

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

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

4 <sup>1</sup> ETH Zürich

5 Author Note

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

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

11

## Abstract

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

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

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

34

## 1. Introduction

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

37 In experimental psychology, it is standard practice to analyse response times (RTs),  
38 saccade latencies, and fixation durations by calculating average performance across a series  
39 of trials. Such comparisons between means have been the workhorse of experimental  
40 psychology over the last century, and have had a substantial impact on theory development  
41 as well as our understanding of the structure of cognition and brain function. Indeed, the  
42 view that mean values are truth and variations around the mean are error is deeply  
43 ingrained in experimental psychology (Bolger, Zee, Rossignac-Milon, & Hassin, 2019).

44 However, differences in mean RT conceal important pieces of information, such as when an  
45 experimental effect starts, how it evolves with increasing waiting time, and whether its  
46 onset is time-locked to other events (Panis, 2020; Panis, Moran, Wolkersdorfer, & Schmidt,  
47 2020; Panis & Schmidt, 2016, 2022; Panis, Torfs, Gillebert, Wagemans, & Humphreys,  
48 2017; Panis & Wagemans, 2009; Wolkersdorfer, Panis, & Schmidt, 2020). Such information  
49 is useful not only for the interpretation of experimental effects under investigation, but also  
50 for cognitive psychophysiology and computational model selection (Panis, Schmidt,  
51 Wolkersdorfer, & Schmidt, 2020).

52 As a simple illustration, Figure 1 summarises simulated single-subject data (200 trials  
53 per condition) that shows how comparing means between two conditions can conceal the  
54 shapes of the underlying RT and accuracy distributions. Indeed, compared to the  
55 aggregation of data across trials (Figure 1A), a distributional approach offers the  
56 possibility to reveal the time course of psychological states (Figure 1B).

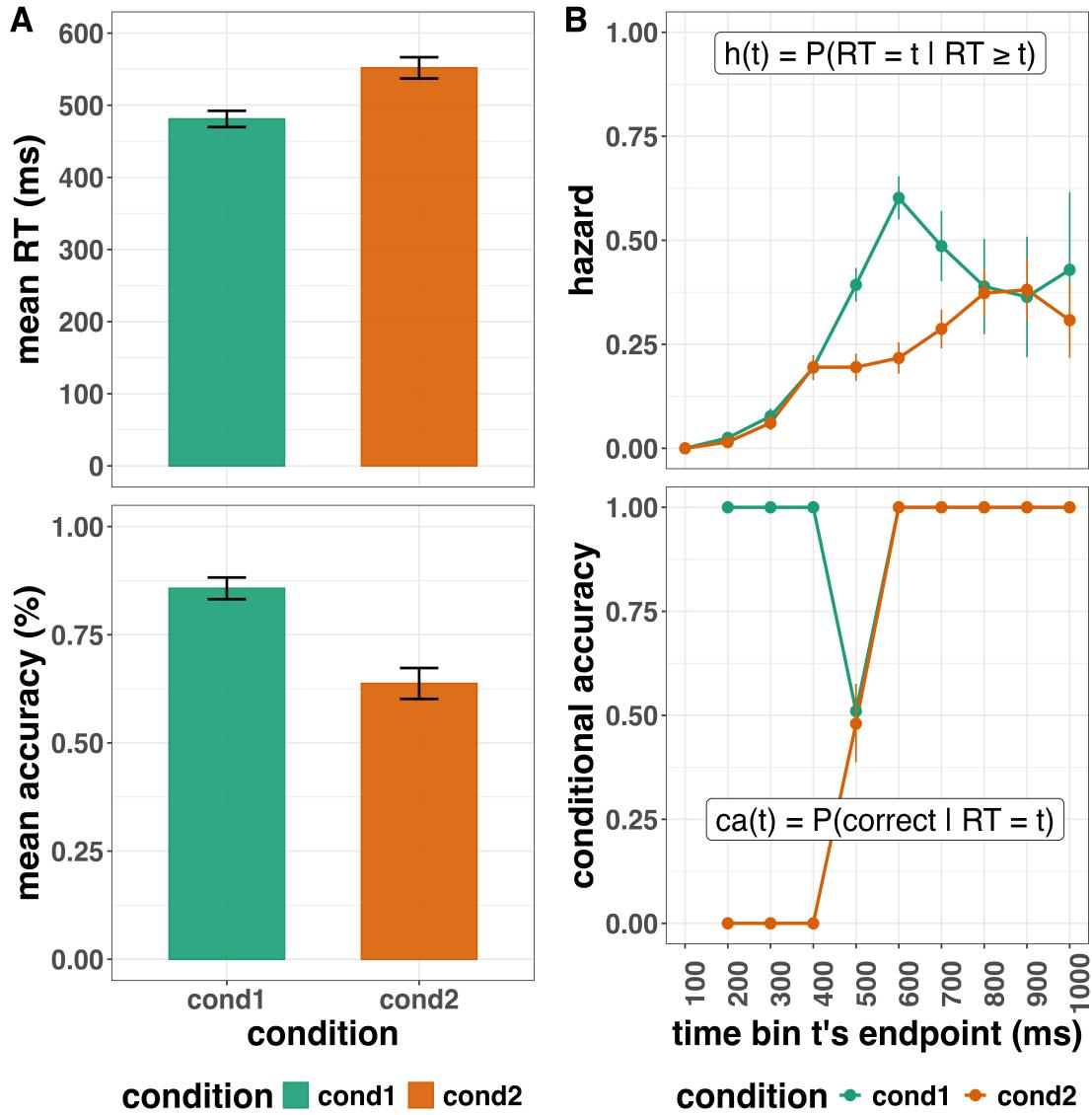


Figure 1. Simulated data showing mean performance versus distributional (EHA/SAT) analyses. (A) The mean RT (top) and overall accuracy (bottom) for two conditions are plotted. (B) The discrete-time hazard functions (top) and conditional accuracy functions (bottom) are plotted for the same data. The first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

57 For example, Figure 1B shows a first state (up to 400 ms after target onset) for which  
58 the early upswing in hazard is equal for both conditions, and the emitted responses are  
59 always correct in condition 1 and always incorrect in condition 2. In a second state (400 to  
60 500 ms), hazard is higher in condition 1, and conditional accuracies are close to .5 in both  
61 conditions. In a third state (>500 ms), the effect disappears in hazard, and all conditional  
62 accuracies are equal to 1 (see also Panis & Schmidt, 2016).

63 Why does this matter for research in psychology? For many psychological questions,  
64 the estimation of such “temporal states” information can be theoretically meaningful by  
65 leading to more fine-grained understanding of psychological processes. Because EHA adds  
66 a relatively under-used but ever-present dimension – the passage of time – to the theory  
67 building toolkit, it provides one possible answer to the recent call for a temporal science of  
68 behavior (Abney, Fausey, Suarez-Rivera, & Tamis-LeMonda, 2025).

## 69 1.2 Aims

70 Our ultimate aim in this paper is twofold: first, we want to convince readers of the  
71 many benefits of using EHA when dealing with psychological RT data, and second, we  
72 want to provide a set of practical tutorials, which provide step-by-step instructions on how  
73 you actually perform a discrete-time EHA on RT data, as well as a complementary  
74 discrete-time speed-accuracy tradeoff (SAT) analysis on timed accuracy data in case of  
75 choice RT data (Figure 1B).

76 Even though EHA is a widely used statistical tool and there already exist many  
77 excellent reviews (Allison, 1982; Blossfeld & Rohwer, 2002; Box-Steffensmeier, 2004;  
78 Hosmer, Lemeshow, & May, 2011; e.g., Singer & Willett, 2003; Teachman, 1983) and  
79 tutorials (e.g., Allison, 2010; Landes, Engelhardt, & Pelletier, 2020), we are not aware of  
80 any tutorials that are aimed specifically at psychological RT (+ accuracy) data, and which  
81 provide worked examples of the key data processing and Bayesian multilevel regression

82 modelling steps. Set within this context, our overall aim is to introduce a set of tutorials,  
83 which explain **how** to do such analyses in the context of experimental psychology, rather  
84 than repeat in any detail **why** you may do them. Therefore, we hope that our tutorials will  
85 provide a pathway for research avenues in experimental psychology that have the potential  
86 to benefit from using EHA in the future.

87 **1.3 Structure**

88 In what follows, the paper is organised in three main sections. In Section 2, we  
89 provide a brief overview of EHA to orient the reader to the basic concepts that we will use  
90 throughout the paper and why such an approach might be relevant for research in  
91 experimental psychology. In Section 3, we outline a series of tutorials, which are written in  
92 the R programming language and publicly available on our Github page  
93 ([https://github.com/sven-panis/Tutorial\\_Event\\_History\\_Analysis](https://github.com/sven-panis/Tutorial_Event_History_Analysis)), along with all of the  
94 other code and material associated with the project. The tutorials provide hands-on,  
95 concrete examples of key parts of the analytical process, such as data wrangling, model  
96 fitting and planning future studies, so that others can apply EHA to their own  
97 time-to-event data measured in RT tasks. In Section 4, we discuss the strengths and  
98 weaknesses of the approach for researchers in experimental psychology.

99 **2. What is event history analysis and why is it relevant to research in**  
100 **experimental psychology?**

101 **2.1 A brief introduction to event history analysis**

102 EHA is a class of statistical approaches to study the occurrence and timing of events,  
103 such as disease onset, marriages, arrests, and job terminations (Allison, 2010). In this  
104 section, we want to provide an intuition regarding how EHA works in general, as well as in  
105 the context of experimental psychology. For those who want more detailed treatment of

106 EHA and/or regression equations, we refer the reader to several excellent textbooks on  
107 these topics (Allison, 2010; Gelman, Hill, & Vehtari, 2020; Singer & Willett, 2003; Winter,  
108 2019). We also visualize and discuss the types of time-to-event data that are obtained in  
109 typical RT tasks and supply relevant regression equations in sections A and E of the  
110 Supplemental Material, respectively.

111 **2.1.1 Terminology and minimum requirements for EHA.** To avoid possible  
112 confusion in terminology used, it is worth noting that EHA is known by various labels,  
113 such as survival analysis, hazard analysis, duration analysis, failure-time analysis, and  
114 transition analysis (Singer & Willett, 2003). In this paper, we choose to use the term EHA  
115 throughout.

116 In terms of minimum requirements to apply a single-event EHA, one must be able to:

- 117 1. define an event of interest that represents a qualitative change - a transition from one  
118 discrete state to another - that can be situated in time (e.g., a button press, a  
119 saccade onset, a fixation offset, etc.);
- 120 2. define time point zero in each trial (e.g., target stimulus onset, fixation onset, etc.);
- 121 3. measure the passage of time between time point zero and event occurrence in discrete  
122 or continuous time units in each trial.

123 **2.1.2 Types of EHA.** There are different types of EHA. For example, the  
124 definition of hazard and the type of models employed depend on whether one is using  
125 continuous or discrete time units. As a lab, and mainly for practical reasons, we have much  
126 more experience using discrete-time EHA, and that is the approach that we describe and  
127 focus on in this paper. This choice may seem counter-intuitive, given that RT is typically  
128 treated as a continuous variable. However, continuous forms of EHA require much more  
129 data to estimate the continuous-time hazard (rate) function well (REFS). Thus, by trading  
130 a bit of temporal resolution for a lower number of trials, discrete-time methods seem ideal

131 for dealing with typical psychological RT data sets for which there are less than ~200 trials  
132 per condition per experiment (REF PANIS). Moreover, as indicated by Allison (2010),  
133 learning discrete-time EHA methods first will help in learning continuous-time methods, so  
134 it seems like a good starting point.

135 To apply discrete-time EHA, one divides the within-trial time in discrete, contiguous  
136 time bins indexed by  $t$  (e.g.,  $t = 1:10$  time bins; Figure 1B). Then let  $RT$  be a discrete  
137 random variable denoting the rank of the time bin in which a particular person's response  
138 occurs in a particular trial (i.e., repeated measure). For example, a response in one trial  
139 might occur at 546 ms and it would be in time bin 6 (any RTs from 501 ms to 600 ms).  
140 One then calculates the sample-based estimate of the discrete-time hazard function of  
141 event occurrence for each experimental condition (Figure 1B top). The discrete-time  
142 hazard function gives you, for each time bin, the conditional probability that the event  
143 occurs (sometime) in bin  $t$ , given that the event does not occur in previous bins. In other  
144 words, it reflects the instantaneous risk that the event occurs in the current bin  $t$ , given  
145 that it has not yet occurred in the past, i.e., in one of the prior bins ( $t-1, t-2, \dots, 1$ ).

146 In the context of experimental psychology, it is often (but not always), the case that  
147 responses can be classified as correct or incorrect. In those cases, one can also calculate the  
148 conditional accuracy function (Figure 1B bottom). The conditional accuracy function gives  
149 you for each time bin the conditional probability that a response is correct given that it is  
150 emitted in time bin  $t$  (Allison, 2010; Kantowitz & Pachella, 2021; Wickelgren, 1977). The  
151  $ca(t)$  function is also known as the micro-level speed-accuracy tradeoff (SAT) function. We  
152 refer to this extended (hazard + conditional accuracy) analysis for choice RT data as  
153 EHA/SAT. The definitions of these and other discrete-time functions are given in section B  
154 of the Supplemental Material.

**155 2.2 Benefits of event history analysis for research in experimental psychology**

156 Statisticians and mathematical psychologists recommend focusing on the hazard  
157 function when analyzing time-to-event data for various reasons (Holden, Van Orden, &  
158 Turvey, 2009; Luce, 1991; Townsend, 1990). We do not cover these benefits in detail here,  
159 as these are more general topics that have been covered elsewhere in textbooks (see also  
160 section F of the Supplemental Material). Instead, here we focus on the benefits as we see  
161 them for common research programmes in experimental psychology.

162 We highlight three benefits that we think are relevant to the domain of experimental  
163 psychology. First, as illustrated in Figure 1, compared to averaging data across trials,  
164 integrating results between hazard functions and their associated conditional accuracy  
165 functions for choice RT data can be informative for understanding psychological processes,  
166 in terms of inferences about the microgenesis and temporal organization of cognition and  
167 theoretical development. As such, the approach permits different kinds of questions to be  
168 asked, different inferences to be made, and it holds the potential to discriminate between  
169 theoretical accounts of psychological and/or brain-based processes. For example, what kind  
170 of theory or set of mechanisms could account for the shape of the functions and the  
171 temporally localized effects reported in Figure 1B (Panis & Schmidt, 2016)? Are there new  
172 auxiliary assumptions that computational models need to adopt (Panis, Moran, et al.,  
173 2020)? Will the temporal effect patterns align nicely with EEG findings (Panis & Schmidt,  
174 2022)? And are there new experiments that need to be performed to test the novel  
175 predictions that follow from these analyses?

176 Second, compared to more conventional analytical approaches, EHA uses more of the  
177 data because it deals with missing data differently. It is conventional with RT data to either  
178 (a) use a response deadline and discard all trials without a response, or (b) wait in each  
179 trial until a response occurs and then apply data trimming techniques, i.e., discarding too  
180 short or too long RTs (and perhaps also erroneous responses) before calculating a mean RT

181 (Berger & Kiefer, 2021). Discarding data can introduce biases, however. Rather than treat  
182 non-responses as missing data, EHA treats such trials as *right-censored* observations on the  
183 variable RT, because all we know is that RT is greater than some value. Right-censoring is  
184 a type of missing data problem and a nearly universal feature of survival data including RT  
185 data. For example, if the censoring time was 1 second, then some trials result in observed  
186 event times (those with a RT below 1 second), while the other trials result in response  
187 times that are right-censored at 1 second. The fact that EHA can deal with  
188 right-censoring, therefore, presents a analytical strength of the approach compared to many  
189 common approaches in experimental psychology (ANOVA, linear regression, delta plots).

190 Third, the approach is generalisable and applicable to many tasks that are commonly  
191 used in experimental psychology, such as detection, discrimination and bistable perception  
192 tasks, and to a range of common experimental manipulations, such as  
193 stimulus-onset-asynchrony (see section A of the Supplemental Material). The upshot is  
194 that one general analytical approach, which holds several potential advantages, is widely  
195 applicable to many substantive use-cases in the domain of experimental psychology,  
196 irrespective of the analyst's current view on the nature of cognition (Barack & Krakauer,  
197 2021).

### 198 **2.3 Implications for research design in experimental psychology**

199 Performing EHA in experimental psychology has implications for how experiments  
200 are designed. More specifically, we consider three implications that researchers will need to  
201 consider when using discrete-time EHA.

202 First, one can use a response deadline in each trial because EHA deals with  
203 right-censored observations.

204 Second, since the number of trials per condition are spread across bins, it is  
205 important to have a relatively large number of trial repetitions per participant and per

206 condition. Accordingly, experimental designs using this approach typically focus on  
207 factorial, within-subject designs, in which a large number of observations are made on a  
208 relatively small number of participants (so-called small- $N$  designs). This approach  
209 emphasizes the precision and reproducibility of data patterns at the individual participant  
210 level to increase the inferential validity of the design (Baker et al., 2021; Smith & Little,  
211 2018). Note that because statistical power derives both from the number of participants  
212 and from the number of repeated measures per participant and condition, small- $N$  designs  
213 can still achieve what are generally considered acceptable levels of statistical power, if they  
214 have a sufficient amount of data overall (Baker et al., 2021; Smith & Little, 2018).

215 Third, the width of each time bin will need to be determined. For instance, in Figure  
216 1B we chose 100ms in an arbitrary manner. In reality, however, bin width will need to be  
217 set by considering a number of factors simultaneously. The optimal bin width will depend  
218 on (a) the length of the observation period in each trial, (b) the rarity of event occurrence,  
219 (c) the number of repeated measures (or trials) per condition per participant, and (d) the  
220 shape of the hazard function. Finding an appropriate bin width in a given user case before  
221 fitting models will require testing a number of options, when calculating and plotting the  
222 descriptive statistics (see section 3.1). The goal is to find the smallest bin width that is  
223 supported by the amount of data available. Based on our experience, a bin width of 50 ms  
224 is a good starting value when the number of repeated measures is 100 or less. Too small  
225 bin widths will result in erratic hazard functions as many bins will have no events, and  
226 thus hazard estimates of zero. Interestingly, the time bins do not need to have the same  
227 width. For example, Panis (2020) used larger bins towards the end of the observation  
228 period, as fewer events occurred there.

### 229 3. Tutorials

230 Tutorials 1a and 1b show how to calculate and plot the descriptive statistics of  
231 EHA/SAT when there are one or two independent variables, respectively. Tutorials 2a and

232 2b illustrate how to use Bayesian multilevel modeling to fit hazard and conditional  
233 accuracy models, respectively. Tutorials 3a and 3b show how to implement, respectively,  
234 multilevel models for hazard and conditional accuracy in the frequentist framework.  
235 Tutorial 4 shows how to use simulation and power analysis for planning experiments.  
236 Additionally, to further simplify the process for other users, the first two tutorials rely on a  
237 set of our own custom functions that make sub-processes easier to automate, such as data  
238 wrangling and plotting functions (see section C in the Supplemental Material for a list of  
239 the custom functions).

240 Our list of tutorials is as follows:

- 241 • 1a. Wrangle raw data and calculate descriptive stats for one independent variable
- 242 • 1b. Wrangle raw data and calculate descriptive stats for two independent variables
- 243 • 2a. Bayesian multilevel modeling for  $h(t)$
- 244 • 2b. Bayesian multilevel modeling for  $ca(t)$
- 245 • 3a. Frequentist multilevel modeling for  $h(t)$
- 246 • 3b. Frequentist multilevel modeling for  $ca(t)$
- 247 • 4. Simulation and power analysis for planning experiments

248 We used R (Version 4.4.0; R Core Team, 2024) and the R-packages *bayesplot* (Version  
249 1.11.1; Gabry, Simpson, Vehtari, Betancourt, & Gelman, 2019), *brms* (Version 2.22.0;  
250 Bürkner, 2017, 2018, 2021), *citr* (Version 0.3.2; Aust, 2019), *cmdstanr* (Version 0.8.1.9000;  
251 Gabry, Češnovar, Johnson, & Brønner, 2024), *dplyr* (Version 1.1.4; Wickham, François,  
252 Henry, Müller, & Vaughan, 2023), *forcats* (Version 1.0.0; Wickham, 2023a), *ggplot2*  
253 (Version 3.5.1; Wickham, 2016), *lme4* (Version 1.1.35.5; Bates, Mächler, Bolker, & Walker,  
254 2015), *lubridate* (Version 1.9.3; Grolemund & Wickham, 2011), *Matrix* (Version 1.7.1;  
255 Bates, Maechler, & Jagan, 2024), *nlme* (Version 3.1.166; Pinheiro & Bates, 2000), *papaja*  
256 (Version 0.1.3; Aust & Barth, 2024), *patchwork* (Version 1.3.0; Pedersen, 2024), *purrr*  
257 (Version 1.0.2; Wickham & Henry, 2023), *RColorBrewer* (Version 1.1.3; Neuwirth, 2022),

258 *Rcpp* (Eddelbuettel & Balamuta, 2018; Version 1.0.13.1; Eddelbuettel & François, 2011),  
 259 *readr* (Version 2.1.5; Wickham, Hester, & Bryan, 2024), *RJ-2021-048* (Bengtsson, 2021),  
 260 *rstan* (Version 2.32.6; Stan Development Team, 2024), *standist* (Version 0.0.0.9000; Girard,  
 261 2024), *StanHeaders* (Version 2.32.10; Stan Development Team, 2020), *stringr* (Version  
 262 1.5.1; Wickham, 2023b), *tibble* (Version 3.2.1; Müller & Wickham, 2023), *tidybayes*  
 263 (Version 3.0.7; Kay, 2024), *tidyverse* (Version 2.0.0; Wickham et al., 2019) and *tinylabels* (Version 0.2.4; Barth, 2023)  
 264 for all reported analyses. The content of the tutorials, in terms of EHA and multilevel  
 265 regression modelling, is mainly based on Allison (2010), Singer and Willett (2003),  
 266 McElreath (2020), Heiss (2021), Kurz (2023a), and Kurz (2023b).

268 **3.1 Tutorial 1a: Calculating descriptive statistics using a life table**

269 **3.1.1 Data wrangling aims.** Our data wrangling procedures serve two related  
 270 purposes. First, we want to summarise and visualise descriptive statistics using a life table.  
 271 A life table includes for each time bin, the risk set (i.e., the number of trials that are  
 272 event-free at the start of the bin), the number of observed events, and the estimates of the  
 273 discrete-time hazard function  $h(t)$ , survivor function  $S(t)$ , probability mass function  $P(t)$ ,  
 274 possibly the conditional accuracy function  $ca(t)$ , and their estimated standard errors (se).  
 275 The definitions of these functions are provided in section A of the Supplemental Material.

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

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

283 ## Use 'xfun::attr2()' instead.  
 284 ## See help("Deprecated")

Table 1

*Data structure for ‘person-trial’ data*

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

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

285 In contrast, the ‘person-trial-bin’ data (Table 2) has a different, more extended  
 286 structure, which indicates in which bin a response occurred, if at all, in each trial.  
 287 Therefore, the ‘person-trial-bin’ data generates a 0 in each bin until an event occurs and  
 288 then it generates a 1 to signal an event has occurred in that bin. This data set is used  
 289 when fitting hazard models (Tutorials 2a and 3a). It is worth pointing out that there is no  
 290 requirement for an event to occur at all (in any bin), as maybe there was no response on

291 that trial or the event occurred after the time window of interest. Likewise, when the event  
292 occurs in bin 1 there would only be one row of data for that trial in the person-trial-bin  
293 data set.

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

295 ## Use 'xfun::attr2()' instead.  

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

Table 2

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

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

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

297 **3.1.2 A real data wrangling example.** To illustrate how to quickly set up life  
298 tables for calculating the descriptive statistics (functions of discrete time), we use a

299 published data set on masked response priming from Panis and Schmidt (2016). In their  
 300 first experiment, Panis and Schmidt (2016) presented a double arrow for 94 ms that  
 301 pointed left or right as the target stimulus with an onset at time point zero in each trial.  
 302 Participants had to indicate the direction in which the double arrow pointed using their  
 303 corresponding index finger, within 800 ms after target onset. Response time and accuracy  
 304 were recorded on each trial. Prime type (blank, congruent, incongruent) and mask type  
 305 were manipulated. Here we focus on the subset of trials in which no mask was presented.  
 306 The 13-ms prime stimulus was a double arrow presented 187 ms before target onset in the  
 307 congruent (same direction as target) and incongruent (opposite direction as target) prime  
 308 conditions.

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

312 The required column names are as follows:

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

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

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

```
  levels=c(1,2,3),
  labels=c("blank","congruent","incongruent")))
```

320        Next, we can set up the life tables and plots of the discrete-time functions  $h(t)$ ,  $S(t)$ ,  
 321  $ca(t)$ , and  $P(t)$ . To do so using a functional programming approach, one has to nest the  
 322 data within participants using the `group_nest()` function, and supply a user-defined  
 323 censoring time and bin width 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))
```

324        Note that the censoring time (here: 600 ms) should be a multiple of the bin width  
 325 (here: 40 ms). The censoring time should be a time point after which no informative  
 326 responses are expected anymore. In experiments that implement a response deadline in  
 327 each trial the censoring time can equal that deadline time point. Trials with a RT larger  
 328 than the censoring time, or trials in which no response is emitted during the data collection  
 329 period, are treated as right-censored observations in EHA. In other words, these trials are  
 330 not discarded, because they contain the information that the event did not occur before the

331 censoring time. Removing such trials before calculating the mean event time will result in  
332 underestimation of the true mean.

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

337 The next step is to set up the life table using our custom function setup\_lt(),  
338 calculate the conditional accuracies using our custom function calc\_ca(), add the ca(t)  
339 estimates to the life table using our custom function join\_lt\_ca(), and then plot the  
340 descriptive statistics using our custom function plot\_eha(). One can now inspect different  
341 aspects, including the life table for a particular condition of a particular subject, and a plot  
342 of the different functions for a particular participant.

343 In general, it is important to visually inspect the functions first for each participant,  
344 in order to identify individuals that may be guessing (e.g., a flat conditional accuracy  
345 function at .5 indicates that someone is just guessing), outlying individuals, and/or  
346 different groups with qualitatively different behavior. Also, to select a bin width for fitting  
347 models, one should test and compare various bin widths in the censor function, and select  
348 the smallest one that is supported by the data. Too small bin widths will result in erratic  
349 hazard functions because many bins will have estimates equal to zero.

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

```
352 ## Warning in attr(x, "align"): 'xfun::attr()' is deprecated.  
353 ## Use 'xfun::attr2()' instead.  
354 ## See help("Deprecated")
```

Table 3

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

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

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

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

## Descriptive stats for subject 6

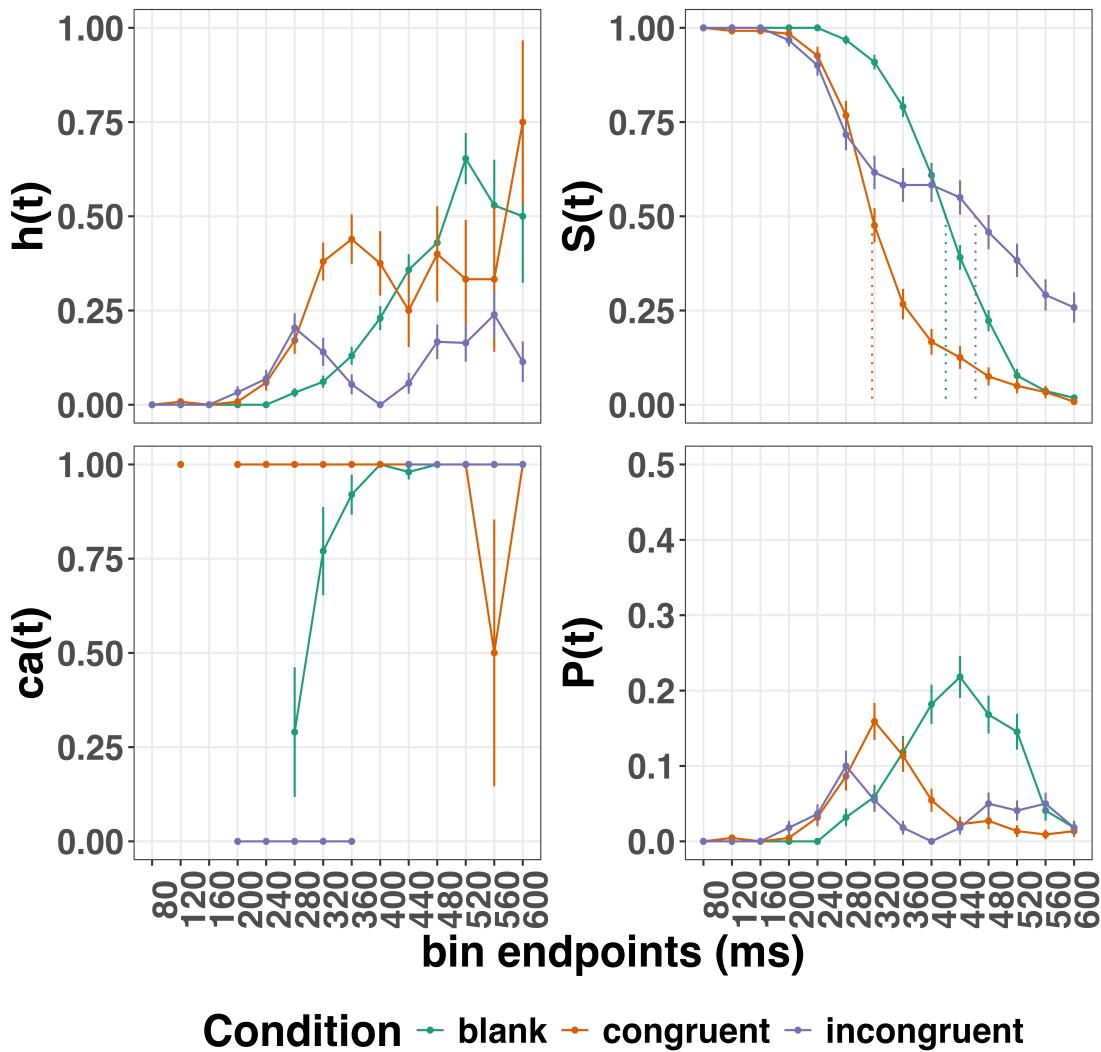


Figure 2. Estimated discrete-time hazard ( $h$ ), survivor ( $S$ ), conditional accuracy ( $ca$ ) and probability mass ( $P$ ) functions for participant 6. Vertical dotted lines indicate the estimated median RTs. Error bars represent +/- 1 standard error of the respective proportion.

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

362 0.17, and 0.20 for the blank, congruent, and incongruent prime conditions, respectively. In

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

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

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

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

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

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

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

370 respectively.

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

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

373 0.36, 0.25, and 0.06 for the blank, congruent, and incongruent prime conditions,

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

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

376 conditions, respectively.

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

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

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

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

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

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

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

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

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

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

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

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

394 **3.2 Tutorial 1b: Generalising to a more complex design**

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

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

406 **3.3 Tutorial 2a: Fitting Bayesian hazard models to discrete time-to-event data**

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

411 Willett, 2003).

412 In general, when fitting regression models, our lab adopts an estimation approach to  
413 multilevel regression (Kruschke & Liddell, 2018; Winter, 2019), which is heavily influenced  
414 by the Bayesian framework as suggested by Richard McElreath (Kurz, 2023b; McElreath,  
415 2020). We also use a “keep it maximal” approach to specifying varying (or random) effects  
416 (Barr, Levy, Scheepers, & Tily, 2013). This means that wherever possible we include  
417 varying intercepts and slopes per participant. To make inferences, we use two main  
418 approaches. We compare models of different complexity, using information criteria (e.g.,  
419 WAIC) and cross-validation (e.g., LOO), to evaluate out-of-sample predictive accuracy  
420 (McElreath, 2020). We also take the most complex model and evaluate key parameters of  
421 interest using point and interval estimates.

422 **3.3.1 Hazard model considerations.** There are several analytic decisions one  
423 has to make when fitting a discrete-time hazard model. First, one has to select an analysis  
424 time window, i.e., a contiguous set of bins for which there is enough data for each  
425 participant. Second, given that the dependent variable (event occurrence) is binary, one  
426 has to select a link function (see section C in the Supplemental Material). The cloglog link  
427 is preferred over the logit link when events can occur in principle at any time point within  
428 a bin, which is the case for RT data (Singer & Willett, 2003). Third, one has to choose  
429 whether to treat TIME (i.e., the time bin index  $t$ ) as a categorical or continuous predictor.  
430 And when you treat a variable as a categorical predictor, you can choose between reference  
431 coding and index coding. With reference coding, one defines the variable as a factor and  
432 selects one of the  $k$  categories as the reference level. Brm() will then construct  $k-1$   
433 indicator variables (see model M1d in Tutorial\_2a.Rmd for an example). With index  
434 coding, one constructs an index variable that contains integers that correspond to different  
435 categories (see models M0i and M1i below). As explained by McElreath (2020), the  
436 advantage of index coding is that the same prior can be assigned to each level of the index  
437 variable, so that each category has the same prior uncertainty.

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

In general, there are three assumptions one can make or relax when adding experimental predictor variables and other covariates: The linearity assumption for continuous predictors (the effect of a 1 unit change is the same anywhere on the scale), the additivity assumption (predictors do not interact), and the proportionality assumption (predictors do not interact with TIME).

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

**455 3.3.2 Prior distributions.** To get the posterior distribution of each model

**456** parameter given the data, we need to specify prior distributions for the model parameters  
**457** which reflect our prior beliefs. In Tutorial\_2a.Rmd we perform a few prior predictive  
**458** checks to make sure our selected prior distributions reflect our prior beliefs (Gelman,  
**459** Vehtari, et al., 2020).

**460** The middle column of Supplementary Figure 2 (section E of the Supplemental  
**461** Material) shows six examples of prior distributions for an intercept on the logit and/or  
**462** cloglog scales. While a normal distribution with relatively large variance is often used as a  
**463** weakly informative prior for continuous dependent variables, rows A and B of  
**464** Supplementary Figure 2 show that specifying such distributions on the logit and cloglog  
**465** scales actually leads to rather informative distributions on the original probability scale, as  
**466** most mass is pushed to probabilities of 0 and 1.

**467 3.3.3 Model M0i: A null model with index coding.** When you do not want to  
**468** make assumptions about the shape of the hazard function, or its shape is not smooth but  
**469** irregular, then you can use a general specification of TIME, i.e., fit one grand intercept per  
**470** time bin. In this first model, we use a general specification of TIME using index coding,  
**471** and do not include experimental predictors. We call this model “M0i”.

**472** Before we fit model M0i, we select the necessary columns from the data, and specify  
**473** 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")

```

474 After selecting the bernoulli family and the cloglog link, the model formula is  
 475 specified. The specification “0 + …” removes the default intercept in brm(). The fixed  
 476 effects include an intercept for each level of timebin. Each of these intercepts is allowed to  
 477 vary across individuals (variable pid). We request 2000 samples from the posterior  
 478 distribution for each of four chains. Estimating model M0i took about 30 minutes on a  
 479 MacBook Pro (Sonoma 14.6.1 OS, 18GB Memory, M3 Pro Chip).

480 **3.3.4 Model M1i: Adding the effects of prime-target congruency.** Previous  
 481 research has shown that psychological effects typically change over time (Panis, 2020;  
 482 Panis, Moran, et al., 2020; Panis & Schmidt, 2022; Panis et al., 2017; Panis & Wagemans,  
 483 2009). In the next model, therefore, we use index coding for both TIME (variable  
 484 “timebin”) and the categorical predictor prime-target-congruency (variable “prime”), so  
 485 that we get 30 grand intercepts, one for each combination of timebin level and prime level.  
 486 Here is the model formula of this model that we call “M1i”.

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

487 Estimating model M1i took about 124 minutes.

488 **3.3.5 Compare the models.** We can compare the two models using the Widely  
 489 Applicable Information Criterion (WAIC) and Leave-One-Out (LOO) cross-validation, and

490 look at model weights for both criteria (Kurz, 2023a; McElreath, 2020).

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

```
491 ## model_M0i model_M1i  
492 ## 0 1
```

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

```
493 ## model_M0i model_M1i  
494 ## 0 1
```

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

497 **3.3.6 Evaluating parameter estimates in model M1i.** To make inferences  
498 from the parameter estimates in model M1i, we first plot the densities of the draws from  
499 the posterior distributions of its population-level parameters in Figure 5, together with  
500 point (median) and interval estimates (80% and 95% credible intervals).

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

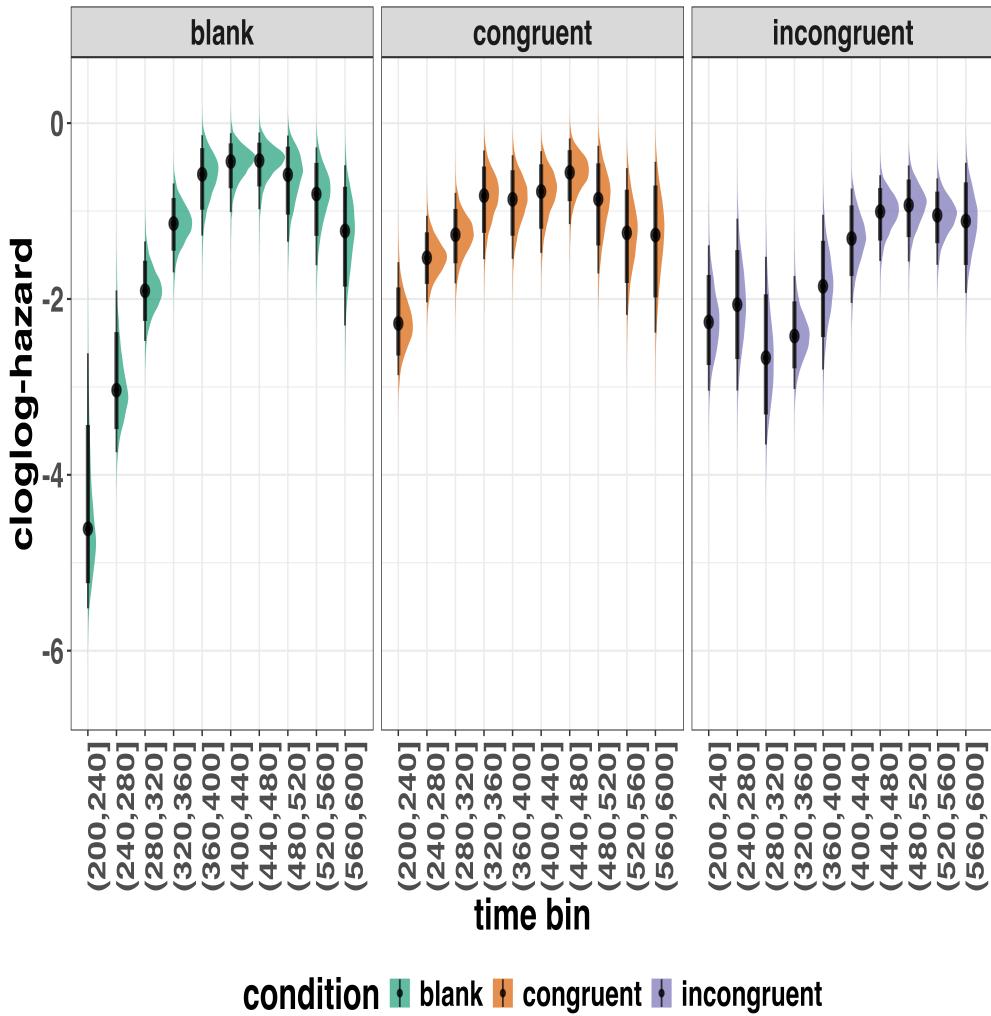


Figure 3. Medians and 80/95% credible intervals of the posterior distributions of the population-level parameters of model M1i.

501 Because the parameter estimates are on the cloglog-hazard scale, we can ease our  
 502 interpretation by plotting the expected value of the posterior predictive distribution – the  
 503 predicted hazard values – at the population level (Figure 6A), and for each participant in  
 504 the data set (Figure 6B).

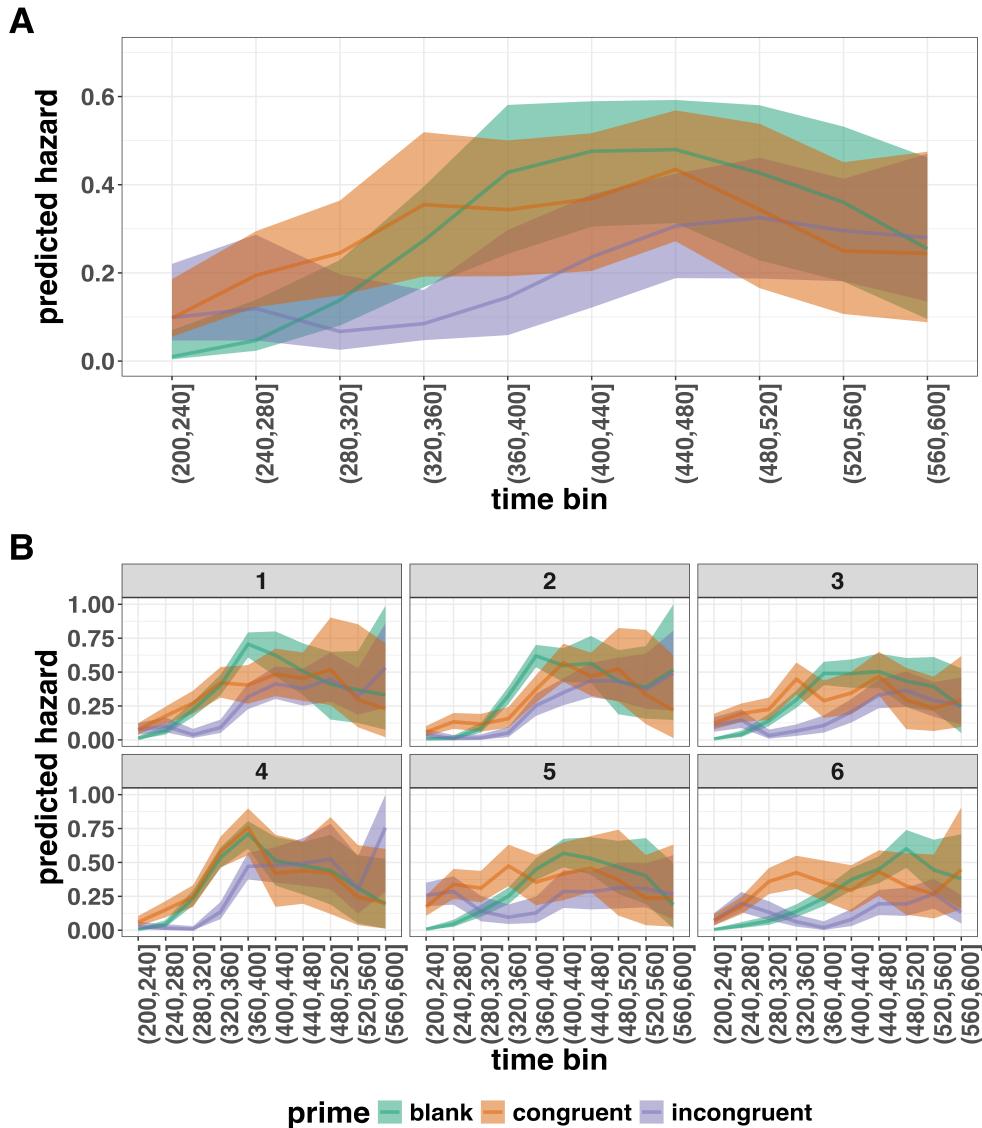


Figure 4. Point (median) and 80/95% credible interval summaries of the hazard estimates (expected values of the draws from the posterior predictive distributions) in each time bin at the population level (A), and for each participant (B).

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

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

507 incongruent-blank), and plot the posterior distributions of these contrast effects, both at

508 the population level (Figure 7A; grand average marginal effect) and at the participant level

509 (Figure 7B; subject-specific average marginal effect).

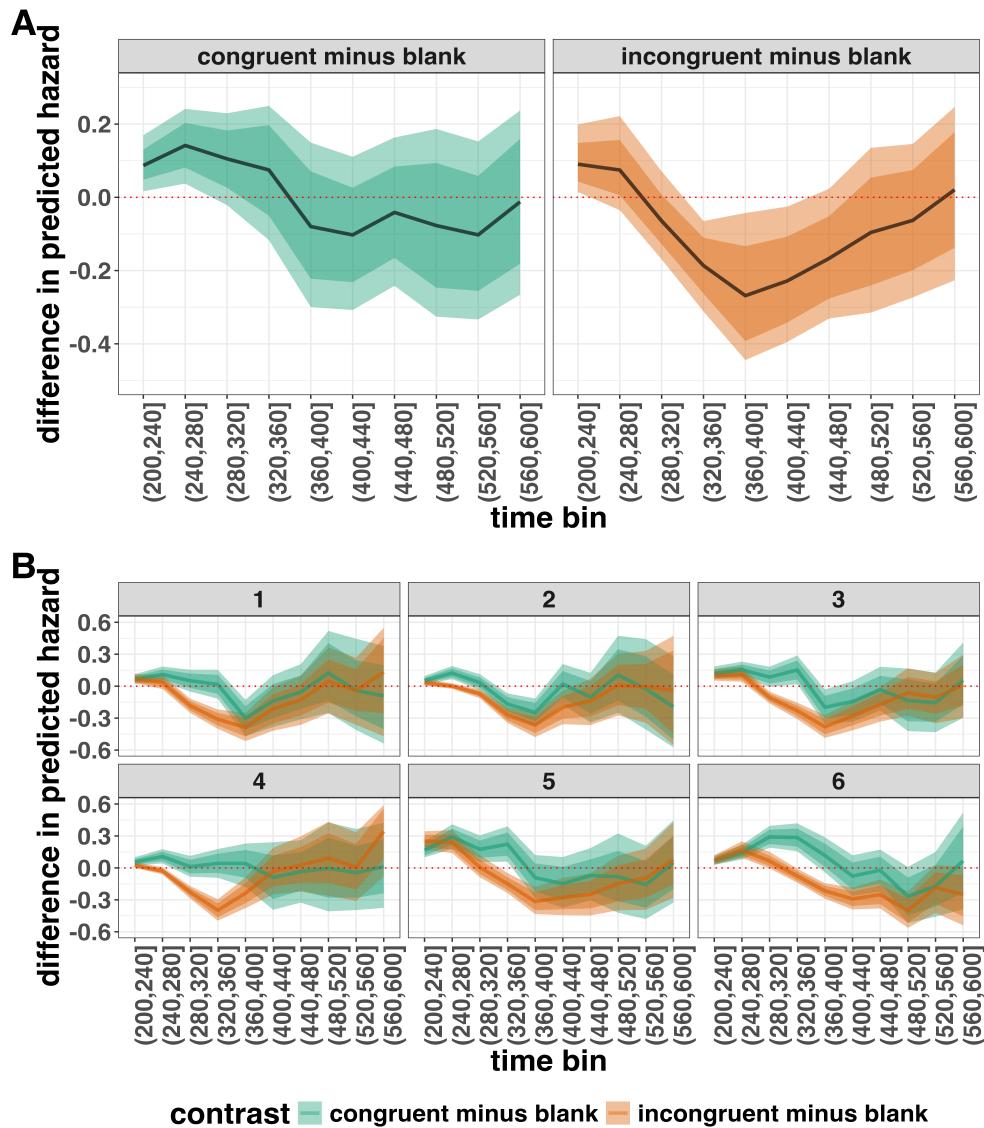


Figure 5. Point (mean) and 80/95% credible interval summaries of estimated differences in hazard in each time bin at the population level (A), and for each participant (B).

510 The point estimates and quantile intervals can be reported in a table (see  
 511 Tutorial\_2a.Rmd for details).

512 **Example conclusions for M1i.** What can we conclude from model M1i about  
 513 our research question, i.e., the temporal dynamics of the effect of prime-target congruency

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

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

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

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

537 Future studies could (a) increase the number of participants to estimate the  
538 proportion of “dippers” in the subject population, and/or (b) try to explain why this dip  
539 occurs. For example, Panis and Schmidt (2016) concluded that active, top-down,

540 task-guided response inhibition effects emerge around 360 ms after the onset of the stimulus  
541 following the prime (here: the target stimulus). Such a top-down inhibitory effect might  
542 exist in our priming data set, because after some time participants will learn that the first  
543 stimulus is not the one they have to respond to. To prevent a premature overt response to  
544 the prime they thus might gradually increase a global response threshold during the  
545 remainder of the experiment, which could result in a lower hazard in congruent trials  
546 compared to blank trials, for bins after ~360 ms, and towards the end of the experiment.  
547 This effect might be masked for incongruent primes by the response competition effect.

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

### 554 3.4 Tutorial 2b: Fitting Bayesian conditional accuracy models

555 In this fourth tutorial, we illustrate how to fit a Bayesian multilevel regression model  
556 to the timed accuracy data from the masked response priming data used in Tutorial 1a.  
557 The general process is similar to Tutorial 2a, except that (a) we use the person-trial data,  
558 (b) we use the logit link function, and (c) we change the priors. To keep the tutorial short,  
559 we only fit one conditional accuracy model, which was based on model M1i from Tutorial  
560 2a and labelled M1i\_ca.

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

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

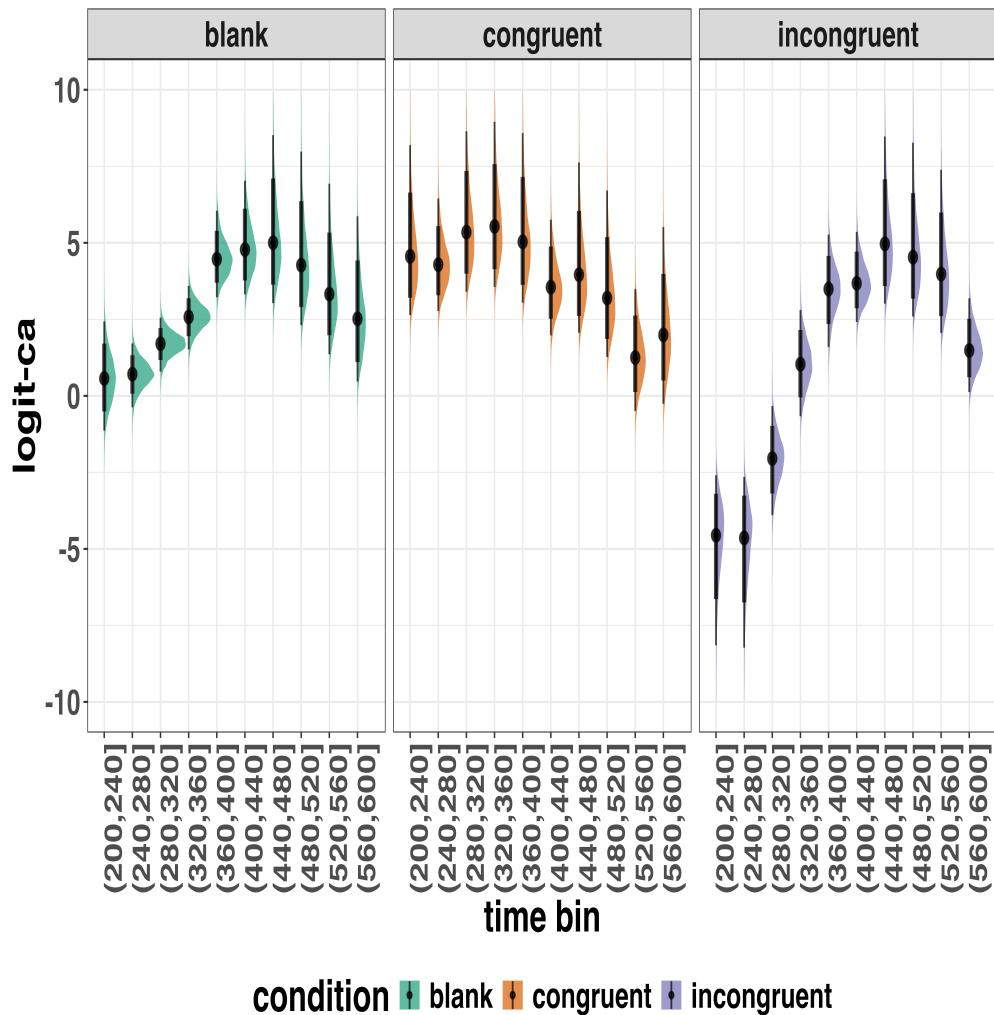


Figure 6. Medians and 80/95% credible intervals of the posterior distributions of the population-level parameters of model M1i\_ca. ca = conditional accuracy.

Because the parameter estimates are on the logit-ca scale, we can ease our

interpretation by plotting the expected value of the posterior predictive distribution – the predicted conditional accurcies – at the population level (Figure 9A), and for each participant in the data set (Figure 9B).

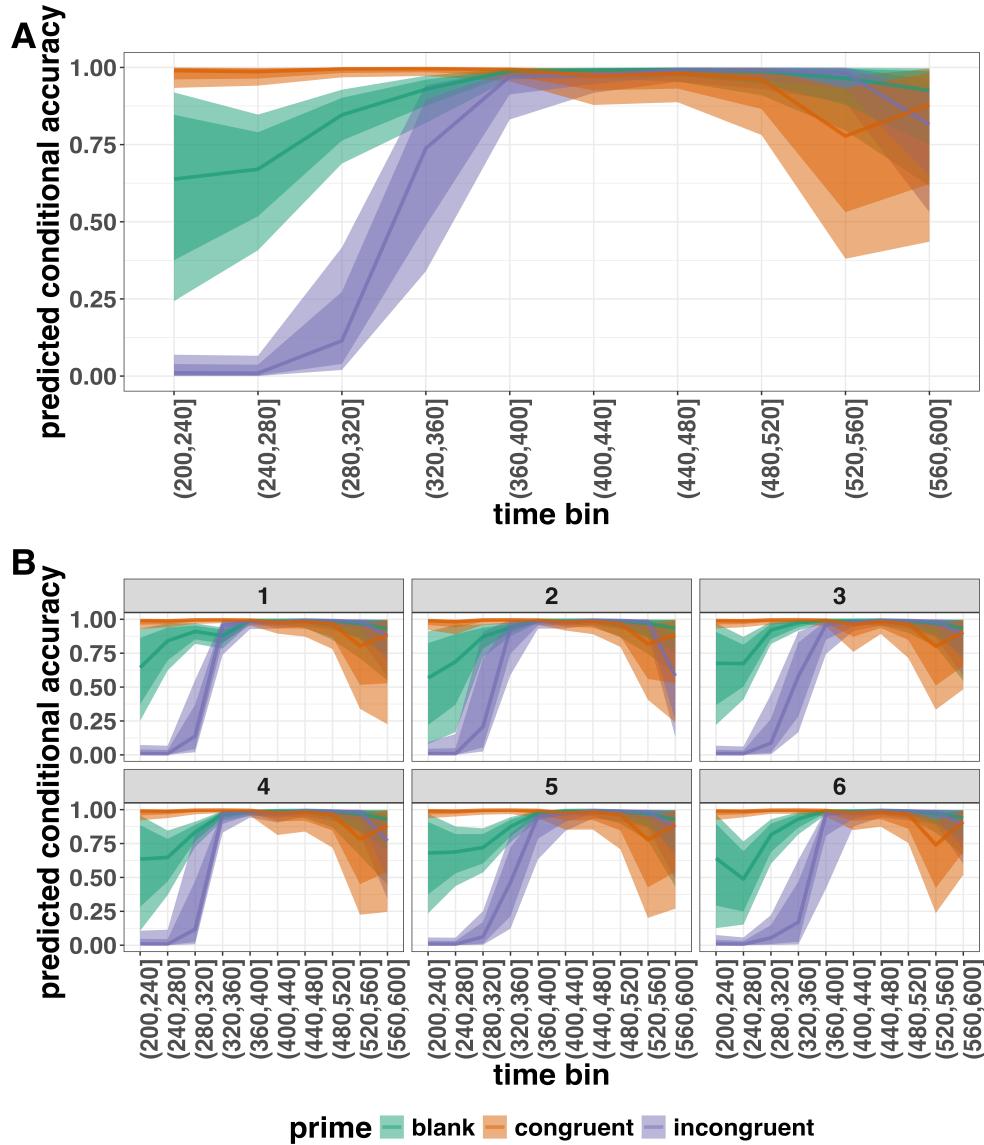


Figure 7. Point (median) and 80/95% credible interval summaries of the conditional accuracy estimates (expected values of the draws from the posterior predictive distributions) in each time bin at the population level (A), and for each participant (B).

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

570 relative to the blank prime condition, we can construct two contrasts (congruent-blank,  
 571 incongruent-blank), and plot the posterior distributions of these contrast effects at the  
 572 population level (Figure 10A) and for each participant (Figure 10B).

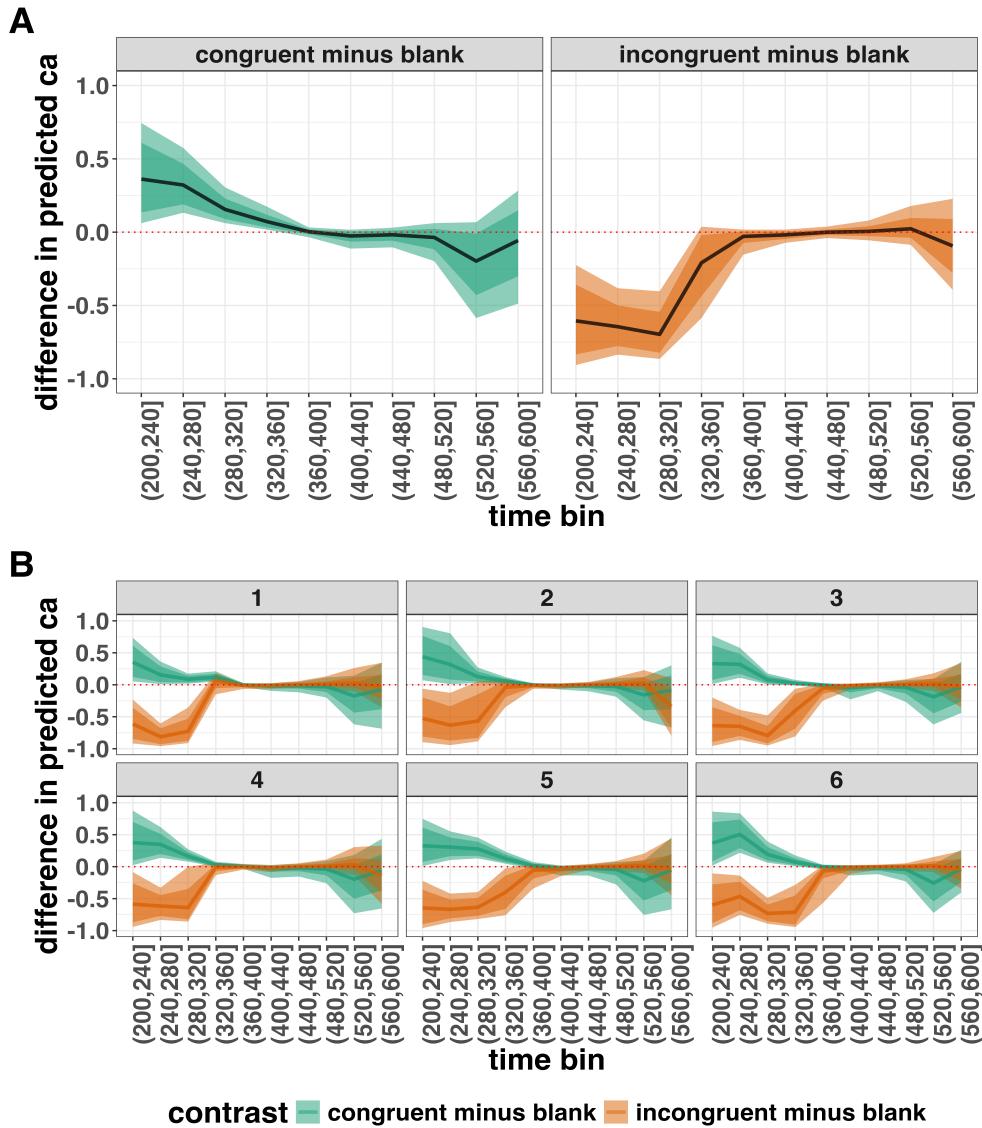


Figure 8. Point (mean) and 80/95% credible interval summaries of estimated differences in conditional accuracy in each time bin at the population level (A), and for each participant (B).

Based on Figure 10A we see that on the population level congruent primes have a

positive effect on the conditional accuracy of emitted responses in time bins (200,240],

(240,280], (280,320], and (320,360], relative to the estimates in the baseline condition

(blank prime; red dashed lines in Figure 10A). Incongruent primes have a negative effect on

577 the conditional accuracy of emitted responses in the first time bins, relative to the  
578 estimates in the baseline condition.

579 **3.5 Tutorial 4: Planning**

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

581 **3.5.1 Background.** The general approach to planning that we adopt here involves  
582 simulating reasonably structured data to help guide what you might be able to expect from  
583 your data once you collect it (Gelman, Vehtari, et al., 2020). The basic structure and code  
584 follows the examples outlined by Solomon Kurz in his ‘power’ blog posts  
585 (<https://solomonkurz.netlify.app/blog/bayesian-power-analysis-part-i/>) and Lisa  
586 DeBruine’s R package `faux{}` (<https://debruine.github.io/faux/>) as well as these related  
587 papers (DeBruine & Barr, 2021; Pargent, Koch, Kleine, Lermer, & Gaube, 2024).

588 **3.5.2 Basic workflow.** The basic workflow is as follows:

- 589 1. Fit a regression model to existing data.
- 590 2. Use the regression model parameters to simulate new data.
- 591 3. Write a function to create 1000s of datasets and vary parameters of interest (e.g.,  
592 sample size, trial count, effect size).
- 593 4. Summarise the simulated data to estimate likely power or precision of the research  
594 design options.

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

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

**3.5.3 Fit a regression model and simulate one dataset.**

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

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

We then took parameters from this model and used them to create a single dataset

with 200 trials per condition for 10 individual participants. The raw data and the simulated data are plotted in Figure 12 and show quite close correspondence, which is re-assuring. But, this is only one dataset. What we really want to do is simulate many datasets and vary parameters of interest, which is what we turn to in the next section.

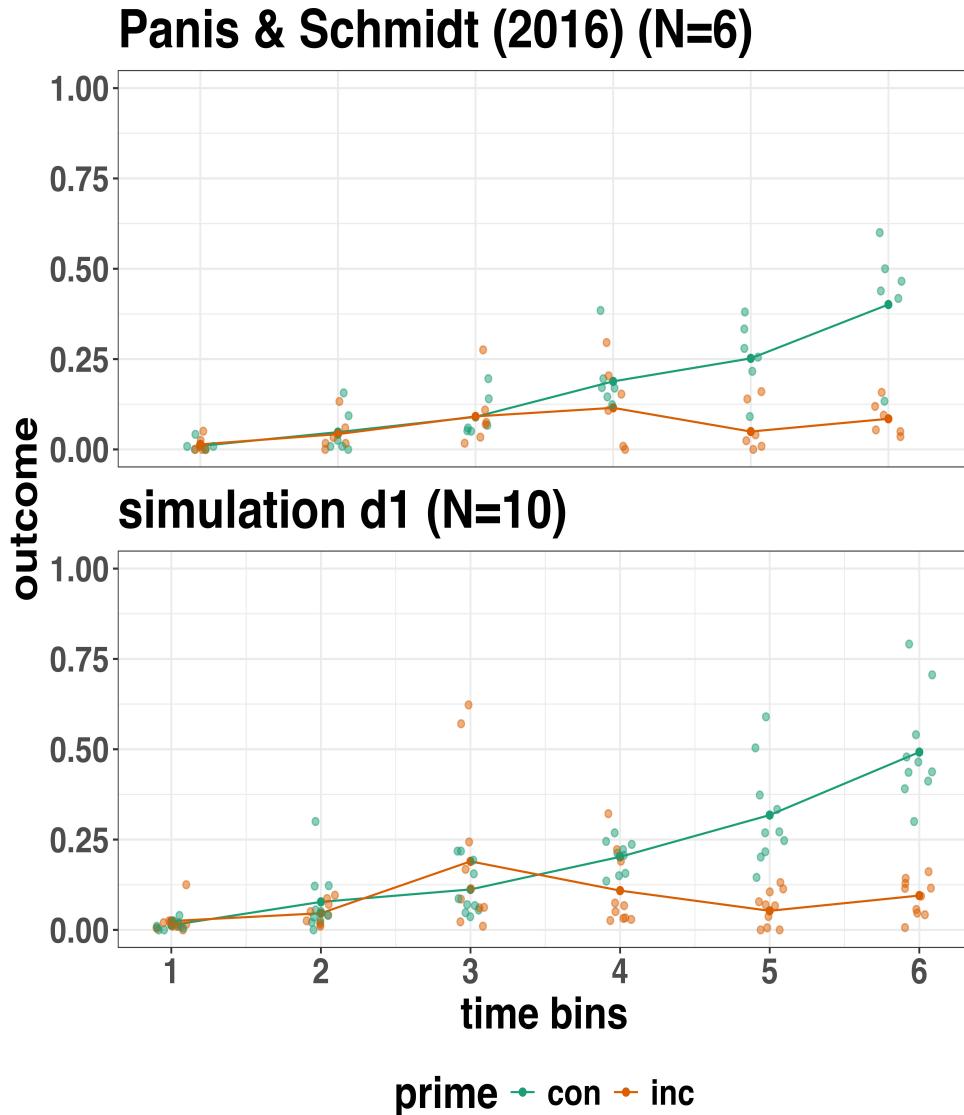


Figure 9. Raw data from Panis and Schmidt (2016) and simulated data from 10 participants.

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

613        Here we use the same data simulation process as used above, but instead of simulating one  
 614        dataset, we simulate 1000 datasets per variation in parameter values. Specifically, in  
 615        Simulation 1, we vary the number of trials per condition (100, 200, and 400), as well as the  
 616        effect size in bin 6. We focus on bin 6 only, in terms of varying the effect size, just to make  
 617        things simpler and easier to understand. The effect size observed in bin 6 in this subsample  
 618

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.

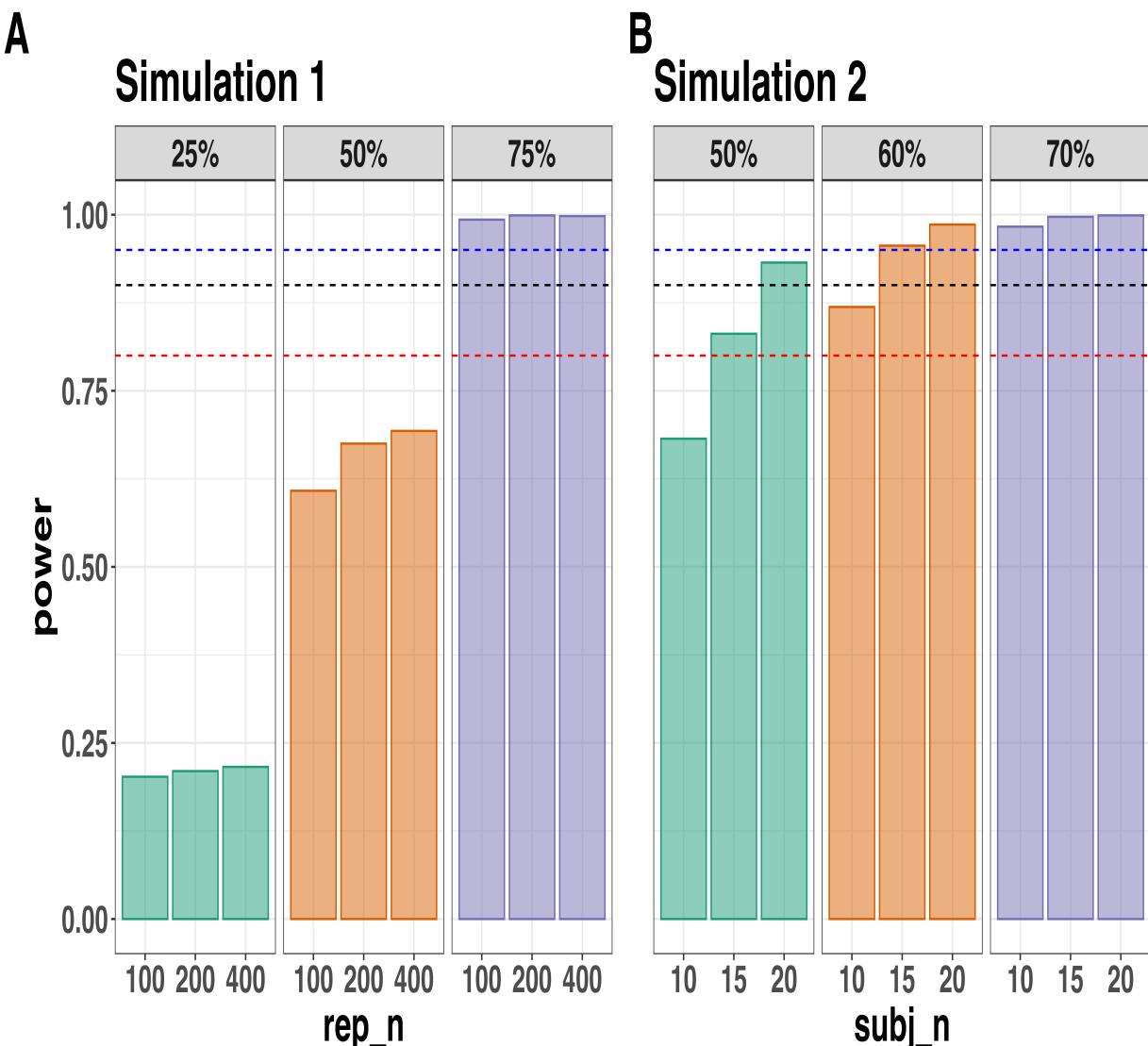
Summary results from Simulation 1 are shown in Figure 13A. Figure 13A depicts statistical “power” as calculated by the percentage of lower-bound 95% confidence intervals that exclude zero when the difference between prime condition is calculated (congruent - incongruent). In other words, what fraction 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

645 calculations might be as follows (trial count = 200 per condition in all cases):

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



*Figure 10.* 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.

649       **3.5.5 Planning decisions.** Now that we have summarised our simulated data,

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

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

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

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

654 Some considerations might be as follows:

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

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

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

658 personnel?

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

675

#### 4. Discussion

676 This main motivation for writing this paper is the observation that EHA and SAT  
677 analysis remain under-used in psychological research. As a consequence, the field of  
678 psychological research is not taking full advantage of the many benefits EHA/SAT provides  
679 compared to more conventional analyses. By providing a freely available set of tutorials,  
680 which provide step-by-step guidelines and ready-to-use R code, we hope that researchers  
681 will feel more comfortable using EHA/SAT in the future. Indeed, we hope that our  
682 tutorials may help to overcome a barrier to entry with EHA/SAT, which is that such  
683 approaches require more analytical complexity compared to mean-average comparisons.  
684 While we have focused here on within-subject, factorial, small- $N$  designs, it is important to  
685 realize that EHA/SAT can be applied to other designs as well (large- $N$  designs with only  
686 one measurement per subject, between-subject designs, etc.). As such, the general workflow  
687 and associated code can be modified and applied more broadly to other contexts and  
688 research questions. In the following, we discuss the main use-cases, issues relating to model  
689 complexity and interpretability, as well as limitations of the approach and future  
690 extensions.

691 **4.1 What are the main use-cases of EHA for understanding cognition and brain  
692 function?**

693 For those researchers, like ourselves, who are primarily interested in understanding  
694 human cognitive and brain systems, we consider two broadly-defined, main use-cases of  
695 EHA. First, as we hope to have made clear by this point, EHA is one way to investigating  
696 a “temporal states” approach to cognitive processes. EHA provides one way to uncover the  
697 microgenesis of cognitive effects, by revealing when cognitive states may start and stop,  
698 how states are replaced with others, as well as what they may be tied to or interact with.  
699 Therefore, if your research questions concern **when psychological states occur, and**

700 **how they are temporally organized**, our EHA tutorials could be useful tools for you to  
701 use.

702 Second, even if you are not primarily interested in studying the temporal  
703 organization of cognitive states, EHA could still be a useful tool to consider using, in order  
704 to qualify inferences that are being made based on comparisons between means. Given that  
705 distinctly different inferences can be made from the same data based on whether one  
706 computes a mean across trials or a RT distribution of events (Figure 1), it may be  
707 important for researchers to supplement comparisons between means with EHA. One could  
708 envisage scenarios where the implicit assumption of an effect in mean RTs manifesting  
709 across all of the time bins measured would not be supported by EHA. Therefore, the  
710 conclusion of interest would not apply to all responses, but instead it would be restricted to  
711 certain periods of within-trial time.

712 **4.2 Model complexity versus interpretability**

713 EHA can quickly become very complex when adding more than one time scale, due to  
714 the many possible higher-order interactions. For example, some of the models discussed in  
715 Tutorial 2a, which we did not focus on in the main text, contain two time scales as  
716 covariates: the passage of time on the within-trial time scale, and the passage of time on  
717 the across-trial (or within-experiment) time scale. However, when trials are presented in  
718 blocks, and blocks of trials within sessions, and when the experiment comprises three  
719 sessions, then four time scales can be defined (within-trial, within-block, within-session,  
720 and within-experiment). From a theoretical perspective, adding more than one time scale –  
721 and their interactions – can be important to capture plasticity and other learning effects  
722 that may play out on such longer time scales, and that are probably present in each  
723 experiment in general (REF DFT). From a practical perspective, therefore, some choices  
724 need to be made to balance the amount of data that is being collected per participant,  
725 condition and across the varying timescales. As one example, if there are several timescales

726 of relevance, then it might be prudent for interpretational purposes to limit the number of  
727 experimental predictor variables (conditions). This is of course where planning and data  
728 simulation efforts would be important to provide a guide to experimental design choices  
729 (see Tutorial 4).

730 **4.3 Limitations**

731 Compared to the orthodox method – comparing means between conditions – the  
732 most important limitation of multilevel hazard and conditional accuracy modeling is that it  
733 might take a long time to estimate the parameters using Bayesian methods or the model  
734 might have to be simplified significantly to use frequentist methods.

735 Another issue is that you need a relatively large number of trials per condition to  
736 estimate the hazard function with high temporal resolution, which is required when testing  
737 predictions of process models of cognition. Indeed, in general, there is a trade-off between  
738 the number of trials per condition and the temporal resolution (i.e., bin width) of the  
739 discrete-time hazard function. Therefore, we recommend researchers to collect as many  
740 trials as possible per experimental condition, given the available resources and considering  
741 the participant experience (e.g., fatigue and boredom). For instance, if the maximum  
742 session length deemed reasonable is between 1 and 2 hours, what is the maximum number  
743 of trials per condition that you could reasonably collect? After consideration, it might be  
744 worth conducting multiple testing sessions per participant and/or reducing the number of  
745 experimental conditions. Finally, there is a user-friendly online tool for calculating  
746 statistical power as a function of the number of trials as well as the number of participants,  
747 and this might be worth consulting to guide the research design process (Baker et al., 2021).

748

## 5. Conclusions

749       Estimating the temporal distributions of RT and accuracy provide a rich source of  
750      information on the time course of cognitive processing, which have been largely  
751      undervalued in the history of experimental psychology and cognitive neuroscience. We  
752      hope that by providing a set of hands-on, step-by-step tutorials, which come with  
753      custom-built and freely available code, researchers will feel more comfortable embracing  
754      EHA and investigating the shape of empirical hazard functions and the temporal profile of  
755      cognitive states. On a broader level, we think that wider adoption of such approaches will  
756      have a meaningful impact on the inferences drawn from data, as well as the development of  
757      theories regarding the structure of cognition.

758

**Author contributions**

759       Conceptualization: S. Panis and R. Ramsey; Software: S. Panis and R. Ramsey;  
760      Writing - Original Draft Preparation: S. Panis; Writing - Review & Editing: S. Panis and  
761      R. Ramsey; Supervision: R. Ramsey.

762

**Conflicts of Interest**

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

765

**Prior versions**

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

768

**Supplemental Material**

769

**Disclosures****770 Data, materials, and online resources**

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

**774 Ethical approval**

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

**References**

- 777
- 778 Abney, D. H., Fausey, C. M., Suarez-Rivera, C., & Tamis-LeMonda, C. S. (2025).  
779 Advancing a temporal science of behavior. *Trends in Cognitive Sciences*.  
780 <https://doi.org/10.1016/j.tics.2025.05.010>
- 781 Allison, P. D. (1982). Discrete-Time Methods for the Analysis of Event Histories.  
782 *Sociological Methodology*, 13, 61. <https://doi.org/10.2307/270718>
- 783 Allison, P. D. (2010). *Survival analysis using SAS: A practical guide* (2. ed). Cary, NC:  
784 SAS Press.
- 785 Aust, F. (2019). *Citr: 'RStudio' add-in to insert markdown citations*. Retrieved from  
786 <https://github.com/crsh/citr>
- 787 Aust, F., & Barth, M. (2024). *papaja: Prepare reproducible APA journal articles with R*  
788 *Markdown*. <https://doi.org/10.32614/CRAN.package.papaja>
- 789 Baker, D. H., Vilidaite, G., Lygo, F. A., Smith, A. K., Flack, T. R., Gouws, A. D., &  
790 Andrews, T. J. (2021). Power contours: Optimising sample size and precision in  
791 experimental psychology and human neuroscience. *Psychological Methods*, 26(3),  
792 295–314. <https://doi.org/10.1037/met0000337>
- 793 Barack, D. L., & Krakauer, J. W. (2021). Two views on the cognitive brain. *Nature*  
794 *Reviews Neuroscience*, 22(6), 359–371. <https://doi.org/10.1038/s41583-021-00448-6>
- 795 Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for  
796 confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*,  
797 68(3), 10.1016/j.jml.2012.11.001. <https://doi.org/10.1016/j.jml.2012.11.001>
- 798 Barth, M. (2023). *tinylabes: Lightweight variable labels*. Retrieved from  
799 <https://cran.r-project.org/package=tinylabes>
- 800 Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects  
801 models using lme4. *Journal of Statistical Software*, 67(1), 1–48.  
802 <https://doi.org/10.18637/jss.v067.i01>
- 803 Bates, D., Maechler, M., & Jagan, M. (2024). *Matrix: Sparse and dense matrix classes and*

- 804       methods. Retrieved from <https://Matrix.R-forge.R-project.org>
- 805   Bengtsson, H. (2021). A unifying framework for parallel and distributed processing in r  
806       using futures. *The R Journal*, 13(2), 208–227. <https://doi.org/10.32614/RJ-2021-048>
- 807   Berger, A., & Kiefer, M. (2021). Comparison of Different Response Time Outlier Exclusion  
808       Methods: A Simulation Study. *Frontiers in Psychology*, 12, 675558.  
809       <https://doi.org/10.3389/fpsyg.2021.675558>
- 810   Blossfeld, H.-P., & Rohwer, G. (2002). *Techniques of event history modeling: New  
811       approaches to causal analysis*, 2nd ed (pp. x, 310). Mahwah, NJ, US: Lawrence  
812       Erlbaum Associates Publishers.
- 813   Bolger, N., Zee, K. S., Rossignac-Milon, M., & Hassin, R. R. (2019). Causal processes in  
814       psychology are heterogeneous. *Journal of Experimental Psychology: General*, 148(4),  
815       601–618. <https://doi.org/10.1037/xge0000558>
- 816   Box-Steffensmeier, J. M. (2004). Event history modeling: A guide for social scientists.  
817       Cambridge: University Press.
- 818   Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan.  
819       *Journal of Statistical Software*, 80(1), 1–28. <https://doi.org/10.18637/jss.v080.i01>
- 820   Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms.  
821       *The R Journal*, 10(1), 395–411. <https://doi.org/10.32614/RJ-2018-017>
- 822   Bürkner, P.-C. (2021). Bayesian item response modeling in R with brms and Stan. *Journal  
823       of Statistical Software*, 100(5), 1–54. <https://doi.org/10.18637/jss.v100.i05>
- 824   DeBruine, L. M., & Barr, D. J. (2021). Understanding Mixed-Effects Models Through  
825       Data Simulation. *Advances in Methods and Practices in Psychological Science*, 4(1),  
826       2515245920965119. <https://doi.org/10.1177/2515245920965119>
- 827   Eddelbuettel, D., & Balamuta, J. J. (2018). Extending R with C++: A Brief Introduction  
828       to Rcpp. *The American Statistician*, 72(1), 28–36.  
829       <https://doi.org/10.1080/00031305.2017.1375990>
- 830   Eddelbuettel, D., & François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal*

- 831        *of Statistical Software*, 40(8), 1–18. <https://doi.org/10.18637/jss.v040.i08>
- 832    Gabry, J., Češnovar, R., Johnson, A., & Brodner, S. (2024). *Cmdstanr: R interface to*  
833    '*CmdStan*'. Retrieved from <https://github.com/stan-dev/cmdstanr>
- 834    Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2019). Visualization  
835    in bayesian workflow. *J. R. Stat. Soc. A*, 182, 389–402.  
836    <https://doi.org/10.1111/rssc.12378>
- 837    Gelman, A., Hill, J., & Vehtari, A. (2020). Regression and Other Stories.  
838    <https://www.cambridge.org/highereducation/books/regression-and-other-stories/DD20DD6C9057118581076E54E40C372C>; Cambridge University Press.  
839    <https://doi.org/10.1017/9781139161879>
- 840    Gelman, A., Vehtari, A., Simpson, D., Margossian, C. C., Carpenter, B., Yao, Y., ...  
841    Modrák, M. (2020). *Bayesian Workflow*. arXiv.  
842    <https://doi.org/10.48550/arXiv.2011.01808>
- 843    Girard, J. (2024). *Standist: What the package does (one line, title case)*. Retrieved from  
844    <https://github.com/jmgirard/standist>
- 845    Golemund, G., & Wickham, H. (2011). Dates and times made easy with lubridate.  
846    *Journal of Statistical Software*, 40(3), 1–25. Retrieved from  
847    <https://www.jstatsoft.org/v40/i03/>
- 848    Heiss, A. (2021, November 10). A Guide to Correctly Calculating Posterior Predictions  
849    and Average Marginal Effects with Multilevel Bayesian Models.  
850    <https://doi.org/10.59350/wbn93-edb02>
- 851    Holden, J. G., Van Orden, G. C., & Turvey, M. T. (2009). Dispersion of response times  
852    reveals cognitive dynamics. *Psychological Review*, 116(2), 318–342.  
853    <https://doi.org/10.1037/a0014849>
- 854    Hosmer, D. W., Lemeshow, S., & May, S. (2011). *Applied Survival Analysis: Regression*  
855    *Modeling of Time to Event Data* (2nd ed). Hoboken: John Wiley & Sons.
- 856    Kantowitz, B. H., & Pachella, R. G. (2021). The Interpretation of Reaction Time in

- 858 Information-Processing Research 1. *Human Information Processing*, 41–82.
- 859 <https://doi.org/10.4324/9781003176688-2>
- 860 Kay, M. (2024). *tidybayes: Tidy data and geoms for Bayesian models*.
- 861 <https://doi.org/10.5281/zenodo.1308151>
- 862 Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing,
- 863 estimation, meta-analysis, and power analysis from a Bayesian perspective.
- 864 *Psychonomic Bulletin & Review*, 25(1), 178–206.
- 865 <https://doi.org/10.3758/s13423-016-1221-4>
- 866 Kurz, A. S. (2023a). *Applied longitudinal data analysis in brms and the tidyverse* (version
- 867 0.0.3). Retrieved from <https://bookdown.org/content/4253/>
- 868 Kurz, A. S. (2023b). *Statistical rethinking with brms, ggplot2, and the tidyverse: Second*
- 869 *edition* (version 0.4.0). Retrieved from <https://bookdown.org/content/4857/>
- 870 Lakens, D. (2022). Sample Size Justification. *Collabra: Psychology*, 8(1), 33267.
- 871 <https://doi.org/10.1525/collabra.33267>
- 872 Landes, J., Engelhardt, S. C., & Pelletier, F. (2020). An introduction to event history
- 873 analyses for ecologists. *Ecosphere*, 11(10), e03238. <https://doi.org/10.1002/ecs2.3238>
- 874 Luce, R. D. (1991). *Response times: Their role in inferring elementary mental organization*
- 875 (1. issued as paperback). Oxford: Univ. Press.
- 876 McElreath, R. (2020). *Statistical Rethinking: A Bayesian Course with Examples in R and*
- 877 *STAN* (2nd ed.). New York: Chapman and Hall/CRC.
- 878 <https://doi.org/10.1201/9780429029608>
- 879 Müller, K., & Wickham, H. (2023). *Tibble: Simple data frames*. Retrieved from
- 880 <https://CRAN.R-project.org/package=tibble>
- 881 Nelder, J. A. (1999). From Statistics to Statistical Science. *Journal of the Royal Statistical*
- 882 *Society. Series D (The Statistician)*, 48(2), 257–269. Retrieved from
- 883 <https://www.jstor.org/stable/2681191>
- 884 Neuwirth, E. (2022). *RColorBrewer: ColorBrewer palettes*. Retrieved from

- 885        <https://CRAN.R-project.org/package=RColorBrewer>
- 886        Panis, S. (2020). How can we learn what attention is? Response gating via multiple direct  
887        routes kept in check by inhibitory control processes. *Open Psychology*, 2(1), 238–279.  
888        <https://doi.org/10.1515/psych-2020-0107>
- 889        Panis, S., Moran, R., Wolkersdorfer, M. P., & Schmidt, T. (2020). Studying the dynamics  
890        of visual search behavior using RT hazard and micro-level speed–accuracy tradeoff  
891        functions: A role for recurrent object recognition and cognitive control processes.  
892        *Attention, Perception, & Psychophysics*, 82(2), 689–714.  
893        <https://doi.org/10.3758/s13414-019-01897-z>
- 894        Panis, S., Schmidt, F., Wolkersdorfer, M. P., & Schmidt, T. (2020). Analyzing Response  
895        Times and Other Types of Time-to-Event Data Using Event History Analysis: A Tool  
896        for Mental Chronometry and Cognitive Psychophysiology. *I-Perception*, 11(6),  
897        2041669520978673. <https://doi.org/10.1177/2041669520978673>
- 898        Panis, S., & Schmidt, T. (2016). What Is Shaping RT and Accuracy Distributions? Active  
899        and Selective Response Inhibition Causes the Negative Compatibility Effect. *Journal of*  
900        *Cognitive Neuroscience*, 28(11), 1651–1671. [https://doi.org/10.1162/jocn\\_a\\_00998](https://doi.org/10.1162/jocn_a_00998)
- 901        Panis, S., & Schmidt, T. (2022). When does “inhibition of return” occur in spatial cueing  
902        tasks? Temporally disentangling multiple cue-triggered effects using response history  
903        and conditional accuracy analyses. *Open Psychology*, 4(1), 84–114.  
904        <https://doi.org/10.1515/psych-2022-0005>
- 905        Panis, S., Torfs, K., Gillebert, C. R., Wageman, J., & Humphreys, G. W. (2017).  
906        Neuropsychological evidence for the temporal dynamics of category-specific naming.  
907        *Visual Cognition*, 25(1-3), 79–99. <https://doi.org/10.1080/13506285.2017.1330790>
- 908        Panis, S., & Wageman, J. (2009). Time-course contingencies in perceptual organization  
909        and identification of fragmented object outlines. *Journal of Experimental Psychology:*  
910        *Human Perception and Performance*, 35(3), 661–687.  
911        <https://doi.org/10.1037/a0013547>

- 912 Pargent, F., Koch, T. K., Kleine, A.-K., Lermer, E., & Gaube, S. (2024). A Tutorial on  
913 Tailored Simulation-Based Sample-Size Planning for Experimental Designs With  
914 Generalized Linear Mixed Models. *Advances in Methods and Practices in Psychological  
915 Science*, 7(4), 25152459241287132. <https://doi.org/10.1177/25152459241287132>
- 916 Pedersen, T. L. (2024). *Patchwork: The composer of plots*. Retrieved from  
917 <https://patchwork.data-imaginist.com>
- 918 Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in s and s-PLUS*. New York:  
919 Springer. <https://doi.org/10.1007/b98882>
- 920 R Core Team. (2024). *R: A language and environment for statistical computing*. Vienna,  
921 Austria: R Foundation for Statistical Computing. Retrieved from  
922 <https://www.R-project.org/>
- 923 Singer, J. D., & Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling  
924 Change and Event Occurrence*. Oxford, New York: Oxford University Press.
- 925 Smith, P. L., & Little, D. R. (2018). Small is beautiful: In defense of the small-N design.  
926 *Psychonomic Bulletin & Review*, 25(6), 2083–2101.  
927 <https://doi.org/10.3758/s13423-018-1451-8>
- 928 Stan Development Team. (2020). *StanHeaders: Headers for the R interface to Stan*.  
929 Retrieved from <https://mc-stan.org/>
- 930 Stan Development Team. (2024). *RStan: The R interface to Stan*. Retrieved from  
931 <https://mc-stan.org/>
- 932 Teachman, J. D. (1983). Analyzing social processes: Life tables and proportional hazards  
933 models. *Social Science Research*, 12(3), 263–301.  
934 [https://doi.org/10.1016/0049-089X\(83\)90015-7](https://doi.org/10.1016/0049-089X(83)90015-7)
- 935 Tong, C. (2019). Statistical Inference Enables Bad Science; Statistical Thinking Enables  
936 Good Science. *The American Statistician*, 73(sup1), 246–261.  
937 <https://doi.org/10.1080/00031305.2018.1518264>
- 938 Townsend, J. T. (1990). Truth and consequences of ordinal differences in statistical

- 939 distributions: Toward a theory of hierarchical inference. *Psychological Bulletin*, 108(3),  
940 551–567. <https://doi.org/10.1037/0033-2909.108.3.551>
- 941 Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics.  
942 *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)
- 943 Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New  
944 York. Retrieved from <https://ggplot2.tidyverse.org>
- 945 Wickham, H. (2023a). *Forcats: Tools for working with categorical variables (factors)*.  
946 Retrieved from <https://forcats.tidyverse.org/>
- 947 Wickham, H. (2023b). *Stringr: Simple, consistent wrappers for common string operations*.  
948 Retrieved from <https://stringr.tidyverse.org>
- 949 Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., ...  
950 Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43),  
951 1686. <https://doi.org/10.21105/joss.01686>
- 952 Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *Dplyr: A  
953 grammar of data manipulation*. Retrieved from <https://dplyr.tidyverse.org>
- 954 Wickham, H., & Henry, L. (2023). *Purrr: Functional programming tools*. Retrieved from  
955 [https://purrr.tidyverse.org/](https://purrr.tidyverse.org)
- 956 Wickham, H., Hester, J., & Bryan, J. (2024). *Readr: Read rectangular text data*. Retrieved  
957 from <https://readr.tidyverse.org>
- 958 Wickham, H., Vaughan, D., & Girlich, M. (2024). *Tidyr: Tidy messy data*. Retrieved from  
959 <https://tidyr.tidyverse.org>
- 960 Winter, B. (2019). *Statistics for Linguists: An Introduction Using R*. New York:  
961 Routledge. <https://doi.org/10.4324/9781315165547>
- 962 Wolkersdorfer, M. P., Panis, S., & Schmidt, T. (2020). Temporal dynamics of sequential  
963 motor activation in a dual-prime paradigm: Insights from conditional accuracy and  
964 hazard functions. *Attention, Perception, & Psychophysics*, 82(5), 2581–2602.  
965 <https://doi.org/10.3758/s13414-020-02010-5>