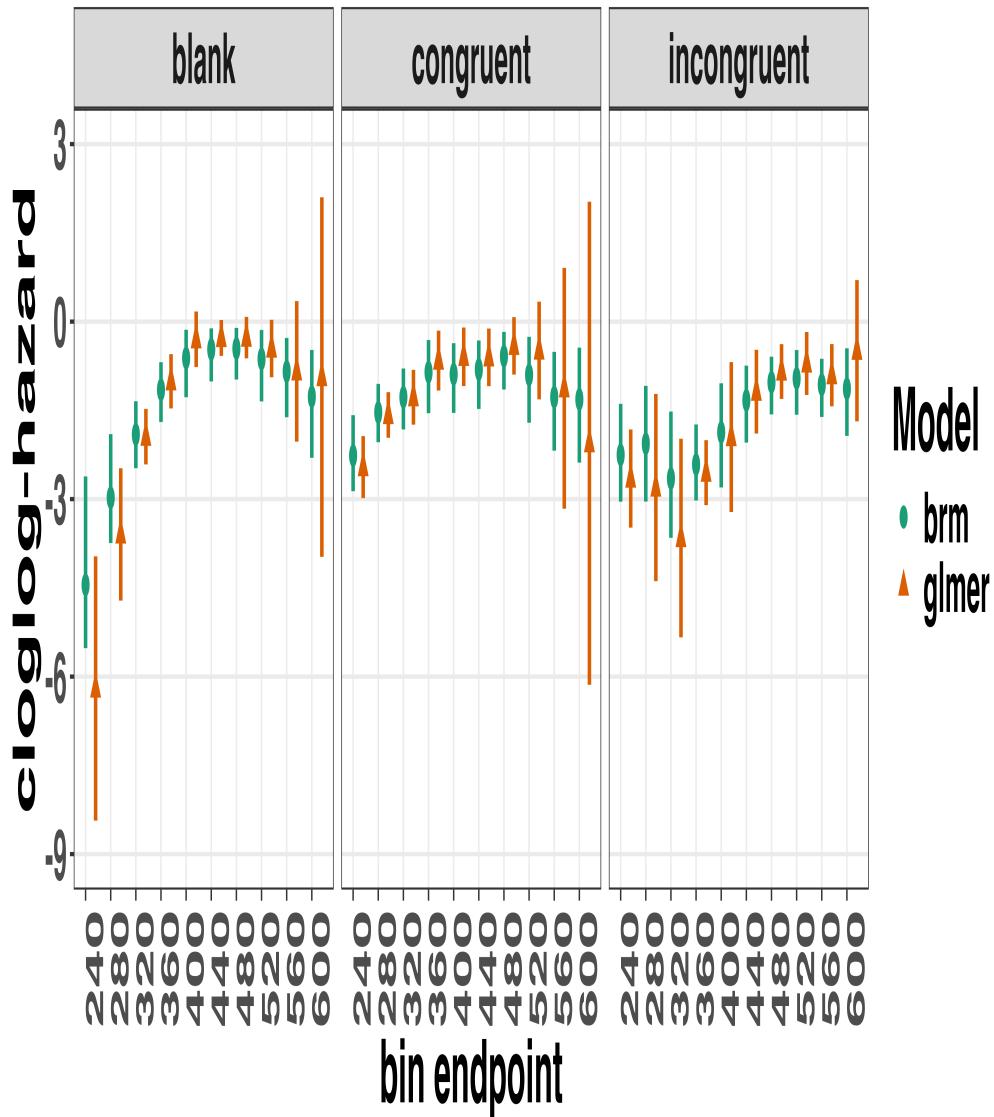
574 (240,280], (280,320], and (320,360], relative to the estimates in the baseline condition  
575 (blank prime; red dashed lines in Figure 10A). Incongruent primes have a negative effect on  
576 the conditional accuracy of emitted responses in the first time bins, relative to the  
577 estimates in the baseline condition.

578 **4.5 Tutorial 3a: Fitting Frequentist hazard models**

579 In this fifth tutorial we illustrate how to fit a multilevel regression model to RT data  
580 in the frequentist framework, for the data used in Tutorial 1a. The general process is  
581 similar to that in Tutorial 2a, except that there are no priors to set.

582 Again, to keep the tutorial concise, we only fit model M1i (see Tutorial 2a) using the  
583 function `glmer()` from the R package `lme4`. Alternatively, one could also use the function  
584 `glmmPQL()` from the R package `MASS` (Ripley et al., 2024). The resulting hazard model  
585 is called `M1i_f` with the appended “`_f`” denoting a frequentist model.

586 In Figure 11 we compare the parameter estimates from the Bayesian regression model  
587 `M1i` with those from the frequentist model `M1i_f`.



*Figure 9.* Parameter estimates for model M1i from brm() – means and 95% credible intervals – and model M1i\_f from glmer() – maximum likelihood estimates and 95% confidence intervals.

588       Figure 11 confirms that the parameter estimates from both Bayesian and frequentist

589       models are pretty similar, which makes sense given the close similarity in model structure.

590       However, model M1i\_f did not converge and resulted in a singular fit. This is of course one

591       of the reasons why Bayesian modeling has become so popular in recent years. But the price

592 you pay for being able to fit models with more complex varying effects structures via a  
593 Bayesian framework is increased computation time. In other words, as we have noted  
594 throughout, some of the Bayesian models in Tutorials 2a took several hours to build.

595 **4.6 Tutorial 3b: Fitting Frequentist conditional accuracy models**

596 In this sixth tutorial we illustrate how to fit a multilevel regression model to the  
597 timed accuracy data in the frequentist framework, for the data used in Tutorial 1a. To be  
598 concise, we only fit effects from model M1i\_ca (see Tutorial 2b) using the function glmer()  
599 from the R package lme4. Alternatively, one could also use the function glmmPQL() from  
600 the R package MASS (Ripley et al., 2024). The resulting conditional accuracy model,  
601 which we labelled M1i\_ca\_f, did not converge and resulted in a singular fit. Again, this  
602 just highlights some of the difficulties in fitting reasonably complex varying/random effects  
603 structures in frequentist workflows.

604 **4.7 Tutorial 4: Planning**

605 In the final tutorial, we look at planning a future experiment, which uses EHA.

606 **4.7.1 Background.** The general approach to planning that we adopt here involves  
607 simulating reasonably structured data to help guide what you might be able to expect from  
608 your data once you collect it (Gelman, Vehtari, et al., 2020). The basic structure and code  
609 follows the examples outlined by Solomon Kurz in his ‘power’ blog posts  
610 (<https://solomonkurz.netlify.app/blog/bayesian-power-analysis-part-i/>) and Lisa  
611 Debruine’s R package faux{} (<https://debruine.github.io/faux/>) as well as these related  
612 papers (DeBruine & Barr, 2021; Pargent, Koch, Kleine, Lemer, & Gaube, 2024).

613 **4.7.2 Basic workflow.** The basic workflow is as follows:

- 614 1. Fit a regression model to existing data.

- 615     2. Use the regression model parameters to simulate new data.
- 616     3. Write a function to create 1000s of datasets and vary parameters of interest (e.g.,  
617       sample size, trial count, effect size).
- 618     4. Summarise the simulated data to estimate likely power or precision of the research  
619       design options.

620       Ideally, in the above workflow, we would also fit a model to each dataset and  
621       summarise the model output, rather than the raw data. However, when each model takes  
622       several hours to build, and we may want to simulate many 1000s of datasets, it can be  
623       computationally demanding for desktop machines. So, for ease, here we just use the raw  
624       simulated datasets to guide future expectations.

625       In the below, we only provide a high-level summary of the process and let readers  
626       dive into the details within the tutorial should they feel so inclined.

627       **4.7.3 Fit a regression model and simulate one dataset.** We again use the  
628       data from Panis and Schmidt (2016) to provide a worked example. We fit an index coding  
629       model on a subset of time bins (six time bins in total) and for two prime conditions  
630       (congruent and incongruent). We chose to focus on a subsample of the data to ease the  
631       computational burden. We also used a full varying effects structure, with the model  
632       formula as follows:

```
event ~ 0 + timebin:prime + (0 + timebin:prime | pid)
```

633       We then took parameters from this model and used them to create a single dataset  
634       with 200 trials per condition for 10 individual participants. The raw data and the  
635       simulated data are plotted in Figure 12 and show quite close correspondence, which is  
636       re-assuring. But, this is only one dataset. What we really want to do is simulate many  
637       datasets and vary parameters of interest, which is what we turn to in the next section.

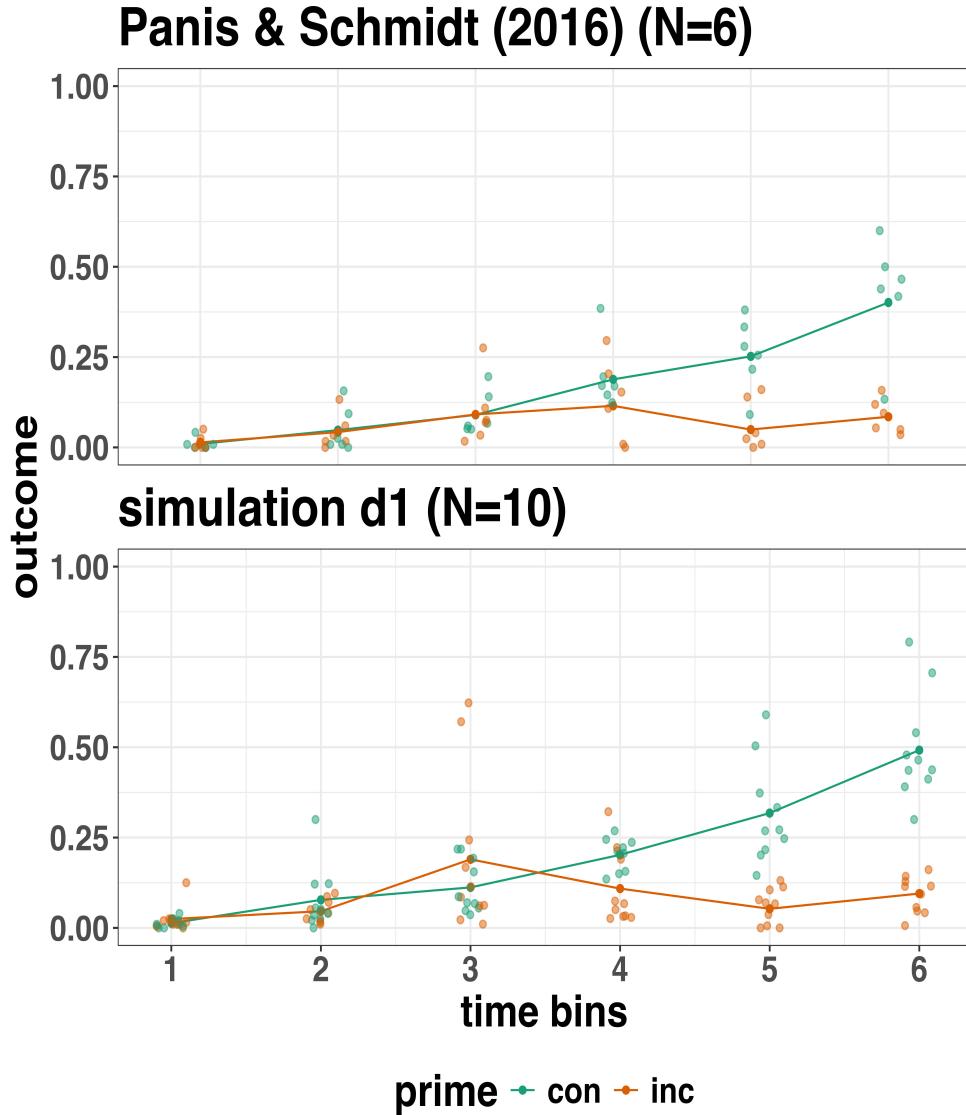


Figure 10. Raw data from Panis and Schmidt (2016) and simulated data from 10 participants.

#### 638        4.7.4 Simulate and summarise data across a range of parameter values.

639        Here we use the same data simulation process as used above, but instead of simulating one  
 640        dataset, we simulate 1000 datasets per variation in parameter values. Specifically, in  
 641        Simulation 1, we vary the number of trials per condition (100, 200, and 400), as well as the  
 642        effect size in bin 6. We focus on bin 6 only, in terms of varying the effect size, just to make

things simpler and easier to understand. The effect size observed in bin 6 in this subsample of data was a 79% reduction in hazard value from the congruent prime (0.401 hazard value) to the incongruent prime condition (0.085 hazard value). In other words, a hazard ratio of 0.21 (e.g.,  $0.085/0.401 = 0.21$ ). As a starting point, we chose three effect sizes, which covered a fairly broad range of hazard ratios (0.25, 0.5, 0.75), which correspond to a 75%, 50% and 25% reduction in hazard value as a function of prime condition.

Summary results from Simulation 1 are shown in Figure 13A. Figure 13A depicts statistical “power” as calculated by the percentage of lower-bound 95% confidence intervals that exclude zero when the difference between prime condition is calculated (congruent - incongruent). In other words, what fraction of the simulated datasets generated an effect of prime that excludes the criterion mark of zero. We are aware that “power” is not part of a Bayesian analytical workflow, but we choose to include it here, as it is familiar to most researchers in experimental psychology.

The results of Simulation 1 show that if we were targeting an effect size similar to the one reported in the original study, then testing 10 participants and collecting 100 trials per condition would be enough to provide over 95% power. However, we could not be as confident about smaller effects, such as a hazard ratio of 50% or 25%. From this simulation, we can see that somewhere between an effect size of a 50% and 75% reduction in hazard value, power increases to a range that most researchers would consider acceptable (i.e., >95% power). To probe this space a little further, we decided to run a second simulation, which varied different parameters.

In Simulation 2, we varied the effect size between a different range of values (0.5, 0.4, 0.3), which correspond to a 50%, 60% and 70% reduction in hazard value as a function of prime condition. In addition, we varied the number of participants per experiment between 10, 15, and 20 participants. Given that trial count per condition made little difference to power in Simulation 1, we fixed trial count at 200 trials per condition in Simulation 2.

669 Summary results from Simulation 2 are shown in Figure 13B. A summary of these power  
670 calculations might be as follows (trial count = 200 per condition in all cases):

- 671 • For a 70% reduction (0.3 hazard ratio), N=10 would give nearly 100% power.  
672 • For a 60% reduction (0.4 hazard ratio), N=10 would give nearly 90% power.  
673 • For a 50% reduction (0.5 hazard ratio), N=15 would give over 80% power.

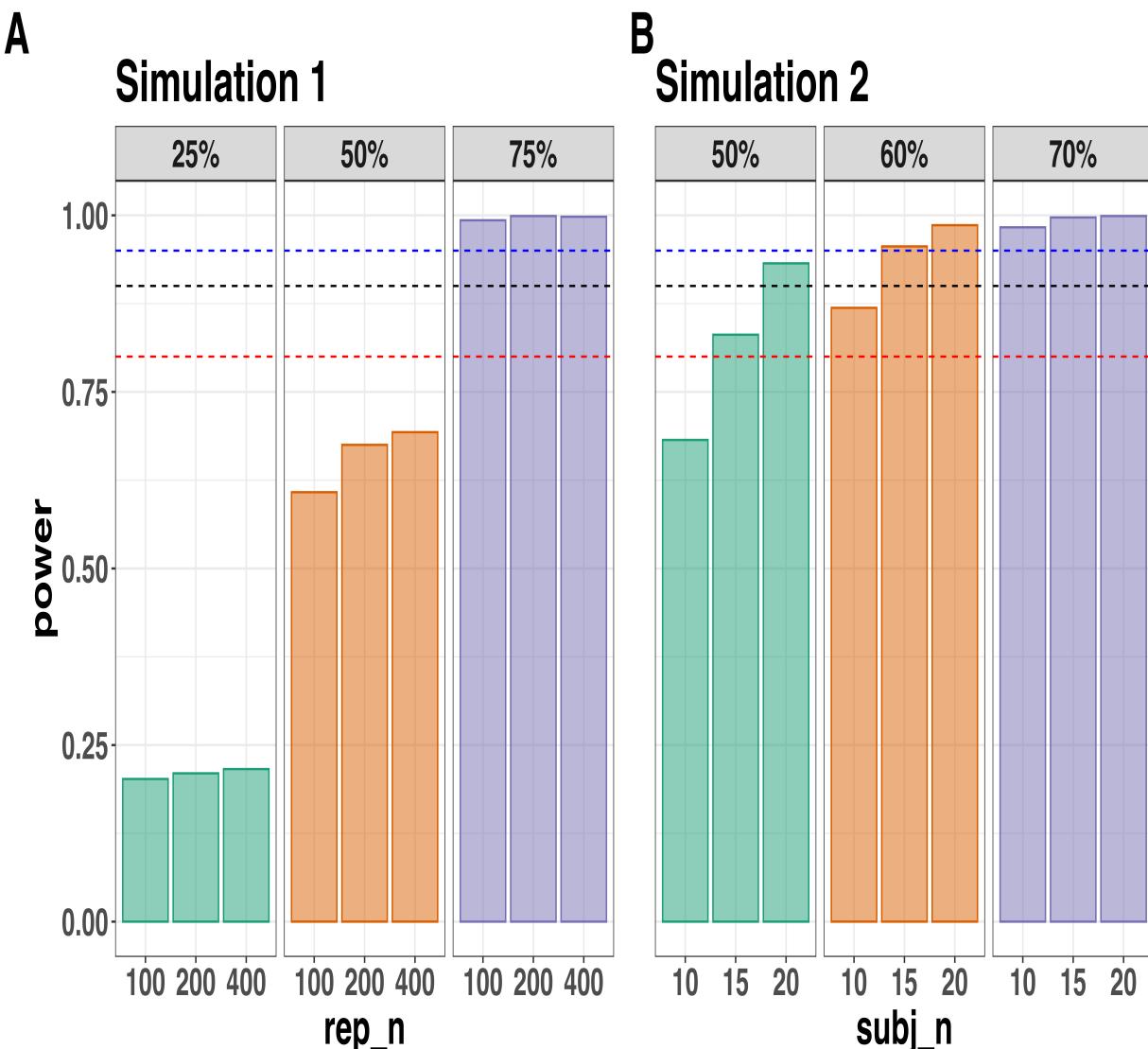


Figure 11. Statistical power across data Simulation 1 (A) and Simulation 2 (B). Power was calculated as the percentage of lower-bound 95% confidence intervals that exclude zero when the difference between prime condition is calculated (congruent - incongruent). In Simulation 1, the effect size was varied between a 25%, 50% and 75% reduction in hazard value, whereas the trial count was varied between 100, 200 and 400 trials per condition (the number of participants was fixed at N=10). In Simulation 2, the effect size was varied between a 50%, 60% and 70% reduction in hazard value, whereas the number of participants was varied between N=10, 15 and 20 (the number of trials per condition was fixed at 200). The dashed lines represent 80% (red), 90% (black) and 95% (blue) power. Abbreviations: rep\_n = the number of trials per experimental condition; subj\_n = the number of participants per simulated experiment.

674       **4.7.5 Planning decisions.** Now that we have summarised our simulated data,

675 what planning decisions could we make about a future study? More concretely, how many

676 trials per condition should we collect and how many participants should we test? Like

677 almost always when planning future studies, the answer depends on your objectives, as well

678 as the available resources (Lakens, 2022). There is no straightforward and clear-cut answer.

679 Some considerations might be as follows:

- 680       • How much power or precision are you looking to obtain in this particular study?

- 681       • Are you running multiple studies that have some form of replication built in?

- 682       • What level of resources do you have at your disposal, such as time, money and

683       personnel?

- 684       • How easy or difficult is it to obtain the specific type of sample?

685       If we were running this kind of study in our lab, what would we do? We might pick a

686 hazard ratio of 0.4 or 0.5 as a target effect size since this is much smaller than that

687 observed previously (Panis & Schmidt, 2016). Then we might pick the corresponding

688 combination of trial count per condition (e.g., 200) and participant sample size (e.g., N=10

689 or N=15) that takes you over the 80% power mark. If we wanted to maximise power based

690 on these simulations, and we had the time and resources available, then we would test

691 N=20 participants, which would provide >90% power for an effect size of 0.5.

692       **But**, and this is an important “but”, unless there are unavoidable reasons, no matter

693 what planning choices we made based on these data simulations, we would not solely rely

694 on data collected from one single study. Instead, we would run a follow-up experiment that

695 replicates and extends the initial result. By doing so, we would aim to avoid the Cult of

696 the Isolated Single Study (Nelder, 1999; Tong, 2019), and thus reduce the reliance on any

697 one type of planning tool, such as a power analysis. Then, we would look for common

698 patterns across two or more experiments, rather than trying to make the case that a single

699 study on its own has sufficient evidential value to hit some criterion mark.

700

## 5. Discussion

701 This main motivation for writing this paper is the observation that EHA and SAT  
702 analysis remain under-used in psychological research. As a consequence, the field of  
703 psychological research is not taking full advantage of the many benefits EHA/SAT provides  
704 compared to more conventional analyses. By providing a freely available set of tutorials,  
705 which provide step-by-step guidelines and ready-to-use R code, we hope that researchers  
706 will feel more comfortable using EHA/SAT in the future. Indeed, we hope that our  
707 tutorials may help to overcome a barrier to entry with EHA/SAT, which is that such  
708 approaches require more analytical complexity compared to mean-average comparisons.  
709 While we have focused here on within-subject, factorial, small- $N$  designs, it is important to  
710 realize that EHA/SAT can be applied to other designs as well (large- $N$  designs with only  
711 one measurement per subject, between-subject designs, etc.). As such, the general workflow  
712 and associated code can be modified and applied more broadly to other contexts and  
713 research questions. In the following, we discuss issues relating to model complexity and  
714 interpretability, individual differences, as well as limitations of the approach and future  
715 extensions.

716 **5.1 What are the main use-cases of EHA for understanding cognition and brain  
717 function?**

718 For those researchers, like ourselves, who are primarily interested in understanding  
719 human cognitive and brain systems, we consider two broadly-defined, main use-cases of  
720 EHA. First, as we hope to have made clear by this point, EHA is one way to investigating  
721 a “temporal states” approach to cognitive processes. EHA provides one way to uncover  
722 when cognitive states may start and stop, as well as what they may be tied to or interact  
723 with. Therefore, if your research questions concern **when** and **for how long** psychological  
724 states occur, our EHA tutorials could be useful tools for you to use.

725        Second, even if you are not primarily interested in studying the temporal states of  
726 cognition, EHA could still be a useful tool to consider using, in order to qualify inferences  
727 that are being made based on mean-average comparisons. Given that distinctly different  
728 inferences can be made from the same data based on whether one computes a  
729 mean-average across trials or a RT distribution of events (Figure 1), it may be important  
730 for researchers to supplement mean-average comparisons with EHA. One could envisage  
731 scenarios where the implicit assumption of an effect manifesting across all of the time bins  
732 measured would not be supported by EHA. Therefore, the conclusion of interest would not  
733 apply to all responses, but instead it would be restricted to certain aspects of time.

## 734 5.2 Model complexity versus interpretability

735        EHA can quickly become very complex when adding more than one time scale, due to  
736 the many possible higher-order interactions. For example, some of the models discussed in  
737 Tutorial 2a, which we did not focus on in the main text, contain two time scales as  
738 covariates: the passage of time on the within-trial time scale, and the passage of time on  
739 the across-trial (or within-experiment) time scale. However, when trials are presented in  
740 blocks, and blocks of trials within sessions, and when the experiment comprises three  
741 sessions, then four time scales can be defined (within-trial, within-block, within-session,  
742 and within-experiment). From a theoretical perspective, adding more than one time scale –  
743 and their interactions – can be important to capture plasticity and other learning effects  
744 that may play out on such longer time scales, and that are probably present in each  
745 experiment in general. From a practical perspective, therefore, some choices need to be  
746 made to balance the amount of data that is being collected per participant, condition and  
747 across the varying timescales. As one example, if there are several timescales of relevance,  
748 then it might be prudent for interpretational purposes to limit the number of experimental  
749 predictor variables (conditions). This is of course where planning and data simulation  
750 efforts would be important to provide a guide to experimental design choices (see Tutorial

751 4).

752 **5.3 Individual differences**

753 One important issue is that of possible individual differences in the overall location of  
754 the distribution, and the time course of psychological effects. For example, when you wait  
755 for a response of the participant on each trial, you allow the participant to have control  
756 over the trial duration, and some participants might respond only when they are confident  
757 that their emitted response will be correct. These issues can be avoided by introducing a  
758 (relatively short) response deadline in each trial, e.g., 500 ms for simple detection tasks,  
759 800 ms for more difficult discrimination tasks, or 2 s for tasks requiring extended high-level  
760 processing. Because EHA can deal in a straightforward fashion with right-censored  
761 observations (i.e., trials without an observed response in the analysis time window),  
762 introducing a response deadline is recommended when designing RT experiments.  
763 Furthermore, introducing a response deadline and asking participants to respond before the  
764 deadline as much as possible, will also lead to individual distributions that overlap in time,  
765 which is important when selecting a common analysis time window when fitting hazard  
766 and conditional accuracy models.

767 But even when using a response deadline, participants can differ qualitatively in the  
768 effects they display (see Panis, 2020). One way to deal with this is to describe and  
769 interpret the different patterns. Another way is to run a clustering algorithm on the  
770 individual hazard estimates across all bins and conditions. The obtained dendrogram can  
771 then be used to identify a (hopefully big) cluster of participants that behave similarly, and  
772 to identify a (hopefully small) cluster of participants with different behavioral patterns.  
773 One might then exclude the smaller sub-group of participants before fitting a hazard model  
774 or consider the possibility that different cognitive processes may be at play during task  
775 performance across the different sub-groups.

776 Another approach to deal with individual differences is Bayesian prevalence (Ince,

777 Paton, Kay, & Schyns, 2021), which is a form of small- $N$  approach (Smith & Little, 2018).

778 This method looks at effects within each individual in the study and asks how likely it

779 would be to see the same result if the experiment was repeated with a new person chosen

780 from the wider population at random. This approach allows one to quantify how typical or

781 uncommon an observed effect is in the population, and the uncertainty around this

782 estimate.

#### 783 5.4 Limitations

784 Compared to the orthodox method – comparing mean-averages between conditions –

785 the most important limitation of multilevel hazard and conditional accuracy modeling is

786 that it might take a long time to estimate the parameters using Bayesian methods or the

787 model might have to be simplified significantly to use frequentist methods.

788 Another issue is that you need a relatively large number of trials per condition to

789 estimate the hazard function with high temporal resolution, which is required when testing

790 predictions of process models of cognition. Indeed, in general, there is a trade-off between

791 the number of trials per condition and the temporal resolution (i.e., bin width) of the

792 hazard function. Therefore, we recommend researchers to collect as many trials as possible

793 per experimental condition, given the available resources and considering the participant

794 experience (e.g., fatigue and boredom). For instance, if the maximum session length

795 deemed reasonable is between 1 and 2 hours, what is the maximum number of trials per

796 condition that you could reasonably collect? After consideration, it might be worth

797 conducting multiple testing sessions per participant and/or reducing the number of

798 experimental conditions. Finally, there is a user-friendly online tool for calculating

799 statistical power as a function of the number of trials as well as the number of participants,

800 and this might be worth consulting to guide the research design process (Baker et al., 2021).

We did not discuss continuous-time EHA, nor continuous-time SAT analysis. As indicated by Allison (2010), learning discrete-time EHA methods first will help in learning continuous-time methods. Given that RT is typically treated as a continuous variable, it is possible that continuous-time methods will ultimately prevail. However, they require much more data to estimate the continuous-time hazard (rate) function well. Thus, by trading a bit of temporal resolution for a lower number of trials, discrete-time methods seem ideal for dealing with typical psychological time-to-event data sets for which there are less than ~200 trials per condition per experiment.

## 5.5 Extensions

The hazard models in this tutorial assume that there is one event of interest. For RT data, this button-press event constitutes a single transition between an “idle” state and a “responded” state. However, in certain situations, more than one event of interest might exist. For example, in a medical or health-related context, an individual might transition back and forth between a “healthy” state and a “depressed” state, before being absorbed into a final “death” state. When you have data on the timing of these transitions, one can apply multi-state hazard models, which generalize EHA to transitions between three or more states (Steele, Goldstein, & Browne, 2004). Also, the predictor variables in this tutorial are time-invariant, i.e., their value did not change over the course of a trial. Thus, another extension is to include time-varying predictors, i.e., predictors whose value can change across the time bins within a trial (Allison, 2010). For example, when gaze position is tracked during a visual search trial, the gaze-target distance will vary during a trial when the eyes move around before a manual response is given; shorter gaze-target distances should be associated with a higher hazard of response occurrence. Note that the effect of a time-varying predictor (e.g., an occipital EEG signal) can itself vary over time.

825

## 6. Conclusions

826       Estimating the temporal distributions of RT and accuracy provide a rich source of  
827      information on the time course of cognitive processing, which have been largely  
828      undervalued in the history of experimental psychology and cognitive neuroscience. We hope  
829      that by providing a set of hands-on, step-by-step tutorials, which come with custom-built  
830      and freely available code, researchers will feel more comfortable embracing EHA and  
831      investigating the temporal profile of cognitive states. On a broader level, we think that  
832      wider adoption of such approaches will have a meaningful impact on the inferences drawn  
833      from data, as well as the development of theories regarding the structure of cognition.

834

**Author contributions**

835       Conceptualization: S. Panis and R. Ramsey; Software: S. Panis and R. Ramsey;  
836      Writing - Original Draft Preparation: S. Panis; Writing - Review & Editing: S. Panis and  
837      R. Ramsey; Supervision: R. Ramsey.

838

**Conflicts of Interest**

839       The author(s) declare that there were no conflicts of interest with respect to the  
840      authorship or the publication of this article.

841

**Prior versions**

842       All of the submitted manuscript and Supplemental Material was previously posted to  
843      a preprint archive: <https://doi.org/10.31234/osf.io/57bh6>

844

**Supplemental Material**

845

**Disclosures****846 Data, materials, and online resources**

847       Link to public archive:  
848      [https://github.com/sven-panis/Tutorial\\_Event\\_History\\_Analysis](https://github.com/sven-panis/Tutorial_Event_History_Analysis)  
849       Supplemental Material: Panis\_Ramsey\_suppl\_material.pdf

**850 Ethical approval**

851       Ethical approval was not required for this tutorial in which we reanalyze existing  
852      data sets.

853

## References

- 854 Allison, P. D. (1982). Discrete-Time Methods for the Analysis of Event Histories.  
855       *Sociological Methodology*, 13, 61. <https://doi.org/10.2307/270718>
- 856 Allison, P. D. (2010). *Survival analysis using SAS: A practical guide* (2. ed). Cary, NC:  
857       SAS Press.
- 858 Aust, F. (2019). *Citr: 'RStudio' add-in to insert markdown citations*. Retrieved from  
859       <https://github.com/crsh/citr>
- 860 Aust, F., & Barth, M. (2024a). *papaja: Prepare reproducible APA journal articles with R*  
861       *Markdown*. <https://doi.org/10.32614/CRAN.package.papaja>
- 862 Aust, F., & Barth, M. (2024b). *papaja: Prepare reproducible APA journal articles with R*  
863       *Markdown*. <https://doi.org/10.32614/CRAN.package.papaja>
- 864 Barth, M. (2023). *tinylabes: Lightweight variable labels*. Retrieved from  
865       <https://cran.r-project.org/package=tinylabes>
- 866 Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects  
867       models using lme4. *Journal of Statistical Software*, 67(1), 1–48.  
868       <https://doi.org/10.18637/jss.v067.i01>
- 869 Bates, D., Maechler, M., & Jagan, M. (2024). *Matrix: Sparse and dense matrix classes and*  
870       *methods*. Retrieved from <https://Matrix.R-forge.R-project.org>
- 871 Bengtsson, H. (2021). A unifying framework for parallel and distributed processing in r  
872       using futures. *The R Journal*, 13(2), 208–227. <https://doi.org/10.32614/RJ-2021-048>
- 873 Blossfeld, H.-P., & Rohwer, G. (2002). *Techniques of event history modeling: New*  
874       *approaches to causal analysis*, 2nd ed (pp. x, 310). Mahwah, NJ, US: Lawrence  
875       Erlbaum Associates Publishers.
- 876 Box-Steffensmeier, J. M. (2004). Event history modeling: A guide for social scientists.  
877       Cambridge: University Press.
- 878 Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan.  
879       *Journal of Statistical Software*, 80(1), 1–28. <https://doi.org/10.18637/jss.v080.i01>

- 880 Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms.  
881     *The R Journal*, 10(1), 395–411. <https://doi.org/10.32614/RJ-2018-017>
- 882 Bürkner, P.-C. (2021). Bayesian item response modeling in R with brms and Stan. *Journal*  
883     *of Statistical Software*, 100(5), 1–54. <https://doi.org/10.18637/jss.v100.i05>
- 884 DeBruine, L. M., & Barr, D. J. (2021). Understanding Mixed-Effects Models Through  
885     Data Simulation. *Advances in Methods and Practices in Psychological Science*, 4(1),  
886     2515245920965119. <https://doi.org/10.1177/2515245920965119>
- 887 Eddelbuettel, D., & Balamuta, J. J. (2018). Extending R with C++: A Brief Introduction  
888     to Rcpp. *The American Statistician*, 72(1), 28–36.  
889     <https://doi.org/10.1080/00031305.2017.1375990>
- 890 Eddelbuettel, D., & François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal*  
891     *of Statistical Software*, 40(8), 1–18. <https://doi.org/10.18637/jss.v040.i08>
- 892 Gabry, J., Češnovar, R., Johnson, A., & Broder, S. (2024). *Cmdstanr: R interface to*  
893     *'CmdStan'*. Retrieved from <https://github.com/stan-dev/cmdstanr>
- 894 Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2019). Visualization  
895     in bayesian workflow. *J. R. Stat. Soc. A*, 182, 389–402.  
896     <https://doi.org/10.1111/rssa.12378>
- 897 Gelman, A., Hill, J., & Vehtari, A. (2020). Regression and Other Stories.  
898     [https://www.cambridge.org/highereducation/books/regression-and-other-](https://www.cambridge.org/highereducation/books/regression-and-other-stories/DD20DD6C9057118581076E54E40C372C)  
899     stories/DD20DD6C9057118581076E54E40C372C; Cambridge University Press.  
900     <https://doi.org/10.1017/9781139161879>
- 901 Gelman, A., Vehtari, A., Simpson, D., Margossian, C. C., Carpenter, B., Yao, Y., ...  
902     Modrák, M. (2020). *Bayesian Workflow*. arXiv.  
903     <https://doi.org/10.48550/arXiv.2011.01808>
- 904 Girard, J. (2024). *Standist: What the package does (one line, title case)*. Retrieved from  
905     <https://github.com/jmgirard/standist>
- 906 Grolemund, G., & Wickham, H. (2011). Dates and times made easy with lubridate.

- 907        *Journal of Statistical Software*, 40(3), 1–25. Retrieved from  
908        <https://www.jstatsoft.org/v40/i03/>
- 909        Halley, E. (1693). VI. An estimate of the degrees of the mortality of mankind; drawn from  
910        curious tables of the births and funerals at the city of breslaw; with an attempt to  
911        ascertain the price of annuities upon lives. *Philosophical Transactions of the Royal*  
912        *Society of London*, 17(196), 596–610. <https://doi.org/10.1098/rstl.1693.0007>
- 913        Heiss, A. (2021, November 10). A Guide to Correctly Calculating Posterior Predictions  
914        and Average Marginal Effects with Multilievel Bayesian Models.  
915        <https://doi.org/10.59350/wbn93-edb02>
- 916        Hosmer, D. W., Lemeshow, S., & May, S. (2011). *Applied Survival Analysis: Regression*  
917        *Modeling of Time to Event Data* (2nd ed). Hoboken: John Wiley & Sons.
- 918        Ince, R. A., Paton, A. T., Kay, J. W., & Schyns, P. G. (2021). Bayesian inference of  
919        population prevalence. *eLife*, 10, e62461. <https://doi.org/10.7554/eLife.62461>
- 920        Kantowitz, B. H., & Pachella, R. G. (2021). The Interpretation of Reaction Time in  
921        Information-Processing Research 1. *Human Information Processing*, 41–82.  
922        <https://doi.org/10.4324/9781003176688-2>
- 923        Kay, M. (2024). *tidybayes: Tidy data and geoms for Bayesian models*.  
924        <https://doi.org/10.5281/zenodo.1308151>
- 925        Kurz, A. S. (2023a). *Applied longitudinal data analysis in brms and the tidyverse* (version  
926        0.0.3). Retrieved from <https://bookdown.org/content/4253/>
- 927        Kurz, A. S. (2023b). *Statistical rethinking with brms, ggplot2, and the tidyverse: Second*  
928        *edition* (version 0.4.0). Retrieved from <https://bookdown.org/content/4857/>
- 929        Lakens, D. (2022). Sample Size Justification. *Collabra: Psychology*, 8(1), 33267.  
930        <https://doi.org/10.1525/collabra.33267>
- 931        Landes, J., Engelhardt, S. C., & Pelletier, F. (2020). An introduction to event history  
932        analyses for ecologists. *Ecosphere*, 11(10), e03238. <https://doi.org/10.1002/ecs2.3238>
- 933        Makeham, W. M. (1860). *On the Law of Mortality and the Construction of Annuity Tables*.

- 934        The Assurance Magazine, and Journal of the Institute of Actuaries.
- 935        McElreath, R. (2020). *Statistical Rethinking: A Bayesian Course with Examples in R and*  
936        *STAN* (2nd ed.). New York: Chapman and Hall/CRC.  
937        <https://doi.org/10.1201/9780429029608>
- 938        Müller, K., & Wickham, H. (2023). *Tibble: Simple data frames*. Retrieved from  
939        <https://CRAN.R-project.org/package=tibble>
- 940        Nelder, J. A. (1999). From Statistics to Statistical Science. *Journal of the Royal Statistical*  
941        *Society. Series D (The Statistician)*, 48(2), 257–269. Retrieved from  
942        <https://www.jstor.org/stable/2681191>
- 943        Neuwirth, E. (2022). *RColorBrewer: ColorBrewer palettes*. Retrieved from  
944        <https://CRAN.R-project.org/package=RColorBrewer>
- 945        Panis, S. (2020). How can we learn what attention is? Response gating via multiple direct  
946        routes kept in check by inhibitory control processes. *Open Psychology*, 2(1), 238–279.  
947        <https://doi.org/10.1515/psych-2020-0107>
- 948        Panis, S., Moran, R., Wolkersdorfer, M. P., & Schmidt, T. (2020). Studying the dynamics  
949        of visual search behavior using RT hazard and micro-level speed–accuracy tradeoff  
950        functions: A role for recurrent object recognition and cognitive control processes.  
951        *Attention, Perception, & Psychophysics*, 82(2), 689–714.  
952        <https://doi.org/10.3758/s13414-019-01897-z>
- 953        Panis, S., Schmidt, F., Wolkersdorfer, M. P., & Schmidt, T. (2020). Analyzing Response  
954        Times and Other Types of Time-to-Event Data Using Event History Analysis: A Tool  
955        for Mental Chronometry and Cognitive Psychophysiology. *I-Perception*, 11(6),  
956        2041669520978673. <https://doi.org/10.1177/2041669520978673>
- 957        Panis, S., & Schmidt, T. (2016). What Is Shaping RT and Accuracy Distributions? Active  
958        and Selective Response Inhibition Causes the Negative Compatibility Effect. *Journal of*  
959        *Cognitive Neuroscience*, 28(11), 1651–1671. [https://doi.org/10.1162/jocn\\_a\\_00998](https://doi.org/10.1162/jocn_a_00998)
- 960        Panis, S., & Schmidt, T. (2022). When does “inhibition of return” occur in spatial cueing

- 961 tasks? Temporally disentangling multiple cue-triggered effects using response history  
962 and conditional accuracy analyses. *Open Psychology*, 4(1), 84–114.  
963 <https://doi.org/10.1515/psych-2022-0005>
- 964 Panis, S., Torfs, K., Gillebert, C. R., Wagemans, J., & Humphreys, G. W. (2017).  
965 Neuropsychological evidence for the temporal dynamics of category-specific naming.  
966 *Visual Cognition*, 25(1-3), 79–99. <https://doi.org/10.1080/13506285.2017.1330790>
- 967 Panis, S., & Wagemans, J. (2009). Time-course contingencies in perceptual organization  
968 and identification of fragmented object outlines. *Journal of Experimental Psychology:  
969 Human Perception and Performance*, 35(3), 661–687.  
970 <https://doi.org/10.1037/a0013547>
- 971 Pargent, F., Koch, T. K., Kleine, A.-K., Lermer, E., & Gaube, S. (2024). A Tutorial on  
972 Tailored Simulation-Based Sample-Size Planning for Experimental Designs With  
973 Generalized Linear Mixed Models. *Advances in Methods and Practices in Psychological  
974 Science*, 7(4), 25152459241287132. <https://doi.org/10.1177/25152459241287132>
- 975 Pedersen, T. L. (2024). *Patchwork: The composer of plots*. Retrieved from  
976 <https://patchwork.data-imaginist.com>
- 977 Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in s and s-PLUS*. New York:  
978 Springer. <https://doi.org/10.1007/b98882>
- 979 R Core Team. (2024). *R: A language and environment for statistical computing*. Vienna,  
980 Austria: R Foundation for Statistical Computing. Retrieved from  
981 <https://www.R-project.org/>
- 982 Ripley, B., Venables, B., Bates, D. M., ca 1998), K. H. (partial. port, ca 1998), A. G.  
983 (partial. port, & polr), D. F. (support. functions for. (2024). *MASS: Support Functions  
984 and Datasets for Venables and Ripley's MASS*.
- 985 Singer, J. D., & Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling  
986 Change and Event Occurrence*. Oxford, New York: Oxford University Press.
- 987 Stan Development Team. (2020). *StanHeaders: Headers for the R interface to Stan*.

- 988        Retrieved from <https://mc-stan.org/>
- 989    Stan Development Team. (2024). *RStan: The R interface to Stan*. Retrieved from  
990        <https://mc-stan.org/>
- 991    Steele, F., Goldstein, H., & Browne, W. (2004). A general multilevel multistate competing  
992        risks model for event history data, with an application to a study of contraceptive use  
993        dynamics. *Statistical Modelling*, 4(2), 145–159.  
994        <https://doi.org/10.1191/1471082X04st069oa>
- 995    Teachman, J. D. (1983). Analyzing social processes: Life tables and proportional hazards  
996        models. *Social Science Research*, 12(3), 263–301.  
997        [https://doi.org/10.1016/0049-089X\(83\)90015-7](https://doi.org/10.1016/0049-089X(83)90015-7)
- 998    Tong, C. (2019). Statistical Inference Enables Bad Science; Statistical Thinking Enables  
999        Good Science. *The American Statistician*, 73(sup1), 246–261.  
1000        <https://doi.org/10.1080/00031305.2018.1518264>
- 1001    Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics.  
1002        *Acta Psychologica*, 41(1), 67–85. [https://doi.org/10.1016/0001-6918\(77\)90012-9](https://doi.org/10.1016/0001-6918(77)90012-9)
- 1003    Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New  
1004        York. Retrieved from <https://ggplot2.tidyverse.org>
- 1005    Wickham, H. (2023a). *Forcats: Tools for working with categorical variables (factors)*.  
1006        Retrieved from <https://forcats.tidyverse.org/>
- 1007    Wickham, H. (2023b). *Stringr: Simple, consistent wrappers for common string operations*.  
1008        Retrieved from <https://stringr.tidyverse.org>
- 1009    Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., ...  
1010        Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43),  
1011        1686. <https://doi.org/10.21105/joss.01686>
- 1012    Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *Dplyr: A  
1013        grammar of data manipulation*. Retrieved from <https://dplyr.tidyverse.org>
- 1014    Wickham, H., & Henry, L. (2023). *Purrr: Functional programming tools*. Retrieved from

- 1015 https://purrr.tidyverse.org/
- 1016 Wickham, H., Hester, J., & Bryan, J. (2024). *Readr: Read rectangular text data*. Retrieved  
1017 from https://readr.tidyverse.org
- 1018 Wickham, H., Vaughan, D., & Girlich, M. (2024). *Tidyr: Tidy messy data*. Retrieved from  
1019 https://tidyr.tidyverse.org
- 1020 Winter, B. (2019). *Statistics for Linguists: An Introduction Using R*. New York:  
1021 Routledge. <https://doi.org/10.4324/9781315165547>
- 1022 Wolkersdorfer, M. P., Panis, S., & Schmidt, T. (2020). Temporal dynamics of sequential  
1023 motor activation in a dual-prime paradigm: Insights from conditional accuracy and  
1024 hazard functions. *Attention, Perception, & Psychophysics*, 82(5), 2581–2602.  
1025 <https://doi.org/10.3758/s13414-020-02010-5>