

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

3 Sven Panis¹ & Richard Ramsey¹

4 ¹ ETH Zürich

5 Author Note

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

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

11

Abstract

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

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

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

34

1. Introduction

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

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

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

60 accuracies (0, 1, 0.51, 0.48) are never even close to the mean accuracies.

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

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

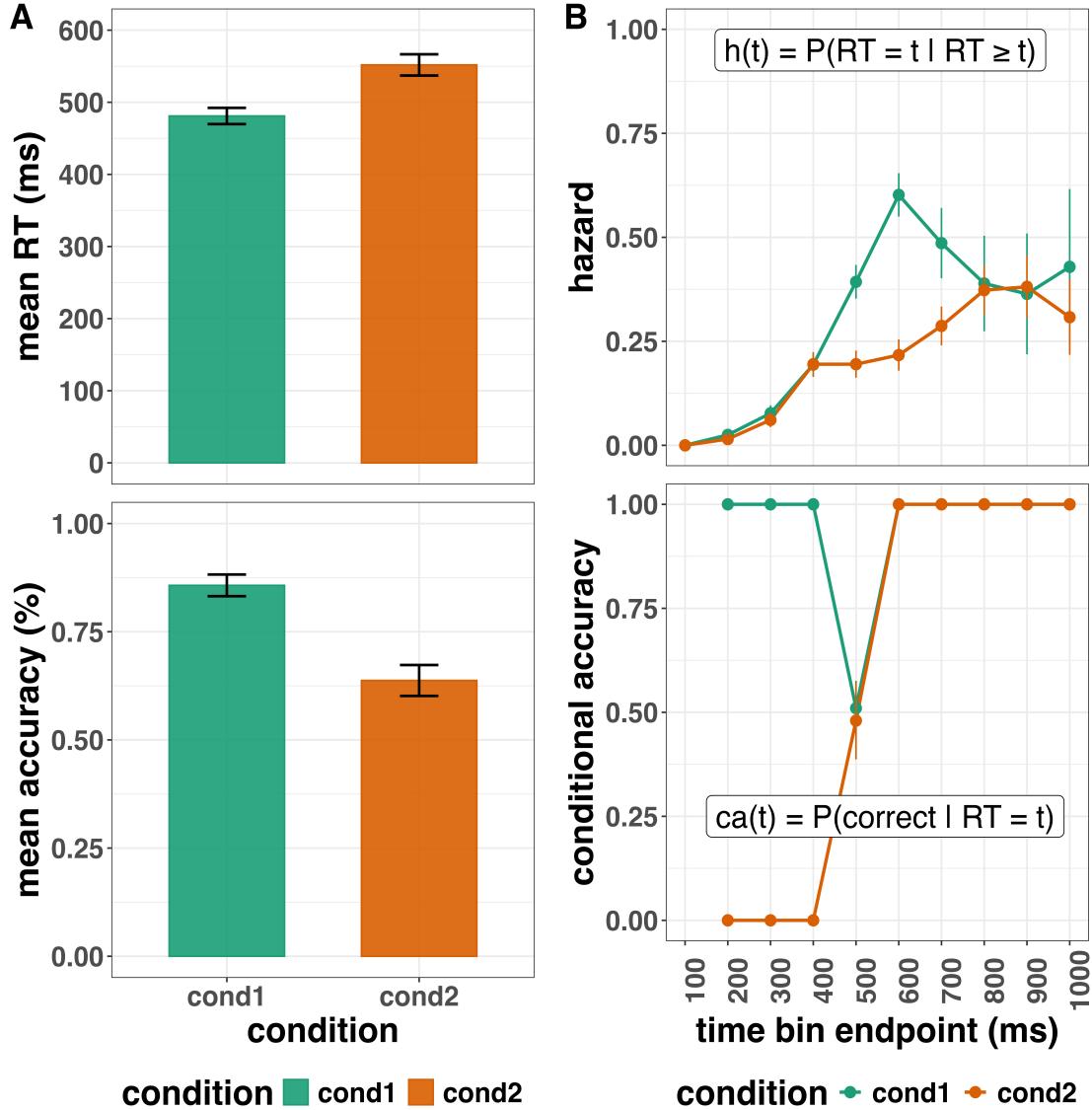


Figure 1. Mean performance versus distributional analyses. (A) The mean RT (top) and overall accuracy (bottom) for two conditions are plotted. (B) The discrete-time hazard functions (top) and conditional accuracy functions (bottom) are plotted for the same data. The first second after target stimulus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

75 **1.2 Aims and structure of the paper**

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

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

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

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

101 discrimination or categorization, bistable perception) are treated in EHA.

102 We then provide seven different tutorials, which are written in the R programming
103 language and publicly available on our Github page (<https://github.com/sven-panis/>)
104 Tutorial_Event_History_Analysis), along with all of the other code and material
105 associated with the project. The tutorials provide hands-on, concrete examples of key parts
106 of the analytical process, so that others can apply EHA to their own time-to-event data.
107 Each tutorial is provided as an RMarkdown file, so that others can download and adapt
108 the code to fit their own purposes. Additionally, each tutorial is made available as a .html
109 file, so that it can be viewed by any web browser, and thus available to those that do not
110 use R. Finally, the manuscript itself is written in R using the papaja package (Aust &
111 Barth, 2024a), which makes it computationally reproducible, in terms of the underlying
112 data and figures.

113 .

114 2. A brief introduction to event history analysis

115 To apply EHA, one must be able to:

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

121 2.1 Single, repeatable, and recurrent events

122 While people can die only once, in experimental RT tasks the events of interest -
123 button presses, saccade onsets, etc. - are typically repeatable. For example, in a one-button

¹²⁴ detection task the participant is presented in each trial with a faint target stimulus that
¹²⁵ (s)he has to detect by pressing a button within a certain time (e.g., 1 second).

¹²⁶ Detection: single event, repeatable: hazard function + clustering Discrimination:
¹²⁷ Two events, repeatable: competing risks Bistable perception task: Two events, repeatable,
¹²⁸ and recurrent

¹²⁹ **2.2 Right censoring vs data trimming**

¹³⁰ **2.3 Discrete vs continuous time units**

¹³¹ All measurements of duration are discrete in nature. However, when the temporal
¹³² resolution is high relative to the duration of the observation window, - interest in the shape
¹³³ of the hazard function <-> Cox - non-smooth functions <-> parametric - definition hazard
¹³⁴ prob and logistic regression

¹³⁵ **2.4 Discrete-time Hazard functions vs cumulative distribution functions**

¹³⁶ def hazard

¹³⁷ def survival

¹³⁸ def cumulative distribution

¹³⁹ def prob mass function

¹⁴⁰ def conditional accuracz function

¹⁴¹ **2.5 Number of time bins, repeated measures, and samples**

¹⁴² **2.6 Bayesian vs. frequentist approach**

144 We recommend several excellent textbooks for a comprehensive background context

145 to EHA (Allison, 2010; Singer & Willett, 2003) and for a more general introduction to

146 understanding regression equations (Gelman, Hill, & Vehtari, 2020; Winter, 2019). Our

147 focus here is not on providing a detailed account of the underlying regression equations,

148 since this topic has been comprehensively covered many times before. Instead, we want to

149 provide an intuition regarding how EHA works in general, as well as in the context of

150 experimental psychology. As such, we only supply regression equations in section D of the

151 Supplemental Material.

152 **2.1 Basic features of event history analysis**

153 In EHA, the definition of hazard and the type of models employed depend on

154 whether one is using continuous or discrete time units.

155 Since our focus here is on hazard models that use discrete time units, we describe

156 that approach. After dividing time in discrete, contiguous time bins indexed by t (e.g., $t =$

157 1:10 time bins), let RT be a discrete random variable denoting the rank of the time bin in

158 which a particular person's response occurs in a particular experimental condition. For

159 example, the first response might occur at 546 ms and it would be in time bin 6 (any RTs

160 from 501 ms to 600 ms).

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

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

163 functions are reported in section A of the Supplemental Material.

164 The discrete-time hazard function gives you, for each time bin, the probability that

165 the event occurs (sometime) in bin t , given that the event does not occur in previous bins.

166 In other words, it reflects the instantaneous risk that the event occurs in the current bin,

167 given that it has not yet occurred in the past, i.e., in one of the prior bins. In contrast, the

168 discrete-time survivor function cumulates the bin-by-bin risks of event *nonoccurrence* to

₁₆₉ obtain the survival probability, the probability that the event occurs after bin t. In other
₁₇₀ words, the survivor function gives you for each time bin the likelihood that the event
₁₇₁ occurs in the future, i.e., in one of the subsequent time bins.

hazard and survivor functions

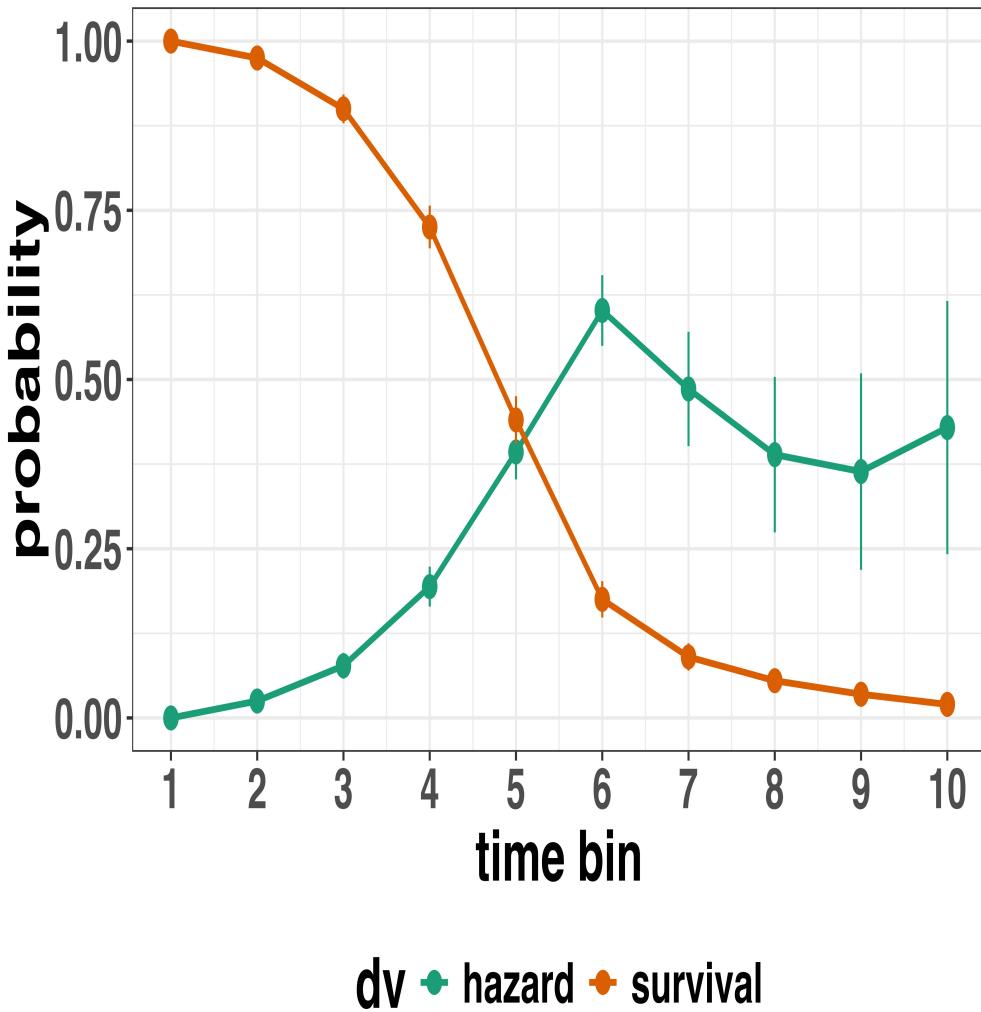


Figure 2. Discrete-time hazard and survivor functions. Discrete time-to-event data were simulated for 200 trials of 1 experimental condition. Error bars represent ± 1 standard error of the respective proportion. While the hazard function is the vehicle for inferring the time course of cognitive processes, 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.

2.3 Event history analysis in the context of experimental psychology

To make EHA more relevant to researchers studying cognitive psychology and

cognitive neuroscience, in this section we provide a relevant worked example and consider

implications that are relevant to that domain of research.

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

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

(Figure 3, upper panel), and that early emitted responses are always correct in the

congruent condition and always incorrect in the incongruent condition (Figure 3, lower

panel). These results show that for short waiting times ($<$ bin 6), responses always follow

the prime (and not the target, as instructed). During time bin 6 the target-triggered

198 response channel is activated and causes response competition – $ca(6) = .5$ – and a lower
199 hazard probability in the incongruent condition. For waiting times of 600 ms or more, the
200 hazard of response occurrence is lower in incongruent compared to congruent trials, and all
201 responses emitted in these late bins are correct.

202 This joint pattern of results is interesting because it can provide meaningfully
203 different conclusions about psychological processes compared to conventional analyses, such
204 as computing mean-average RT and accuracy across trials. Mean-average RT would only
205 represent the overall ability of cognition to overcome interference, on average, across trials.
206 For instance, if mean-average RT was higher in incongruent than congruent trials, one may
207 conclude that cognitive mechanisms that support interference control are working as
208 expected across trials, and are indexed by each recorded response. But such a conclusion is
209 not supported when the effects are explored over a timeline. Instead, the psychological
210 conclusion is much more nuanced and suggests that multiple states start, stop and possibly
211 interact over a particular temporal window.

212 Unlocking the temporal states of cognitive processes can be revealing for theory
213 development and the understanding of basic psychological processes. Possibly more
214 importantly, however, is that it simultaneously opens the door to address many new and
215 previously unanswered questions. Do all participants show similar temporal states or are
216 there individual differences? Do such individual differences extend to those individuals that
217 have been diagnosed with some form of psychopathology? How do temporal states relate to
218 brain-based mechanisms that might be studied using other methods from cognitive
219 neuroscience? And how much of theory in cognitive psychology would be in need of
220 revision if mean-average comparisons were supplemented with a temporal states approach?

221 **2.3.2 Implications for designing experiments.** Performing EHA in
222 experimental psychology has implications for how experiments are designed. Indeed, if
223 trials are categorised as a function of when responses occur, then each time bin will only
224 include a subset of the total number of trials. For example, let's consider an experiment

225 where each participant performs 2 conditions and there are 100 trial repetitions per
226 condition. Those 100 trials must be distributed in some manner across the chosen number
227 of bins.

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

236 In contrast to the large- N design that typically average across many participants
237 without being able to scrutinize individual data patterns, small- N designs retain crucial
238 information about the data patterns of individual observers. This can be advantageous
239 whenever participants differ systematically in their strategies or in the time courses of their
240 effects, so that averaging them would lead to misleading data patterns. Note that because
241 statistical power derives both from the number of participants and from the number of
242 repeated measures per participant and condition, small- N designs can still achieve what
243 are generally considered acceptable levels of statistical power, if they have a sufficient
244 amount of data overall (Baker et al., 2021; Smith & Little, 2018).

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

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

249 3.1 Data science workflow and descriptive statistics

250 We perform data wrangling following tidyverse principles and a functional
251 programming approach (Wickham, Çetinkaya-Rundel, & Grolemund, 2023). In short,
252 functional programming means that you avoid writing your own loops and instead use
253 functions that have been built and tested by others. In addition, we also supply a set of
254 custom-built functions, which make the process of data wrangling in the context of data
255 preparation and descriptive statistics a lot quicker and more efficient.

256 3.2 Inferential statistical approach

257 Our lab adopts an estimation approach to multilevel regression (Kruschke & Liddell,
258 2018; Winter, 2019), which is heavily influenced by the Bayesian framework as suggested
259 by Richard McElreath (Kurz, 2023b; McElreath, 2020). We also use a “keep it maximal”
260 approach to specifying varying (or random) effects (Barr, Levy, Scheepers, & Tily, 2013).
261 This means that wherever possible we include varying intercepts and slopes per participant.
262 To make inferences, we use two main approaches. We compare models of different
263 complexity, using information criteria (e.g., WAIC) and cross-validation (e.g., LOO), to
264 evaluate out-of-sample predictive accuracy (McElreath, 2020). We also take the most
265 complex model and evaluate key parameters of interest using point and interval estimates.

266 3.3 Implementation

267 We used R (Version 4.4.0; R Core Team, 2024)¹ for all reported analyses. The
268 content of the tutorials, in terms of EHA and multilevel regression modelling, is mainly

¹ We, furthermore, used the R-packages *bayesplot* (Version 1.11.1; Gabry, Simpson, Vehtari, Betancourt, & Gelman, 2019), *brms* (Version 2.22.0; Bürkner, 2017, 2018, 2021), *citr* (Version 0.3.2; Aust, 2019), *cmdstanr* (Version 0.8.1.9000; Gabry, Češnovar, Johnson, & Brander, 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;

269 based on Allison (2010), Singer and Willett (2003), McElreath (2020), Heiss (2021), Kurz
270 (2023a), and Kurz (2023b).

271

4. Tutorials

272 Tutorials 1a and 1b show how to calculate and plot the descriptive statistics of
273 EHA/SAT when there are one or two independent variables, respectively. Tutorials 2a and
274 2b illustrate how to use Bayesian multilevel modeling to fit hazard and conditional
275 accuracy models, respectively. Tutorials 3a and 3b show how to implement, respectively,
276 multilevel models for hazard and conditional accuracy in the frequentist framework.
277 Additionally, to further simplify the process for other users, the first two tutorials rely on a
278 set of our own custom functions that make sub-processes easier to automate, such as data
279 wrangling and plotting functions (see section B in the Supplemental Material for a list of
280 the custom functions).

281 Our list of tutorials is as follows:

- 282 • 1a. Wrangle raw data and calculate descriptive stats for one independent variable
283 • 1b. Wrangle raw data and calculate descriptive stats for two independent variables
284 • 2a. Bayesian multilevel modeling for $h(t)$
285 • 2b. Bayesian multilevel modeling for $ca(t)$

Grolemund & Wickham, 2011), *Matrix* (Version 1.7.1; Bates, Maechler, & Jagan, 2024), *nlme* (Version 3.1.166; Pinheiro & Bates, 2000), *papaja* (Version 0.1.3; Aust & Barth, 2024b), *patchwork* (Version 1.3.0; Pedersen, 2024), *purrr* (Version 1.0.2; Wickham & Henry, 2023), *RColorBrewer* (Version 1.1.3; Neuwirth, 2022), *Rcpp* (Eddelbuettel & Balamuta, 2018; Version 1.0.13.1; Eddelbuettel & François, 2011), *readr* (Version 2.1.5; Wickham, Hester, & Bryan, 2024), *RJ-2021-048* (Bengtsson, 2021), *rstan* (Version 2.32.6; Stan Development Team, 2024), *standist* (Version 0.0.0.9000; Girard, 2024), *StanHeaders* (Version 2.32.10; Stan Development Team, 2020), *stringr* (Version 1.5.1; Wickham, 2023b), *tibble* (Version 3.2.1; Müller & Wickham, 2023), *tidybayes* (Version 3.0.7; Kay, 2024), *tidyverse* (Version 2.0.0; Wickham et al., 2019) and *tinylabels* (Version 0.2.4; Barth, 2023).

- 286 • 3a. Frequentist multilevel modeling for $h(t)$
287 • 3b. Frequentist multilevel modeling for $ca(t)$
288 • 4. Simulation and power analysis for planning experiments

289 **4.1 Tutorial 1a: Calculating descriptive statistics using a life table**

290 **4.1.1 Data wrangling aims.** Our data wrangling procedures serve two related
291 purposes. First, we want to summarise and visualise descriptive statistics that relate to our
292 main research questions about the time course of psychological processes, using a life table.
293 A life table includes for each time bin, the risk set (i.e., the number of trials that are
294 event-free at the start of the bin), the number of observed events, and the estimates of
295 $h(t)$, $S(t)$, $P(t)$, possibly $ca(t)$, and their estimated standard errors (se).

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

302 `## Warning in attr(x, "align"): 'xfun::attr()' is deprecated.`
303 `## Use 'xfun::attr2()' instead.`
304 `## 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.

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

```

314 ## Warning in attr(x, "align"): 'xfun::attr()' is deprecated.
315 ## Use 'xfun::attr2()' instead.
316 ## 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.

317 **4.1.2 A real data wrangling example.** To illustrate how to quickly set up life
 318 tables for calculating the descriptive statistics (functions of discrete time), we use a
 319 published data set on masked response priming from Panis and Schmidt (2016). In their
 320 first experiment, Panis and Schmidt (2016) presented a double arrow for 94 ms that
 321 pointed left or right as the target stimulus with an onset at time point zero in each trial.

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

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

332 The required column names are as follows:

- 333 • “pid”, indicating unique participant IDs;
- 334 • “trial”, indicating each unique trial per participant;
- 335 • “condition”, a factor indicating the levels of the independent variable (1, 2, ...) and
 the corresponding labels;
- 337 • “rt”, indicating the response times in ms;
- 338 • “acc”, indicating the accuracies (1/0).

339 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")))
```

340 Next, we can set up the life tables and plots of the discrete-time functions $h(t)$, $S(t)$,
 341 $ca(t)$, and $P(t)$ – see section A of the Supplemental Material for their definitions. To do so
 342 using a functional programming approach, one has to nest the data within participants
 343 using the `group_nest()` function, and supply a user-defined censoring time and bin width
 344 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))
```

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

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

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

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

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

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

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

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

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

362 warning messages will likely be generated, like these:

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

364 (`geom_line()`).

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

366 (`geom_point()`).

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

368 (`geom_segment()`).

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

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

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

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

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

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

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

376 different behavior.

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

378 participant 6.

```
379 ## Warning in attr(x, "align"): 'xfun::attr()' is deprecated.  
380 ## Use 'xfun::attr2()' instead.  
381 ## See help("Deprecated")
```

Table 3

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

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

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

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

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

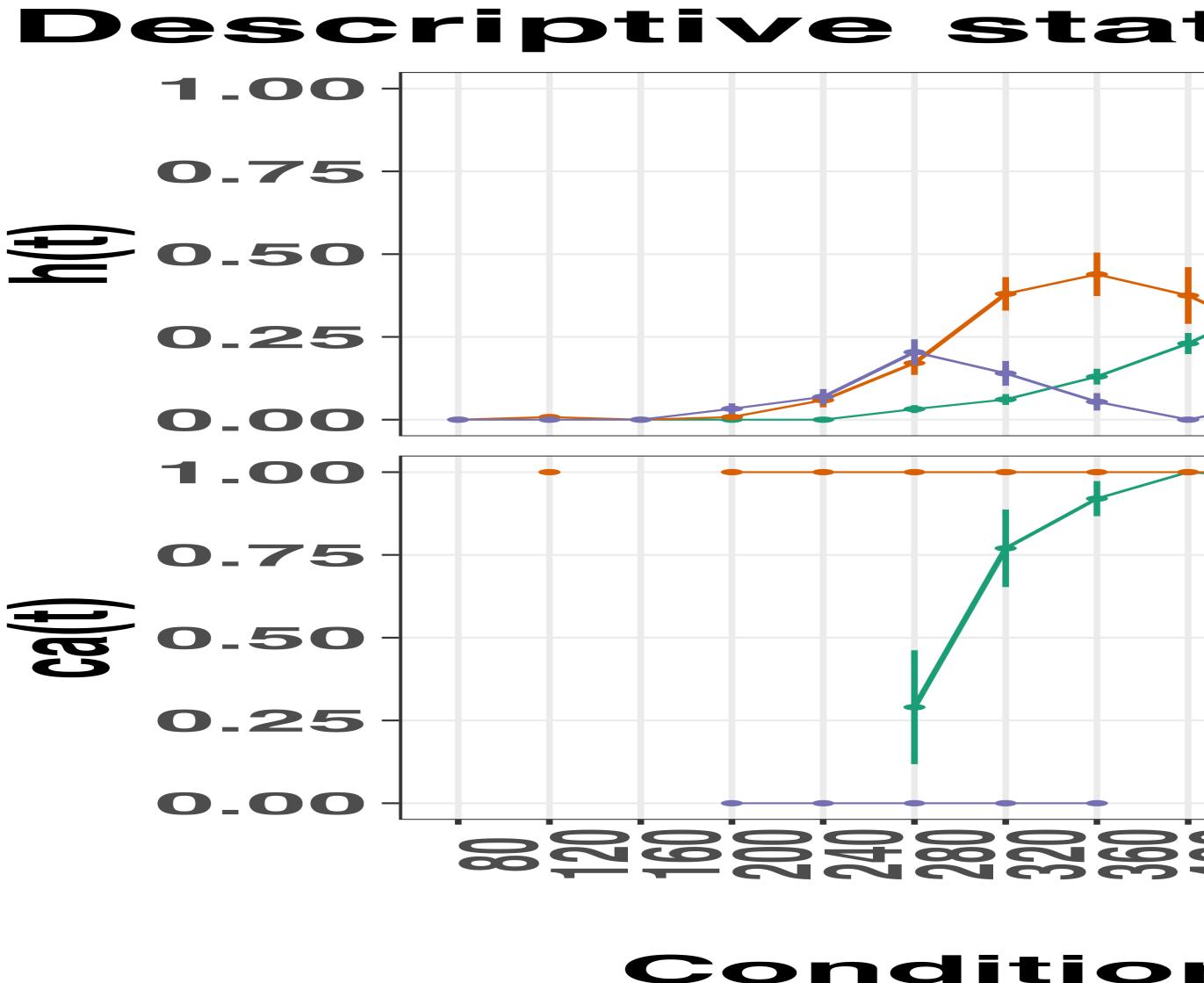


Figure 3. Estimated discrete-time hazard (h), survivor (S), conditional accuracy (ca) and probability mass (P) functions for participant 6. Vertical dotted lines indicate the estimated median RTs. Error bars represent ± 1 standard error of the respective proportion.

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

389 0.17, and 0.20 for the blank, congruent, and incongruent prime conditions, respectively. In

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

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

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

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

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

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

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

397 respectively.

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

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

400 0.36, 0.25, and 0.06 for the blank, congruent, and incongruent prime conditions,

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

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

403 conditions, respectively.

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

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

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

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

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

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

410 group-average distribution per condition (see Tutorial_1a.Rmd).

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

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

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

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

415 At this point, we have calculated, summarised and plotted descriptive statistics for
416 the key variables in EHA/SAT. As we will show in later Tutorials, statistical models for
417 $h(t)$ and $ca(t)$ can be implemented as generalized linear mixed regression models predicting
418 event occurrence (1/0) and conditional accuracy (1/0) in each bin of a selected time
419 window for analysis. But first we consider calculating the descriptive statistics for two
420 independent variables.

421 **4.2 Tutorial 1b: Generalising to a more complex design**

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

428 To this end, Tutorial 1b illustrates how to calculate and plot the descriptive statistics
429 for the full data set of Experiment 1 of Panis and Schmidt (2016), which includes two
430 independent variables: mask type and prime type. As we use the same functional
431 programming approach as in Tutorial 1a, we simply present the sample-based functions for
432 each participant as part of Tutorial_1b.Rmd for those that are interested.

433 **4.3 Tutorial 2a: Fitting Bayesian hazard models to discrete time-to-event data**

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

438 Willett, 2003).

439 **4.3.1 Hazard model considerations.** There are several analytic decisions one
440 has to make when fitting a discrete-time hazard model. First, one has to select an analysis
441 time window, i.e., a contiguous set of bins for which there is enough data for each
442 participant. Second, given that the dependent variable (event occurrence) is binary, one
443 has to select a link function (see section C in the Supplemental Material). The cloglog link
444 is preferred over the logit link when events can occur in principle at any time point within
445 a bin, which is the case for RT data (Singer & Willett, 2003). Third, one has to choose
446 whether to treat TIME (i.e., the time bin index t) as a categorical or continuous predictor.
447 And when you treat a variable as a categorical predictor, you can choose between reference
448 coding and index coding. With reference coding, one defines the variable as a factor and
449 selects one of the k categories as the reference level. `Brm()` will then construct $k-1$
450 indicator variables (see model M1d in Tutorial_2a.Rmd for an example). With index
451 coding, one constructs an index variable that contains integers that correspond to different
452 categories (see models M0i and M1i below). As explained by McElreath (2020), the
453 advantage of index coding is that the same prior can be assigned to each level of the index
454 variable, so that each category has the same prior uncertainty.

455 In the case of a large- N design without repeated measurements, the parameters of a
456 discrete-time hazard model can be estimated using standard logistic regression software
457 after expanding the typical person-trial data set into a person-trial-bin data set (Allison,
458 2010). When there is clustering in the data, as in the case of a small- N design with
459 repeated measurements, the parameters of a discrete-time hazard model can be estimated
460 using population-averaged methods (e.g., Generalized Estimating Equations), and Bayesian
461 or frequentist generalized linear mixed models (Allison, 2010).

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

465 additivity assumption (predictors do not interact), and the proportionality assumption
 466 (predictors do not interact with TIME).

467 In tutorial_2a.Rmd we fit several Bayesian multilevel models (i.e., generalized linear
 468 mixed models) that differ in complexity to the person-trial-bin oriented data set that we
 469 created in Tutorial 1a. We decided to select the analysis time window (200,600] and the
 470 cloglog link. Below, we shortly discuss two of these models. The person-trial-bin data set is
 471 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)))
```

472 **4.3.2 Prior distributions.** To get the posterior distribution of each model
 473 parameter given the data, we need to specify prior distributions for the model parameters
 474 which reflect our prior beliefs. In Tutorial_2a.Rmd we perform a few prior predictive
 475 checks to make sure our selected prior distributions reflect our prior beliefs (Gelman,
 476 Vehtari, et al., 2020).

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

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

484 **4.3.3 Model M0i: A null model with index coding.** When you do not want to
 485 make assumptions about the shape of the hazard function, or its shape is not smooth but
 486 irregular, then you can use a general specification of TIME, i.e., fit one grand intercept per
 487 time bin. In this first model, we use a general specification of TIME using index coding,
 488 and do not include experimental predictors. We call this model “M0i”.

489 Before we fit model M0i, we select the necessary columns from the data, and specify
 490 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")
```

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

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

497 **4.3.4 Model M1i: Adding the effects of prime-target congruency.** Previous
 498 research has shown that psychological effects typically change over time (Panis, 2020;
 499 Panis, Moran, et al., 2020; Panis & Schmidt, 2022; Panis et al., 2017; Panis & Wagemans,
 500 2009). In the next model, therefore, we use index coding for both TIME (variable
 501 “timebin”) and the categorical predictor prime-target-congruency (variable “prime”), so
 502 that we get 30 grand intercepts, one for each combination of timebin level and prime level.
 503 Here is the model formula of this model that we call “M1i”.

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

504 Estimating model M1i took about 124 minutes.

505 **4.3.5 Compare the models.** We can compare the two models using the Widely
 506 Applicable Information Criterion (WAIC) and Leave-One-Out (LOO) cross-validation, and
 507 look at model weights for both criteria (Kurz, 2023a; McElreath, 2020).

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

508 ## model_M0i model_M1i
 509 ## 0 1

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

510 ## model_M0i model_M1i
 511 ## 0 1

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

514 **4.3.6 Evaluating parameter estimates in model M1i.** To make inferences

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

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

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

Posterior distributions for population-level effects in Model M1i

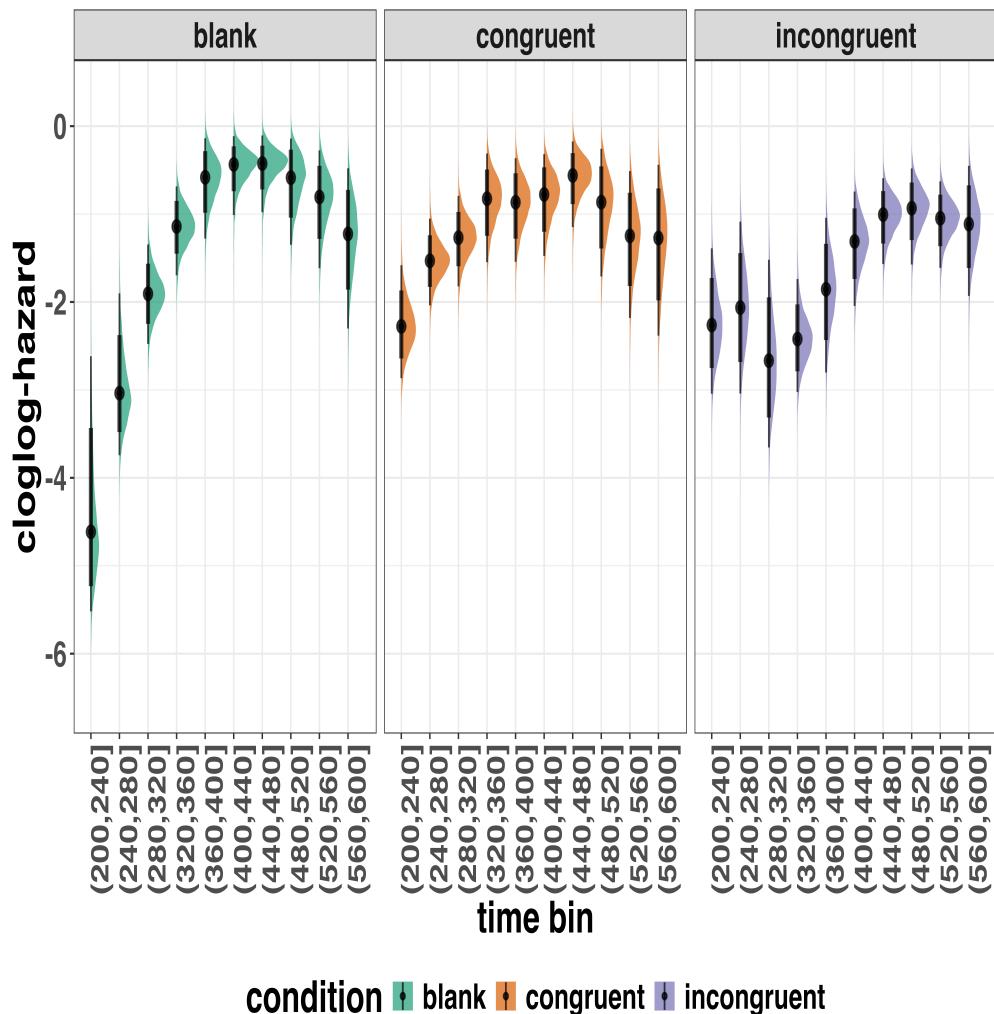


Figure 4. Medians and 80/95% credible intervals of the posterior distributions of the population-level parameters of model M1i.

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

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

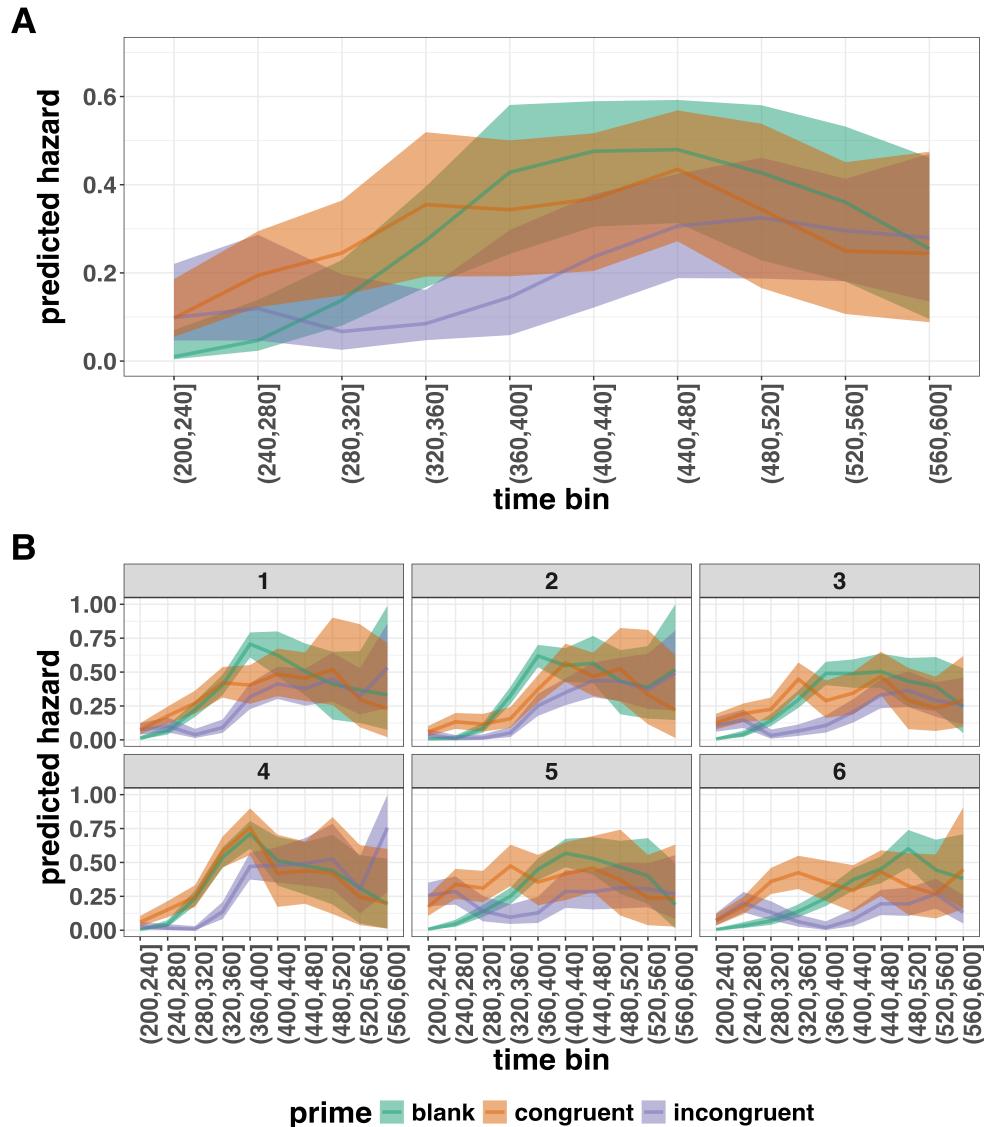


Figure 5. Point (median) and 80/95% credible interval summaries of the hazard estimates (expected values of the draws from the posterior predictive distributions) in each time bin at the population level (A), and for each participant (B).

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

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

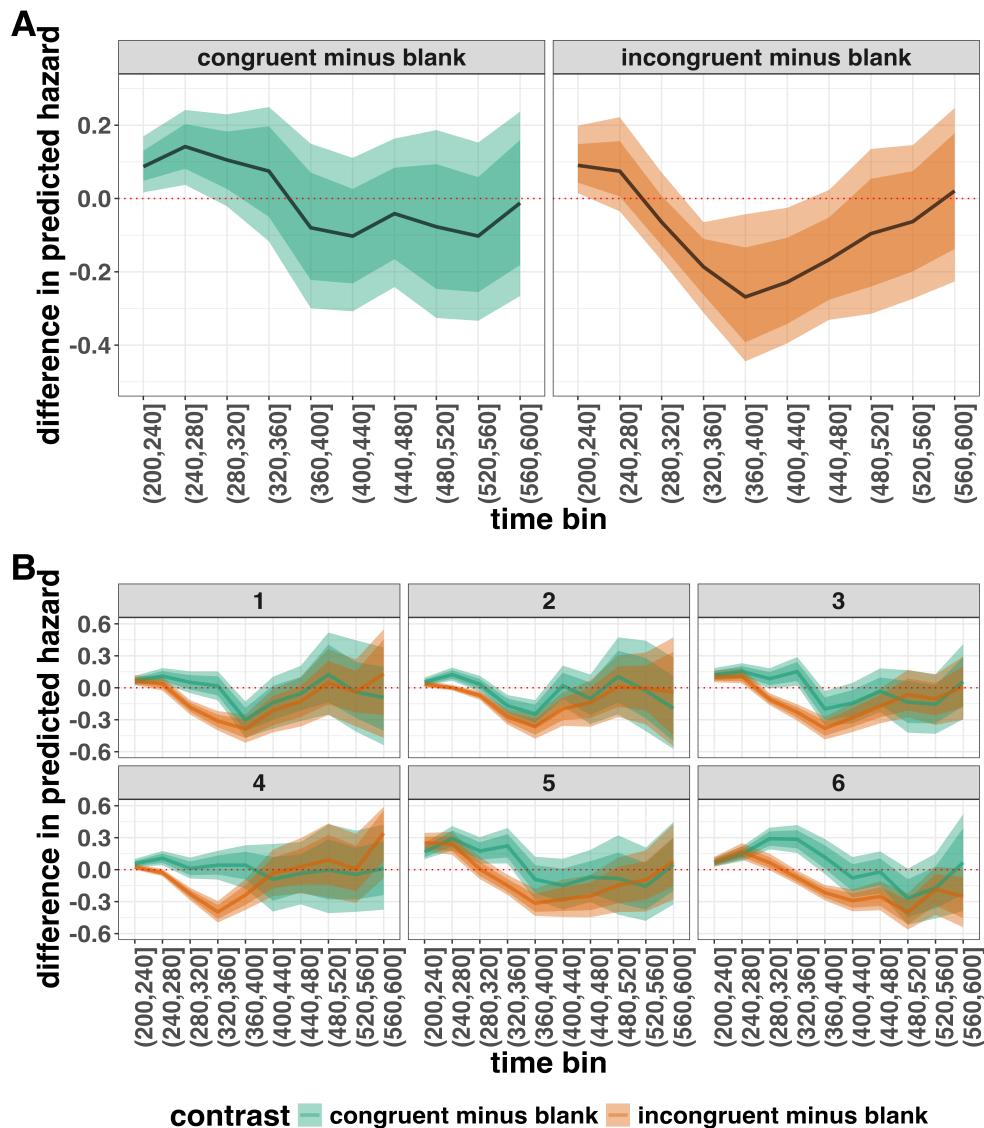


Figure 6. Point (mean) and 80/95% credible interval summaries of estimated differences in hazard in each time bin at the population level (A), and for each participant (B).

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

528 Tutorial_2a.Rmd for details).

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

534 If we want to estimate the population-level effect of prime type on hazard, we can
535 base our conclusion on Figure 7A. The contrast “congruent minus blank” was estimated to
536 be 0.09 hazard units in bin (200,240] (95% CrI = [0.02, 0.17]), and 0.14 hazard units in bin
537 (240,280]) (95% CrI = [0.04, 0.25]). For the other bins, the 95% credible interval contained
538 zero. The contrast “incongruent minus blank” was estimated to be 0.09 hazard units in bin
539 (200,240] (95% CrI = [0.01, 0.21]), -0.19 hazard units in bin (320,360] (95% CrI = [-0.31,
540 -0.06]), -0.27 hazard units in bin (360,400] (95% CrI = [-0.45, -0.04]), and -0.23 hazard
541 units in bin (400,440] (95% CrI = [-0.40, -0.03]). For the other bins, the 95% credible
542 interval contained zero.

543 There are thus two phases of performance for the average person between 200 and
544 600 ms after target onset. In the first phase, the addition of a congruent or incongruent
545 prime stimulus increases the hazard of response occurrence compared to blank prime trials
546 in the time period (200, 240]. In the second phase, only the incongruent prime decreases
547 the hazard of response occurrence compared to blank primes, in the time period (320,440].
548 The sign of the effect of incongruent primes on the hazard of response occurrence thus
549 depends on how much waiting time has passed since target onset.

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

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

565 Interestingly, all subjects show a tendency in their mean difference (congruent minus
566 blank) to “dip” around that time (Figure 7B). Therefore, future modeling efforts could
567 incorporate the trial number into the model formula, in order to also study how the effects
568 of prime type on hazard change on the long experiment-wide time scale, next to the short
569 trial-wide time scale. In Tutorial_2a.Rmd we provide a number of model formulae that
570 should get you going.

571 4.4 Tutorial 2b: Fitting Bayesian conditional accuracy models

572 In this fourth tutorial, we illustrate how to fit a Bayesian multilevel regression model
573 to the timed accuracy data from the masked response priming data used in Tutorial 1a.
574 The general process is similar to Tutorial 2a, except that (a) we use the person-trial data,
575 (b) we use the logit link function, and (c) we change the priors. To keep the tutorial short,
576 we only fit one conditional accuracy model, which was based on model M1i from Tutorial
577 2a and labelled M1i_ca.

578 To make inferences from the parameter estimates in model M1i_ca, we first plot the

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

Posterior distributions for population-level effects in Model M1i_ca

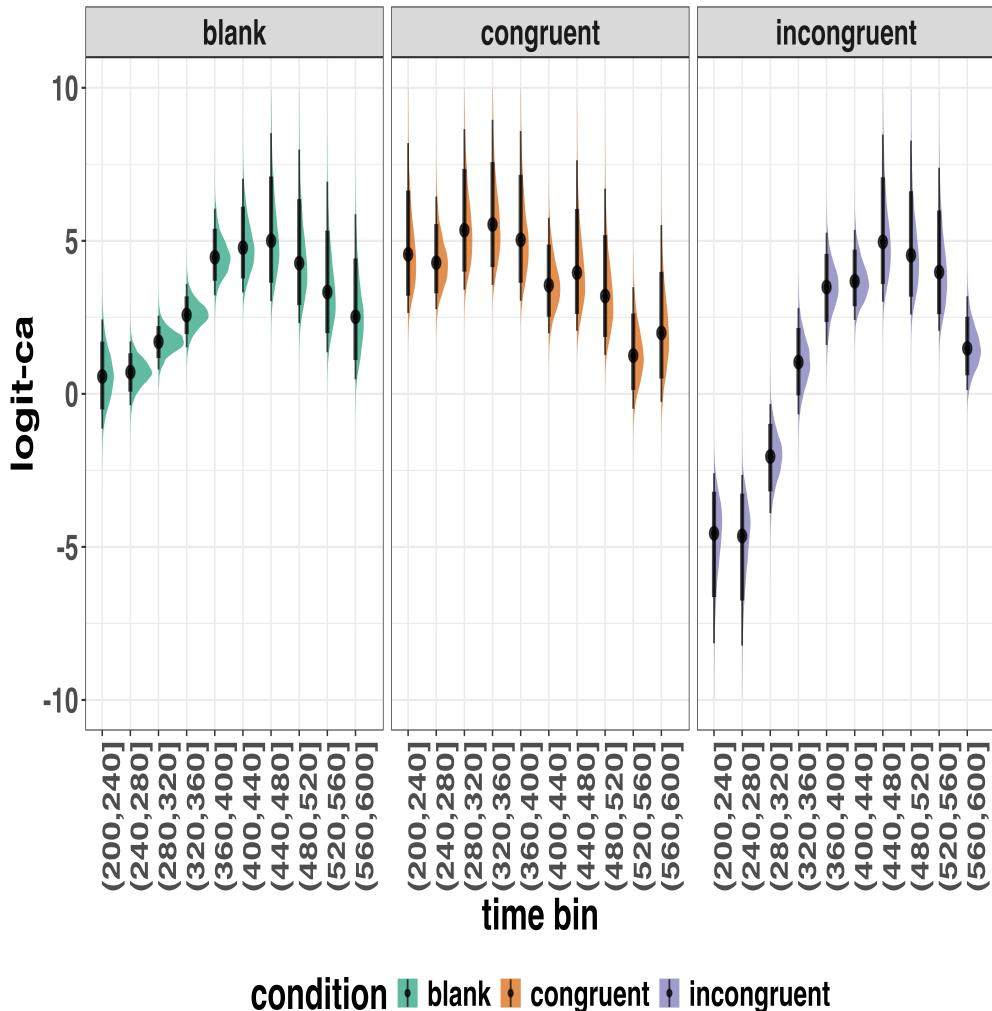


Figure 7. Medians and 80/95% credible intervals of the posterior distributions of the population-level parameters of model M1i_ca. ca = conditional accuracy.

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

⁵⁸⁴ predicted conditional accuracies – at the population level (Figure 9A), and for each
⁵⁸⁵ participant in the data set (Figure 9B).

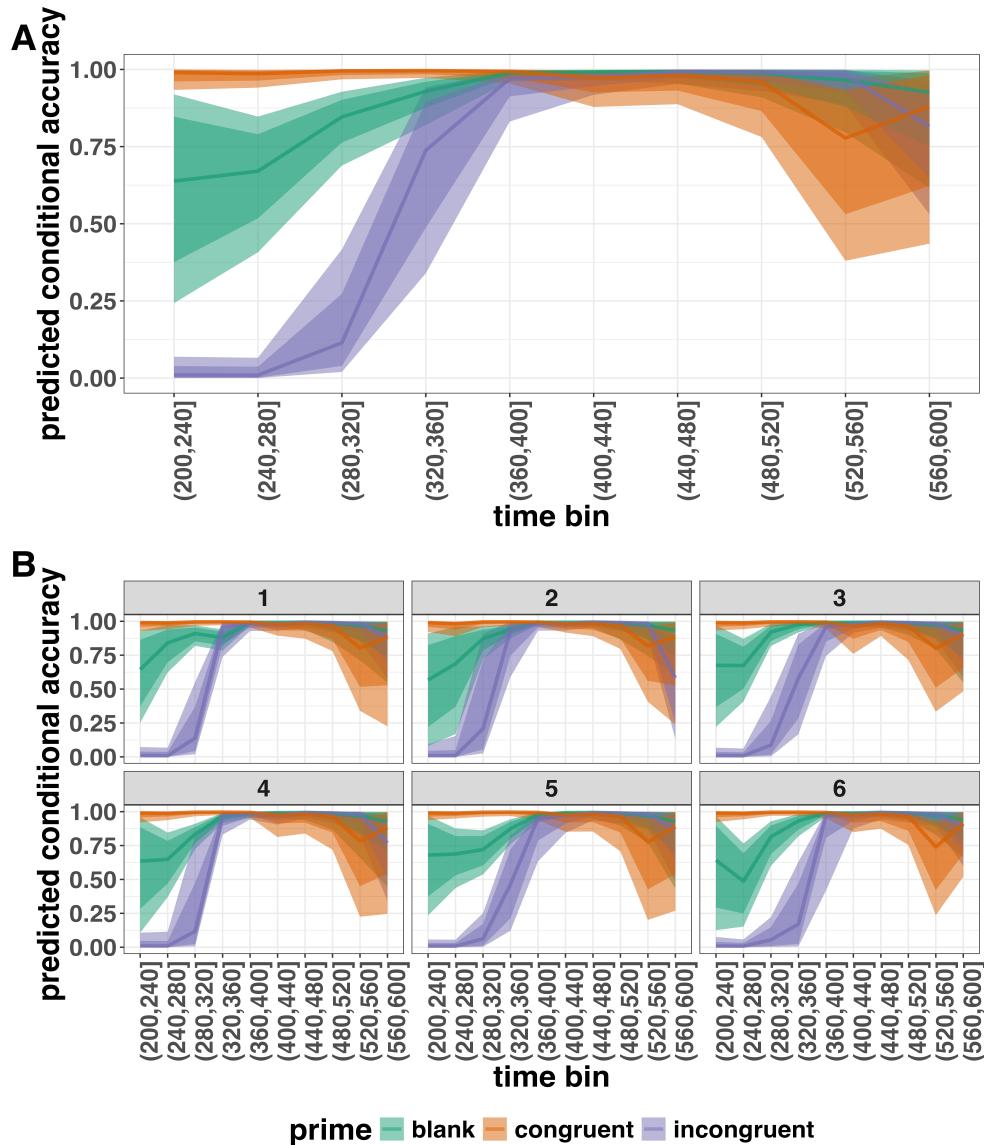


Figure 8. Point (median) and 80/95% credible interval summaries of the conditional accuracy estimates (expected values of the draws from the posterior predictive distributions) in each time bin at the population level (A), and for each participant (B).

⁵⁸⁶ As we are actually interested in the effects of congruent and incongruent primes,
⁵⁸⁷ relative to the blank prime condition, we can construct two contrasts (congruent-blank,

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

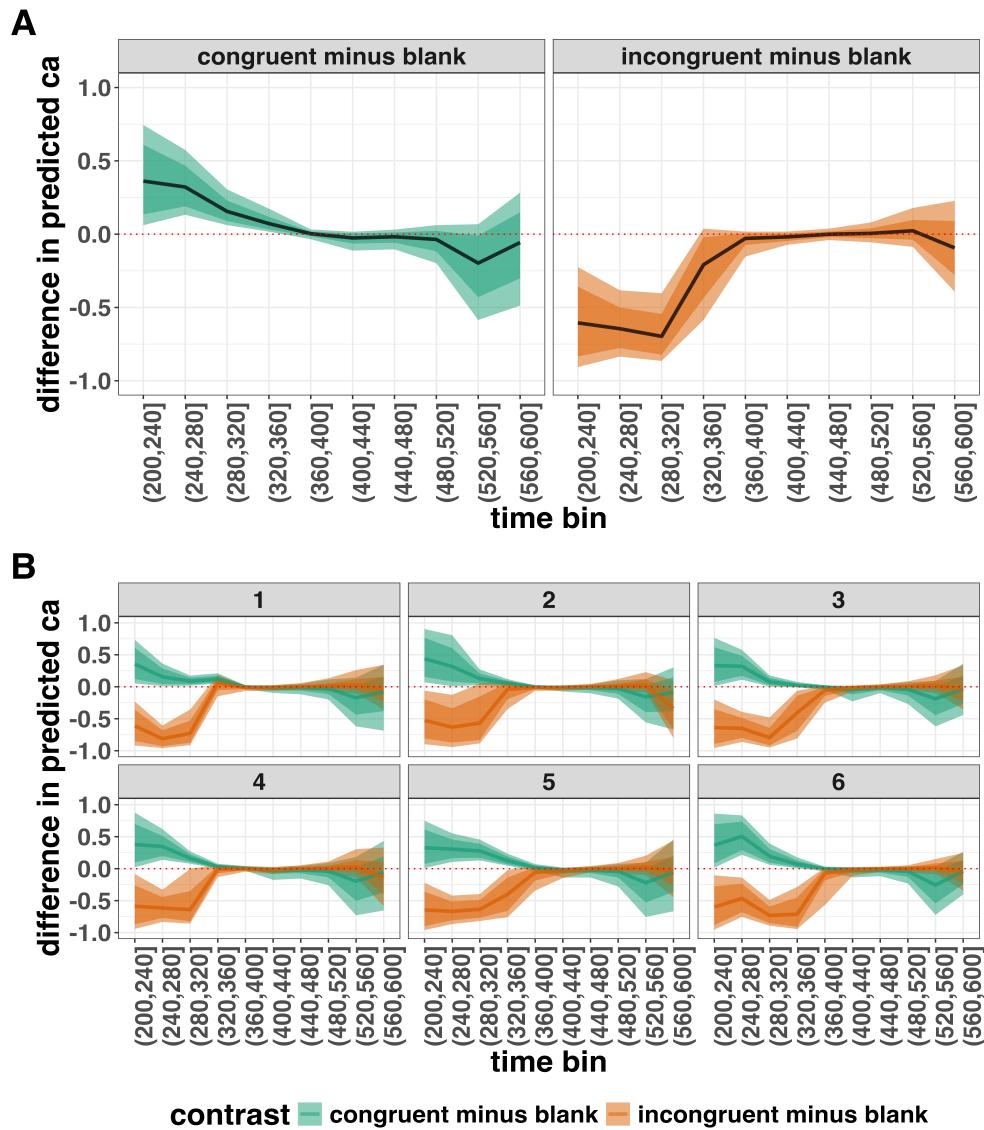


Figure 9. Point (mean) and 80/95% credible interval summaries of estimated differences in conditional accuracy in each time bin at the population level (A), and for each participant (B).

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

592 (240,280], (280,320], and (320,360], relative to the estimates in the baseline condition
593 (blank prime; red dashed lines in Figure 10A). Incongruent primes have a negative effect on
594 the conditional accuracy of emitted responses in the first time bins, relative to the
595 estimates in the baseline condition.

596 **4.5 Tutorial 3a: Fitting Frequentist hazard models**

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

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

604 In Figure 11 we compare the parameter estimates from the Bayesian regression model
605 M1i with those from the frequentist model M1i_f.

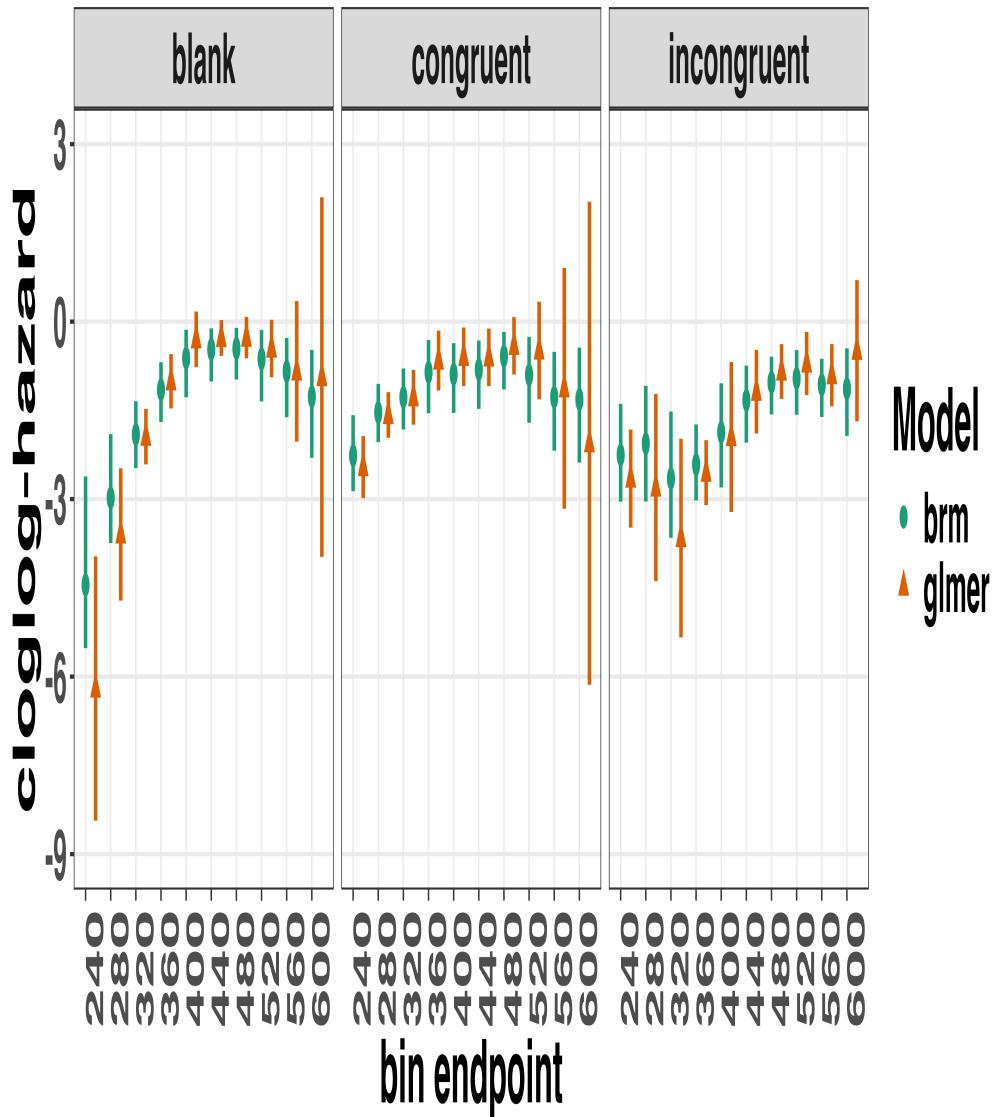


Figure 10. Parameter estimates for model M1i from brm() – means and 95% credible intervals – and model M1i_f from glmer() – maximum likelihood estimates and 95% confidence intervals.

606 Figure 11 confirms that the parameter estimates from both Bayesian and frequentist
 607 models are pretty similar, which makes sense given the close similarity in model structure.
 608 However, model M1i_f did not converge and resulted in a singular fit. This is of course one
 609 of the reasons why Bayesian modeling has become so popular in recent years. But the price

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

613 **4.6 Tutorial 3b: Fitting Frequentist conditional accuracy models**

614 In this sixth tutorial we illustrate how to fit a multilevel regression model to the
615 timed accuracy data in the frequentist framework, for the data used in Tutorial 1a. To be
616 concise, we only fit effects from model M1i_ca (see Tutorial 2b) using the function glmer()
617 from the R package lme4. Alternatively, one could also use the function glmmPQL() from
618 the R package MASS (Ripley et al., 2024). The resulting conditional accuracy model,
619 which we labelled M1i_ca_f, did not converge and resulted in a singular fit. Again, this
620 just highlights some of the difficulties in fitting reasonably complex varying/random effects
621 structures in frequentist workflows.

622 **4.7 Tutorial 4: Planning**

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

624 **4.7.1 Background.** The general approach to planning that we adopt here involves
625 simulating reasonably structured data to help guide what you might be able to expect from
626 your data once you collect it (Gelman, Vehtari, et al., 2020). The basic structure and code
627 follows the examples outlined by Solomon Kurz in his ‘power’ blog posts
628 (<https://solomonkurz.netlify.app/blog/bayesian-power-analysis-part-i/>) and Lisa
629 Debruine’s R package faux{} (<https://debruine.github.io/faux/>) as well as these related
630 papers (DeBruine & Barr, 2021; Pargent, Koch, Kleine, Lemer, & Gaube, 2024).

631 **4.7.2 Basic workflow.** The basic workflow is as follows:

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

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

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

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

645 **4.7.3 Fit a regression model and simulate one dataset.** We again use the
646 data from Panis and Schmidt (2016) to provide a worked example. We fit an index coding
647 model on a subset of time bins (six time bins in total) and for two prime conditions
648 (congruent and incongruent). We chose to focus on a subsample of the data to ease the
649 computational burden. We also used a full varying effects structure, with the model
650 formula as follows:

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

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

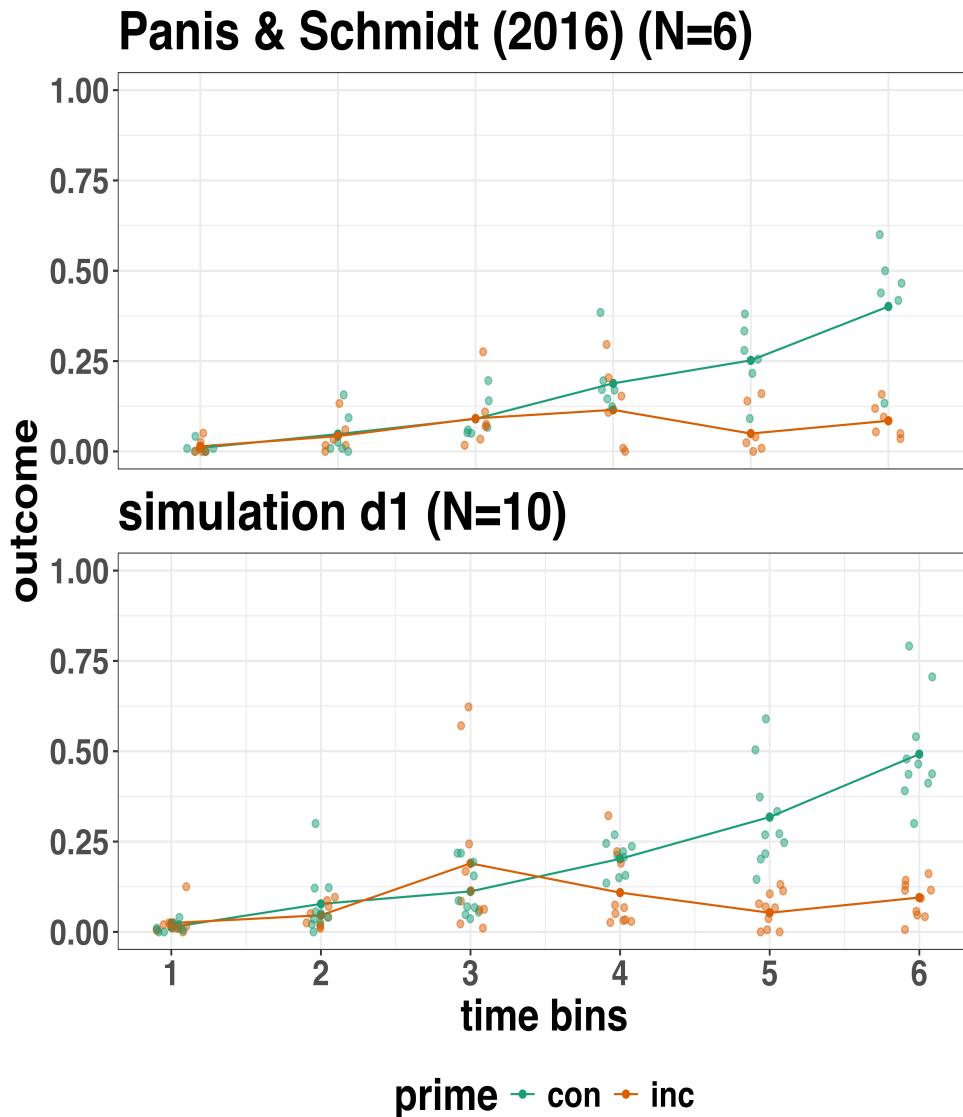


Figure 11. Raw data from Panis and Schmidt (2016) and simulated data from 10 participants.

656 **4.7.4 Simulate and summarise data across a range of parameter values.**

657 Here we use the same data simulation process as used above, but instead of simulating one
 658 dataset, we simulate 1000 datasets per variation in parameter values. Specifically, in
 659 Simulation 1, we vary the number of trials per condition (100, 200, and 400), as well as the
 660 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.

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

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

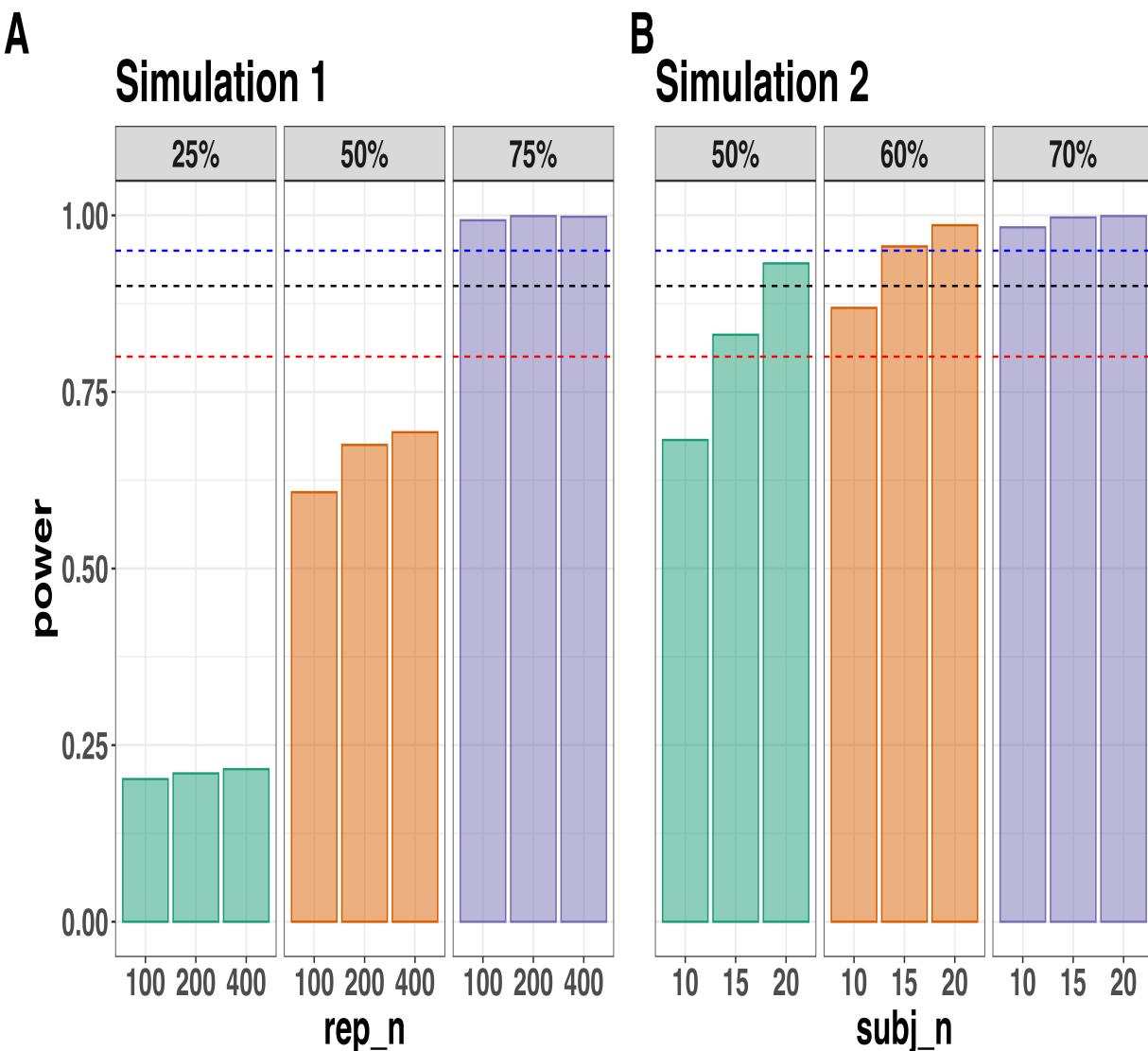


Figure 12. Statistical power across data Simulation 1 (A) and Simulation 2 (B). Power was calculated as the percentage of lower-bound 95% confidence intervals that exclude zero when the difference between prime condition is calculated (congruent - incongruent). In Simulation 1, the effect size was varied 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.

692 **4.7.5 Planning decisions.** Now that we have summarised our simulated data,

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

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

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

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

697 Some considerations might be as follows:

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

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

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

701 personnel?

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

718

5. Discussion

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

734 **5.1 What are the main use-cases of EHA for understanding cognition and brain
735 function?**

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

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

752 5.2 Model complexity versus interpretability

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

769 4).

770 **5.3 Individual differences**

771 One important issue is that of possible individual differences in the overall location of
772 the distribution, and the time course of psychological effects. For example, when you wait
773 for a response of the participant on each trial, you allow the participant to have control
774 over the trial duration, and some participants might respond only when they are confident
775 that their emitted response will be correct. These issues can be avoided by introducing a
776 (relatively short) response deadline in each trial, e.g., 500 ms for simple detection tasks,
777 800 ms for more difficult discrimination tasks, or 2 s for tasks requiring extended high-level
778 processing. Because EHA can deal in a straightforward fashion with right-censored
779 observations (i.e., trials without an observed response in the analysis time window),
780 introducing a response deadline is recommended when designing RT experiments.
781 Furthermore, introducing a response deadline and asking participants to respond before the
782 deadline as much as possible, will also lead to individual distributions that overlap in time,
783 which is important when selecting a common analysis time window when fitting hazard
784 and conditional accuracy models.

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

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

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

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

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

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

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

800 estimate.

801 5.4 Limitations

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

5.5 Extensions

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

843

6. Conclusions

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

852

Author contributions

853 Conceptualization: S. Panis and R. Ramsey; Software: S. Panis and R. Ramsey;
854 Writing - Original Draft Preparation: S. Panis; Writing - Review & Editing: S. Panis and
855 R. Ramsey; Supervision: R. Ramsey.

856

Conflicts of Interest

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

859

Prior versions

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

862

Supplemental Material

863

Disclosures**864 Data, materials, and online resources**

865 Link to public archive:
866 https://github.com/sven-panis/Tutorial_Event_History_Analysis
867 Supplemental Material: Panis_Ramsey_suppl_material.pdf

868 Ethical approval

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

871

References

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

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

- 925 https://doi.org/10.1017/9781139161879
- 926 Gelman, A., Vehtari, A., Simpson, D., Margossian, C. C., Carpenter, B., Yao, Y., ...
- 927 Modrák, M. (2020). *Bayesian Workflow*. arXiv.
- 928 https://doi.org/10.48550/arXiv.2011.01808
- 929 Girard, J. (2024). *Standist: What the package does (one line, title case)*. Retrieved from
- 930 https://github.com/jmgirard/standist
- 931 Grolemund, G., & Wickham, H. (2011). Dates and times made easy with lubridate.
- 932 *Journal of Statistical Software*, 40(3), 1–25. Retrieved from
- 933 https://www.jstatsoft.org/v40/i03/
- 934 Halley, E. (1693). VI. An estimate of the degrees of the mortality of mankind; drawn from
- 935 curious tables of the births and funerals at the city of breslaw; with an attempt to
- 936 ascertain the price of annuities upon lives. *Philosophical Transactions of the Royal*
- 937 *Society of London*, 17(196), 596–610. https://doi.org/10.1098/rstl.1693.0007
- 938 Heiss, A. (2021, November 10). A Guide to Correctly Calculating Posterior Predictions
- 939 and Average Marginal Effects with Multilevel Bayesian Models.
- 940 https://doi.org/10.59350/wbn93-edb02
- 941 Hosmer, D. W., Lemeshow, S., & May, S. (2011). *Applied Survival Analysis: Regression*
- 942 *Modeling of Time to Event Data* (2nd ed). Hoboken: John Wiley & Sons.
- 943 Ince, R. A., Paton, A. T., Kay, J. W., & Schyns, P. G. (2021). Bayesian inference of
- 944 population prevalence. *eLife*, 10, e62461. https://doi.org/10.7554/eLife.62461
- 945 Kantowitz, B. H., & Pachella, R. G. (2021). The Interpretation of Reaction Time in
- 946 Information-Processing Research 1. *Human Information Processing*, 41–82.
- 947 https://doi.org/10.4324/9781003176688-2
- 948 Kay, M. (2024). *tidybayes: Tidy data and geoms for Bayesian models*.
- 949 https://doi.org/10.5281/zenodo.1308151
- 950 Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing,
- 951 estimation, meta-analysis, and power analysis from a Bayesian perspective.

- 952 *Psychonomic Bulletin & Review*, 25(1), 178–206.
- 953 <https://doi.org/10.3758/s13423-016-1221-4>
- 954 Kurz, A. S. (2023a). *Applied longitudinal data analysis in brms and the tidyverse* (version 0.0.3). Retrieved from <https://bookdown.org/content/4253/>
- 955 Kurz, A. S. (2023b). *Statistical rethinking with brms, ggplot2, and the tidyverse: Second edition* (version 0.4.0). Retrieved from <https://bookdown.org/content/4857/>
- 956 Lakens, D. (2022). Sample Size Justification. *Collabra: Psychology*, 8(1), 33267.
- 957 <https://doi.org/10.1525/collabra.33267>
- 958 Landes, J., Engelhardt, S. C., & Pelletier, F. (2020). An introduction to event history analyses for ecologists. *Ecosphere*, 11(10), e03238. <https://doi.org/10.1002/ecs2.3238>
- 959 Makeham, W. M. (1860). *On the Law of Mortality and the Construction of Annuity Tables*. The Assurance Magazine, and Journal of the Institute of Actuaries.
- 960 McElreath, R. (2020). *Statistical Rethinking: A Bayesian Course with Examples in R and STAN* (2nd ed.). New York: Chapman and Hall/CRC.
- 961 <https://doi.org/10.1201/9780429029608>
- 962 Müller, K., & Wickham, H. (2023). *Tibble: Simple data frames*. Retrieved from <https://CRAN.R-project.org/package=tibble>
- 963 Nelder, J. A. (1999). From Statistics to Statistical Science. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 48(2), 257–269. Retrieved from <https://www.jstor.org/stable/2681191>
- 964 Neuwirth, E. (2022). *RColorBrewer: ColorBrewer palettes*. Retrieved from <https://CRAN.R-project.org/package=RColorBrewer>
- 965 Panis, S. (2020). How can we learn what attention is? Response gating via multiple direct routes kept in check by inhibitory control processes. *Open Psychology*, 2(1), 238–279.
- 966 <https://doi.org/10.1515/psych-2020-0107>
- 967 Panis, S., Moran, R., Wolkersdorfer, M. P., & Schmidt, T. (2020). Studying the dynamics of visual search behavior using RT hazard and micro-level speed–accuracy tradeoff
- 968 <https://doi.org/10.1515/psych-2020-0107>
- 969 <https://doi.org/10.1515/psych-2020-0107>
- 970 <https://doi.org/10.1515/psych-2020-0107>
- 971 <https://doi.org/10.1515/psych-2020-0107>
- 972 <https://doi.org/10.1515/psych-2020-0107>
- 973 <https://doi.org/10.1515/psych-2020-0107>
- 974 <https://doi.org/10.1515/psych-2020-0107>
- 975 <https://doi.org/10.1515/psych-2020-0107>
- 976 <https://doi.org/10.1515/psych-2020-0107>
- 977 <https://doi.org/10.1515/psych-2020-0107>
- 978 <https://doi.org/10.1515/psych-2020-0107>

- 979 functions: A role for recurrent object recognition and cognitive control processes.
- 980 *Attention, Perception, & Psychophysics*, 82(2), 689–714.
- 981 <https://doi.org/10.3758/s13414-019-01897-z>
- 982 Panis, S., Schmidt, F., Wolkersdorfer, M. P., & Schmidt, T. (2020). Analyzing Response
983 Times and Other Types of Time-to-Event Data Using Event History Analysis: A Tool
984 for Mental Chronometry and Cognitive Psychophysiology. *I-Perception*, 11(6),
985 2041669520978673. <https://doi.org/10.1177/2041669520978673>
- 986 Panis, S., & Schmidt, T. (2016). What Is Shaping RT and Accuracy Distributions? Active
987 and Selective Response Inhibition Causes the Negative Compatibility Effect. *Journal of
988 Cognitive Neuroscience*, 28(11), 1651–1671. https://doi.org/10.1162/jocn_a_00998
- 989 Panis, S., & Schmidt, T. (2022). When does “inhibition of return” occur in spatial cueing
990 tasks? Temporally disentangling multiple cue-triggered effects using response history
991 and conditional accuracy analyses. *Open Psychology*, 4(1), 84–114.
992 <https://doi.org/10.1515/psych-2022-0005>
- 993 Panis, S., Torfs, K., Gillebert, C. R., Wagemans, J., & Humphreys, G. W. (2017).
994 Neuropsychological evidence for the temporal dynamics of category-specific naming.
995 *Visual Cognition*, 25(1-3), 79–99. <https://doi.org/10.1080/13506285.2017.1330790>
- 996 Panis, S., & Wagemans, J. (2009). Time-course contingencies in perceptual organization
997 and identification of fragmented object outlines. *Journal of Experimental Psychology:
998 Human Perception and Performance*, 35(3), 661–687.
999 <https://doi.org/10.1037/a0013547>
- 1000 Pargent, F., Koch, T. K., Kleine, A.-K., Lermer, E., & Gaube, S. (2024). A Tutorial on
1001 Tailored Simulation-Based Sample-Size Planning for Experimental Designs With
1002 Generalized Linear Mixed Models. *Advances in Methods and Practices in Psychological
1003 Science*, 7(4), 25152459241287132. <https://doi.org/10.1177/25152459241287132>
- 1004 Pedersen, T. L. (2024). *Patchwork: The composer of plots*. Retrieved from
1005 <https://patchwork.data-imaginist.com>

- 1006 Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in s and s-PLUS*. New York:
1007 Springer. <https://doi.org/10.1007/b98882>
- 1008 R Core Team. (2024). *R: A language and environment for statistical computing*. Vienna,
1009 Austria: R Foundation for Statistical Computing. Retrieved from
1010 <https://www.R-project.org/>
- 1011 Ripley, B., Venables, B., Bates, D. M., ca 1998), K. H. (partial. port, ca 1998), A. G.
1012 (partial. port, & polr), D. F. (support. functions for. (2024). *MASS: Support Functions*
1013 and *Datasets for Venables and Ripley's MASS*.
- 1014 Singer, J. D., & Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling*
1015 *Change and Event Occurrence*. Oxford, New York: Oxford University Press.
- 1016 Smith, P. L., & Little, D. R. (2018). Small is beautiful: In defense of the small-N design.
1017 *Psychonomic Bulletin & Review*, 25(6), 2083–2101.
1018 <https://doi.org/10.3758/s13423-018-1451-8>
- 1019 Stan Development Team. (2020). *StanHeaders: Headers for the R interface to Stan*.
1020 Retrieved from <https://mc-stan.org/>
- 1021 Stan Development Team. (2024). *RStan: The R interface to Stan*. Retrieved from
1022 <https://mc-stan.org/>
- 1023 Steele, F., Goldstein, H., & Browne, W. (2004). A general multilevel multistate competing
1024 risks model for event history data, with an application to a study of contraceptive use
1025 dynamics. *Statistical Modelling*, 4(2), 145–159.
1026 <https://doi.org/10.1191/1471082X04st069oa>
- 1027 Teachman, J. D. (1983). Analyzing social processes: Life tables and proportional hazards
1028 models. *Social Science Research*, 12(3), 263–301.
1029 [https://doi.org/10.1016/0049-089X\(83\)90015-7](https://doi.org/10.1016/0049-089X(83)90015-7)
- 1030 Tong, C. (2019). Statistical Inference Enables Bad Science; Statistical Thinking Enables
1031 Good Science. *The American Statistician*, 73(sup1), 246–261.
1032 <https://doi.org/10.1080/00031305.2018.1518264>

- 1033 Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics.
- 1034 *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)
- 1035 Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New
- 1036 York. Retrieved from <https://ggplot2.tidyverse.org>
- 1037 Wickham, H. (2023a). *Forcats: Tools for working with categorical variables (factors)*.
- 1038 Retrieved from <https://forcats.tidyverse.org/>
- 1039 Wickham, H. (2023b). *Stringr: Simple, consistent wrappers for common string operations*.
- 1040 Retrieved from <https://stringr.tidyverse.org>
- 1041 Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., ...
- 1042 Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43),
- 1043 1686. <https://doi.org/10.21105/joss.01686>
- 1044 Wickham, H., Çetinkaya-Rundel, M., & Grolemund, G. (2023). *R for data science: Import,*
- 1045 *tidy, transform, visualize, and model data* (2nd edition). Beijing Boston Farnham
- 1046 Sebastopol Tokyo: O'Reilly.
- 1047 Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *Dplyr: A*
- 1048 *grammar of data manipulation*. Retrieved from <https://dplyr.tidyverse.org>
- 1049 Wickham, H., & Henry, L. (2023). *Purrr: Functional programming tools*. Retrieved from
- 1050 <https://purrr.tidyverse.org/>
- 1051 Wickham, H., Hester, J., & Bryan, J. (2024). *Readr: Read rectangular text data*. Retrieved
- 1052 from <https://readr.tidyverse.org>
- 1053 Wickham, H., Vaughan, D., & Girlich, M. (2024). *Tidyr: Tidy messy data*. Retrieved from
- 1054 <https://tidyr.tidyverse.org>
- 1055 Winter, B. (2019). *Statistics for Linguists: An Introduction Using R*. New York:
- 1056 Routledge. <https://doi.org/10.4324/9781315165547>
- 1057 Wolkersdorfer, M. P., Panis, S., & Schmidt, T. (2020). Temporal dynamics of sequential
- 1058 motor activation in a dual-prime paradigm: Insights from conditional accuracy and
- 1059 hazard functions. *Attention, Perception, & Psychophysics*, 82(5), 2581–2602.

1060 <https://doi.org/10.3758/s13414-020-02010-5>