

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 The authors made the following contributions. Sven Panis: Conceptualization,  
10 Writing - Original Draft Preparation, Writing - Review & Editing; Richard Ramsey:  
11 Conceptualization, Writing - Review & Editing, Supervision.

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

14

## Abstract

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

37       *Keywords:* response times, event history analysis, Bayesian multi-level regression  
38       models, experimental psychology, cognitive psychology

39       Word count: X

40

## 1. Introduction

### 41 1.1 Motivation and background context: Comparing means versus 42 distributional shapes

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

56 As a simple illustration, Figure 1 shows the results of several simulated RT data sets,  
57 which show how mean-average comparisons between two conditions can conceal the shape  
58 of the underlying RT distributions. For instance, in examples 1-3, mean RT is always  
59 comparable between two conditions, while the distributions differ (Figure 1, left). In  
60 contrast, in examples 4-6, mean RT is lower in condition 2 compared to condition 1, but  
61 the RT distributions differ in each case (Figure 1, right). Therefore, a comparison of means  
62 would lead to a similar conclusion in examples 1-3, as well as examples 4-6, whereas a  
63 comparison of the distributions would lead to a different conclusion in every case.

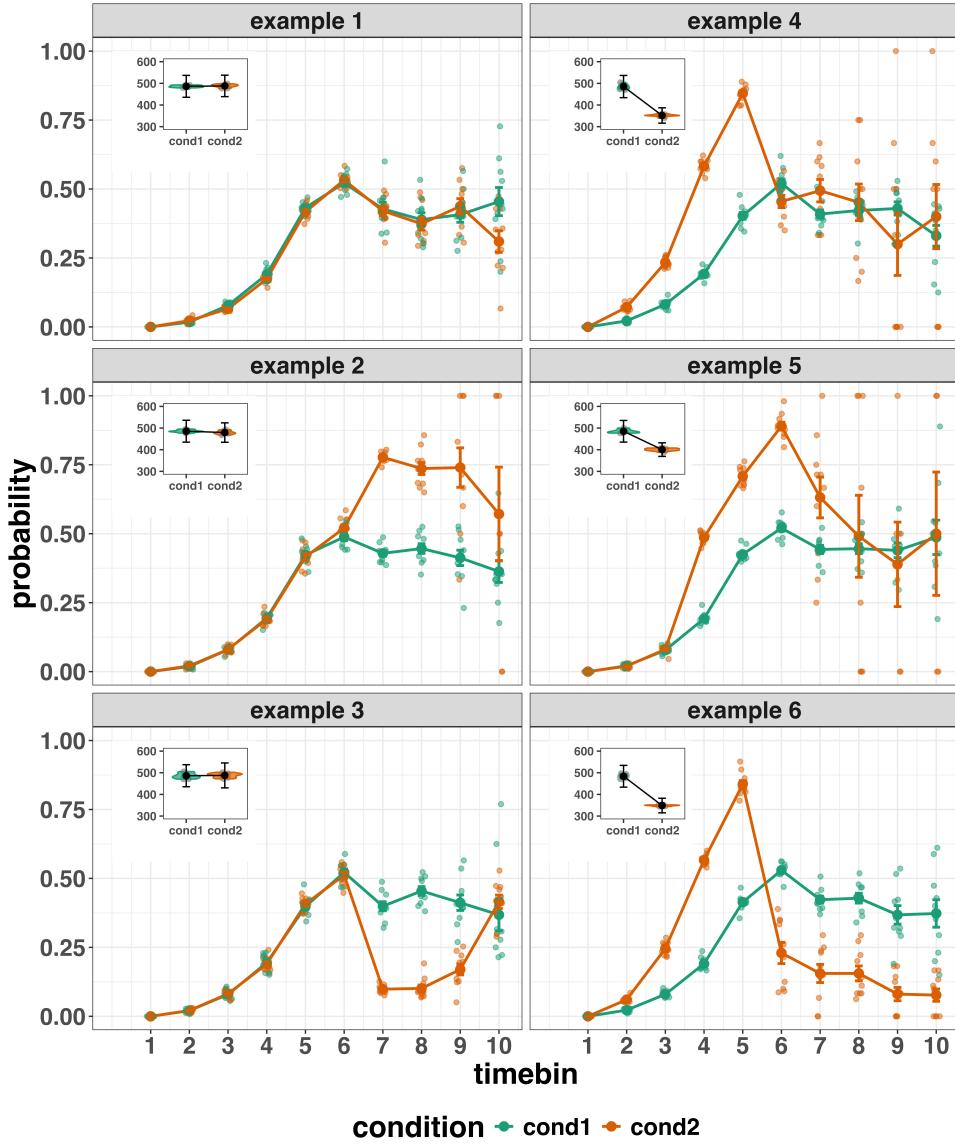


Figure 1. Means versus distributional shapes for six different simulated data set examples.

The first second after stimulus onset is divided in ten bins of 100 ms. Timebin indicates the bin rank. The first bin is (0,100], the last bin is (900,1000]. For our purposes here, it is enough to know that the distributions plotted represent the probability of an event occurring in that timebin, given that it has not yet occurred. Insets show mean response time per condition.

65 data across trials, a distributional approach offers the possibility to reveal the time course  
66 of psychological states. As such, the approach permits different kinds of questions to be  
67 asked, different inferences to be made, and it holds the potential to discriminate between  
68 different theoretical accounts of psychological and/or brain-based processes. For example,  
69 the distributions in Example 4 show that the effect starts between 100 and 200 ms (in  
70 timebin 2) and is gone when the waiting time reaches 500 ms or more. In contrast, in  
71 Example 5, the effect starts around 300 ms and is gone by 700 ms. And in the Example 6,  
72 the effect reverses between 500 and 600 ms. What kind of theory or theories could account  
73 for such effects? Are there new auxiliary assumptions that theories need to adopt? And are  
74 there new experiments that need to be performed to test the novel predictions that follow  
75 from these analyses? As we show later using published examples, for many psychological  
76 questions, such “temporal states” information can be theoretically meaningful by leading to  
77 more fine-grained understanding of psychological processes, as well as adding a relatively  
78 under-used dimension – the passage of time – to the theory building toolkit.

79 From a historical perspective, it is worth noting that the development of analytical  
80 tools that can estimate or predict whether and when events will occur is not a new  
81 endeavour. Indeed, hundreds of years ago, analytical methods were developed to predict  
82 the duration of time until people died (e.g., Halley, 1693; Makeham, 1860). The same logic  
83 has been applied to psychological time-to-event data, as previously demonstrated (Panis,  
84 Schmidt, et al., 2020).

## 85 1.2 Aims and structure of the paper

86 In this paper, we focus on a distributional method for time-to-event data known as  
87 discrete-time Event History Analysis (EHA), a.k.a. survival analysis, hazard analysis,  
88 duration analysis, failure-time analysis, and transition analysis (Singer & Willett, 2003).  
89 We hope to show the value of EHA for knowledge and theory building in cognitive  
90 psychology and related areas of research, such as cognitive neuroscience. Most importantly,

91 we provide tutorials that provide step-by-step code and instructions in the hope that we  
92 can enable others to use EHA in a more routine, efficient and effective manner.

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

98 We provide six different tutorials, which are written in the R programming language  
99 and publicly available on our Github and the Open Science Framework (OSF) pages, along  
100 with all of the other code and material associated with the project. The tutorials provide  
101 hands-on, concrete examples of key parts of the analytical process, so that others can apply  
102 EHA to their own time-to-event data sets. Each tutorial is provided as an RMarkdown file,  
103 so that others can download and adapt the code to fit their own purposes. Additionally,  
104 each tutorial is made available as a .html file, so that it can be viewed by any web browser,  
105 and thus available to those that do not use R. Finally, the manuscript itself is written in R  
106 using the papaja package (Aust & Barth, 2024), which makes it computationally  
107 reproducible, in terms of the underlying data and figures.

108 In Tutorial 1a, we illustrate how to process or “wrangle” a previously published RT +  
109 accuracy data set to calculate descriptive statistics when there is one independent variable.  
110 The descriptive statistics are plotted, and we comment on their interpretation. In Tutorial  
111 1b we provide a generalisation of this approach to illustrate how one can calculate the  
112 descriptive statistics when using a more complex design, such as when there are two  
113 independent variables.

114 In Tutorial 2a, we illustrate how one can fit Bayesian multi-level regression models to  
115 RT data using the R package brms. We also perform prior predictive checks, compare  
116 models, and interpret the plots of the predicted hazard functions for the selected model,

117 and the posterior distributions of our contrasts of interest. In Tutorial 2b we fit Bayesian  
118 multi-level regression models to *timed* accuracy data to perform a micro-level  
119 speed-accuracy tradeoff (SAT) analysis, which complements the EHA of RT data for choice  
120 RT data.

121 In Tutorial 3a, we shortly illustrate how to fit similar multilevel regression models for  
122 RT data in a frequentist framework using the R package lme4. We then briefly compare  
123 and contrast these inferential frameworks when applied to EHA. In Tutorial 3b, we  
124 illustrate how to perform the SAT analysis in a frequentist framework.

125 In tutorial 4, we illustrate one approach to planning how much data to collect in an  
126 experiment using EHA. We use data simulation techniques to vary sample size and trial  
127 count per condition until a certain degree of statistical power or precision is reached.  
128 [[more to come here, once we have written the tutorial]].

129 In summary, even though EHA is a widely used statistical tool and there already exist  
130 many excellent reviews (e.g., Blossfeld & Rohwer, 2002; Box-Steffensmeier, 2004; Hosmer,  
131 Lemeshow, & May, 2011; Teachman, 1983) and tutorials (e.g., Allison, 2010; Landes,  
132 Engelhardt, & Pelletier, 2020) on its general use-cases, we are not aware of any tutorials  
133 that are aimed at psychological time-to-event data, and which provide worked examples of  
134 the key data processing and multi-level regression modelling steps. Therefore, our ultimate  
135 goal is twofold: first, we want to convince readers of the many benefits of using EHA when  
136 dealing with time-to-event data with a focus on psychological time-to-event data, and  
137 second, we want to provide a set of practical tutorials, which provide step-by-step  
138 instructions on how you actually perform a discrete-time EHA on time-to-event data such  
139 as RT data, as well as a complementary discrete-time SAT analysis on timed accuracy data.

**2. A brief introduction to event history analysis**

For a comprehensive background context to EHA, we recommend several excellent textbooks (Allison, 2010; Singer & Willett, 2003). Likewise, for a general introduction to understanding regression equations, we recommend several excellent textbooks (Gelman, Hill, & Vehtari, 2020; Winter, 2019). Our focus here is not on providing a detailed account of the underlying regression equations, since this topic has been comprehensively covered many times before. Instead, we want to provide an intuition regarding how EHA works in general, as well as in the context of experimental psychology. As such, we only supply regression equations in part D of the supplementary material.

**2.1 Basic features of event history analysis**

To apply EHA, one must be able to:

1. define an event of interest that represents a qualitative change that can be situated in time (e.g., a button press, a saccade onset, a fixation offset, etc.);
2. define time point zero (e.g., target stimulus onset, fixation onset, etc.);
3. measure the passage of time between time point zero and event occurrence in discrete or continuous time units.

In EHA, the definition of hazard and the type of models employed depend on whether one is using continuous or discrete time units. Since our focus here is on hazard models that use discrete time units, we describe that approach. After dividing time in discrete, contiguous time bins indexed by  $t$  (e.g.,  $t = 1:10$  timebins), let  $RT$  be a discrete random variable denoting the rank of the time bin in which a particular person's response occurs in a particular experimental condition. For example, the first response might occur at 546 ms and it would be in timebin 6 (any RTs from 501 ms to 600 ms).

163 Discrete-time EHA focuses on the discrete-time hazard function of event occurrence

164 and the discrete-time survivor function (Figure 2). The equations that define both of these

165 functions are reported in part A of the supplementary material. The discrete-time hazard

166 function gives you, for each time bin, the probability that the event occurs (sometime) in

167 bin  $t$ , given that the event does not occur in previous bins. In other words, it reflects the

168 instantaneous likelihood that the event occurs in the current bin, given that it has not yet

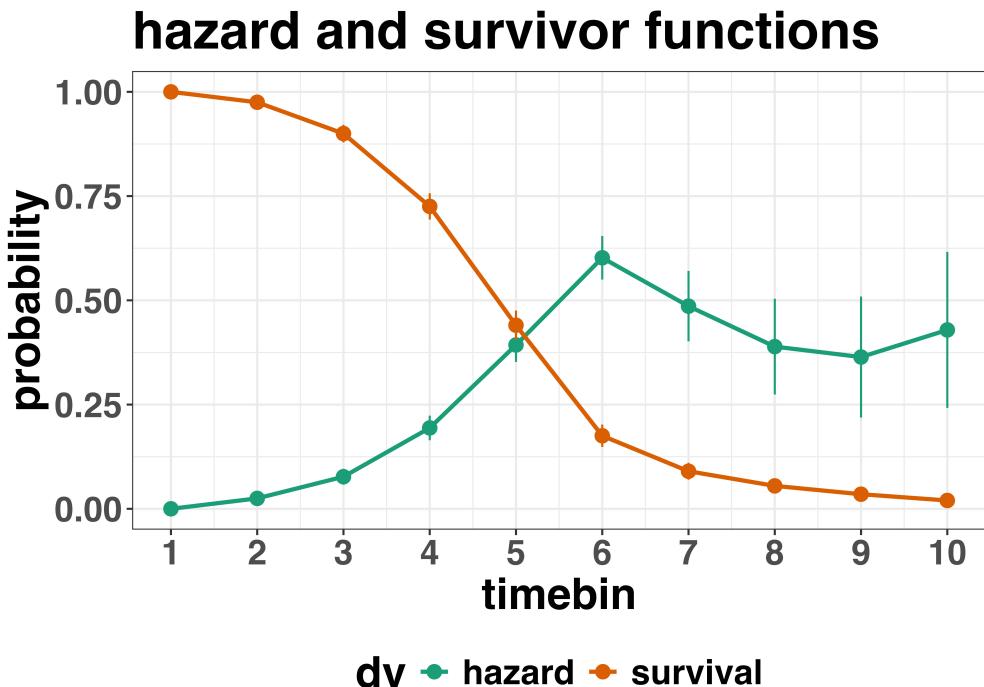
169 occurred in the past, i.e., in one of the prior bins. In contrast, the discrete-time survivor

170 function cumulates the bin-by-bin risks of event *nonoccurrence* to obtain the survival

171 probability, the probability that the event occurs after bin  $t$ . In other words, the survivor

172 function gives you for each time bin the likelihood that the event occurs in the future, i.e.,

173 in one of the subsequent timebins.



*Figure 2.* Discrete-time hazard and survivor functions. Discrete time-to-event data were simulated for 200 trials of 1 experimental condition. While the hazard function is the vehicle for inferring the time course of cognitive processes, the survival probability  $S(t-1)$  can help to qualify or provide context to the interpretation of the hazard probability  $h(t)$ . For example, the high hazard of  $.60 = h(t=6)$  is experienced only by 44 percent of the trials, as  $S(t=5) = .44$ . Because the survivor function is a decreasing function of time, the error bars in later parts of the hazard function will always be wider and less precise compared to earlier parts.

<sup>174</sup> **2.2 Benefits of event history analysis**

<sup>175</sup> Statisticians and mathematical psychologists recommend focusing on the hazard  
<sup>176</sup> function when analyzing time-to-event data for various reasons. We do not cover these  
<sup>177</sup> benefits in detail here, as these are more general topics that have been covered elsewhere in  
<sup>178</sup> textbooks. Instead, we briefly summarise list the benefits below, and refer the reader to  
<sup>179</sup> section F of Supplementary Materials for more detailed coverage of the benefits. A  
<sup>180</sup> summary of the benefits are as follows:

- 181 1. Hazard functions are more diagnostic than density functions when one is interested in  
182 studying the detailed shape of a RT distribution (Holden et al., 2009).
- 183 2. RT distributions may differ from each other in multiple ways, and hazard functions  
184 allow one to capture these differences which mean-average comparisons may conceal  
185 (Townsend, 1990).
- 186 3. EHA takes account of more of the data collected in a typical speeded response  
187 experiment, by virtue of not discarding right-censored observations. Trials with very  
188 long RTs are not discarded, but instead contribute to the risk set in each time bin  
189 (see below).
- 190 4. Hazard modeling allows one to incorporate time-varying explanatory covariates, such  
191 as heart rate, electroencephalogram (EEG) signal amplitude, gaze location, etc.  
192 (Allison, 2010). This is useful for linking physiological effects to behavioral effects  
193 when performing cognitive psychophysiology (Meyer, Osman, Irwin, & Yantis, 1988).
- 194 5. EHA can help to solve the problem of model mimicry, i.e., the fact that different  
195 computational models can often predict the same mean RTs as observed in the  
196 empirical data, but not necessarily the detailed shapes of the empirical RT hazard  
197 distributions. As such, EHA can be a tool to help distinguish between competing  
198 theories of cognition and brain function.

### 199 2.3 Event history analysis in the context of experimental psychology

200 To make EHA more relevant to researchers studying cognitive psychology and

201 cognitive neuroscience, in this section we provide a relevant worked example and consider  
202 implications that are relevant to that domain of research.

203 **2.3.1 A worked example.** In the context of experimental psychology, it is

204 common for participants to be presented with either a 1-button detection task or a

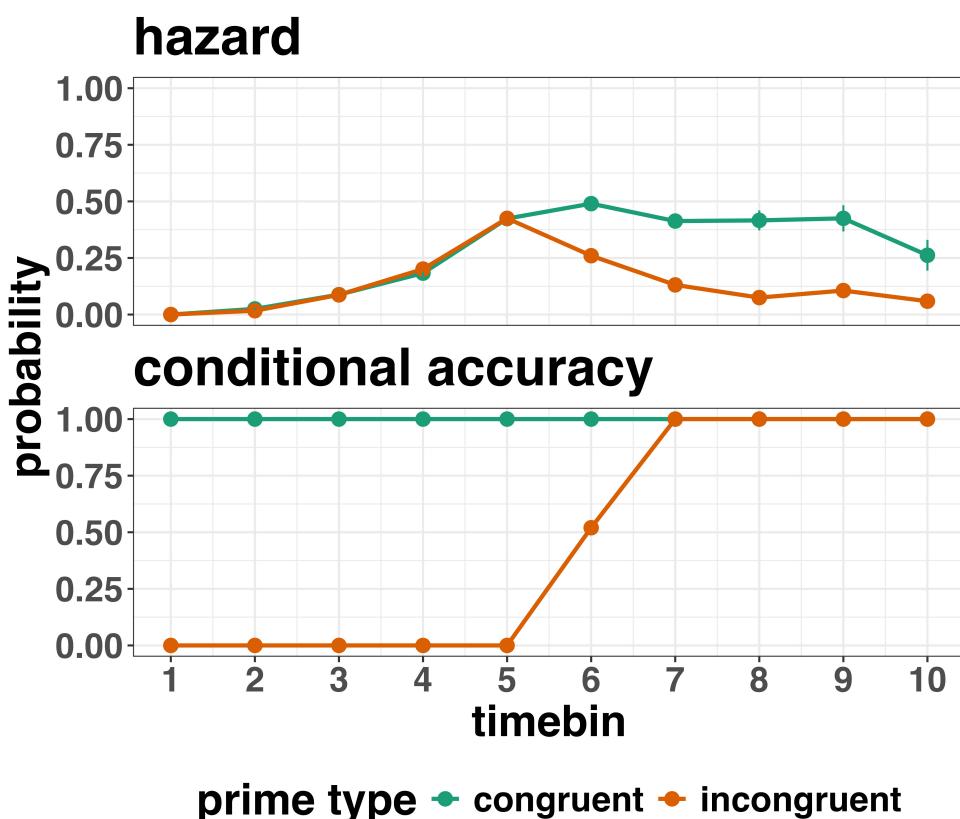
discrimination task. For example, a task may involve choosing between two response options with only one of them being correct. For such two-choice RT data, the discrete-time EHA of the RT data (hazard and survivor functions) can be extended with a discrete-time SAT analysis of the timed accuracy data. Specifically, the hazard function of event occurrence can be extended with the discrete-time conditional accuracy function, which gives you the probability that a response is correct given that it is emitted in time bin  $t$  (Allison, 2010; Kantowitz & Pachella, 2021; Wickelgren, 1977). We refer to this extended (hazard + conditional accuracy) analysis for choice RT data as EHA/SAT.

Integrating results between hazard and conditional accuracy functions for choice RT data can be informative for understanding psychological processes. To illustrate, we consider a hypothetical choice RT example that is inspired by real data (Panis & Schmidt, 2016), but simplified to make the main point clearer (Figure 3). In a standard priming paradigm, there is a prime stimulus (e.g., an arrow pointing left or right) followed by a target stimulus (another arrow pointing left or right). The prime can then be congruent or incongruent with the target.

Figure 3 shows that the early upswing in hazard is equal for both priming conditions, and that early emitted responses are always correct in the congruent condition and always incorrect in the incongruent condition. These results show that for short waiting times (< bin 6), responses always follow the prime (and not the target, as instructed). During timebin 6 the target-triggered response channel is activated and causes response competition –  $ca(6) = .5$  – and a lower hazard probability in the incongruent condition. For waiting times of 600 ms or more, the hazard of response occurrence is lower in incongruent compared to congruent trials, and all responses emitted in these late bins are correct.

This joint pattern of results is interesting because it can provide meaningfully different conclusions about psychological processes compared to conventional analyses, such as computing mean-average RT and accuracy across trials. Mean-average RT would only

<sup>231</sup> represent the overall ability of cognition to overcome interference, on average, across trials.  
<sup>232</sup> For instance, if mean-average RT was higher in incongruent than congruent trials, one may  
<sup>233</sup> conclude that cognitive mechanisms that support interference control are working as  
<sup>234</sup> expected across trials, and are indexed by each recorded response. But such a conclusion is  
<sup>235</sup> not supported when the effects are explored over a timeline. Instead, the psychological  
<sup>236</sup> conclusion is much more nuanced and suggests that multiple states start, stop and possibly  
<sup>237</sup> interact over a particular temporal window.



*Figure 3.* Discrete-time hazard and conditional accuracy functions. Discrete time-to-event and conditional accuracy data were simulated for 2000 trials for each of two priming conditions (congruent and incongruent prime stimuli). Bin width equals 100 ms.

<sup>238</sup> Unlocking the temporal states of cognitive processes can be revealing for theory  
<sup>239</sup> development and the understanding of basic psychological processes. Possibly more

240 importantly, however, is that it simultaneously opens the door to address many new and  
241 previously unanswered questions. Do all participants show similar temporal states or are  
242 there individual differences? Do such individual differences extend to those individuals that  
243 have been diagnosed with some form of psychopathology? How do temporal states relate to  
244 brain-based mechanisms that might be studied using other methods from cognitive  
245 neuroscience? And how much of theory in cognitive psychology would be in need of  
246 revision if mean-average comparisons were supplemented with a temporal states approach?

247 **2.3.2 Implications for designing experiments.** Performing EHA in  
248 experimental psychology has implications for how experiments are designed. Indeed, if  
249 trials are categorised as a function of when responses occur, then each timebin will only  
250 include a subset of the total number of trials. For example, let's consider an experiment  
251 where each participant performs 2 conditions and there are 100 trial repetitions per  
252 condition. Those 100 trials must be distributed in some manner across the chosen number  
253 of bins.

254 In such experimental designs, since the number of trials per condition are spread  
255 across bins, it is important to have a relatively large number of trial repetitions per  
256 participant and per condition. Accordingly, experimental designs using this approach  
257 typically focus on factorial, within-subject designs, in which a large number of observations  
258 are made on a relatively small number of participants (so-called small- $N$  designs). This  
259 approach emphasizes the precision and reproducibility of data patterns at the individual  
260 participant level to increase the inferential validity of the design (Baker et al., 2021; Smith  
261 & Little, 2018).

262 In contrast to the large- $N$  design that typically average across many participants  
263 without being able to scrutinize individual data patterns, small- $N$  designs retain crucial  
264 information about the data patterns of individual observers. This can be advantageous  
265 whenever participants differ systematically in their strategies or in the time courses of their  
266 effects, so that averaging them would lead to misleading data patterns. Note that because

267 statistical power derives both from the number of participants and from the number of  
268 repeated measures per participant and condition, small- $N$  designs can still achieve what  
269 are generally considered acceptable levels of statistical power, if they have a sufficient  
270 amount of data overall (Baker et al., 2021; Smith & Little, 2018).

271 **3. An overview of the general analytical workflow**

272 Although the focus is on EHA/SAT, we also want to briefly comment on broader  
273 aspects of our general analytical workflow, which relate more to data science and data  
274 analysis workflows.

275 **3.1 Data science workflow and descriptive statistics**

276 We perform data wrangling following tidyverse principles and a functional  
277 programming approach (Wickham, Çetinkaya-Rundel, & Grolemund, 2023). Functional  
278 programming basically means that you avoid writing your own loops but instead use  
279 functions that have been built and tested by others. In addition, we also supply a set of  
280 custom-built functions, which make the process of data wrangling in the context of  
281 descriptive EHA a lot quicker and more efficient.

282 **3.2 Inferential statistical approach**

283 Our lab adopts a estimation approach to multi-level regression (Kruschke & Liddell,  
284 2018; Winter, 2019), which is heavily influenced by the Bayesian framework as suggested  
285 by Richard McElreath (Kurz, 2023b; McElreath, 2020). We also use a “keep it maximal”  
286 approach to specifying varying (or random) effects (Barr, Levy, Scheepers, & Tily, 2013).  
287 This means that wherever possible we include varying intercepts and slopes per participant  
288 To make inferences, we use two main approaches. We compare models of different  
289 complexity, using information criteria (e.g., WAIC) and cross-validation (e.g., LOO), to

290 evaluate out-of-sample predictive accuracy (McElreath, 2020). We also take the most  
291 complex model and evaluate key parameters of interest using point and interval estimates.

292 **3.3 Implementation**

293 We used R (Version 4.4.0; R Core Team, 2024)<sup>1</sup> for all reported analyses. The  
294 content of the tutorials, in terms of EHA and multi-level regression modelling, is mainly  
295 based on Allison (2010), Singer and Willett (2003), McElreath (2020), Heiss (2021), Kurz  
296 (2023a), and Kurz (2023b).

297 **4. Tutorials**

298 Tutorials 1a and 1b show how to calculate and plot the descriptive statistics of  
299 EHA/SAT when there are one and two independent variables, respectively. Tutorials 2a  
300 and 2b illustrate how to use Bayesian multilevel modeling to fit hazard and conditional  
301 accuracy models, respectively. Tutorials 3a and 3b show how to implement, respectively,  
302 multilevel models for hazard and conditional accuracy in the frequentist framework.

---

<sup>1</sup> We, furthermore, used the R-packages *bayesplot* (Version 1.11.1; Gabry, Simpson, Vehtari, Betancourt, & Gelman, 2019), *brms* (Version 2.21.0; 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 1.7.0; Bates, Maechler, & Jagan, 2024), *nlme* (Version 3.1.166; Pinheiro & Bates, 2000), *papaja* (Version 0.1.2.9000; Aust & Barth, 2023), *patchwork* (Version 1.2.0; Pedersen, 2024), *purrr* (Version 1.0.2; Wickham & Henry, 2023), *RColorBrewer* (Version 1.1.3; Neuwirth, 2022), *Rcpp* (Eddelbuettel & Balamuta, 2018; Version 1.0.12; Eddelbuettel & François, 2011), *readr* (Version 2.1.5; Wickham, Hester, & Bryan, 2024), *RJ-2021-048* (Bengtsson, 2021), *standist* (Version 0.0.0.9000; Girard, 2024), *stringr* (Version 1.5.1; Wickham, 2023b), *tibble* (Version 3.2.1; Müller & Wickham, 2023), *tidybayes* (Version 3.0.6; Kay, 2023), *tidyrr* (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).

303 Additionally, to further simplify the process for other users, the first two tutorials rely on a  
304 set of our own custom functions that make sub-processes easier to automate, such as data  
305 wrangling and plotting functions (see part B in the supplemental material for a list of the  
306 custom functions).

307 Our list of tutorials is as follows:

- 308 • 1a. Wrangle raw data and calculate descriptive stats for one independent variable
- 309 • 1b. Wrangle raw data and calculate descriptive stats for two independent variables
- 310 • 2a. Bayesian multilevel modeling for  $h(t)$
- 311 • 2b. Bayesian multilevel modeling for  $ca(t)$
- 312 • 3a. Frequentist multilevel modeling for  $h(t)$
- 313 • 3b. Frequentist multilevel modeling for  $ca(t)$
- 314 • 4. Simulation and power analysis for planning experiments

315 **4.1 Tutorial 1a: Calculating descriptive statistics using a life table**

316 **4.1.1 Data wrangling aims.** Our data wrangling procedures serve two related  
317 purposes. First, we want to summarise and visualise descriptive statistics that relate to our  
318 main research questions about the time course of psychological processes, using a life table.  
319 A life table includes for each time bin, the risk set (i.e., the number of trials that are  
320 event-free at the start of the bin), the number of observed events, and the estimates of  
321  $h(t)$ ,  $S(t)$ ,  $P(t)$ , possibly  $ca(t)$ , and their estimated standard errors (se).

322 Second, we want to produce two different data sets that can each be submitted to  
323 different types of inferential modelling approaches. The two types of data structure we  
324 label as ‘person-trial’ data and ‘person-trial-bin’ data. The ‘person-trial’ data (Table 1)  
325 will be familiar to most researchers who record behavioural responses from participants, as  
326 it represents the measured RT and accuracy per trial within an experiment. This data set  
327 is used when fitting conditional accuracy models (Tutorials 2b and 3b).

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.

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

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.

337       **4.1.2 A real data wrangling example.** To illustrate how to quickly set up life  
 338       tables for calculating the descriptive statistics (functions of discrete time), we use a  
 339       published data set on masked response priming from Panis and Schmidt (2016). In their  
 340       first experiment, Panis and Schmidt (2016) presented a double arrow for 94 ms that  
 341       pointed left or right as the target stimulus with an onset at time point zero in each trial.  
 342       Participants had to indicate the direction in which the double arrow pointed using their  
 343       corresponding index finger, within 800 ms after target onset. Response time and accuracy  
 344       were recorded on each trial. Prime type (blank, congruent, incongruent) and mask type  
 345       were manipulated. Here we focus on the subset of trials in which no mask was presented.

346 The 13-ms prime stimulus was a double arrow presented 187 ms before target onset in the  
 347 congruent (same direction as target) and incongruent (opposite direction as target) prime  
 348 conditions.

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

352 The required column names are as follows:

- 353 • “pid”, indicating unique participant IDs;
- 354 • “trial”, indicating each unique trial per participant;
- 355 • “condition”, a factor indicating the levels of the independent variable (1, 2, ...) and  
 356 the corresponding labels;
- 357 • “rt”, indicating the response times in ms;
- 358 • “acc”, indicating the accuracies (1/0).

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

```
data_wr<-read_csv("../Tutorial_1_descriptive_stats/data/DataExp1_6subjects_wrangled.csv")
colnames(data_wr) <- c("pid","bl","tr","condition","resp","acc","rt","trial")
data_wr <- data_wr %>%
  mutate(condition = condition + 1, # original levels were 0, 1, 2.
        condition = factor(condition,
                             levels=c(1,2,3),
                             labels=c("blank","congruent","incongruent")))
```

360 Next, we can set up the life tables and plots of the discrete-time functions  $h(t)$ ,  $S(t)$ ,  
 361  $ca(t)$ , and  $P(t)$  – see part A of the supplementary material for their definitions. To do so  
 362 using a functional programming approach, one has to nest the data within participants  
 363 using the `group_nest()` function, and supply a user-defined censoring time and bin width  
 364 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))

```

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

374 The person-trial-bin oriented data set is created by our custom function ptb(), and it  
 375 has one row for each time bin (of each trial) that is at risk for event occurrence. The  
 376 variable “event” in the person-trial-bin oriented data set indicates whether a response  
 377 occurs (1) or not (0) for each bin.

378 The next step is to set up the life table using our custom function setup\_lt(),

379 calculate the conditional accuracies using our custom function `calc_ca()`, add the `ca(t)`  
380 estimates to the life table using our custom function `join_lt_ca()`, and then plot the  
381 descriptive statistics using our custom function `plot_eha()`. When creating the plots, some  
382 warning messages will likely be generated, like these:

- 383 • Removed 2 rows containing missing values or values outside the scale range  
384     (`geom_line()`).  
385 • Removed 2 rows containing missing values or values outside the scale range  
386     (`geom_point()`).  
387 • Removed 2 rows containing missing values or values outside the scale range  
388     (`geom_segment()`).

389 The warning messages are generated because some bins have no hazard and `ca(t)`  
390 estimates, and no error bars. They can thus safely be ignored. One can now inspect  
391 different aspects, including the life table for a particular condition of a particular subject,  
392 and a plot of the different functions for a particular participant. In general, it is important  
393 to visually inspect the functions first for each participant, in order to identify individuals  
394 that may be guessing (e.g., a flat conditional accuracy function at .5 indicates that  
395 someone is just guessing), outlying individuals, and/or different groups with qualitatively  
396 different behavior.

397 Table 3 shows the life table for condition “blank” (no prime stimulus presented) for  
398 participant 6.

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

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

## Descriptive stats for subject 6

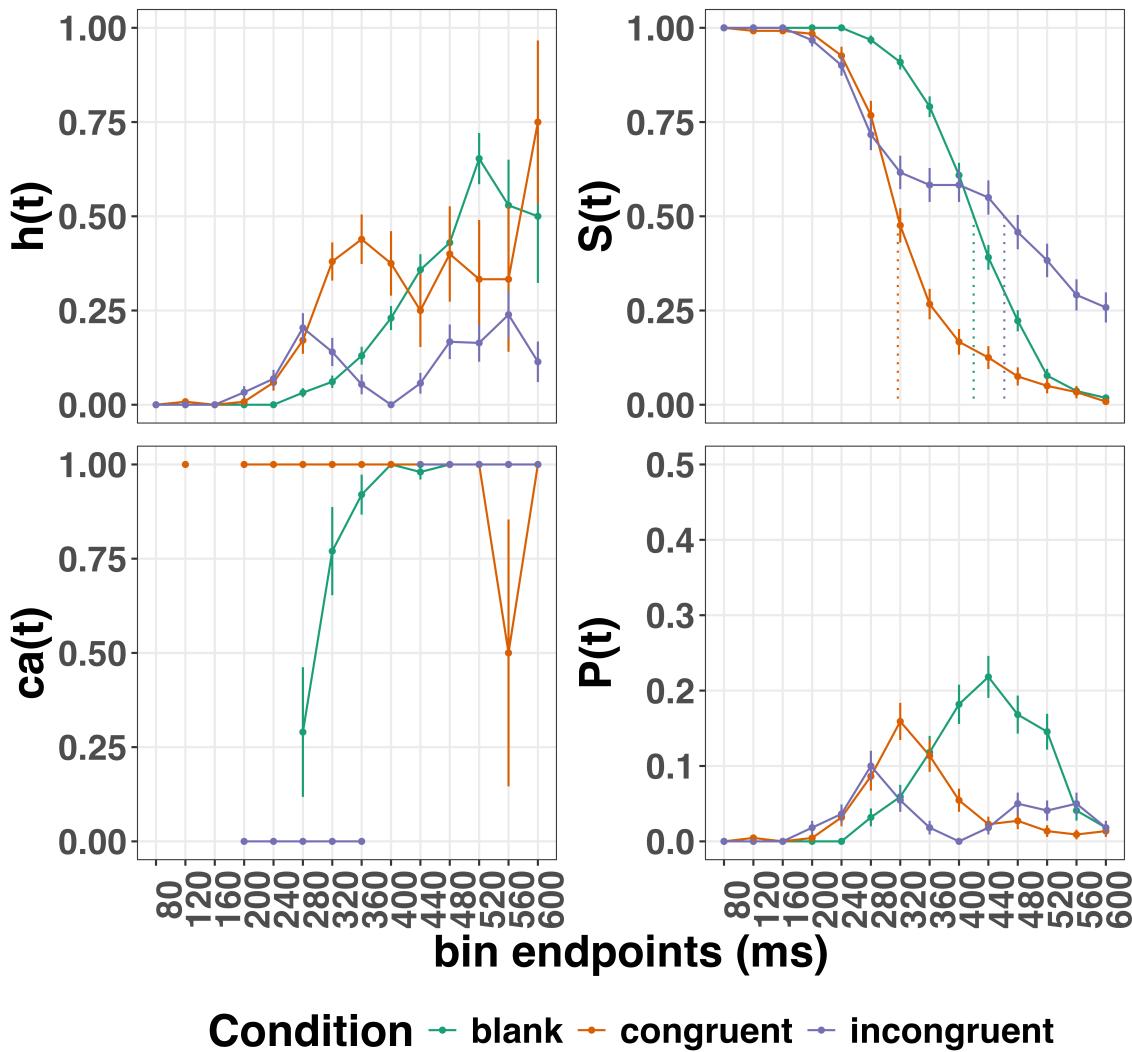


Figure 4. Estimated discrete-time hazard, survivor, probability mass, and conditional accuracy functions for participant 6. Vertical dotted lines indicate the estimated median RTs. Error bars represent +/- 1 Standard Error of the respective proportion.

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

406 0.17, and 0.20 for the blank, congruent, and incongruent prime conditions, respectively. In

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

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

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

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

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

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

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

414 respectively.

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

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

417 0.36, 0.25, and 0.06 for the blank, congruent, and incongruent prime conditions,

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

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

420 conditions, respectively.

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

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

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

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

425 were obtained for the other five participants. When participants show qualitatively the

426 same distributional patterns, one might consider to aggregate their data and make one plot

427 (see Tutorial\_1a.Rmd).

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

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

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

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

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

438 **4.2 Tutorial 1b: Generalising to a more complex design**

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

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

450 **4.3 Tutorial 2a: Fitting Bayesian hazard models to discrete time-to-event data**

451 In this third tutorial, we illustrate how to fit Bayesian multi-level regression models  
452 to the RT data of the masked response priming data set used in Tutorial 1a. Fitting  
453 (Bayesian or non-Bayesian) regression models to time-to-event data is important when you  
454 want to study how the shape of the hazard function depends on various predictors (Singer

455 & Willett, 2003).

456 **4.3.1 Hazard model considerations.** There are several analytic decisions one  
457 has to make when fitting a discrete-time hazard model. First, one has to select an analysis  
458 time window, i.e., a contiguous set of bins for which there is enough data for each  
459 participant. Second, given that the dependent variable (event occurrence) is binary, one  
460 has to select a link function (see part C in the supplementary material). The cloglog link is  
461 preferred over the logit link when events can occur in principle at any time point within a  
462 bin, which is the case for RT data (Singer & Willett, 2003). Third, one has to choose  
463 whether to treat TIME (i.e., the time bin index  $t$ ) as a categorical or continuous predictor.  
464 And when you treat a variable as a categorical predictor, you can choose between reference  
465 coding and index coding. With reference coding, one defines the variable as a factor and  
466 selects one of the  $k$  categories as the reference level. `Brm()` will then construct  $k-1$   
467 indicator variables (see model M1d in Tutorial\_2a.Rmd for an example). With index  
468 coding, one constructs an index variable which contains integers that correspond to  
469 different categories (see models M0i and M1i below). As explained by McElreath (2020),  
470 the advantage of index coding is that the same prior can be assigned to each level of the  
471 index variable, so that each category has the same prior uncertainty.

472 In the case of a large- $N$  design without repeated measurements, the parameters of a  
473 discrete-time hazard model can be estimated using standard logistic regression software  
474 after expanding the typical person-trial data set into a person-trial-bin data set (Allison,  
475 2010). When there is clustering in the data, as in the case of a small- $N$  design with  
476 repeated measurements, the parameters of a discrete-time hazard model can be estimated  
477 using population-averaged methods (e.g., Generalized Estimating Equations), and Bayesian  
478 or frequentist generalized linear mixed models (Allison, 2010).

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

482 additivity assumption (predictors do not interact), and the proportionality assumption  
 483 (predictors do not interact with TIME).

484 In tutorial\_2a.Rmd we fit several Bayesian multilevel models (i.e., generalized linear  
 485 mixed models) that differ in complexity to the person-trial-bin oriented data set that we  
 486 created in Tutorial 1a. We decided to select the analysis time window (200,600] and the  
 487 cloglog link. Below, we shortly discuss two of these models. The person-trial-bin data set is  
 488 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)))
```

489 **4.3.2 Prior distributions.** To get the posterior distribution of each model  
 490 parameter given the data, we need to specify prior distributions for the model parameters  
 491 which reflect our prior beliefs. In Tutorial\_2a.Rmd we perform a few prior predictive  
 492 checks to make sure our selected prior distributions reflect our prior beliefs (Gelman,  
 493 Vehtari, et al., 2020).

494 The middle column of Figure 16 in part E of the supplementary material shows six  
 495 examples of prior distributions for an intercept on the logit and/or cloglog scales. While a  
 496 normal distribution with relatively large variance is often used as a weakly informative

497 prior for continuous dependent variables, rows A and B in Figure 16 show that specifying  
 498 such distributions on the logit and cloglog scales actually leads to rather informative  
 499 distributions on the original probability scale, as most mass is pushed to probabilities of 0  
 500 and 1.

501       **4.3.3 Model M0i: A null model with index coding.** When you do not want to  
 502 make assumptions about the shape of the hazard function, or its shape is not smooth but  
 503 irregular, then you can use a general specification of TIME, i.e., fit one grand intercept per  
 504 time bin. In this first model, we use a general specification of TIME using index coding,  
 505 and do not include experimental predictors. We call this model “M0i”.

506       Before we fit model M0i, we select the necessary columns from the data, and specify  
 507 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")
```

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

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

514 **4.3.4 Model M1i: Adding the effects of prime-target congruency.** Previous  
 515 research has shown that psychological effects typically change over time (Panis, 2020;  
 516 Panis, Moran, et al., 2020; Panis & Schmidt, 2022; Panis et al., 2017; Panis & Wagemans,  
 517 2009). In the next model, therefore, we use index coding for both TIME (variable  
 518 “timebin”) and the categorical predictor prime-target-congruency (variable “prime”), so  
 519 that we get 30 grand intercepts, one for each combination of timebin level and prime level.  
 520 Here is the model formula of this model that we call “M1i”.

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

521 Estimating model M1i took about 124 minutes.

522 **4.3.5 Compare the models.** We can compare the two models using the Widely  
 523 Applicable Information Criterion (WAIC) and Leave-One-Out (LOO) cross-validation, and  
 524 look at model weights for both criteria (Kurz, 2023a; McElreath, 2020).

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

525 ## model\_M0i model\_M1i  
 526 ## 0 1

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

527 ## model\_M0i model\_M1i  
 528 ## 0 1

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

531        **4.3.6 Evaluating parameter estimates in model M1i.** To make inferences

532    from the parameter estimates in model M1i, we first plot the densities of the draws from  
 533    the posterior distributions of its population-level parameters in Figure 5, together with  
 534    point (median) and interval estimates (80% and 95% credible intervals).

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

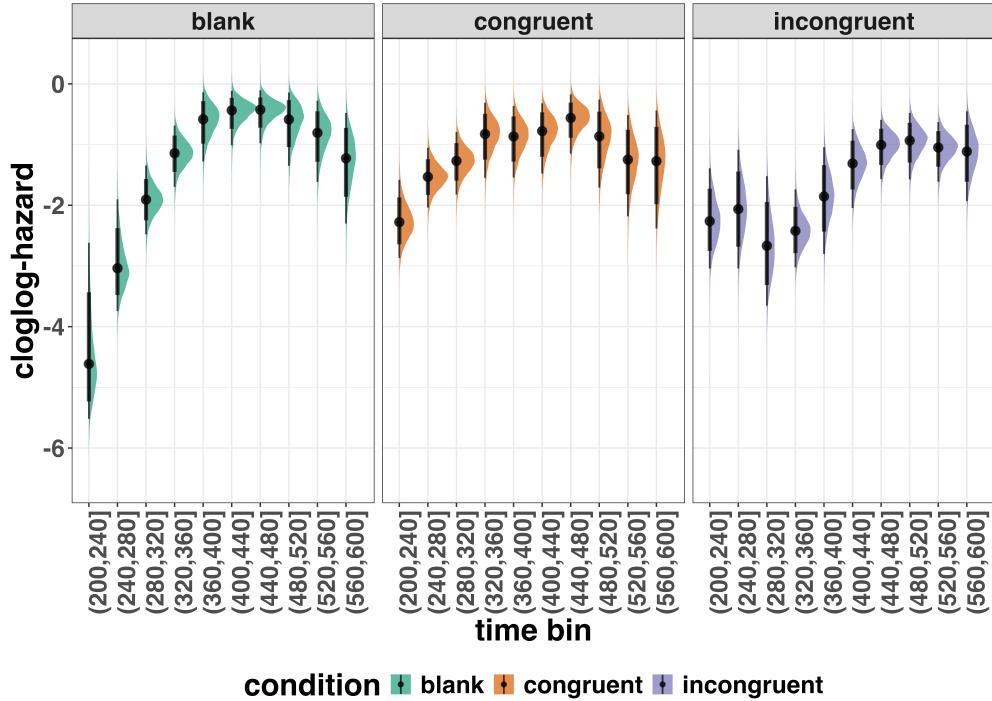


Figure 5. Medians and 80/95% credible intervals of the posterior distributions of the population-level parameters of model M1i.

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

536    interpretation by plotting the expected value of the posterior predictive distribution – the  
 537    predicted hazard values – for the average participant (Figure 6A), and for each participant  
 538    in the data set (Figure 6B).

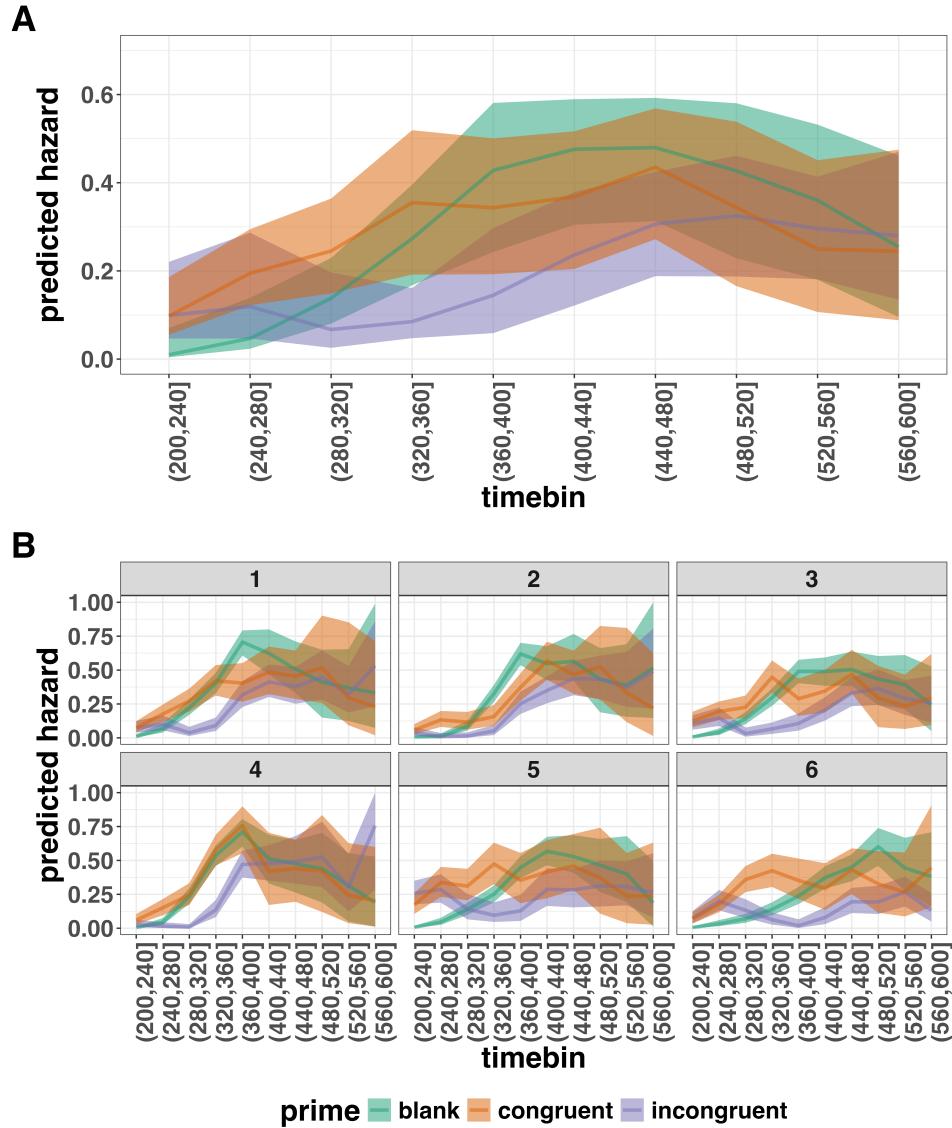


Figure 6. Point (median) and 80/95% credible interval summaries of the hazard estimates (expected values of the draws from the posterior predictive distributions) in each time bin for the average participant (A), and for each participant (B).

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

540 relative to the blank prime condition, we can construct two contrasts (congruent-blank,

541 incongruent-blank), and plot the posterior distributions of these contrast effects, both for

542 the average participant (Figure 7A; grand average marginal effect) and for each participant

<sup>543</sup> in the data set (Figure 7B; subject-specific average marginal effect).

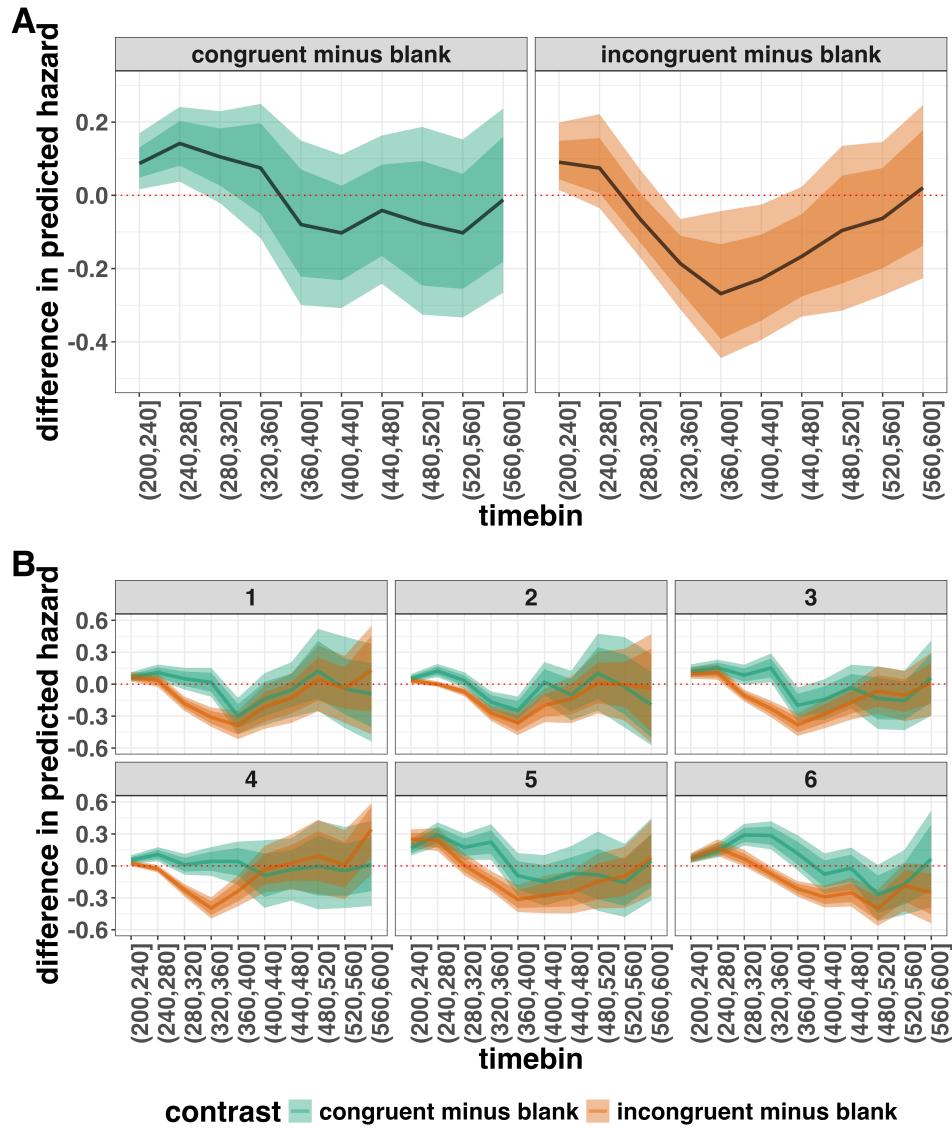


Figure 7. Point (mean) and 80/95% credible interval summaries of estimated differences in hazard in each time bin for the average participant (A), and for each participant (B).

<sup>544</sup> The point estimates and quantile intervals can be reported in a table (see  
<sup>545</sup> Tutorial\_2a.Rmd for details).

<sup>546</sup> **Example conclusions for M1i.** What can we conclude from model M1i about  
<sup>547</sup> our research question, i.e., the temporal dynamics of the effect of prime-target congruency

548 on RT? In other words, in which of the 40-ms time bins between 200 and 600 ms after  
549 target onset does changing the prime from blank to congruent or incongruent affect the  
550 hazard of response occurrence (for a prime-target SOA of 187 ms)?

551 If we want to study the average effect of prime type on hazard, uncontaminated by  
552 inter-individual differences, we can base our conclusion on Figure 8 and Table 4. The  
553 contrast “congruent minus blank” was estimated to be 0.09 hazard units in bin 6 (95% CrI  
554 = [0.02, 0.17]), and 0.14 hazard units in bin 7 (95% CrI = [0.04, 0.25]). For the other bins,  
555 the 95% credible interval contained zero. The contrast “incongruent minus blank” was  
556 estimated to be 0.09 hazard units in bin 6 (95% CrI = [0.01, 0.21]), -0.19 hazard units in  
557 bin 9 (95% CrI = [-0.31, -0.06]), -0.27 hazard units in bin 10 (95% CrI = [-0.45, -0.04]),  
558 and -0.23 hazard units in bin 11 (95% CrI = [-0.40, -0.03]). For the other bins, the 95%  
559 credible interval contained zero. Note that we could also have calculated hazard ratios  
560 instead of hazard differences.

561 There are thus two phases of performance for the average person between 200 and  
562 600 ms after target onset. In the first phase, the addition of a congruent or incongruent  
563 prime stimulus increases the hazard of response occurrence compared to blank prime trials  
564 in the time period (200, 240]. In the second phase, only the incongruent prime decreases  
565 the hazard of response occurrence compared to blank primes, in the time period (320,440].  
566 The sign of the effect of incongruent primes on the hazard of response occurrence thus  
567 depends on how much waiting time has passed since target onset.

568 If we want to focus more on inter-individual differences, we can study the  
569 subject-specific hazard functions in Figure 9. Note that three participants (1, 2, and 3)  
570 show a negative difference for the contrast “congruent minus incongruent” in bin (360,400]  
571 – subject 2 also in bin (320,360].

572 Future studies could (a) increase the number of participants to estimate the  
573 proportion of “dippers” in the subject population, and/or (b) try to explain why this dip

574 occurs. For example, Panis and Schmidt (2016) concluded that active, top-down,  
575 task-guided response inhibition effects emerge around 360 ms after the onset of the stimulus  
576 following the prime (here: the target stimulus). Such a top-down inhibitory effect might  
577 exist in our priming data set, because after some time participants will learn that the first  
578 stimulus is not the one they have to respond to. To prevent a premature overt response to  
579 the prime they thus might gradually increase a global response threshold during the  
580 remainder of the experiment, which could result in a lower hazard in congruent trials  
581 compared to blank trials, for bins after ~360 ms, and towards the end of the experiment.  
582 This effect might be masked for incongruent primes by the response competition effect.

583 Interestingly, all subjects show a tendency in their mean difference (congruent minus  
584 blank) to “dip” around that time (Figure 9). Therefore, future modeling efforts could  
585 incorporate the trial number into the model formula, in order to also study how the effects  
586 of prime type on hazard change on the long experiment-wide time scale, next to the short  
587 trial-wide time scale. In Tutorial\_2a.Rmd we provide a number of model formula that  
588 should get you going.

#### 589 **4.4 Tutorial 2b: Fitting Bayesian conditional accuracy models**

590 In this fourth tutorial, we illustrate how to fit a Bayesian multi-level regression model  
591 to the timed accuracy data from the masked response priming data set used in Tutorial 1a.  
592 The general process is similar to Tutorial 2a, except that (a) we use the person-trial data  
593 set, (b) we use the logit link function, and (c) we change the priors. To keep the tutorial  
594 short, we only fitted the effects of model M1i (see Tutorial 2a) in the conditional accuracy  
595 model called M1i\_ca.

596 To make inferences from the parameter estimates in model M1i\_ca, we first plot the  
597 densities of the draws from the posterior distributions of its population-level parameters in  
598 Figure 8, together with point (median) and interval estimates (80% and 95% credible

599 intervals).

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

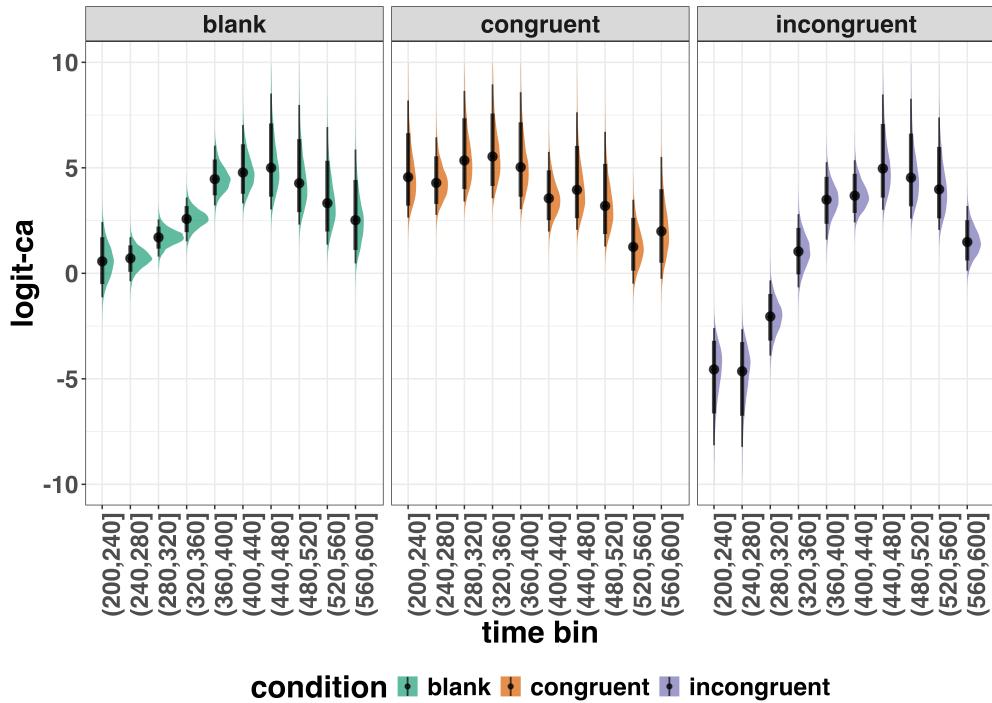


Figure 8. Medians and 80/95% credible intervals of the posterior distributions of the population-level parameters of model M1i\_ca.

600 Because the parameter estimates are on the logit-ca scale, we can ease our  
 601 interpretation by plotting the expected value of the posterior predictive distribution – the  
 602 predicted conditional accuracies – for the average participant (Figure 9A), and for each  
 603 participant in the data set (Figure 9B).

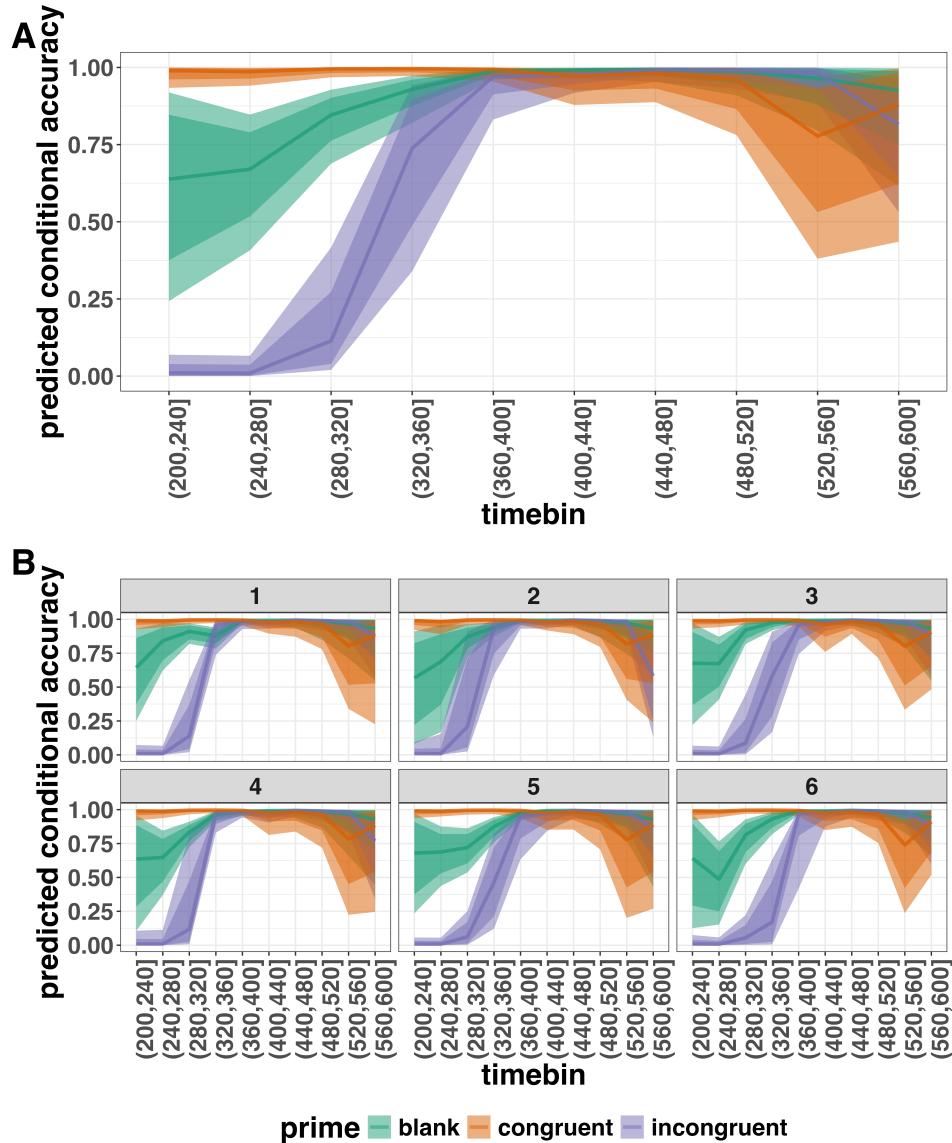
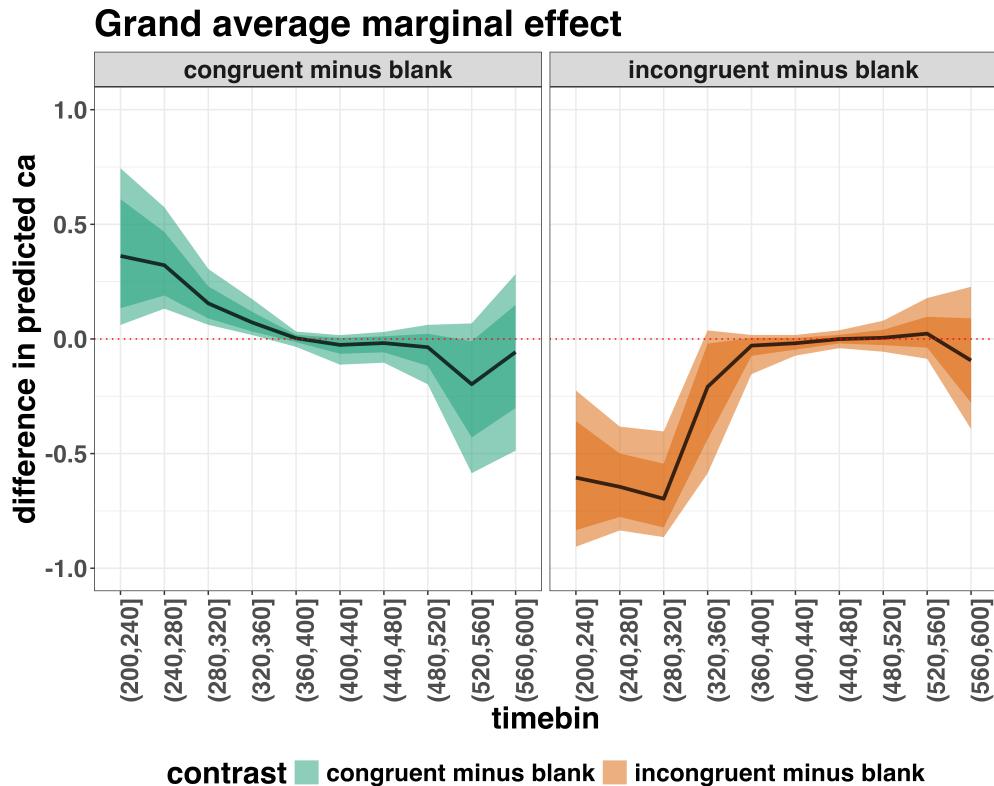


Figure 9. Point (median) and 80/95% credible interval summaries of the conditional accuracy estimates (expected values of the draws from the posterior predictive distributions) in each time bin for the average participant (A), and for each participant (B).

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

relative to the blank prime condition, we can construct two contrasts (congruent-blank, incongruent-blank), and plot the posterior distributions of these contrast effects for the average participant (Figure 10; grand average marginal effect).



*Figure 10.* Point (mean) and 80/95% credible interval summaries of estimated differences in conditional accuracy in each time bin for the average participant.

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

609        Tutorial\_2b.Rmd for details).

610        Based on Figure 10 we see that congruent primes have a positive effect on the

611        conditional accuracy of emitted responses in time bins (200,240], (240,280], and (280,320],

612        relative to the estimates in the baseline condition (blank prime; red dashed lines in Figure

613        14). Incongruent primes have a negative effect on the conditional accuracy of emitted

614        responses in those time bins, relative to the estimates in the baseline condition.

#### 615        4.5 Tutorial 3a: Fitting Frequentist hazard models

616        In this fifth tutorial we illustrate how to fit a multilevel regression model to RT data

617        in the frequentist framework, for the data set used in Tutorial 1a. The general process is

618 similar to that in Tutorial 2a, except that there are no priors to set.

619 To keep this tutorial short, we only fitted the effects from model M1i (see Tutorial  
 620 2a) using the function `glmer()` from the R package `lme4`. Alternatively, one could also use  
 621 the function `glmmPQL()` from the R package `MASS` (Ripley et al., 2024). The resulting  
 622 hazard model is called `M1i_f`.

623 In Figure 11 we compare the parameter estimates of model M1i from `brm()` with  
 624 those of model `M1i_f` from `glmer()`.

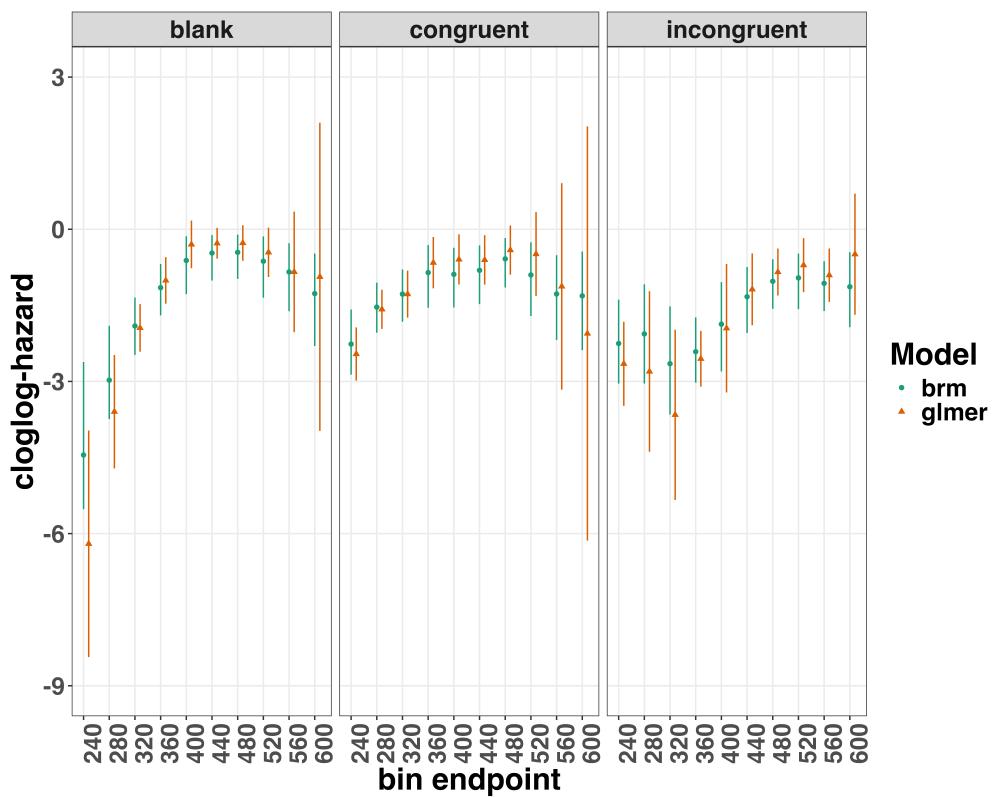


Figure 11. Parameter estimates for model M1i from `brm()` – means and 95% credible intervals – and model `M1i_f` from `glmer()` – maximum likelihood estimates and 95% confidence intervals.

625 Figure 11 confirms that the parameter estimates from both Bayesian and frequentist  
 626 models are pretty similar, which makes sense given the close similarity in model structure.  
 627 However, model `M1i_f` did not converge and resulted in a singular fit. This is of course one

of the reasons why Bayesian modeling has become so popular in recent years. But the price you pay for being able to fit more complex random effects models in a Bayesian framework is computation time. In other words, as we have noted throughout, some of the Bayesian models in Tutorials 2a took several hours to build.

#### 4.6 Tutorial 3b: Fitting Frequentist conditional accuracy models

In this sixth tutorial we illustrate how to fit a multilevel regression model to the timed accuracy data in the frequentist framework, for the data set used in Tutorial 1a. To keep it short, we only fitted the effects from model M1i\_ca (see Tutorial 2b) using the function glmer() from the R package lme4. Alternatively, one could also use the function glmmPQL() from the R package MASS (Ripley et al., 2024). Again, the resulting conditional accuracy model M1i\_ca\_f did not converge and resulted in a singular fit.

#### 4.7 Tutorial 4: Planning

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

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

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

1. Fit a regression model to an existing dataset.
2. Use the regression model parameters to simulate one new dataset.

- 651     3. Write a function to create 1000s of datasets and vary parameters of interest (e.g.,  
 652       sample size, trial count, effect size).
- 653     4. Summarise the simulated data to estimate likely power or precision of the research  
 654       design options.

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

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

662       **4.7.3 Fit a regression model and simulate one dataset.** We again use the  
 663       data from Panis and Schmidt (2016) to provide a worked example. We fit an index coding  
 664       model on a subset of timebins (six timebins in total) and for two prime conditions  
 665       (congruent and incongruent). We chose to focus on a subsample of the data to ease the  
 666       computational burden. We also used a full varying effects structure, with the model  
 667       formula, as follows:

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

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

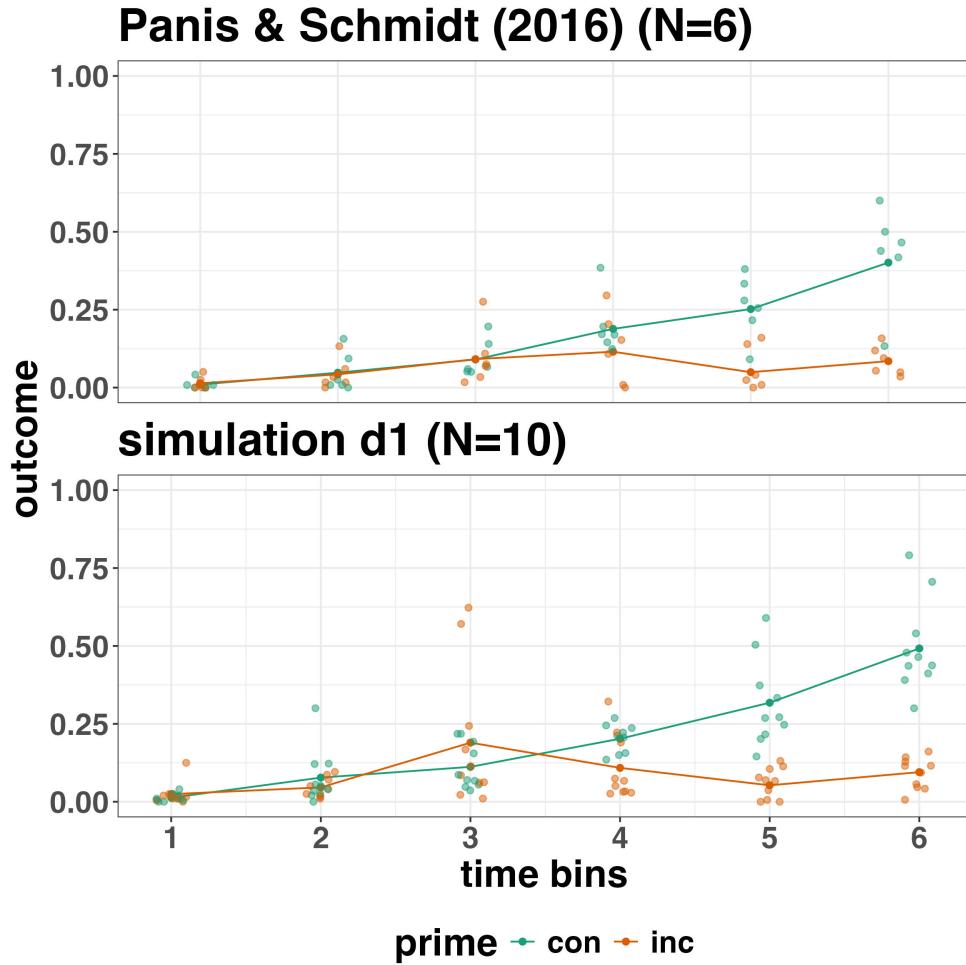


Figure 12. Raw data from Panis and Schmidt (2016) and simulated data from 10 participants.

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

Here we use the same data simulation process as used above, but instead of simulating one dataset, we simulate 1000 datasets per variation in parameter values. Specifically, in Simulation 1, we vary the number of trials per condition (100, 200, and 400), as well as the 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. Summary results from Simulation 2 are shown in Figure 13B. A summary of these power calculations might be as follows (trial count = 200 per condition in all cases):

- For a 70% reduction (0.3 hazard ratio), N=10 would give nearly 100% power.

- 707 • For a 60% reduction (0.4 hazard ratio), N=10 would give nearly 90% power.
- 708 • For a 50% reduction (0.5 hazard ratio), N=15 would give over 80% power.

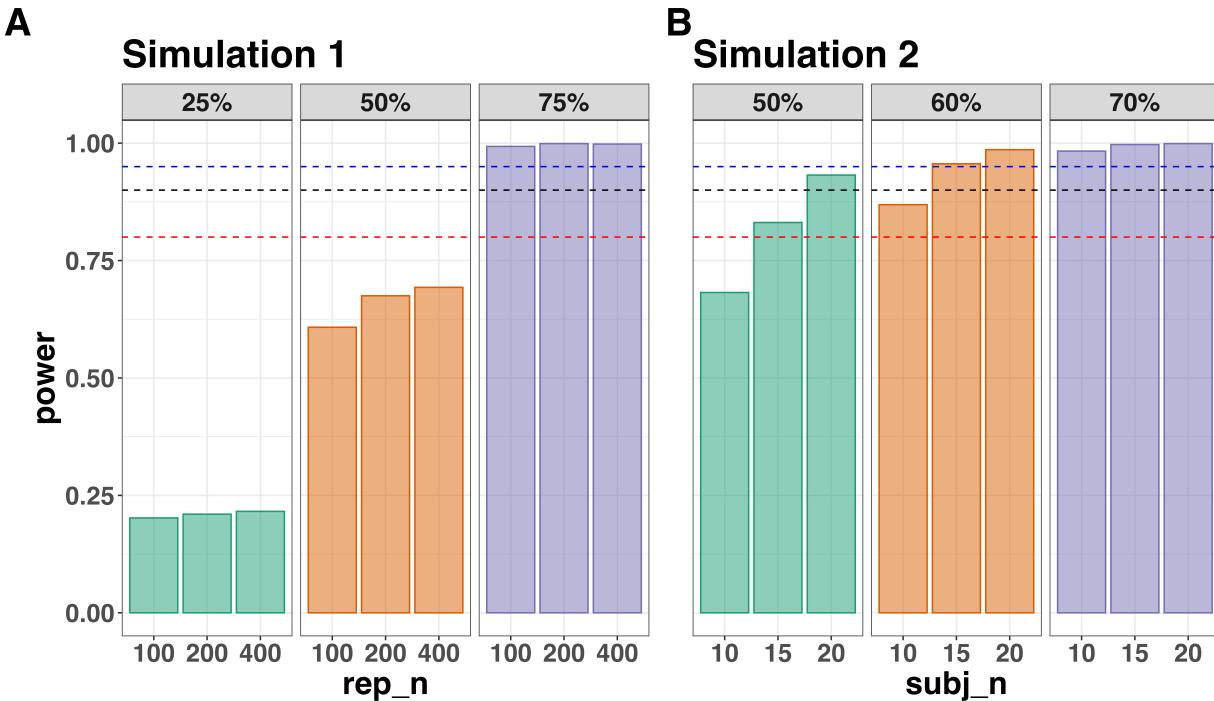


Figure 13. Statistical power across data Simulation 1 (A) and Simulation 2 (B). Power was calculated as the percentage of lower-bound 95% confidence intervals that exclude zero when the difference between prime condition is calculated (congruent - incongruent). In Simulation 1, the effect size was varied 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.

709       **4.7.5 Planning decisions.** Now that we have summarised our simulated data,

710       what planning decisions could we make about a future study? How many trials per  
711       condition should we collect and how many participants should we test? Like almost always  
712       when planning future studies, the answer depends on your objectives, as well as the  
713       available resources (Lakens, 2022). There is no straightforward and clear-cut answer. Some  
714       considerations might be...

- 715       • How much power or precision are you looking to obtain in this particular study?
- 716       • Are you running multiple studies that have some form of replication built in?
- 717       • What resources do you have at your disposal, such as time, money and personnel?
- 718       • How easy or difficult is it to obtain the specific type of sample?

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

720       hazard ratio of 0.4 or 0.5 as a target effect size since this is much smaller than that  
721       observed in the previously published study that this work is building upon (Panis &  
722       Schmidt, 2016). Then we might pick the corresponding N value (i.e., N=10 or N=15) that  
723       takes you over the 80% power mark. If we wanted to maximise power based on these  
724       simulations, and we had the time and resources available, then we test N=20 participants,  
725       which would provide >90% power for an effect size of 0.5.

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

727       what planning choices we made based on these data simulations, we would not solely rely  
728       on data collected from one single study. Instead, we would run a follow-up experiment that  
729       replicates and extends the initial result. By doing so, we would aim to avoid the Cult of  
730       the Isolated Single Study (Nelder, 1999; Tong, 2019), and thus reduce the reliance on any  
731       one type of planning tool, such as a power analysis. Then, we would look for common  
732       patterns across two or more experiments, rather than trying to make the case that a single  
733       study on its own has sufficient evidential value to hit some criterion mark.

734

## 5. Discussion

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

750 **5.1 What are the main use-cases of EHA for understanding cognition and brain  
751 function?**

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

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

## 768        5.2 Model complexity versus interpretability

769        EHA can quickly become very complex when adding more than 1 time scale, due to  
770 the many possible higher-order interactions. For example, some of the models discussed in  
771 Tutorial 2a (M2) contain two time scales as covariates: the passage of time on the  
772 within-trial time scale, and the passage of time on the across-trial (or within-experiment)  
773 time scale. However, when trials are presented in blocks, and blocks of trials within  
774 sessions, and when the experiment comprises three sessions, then four time scales can be  
775 defined (within-trial, within-block, within-session, and within-experiment). From a  
776 theoretical perspective, adding more than 1 time scale – and their interactions – can be  
777 important to capture plasticity and other learning effects that may play out on such longer  
778 time scales, and that are probably present in each experiment in general. From a practical  
779 perspective, therefore, some choices need to be made to balance the amount of data that is  
780 being collected per participant, condition and across the varying timescales. As one  
781 example, if there are several timescales of relevance, then it might be prudent for  
782 interpretational purposes to limit the number of experimental predictor variables  
783 (conditions). This is of course where planning and data simulation efforts would be  
784 important to provide a guide to experimental design choices (see Tutorial 4).

**785 5.3 Individual differences**

786 One important issue is that of possible individual differences in the overall location of  
787 the distribution, and the time course of psychological effects. For example, when you wait  
788 for a response of the participant on each trial, you allow the participant to have control  
789 over the trial duration, and some participants might respond only when they are confident  
790 that their emitted response will be correct. These issues can be avoided by introducing a  
791 (relatively short) response deadline in each trial, e.g., 600 ms for simple detection tasks,  
792 1000 ms for more difficult discrimination tasks, or 2 s for tasks requiring extended high-level  
793 processing. Because EHA can deal in a straightforward fashion with right-censored  
794 observations (i.e., trials without an observed response), introducing a response deadline is  
795 recommended when designing RT experiments. Furthermore, introducing a response  
796 deadline and asking participants to respond before the deadline as much as possible, will  
797 also lead to individual distributions that overlap in time, which is important when selecting  
798 a common analysis time window when fitting hazard and conditional accuracy models.

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

808 Another approach to deal with individual differences is Bayesian prevalence (Ince,  
809 Paton, Kay, & Schyns, 2021), which is a from of Small-N approach (Smith & Little, 2018).  
810 This method looks at effects within each individual in the study and asks how likely it

would be to see the same result if the experiment was repeated with a new person chosen from the wider population at random. This approach allows one to quantify how typical or uncommon an observed effect is in the population, and the uncertainty around this estimate.

#### 5.4 Limitations

Compared to the orthodox method – comparing mean-averages between conditions – the most important limitation of multi-level hazard and conditional accuracy modeling is that it might take a long time to estimate the parameters using Bayesian methods or the model might have to be simplified significantly to use frequentist methods.

Another issue is that you need a relatively large number of trials per condition to estimate the hazard function with high temporal resolution, which is required when testing predictions of process models of cognition. Indeed, in general, there is a trade-off between the number of trials per condition and the temporal resolution (i.e., bin width) of the hazard function. Therefore, we recommend researchers to collect as many trials as possible per experimental condition, given the available resources and considering the participant experience (e.g., fatigue and boredom). For instance, if the maximum session length deemed reasonable is between 1 and 2 hours, what is the maximum number of trials per condition that you could reasonably collect? After consideration, it might be worth conducting multiple testing sessions per participant and/or reducing the number of experimental conditions. Finally, there is a user-friendly online tool for calculating statistical power as a function of the number of trials as well as the number of participants, 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 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.

## 6. Conclusions

Estimating the temporal distributions of RT and accuracy provide a rich source of information on the time course of cognitive processing, which have been largely undervalued in the history of experimental psychology and cognitive neuroscience.

861 Statistically controlling for the passage of time during data analysis is equally important as  
862 experimental control during the design of an experiment, to better understand human  
863 behavior in experimental paradigms. We hope that by providing a set of hands-on,  
864 step-by-step tutorials, which come with custom-built and freely available code, researchers  
865 will feel more comfortable embracing EHA and investigating the temporal profile of  
866 cognitive states. On a broader level, we think that wider adoption of such approaches will  
867 have a meaningful impact on the inferences drawn from data, as well as the development of  
868 theories regarding the structure of cognition.

869

## References

- 870 Allison, P. D. (1982). Discrete-Time Methods for the Analysis of Event Histories.  
871     *Sociological Methodology*, 13, 61. <https://doi.org/10.2307/270718>
- 872 Allison, P. D. (2010). *Survival analysis using SAS: A practical guide* (2. ed). Cary, NC:  
873     SAS Press.
- 874 Aust, F. (2019). *Citr: 'RStudio' add-in to insert markdown citations*. Retrieved from  
875     <https://github.com/crsh/citr>
- 876 Aust, F., & Barth, M. (2023). *papaja: Prepare reproducible APA journal articles with R*  
877     *Markdown*. Retrieved from <https://github.com/crsh/papaja>
- 878 Aust, F., & Barth, M. (2024). *papaja: Prepare reproducible APA journal articles with R*  
879     *Markdown*. <https://doi.org/10.32614/CRAN.package.papaja>
- 880 Baker, D. H., Vilidaite, G., Lygo, F. A., Smith, A. K., Flack, T. R., Gouws, A. D., &  
881 Andrews, T. J. (2021). Power contours: Optimising sample size and precision in  
882 experimental psychology and human neuroscience. *Psychological Methods*, 26(3),  
883 295–314. <https://doi.org/10.1037/met0000337>
- 884 Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for  
885 confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*,  
886 68(3), 10.1016/j.jml.2012.11.001. <https://doi.org/10.1016/j.jml.2012.11.001>
- 887 Barth, M. (2023). *tinylabes: Lightweight variable labels*. Retrieved from  
888     <https://cran.r-project.org/package=tinylabes>
- 889 Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects  
890 models using lme4. *Journal of Statistical Software*, 67(1), 1–48.  
891     <https://doi.org/10.18637/jss.v067.i01>
- 892 Bates, D., Maechler, M., & Jagan, M. (2024). *Matrix: Sparse and dense matrix classes and*  
893 *methods*. Retrieved from <https://CRAN.R-project.org/package=Matrix>
- 894 Bengtsson, H. (2021). A unifying framework for parallel and distributed processing in r  
895 using futures. *The R Journal*, 13(2), 208–227. <https://doi.org/10.32614/RJ-2021-048>

- 896 Blossfeld, H.-P., & Rohwer, G. (2002). *Techniques of event history modeling: New*  
897       *approaches to causal analysis, 2nd ed* (pp. x, 310). Mahwah, NJ, US: Lawrence  
898       Erlbaum Associates Publishers.
- 899 Box-Steffensmeier, J. M. (2004). Event history modeling: A guide for social scientists.  
900       Cambridge: University Press.
- 901 Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan.  
902       *Journal of Statistical Software*, 80(1), 1–28. <https://doi.org/10.18637/jss.v080.i01>
- 903 Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms.  
904       *The R Journal*, 10(1), 395–411. <https://doi.org/10.32614/RJ-2018-017>
- 905 Bürkner, P.-C. (2021). Bayesian item response modeling in R with brms and Stan. *Journal*  
906       *of Statistical Software*, 100(5), 1–54. <https://doi.org/10.18637/jss.v100.i05>
- 907 DeBruine, L. M., & Barr, D. J. (2021). Understanding Mixed-Effects Models Through  
908       Data Simulation. *Advances in Methods and Practices in Psychological Science*, 4(1),  
909       2515245920965119. <https://doi.org/10.1177/2515245920965119>
- 910 Eddelbuettel, D., & Balamuta, J. J. (2018). Extending R with C++: A Brief Introduction  
911       to Rcpp. *The American Statistician*, 72(1), 28–36.  
912       <https://doi.org/10.1080/00031305.2017.1375990>
- 913 Eddelbuettel, D., & François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal*  
914       *of Statistical Software*, 40(8), 1–18. <https://doi.org/10.18637/jss.v040.i08>
- 915 Gabry, J., Češnovar, R., Johnson, A., & Broder, S. (2024). *Cmdstanr: R interface to*  
916       *'CmdStan'*. Retrieved from <https://github.com/stan-dev/cmdstanr>
- 917 Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2019). Visualization  
918       in bayesian workflow. *J. R. Stat. Soc. A*, 182, 389–402.  
919       <https://doi.org/10.1111/rssa.12378>
- 920 Gelman, A., Hill, J., & Vehtari, A. (2020). Regression and Other Stories.  
921       <https://www.cambridge.org/highereducation/books/regression-and-other-stories/DD20DD6C9057118581076E54E40C372C>; Cambridge University Press.

- 923 https://doi.org/10.1017/9781139161879
- 924 Gelman, A., Vehtari, A., Simpson, D., Margossian, C. C., Carpenter, B., Yao, Y., ...
- 925 Modrák, M. (2020). *Bayesian Workflow*. arXiv.
- 926 https://doi.org/10.48550/arXiv.2011.01808
- 927 Girard, J. (2024). *Standist: What the package does (one line, title case)*. Retrieved from
- 928 https://github.com/jmgirard/standist
- 929 Grolemund, G., & Wickham, H. (2011). Dates and times made easy with lubridate.
- 930 *Journal of Statistical Software*, 40(3), 1–25. Retrieved from
- 931 https://www.jstatsoft.org/v40/i03/
- 932 Halley, E. (1693). VI. An estimate of the degrees of the mortality of mankind; drawn from
- 933 curious tables of the births and funerals at the city of breslaw; with an attempt to
- 934 ascertain the price of annuities upon lives. *Philosophical Transactions of the Royal*
- 935 *Society of London*, 17(196), 596–610. https://doi.org/10.1098/rstl.1693.0007
- 936 Heiss, A. (2021, November 10). A Guide to Correctly Calculating Posterior Predictions
- 937 and Average Marginal Effects with Multilevel Bayesian Models.
- 938 https://doi.org/10.59350/wbn93-edb02
- 939 Holden, J. G., Van Orden, G. C., & Turvey, M. T. (2009). Dispersion of response times
- 940 reveals cognitive dynamics. *Psychological Review*, 116(2), 318–342.
- 941 https://doi.org/10.1037/a0014849
- 942 Hosmer, D. W., Lemeshow, S., & May, S. (2011). *Applied Survival Analysis: Regression*
- 943 *Modeling of Time to Event Data* (2nd ed). Hoboken: John Wiley & Sons.
- 944 Ince, R. A., Paton, A. T., Kay, J. W., & Schyns, P. G. (2021). Bayesian inference of
- 945 population prevalence. *eLife*, 10, e62461. https://doi.org/10.7554/eLife.62461
- 946 Kantowitz, B. H., & Pachella, R. G. (2021). The Interpretation of Reaction Time in
- 947 Information-Processing Research 1. *Human Information Processing*, 41–82.
- 948 https://doi.org/10.4324/9781003176688-2
- 949 Kay, M. (2023). *tidybayes: Tidy data and geoms for Bayesian models*.

- 950 https://doi.org/10.5281/zenodo.1308151
- 951 Kelso, J. A. S., Dumas, G., & Tognoli, E. (2013). Outline of a general theory of behavior  
952 and brain coordination. *Neural Networks: The Official Journal of the International*  
953 *Neural Network Society*, 37, 120–131. https://doi.org/10.1016/j.neunet.2012.09.003
- 954 Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing,  
955 estimation, meta-analysis, and power analysis from a Bayesian perspective.  
956 *Psychonomic Bulletin & Review*, 25(1), 178–206.
- 957 https://doi.org/10.3758/s13423-016-1221-4
- 958 Kurz, A. S. (2023a). *Applied longitudinal data analysis in brms and the tidyverse* (version  
959 0.0.3). Retrieved from https://bookdown.org/content/4253/
- 960 Kurz, A. S. (2023b). *Statistical rethinking with brms, ggplot2, and the tidyverse: Second*  
961 *edition* (version 0.4.0). Retrieved from https://bookdown.org/content/4857/
- 962 Lakens, D. (2022). Sample Size Justification. *Collabra: Psychology*, 8(1), 33267.  
963 https://doi.org/10.1525/collabra.33267
- 964 Landes, J., Engelhardt, S. C., & Pelletier, F. (2020). An introduction to event history  
965 analyses for ecologists. *Ecosphere*, 11(10), e03238. https://doi.org/10.1002/ecs2.3238
- 966 Luce, R. D. (1991). *Response times: Their role in inferring elementary mental organization*  
967 (1. issued as paperback). Oxford: Univ. Press.
- 968 Makeham, W. M. (1860). *On the Law of Mortality and the Construction of Annuity Tables*.  
969 The Assurance Magazine, and Journal of the Institute of Actuaries.
- 970 McElreath, R. (2020). *Statistical Rethinking: A Bayesian Course with Examples in R and*  
971 *STAN* (2nd ed.). New York: Chapman and Hall/CRC.
- 972 https://doi.org/10.1201/9780429029608
- 973 Meyer, D. E., Osman, A. M., Irwin, D. E., & Yantis, S. (1988). Modern mental  
974 chronometry. *Biological Psychology*, 26(1-3), 3–67.  
975 https://doi.org/10.1016/0301-0511(88)90013-0
- 976 Müller, K., & Wickham, H. (2023). *Tibble: Simple data frames*. Retrieved from

- 977 https://CRAN.R-project.org/package=tibble
- 978 Nelder, J. A. (1999). From Statistics to Statistical Science. *Journal of the Royal Statistical  
979 Society. Series D (The Statistician)*, 48(2), 257–269. Retrieved from  
980 https://www.jstor.org/stable/2681191
- 981 Neuwirth, E. (2022). *RColorBrewer: ColorBrewer palettes*. Retrieved from  
982 https://CRAN.R-project.org/package=RColorBrewer
- 983 Panis, S. (2020). How can we learn what attention is? Response gating via multiple direct  
984 routes kept in check by inhibitory control processes. *Open Psychology*, 2(1), 238–279.  
985 https://doi.org/10.1515/psych-2020-0107
- 986 Panis, S., Moran, R., Wolkersdorfer, M. P., & Schmidt, T. (2020). Studying the dynamics  
987 of visual search behavior using RT hazard and micro-level speed–accuracy tradeoff  
988 functions: A role for recurrent object recognition and cognitive control processes.  
989 *Attention, Perception, & Psychophysics*, 82(2), 689–714.  
990 https://doi.org/10.3758/s13414-019-01897-z
- 991 Panis, S., Schmidt, F., Wolkersdorfer, M. P., & Schmidt, T. (2020). Analyzing Response  
992 Times and Other Types of Time-to-Event Data Using Event History Analysis: A Tool  
993 for Mental Chronometry and Cognitive Psychophysiology. *I-Perception*, 11(6),  
994 2041669520978673. https://doi.org/10.1177/2041669520978673
- 995 Panis, S., & Schmidt, T. (2016). What Is Shaping RT and Accuracy Distributions? Active  
996 and Selective Response Inhibition Causes the Negative Compatibility Effect. *Journal of  
997 Cognitive Neuroscience*, 28(11), 1651–1671. https://doi.org/10.1162/jocn\_a\_00998
- 998 Panis, S., & Schmidt, T. (2022). When does “inhibition of return” occur in spatial cueing  
999 tasks? Temporally disentangling multiple cue-triggered effects using response history  
1000 and conditional accuracy analyses. *Open Psychology*, 4(1), 84–114.  
1001 https://doi.org/10.1515/psych-2022-0005
- 1002 Panis, S., Torfs, K., Gillebert, C. R., Wageman, J., & Humphreys, G. W. (2017).  
1003 Neuropsychological evidence for the temporal dynamics of category-specific naming.

- 1004        *Visual Cognition*, 25(1-3), 79–99. <https://doi.org/10.1080/13506285.2017.1330790>
- 1005      Panis, S., & Wagemans, J. (2009). Time-course contingencies in perceptual organization  
1006      and identification of fragmented object outlines. *Journal of Experimental Psychology:  
1007      Human Perception and Performance*, 35(3), 661–687.  
1008      <https://doi.org/10.1037/a0013547>
- 1009      Pedersen, T. L. (2024). *Patchwork: The composer of plots*. Retrieved from  
1010      <https://CRAN.R-project.org/package=patchwork>
- 1011      Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in s and s-PLUS*. New York:  
1012      Springer. <https://doi.org/10.1007/b98882>
- 1013      R Core Team. (2024). *R: A language and environment for statistical computing*. Vienna,  
1014      Austria: R Foundation for Statistical Computing. Retrieved from  
1015      <https://www.R-project.org/>
- 1016      Ripley, B., Venables, B., Bates, D. M., ca 1998), K. H. (partial. port, ca 1998), A. G.  
1017      (partial. port, & polr), D. F. (support. functions for. (2024). *MASS: Support Functions  
1018      and Datasets for Venables and Ripley's MASS*.
- 1019      Singer, J. D., & Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling  
1020      Change and Event Occurrence*. Oxford, New York: Oxford University Press.
- 1021      Smith, P. L., & Little, D. R. (2018). Small is beautiful: In defense of the small-N design.  
1022      *Psychonomic Bulletin & Review*, 25(6), 2083–2101.  
1023      <https://doi.org/10.3758/s13423-018-1451-8>
- 1024      Steele, F., Goldstein, H., & Browne, W. (2004). A general multilevel multistate competing  
1025      risks model for event history data, with an application to a study of contraceptive use  
1026      dynamics. *Statistical Modelling*, 4(2), 145–159.  
1027      <https://doi.org/10.1191/1471082X04st069oa>
- 1028      Teachman, J. D. (1983). Analyzing social processes: Life tables and proportional hazards  
1029      models. *Social Science Research*, 12(3), 263–301.  
1030      [https://doi.org/10.1016/0049-089X\(83\)90015-7](https://doi.org/10.1016/0049-089X(83)90015-7)

- 1031 Tong, C. (2019). Statistical Inference Enables Bad Science; Statistical Thinking Enables  
1032 Good Science. *The American Statistician*, 73(sup1), 246–261.  
1033 <https://doi.org/10.1080/00031305.2018.1518264>
- 1034 Townsend, J. T. (1990). Truth and consequences of ordinal differences in statistical  
1035 distributions: Toward a theory of hierarchical inference. *Psychological Bulletin*, 108(3),  
1036 551–567. <https://doi.org/10.1037/0033-2909.108.3.551>
- 1037 Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics.  
1038 *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)
- 1039 Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New  
1040 York. Retrieved from <https://ggplot2.tidyverse.org>
- 1041 Wickham, H. (2023a). *Forcats: Tools for working with categorical variables (factors)*.  
1042 Retrieved from <https://forcats.tidyverse.org/>
- 1043 Wickham, H. (2023b). *Stringr: Simple, consistent wrappers for common string operations*.  
1044 Retrieved from <https://stringr.tidyverse.org>
- 1045 Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., ...  
1046 Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43),  
1047 1686. <https://doi.org/10.21105/joss.01686>
- 1048 Wickham, H., Çetinkaya-Rundel, M., & Grolemund, G. (2023). *R for data science: Import,  
1049 tidy, transform, visualize, and model data* (2nd edition). Beijing Boston Farnham  
1050 Sebastopol Tokyo: O'Reilly.
- 1051 Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *Dplyr: A  
1052 grammar of data manipulation*. Retrieved from <https://dplyr.tidyverse.org>
- 1053 Wickham, H., & Henry, L. (2023). *Purrr: Functional programming tools*. Retrieved from  
1054 [https://purrr.tidyverse.org/](https://purrr.tidyverse.org)
- 1055 Wickham, H., Hester, J., & Bryan, J. (2024). *Readr: Read rectangular text data*. Retrieved  
1056 from <https://readr.tidyverse.org>
- 1057 Wickham, H., Vaughan, D., & Girlich, M. (2024). *Tidyr: Tidy messy data*. Retrieved from

- 1058 https://tidyverse.org
- 1059 Winter, B. (2019). *Statistics for Linguists: An Introduction Using R*. New York:  
1060 Routledge. <https://doi.org/10.4324/9781315165547>
- 1061 Wolkersdorfer, M. P., Panis, S., & Schmidt, T. (2020). Temporal dynamics of sequential  
1062 motor activation in a dual-prime paradigm: Insights from conditional accuracy and  
1063 hazard functions. *Attention, Perception, & Psychophysics*, 82(5), 2581–2602.  
1064 <https://doi.org/10.3758/s13414-020-02010-5>

1065

**Supplementary material**

1066 **A. Definitions of discrete-time hazard, survivor, probability mass, and**  
 1067 **conditional accuracy functions**

1068 The shape of a distribution of waiting times can be described in multiple ways (Luce,  
 1069 1991). After dividing time in discrete, contiguous time bins indexed by  $t$ , let  $RT$  be a  
 1070 discrete random variable denoting the rank of the time bin in which a particular person's  
 1071 response occurs in a particular experimental condition. Because waiting times can only  
 1072 increase, discrete-time EHA focuses on the discrete-time hazard function

$$1073 \quad h(t) = P(RT = t | RT \geq t) \quad (1)$$

1074 and the discrete-time survivor function

$$1075 \quad S(t) = P(RT > t) = [1-h(t)].[1-h(t-1)].[1-h(t-2)] \dots [1-h(1)] \quad (2)$$

1076 and not on the probability mass function

$$1077 \quad P(t) = P(RT = t) = h(t).S(t-1) \quad (3)$$

1078 nor the cumulative distribution function

$$1079 \quad F(t) = P(RT \leq t) = 1-S(t) \quad (4)$$

1080 The discrete-time hazard function of event occurrence gives you for each bin the  
 1081 probability that the event occurs (sometime) in that bin, given that the event has not  
 1082 occurred yet in previous bins. This conditionality in the definition of hazard is what makes  
 1083 the hazard function so diagnostic for studying event occurrence, as an event can physically  
 1084 not occur when it has already occurred before. While the discrete-time hazard function  
 1085 assesses the unique risk of event occurrence associated with each time bin, the  
 1086 discrete-time survivor function cumulates the bin-by-bin risks of event *nonoccurrence* to  
 1087 obtain the probability that the event occurs after bin  $t$ . The probability mass function  
 1088 cumulates the risk of event occurrence in bin  $t$  with the risks of event nonoccurrence in

1089 bins 1 to t-1. From equation 3 we find that hazard in bin t is equal to  $P(t)/S(t-1)$ .

1090 For two-choice RT data, the discrete-time hazard function can be extended with the  
 1091 discrete-time conditional accuracy function

1092  $ca(t) = P(\text{correct} \mid RT = t)$  (5)

1093 which gives you for each bin the probability that a response is correct given that it is  
 1094 emitted in time bin t (Allison, 2010; Kantowitz & Pachella, 2021; Wickelgren, 1977). This  
 1095 latter function is also known as the micro-level speed-accuracy tradeoff (SAT) function.

1096 The survivor function provides a context for the hazard function, as  $S(t-1) = P(RT >$   
 1097  $t-1) = P(RT \geq t)$  tells you on how many percent of the trials the estimate  $h(t) = P(RT =$   
 1098  $t \mid RT \geq t)$  is based. The probability mass function provides a context for the conditional  
 1099 accuracy function, as  $P(t) = P(RT = t)$  tells you on how many percent of the trials the  
 1100 estimate  $ca(t) = P(\text{correct} \mid RT = t)$  is based.

1101 While psychological RT data is typically measured in small, continuous units (e.g.,  
 1102 milliseconds), discrete-time EHA treats the RT data as interval-censored data, because it  
 1103 only uses the information that the response occurred sometime in a particular bin of time  
 1104  $(x,y]: x < RT \leq y$ . If we want to use the exact event times, then we treat time as a  
 1105 continuous variable, and let RT be a continuous random variable denoting a particular  
 1106 person's response time in a particular experimental condition. Continuous-time EHA does  
 1107 not focus on the cumulative distribution function  $F(t) = P(RT \leq t)$  and its derivative, the  
 1108 probability density function  $f(t) = F(t)'$ , but on the survivor function  $S(t) = P(RT > t)$   
 1109 and the hazard rate function  $\lambda(t) = f(t)/S(t)$ . The hazard rate function gives you the  
 1110 instantaneous *rate* of event occurrence at time point t, given that the event has not  
 1111 occurred yet.

1112 **B. Custom functions for descriptive discrete-time hazard analysis**

1113 We defined 12 custom functions that we list here.

- censor(df,timeout,bin\_width) : divide the time segment  $(0, \text{timeout}]$  in bins, identify any right-censored observations, and determine the discrete RT (time bin rank)
- ptb(df) : transform the person-trial data set to the person-trial-bin data set
- setup\_lt(ptb) : set up a life table for each level of 1 independent variable
- setup\_lt\_2IV(ptb) : set up a life table for each combination of levels of 2 independent variables
- calc\_ca(df) : estimate the conditional accuracies when there is 1 independent variable
- calc\_ca\_2IV(df) : estimate the conditional accuracies when there are 2 independent variables
- join\_lt\_ca(df1,df2) : add the  $\text{ca}(t)$  estimates to the life tables (1 independent variable)
- join\_lt\_ca\_2IV(df1, df2) : add the  $\text{ca}(t)$  estimates to the life tables (2 independent variables)
- extract\_median(df) : estimate quantiles  $S(t)._{50}$  (1 independent variable)
- extract\_median\_2IV(df) : estimate quantiles  $S(t)._{50}$  (2 independent variables)
- plot\_eha(df, subj, haz\_yaxis=1, first\_bin\_shown=1, aggregated\_data=F, Nsubj=6) : create plots of the discrete-time functions (1 independent variable), and specify the upper limit of the y-axis in the hazard plot, with which bin to start plotting, whether the data is aggregated across participants, and across how many participants
- plot\_eha\_2IV(df, subj, haz\_yaxis=1, first\_bin\_shown=1, aggregated\_data=F, Nsubj=6) : create plots of the discrete-time functions (2 independent variables), and specify the upper limit of the y-axis in the hazard plot, with which bin to start plotting, whether the data is aggregated across participants, and across how many participants

When you want to analyse simple RT data from a detection experiment with one independent variable, the functions calc\_ca() and join\_lt\_ca() should not be used, and the code to plot the conditional accuracy functions should be removed from the function

<sub>1141</sub> plot\_eha(). When you want to analyse simple RT data from a detection experiment with  
<sub>1142</sub> two independent variables, the functions calc\_ca\_2IV() and join\_lt\_ca\_2IV() should not  
<sub>1143</sub> be used, and the code to plot the conditional accuracy functions should be removed from  
<sub>1144</sub> the function plot\_eha\_2IV().

<sub>1145</sub> **C. Link functions**

<sub>1146</sub> Popular link functions include the logit link and the complementary log-log link, as  
<sub>1147</sub> shown in Figure 15.

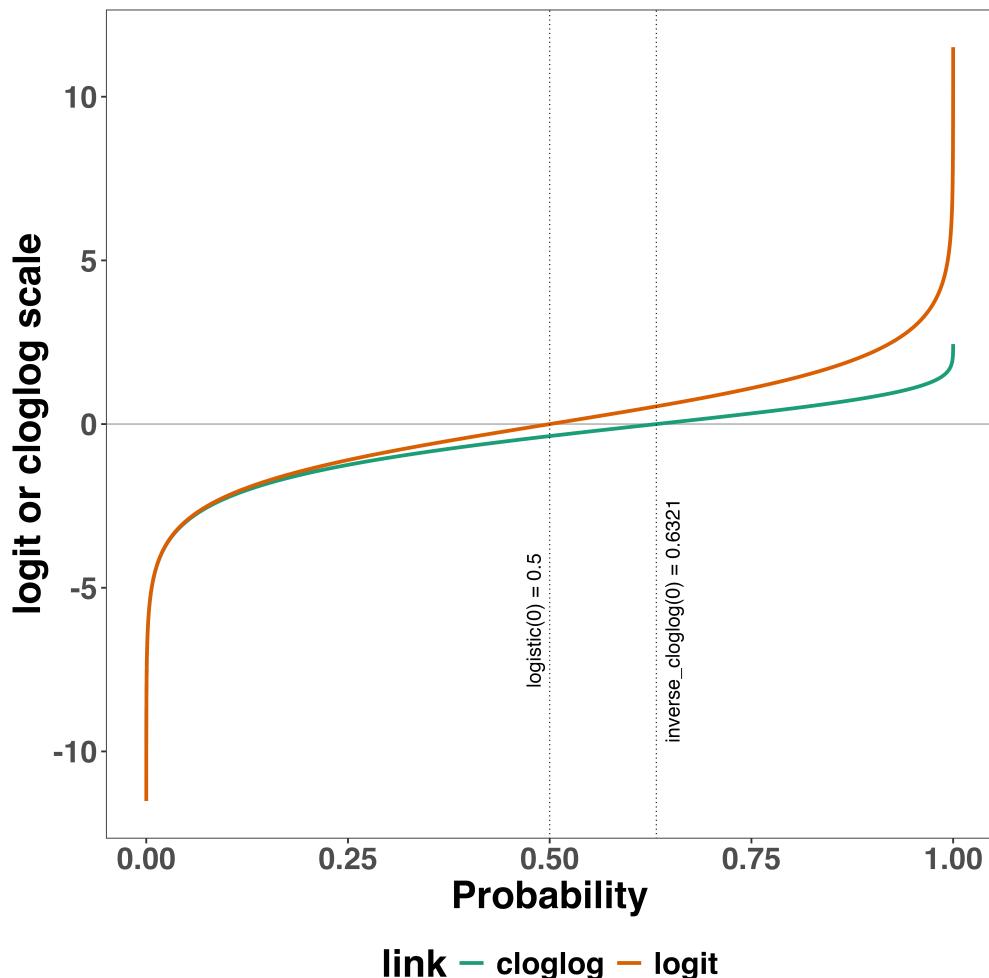


Figure 14. The logit and cloglog link functions.

<sub>1148</sub> **D. Regression equations**

<sub>1149</sub> An example (single-level) discrete-time hazard model with three predictors (TIME,  
<sub>1150</sub> X<sub>1</sub>, X<sub>2</sub>), the cloglog link function, and a second-order polynomial specification for TIME  
<sub>1151</sub> can be written as follows:

$$\begin{aligned} \text{cloglog}[h(t)] &= \ln(-\ln[1-h(t)]) = [\beta_0 \text{ONE} + \beta_1(\text{TIME}-9) + \beta_2(\text{TIME}-9)^2] + [\beta_3 X_1 + \beta_4 X_2 \\ &\quad + \beta_5 X_2(\text{TIME}-9)] \end{aligned} \quad (6)$$

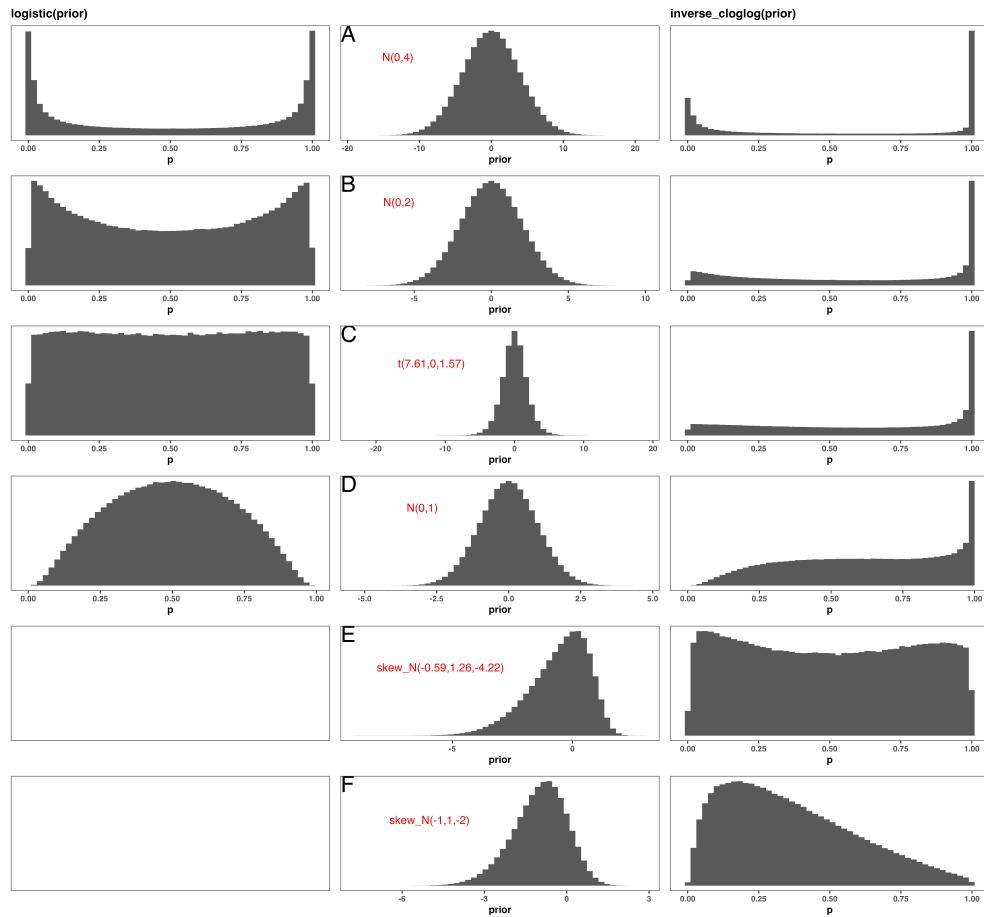
<sub>1154</sub> The main predictor variable TIME is the time bin index t that is centered on value 9  
<sub>1155</sub> in this example. The first set of terms within brackets, the parameters  $\beta_0$  to  $\beta_2$  multiplied  
<sub>1156</sub> by their polynomial specifications of (centered) time, represents the shape of the baseline  
<sub>1157</sub> cloglog-hazard function (i.e., when all predictors X<sub>i</sub> take on a value of zero). The second  
<sub>1158</sub> set of terms (the beta parameters  $\beta_3$  to  $\beta_5$ ) represents the vertical shift in the baseline  
<sub>1159</sub> cloglog-hazard for a 1 unit increase in the respective predictor variable. Predictors can be  
<sub>1160</sub> discrete, continuous, and time-varying or time-invariant. For example, the effect of a 1 unit  
<sub>1161</sub> increase in X<sub>1</sub> is to vertically shift the whole baseline cloglog-hazard function by  $\beta_3$   
<sub>1162</sub> cloglog-hazard units. However, if the predictor interacts linearly with TIME (see X<sub>2</sub> in the  
<sub>1163</sub> example), then the effect of a 1 unit increase in X<sub>2</sub> is to vertically shift the predicted  
<sub>1164</sub> cloglog-hazard in bin 9 by  $\beta_4$  cloglog-hazard units (when TIME-9 = 0), in bin 10 by  $\beta_4 +$   
<sub>1165</sub>  $\beta_5$  cloglog-hazard units (when TIME-9 = 1), and so forth. To interpret the effects of a  
<sub>1166</sub> predictor, its  $\beta$  parameter is exponentiated, resulting in a hazard ratio (due to the use of  
<sub>1167</sub> the cloglog link). When using the logit link, exponentiating a  $\beta$  parameter results in an  
<sub>1168</sub> odds ratio.

<sub>1169</sub> An example (single-level) discrete-time hazard model with a general specification for  
<sub>1170</sub> TIME (separate intercepts for each of six bins, where D1 to D6 are binary variables  
<sub>1171</sub> identifying each bin) and a single predictor (X<sub>1</sub>) can be written as follows:

$$\text{cloglog}[h(t)] = [\beta_0 D1 + \beta_1 D2 + \beta_2 D3 + \beta_3 D4 + \beta_4 D5 + \beta_5 D6] + [\beta_6 X_1] \quad (7)$$

<sub>1173</sub> **E. Prior distributions**

<sub>1174</sub> To gain a sense of what prior *logit* values would approximate a uniform distribution  
<sub>1175</sub> on the probability scale, Kurz (2023a) simulated a large number of draws from the  
<sub>1176</sub> Uniform(0,1) distribution, converted those draws to the log-odds metric, and fitted a  
<sub>1177</sub> Student's t distribution. Row C in Figure 16 shows that using a t-distribution with 7.61  
<sub>1178</sub> degrees of freedom and a scale parameter of 1.57 as a prior on the logit scale, approximates  
<sub>1179</sub> a uniform distribution on the probability scale. According to Kurz (2023a), such a prior  
<sub>1180</sub> might be a good prior for the intercept(s) in a logit-hazard model, while the N(0,1) prior in  
<sub>1181</sub> row D might be a good prior for the non-intercept parameters in a logit-hazard model, as it  
<sub>1182</sub> gently regularizes p towards .5 (i.e., a zero effect on the logit scale).



*Figure 15.* Prior distributions for the Intercept on the logit and/or cloglog scales (middle column), and their implications on the probability scale after applying the inverse-logit (or logistic) transformation (left column), and the inverse-cloglog transformation (right column).

1183 To gain a sense of what prior *cloglog* values would approximate a uniform distribution  
 1184 on the hazard probability scale, we followed Kurz's approach and simulated a large number  
 1185 of draws from the Uniform(0,1) distribution, converted them to the cloglog metric, and  
 1186 fitted a skew-normal model (due to the asymmetry of the cloglog link function). Row E  
 1187 shows that using a skew-normal distribution with a mean of -0.59, a standard deviation of  
 1188 1.26, and a skewness of -4.22 as a prior on the cloglog scale, approximates a uniform  
 1189 distribution on the probability scale. However, because hazard values below .5 are more  
 1190 likely in RT studies, using a skew-normal distribution with a mean of -1, a standard

1191 deviation of 1, and a skewness of -2 as a prior on the cloglog scale (row F), might be a good  
1192 weakly informative prior for the intercept(s) in a cloglog-hazard model.

1193 **F. Advantages of hazard analysis**

1194 Statisticians and mathematical psychologists recommend focusing on the hazard  
1195 function when analyzing time-to-event data for various reasons. First, as discussed by  
1196 Holden, Van Orden, and Turvey (2009), “probability density [and mass] functions can  
1197 appear nearly identical, both statistically and to the naked eye, and yet are clearly different  
1198 on the basis of their hazard functions (but not vice versa). Hazard functions are thus more  
1199 diagnostic than density functions” (p. 331) when one is interested in studying the detailed  
1200 shape of a RT distribution (see also Figure 1 in Panis, Schmidt, et al., 2020). Therefore,  
1201 when the goal is to study how psychological effects change over time, hazard and  
1202 conditional accuracy functions are the preferred ways to describe the RT + accuracy data.

1203 Second, because RT distributions may differ from one another in multiple ways,  
1204 Townsend (1990) developed a dominance hierarchy of statistical differences between two  
1205 arbitrary distributions A and B. For example, if  $h_A(t) > h_B(t)$  for all t, then both hazard  
1206 functions are said to show a complete ordering. Townsend (1990) concluded that stronger  
1207 conclusions can be drawn from data when comparing the hazard functions using EHA. For  
1208 example, when mean A < mean B, the hazard functions might show a complete ordering  
1209 (i.e., for all t), a partial ordering (e.g., only for  $t > 300$  ms, or only for  $t < 500$  ms), or they  
1210 may cross each other one or more times.

1211 Third, EHA does not discard right-censored observations when estimating hazard  
1212 functions, that is, trials for which we do not observe a response during the data collection  
1213 period in a trial so that we only know that the RT must be larger than some value (e.g.,  
1214 the response deadline). This is important because although a few right-censored  
1215 observations are inevitable in most RT tasks, a lot of right-censored observations are

1216 expected in experiments on masking, the attentional blink, and so forth. In other words, by  
1217 using EHA you can analyze RT data from experiments that typically do not measure  
1218 response times. As a result, EHA can also deal with long RTs in experiments without a  
1219 response deadline, which are typically treated as outliers and are discarded before  
1220 calculating a mean. This orthodox procedure leads to underestimation of the true mean.  
1221 By introducing a fixed censoring time for all trials at the end of the analysis time window,  
1222 trials with long RTs are not discarded but contribute to the risk set of each bin.

1223 Fourth, hazard modeling allows incorporating time-varying explanatory covariates  
1224 such as heart rate, electroencephalogram (EEG) signal amplitude, gaze location, etc.  
1225 (Allison, 2010). This is useful for linking physiological effects to behavioral effects when  
1226 performing cognitive psychophysiology (Meyer et al., 1988).

1227 Finally, as explained by Kelso, Dumas, and Tognoli (2013), it is crucial to first have a  
1228 precise description of the macroscopic behavior of a system (here:  $h(t)$  and possibly  $ca(t)$ )  
1229 functions) in order to know what to derive on the microscopic level. EHA can thus solve  
1230 the problem of model mimicry, i.e., the fact that different computational models can often  
1231 predict the same mean RTs as observed in the empirical data, but not necessarily the  
1232 detailed shapes of the empirical RT hazard distributions. Also, fitting parametric functions  
1233 or computational models to data without studying the shape of the empirical discrete-time  
1234  $h(t)$  and  $ca(t)$  functions can miss important features in the data (Panis, Moran, et al.,  
1235 2020; Panis & Schmidt, 2016).