

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 how they evolve with increasing waiting
17 time. The ability to reveal finer-grained, “temporal states” of cognitive processes can have
18 important consequences for theory development by qualitatively changing the key
19 inferences that are drawn from psychological data. Luckily, well-established analytical
20 approaches, such as event history analysis (EHA), are able to evaluate the detailed shape
21 of time-to-event distributions, and thus characterize the time course of psychological states.
22 One barrier to wider use of EHA, however, is that the analytical workflow is typically more
23 time-consuming and complex than orthodox approaches. To help achieve broader uptake of
24 EHA, in this paper we outline a set of tutorials that detail one distributional method
25 known as discrete-time EHA. We touch upon several key aspects of the workflow, such as
26 how to process raw data and specify regression models, and we also consider the
27 implications for experimental design. We finish the article by considering the benefits of
28 the approach for understanding psychological states, as well as its limitations. Finally, the
29 project is written in R and freely available, which means the approach can easily be
30 adapted to other data sets.

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

33 Word count: 10167 (body) + 1742 (references) + 3458 (body supplemental material)
34 + 539 (refs suppl. mat.)

35 **1. Introduction**

36 **1.1 Motivation and background context: Comparing means versus 37 distributional shapes**

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

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

53 As a simple illustration, Figure 1 summarises simulated data for one subject that
54 shows how comparing means between two conditions can conceal the shapes of the
55 underlying RT and accuracy distributions. Indeed, compared to the aggregation of data
56 across trials (Figure 1A), a distributional approach offers the possibility to reveal the time
57 course of psychological states (Figure 1B). Here we apply a distributional method known as
58 event history analysis (EHA) extended with speed-accuracy tradeoff (SAT) analysis. For
59 example, Figure 1B shows a first state (up to 400 ms after target onset) for which the early
60 upswing in the hazard of response occurrence is equal for both conditions, and the emitted

61 responses are always correct in condition 1 and always incorrect in condition 2. In a second
62 state (400 to 500 ms), the hazard of response occurrence is higher in condition 1, and
63 conditional accuracies are close to .5 in both conditions. In a third state (>500 ms), the
64 effect disappears in hazard, and all conditional accuracies are equal to 1. Note that we will
65 always refer to a time bin by its upper bound. For example, time bin “500” in Figure 1B
66 refers to the time interval running from 400 ms to 500 ms, with the lower bound of 400 ms
67 excluded, and the upper bound of 500 ms included. Importantly from a face-validity
68 perspective, this pattern of simulated data can be seen in the experimental psychology
69 literature (Panis, 2020; Panis & Schmidt, 2016, 2022).

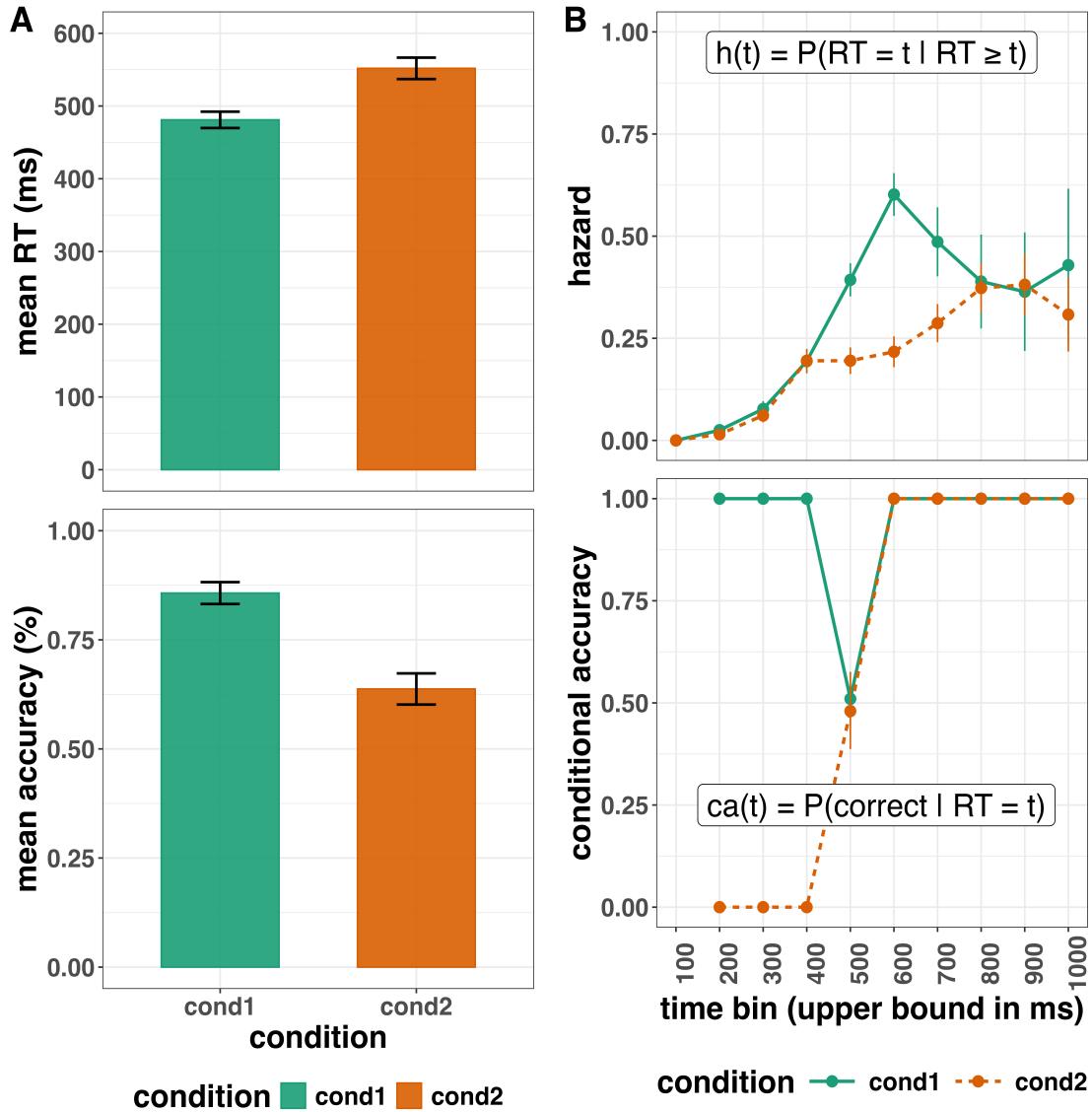


Figure 1. Simulated single-subject data showing mean performance versus a distributional analysis. (A) The mean RT (top) and overall accuracy (bottom) for two conditions are plotted. Two hundred trials are simulated in each condition. (B) The discrete-time hazard functions (top) and conditional accuracy functions (bottom) are plotted for the same data. The first second after target stimulus onset (time zero) is divided in ten time bins of 100 ms (indexed by $t = 1$ to 10). The hazard and conditional accuracy estimates are plotted at the upper bound of each time bin. The definitions of discrete-time hazard and conditional accuracy are further explained in section 2.1.2. Error bars represent ± 1 standard error of the mean (A) or proportion (B).

70 Why does this matter for research in psychology? For many psychological questions,
 71 the estimation of such “temporal states” information can be theoretically meaningful by

72 leading to more fine-grained understanding of psychological processes. Because EHA adds
73 a relatively under-used but ever-present dimension – the passage of time – to the theory
74 building toolkit, it provides one possible response to the recent call for a temporal science
75 of behavior (Abney, Fausey, Suarez-Rivera, & Tamis-LeMonda, 2025).

76 **1.2 Aims**

77 Our ultimate aim in this paper is twofold. First, we want to convince readers of the
78 many benefits of using EHA when dealing with psychological RT data. Second, we want to
79 provide a set of practical tutorials, which provide step-by-step instructions on how you
80 actually perform a (single event) discrete-time EHA on RT data, as well as a
81 complementary discrete-time SAT analysis on timed accuracy data in case of choice RT
82 data (Figure 1B).

83 Even though EHA is a widely used statistical tool and there already exist many
84 excellent reviews (Allison, 1982; Blossfeld & Rohwer, 2002; Box-Steffensmeier, 2004;
85 Hosmer, Lemeshow, & May, 2011; Mills, 2011; Singer & Willett, 2003; Teachman, 1983)
86 and tutorials (Allison, 2010; Elmer, Van Duijn, Ram, & Bringmann, 2023; Landes,
87 Engelhardt, & Pelletier, 2020; Lougheed, Benson, Cole, & Ram, 2019; Stoolmiller, 2015;
88 Stoolmiller & Snyder, 2006), we are not aware of any tutorials that are aimed specifically
89 at psychological RT (+ accuracy) data, and which provide worked examples of the key
90 data processing and Bayesian multilevel regression modelling steps.

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

96 1.3 Structure

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

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

110 **2.1 A brief introduction to event history analysis**

111 EHA is a class of statistical approaches to study the occurrence and timing of events,
112 such as disease onset, marriages, arrests, and job terminations (Allison, 2010). In this
113 section, we want to provide an intuition regarding how EHA works in general, as well as in
114 the context of experimental psychology. For those who want more detailed treatment of
115 EHA and/or regression equations, we refer the reader to several excellent textbooks on
116 these topics (Allison, 2010; Gelman, Hill, & Vehtari, 2020; Mills, 2011; Singer & Willett,
117 2003; Winter, 2019). We also supply relevant regression equations in section E of the
118 Supplemental Material.

119 **2.1.1 Terminology and minimum requirements for EHA.** To avoid possible
120 confusion in terminology used, it is worth noting that EHA is known by various labels,

121 such as survival analysis, hazard analysis, duration analysis, failure-time analysis, and
122 transition analysis (Singer & Willett, 2003). In this paper, we choose to use the term EHA
123 throughout.

124 In terms of minimum requirements to apply EHA, one must be able to:

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

131 These minimal requirements are fulfilled by the RT data obtained in single-button
132 detection tasks, where the time-to-response is repeatedly measured in different trials in the
133 same individual. In section A of the Supplemental Material we visualize this and other
134 types of time-to-event data which are typically obtained in discrimination and bistable
135 perception tasks.

136 **2.1.2 Types of EHA.** There are different types of modeling approaches in EHA.
137 For example, the definition of hazard and the type of models employed depend on whether
138 one is using continuous or discrete time units. As a lab, and mainly for practical reasons,
139 we have much more experience using discrete-time EHA, and that is the approach that we
140 describe and focus on in this paper. This choice may seem counter-intuitive, given that RT
141 is typically treated as a continuous variable. However, continuous forms of EHA require
142 much more data to reliably estimate the continuous-time hazard (rate) function (Bloxom,
143 1984; Luce, 1991; Van Zandt, 2000). Thus, by trading a bit of temporal resolution for a
144 lower number of trials, discrete-time methods seem ideal for dealing with typical
145 psychological RT data sets for which there are less than ~200 trials per condition per

¹⁴⁶ participant (Panis, Schmidt, et al., 2020). Moreover, as indicated by Singer and Willett
¹⁴⁷ (2003), learning discrete-time EHA methods first will help in learning continuous-time
¹⁴⁸ methods, so it seems like a good starting point.

¹⁴⁹ To apply discrete-time EHA, one divides the within-trial time in discrete, contiguous
¹⁵⁰ time bins indexed by t (e.g., $t = 1$ to 10; Figure 1B). Then let RT be a discrete random
¹⁵¹ variable denoting the rank of the time bin in which a particular person's response occurs in
¹⁵² a particular trial across a repeated measures design. For example, a response in one trial
¹⁵³ might occur at 546 ms and it would be in time bin 6 (any RTs from 501 ms to 600 ms).
¹⁵⁴ One then calculates the sample-based estimate of the discrete-time hazard function of
¹⁵⁵ event occurrence for each experimental condition (Figure 1B upper panel). The
¹⁵⁶ discrete-time hazard function gives you, for each time bin, the conditional probability that
¹⁵⁷ the event occurs (sometime) in bin t , given that the event does not occur in previous bins.
¹⁵⁸ In other words, it reflects the instantaneous risk that the event occurs in the current bin t ,
¹⁵⁹ given that it has not yet occurred in the past, i.e., in one of the prior bins ($t-1, t-2, \dots, 1$).

¹⁶⁰ In the context of experimental psychology, it is often (but not always), the case that
¹⁶¹ responses can be classified as correct or incorrect. In those cases, one can also calculate the
¹⁶² conditional accuracy function (Figure 1B lower panel). The conditional accuracy function
¹⁶³ gives you for each time bin the conditional probability that a response is correct given that
¹⁶⁴ it is emitted in time bin t (Allison, 2010; Kantowitz & Pachella, 2021; Wickelgren, 1977).
¹⁶⁵ The conditional accuracy function is also known as the micro-level speed-accuracy tradeoff
¹⁶⁶ (SAT) function. We refer to this extended (hazard + conditional accuracy) analysis for
¹⁶⁷ choice RT data as EHA/SAT. The definitions of these and other discrete-time functions are
¹⁶⁸ given in section B of the Supplemental Material.

169 2.2 Benefits of event history analysis for research in experimental psychology

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

176 We highlight three benefits that we think are relevant to the domain of experimental
177 psychology. First, as illustrated in Figure 1, compared to averaging data across trials,
178 integrating results between hazard functions and their associated conditional accuracy
179 functions for choice RT data can be informative for understanding psychological processes,
180 in terms of inferences about the microgenesis and temporal organization of cognition and
181 theoretical development. As such, the approach permits different kinds of questions to be
182 asked, different inferences to be made, and it holds the potential to discriminate between
183 theoretical accounts of psychological and/or brain-based processes. For example, what kind
184 of theory or set of mechanisms could account for the shape of the functions and the
185 temporally localized effects reported in Figure 1B (Panis & Schmidt, 2016)? Are there new
186 auxiliary assumptions that computational models need to adopt (Panis, Moran, et al.,
187 2020)? Will the temporal effect patterns align nicely with electroencephalography (EEG)
188 findings (Panis & Schmidt, 2022)? And are there new experiments that need to be
189 performed to test the novel predictions that follow from these analyses?

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

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

205 Third, the approach is generalisable and applicable to many tasks that are commonly
206 used in experimental psychology, such as detection, discrimination and bistable perception
207 tasks, and to a range of common experimental manipulations, such as
208 stimulus-onset-asynchrony (see section A of the Supplemental Material). The upshot is
209 that one general analytical approach, which holds several potential advantages, is widely
210 applicable to many substantive use-cases in the RT domain of experimental psychology,
211 irrespective of the analyst's current view on the nature of cognition (Barack & Krakauer,
212 2021).

213 2.3 Implications for research design in experimental psychology

214 Performing EHA in experimental psychology has implications for how experiments
215 are designed. More specifically, we consider three implications that researchers will need to
216 consider when using discrete-time EHA. First, because EHA deals with right-censored
217 observations, one can use a fixed response deadline in each trial. This will increase design
218 efficiency as one does not need to wait for very long RTs that would be trimmed anyway.

219 Second, since the number of trials per condition are spread across bins, it is

220 important to have a relatively large number of trial repetitions per participant and per
221 condition. Accordingly, experimental designs using this approach typically focus on
222 factorial, within-subject designs, in which a large number of observations are made on a
223 relatively small number of participants (so-called small-*N* designs). This approach
224 emphasizes the precision and reproducibility of data patterns at the individual participant
225 level to increase the inferential validity of the design (Baker et al., 2021; Smith & Little,
226 2018). Note that because statistical power derives both from the number of participants
227 and from the number of repeated measures per participant and condition, small-*N* designs
228 can still achieve what are generally considered acceptable levels of statistical power, if they
229 have a sufficient amount of data overall (Baker et al., 2021; Smith & Little, 2018).

230 Third, the width of each time bin will need to be determined. For instance, in Figure
231 1B we chose 100 ms in an arbitrary manner. In reality, however, bin width will need to be
232 set by considering a number of factors simultaneously. The optimal bin width will depend
233 on (a) the length of the observation period in each trial, (b) the rarity of event occurrence,
234 (c) the number of repeated measures (or trials) per condition per participant, and (d) the
235 shape of the hazard function. Finding an appropriate bin width in a given user case before
236 fitting models will require testing a number of options, when calculating and plotting the
237 descriptive statistics (see section 3.1). The goal is to find the smallest bin width that is
238 supported by the amount of data available. Based on our experience, a bin width of 50 ms
239 is a good starting value when the number of repeated measures is 100 or less. Overly small
240 bin widths will result in erratic hazard functions as many bins will have no events, and
241 thus hazard estimates of zero. Of note, however, is that time bins do not need to have the
242 same width. For example, Panis (2020) used larger bins towards the end of the observation
243 period, as fewer events occurred there.

244

3. Tutorials

245 Tutorials 1a and 1b show how to calculate and plot the descriptive statistics of
246 EHA/SAT when there are one or two independent variables, respectively. Tutorials 2a and
247 2b illustrate how to use Bayesian multilevel modeling to fit hazard and conditional
248 accuracy models, respectively. Tutorials 3a and 3b show how to implement, respectively,
249 multilevel models for hazard and conditional accuracy in the frequentist framework.
250 Tutorial 4 shows how to use simulation and power analysis for planning experiments.
251 Additionally, to further simplify the process for other users, the first two tutorials rely on a
252 set of our own custom functions that make sub-processes easier to automate, such as data
253 wrangling and plotting functions (see section C of the Supplemental Material for a list of
254 the custom functions).

255 The content of the tutorials, in terms of EHA and multilevel regression modelling, is
256 mainly based on Allison (2010), Singer and Willett (2003), McElreath (2020), Heiss (2021),
257 Kurz (2023a), and Kurz (2023b). We used R (Version 4.5.1; R Core Team, 2024)¹,
258 for all reported analyses.

¹ We, furthermore, used the R-packages *bayesplot* (Version 1.13.0; Gabry, Simpson, Vehtari, Betancourt, & Gelman, 2019), *brms* (Version 2.22.0; Bürkner, 2017, 2018, 2021), *citr* (Version 0.3.2; Aust, 2019), *cmdstanr* (Version 0.9.0.9000; Gabry, Češnovar, Johnson, & Brondum, 2024), *dplyr* (Version 1.1.4; Wickham, François, Henry, Müller, & Vaughan, 2023), *forcats* (Version 1.0.0; Wickham, 2023a), *futures* (Bengtsson, 2021), *ggplot2* (Version 3.5.2; Wickham, 2016), *lme4* (Version 1.1.37; Bates, Mächler, Bolker, & Walker, 2015), *lubridate* (Version 1.9.4; Grolemund & Wickham, 2011), *Matrix* (Version 1.7.3; Bates, Maechler, & Jagan, 2024), *nlme* (Version 3.1.168; Pinheiro & Bates, 2000), *papaja* (Version 0.1.3; Aust & Barth, 2024), *patchwork* (Version 1.3.0; Pedersen, 2024), *purrr* (Version 1.0.4; Wickham & Henry, 2023), *RColorBrewer* (Version 1.1.3; Neuwirth, 2022), *Rcpp* (Eddelbuettel & Balamuta, 2018; Version 1.0.14; Eddelbuettel & François, 2011), *readr* (Version 2.1.5; Wickham, Hester, & Bryan, 2024), *rstan* (Version 2.32.7; 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 3.3.0; Müller & Wickham, 2023), *tidybayes* (Version 3.0.7; Kay, 2024), *tidyverse* (Version 1.3.1; Wickham, Vaughan, & Girlich, 2024), *tidyverse* (Version 2.0.0; Wickham et al., 2019) and *tinylabels* (Version 0.2.5; Barth, 2023).

259 **3.1 Tutorial 1a: Calculating descriptive statistics using a life table**

260 **3.1.1 Data wrangling aims.** Our data wrangling procedures serve two related
 261 purposes. First, we want to calculate descriptive statistics for each condition in each
 262 individual using a life table. A life table (see Table 3) includes for each time bin indexed by
 263 t , the risk set (i.e., the number of trials that are event-free at the start of the bin), the
 264 number of observed events, and the estimates of the discrete-time hazard probability $h(t)$,
 265 survival probability $S(t)$, probability mass $P(t)$, possibly the conditional accuracy $ca(t)$,
 266 and their estimated standard errors (se). The definitions of these quantities are provided in
 267 section B of the Supplemental Material.

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

Table 1
Data structure for ‘person-trial’ data

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

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

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

Table 2
Data structure for ‘person-trial-bin’ data

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

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

283 **3.1.2 A real data wrangling example.** To illustrate how to quickly set up life
 284 tables for calculating the descriptive statistics (functions of discrete time), we use a
 285 published data set on masked response priming from Panis and Schmidt (2016), who were
 286 interested in the temporal dynamics of the effect of prime-target congruency in RT and
 287 accuracy data. In their first experiment, Panis and Schmidt (2016) presented a double
 288 arrow for 94 ms that pointed left or right as the target stimulus with an onset at time

289 point zero in each trial. Participants had to indicate the direction in which the double
 290 arrow pointed using their corresponding index finger, within 800 ms after target onset.
 291 Response time and accuracy were recorded on each trial. Prime type (blank, congruent,
 292 incongruent) and mask type were manipulated across trials (i.e., repeated measures of
 293 time-to-response). Here we focus for each participant on the subset of 220 trials in which
 294 no mask was presented. The 13-ms prime stimulus was a double arrow presented 187 ms
 295 before target onset in the congruent (same direction as target) and incongruent (opposite
 296 direction as target) prime conditions.

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

300 The required column names are as follows:

- 301 • “pid”, indicating unique participant IDs;
- 302 • “trial”, indicating each unique trial per participant;
- 303 • “condition”, a factor indicating the levels of the independent variable (1, 2, ...) and
 the corresponding labels;
- 304 • “rt”, indicating the response times in ms;
- 305 • “acc”, indicating the accuracies (1/0).

307 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")))
```

308 Next, we can set up the life tables and plot for each participant and condition the
 309 discrete-time hazard function $h(t)$, survivor function $S(t)$, probability mass function $P(t)$,
 310 and conditional accuracy function $ca(t)$. To do so using a functional programming
 311 approach, one has to nest the person-trial data within participants using the `group_nest()`
 312 function, and supply a user-defined censoring time and bin width to our custom function
 313 “`censor()`”, as follows.

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

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

314 Note that the censoring time (here: 600 ms) should be a multiple of the bin width
 315 (here: 40 ms). The censoring time should be a time point after which no informative
 316 responses are expected anymore, in case one waits for a response in each trial. In
 317 experiments that implement a response deadline in each trial the censoring time can equal
 318 that deadline time point. Trials with a RT larger than the censoring time, or trials in
 319 which no response is emitted during the data observation period, are treated as
 320 right-censored observations in EHA. In other words, these trials are not discarded, because
 321 they contain the information that the event did not occur before the censoring time.
 322 Removing such trials before calculating the mean event time would result in

323 underestimation of the true mean.

324 The person-trial-bin oriented data set is created by our custom function ptb(), and it
325 has one row for each time bin (of each trial) that is at risk for event occurrence. The
326 variable “event” in the person-trial-bin oriented data set indicates whether a response
327 occurs (1) or not (0) for each bin. The next steps are to set up the life table using our
328 custom function setup_lt(), calculate the conditional accuracies using our custom function
329 calc_ca(), add the ca(t) estimates to the life table using our custom function join_lt_ca(),
330 and then plot the descriptive statistics using our custom function plot_eha(). One can now
331 inspect different aspects, including the life table for a particular condition of a particular
332 subject, and a plot of the different functions for a particular participant.

333 In general, it is important to visually inspect the functions first for each participant,
334 in order to identify individuals that may not be following task instructions (e.g., a flat
335 conditional accuracy function at .5 indicates that someone is just guessing), outlying
336 individuals, and/or different groups with qualitatively different behavior. Also, to select a
337 suited bin width for model fitting, one can test and compare various bin widths in the
338 censor function, and select the smallest one that is supported by the data.

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

Table 3

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

bin	index t	RS	#events	h(t)	se[h(t)]	S(t)	se[S(t)]	ca(t)	se[ca(t)]	P(t)	se[P(t)]
0	0	220	NA	NA	NA	1.00	0.00	NA	NA	0.00	0.00
40	1	220	0	0.00	0.00	1.00	0.00	NA	NA	0.00	0.00
80	2	220	0	0.00	0.00	1.00	0.00	NA	NA	0.00	0.00
120	3	220	0	0.00	0.00	1.00	0.00	NA	NA	0.00	0.00
160	4	220	0	0.00	0.00	1.00	0.00	NA	NA	0.00	0.00
200	5	220	0	0.00	0.00	1.00	0.00	NA	NA	0.00	0.00
240	6	220	0	0.00	0.00	1.00	0.00	NA	NA	0.00	0.00
280	7	220	7	0.03	0.01	0.97	0.01	0.29	0.17	0.03	0.01
320	8	213	13	0.06	0.02	0.91	0.02	0.77	0.12	0.06	0.02
360	9	200	26	0.13	0.02	0.79	0.03	0.92	0.05	0.12	0.02
400	10	174	40	0.23	0.03	0.61	0.03	1.00	0.00	0.18	0.03
440	11	134	48	0.36	0.04	0.39	0.03	0.98	0.02	0.22	0.03
480	12	86	37	0.43	0.05	0.22	0.03	1.00	0.00	0.17	0.03
520	13	49	32	0.65	0.07	0.08	0.02	1.00	0.00	0.15	0.02
560	14	17	9	0.53	0.12	0.04	0.01	1.00	0.00	0.04	0.01
600	15	8	4	0.50	0.18	0.02	0.01	1.00	0.00	0.02	0.01

Note. The column named “bin” indicates the upper bound of each time bin (in ms), and includes time point zero. At time point zero, no events can occur and therefore both the discrete-time hazard $h(t=0)$ and the conditional accuracy $ca(t=0)$ are undefined. RS = risk set; se = standard error; NA = undefined.

341 Figure 2 displays the discrete-time hazard, survivor, conditional accuracy, and

342 probability mass functions for each prime condition for participant 6. By using

343 discrete-time hazard functions of event occurrence – in combination with conditional

344 accuracy functions for two-choice tasks – one can provide an unbiased, time-varying, and

345 probabilistic description of the latency and accuracy of responses based on all trials of any

346 RT data set.

Descriptive stats for subject 6

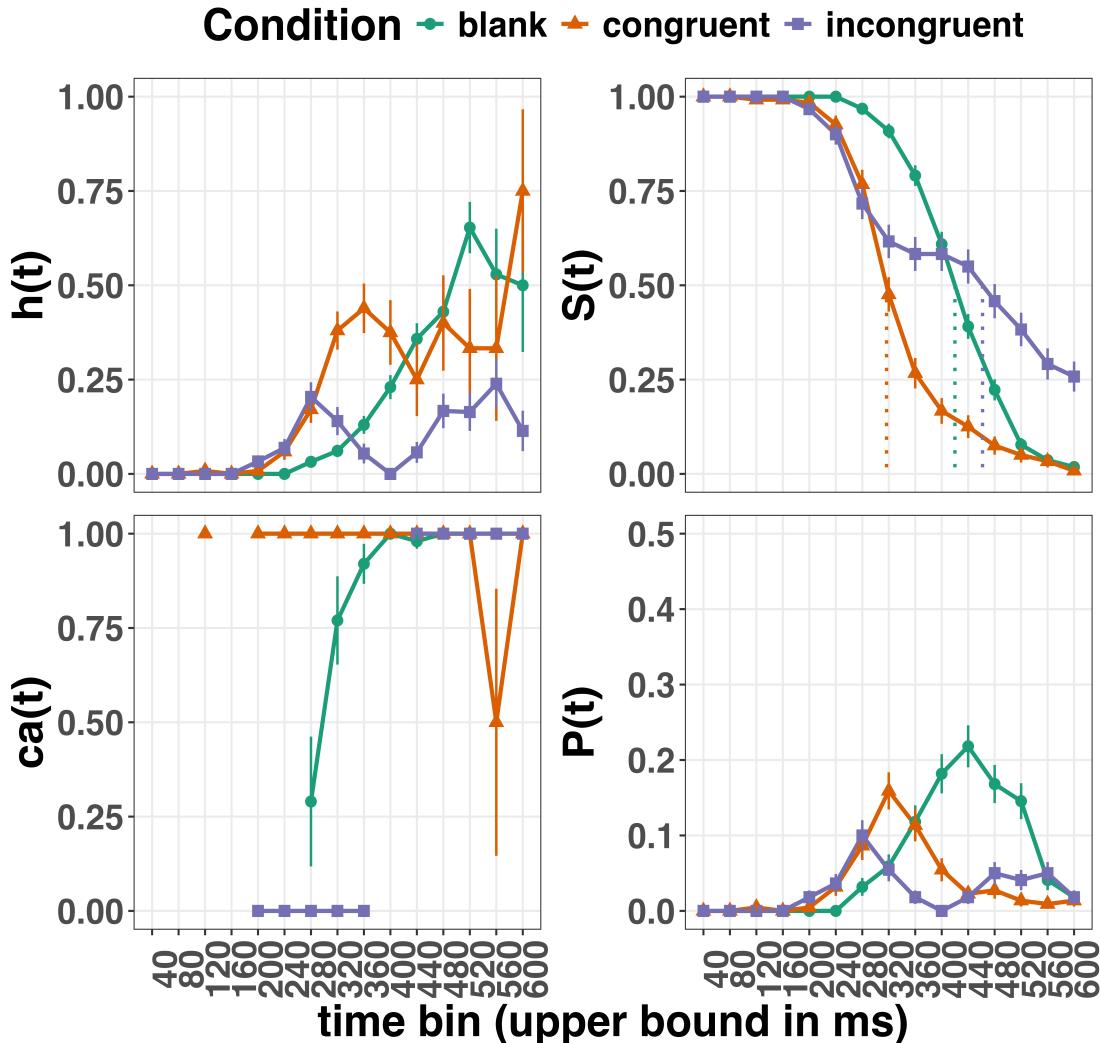


Figure 2. Estimated sample-based discrete-time hazard (h), survivor (S), conditional accuracy (ca) and probability mass (P) functions for participant 6. Vertical dotted lines indicate the estimated median RTs. Error bars represent ± 1 standard error of the respective proportion.

347 For example, for participant 6, the estimated hazard values in bin 280 are 0.03, 0.17,
 348 and 0.20 for the blank, congruent, and incongruent prime conditions, respectively. In other
 349 words, when the waiting time has increased until 240 ms after target onset, then the
 350 conditional probability of response occurrence in the next 40 ms is more than five times
 351 larger for both prime-present conditions, compared to the blank prime condition.

352 Furthermore, the estimated conditional accuracy values in bin 280 are 0.29, 1, and 0

353 for the blank, congruent, and incongruent prime conditions, respectively. In other words, if

354 a response is emitted in bin 280, then the probability that it is correct is estimated to be

355 0.29, 1, and 0 for the blank, congruent, and incongruent prime conditions, respectively.

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

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

358 0.36, 0.25, and 0.06 for the blank, congruent, and incongruent prime conditions,

359 respectively. And when a response does occur in bin 440, then the probability that it is

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

361 conditions, respectively.

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

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

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

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

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

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

368 group-average distribution per condition (see Tutorial_1a.Rmd). More generally, these

369 results go against the (often implicit) assumption in research on priming that all observed

370 responses are primed responses to the target stimulus. Instead, the distributional data

371 show that fast responses are triggered exclusively by the prime stimulus, while only the

372 slower responses reflect primed responses to the target stimulus.

373 At this point, we have calculated and plotted the descriptive statistics for each type

374 of prime stimulus. As we will show in later Tutorials, statistical models for hazard and

375 conditional accuracy functions can be implemented as generalized linear mixed regression

376 models predicting event occurrence (1/0) and conditional accuracy (1/0) in each bin of a

377 selected time window for analysis. But first we consider calculating the descriptive

378 statistics for within-subject designs with two independent variables.

379 **3.2 Tutorial 1b: Generalising to a more complex design**

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

386 To this end, Tutorial 1b illustrates how to calculate and plot the descriptive statistics
387 for the full data set of Experiment 1 of Panis and Schmidt (2016), which includes two
388 independent variables: mask type and prime type. As we use the same functional
389 programming approach as in Tutorial 1a, we simply refer the reader to Tutorial_1b.Rmd.

390 **3.3 Tutorial 2a: Fitting Bayesian hazard models to interval-censored RT data**

391 In this third tutorial, we illustrate how to fit Bayesian multilevel regression models to
392 the RT data of the masked response priming data used in Tutorial 1a. Fitting (Bayesian or
393 non-Bayesian) regression models to time-to-event data is important when you want to
394 study how the shape of the hazard function depends on various predictors (Singer &
395 Willett, 2003).

396 In general, when fitting regression models, our lab adopts an estimation approach to
397 multilevel regression (Kruschke & Liddell, 2018; Winter, 2019), which is heavily influenced
398 by the Bayesian framework as suggested by Richard McElreath (Kurz, 2023b; McElreath,
399 2020). We also use a “keep it maximal” approach by specifying a full varying (or random)
400 effects structure (Barr, Levy, Scheepers, & Tily, 2013). This means that wherever possible
401 we include varying intercepts and slopes per participant. To make inferences, we use two

402 main approaches. We compare models of different complexity using information criteria
403 and cross-validation, to evaluate out-of-sample predictive accuracy (McElreath, 2020). We
404 also take the most complex model and evaluate key parameters of interest using point and
405 interval estimates.

406 **3.3.1 Hazard model considerations.** There are several analytic decisions one
407 has to make when fitting a discrete-time hazard model. First, because the first few bins
408 often contain no responses, one has to select an analysis time window, i.e., a contiguous set
409 of bins for which there is data for each participant. Second, given that the dependent
410 variable (event occurrence) is binary, one has to select a link function (see section D of the
411 Supplemental Material). The cloglog link is preferred over the logit link when events can
412 occur in principle at any time point within a bin, which is the case for RT data (Singer &
413 Willett, 2003). Third, one has to choose whether to treat TIME (i.e., the time bin index t)
414 as a categorical or continuous predictor (see also section E of the Supplemental Material).
415 For example, when you want to know if cloglog-hazard is changing linearly or quadratically
416 over time, you should treat TIME as a continuous predictor. When you are only interested
417 in the effect of covariates on hazard, you can treat TIME as a categorical predictor (i.e., fit
418 an intercept for each bin), in which case you can choose between reference coding and
419 index coding. With reference coding, one defines the variable as a factor and selects one of
420 the k categories as the reference level. Brm() will then construct $k-1$ indicator variables
421 (see model M1d in Tutorial_2a.Rmd for an example). With index coding, one constructs
422 an index variable that contains integers that correspond to different categories (see models
423 M0i and M1i below). As explained by McElreath (2020), the advantage of index coding is
424 that the same prior can be assigned to each level of the index variable, so that each
425 category has the same prior uncertainty.

426 In the case of a large- N design without repeated measurements, the parameters of a
427 discrete-time hazard model can be estimated using standard logistic regression software
428 after expanding the typical person-trial data set into a person-trial-bin data set (Allison,

429 2010). When there is clustering in the data, as in the case of a small- N design with
 430 repeated measurements, the parameters of a discrete-time hazard model can be estimated
 431 using population-averaged methods (e.g., Generalized Estimating Equations), and Bayesian
 432 or frequentist generalized linear mixed models (Allison, 2010).

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

438 In tutorial_2a.Rmd we fit several Bayesian multilevel models (i.e., generalized linear
 439 mixed models) that differ in complexity to the person-trial-bin oriented data set that we
 440 created in Tutorial 1a. We decided to select the 200-600 ms time window for inferential
 441 analysis, and the cloglog link. Below, we shortly discuss two of these models. The
 442 person-trial-bin data set is prepared as follows.

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

ptb_data <- ptb_data %>%
  # select analysis time range: 200-600 ms, 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)))
```

443 **3.3.2 Prior distributions.** To get the posterior distribution of each model

444 parameter given the data, we need to specify prior distributions for the model parameters
 445 which reflect our prior beliefs. In Tutorial_2a.Rmd we perform a few prior predictive
 446 checks to make sure our selected prior distributions reflect our prior beliefs (Gelman,
 447 Vehtari, et al., 2020).

448 The middle column of Supplementary Figure 3 (section F of the Supplemental

449 Material) shows six examples of prior distributions for an intercept on the logit and/or
 450 cloglog scales. While a normal distribution with relatively large variance is often used as a
 451 weakly informative prior for continuous dependent variables, rows A and B of
 452 Supplementary Figure 3 show that specifying such distributions on the logit and cloglog
 453 scales actually leads to rather informative distributions on the original probability scale, as
 454 most mass is pushed to probabilities of 0 and 1. As such, we modify the prior formulation
 455 in order to make sure that it remains consistent with a weakly informative approach (see
 456 section F of the Supplemental Material).

457 **3.3.3 Model M0i: A null model with index coding.** When you do not want to

458 make assumptions about the shape of the hazard function, or its shape is not smooth but
 459 irregular, then you can use a general specification of TIME, i.e., fit one grand intercept per
 460 time bin. In this first baseline or reference model, we use a general specification of TIME
 461 using index coding, and do not include experimental predictors. We call this model “M0i”.

462 The other model (see section 3.3.4) extends model M0i by including our experimental
 463 predictor prime type.

464 Before we fit model M0i, we select the necessary columns from the data, and specify

465 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")

```

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

472 **3.3.4 Model M1i: Adding the effects of prime-target congruency.** Previous
 473 research has shown that psychological effects typically change over time (Panis, 2020;
 474 Panis, Moran, et al., 2020; Panis & Schmidt, 2022; Panis et al., 2017; Panis & Wagemans,
 475 2009). In the next model, therefore, we use index coding for both TIME (variable
 476 “timebin”) and the categorical predictor prime-target-congruency (variable “prime”), so
 477 that we get 30 grand intercepts, one for each combination of timebin level and prime level.
 478 Here is the model formula of this model that we call “M1i”.

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

479 Estimating model M1i took about 124 minutes using the same MacBook Pro.

480 **3.3.5 Compare the models.** There are two popular strategies to evaluate how

481 well models will perform in predicting new data on average: Leave-One-Out (LOO)
 482 cross-validation and the Widely Applicable Information Criterion or WAIC (McElreath,
 483 2020). LOO-weights represent the optimal linear combination of models for predictive
 484 performance, with higher weights for models with better out-of-sample predictive
 485 performance. WAIC-weights represent the relative evidence for each model, with higher
 486 weights for models with a better fit while accounting for model complexity (Kurz, 2023a;
 487 McElreath, 2020).

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

488 ## model_M0i model_M1i
 489 ## "0.00" "1.00"

```
model_weights(model_M0i, model_M1i, weights = "waic") %>% round(digits = 1) %>% format(nsmall=2)
```

490 ## model_M0i model_M1i
 491 ## "0.00" "1.00"

492 Clearly, both the loo and waic weighting schemes assign a weight of 1 to model M1i,
 493 and a weight of 0 to model M0i.

494 **3.3.6 Evaluating parameter estimates in model M1i.** To make causal

495 inferences from the parameter estimates in model M1i (Frank et al., 2025), we first plot the
 496 densities of the draws from the posterior distributions of its population-level parameters in
 497 Figure 3A, together with point (median) and interval estimates (80% and 95% credible
 498 intervals). A credible interval is a range of values that contains a parameter's true value
 499 with a specified probability, given the observed data and model.

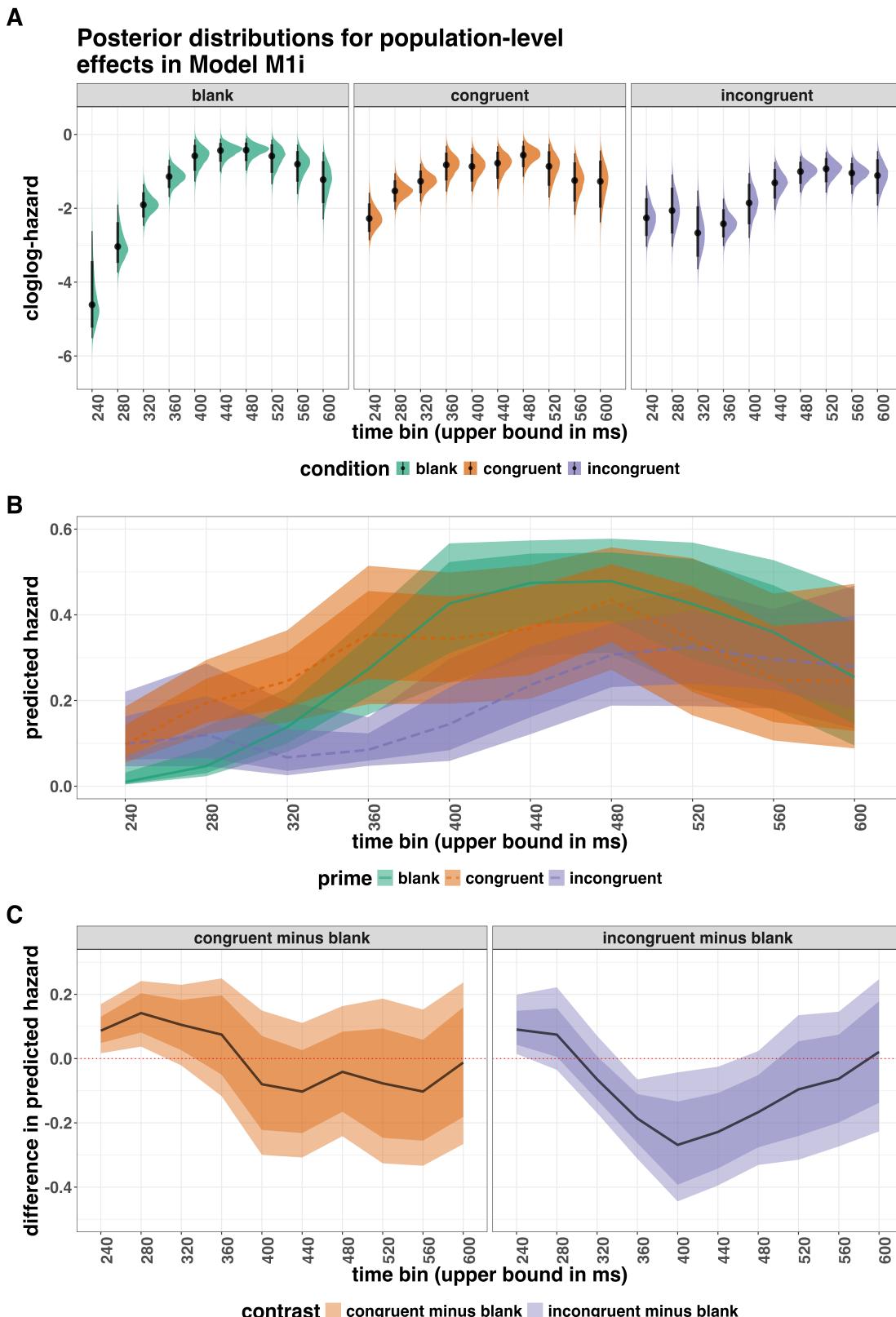


Figure 3. Discrete-time hazard modeling results at the population level. (A) Medians and 80/95% credible intervals of the posterior distributions of the population-level parameters of model M1i. (B) Point (median) and 80/95% credible interval summaries of the hazard estimates (expected values of the draws from the posterior predictive distributions) in each time bin. (C) Point (mean) and 80/95% credible interval summaries of estimated differences in hazard in each time bin.

Because the parameter estimates are on the cloglog-hazard scale, we can ease our interpretation by plotting the expected value of the posterior predictive distribution – the predicted hazard values – at the population level (Figure 3B). As we are actually interested in the effects of congruent and incongruent primes, relative to the blank prime condition, we can construct two contrasts (congruent-blank, incongruent-blank), and plot the posterior distributions of these contrast effects at the population level (Figure 3C). The point estimates and quantile intervals can also be reported in a table (see Tutorial_2a.Rmd for details).

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

If we want to estimate the population-level effect of prime type on hazard, we can base our conclusion on the credible Intervals (CrIs) in Figure 3C. The contrast “congruent minus blank” was estimated to be 0.09 hazard units in bin 240 (95% CrI = [0.02, 0.17]), and 0.14 hazard units in bin 280 (95% CrI = [0.04, 0.25]). For the other bins, the 95% credible interval contained zero. The contrast “incongruent minus blank” was estimated to be 0.09 hazard units in bin 240 (95% CrI = [0.01, 0.21]), -0.19 hazard units in bin 360 (95% CrI = [-0.31, -0.06]), -0.27 hazard units in bin 400 (95% CrI = [-0.45, -0.04]), and -0.23 hazard units in bin 440 (95% CrI = [-0.40, -0.03]). For the other bins, the 95% credible interval contained zero.

There are thus two phases of performance for the average person between 200 and 600 ms after target onset. In the first phase, the addition of a congruent or incongruent prime stimulus increases the hazard of response occurrence compared to blank prime trials in time bin 240. In the second phase, only the incongruent prime decreases the hazard of response occurrence compared to blank primes, in the time period 320-440 ms. The sign of

527 the effect of incongruent primes on the hazard of response occurrence thus depends on
528 how much waiting time has passed since target onset. Future modeling efforts could
529 incorporate the trial number into the model formula, in order to also study how the effects
530 of prime type on hazard change on the long experiment-wide time scale, next to the short
531 trial-wide time scale. In Tutorial_2a.Rmd we provide a number of model formulae that
532 should get you going.

533 **3.4 Tutorial 2b: Fitting Bayesian conditional accuracy models**

534 In this fourth tutorial, we illustrate how to fit a Bayesian multilevel regression model
535 to the timed accuracy data from the masked response priming data used in Tutorial 1a.
536 The general process is similar to Tutorial 2a, except that (a) we use the person-trial data,
537 (b) we use the symmetric logit link function, and (c) we change the priors (our prior belief
538 is that conditional accuracy values between 0 and 1 are equally likely). To keep the tutorial
539 short, we only fit one conditional accuracy model, which was based on model M1i from
540 Tutorial 2a and labelled M1i_ca.

541 To make inferences from the parameter estimates in model M1i_ca, we first plot the
542 densities of the draws from the posterior distributions of its population-level parameters in
543 Figure 4A, together with point (median) and interval estimates (80% and 95% credible
544 intervals).

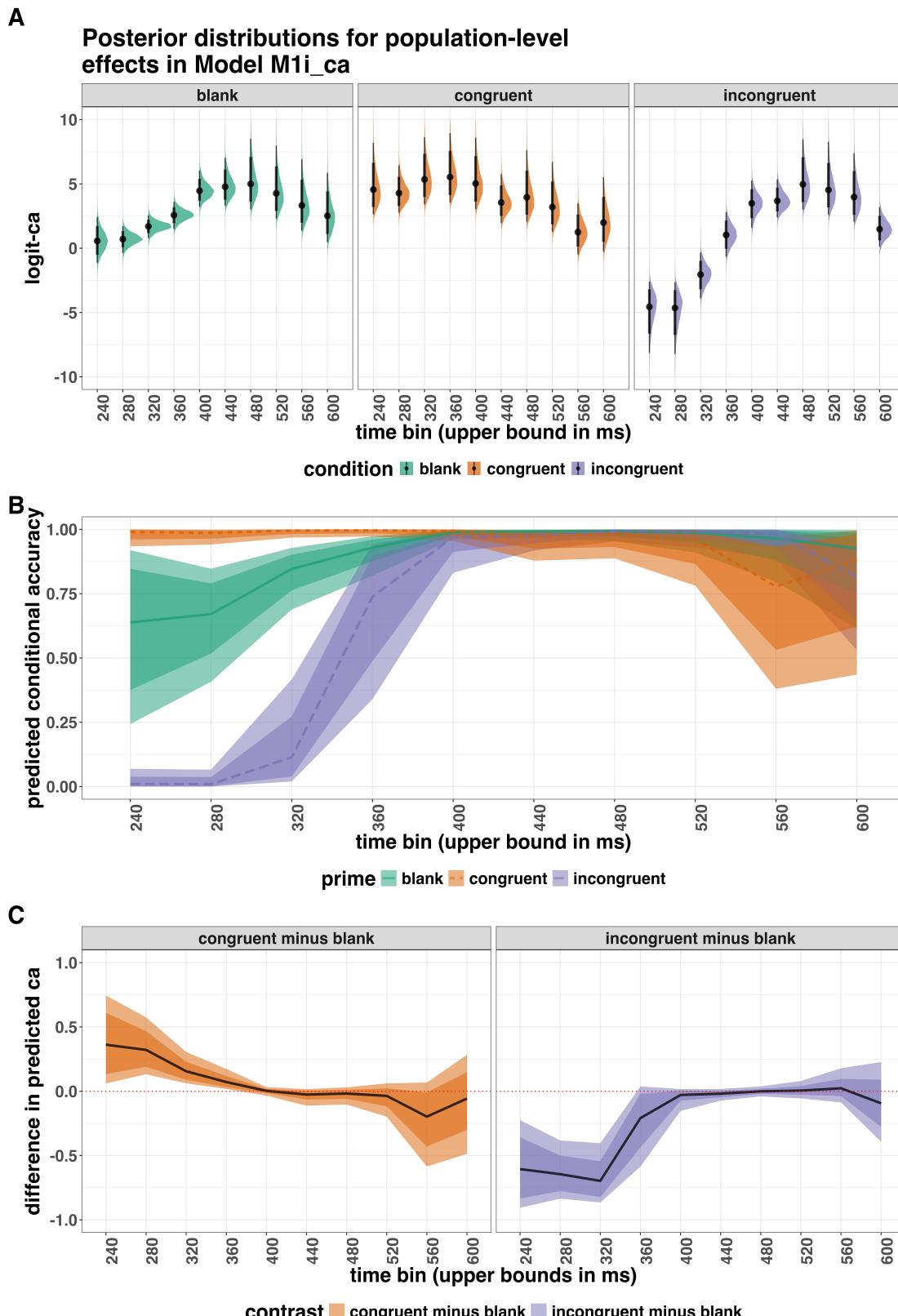


Figure 4. Conditional accuracy modeling results at the population level. (A) Medians and 80/95% credible intervals of the posterior distributions of the population-level parameters of model M1i_ca. (B) Point (median) and 80/95% credible interval summaries of the conditional accuracy (ca) estimates (expected values of the draws from the posterior predictive distributions) in each time bin. (C) Point (mean) and 80/95% credible interval summaries of estimated differences in conditional accuracy in each time bin.

545 Because the parameter estimates are on the logit-ca scale, we can ease our
546 interpretation by plotting the expected value of the posterior predictive distribution – the
547 predicted conditional accurcies – at the population level (Figure 4B). As we are actually
548 interested in the effects of congruent and incongruent primes, relative to the blank prime
549 condition, we can construct two contrasts (congruent-blank, incongruent-blank), and plot
550 the posterior distributions of these contrast effects at the population level (Figure 4C).

551 Based on Figure 4C we see that on the population level congruent primes have a positive
552 effect on the conditional accuracy of emitted responses in time bins 240, 280, 320, and 360,
553 relative to the estimates in the baseline condition (blank prime; red dashed lines in Figure
554 4C). Incongruent primes have a negative effect on the conditional accuracy of emitted
555 responses in the first three time bins, relative to blank primes.

556 Finally, because many researchers will be more familiar with frequentist statistics, we
557 also provide code to fit hazard and conditional accuracy models in the frequentist
558 framework in Tutorial_3a.Rmd and Tutorial_3b.Rmd, using the R package lme4() (Bates
559 et al., 2015).

560 3.5 Tutorial 4: Planning

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

562 **3.5.1 Background.** The general approach to planning that we adopt here involves
563 simulating reasonably structured data to help guide what you might be able to expect from
564 your data once you collect it (Gelman, Vehtari, et al., 2020). The basic structure and code
565 follows the examples outlined by Solomon Kurz in his ‘power’ blog posts
566 (<https://solomonkurz.netlify.app/blog/bayesian-power-analysis-part-i/>) and Lisa
567 Debruine’s R package faux{} (<https://debruine.github.io/faux/>), as well as these related
568 papers (DeBruine & Barr, 2021; Pargent, Koch, Kleine, Lermer, & Gaube, 2024).

569 **3.5.2 Basic workflow.** The basic workflow is as follows:

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

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

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

583 **3.5.3 Fit a regression model and simulate one dataset.** We again use the
 584 data from Panis and Schmidt (2016) to provide a worked example. We fit an index coding
 585 model on a subset of time bins (six time bins in total) and for two prime conditions
 586 (congruent and incongruent). We chose to focus on a subsample of the data to ease the
 587 computational burden. We also used a full varying effects structure, with the model
 588 formula as follows:

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

589 We then took parameters from this model and used them to create a single dataset
 590 with 200 trials per condition for 10 individual participants. The raw data and the
 591 simulated data are plotted in Figure 5 and show quite close correspondence, which is
 592 re-assuring. But, this is only one dataset. What we really want to do is simulate many
 593 datasets and vary parameters of interest, which is what we turn to in the next section.

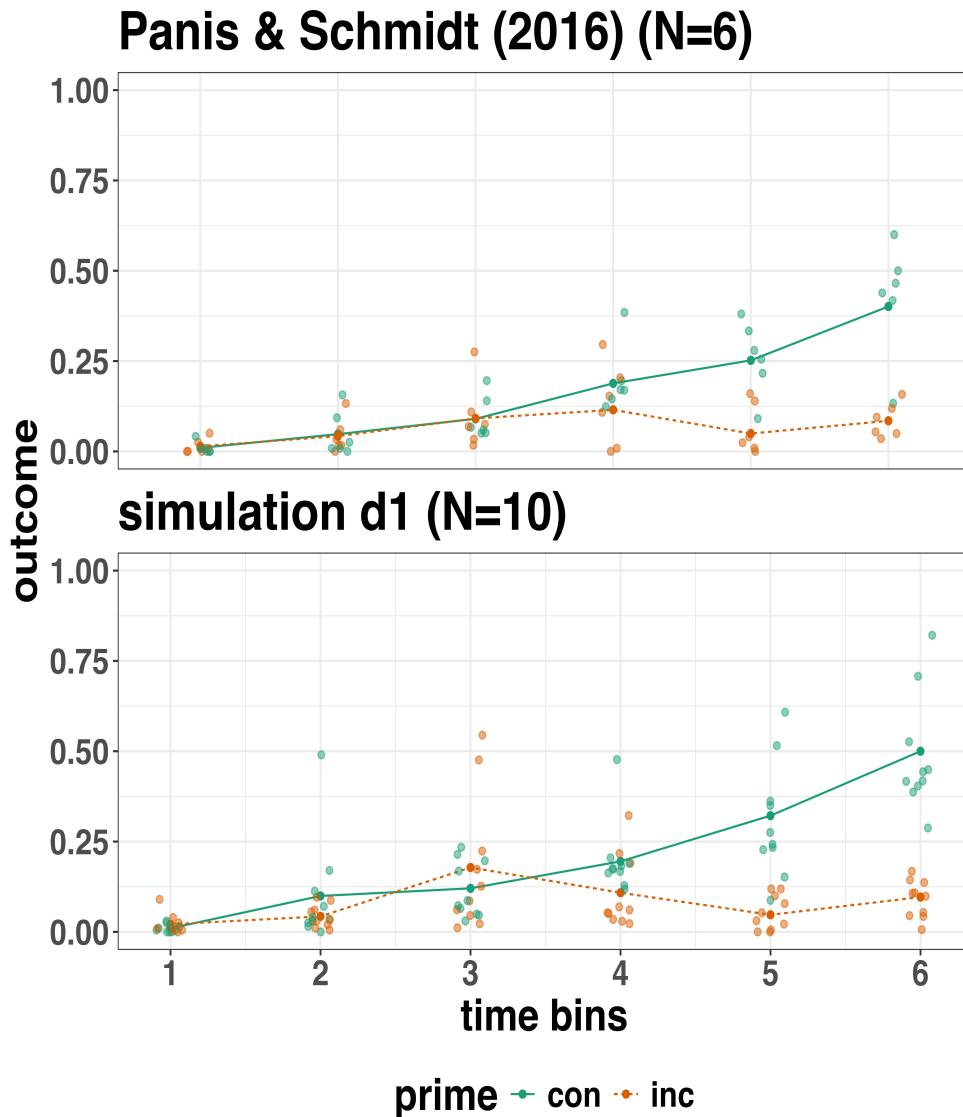


Figure 5. Raw data from Panis and Schmidt (2016) and simulated data from 10 participants.

594 3.5.4 Simulate and summarise data across a range of parameter values.

595 Here we use the same data simulation process as used above, but instead of simulating one
 596 dataset, we simulate 1000 datasets per variation in parameter values. Specifically, in
 597 Simulation 1, we vary the number of trials per condition (100, 200, and 400), as well as the
 598 effect size in bin 6. We focus on bin 6 only, in terms of varying the effect size, just to make
 599 things simpler and easier to understand. The effect size observed in bin 6 in this subsample
 600 of data was a 79% reduction in hazard value from the congruent prime (0.401 hazard

601 value) to the incongruent prime condition (0.085 hazard value). In other words, a hazard
602 ratio of 0.21 (e.g., $0.085/0.401 = 0.21$). As a starting point, we chose three effect sizes,
603 which covered a fairly broad range of hazard ratios (0.25, 0.5, 0.75), which correspond to a
604 75%, 50% and 25% reduction in hazard value as a function of prime condition.

605 Summary results from Simulation 1 are shown in Figure 6A. Figure 6A depicts
606 statistical “power” as calculated by the percentage of lower-bound 95% confidence intervals
607 that exclude zero when the difference between prime condition is calculated (congruent -
608 incongruent). In other words, we calculate the fraction of simulated datasets that
609 generated an effect of prime that excludes the criterion mark of zero. We are aware that
610 “power” is not part of a Bayesian analytical workflow, but we choose to include it here, as
611 it is familiar to most researchers in experimental psychology.

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

620 In Simulation 2, we varied the effect size between a different range of values (0.5, 0.4,
621 0.3), which correspond to a 50%, 60% and 70% reduction in hazard value as a function of
622 prime condition. In addition, we varied the number of participants per experiment between
623 10, 15, and 20 participants. Given that trial count per condition made little difference to
624 power in Simulation 1, we fixed trial count at 200 trials per condition in Simulation 2.
625 Summary results from Simulation 2 are shown in Figure 6B. A summary of these power
626 calculations might be as follows (trial count = 200 per condition in all cases):

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

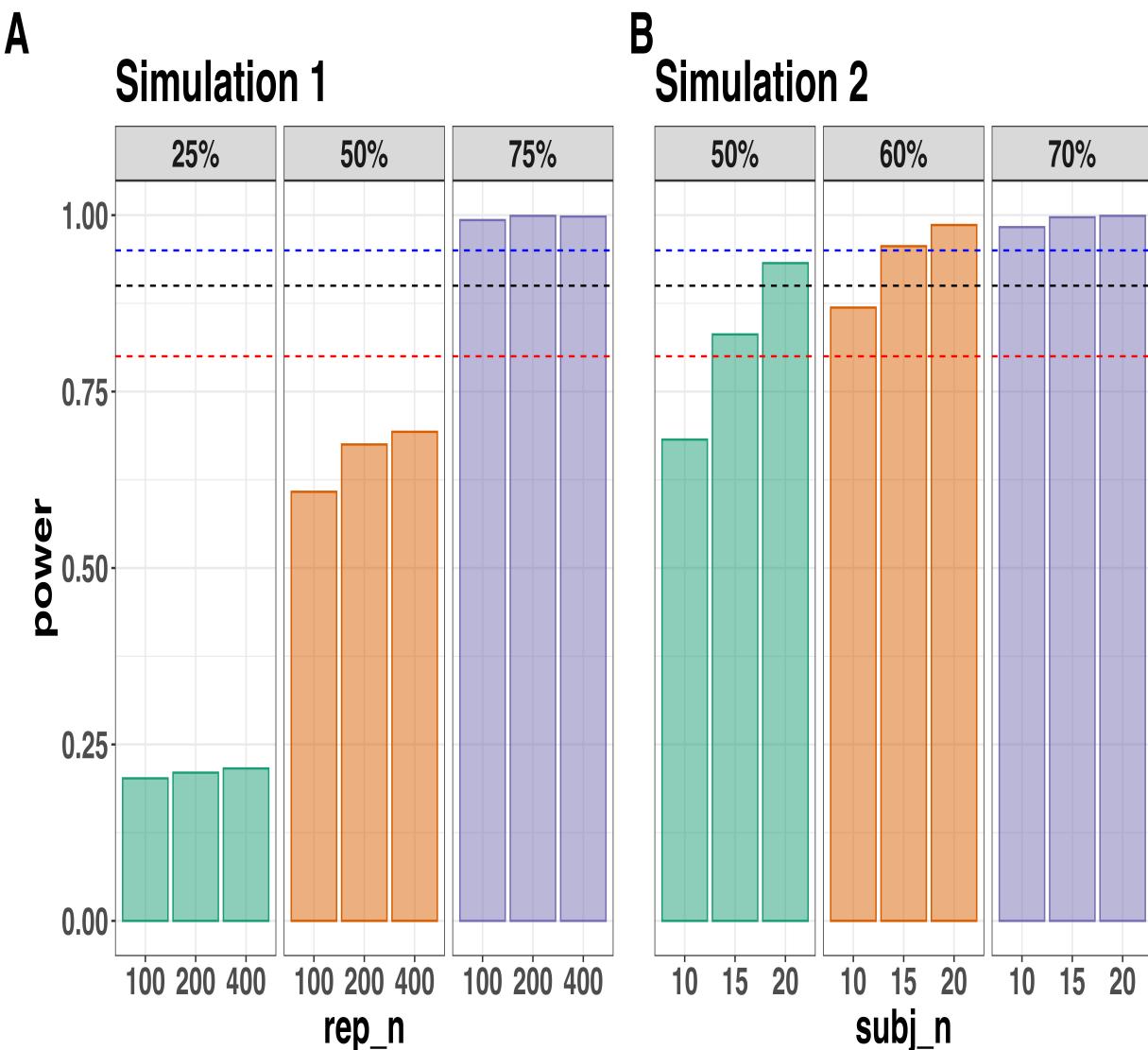


Figure 6. Statistical power across data Simulation 1 (A) and Simulation 2 (B). Power was calculated as the percentage of lower-bound 95% confidence intervals that exclude zero when the difference between prime condition is calculated (congruent - incongruent). In Simulation 1, the effect size was varied 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.

630 **3.5.5 Planning decisions.** Now that we have summarised our simulated data,

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

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

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

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

635 Some considerations might be as follows:

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

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

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

639 personnel?

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

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

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

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

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

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

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

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

648 But, and this is an important caveat, unless there are unavoidable reasons, no matter

649 what kind of planning choices we made based on these data simulations, we would not

650 solely rely on data collected from one single study. Instead, we would run a follow-up

651 experiment that replicates and extends the initial result. By doing so, we would aim to

652 avoid the Cult of the Isolated Single Study (Nelder, 1999; Tong, 2019), and thus reduce the

653 reliance on any one type of planning tool, such as a power analysis. Then, we would look

654 for common patterns across two or more experiments, rather than trying to make the case

655 that a single study on its own has sufficient evidential value to hit some criterion mark.

656

4. Discussion

657 This main motivation for writing this paper is the observation that EHA and SAT
658 analysis remain under-used in psychological research. As a consequence, the field of
659 psychological research is not taking full advantage of the many benefits EHA/SAT provides
660 compared to more conventional analyses. By providing a freely available set of tutorials,
661 which provide step-by-step guidelines and ready-to-use R code, we hope that researchers
662 will feel more comfortable using EHA/SAT in the future. Indeed, we hope that our
663 tutorials may help to overcome a barrier to entry with EHA/SAT, which is that such
664 approaches require more analytical complexity compared to standard approaches. While
665 we have focused here on within-subject, factorial, small- N designs, it is important to realize
666 that EHA/SAT can be applied to other designs as well (large- N designs with only one
667 measurement per subject, between-subject designs, etc.). As such, the general workflow
668 and associated code can be modified and applied more broadly to other contexts and
669 research questions. In the following, we discuss the main use-cases, issues relating to model
670 complexity and interpretability, as well as limitations of the approach.

671 **4.1 What are the main use-cases of EHA for understanding cognition and brain
672 function?**

673 For those researchers, like ourselves, who are primarily interested in understanding
674 human cognitive and brain systems, we consider two broadly-defined, main use-cases of
675 EHA. First, as we hope to have made clear by this point, EHA is one way to investigating
676 a “temporal states” approach to cognitive processes, by tracking behavior as a function of
677 step-wise increases in absolute waiting time. EHA thus provides a way to uncover the
678 microgenesis of cognitive effects, by revealing when cognitive states may start and stop,
679 how states are replaced with others, as well as what they may be tied to or interact with.
680 Therefore, if your research questions concern **when psychological states occur, and**

681 **how they are temporally organized**, our EHA tutorials could be useful tools to use for
682 basic knowledge development, as well as theory building.

683 Second, even if you are not primarily interested in studying the temporal organization
684 of cognitive states, EHA could still be a useful tool to consider using, in order to qualify
685 inferences that are being made based on comparisons between means. Given that distinctly
686 different inferences can be made from the same data based on whether one computes a
687 mean across trials or a RT distribution of events (Figure 1), it may be important for
688 researchers to supplement comparisons between means with EHA. For instance, EHA
689 might reveal that the conclusion of interest based on averaging across trials does not apply
690 to all responses, but is instead restricted to certain periods of within-trial time.

691 4.2 Model complexity versus interpretability

692 Hazard and conditional accuracy models can quickly become very complex when
693 adding more than one time scale, due to the many possible higher-order interactions. For
694 example, some of the models discussed in Tutorial 2a, which we did not focus on in the
695 main text, contain two time scales as covariates: the passage of time on the within-trial
696 time scale, and the passage of time on the across-trial (or within-experiment) time scale.
697 However, when trials are presented in blocks, and blocks of trials within sessions, and when
698 the experiment comprises a number of sessions, then four time scales can be defined
699 (within-trial, within-block, within-session, and within-experiment). From a theoretical
700 perspective, adding more than one time scale – and their interactions – can be important
701 to capture plasticity and other learning effects that may play out on such longer time
702 scales, and that are probably present in each experiment in general (Schöner & Spencer,
703 2016). From a practical perspective, therefore, some choices need to be made to balance
704 the amount of data that is being collected per participant, condition and across the varying
705 timescales. As one example, if there are several timescales of relevance, then it might be
706 prudent for interpretational purposes to limit the number of experimental predictor

707 variables (conditions). This is of course where planning and data simulation efforts would
708 be important to provide a guide to experimental design choices (see Tutorial 4 and section
709 2.3).

710 **4.3 Limitations**

711 Compared to the orthodox method – comparing means between conditions – the
712 most important limitation of multilevel hazard and conditional accuracy modeling is that it
713 might take a long time to estimate the parameters using Bayesian methods or the model
714 might have to be simplified significantly to use frequentist methods. Relatedly, as these
715 models can be quite complex in terms of the number of possible parameters, more thought
716 is required at the model specification and model building stages.

717 Another issue is that you need a relatively large number of trials per condition to
718 estimate the discrete-time hazard function with relatively high temporal resolution (e.g., 20
719 ms), which is required when testing predictions of process models of cognition. Indeed, in
720 general, there is a trade-off between the number of trials per condition and the temporal
721 resolution (i.e., bin width) of the discrete-time hazard function. Therefore, we recommend
722 researchers to collect as many trials as possible per experimental condition, given the
723 available resources and considering the participant experience (e.g., fatigue and boredom).
724 For instance, if the maximum session length deemed reasonable is between 1 and 2 hours,
725 what is the maximum number of trials per condition that you could reasonably collect?
726 After consideration, it might be worth conducting multiple testing sessions per participant
727 and/or reducing the number of experimental conditions. There is a user-friendly online tool
728 for calculating statistical power as a function of the number of trials as well as the number
729 of participants, and this might be worth consulting to guide the research design process
730 (Baker et al., 2021). Finally, if you have a lot of repeated measurements per condition per
731 participant, you can of course also try continuous-time methods (Allison, 2010; Elmer et
732 al., 2023).

733

5. Conclusions

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

743

Author contributions

744 Conceptualization: S. Panis and R. Ramsey; Software: S. Panis and R. Ramsey;

745 Writing - Original Draft Preparation: S. Panis; Writing - Review & Editing: S. Panis and

746 R. Ramsey; Supervision: R. Ramsey.

747

Conflicts of Interest

748 The author(s) declare that there were no conflicts of interest with respect to the

749 authorship or the publication of this article.

750

Prior versions

751 All of the submitted manuscript and Supplemental Material was previously posted to

752 a preprint archive: <https://doi.org/10.31234/osf.io/57bh6>

753

Supplemental Material

754

Disclosures

755 **Data, materials, and online resources**

756 Link to public archive:

757 https://github.com/sven-panis/Tutorial_Event_History_Analysis

758 Supplemental Material: Panis_Ramsey_suppl_material_R1.pdf

759 **Ethical approval**

760 Ethical approval was not required for this tutorial in which we reanalyze existing

761 data sets.

762

References

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

- 789 *methods*. Retrieved from <https://Matrix.R-forge.R-project.org>
- 790 Bengtsson, H. (2021). futures: A unifying framework for parallel and distributed
791 processing in r using futures. *The R Journal*, 13(2), 208–227.
792 <https://doi.org/10.32614/RJ-2021-048>
- 793 Berger, A., & Kiefer, M. (2021). Comparison of Different Response Time Outlier Exclusion
794 Methods: A Simulation Study. *Frontiers in Psychology*, 12, 675558.
795 <https://doi.org/10.3389/fpsyg.2021.675558>
- 796 Blossfeld, H.-P., & Rohwer, G. (2002). *Techniques of event history modeling: New
797 approaches to causal analysis*, 2nd ed (pp. x, 310). Mahwah, NJ, US: Lawrence
798 Erlbaum Associates Publishers.
- 799 Bloxom, B. (1984). Estimating response time hazard functions: An exposition and
800 extension. *Journal of Mathematical Psychology*, 28(4), 401–420.
801 [https://doi.org/10.1016/0022-2496\(84\)90008-7](https://doi.org/10.1016/0022-2496(84)90008-7)
- 802 Bolger, N., Zee, K. S., Rossignac-Milon, M., & Hassin, R. R. (2019). Causal processes in
803 psychology are heterogeneous. *Journal of Experimental Psychology: General*, 148(4),
804 601–618. <https://doi.org/10.1037/xge0000558>
- 805 Box-Steffensmeier, J. M. (2004). Event history modeling: A guide for social scientists.
806 Cambridge: University Press.
- 807 Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan.
808 *Journal of Statistical Software*, 80(1), 1–28. <https://doi.org/10.18637/jss.v080.i01>
- 809 Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms.
810 *The R Journal*, 10(1), 395–411. <https://doi.org/10.32614/RJ-2018-017>
- 811 Bürkner, P.-C. (2021). Bayesian item response modeling in R with brms and Stan. *Journal
812 of Statistical Software*, 100(5), 1–54. <https://doi.org/10.18637/jss.v100.i05>
- 813 DeBruine, L. M., & Barr, D. J. (2021). Understanding Mixed-Effects Models Through
814 Data Simulation. *Advances in Methods and Practices in Psychological Science*, 4(1),
815 2515245920965119. <https://doi.org/10.1177/2515245920965119>

- 816 Eddelbuettel, D., & Balamuta, J. J. (2018). Extending R with C++: A Brief Introduction
817 to Rcpp. *The American Statistician*, 72(1), 28–36.
818 <https://doi.org/10.1080/00031305.2017.1375990>
- 819 Eddelbuettel, D., & François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal
820 of Statistical Software*, 40(8), 1–18. <https://doi.org/10.18637/jss.v040.i08>
- 821 Elmer, T., Van Duijn, M. A. J., Ram, N., & Bringmann, L. F. (2023). Modeling
822 categorical time-to-event data: The example of social interaction dynamics captured
823 with event-contingent experience sampling methods. *Psychological Methods*.
824 <https://doi.org/10.1037/met0000598>
- 825 Frank, M. C., Braginsky, M., Cachia, J., Coles, N. A., Hardwicke, T. E., Hawkins, R. D.,
826 ... Williams, R. (2025). *Experimentology: An Open Science Approach to Experimental
827 Psychology Methods*. Stanford University. <https://doi.org/10.25936/3JP6-5M50>
- 828 Gabry, J., Češnovar, R., Johnson, A., & Broder, S. (2024). *Cmdstanr: R interface to
829 'CmdStan'*. Retrieved from <https://github.com/stan-dev/cmdstanr>
- 830 Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2019). Visualization
831 in bayesian workflow. *J. R. Stat. Soc. A*, 182, 389–402.
832 <https://doi.org/10.1111/rssa.12378>
- 833 Gelman, A., Hill, J., & Vehtari, A. (2020). Regression and Other Stories.
834 [https://www.cambridge.org/highereducation/books/regression-and-other-
835 stories/DD20DD6C9057118581076E54E40C372C](https://www.cambridge.org/highereducation/books/regression-and-other-stories/DD20DD6C9057118581076E54E40C372C); Cambridge University Press.
836 <https://doi.org/10.1017/9781139161879>
- 837 Gelman, A., Vehtari, A., Simpson, D., Margossian, C. C., Carpenter, B., Yao, Y., ...
838 Modrák, M. (2020). *Bayesian Workflow*. arXiv.
839 <https://doi.org/10.48550/arXiv.2011.01808>
- 840 Girard, J. (2024). *Standist: What the package does (one line, title case)*. Retrieved from
841 <https://github.com/jmgirard/standist>
- 842 Grolemund, G., & Wickham, H. (2011). Dates and times made easy with lubridate.

- 843 *Journal of Statistical Software*, 40(3), 1–25. Retrieved from
844 <https://www.jstatsoft.org/v40/i03/>
- 845 Heiss, A. (2021, November 10). A Guide to Correctly Calculating Posterior Predictions
846 and Average Marginal Effects with Multilevel Bayesian Models.
847 <https://doi.org/10.59350/wbn93-edb02>
- 848 Holden, J. G., Van Orden, G. C., & Turvey, M. T. (2009). Dispersion of response times
849 reveals cognitive dynamics. *Psychological Review*, 116(2), 318–342.
850 <https://doi.org/10.1037/a0014849>
- 851 Hosmer, D. W., Lemeshow, S., & May, S. (2011). *Applied Survival Analysis: Regression*
852 *Modeling of Time to Event Data* (2nd ed). Hoboken: John Wiley & Sons.
- 853 Kantowitz, B. H., & Pachella, R. G. (2021). The Interpretation of Reaction Time in
854 Information-Processing Research 1. *Human Information Processing*, 41–82.
855 <https://doi.org/10.4324/9781003176688-2>
- 856 Kay, M. (2024). *tidybayes: Tidy data and geoms for Bayesian models*.
857 <https://doi.org/10.5281/zenodo.1308151>
- 858 Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing,
859 estimation, meta-analysis, and power analysis from a Bayesian perspective.
860 *Psychonomic Bulletin & Review*, 25(1), 178–206.
861 <https://doi.org/10.3758/s13423-016-1221-4>
- 862 Kurz, A. S. (2023a). *Applied longitudinal data analysis in brms and the tidyverse* (version
863 0.0.3). Retrieved from <https://bookdown.org/content/4253/>
- 864 Kurz, A. S. (2023b). *Statistical rethinking with brms, ggplot2, and the tidyverse: Second*
865 *edition* (version 0.4.0). Retrieved from <https://bookdown.org/content/4857/>
- 866 Lakens, D. (2022). Sample Size Justification. *Collabra: Psychology*, 8(1), 33267.
867 <https://doi.org/10.1525/collabra.33267>
- 868 Landes, J., Engelhardt, S. C., & Pelletier, F. (2020). An introduction to event history
869 analyses for ecologists. *Ecosphere*, 11(10), e03238. <https://doi.org/10.1002/ecs2.3238>

- 870 Lougheed, J. P., Benson, L., Cole, P. M., & Ram, N. (2019). Multilevel survival analysis:
871 Studying the timing of children's recurring behaviors. *Developmental Psychology*,
872 55(1), 53–65. <https://doi.org/10.1037/dev0000619>
- 873 Luce, R. D. (1991). *Response times: Their role in inferring elementary mental organization*
874 (1. issued as paperback). Oxford: Univ. Press.
- 875 McElreath, R. (2020). *Statistical Rethinking: A Bayesian Course with Examples in R and*
876 *STAN* (2nd ed.). New York: Chapman and Hall/CRC.
877 <https://doi.org/10.1201/9780429029608>
- 878 Mills, M. (2011). *Introducing Survival and Event History Analysis*. 1 Oliver's Yard, 55 City
879 Road, London EC1Y 1SP United Kingdom: SAGE Publications Ltd.
880 <https://doi.org/10.4135/9781446268360>
- 881 Müller, K., & Wickham, H. (2023). *Tibble: Simple data frames*. Retrieved from
882 <https://CRAN.R-project.org/package=tibble>
- 883 Nelder, J. A. (1999). From Statistics to Statistical Science. *Journal of the Royal Statistical*
884 *Society. Series D (The Statistician)*, 48(2), 257–269. Retrieved from
885 <https://www.jstor.org/stable/2681191>
- 886 Neuwirth, E. (2022). *RColorBrewer: ColorBrewer palettes*. Retrieved from
887 <https://CRAN.R-project.org/package=RColorBrewer>
- 888 Panis, S. (2020). How can we learn what attention is? Response gating via multiple direct
889 routes kept in check by inhibitory control processes. *Open Psychology*, 2(1), 238–279.
890 <https://doi.org/10.1515/psych-2020-0107>
- 891 Panis, S., Moran, R., Wolkersdorfer, M. P., & Schmidt, T. (2020). Studying the dynamics
892 of visual search behavior using RT hazard and micro-level speed–accuracy tradeoff
893 functions: A role for recurrent object recognition and cognitive control processes.
894 *Attention, Perception, & Psychophysics*, 82(2), 689–714.
895 <https://doi.org/10.3758/s13414-019-01897-z>
- 896 Panis, S., Schmidt, F., Wolkersdorfer, M. P., & Schmidt, T. (2020). Analyzing Response

- 897 Times and Other Types of Time-to-Event Data Using Event History Analysis: A Tool
898 for Mental Chronometry and Cognitive Psychophysiology. *I-Perception*, 11(6),
899 2041669520978673. <https://doi.org/10.1177/2041669520978673>
- 900 Panis, S., & Schmidt, T. (2016). What Is Shaping RT and Accuracy Distributions? Active
901 and Selective Response Inhibition Causes the Negative Compatibility Effect. *Journal of*
902 *Cognitive Neuroscience*, 28(11), 1651–1671. https://doi.org/10.1162/jocn_a_00998
- 903 Panis, S., & Schmidt, T. (2022). When does “inhibition of return” occur in spatial cueing
904 tasks? Temporally disentangling multiple cue-triggered effects using response history
905 and conditional accuracy analyses. *Open Psychology*, 4(1), 84–114.
906 <https://doi.org/10.1515/psych-2022-0005>
- 907 Panis, S., Torfs, K., Gillebert, C. R., Wagemans, J., & Humphreys, G. W. (2017).
908 Neuropsychological evidence for the temporal dynamics of category-specific naming.
909 *Visual Cognition*, 25(1-3), 79–99. <https://doi.org/10.1080/13506285.2017.1330790>
- 910 Panis, S., & Wagemans, J. (2009). Time-course contingencies in perceptual organization
911 and identification of fragmented object outlines. *Journal of Experimental Psychology:*
912 *Human Perception and Performance*, 35(3), 661–687.
913 <https://doi.org/10.1037/a0013547>
- 914 Pargent, F., Koch, T. K., Kleine, A.-K., Lermer, E., & Gaube, S. (2024). A Tutorial on
915 Tailored Simulation-Based Sample-Size Planning for Experimental Designs With
916 Generalized Linear Mixed Models. *Advances in Methods and Practices in Psychological*
917 *Science*, 7(4), 25152459241287132. <https://doi.org/10.1177/25152459241287132>
- 918 Pedersen, T. L. (2024). *Patchwork: The composer of plots*. Retrieved from
919 <https://patchwork.data-imaginist.com>
- 920 Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in s and s-PLUS*. New York:
921 Springer. <https://doi.org/10.1007/b98882>
- 922 R Core Team. (2024). *R: A language and environment for statistical computing*. Vienna,
923 Austria: R Foundation for Statistical Computing. Retrieved from

- 924 https://www.R-project.org/
- 925 Schöner, G., & Spencer, J. P. (2016). *Dynamic thinking: A primer on dynamic field theory*.
926 New York, NY: Oxford University Press.
- 927 https://doi.org/10.1093/acprof:oso/9780199300563.001.0001
- 928 Singer, J. D., & Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling
929 Change and Event Occurrence*. Oxford, New York: Oxford University Press.
- 930 Smith, P. L., & Little, D. R. (2018). Small is beautiful: In defense of the small-N design.
931 *Psychonomic Bulletin & Review*, 25(6), 2083–2101.
932 https://doi.org/10.3758/s13423-018-1451-8
- 933 Stan Development Team. (2020). *StanHeaders: Headers for the R interface to Stan*.
934 Retrieved from https://mc-stan.org/
- 935 Stan Development Team. (2024). *RStan: The R interface to Stan*. Retrieved from
936 https://mc-stan.org/
- 937 Stoolmiller, M. (2015). *An Introduction to Using Multivariate Multilevel Survival Analysis
938 to Study Coercive Family Process* (Vol. 1; T. J. Dishion & J. Snyder, Eds.). Oxford
939 University Press. https://doi.org/10.1093/oxfordhb/9780199324552.013.27
- 940 Stoolmiller, M., & Snyder, J. (2006). Modeling heterogeneity in social interaction processes
941 using multilevel survival analysis. *Psychological Methods*, 11(2), 164–177.
942 https://doi.org/10.1037/1082-989X.11.2.164
- 943 Teachman, J. D. (1983). Analyzing social processes: Life tables and proportional hazards
944 models. *Social Science Research*, 12(3), 263–301.
945 https://doi.org/10.1016/0049-089X(83)90015-7
- 946 Tong, C. (2019). Statistical Inference Enables Bad Science; Statistical Thinking Enables
947 Good Science. *The American Statistician*, 73(sup1), 246–261.
948 https://doi.org/10.1080/00031305.2018.1518264
- 949 Townsend, J. T. (1990). Truth and consequences of ordinal differences in statistical
950 distributions: Toward a theory of hierarchical inference. *Psychological Bulletin*, 108(3),

- 951 551–567. <https://doi.org/10.1037/0033-2909.108.3.551>
- 952 Van Zandt, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin & Review*, 7(3), 424–465. <https://doi.org/10.3758/BF03214357>
- 953 Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *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)
- 954 Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from <https://ggplot2.tidyverse.org>
- 955 Wickham, H. (2023a). *Forcats: Tools for working with categorical variables (factors)*. Retrieved from <https://forcats.tidyverse.org/>
- 956 Wickham, H. (2023b). *Stringr: Simple, consistent wrappers for common string operations*. Retrieved from <https://stringr.tidyverse.org>
- 957 Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., ... Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- 958 Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *Dplyr: A grammar of data manipulation*. Retrieved from <https://dplyr.tidyverse.org>
- 959 Wickham, H., & Henry, L. (2023). *Purrr: Functional programming tools*. Retrieved from [https://purrr.tidyverse.org/](https://purrr.tidyverse.org)
- 960 Wickham, H., Hester, J., & Bryan, J. (2024). *Readr: Read rectangular text data*. Retrieved from <https://readr.tidyverse.org>
- 961 Wickham, H., Vaughan, D., & Girlich, M. (2024). *Tidyr: Tidy messy data*. Retrieved from <https://tidyr.tidyverse.org>
- 962 Winter, B. (2019). *Statistics for Linguists: An Introduction Using R*. New York: Routledge. <https://doi.org/10.4324/9781315165547>
- 963 Wolkersdorfer, M. P., Panis, S., & Schmidt, T. (2020). Temporal dynamics of sequential motor activation in a dual-prime paradigm: Insights from conditional accuracy and hazard functions. *Attention, Perception, & Psychophysics*, 82(5), 2581–2602.

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