

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

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

14

Abstract

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

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

39 Word count: X

40

1. Introduction

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

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

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

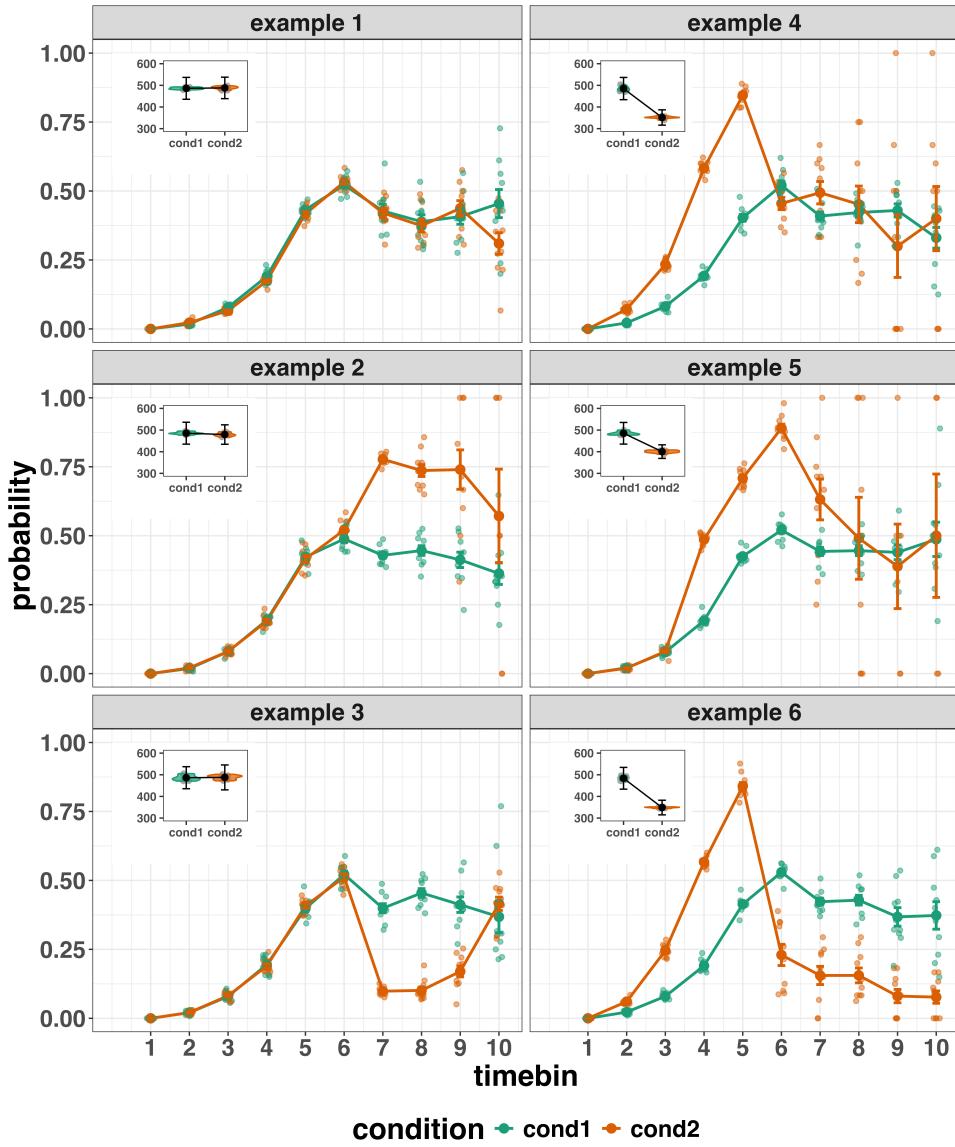


Figure 1. Means versus distributional shapes for six different simulated data set examples. The first second after stimulus onset is divided in ten bins of 100 ms. Timebin indicates the bin rank. The first bin is (0,100], the last bin is (900,1000]. For our purposes here, it is enough to know that the distributions plotted represent the probability of an event occurring in that timebin, given that it has not yet occurred. Insets show mean response time per condition.

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

79 From a historical perspective, it is worth noting that the development of analytical
80 tools that can estimate or predict whether and when events will occur is not a new
81 endeavour. Indeed, hundreds of years ago, analytical methods were developed to predict the
82 duration of time until people died (e.g., Halley, 1997; Makeham, William M., 1860). The
83 same logic has been applied to psychological time-to-event data, as previously demonstrated
84 (Panis, Schmidt, et al., 2020).

85 1.2 Aims and structure of the paper

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

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

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

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

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

114 In Tutorial 2a, we illustrate how one can fit Bayesian multi-level regression models to
115 RT data using the R package brms. We also perform prior predictive checks, compare

models, and interpret the plots of the predicted hazard functions for the selected model, and the posterior distributions of our contrasts of interest. In Tutorial 2b we fit Bayesian multi-level regression models to *timed* accuracy data to perform a micro-level speed-accuracy tradeoff (SAT) analysis, which complements the EHA of RT data for choice RT data.

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

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

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

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

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

148 **2.1 Basic features of event history analysis**

149 To apply EHA, one must be able to:

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

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

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

and the discrete-time survivor function (Figure 2). The equations that define both of these functions are reported in part A of the supplementary material. The discrete-time hazard function gives you, for each time bin, the probability that the event occurs (sometime) in bin t , given that the event does not occur in previous bins. In other words, it reflects the instantaneous likelihood that the event occurs in the current bin, given that it has not yet occurred in the past, i.e., in one of the prior bins. In contrast, the discrete-time survivor function cumulates the bin-by-bin risks of event *nonoccurrence* to obtain the survival probability, the probability that the event occurs after bin t . In other words, the survivor function gives you for each time bin the likelihood that the event occurs in the future, i.e., in one of the subsequent timebins.

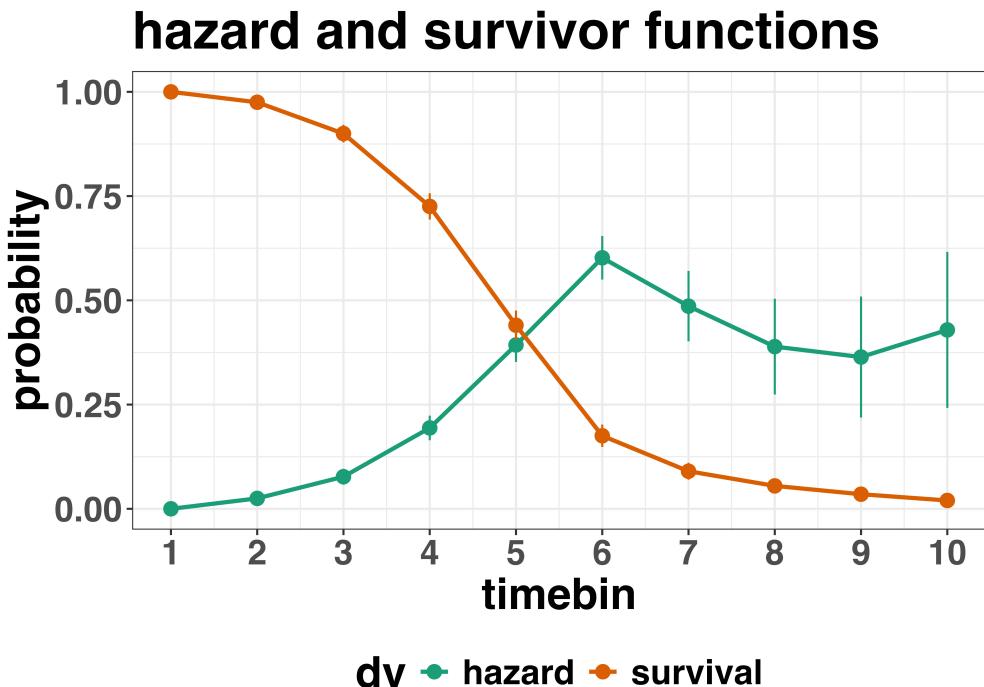


Figure 2. Discrete-time hazard and survivor functions. Discrete time-to-event data were simulated for 200 trials of 1 experimental condition. While the hazard function is the vehicle for inferring the time course of cognitive processes, the survival probability $S(t-1)$ can help to qualify or provide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

173 2.2 Benefits of event history analysis

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

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

198 2.3 Event history analysis in the context of experimental psychology

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

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

202 **2.3.1 A worked example.** In the context of experimental psychology, it is common

203 for participants to be presented with either a 1-button detection task or a discrimination

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

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

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

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

²³⁰ represent the overall ability of cognition to overcome interference, on average, across trials.
²³¹ For instance, if mean-average RT was higher in incongruent than congruent trials, one may
²³² conclude that cognitive mechanisms that support interference control are working as
²³³ expected across trials, and are indexed by each recorded response. But such a conclusion is
²³⁴ not supported when the effects are explored over a timeline. Instead, the psychological
²³⁵ conclusion is much more nuanced and suggests that multiple states start, stop and possibly
²³⁶ interact over a particular temporal window.

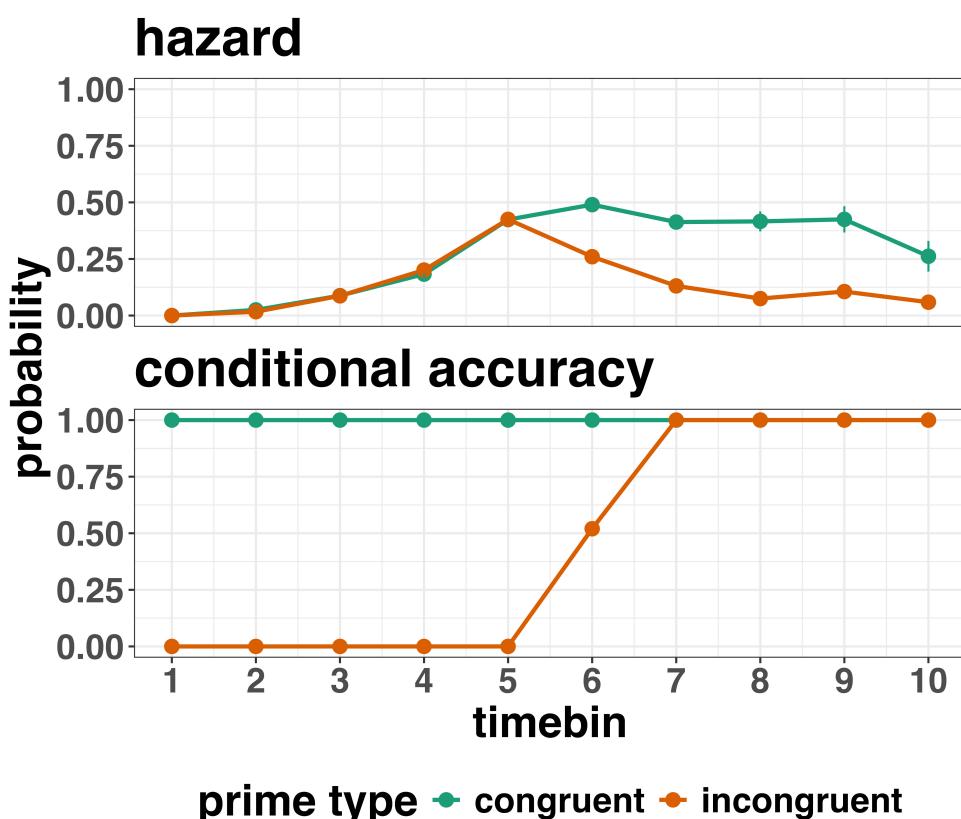


Figure 3. Discrete-time hazard and conditional accuracy functions. Discrete time-to-event and conditional accuracy data were simulated for 2000 trials for each of two priming conditions (congruent and incongruent prime stimuli). Bin width equals 100 ms.

²³⁷ Unlocking the temporal states of cognitive processes can be revealing for theory
²³⁸ development and the understanding of basic psychological processes. Possibly more

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

246 **2.3.2 Implications for designing experiments.** Performing EHA in experimental
247 psychology has implications for how experiments are designed. Indeed, if trials are
248 categorised as a function of when responses occur, then each timebin will only include a
249 subset of the total number of trials. For example, let's consider an experiment where each
250 participant performs 2 conditions and there are 100 trial repetitions per condition. Those
251 100 trials must be distributed in some manner across the chosen number of bins.

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

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

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

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

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

273 **3.1 Data science workflow and descriptive statistics**

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

280 **3.2 Inferential statistical approach**

281 Our lab adopts a estimation approach to multi-level regression (Kruschke & Liddell,
282 2018; Winter, 2019), which is heavily influenced by the Bayesian framework as suggested by
283 Richard McElreath (Kurz, 2023b; McElreath, 2018). We also use a “keep it maximal”
284 approach to specifying varying (or random) effects (Barr, Levy, Scheepers, & Tily, 2013).
285 This means that wherever possible we include varying intercepts and slopes per participant
286 To make inferences, we use two main approaches. We compare models of different
287 complexity, using information criteria (e.g., WAIC) and cross-validation (e.g., LOO), to
288 evaluate out-of-sample predictive accuracy (McElreath, 2018). We also take the most
289 complex model and evaluate key parameters of interest using point and interval estimates.

290 **3.3 Implementation**

291 We used R (Version 4.4.1; R Core Team, 2024)¹ for all reported analyses. The content
292 of the tutorials, in terms of EHA and multi-level regression modelling, is mainly based on
293 Allison (2010), Singer and Willett (2003), McElreath (2018), Heiss (2021), Kurz (2023a), and
294 Kurz (2023b).

295 **4. Tutorials**

296 Tutorials 1a and 1b show how to calculate and plot the descriptive statistics of
297 EHA/SAT when there are one and two independent variables, respectively. Tutorials 2a and
298 2b illustrate how to use Bayesian multilevel modeling to fit hazard and conditional accuracy
299 models, respectively. Tutorials 3a and 3b show how to implement, respectively, multilevel
300 models for hazard and conditional accuracy in the frequentist framework. Additionally, to
301 further simplify the process for other users, the first two tutorials rely on a set of our own
302 custom functions that make sub-processes easier to automate, such as data wrangling and
303 plotting functions (see part B in the supplemental material for a list of the custom functions).

¹ We, furthermore, used the R-packages *bayesplot* (Version 1.11.1; Gabry, Simpson, Vehtari, Betancourt, & Gelman, 2019), *brms* (Version 2.21.0; Bürkner, 2017, 2018, 2021), *citr* (Version 0.3.2; Aust, 2019), *cmdstanr* (Version 0.8.1.9000; Gabry, Češnovar, Johnson, & Broder, 2024), *dplyr* (Version 1.1.4; Wickham, François, Henry, Müller, & Vaughan, 2023), *forcats* (Version 1.0.0; Wickham, 2023a), *ggplot2* (Version 3.5.1; Wickham, 2016), *lme4* (Version 1.1.35.5; Bates, Mächler, Bolker, & Walker, 2015), *lubridate* (Version 1.9.3; Grolemund & Wickham, 2011), *Matrix* (Version 1.7.0; Bates, Maechler, & Jagan, 2024), *nlme* (Version 3.1.166; Pinheiro & Bates, 2000), *papaja* (Version 0.1.2.9000; Aust & Barth, 2023), *patchwork* (Version 1.2.0; Pedersen, 2024), *purrr* (Version 1.0.2; Wickham & Henry, 2023), *RColorBrewer* (Version 1.1.3; Neuwirth, 2022), *Rcpp* (Eddelbuettel & Balamuta, 2018; Version 1.0.12; Eddelbuettel & François, 2011), *readr* (Version 2.1.5; Wickham, Hester, & Bryan, 2024), *RJ-2021-048* (Bengtsson, 2021), *standist* (Version 0.0.0.9000; Girard, n.d.), *stringr* (Version 1.5.1; Wickham, 2023b), *tibble* (Version 3.2.1; Müller & Wickham, 2023), *tidybayes* (Version 3.0.6; Kay, 2023), *tidyR* (Version 1.3.1; Wickham, Vaughan, & Girlich, 2024), *tidyverse* (Version 2.0.0; Wickham et al., 2019), and *tinylabels* (Version 0.2.4; Barth, 2023).

304 Our list of tutorials is as follows:

- 305 • 1a. Wrangle raw data and calculate descriptive stats for one independent variable
- 306 • 1b. Wrangle raw data and calculate descriptive stats for two independent variables
- 307 • 2a. Bayesian multilevel modeling for $h(t)$
- 308 • 2b. Bayesian multilevel modeling for $ca(t)$
- 309 • 3a. Frequentist multilevel modeling for $h(t)$
- 310 • 3b. Frequentist multilevel modeling for $ca(t)$
- 311 • 4. Simulation and power analysis for planning experiments

312 **4.1 Tutorial 1a: Calculating descriptive statistics using a life table**

313 **4.1.1 Data wrangling aims.** Our data wrangling procedures serve two related

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

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

320 different types of inferential modelling approaches. The two types of data structure we label
321 as ‘person-trial’ data and ‘person-trial-bin’ data. The ‘person-trial’ data (Table 1) will be
322 familiar to most researchers who record behavioural responses from participants, as it
323 represents the measured RT and accuracy per trial within an experiment. This data set is
324 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.

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

333 **4.1.2 A real data wrangling example.** To illustrate how to quickly set up life
 334 tables for calculating the descriptive statistics (functions of discrete time), we use a
 335 published data set on masked response priming from Panis and Schmidt (2016). In their first
 336 experiment, Panis and Schmidt (2016) presented a double arrow for 94 ms that pointed left
 337 or right as the target stimulus with an onset at time point zero in each trial. Participants
 338 had to indicate the direction in which the double arrow pointed using their corresponding
 339 index finger, within 800 ms after target onset. Response time and accuracy were recorded on
 340 each trial. Prime type (blank, congruent, incongruent) and mask type were manipulated.
 341 Here we focus on the subset of trials in which no mask was presented. The 13-ms prime
 342 stimulus was a double arrow presented 187 ms before target onset in the congruent (same

343 direction as target) and incongruent (opposite direction as target) prime conditions.

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

347 The required column names are as follows:

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

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

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

355 Next, we can set up the life tables and plots of the discrete-time functions $h(t)$, $S(t)$,
 356 $ca(t)$, and $P(t)$ – see part A of the supplementary material for their definitions. To do so
 357 using a functional programming approach, one has to nest the data within participants using
 358 the group_nest() function, and supply a user-defined censoring time and bin width to our
 359 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))

```

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

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

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

374 the conditional accuracies using our custom function `calc_ca()`, add the $ca(t)$ estimates to

375 the life table using our custom function `join_lt_ca()`, and then plot the descriptive statistics

376 using our custom function `plot_eha()`. When creating the plots, some warning messages will

377 likely be generated, like these:

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

379 (`geom_line()`).

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

381 (`geom_point()`).

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

383 (`geom_segment()`).

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

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

386 aspects, including the life table for a particular condition of a particular subject, and a plot

387 of the different functions for a particular participant. In general, it is important to visually

388 inspect the functions first for each participant, in order to identify individuals that may be

389 guessing (e.g., a flat conditional accuracy function at .5 indicates that someone is just

390 guessing), outlying individuals, and/or different groups with qualitatively different behavior.

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

392 participant 6.

Table 3

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

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

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

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

³⁹⁴ probability mass functions for each prime condition for participant 6. By using discrete-time
³⁹⁵ hazard functions of event occurrence – in combination with conditional accuracy functions
³⁹⁶ for two-choice tasks – one can provide an unbiased, time-varying, and probabilistic
³⁹⁷ description of the latency and accuracy of responses based on all trials of any data set.

Descriptive stats for subject 6

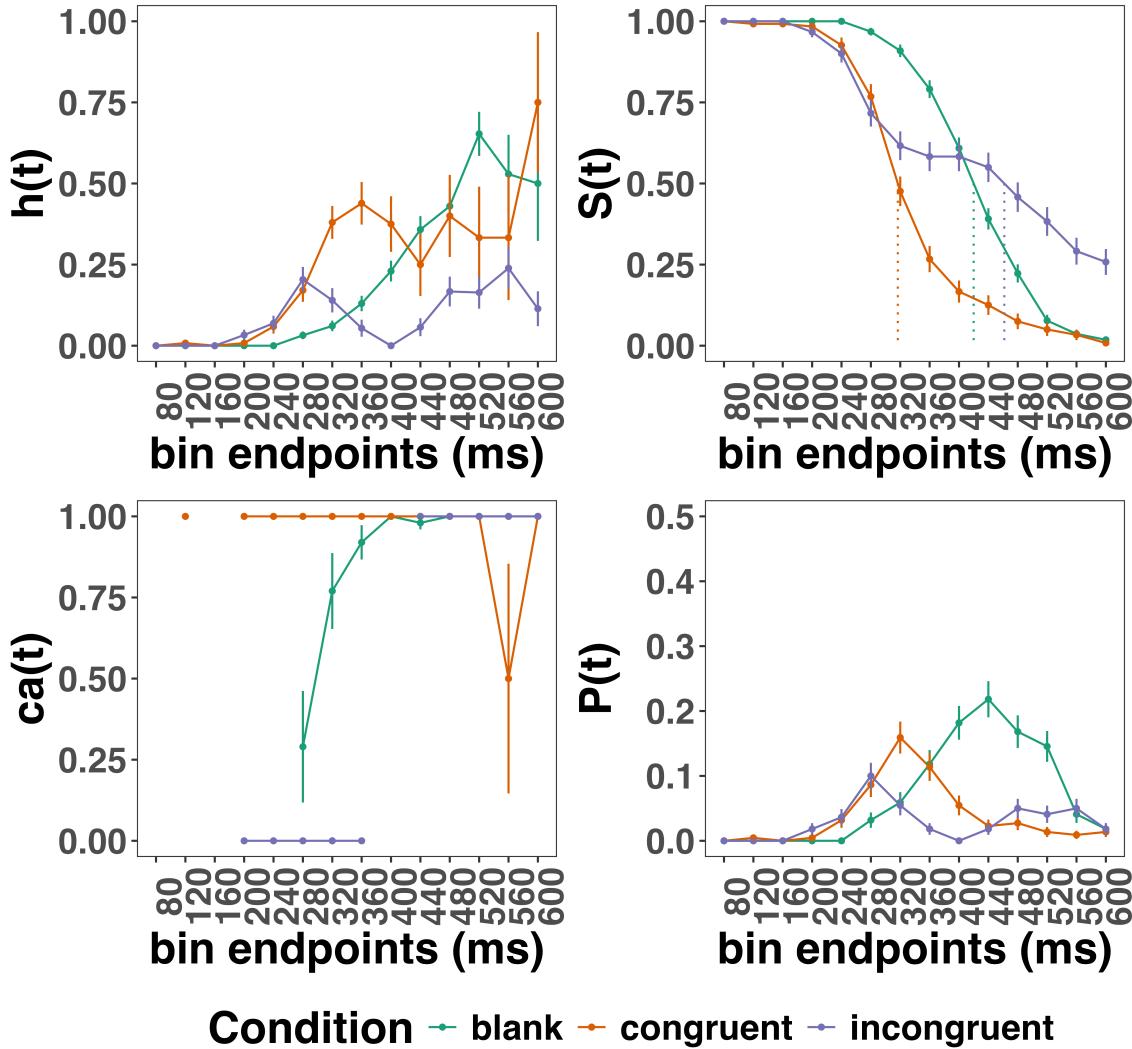


Figure 4. Estimated discrete-time hazard, survivor, probability mass, and conditional accuracy functions for participant 6. Vertical dotted lines indicate the estimated median RTs. Error bars represent +/- 1 Standard Error of the respective proportion.

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

399 0.17, and 0.20 for the blank, congruent, and incongruent prime conditions, respectively. In

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

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

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

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

404 0 for the blank, congruent, and incongruent prime conditions, respectively. In other words, if

405 a response is emitted in bin (240,280], then the probability that it is correct is estimated to

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

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

408 conditional probability of response occurrence in the next 40 ms is estimated to be 0.36, 0.25,

409 and 0.06 for the blank, congruent, and incongruent prime conditions, respectively. And when

410 a response does occur in bin (400,440], then the probability that it is correct is estimated to

411 be 0.98, 1, and 1 for the blank, congruent, and incongruent prime conditions, respectively.

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

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

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

415 responses are fully controlled by the target stimulus. Qualitatively similar results were

416 obtained for the other five participants. When participants show qualitatively the same

417 distributional patterns, one might consider to aggregate their data and make one plot (see

418 Tutorial_1a.Rmd).

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

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

421 distributional data show that early responses are triggered exclusively by the prime stimulus,

422 while only later responses reflect primed responses to the target stimulus.

423 At this point, we have calculated, summarised and plotted descriptive statistics for the
424 key variables in EHA/SAT. As we will show in later Tutorials, statistical models for $h(t)$ and
425 $ca(t)$ can be implemented as generalized linear mixed regression models predicting event
426 occurrence (1/0) and conditional accuracy (1/0) in each bin of a selected time window for
427 analysis. But first we consider calculating the descriptive statistics for two independent
428 variables.

429 **4.2 Tutorial 1b: Generalising to a more complex design**

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

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

441 **4.3 Tutorial 2a: Fitting Bayesian hazard models to discrete time-to-event data**

442 In this third tutorial, we illustrate how to fit Bayesian multi-level regression models to
443 the RT data of the masked response priming data set used in Tutorial 1a. Fitting (Bayesian
444 or non-Bayesian) regression models to time-to-event data is important when you want to
445 study how the shape of the hazard function depends on various predictors (Singer & Willett,
446 2003).

447 4.3.1 Hazard model considerations. There are several analytic decisions one has

448 to make when fitting a discrete-time hazard model. First, one has to select an analysis time

449 window, i.e., a contiguous set of bins for which there is enough data for each participant.

450 Second, given that the dependent variable (event occurrence) is binary, one has to select a

451 link function (see part C in the supplementary material). The cloglog link is preferred over

452 the logit link when events can occur in principle at any time point within a bin, which is the

453 case for RT data (Singer & Willett, 2003). Third, one has to choose whether to treat TIME

454 (i.e., the time bin index t) as a discrete or continuous predictor. And when you treat a

455 variable as a discrete predictor, you can choose between reference coding and index coding.

456 In the case of a large- N design without repeated measurements, the parameters of a

457 discrete-time hazard model can be estimated using standard logistic regression software after

458 expanding the typical person-trial data set into a person-trial-bin data set (Allison, 2010).

459 When there is clustering in the data, as in the case of a small- N design with repeated

460 measurements, the parameters of a discrete-time hazard model can be estimated using

461 population-averaged methods (e.g., Generalized Estimating Equations), and Bayesian or

462 frequentist generalized linear mixed models (Allison, 2010).

463 In general, there are three assumptions one can make or relax when adding

464 experimental predictor variables and other covariates: The linearity assumption for

465 continuous predictors (the effect of a 1 unit change is the same anywhere on the scale), the

466 additivity assumption (predictors do not interact), and the proportionality assumption

467 (predictors do not interact with TIME).

468 In tutorial_2a.Rmd we fit several Bayesian multilevel models (i.e., generalized linear

469 mixed models) that differ in complexity to the person-trial-bin oriented data set that we

470 created in Tutorial 1a. We decided to select the analysis time window (200,600] and the

471 cloglog link. Below, we shortly discuss three of these models. The person-trial-bin data set is

472 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) %>%
    # continuous predictor for TIME named "period_9", centered on bin 9
  mutate(period_9 = period - 9,
        # categorical predictor for TIME named "timebin" with index coding
        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)))
```

4.3.2 Prior distributions.

To get the posterior distribution of each model

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

The middle column of Figure 16 in part E of the supplementary material shows six examples of prior distributions for an intercept on the logit and/or cloglog scales. While a normal distribution with relatively large variance is often used as a weakly informative prior for continuous dependent variables, rows A and B in Figure 16 show that specifying such distributions on the logit and cloglog scales actually leads to rather informative distributions on the original probability scale, as most mass is pushed to probabilities of 0 and 1.

4.3.3 Model M0i: A null model with index coding.

When you do not want to make assumptions about the shape of the hazard function, or its shape is not smooth but 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, and
 488 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 specified.
 492 The specification “0 + …” removes the default intercept in brm(). The fixed effects include
 493 an intercept for each level of timebin. Each of these intercepts is allowed to vary across
 494 individuals (variable pid). We request 2000 samples from the posterior distribution for each
 495 of four chains. Estimating model M0i took about 30 minutes on a MacBook Pro (Sonoma
 496 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; Panis,
 499 Moran, et al., 2020; Panis & Schmidt, 2022; Panis et al., 2017; Panis & Wagemans, 2009). In
 500 the next model, therefore, we use index coding for both TIME (variable “timebin”) and the

501 categorical predictor prime-target-congruency (variable “prime”), so that we get 30 grand
 502 intercepts, one for each combination of timebin level and prime level. Here is the model
 503 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 **4.3.4 Model M1d: A more parsimonious model.** When the shape of the hazard
 506 function is rather smooth, as it is for behavioral RT data, one can fit a more parsimonious
 507 model by treating TIME as a continuous variable, and use a polynomial specification of the
 508 effect of TIME. Thus, if we want to make assumptions about (1) how hazard changes over
 509 TIME in the reference condition (blank prime), and (2) how the effect of congruent and
 510 incongruent primes change over TIME (relax the proportionality assumption), then we can
 511 switch to a dummy coding approach for prime-target congruency (variable “condition”) and
 512 treat TIME as a continuous variable (variable “period_9”).

513 For example, we may assume that hazard can change in a linear + quadratic fashion
 514 over time for a blank prime, and that the effects of congruent and incongruent primes
 515 relative to blank change in a linear + quadratic fashion, and fit the model called “M1d”.
 516 Here is its model formula.

```
event ~ 0 + Intercept + condition*period_9 + condition*period_9_sq +  

          (1 + condition*period_9 + condition*period_9_sq | pid)
```

517 The specification “0 + Intercept + …” removes the default intercept in brm() and
 518 adds an explicit Intercept for which we can set the prior ourselves. The variable
 519 period_9_sq is a squared version of period_9. Note that duplicate terms in the model
 520 formula (e.g., condition) are ignored. Because TIME is centered on bin 9, the Intercept
 521 represents the estimated cloglog-hazard in bin 9 for the blank prime condition. Model M1d
 522 took about 184 minutes to run.

523 **4.3.5 Compare the models.** We can compare the three models using the Widely

524 Applicable Information Criterion (WAIC) and Leave-One-Out (LOO) cross-validation, and

525 look at model weights for both criteria (Kurz, 2023a; McElreath, 2018).

```
model_weights(model_M0i, model_M1i, model_M1d, weights = "loo") %>% round(digits = 2)
```

526 ## model_M0i model_M1i model_M1d

527 ## 0 1 0

```
model_weights(model_M0i, model_M1i, model_M1d, weights = "waic") %>% round(digits = 2)
```

528 ## model_M0i model_M1i model_M1d

529 ## 0 1 0

530 Clearly, both the loo and waic weighting schemes assign a weight of 1 to model M1i,

531 and a weight of 0 to the other two simpler models.

532 **4.3.6 Interpreting model M1i.** To make inferences from the parameter estimates

533 in model M1i, we first plot the densities of the draws from the posterior distributions of its

534 population-level parameters in Figure 5, together with point (median) and interval estimates

535 (80% and 95% credible intervals).

Posterior distributions for population-level effects in Model M1i

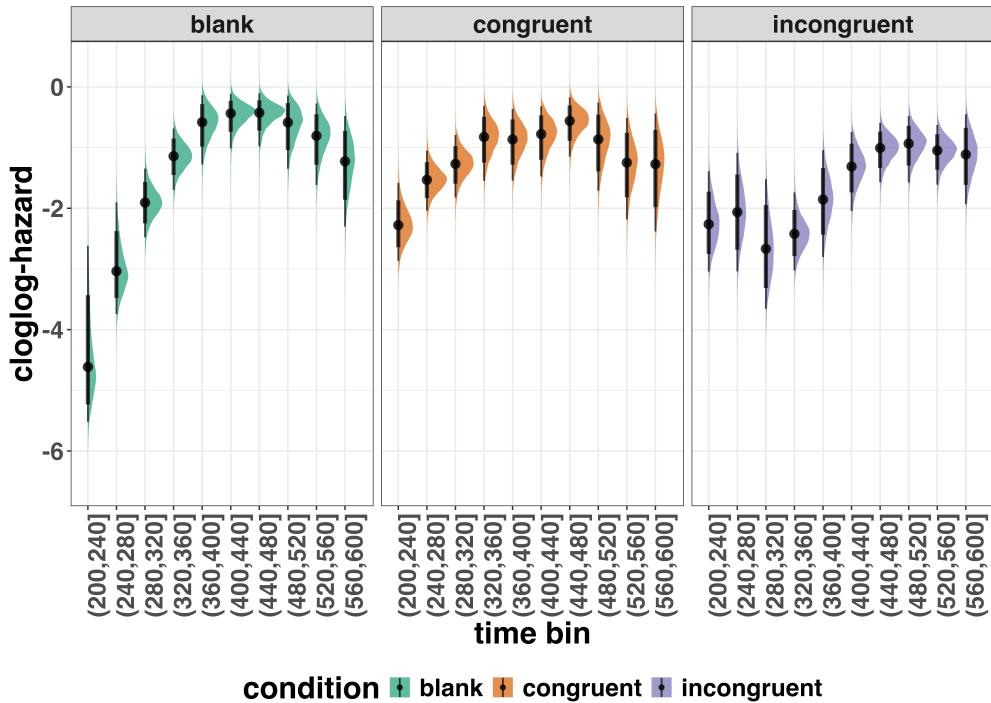


Figure 5. Medians and 80/95% credible intervals of the posterior distributions of the population-level parameters of model M1i.

536 Because the parameter estimates are on the cloglog-hazard scale, we can ease our
 537 interpretation by plotting the expected value of the posterior predictive distribution – the
 538 predicted hazard values – for the average participant (Figure 6), and for each participant in
 539 the data set (Figure 7).

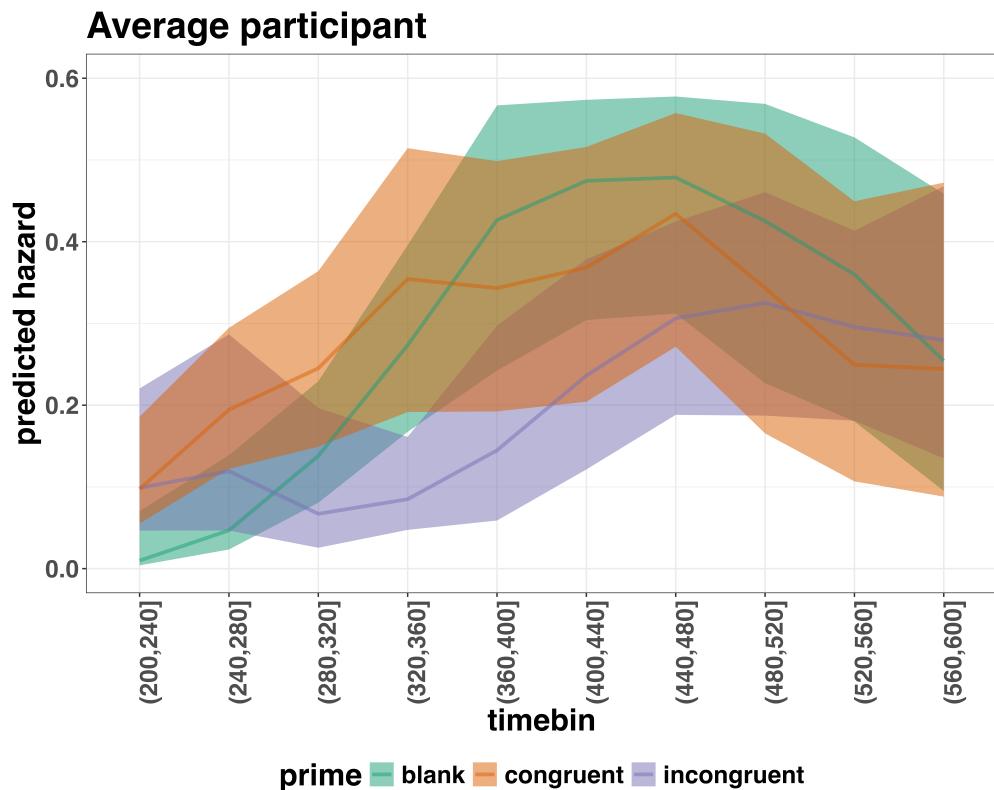


Figure 6. Point (median) and 80/95% credible interval summaries of the hazard estimates (expected values of the draws from the posterior predictive distributions) in each time bin for the average participant.

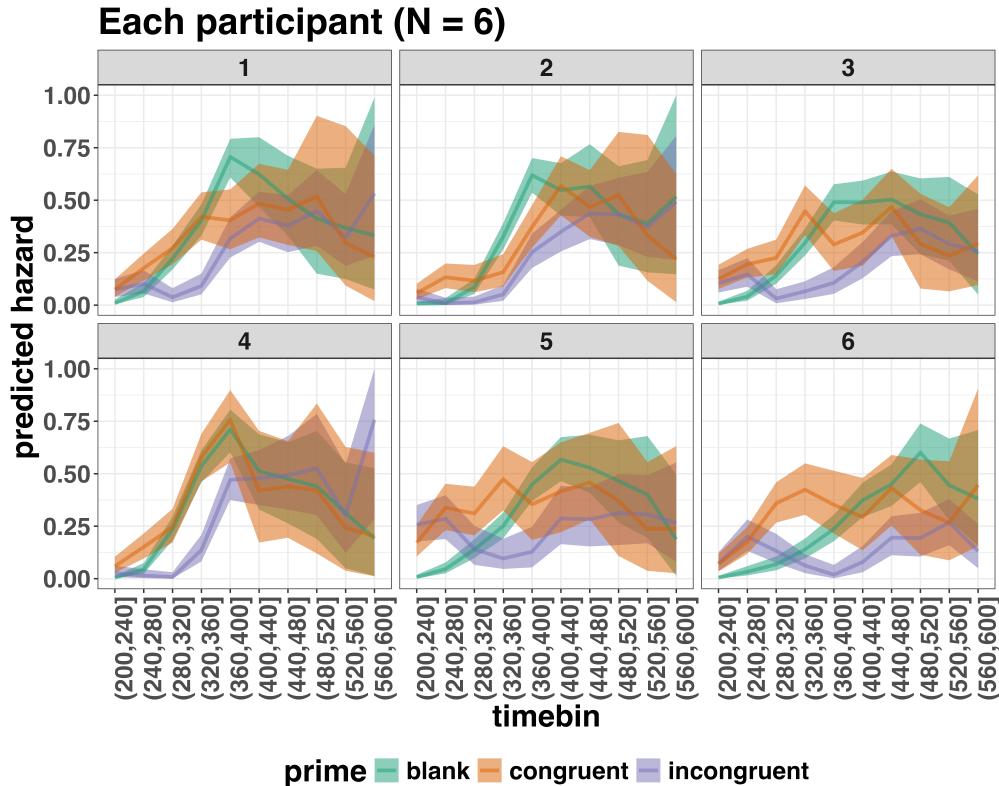


Figure 7. Point (median) and 80/95% credible interval summaries of the hazard estimates (expected values of the draws from the posterior predictive distributions) in each time bin for each participant.

540 As we are actually interested in the effects of congruent and incongruent primes,
 541 relative to the blank prime condition, we can construct two contrasts (congruent-blank,
 542 incongruent-blank), and plot the posterior distributions of these contrast effects, both for the
 543 average participant (Figure 8; grand average marginal effect) and for each participant in the
 544 data set (Figure 9; subject-specific average marginal effect).

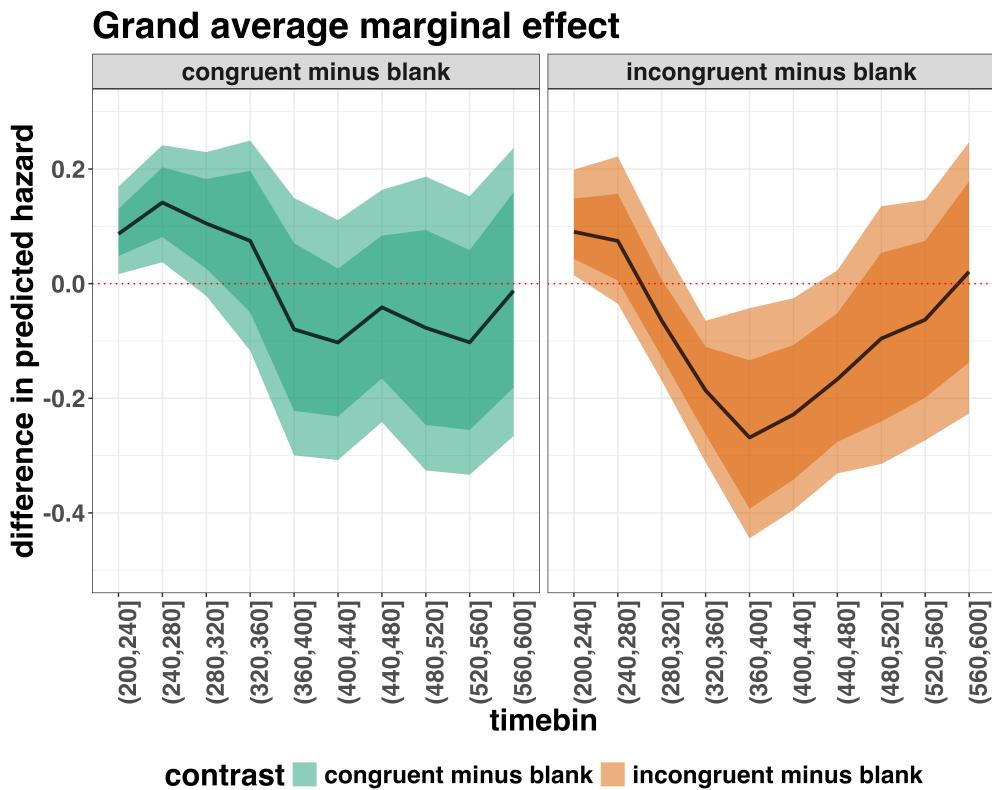


Figure 8. Point (mean) and 80/95% credible interval summaries of estimated differences in hazard in each time bin for the average participant.

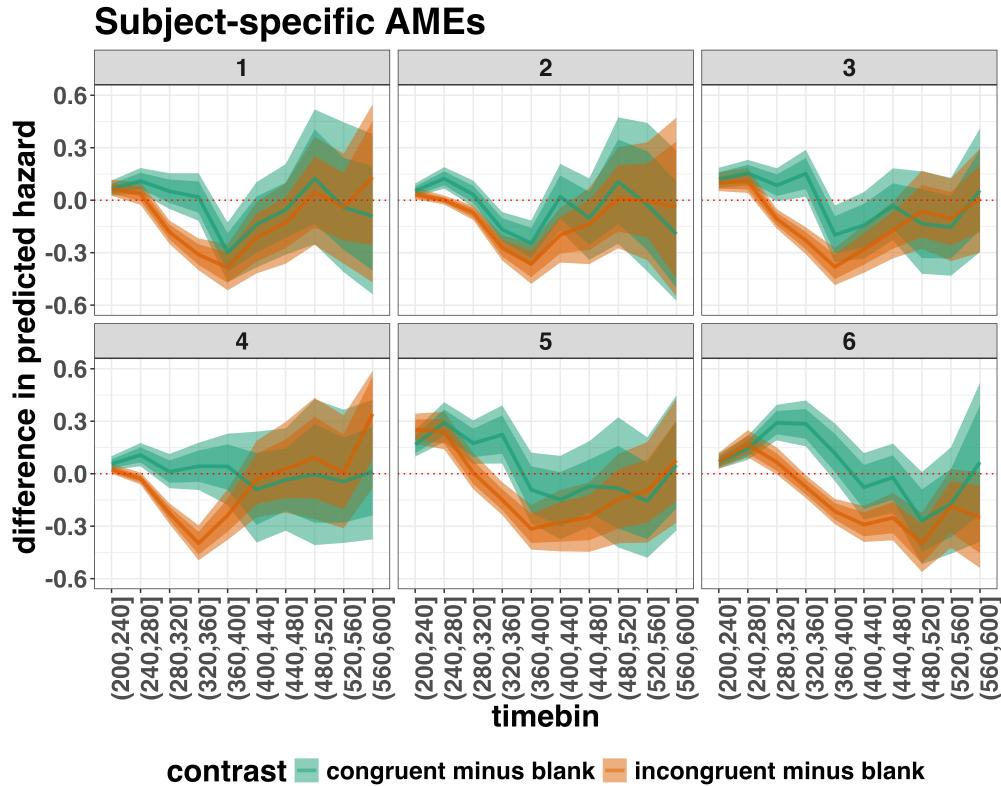


Figure 9. Point (mean) and 80/95% credible interval summaries of estimated differences in hazard in each time bin for each participant.

545 Table 4 shows the summaries of the estimated hazard differences for both contrasts in
 546 terms of a point estimate (the mean) and the upper and lower bounds of the 95% credible
 547 interval, for the average participant.

Table 4

*Point (mean) and 95% credible interval summary of estimated
 differences in hazard, for each time bin and contrast, in the
 average participant.*

contrast	timebin	diff	.lower	.upper
congruent minus blank	6	0.09	0.02	0.17
congruent minus blank	7	0.14	0.04	0.25

Table 4 continued

contrast	timebin	diff	.lower	.upper
congruent minus blank	8	0.11	-0.02	0.24
congruent minus blank	9	0.08	-0.12	0.27
congruent minus blank	10	-0.08	-0.30	0.15
congruent minus blank	11	-0.10	-0.31	0.11
congruent minus blank	12	-0.04	-0.24	0.17
congruent minus blank	13	-0.08	-0.33	0.20
congruent minus blank	14	-0.10	-0.33	0.15
congruent minus blank	15	-0.01	-0.27	0.27
incongruent minus blank	6	0.09	0.01	0.21
incongruent minus blank	7	0.08	-0.03	0.23
incongruent minus blank	8	-0.06	-0.17	0.07
incongruent minus blank	9	-0.19	-0.31	-0.06
incongruent minus blank	10	-0.27	-0.45	-0.04
incongruent minus blank	11	-0.23	-0.40	-0.03
incongruent minus blank	12	-0.17	-0.33	0.02
incongruent minus blank	13	-0.10	-0.31	0.14
incongruent minus blank	14	-0.06	-0.27	0.15
incongruent minus blank	15	0.03	-0.23	0.27

Note. diff = difference in predicted hazard.

548

549 ***Example conclusions for M1i.*** What can we conclude from model M1i about our
 550 research question, i.e., the temporal dynamics of the effect of prime-target congruency on
 551 RT? In other words, in which of the 40-ms time bins between 200 and 600 ms after target

552 onset does changing the prime from blank to congruent or incongruent affect the hazard of
553 response occurrence (for a prime-target SOA of 187 ms)?

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

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

571 The posterior distribution of each contrast can also be summarized by considering its
572 proportion below or above some value, like zero. Table 5 shows the proportion of the
573 posterior distribution below or above zero, for each time bin and contrast.

Table 5

Summarizing the posterior distributions of each contrast by their proportion below and above zero.

timebin	contrast	prop_above	prop_below
6	congruent minus blank	0.99	0.01
7	congruent minus blank	0.99	0.01
8	congruent minus blank	0.95	0.05
9	congruent minus blank	0.79	0.21
10	congruent minus blank	0.24	0.76
11	congruent minus blank	0.15	0.85
12	congruent minus blank	0.33	0.67
13	congruent minus blank	0.28	0.72
14	congruent minus blank	0.19	0.81
15	congruent minus blank	0.47	0.53
6	incongruent minus blank	0.98	0.02
7	incongruent minus blank	0.92	0.08
8	incongruent minus blank	0.12	0.88
9	incongruent minus blank	0.00	1.00
10	incongruent minus blank	0.01	0.99
11	incongruent minus blank	0.02	0.98
12	incongruent minus blank	0.04	0.96
13	incongruent minus blank	0.20	0.80
14	incongruent minus blank	0.27	0.73
15	incongruent minus blank	0.58	0.42

Note. prop_below = proportion below zero; prop_above = proportion above zero.

574

575 Thus, the probability that the contrast “congruent minus blank” is larger than 0, is
576 larger than .9 in bins 6 to 8. And the probability that the contrast “incongruent minus
577 blank” is smaller than 0, is larger than .9 in bins 9 to 12.

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

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

593 Interestingly, all subjects show a tendency in their mean difference (congruent minus
594 blank) to “dip” around that time (Figure 9). Therefore, future modeling efforts could
595 incorporate the trial number into the model formula, in order to also study how the effects of
596 prime type on hazard change on the long experiment-wide time scale, next to the short
597 trial-wide time scale. In Tutorial_2a.Rmd we provide a number of model formula that
598 should get you going.

⁵⁹⁹ **4.4 Tutorial 2b: Fitting Bayesian conditional accuracy models**

⁶⁰⁰ In this fourth tutorial, we illustrate how to fit a Bayesian multi-level regression model
⁶⁰¹ to the timed accuracy data from the masked response priming data set used in Tutorial 1a.
⁶⁰² The general process is similar to Tutorial 2a, except that (a) we use the person-trial data set,
⁶⁰³ (b) we use the logit link function, and (c) we change the priors. To keep the tutorial short,
⁶⁰⁴ we only fitted the effects of model M1i (see Tutorial 2a) in the conditional accuracy model
⁶⁰⁵ called M1i_ca.

⁶⁰⁶ To make inferences from the parameter estimates in model M1i_ca, we first plot the
⁶⁰⁷ densities of the draws from the posterior distributions of its population-level parameters in
⁶⁰⁸ Figure 10, together with point (median) and interval estimates (80% and 95% credible
⁶⁰⁹ intervals).

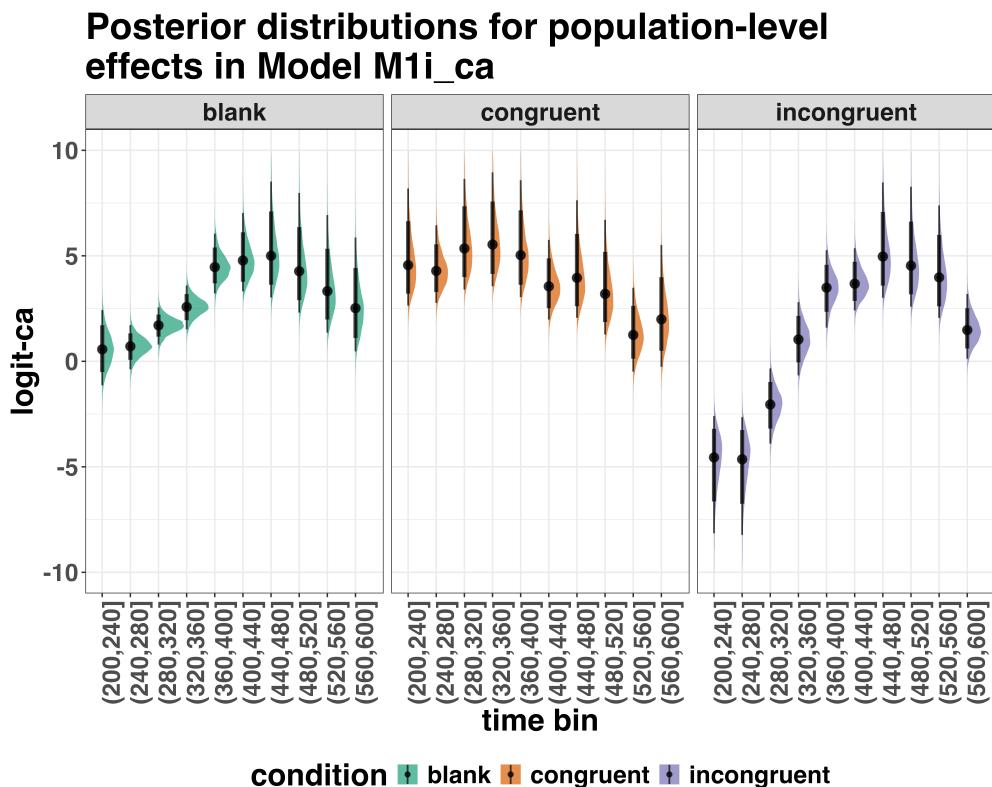


Figure 10. Medians and 80/95% credible intervals of the posterior distributions of the population-level parameters of model M1i_ca.

Because the parameter estimates are on the logit-ca scale, we can ease our interpretation by plotting the expected value of the posterior predictive distribution – the predicted conditional accuracies – for the average participant (Figure 11), and for each participant in the data set (Figure 12).

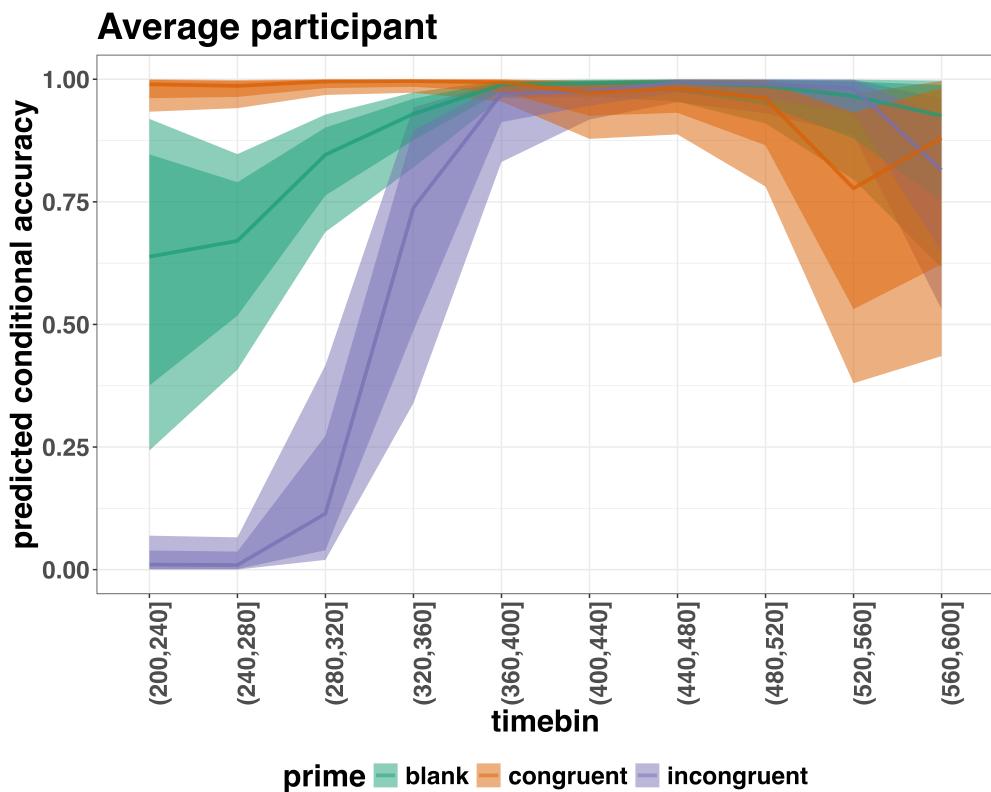


Figure 11. Point (median) and 80/95% credible interval summaries of the conditional accuracy estimates (expected values of the draws from the posterior predictive distributions) in each time bin for the average participant.

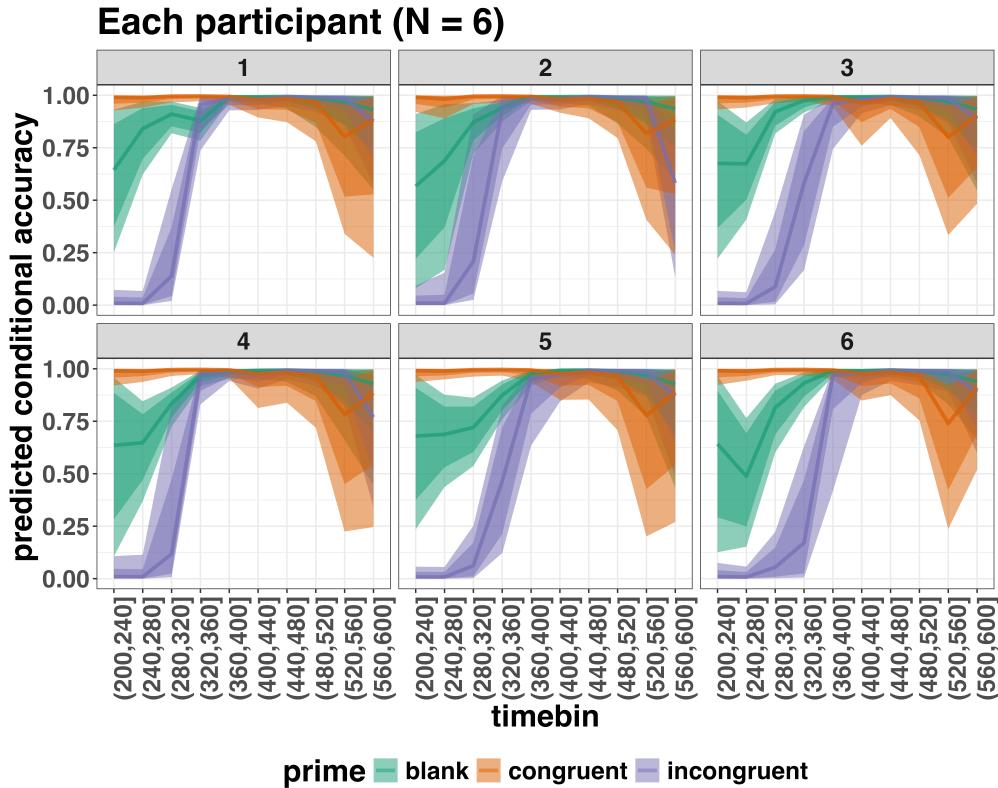


Figure 12. Point (median) and 80/95% credible interval summaries of the conditional accuracy estimates (expected values of the draws from the posterior predictive distributions) in each time bin for each participant.

614 As we are actually interested in the effects of congruent and incongruent primes,
 615 relative to the blank prime condition, we can construct two contrasts (congruent-blank,
 616 incongruent-blank), and plot the posterior distributions of these contrast effects for the
 617 average participant (Figure 13; grand average marginal effect).

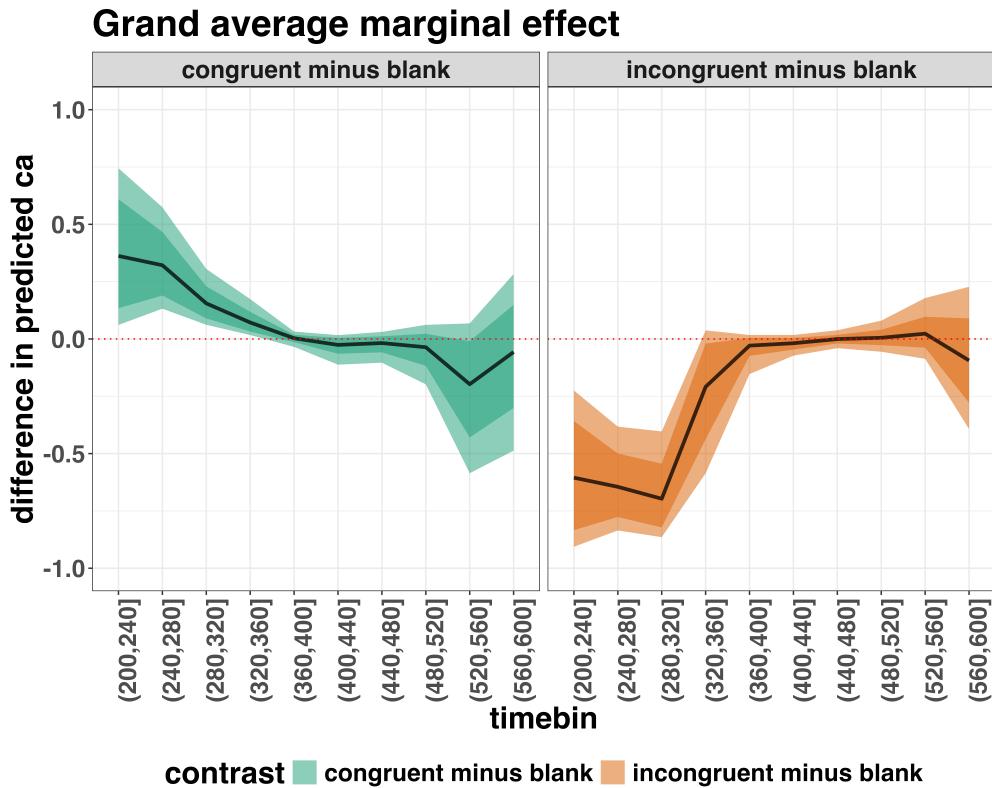


Figure 13. Point (mean) and 80/95% credible interval summaries of estimated differences in conditional accuracy in each time bin for the average participant.

618 Table 6 shows the summaries of the estimated differences in conditional accuracy for
 619 both contrasts in terms of a point estimate (the mean) and the upper and lower bounds of
 620 the 95% credible interval, for the average participant.

Table 6

Point (mean) and 95% credible interval summary of estimated differences in conditional accuracy, for each time bin and contrast, in the average participant.

contrast	timebin	diff_ca	.lower	.upper
congruent minus blank	6	0.36	0.06	0.74
congruent minus blank	7	0.32	0.13	0.57

Table 6 continued

contrast	timebin	diff_ca	.lower	.upper
congruent minus blank	8	0.16	0.06	0.31
congruent minus blank	9	0.07	0.02	0.17
congruent minus blank	10	0.00	-0.03	0.03
congruent minus blank	11	-0.03	-0.11	0.02
congruent minus blank	12	-0.02	-0.10	0.03
congruent minus blank	13	-0.04	-0.20	0.06
congruent minus blank	14	-0.20	-0.59	0.07
congruent minus blank	15	-0.06	-0.49	0.28
incongruent minus blank	6	-0.61	-0.91	-0.22
incongruent minus blank	7	-0.64	-0.84	-0.38
incongruent minus blank	8	-0.70	-0.86	-0.40
incongruent minus blank	9	-0.21	-0.59	0.04
incongruent minus blank	10	-0.03	-0.15	0.02
incongruent minus blank	11	-0.02	-0.07	0.02
incongruent minus blank	12	0.00	-0.04	0.04
incongruent minus blank	13	0.01	-0.06	0.08
incongruent minus blank	14	0.02	-0.09	0.18
incongruent minus blank	15	-0.09	-0.39	0.23

Note. diff = difference in predicted conditional accuracy.

621

622 Based on Figure 13 and Table 6, we see that congruent primes have a positive effect on
 623 the conditional accuracy of emitted responses in time bins (200,240], (240,280], and

624 (280,320], relative to the estimates in the baseline condition (blank prime; red dashed lines in
625 Figure 14). Incongruent primes have a negative effect on the conditional accuracy of emitted
626 responses in those time bins, relative to the estimates in the baseline condition.

627 **4.5 Tutorial 3a: Fitting Frequentist hazard models**

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

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

635 In Figure 14 we compare the parameter estimates of model M1i from `brm()` with those
636 of model `M1i_f` from `glmer()`.

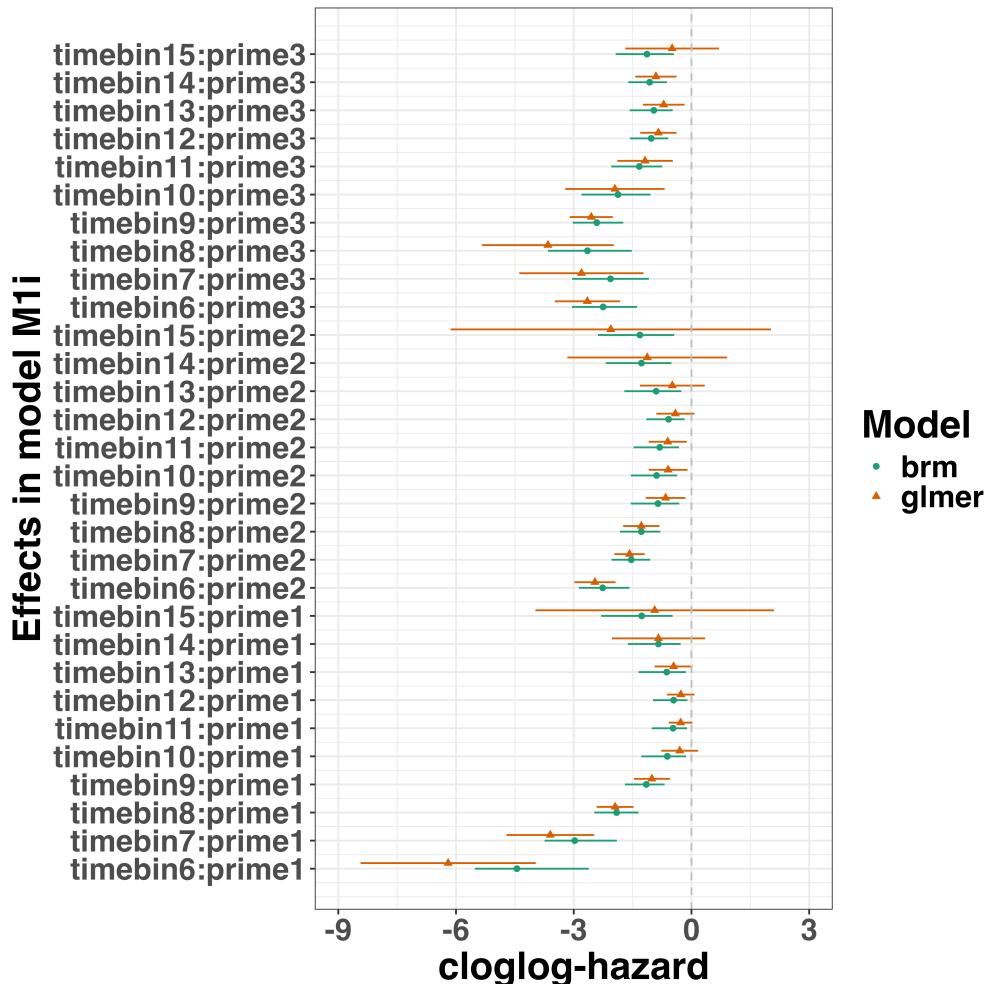


Figure 14. Parameter estimates for model M1i from `brm()` – means and 95% credible intervals – and model M1i_f from `glmer()` – maximum likelihood estimates and 95% confidence intervals.

637 Figure 14 confirms that the parameter estimates from both Bayesian and frequentist
 638 models are pretty similar, which makes sense given the close similarity in model structure.
 639 However, model M1i_f did not converge and resulted in a singular fit. This is of course one
 640 of the reasons why Bayesian modeling has become so popular in recent years. But the price
 641 you pay for being able to fit more complex random effects models in a Bayesian framework is
 642 computation time. In other words, as we have noted throughout, some of the Bayesian
 643 models in Tutorials 2a took several hours to build.

644 4.6 Tutorial 3b: Fitting Frequentist conditional accuracy models

645 In this sixth tutorial we illustrate how to fit a multilevel regression model to the timed
646 accuracy data in the frequentist framework, for the data set used in Tutorial 1a. To keep it
647 short, we only fitted the effects from model M1i_ca (see Tutorial 2b) using the function
648 glmer() from the R package lme4. Alternatively, one could also use the function glmmPQL()
649 from the R package MASS (Ripley et al., 2024). Again, the resulting conditional accuracy
650 model M1i_ca_f did not converge and resulted in a singular fit.

651 4.7 Tutorial 4: Planning

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

653 **4.7.1 Background.** The general approach to planning that we adopt here involves
654 simulating data to help guide what you might be able to expect from your data once you
655 collect it (Gelman et al., 2020). The basic structure and code follows the examples outlined
656 by Solomon Kurz in his ‘power’ blog posts (Kurz, 2019) and Lisa Debruine’s R package
657 faux{} (<https://debruine.github.io/faux/>) as well as the related paper (DeBruine & Barr,
658 2021).

659 **4.7.2 Basic workflow.** The basic workflow is as follows:

- 660 1. Fit a regression model to an existing dataset.
- 661 2. Use the regression model parameters to simulate one new dataset.
- 662 3. Write a function to create 1000s of datasets and vary parameters of interest (e.g.,
663 sample size, trial count, effect size).
- 664 4. Summarise the simulated data to estimate likely power or precision of the research
665 design options.

666 Ideally, in the above workflow, we would also fit a model to each dataset and
667 summarise the model output, rather than the raw data. However, when each model takes
668 several hours to build, and we may want to simulate 1000s of datasets, it can be too

669 computationally demanding for desktop machines. So, for ease, here we just use the raw
670 simulated datasets to set expectations.

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

673 **4.7.3 Fit a regression model and simulate one dataset.** We again use the data
674 from Panis & Schmidt (2016) to provide a worked example. We fit an index coding model on
675 a subset of timebins (six timebins in total) and for two prime conditions (congruent and
676 incongruent). We chose to focus on a subsample of the data to ease the computational
677 burden. We also used a full varying effects structure, with the model formula, as follows:

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

678 We then took parameters from this model and used them to create a single dataset
679 with 200 trials per condition for 10 individual participants. The raw data and the simulated
680 data are plotted in Figure 15 and show quite close correspondence, which is re-assuring. But,
681 of course, this is only one dataset. What we really want to do is simulate many datasets and
682 vary parameters of interest, which is what we turn to in the next section.

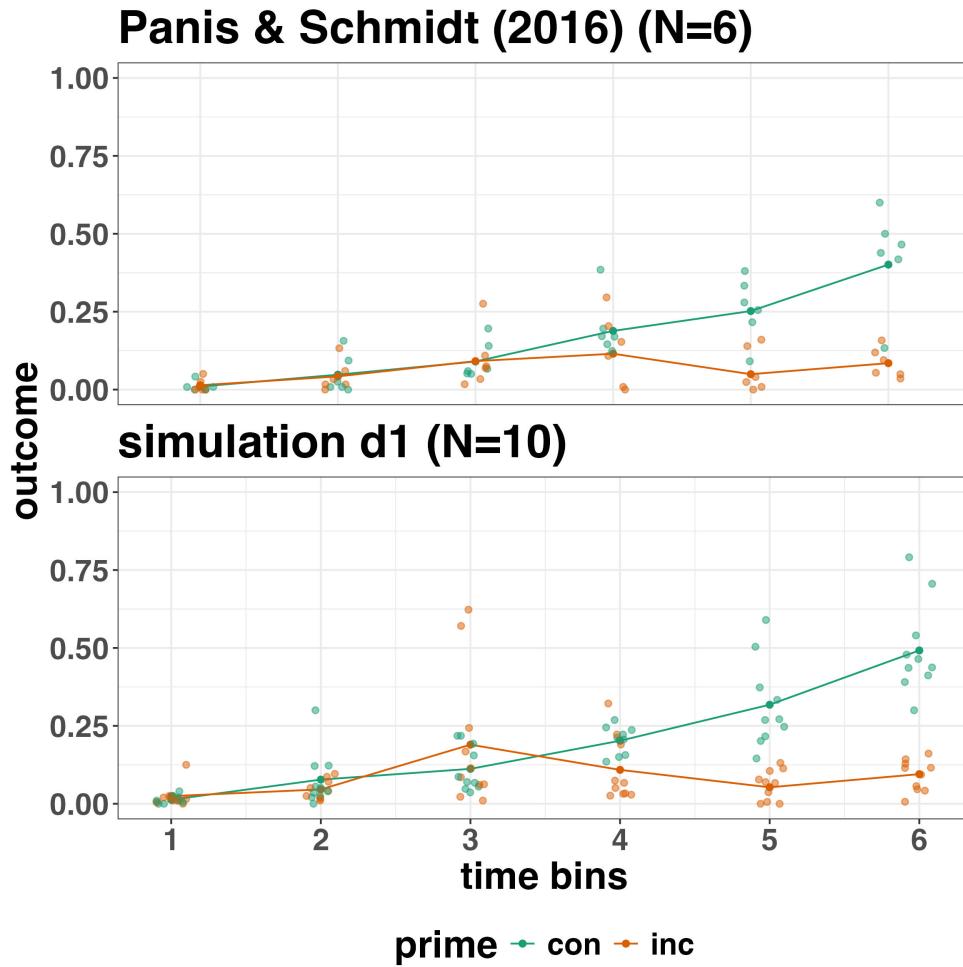


Figure 15. Raw data from Panis & Schmidt (2016) and simulated data from 10 participants.

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

Here we use the same data simulating process as used above, but instead of simulating one dataset, we simulate 1000 datasets per variation in parameter values. Specifically, in Simulation 1, we vary the number of trials per condition (100, 200, and 400), as well as the effect size in bin 6. We focus on bin 6 only, in terms of varying the effect size, just to make things simpler and easier to understand. The effect size observed in bin 6 in this subsample of data was a 79% reduction in hazard value from the congruent prime (0.401 hazard value) to the incongruent prime condition (0.085 hazard value). In other words, a hazard ratio of 0.21. As a starting point, we chose three effect sizes, which covered a fairly broad range of

692 hazard ratios (0.25, 0.5, 0.75), which correspond to a 75%, 50% and 25% reduction in hazard
693 as a function of prime condition.

694 Summary results from Simulation 1 are shown in Figure 16A. Figure 16A which
695 depicts statistical “power” as calculated by the percentage of lower-bound 95% confidence
696 intervals that exclude zero when the difference between prime condition is calculated
697 (congruent - incongruent). We know that “power” is not a Bayesian thing, but we choose to
698 include it here, as it is familiar to most researchers in experimental psychology. In other
699 words, what fraction of the simulated datasets generated an effect of prime that excludes the
700 criterion mark of zero. The results show that if we were targetting an effect size similar to
701 the one reported in the orgininal study, the using 10 participants and 100 trials per condition
702 would be enough to provide over 95% power. However, we could not be as confident about
703 smaller effects, such as a hazard ratio of 50% or 25%. From this simulation, we can see that
704 somewhere between an effect size of 50% and 75%, power increases to a range that most
705 researchers would consider acceptable (i.e., >95% power). To probe this space a little
706 further, we decided to run a second simulation, which varied different parameters

707 In Simulation 2, we varied the effect size across different values (0.5, 0.4, 0.3), which
708 correspond to a 50%, 60% and 70% reduction in hazard as a function of prime condition. In
709 addition, we varied the number of participants in the data (10, 15, and 20 participants).
710 Given that trial count per condition made little difference to power in Simulation 1, we fixed
711 trial count at 200 trials per condition in Simulation 2. Summary results from Simulation 2
712 are shown in Figure 16B. A summary of these power calculations might be as follows:

713 For a 70% reduction (0.3 hazard ratio), N=10, trial=200 per condition, would give
714 nearly 100% power. For a 60% reduction (0.4 hazard ratio), N=10, trial=200 per condition,
715 would give nearly 90% power. For a 50% reduction (0.5 hazard ratio), N=15, trial=200 per
716 condition, would give over 80% power.

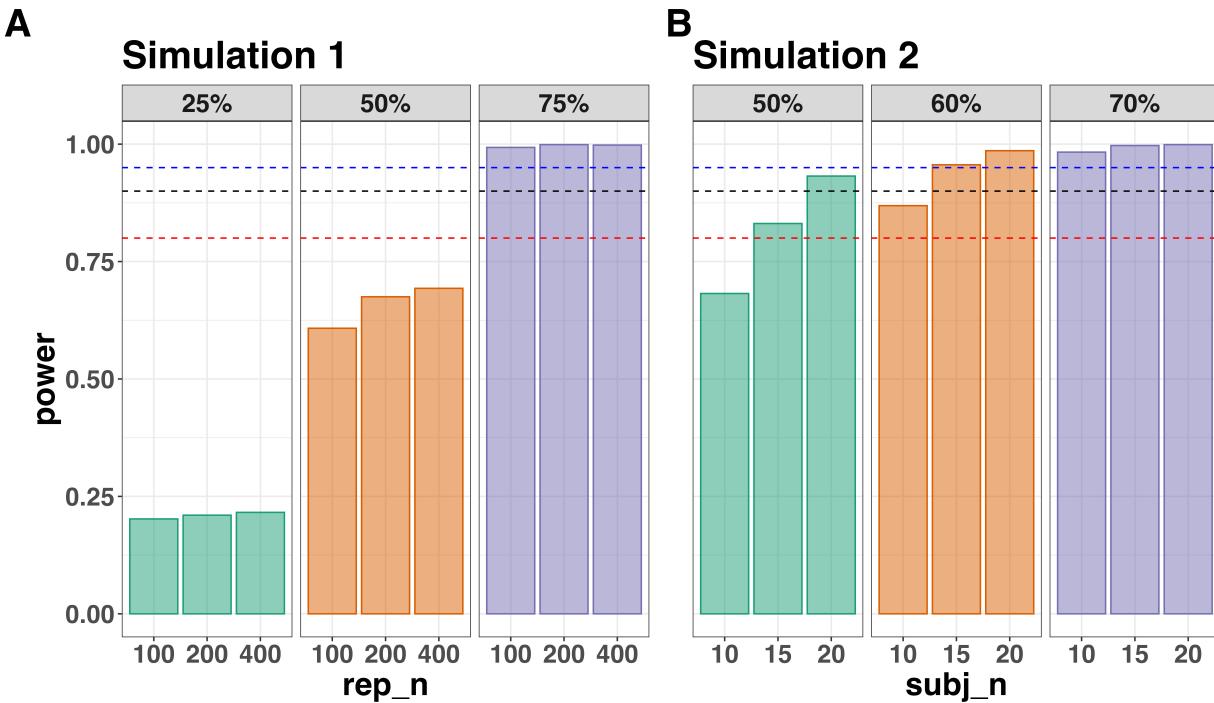


Figure 16. Statistical power across data Simulation 1 (A) and Simulation 2 (B). Power was calculated as the percentage of lower-bound 95% confidence intervals that exclude zero when the difference between prime condition is calculated (congruent - incongruent). In Simulation 1, the effect size was varied 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.

717 **4.7.5 Planning decisions.** Now that we have summarised our simulated data, what
 718 planning decisions do we make about future study? How many trials per condition should we
 719 collect and how many participants should we test? Like almost always when planning future
 720 studies, the answer depends on your objectives, as well as the available resources (Lakens,
 721 2022). There is no straightforward and clear-cut answer. Some considerations might be...

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

726 If we were running this kind of study in our lab, what would we do? We might pick a
727 hazard ratio of 0.4 or 0.5 as a target effect size since this is much smaller than that observed
728 in the previously published study that this work is building upon, then pick the
729 corresponding N value (i.e., N=10 or N=15) that takes you over the 80% power mark. If we
730 wanted to maximise power based on these simulations, and we had the time and resources
731 available, then we test N=20 participants, which would provide >90% power for an effect
732 size of 0.5.

733 **But**, and this is an important “but”, unless there are unavoidable reasons, no matter
734 what planning choices we made, we would not solely rely on one study. Instead, we would
735 run a follow-up experiment that replicates and extends the initial result. By doing so, we
736 would aim to avoid the Cult of the Isolated Single Study (Nelder, 1986; Tong, 2019), and
737 thus reduce the reliance on any one type of power analysis. Then, we would look for common
738 patterns across two or more experiments, rather than trying to make the case that a single
739 study has sufficient evidential value to hit some criterion mark.

740 5. Discussion

741 This main motivation for writing this paper is the observation that EHA and SAT
742 analysis remain under-used in psychological research. As a consequence, the field of
743 psychological research is not taking full advantage of the many benefits EHA/SAT provides
744 compared to more conventional analyses. By providing a freely available set of tutorials,
745 which provide step-by-step guidelines and ready-to-use R code, we hope that researchers will
746 feel more comfortable using EHA/SAT in the future. Indeed, we hope that our tutorials may

747 help to overcome a barrier to entry with EHA/SAT, which is that such approaches require
748 more analytical complexity compared to mean-average comparisons. While we have focused
749 here on within-subject, factorial, small- N designs, it is important to realize that EHA/SAT
750 can be applied to other designs as well (large- N designs with only one measurement per
751 subject, between-subject designs, etc.). As such, the general workflow and associated code
752 can be modified and applied more broadly to other contexts and research questions. In the
753 following, we discuss issues relating to model complexity and interpretability, individual
754 differences, as well as limitations of the approach and future extensions.

755 **5.1 What are the main use-cases of EHA for understanding cognition and brain
756 function?**

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

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

773 5.2 Model complexity versus interpretability

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

790 5.3 Individual differences

791 One important issue is that of possible individual differences in the overall location of
792 the distribution, and the time course of psychological effects. For example, when you wait for
793 a response of the participant on each trial, you allow the participant to have control over the
794 trial duration, and some participants might respond only when they are confident that their
795 emitted response will be correct. These issues can be avoided by introducing a (relatively
796 short) response deadline in each trial, e.g., 600 ms for simple detection tasks, 1000 ms for
797 more difficult discrimination tasks, or 2 s for tasks requiring extended high-level processing.
798 Because EHA can deal in a straightforward fashion with right-censored observations (i.e.,

799 trials without an observed response), introducing a response deadline is recommended when
800 designing RT experiments. Furthermore, introducing a response deadline and asking
801 participants to respond before the deadline as much as possible, will also lead to individual
802 distributions that overlap in time, which is important when selecting a common analysis
803 time window when fitting hazard and conditional accuracy models.

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

813 Another approach to deal with individual differences is Bayesian prevalence (Ince,
814 Paton, Kay, & Schyns, 2021), which is a from of Small-N approach (Smith & Little, 2018).
815 This method looks at effects within each individual in the study and asks how likely it would
816 be to see the same result if the experiment was repeated with a new person chosen from the
817 wider population at random. This approach allows one to quantify how typical or uncommon
818 an observed effect is in the population, and the uncertainty around this estimate.

819 **5.4 Limitations**

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

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

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

5.5 Extensions

The hazard models in this tutorial assume that there is one event of interest. For RT data, this event constitutes a single transition between an “idle” state and a “responded” state. However, in certain situations, more than one event of interest might exist. For example, in a medical or health-related context, an individual might transition back and

850 forth between a “healthy” state and a “depressed” state, before being absorbed into a final
851 “death” state. When you have data on the timing of these transitions, one can apply
852 multi-state hazard models, which generalize EHA to transitions between three or more states
853 (Steele, Goldstein, & Browne, 2004). Also, the predictor variables in this tutorial are
854 time-invariant, i.e., their value did not change over the course of a trial. Thus, another
855 extension is to include time-varying predictors, i.e., predictors whose value can change across
856 the time bins within a trial (Allison, 2010). For example, when gaze position is tracked
857 during a visual search trial, the gaze-target distance will vary during a trial when the eyes
858 move around before a manual response is given; shorter gaze-target distances should be
859 associated with a higher hazard of response occurrence. Note that the effect of a
860 time-varying predictor (e.g., an occipital EEG signal) can itself vary over time.

861

6. Conclusions

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

872

References

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

- 899 Blossfeld, H.-P., & Rohwer, G. (2002). *Techniques of event history modeling: New
900 approaches to causal analysis, 2nd ed* (pp. x, 310). Mahwah, NJ, US: Lawrence Erlbaum
901 Associates Publishers.
- 902 Box-Steffensmeier, J. M. (2004). Event history modeling: A guide for social scientists.
903 Cambridge: University Press.
- 904 Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan.
905 *Journal of Statistical Software*, 80(1), 1–28. <https://doi.org/10.18637/jss.v080.i01>
- 906 Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms.
907 *The R Journal*, 10(1), 395–411. <https://doi.org/10.32614/RJ-2018-017>
- 908 Bürkner, P.-C. (2021). Bayesian item response modeling in R with brms and Stan. *Journal
909 of Statistical Software*, 100(5), 1–54. <https://doi.org/10.18637/jss.v100.i05>
- 910 Eddelbuettel, D., & Balamuta, J. J. (2018). Extending R with C++: A Brief Introduction
911 to Rcpp. *The American Statistician*, 72(1), 28–36.
912 <https://doi.org/10.1080/00031305.2017.1375990>
- 913 Eddelbuettel, D., & François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal of
914 Statistical Software*, 40(8), 1–18. <https://doi.org/10.18637/jss.v040.i08>
- 915 Gabry, J., Češnovar, R., Johnson, A., & Broder, S. (2024). *Cmdstanr: R interface to
916 'CmdStan'*. Retrieved from <https://github.com/stan-dev/cmdstanr>
- 917 Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2019). Visualization in
918 bayesian workflow. *J. R. Stat. Soc. A*, 182, 389–402. <https://doi.org/10.1111/rssa.12378>
- 919 Gelman, A., Hill, J., & Vehtari, A. (2020). Regression and Other Stories.
920 <https://www.cambridge.org/highereducation/books/regression-and-other->
921 stories/DD20DD6C9057118581076E54E40C372C; Cambridge University Press.
922 <https://doi.org/10.1017/9781139161879>
- 923 Gelman, A., Vehtari, A., Simpson, D., Margossian, C. C., Carpenter, B., Yao, Y., ...
924 Modrák, M. (2020). *Bayesian Workflow*. arXiv.
925 <https://doi.org/10.48550/arXiv.2011.01808>

- 926 Girard, J. (n.d.). *Standist: What the package does (one line, title case)*. Retrieved from
927 <https://github.com/jmgirard/standist>
- 928 Grolemund, G., & Wickham, H. (2011). Dates and times made easy with lubridate. *Journal*
929 *of Statistical Software*, 40(3), 1–25. Retrieved from <https://www.jstatsoft.org/v40/i03/>
- 930 Halley, E. (1997). VI. An estimate of the degrees of the mortality of mankind; drawn from
931 curious tables of the births and funerals at the city of Breslaw; with an attempt to
932 ascertain the price of annuities upon lives. *Philosophical Transactions of the Royal*
933 *Society of London*, 17(196), 596–610. <https://doi.org/10.1098/rstl.1693.0007>
- 934 Heiss, A. (2021, November 10). A Guide to Correctly Calculating Posterior Predictions and
935 Average Marginal Effects with Multilevel Bayesian Models.
936 <https://doi.org/10.59350/wbn93-edb02>
- 937 Holden, J. G., Van Orden, G. C., & Turvey, M. T. (2009). Dispersion of response times
938 reveals cognitive dynamics. *Psychological Review*, 116(2), 318–342.
939 <https://doi.org/10.1037/a0014849>
- 940 Hosmer, D. W., Lemeshow, S., & May, S. (2011). *Applied Survival Analysis: Regression*
941 *Modeling of Time to Event Data* (2nd ed). Hoboken: John Wiley & Sons.
- 942 Ince, R. A., Paton, A. T., Kay, J. W., & Schyns, P. G. (2021). Bayesian inference of
943 population prevalence. *eLife*, 10, e62461. <https://doi.org/10.7554/eLife.62461>
- 944 Kantowitz, B. H., & Pachella, R. G. (2021). The Interpretation of Reaction Time in
945 Information-Processing Research 1. *Human Information Processing*, 41–82.
946 <https://doi.org/10.4324/9781003176688-2>
- 947 Kay, M. (2023). *tidybayes: Tidy data and geoms for Bayesian models*.
948 <https://doi.org/10.5281/zenodo.1308151>
- 949 Kelso, J. A. S., Dumas, G., & Tognoli, E. (2013). Outline of a general theory of behavior
950 and brain coordination. *Neural Networks: The Official Journal of the International*
951 *Neural Network Society*, 37, 120–131. <https://doi.org/10.1016/j.neunet.2012.09.003>
- 952 Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing,

- 953 estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic
954 Bulletin & Review*, 25(1), 178–206. <https://doi.org/10.3758/s13423-016-1221-4>
- 955 Kurz, A. S. (2023a). *Applied longitudinal data analysis in brms and the tidyverse* (version
956 0.0.3). Retrieved from <https://bookdown.org/content/4253/>
- 957 Kurz, A. S. (2023b). *Statistical rethinking with brms, ggplot2, and the tidyverse: Second
958 edition* (version 0.4.0). Retrieved from <https://bookdown.org/content/4857/>
- 959 Landes, J., Engelhardt, S. C., & Pelletier, F. (2020). An introduction to event history
960 analyses for ecologists. *Ecosphere*, 11(10), e03238. <https://doi.org/10.1002/ecs2.3238>
- 961 Luce, R. D. (1991). *Response times: Their role in inferring elementary mental organization*
962 (1. issued as paperback). Oxford: Univ. Press.
- 963 Makeham, William M. (1860). *On the Law of Mortality and the Construction of Annuity
964 Tables*. The Assurance Magazine, and Journal of the Institute of Actuaries.
- 965 McElreath, R. (2018). *Statistical Rethinking: A Bayesian Course with Examples in R and
966 Stan* (1st ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9781315372495>
- 967 Meyer, D. E., Osman, A. M., Irwin, D. E., & Yantis, S. (1988). Modern mental chronometry.
968 *Biological Psychology*, 26(1-3), 3–67. [https://doi.org/10.1016/0301-0511\(88\)90013-0](https://doi.org/10.1016/0301-0511(88)90013-0)
- 969 Müller, K., & Wickham, H. (2023). *Tibble: Simple data frames*. Retrieved from
970 <https://CRAN.R-project.org/package=tibble>
- 971 Neuwirth, E. (2022). *RColorBrewer: ColorBrewer palettes*. Retrieved from
972 <https://CRAN.R-project.org/package=RColorBrewer>
- 973 Panis, S. (2020). How can we learn what attention is? Response gating via multiple direct
974 routes kept in check by inhibitory control processes. *Open Psychology*, 2(1), 238–279.
975 <https://doi.org/10.1515/psych-2020-0107>
- 976 Panis, S., Moran, R., Wolkersdorfer, M. P., & Schmidt, T. (2020). Studying the dynamics of
977 visual search behavior using RT hazard and micro-level speed–accuracy tradeoff
978 functions: A role for recurrent object recognition and cognitive control processes.
979 *Attention, Perception, & Psychophysics*, 82(2), 689–714.

- 980 https://doi.org/10.3758/s13414-019-01897-z
- 981 Panis, S., Schmidt, F., Wolkersdorfer, M. P., & Schmidt, T. (2020). Analyzing Response
982 Times and Other Types of Time-to-Event Data Using Event History Analysis: A Tool for
983 Mental Chronometry and Cognitive Psychophysiology. *I-Perception*, 11(6),
984 2041669520978673. https://doi.org/10.1177/2041669520978673
- 985 Panis, S., & Schmidt, T. (2016). What Is Shaping RT and Accuracy Distributions? Active
986 and Selective Response Inhibition Causes the Negative Compatibility Effect. *Journal of*
987 *Cognitive Neuroscience*, 28(11), 1651–1671. https://doi.org/10.1162/jocn_a_00998
- 988 Panis, S., & Schmidt, T. (2022). When does “inhibition of return” occur in spatial cueing
989 tasks? Temporally disentangling multiple cue-triggered effects using response history and
990 conditional accuracy analyses. *Open Psychology*, 4(1), 84–114.
991 https://doi.org/10.1515/psych-2022-0005
- 992 Panis, S., Torfs, K., Gillebert, C. R., Wagemans, J., & Humphreys, G. W. (2017).
993 Neuropsychological evidence for the temporal dynamics of category-specific naming.
994 *Visual Cognition*, 25(1-3), 79–99. https://doi.org/10.1080/13506285.2017.1330790
- 995 Panis, S., & Wagemans, J. (2009). Time-course contingencies in perceptual organization and
996 identification of fragmented object outlines. *Journal of Experimental Psychology: Human*
997 *Perception and Performance*, 35(3), 661–687. https://doi.org/10.1037/a0013547
- 998 Pedersen, T. L. (2024). *Patchwork: The composer of plots*. Retrieved from
999 https://CRAN.R-project.org/package=patchwork
- 1000 Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in s and s-PLUS*. New York:
1001 Springer. https://doi.org/10.1007/b98882
- 1002 R Core Team. (2024). *R: A language and environment for statistical computing*. Vienna,
1003 Austria: R Foundation for Statistical Computing. Retrieved from
1004 https://www.R-project.org/
- 1005 Ripley, B., Venables, B., Bates, D. M., ca 1998), K. H. (partial. port, ca 1998), A. G.
1006 (partial. port, & polr), D. F. (support. functions for. (2024). *MASS: Support Functions*

- 1007 and *Datasets for Venables and Ripley's MASS*.
- 1008 Singer, J. D., & Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling Change*
1009 and Event Occurrence. Oxford, New York: Oxford University Press.
- 1010 Smith, P. L., & Little, D. R. (2018). Small is beautiful: In defense of the small-N design.
1011 *Psychonomic Bulletin & Review*, 25(6), 2083–2101.
1012 <https://doi.org/10.3758/s13423-018-1451-8>
- 1013 Steele, F., Goldstein, H., & Browne, W. (2004). A general multilevel multistate competing
1014 risks model for event history data, with an application to a study of contraceptive use
1015 dynamics. *Statistical Modelling*, 4(2), 145–159.
1016 <https://doi.org/10.1191/1471082X04st069oa>
- 1017 Teachman, J. D. (1983). Analyzing social processes: Life tables and proportional hazards
1018 models. *Social Science Research*, 12(3), 263–301.
1019 [https://doi.org/10.1016/0049-089X\(83\)90015-7](https://doi.org/10.1016/0049-089X(83)90015-7)
- 1020 Townsend, J. T. (1990). Truth and consequences of ordinal differences in statistical
1021 distributions: Toward a theory of hierarchical inference. *Psychological Bulletin*, 108(3),
1022 551–567. <https://doi.org/10.1037/0033-2909.108.3.551>
- 1023 Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics.
1024 *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)
- 1025 Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York.
1026 Retrieved from <https://ggplot2.tidyverse.org>
- 1027 Wickham, H. (2023a). *Forcats: Tools for working with categorical variables (factors)*.
1028 Retrieved from <https://CRAN.R-project.org/package=forcats>
- 1029 Wickham, H. (2023b). *Stringr: Simple, consistent wrappers for common string operations*.
1030 Retrieved from <https://CRAN.R-project.org/package=stringr>
- 1031 Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., ... Yutani,
1032 H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686.
1033 <https://doi.org/10.21105/joss.01686>

- 1034 Wickham, H., Çetinkaya-Rundel, M., & Grolemund, G. (2023). *R for data science: Import,*
1035 *tidy, transform, visualize, and model data* (2nd edition). Beijing Boston Farnham
1036 Sebastopol Tokyo: O'Reilly.
- 1037 Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *Dplyr: A grammar*
1038 *of data manipulation.* Retrieved from <https://CRAN.R-project.org/package=dplyr>
- 1039 Wickham, H., & Henry, L. (2023). *Purrr: Functional programming tools.* Retrieved from
1040 <https://CRAN.R-project.org/package=purrr>
- 1041 Wickham, H., Hester, J., & Bryan, J. (2024). *Readr: Read rectangular text data.* Retrieved
1042 from <https://CRAN.R-project.org/package=readr>
- 1043 Wickham, H., Vaughan, D., & Girlich, M. (2024). *Tidyr: Tidy messy data.* Retrieved from
1044 <https://CRAN.R-project.org/package=tidyr>
- 1045 Winter, B. (2019). *Statistics for Linguists: An Introduction Using R.* New York: Routledge.
1046 <https://doi.org/10.4324/9781315165547>
- 1047 Wolkersdorfer, M. P., Panis, S., & Schmidt, T. (2020). Temporal dynamics of sequential
1048 motor activation in a dual-prime paradigm: Insights from conditional accuracy and
1049 hazard functions. *Attention, Perception, & Psychophysics*, 82(5), 2581–2602.
1050 <https://doi.org/10.3758/s13414-020-02010-5>

1051

Supplementary material

1052 **A. Definitions of discrete-time hazard, survivor, probability mass, and conditional
1053 accuracy functions**

1054 The shape of a distribution of waiting times can be described in multiple ways (Luce,
1055 1991). After dividing time in discrete, contiguous time bins indexed by t , let RT be a
1056 discrete random variable denoting the rank of the time bin in which a particular person's
1057 response occurs in a particular experimental condition. Because waiting times can only
1058 increase, discrete-time EHA focuses on the discrete-time hazard function

1059
$$h(t) = P(RT = t \mid RT \geq t) \quad (1)$$

1060 and the discrete-time survivor function

1061
$$S(t) = P(RT > t) = [1-h(t)].[1-h(t-1)].[1-h(t-2)] \dots [1-h(1)] \quad (2)$$

1062 and not on the probability mass function

1063
$$P(t) = P(RT = t) = h(t).S(t-1) \quad (3)$$

1064 nor the cumulative distribution function

1065
$$F(t) = P(RT \leq t) = 1-S(t) \quad (4)$$

1066 The discrete-time hazard function of event occurrence gives you for each bin the
1067 probability that the event occurs (sometime) in that bin, given that the event has not
1068 occurred yet in previous bins. This conditionality in the definition of hazard is what makes
1069 the hazard function so diagnostic for studying event occurrence, as an event can physically
1070 not occur when it has already occurred before. While the discrete-time hazard function
1071 assesses the unique risk of event occurrence associated with each time bin, the discrete-time
1072 survivor function cumulates the bin-by-bin risks of event *nonoccurrence* to obtain the
1073 probability that the event occurs after bin t . The probability mass function cumulates the

₁₀₇₄ risk of event occurrence in bin t with the risks of event nonoccurrence in bins 1 to t-1. From
₁₀₇₅ equation 3 we find that hazard in bin t is equal to $P(t)/S(t-1)$.

₁₀₇₆ For two-choice RT data, the discrete-time hazard function can be extended with the
₁₀₇₇ discrete-time conditional accuracy function

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

₁₀₇₉ which gives you for each bin the probability that a response is correct given that it is emitted
₁₀₈₀ in time bin t (Allison, 2010; Kantowitz & Pachella, 2021; Wickelgren, 1977). This latter
₁₀₈₁ function is also known as the micro-level speed-accuracy tradeoff (SAT) function.

₁₀₈₂ The survivor function provides a context for the hazard function, as $S(t-1) = P(\text{RT} > t-1) = P(\text{RT} \geq t)$ tells you on how many percent of the trials the estimate $h(t) = P(\text{RT} = t \mid \text{RT} \geq t)$ is based. The probability mass function provides a context for the conditional
₁₀₈₄ accuracy function, as $P(t) = P(\text{RT} = t)$ tells you on how many percent of the trials the
₁₀₈₅ estimate $\text{ca}(t) = P(\text{correct} \mid \text{RT} = t)$ is based.

₁₀₈₇ While psychological RT data is typically measured in small, continuous units (e.g.,
₁₀₈₈ milliseconds), discrete-time EHA treats the RT data as interval-censored data, because it
₁₀₈₉ only uses the information that the response occurred sometime in a particular bin of time
₁₀₉₀ $(x,y]: x < \text{RT} \leq y$. If we want to use the exact event times, then we treat time as a
₁₀₉₁ continuous variable, and let RT be a continuous random variable denoting a particular
₁₀₉₂ person's response time in a particular experimental condition. Continuous-time EHA does
₁₀₉₃ not focus on the cumulative distribution function $F(t) = P(\text{RT} \leq t)$ and its derivative, the
₁₀₉₄ probability density function $f(t) = F(t)'$, but on the survivor function $S(t) = P(\text{RT} > t)$ and
₁₀₉₅ the hazard rate function $\lambda(t) = f(t)/S(t)$. The hazard rate function gives you the
₁₀₉₆ instantaneous *rate* of event occurrence at time point t, given that the event has not occurred
₁₀₉₇ yet.

1098 B. Custom functions for descriptive discrete-time hazard analysis

1099 We defined 12 custom functions that we list here.

- 1100 • censor(df,timeout,bin_width) : divide the time segment (0,timeout] in bins, identify
any right-censored observations, and determine the discrete RT (time bin rank)
- 1101 • ptb(df) : transform the person-trial data set to the person-trial-bin data set
- 1102 • setup_lt(ptb) : set up a life table for each level of 1 independent variable
- 1103 • setup_lt_2IV(ptb) : set up a life table for each combination of levels of 2 independent
variables
- 1104 • calc_ca(df) : estimate the conditinal accuracies when there is 1 independent variable
- 1105 • calc_ca_2IV(df) : estimate the conditional accuraiies when there are 2 independent
variables
- 1106 • join_lt_ca(df1,df2) : add the ca(t) estimates to the life tables (1 independent variable)
- 1107 • join_lt_ca_2IV(df1, df2) : add the ca(t) estimates to the life tables (2 independent
variables)
- 1108 • extract_median(df) : estimate quantiles $S(t)._{50}$ (1 independent variable)
- 1109 • extract_median_2IV(df) : estimate quantiles $S(t)._{50}$ (2 independent variables)
- 1110 • plot_eha(df, subj, haz_yaxis=1, first_bin_shown=1, aggregated_data=F, Nsubj=6) :
create plots of the discrete-time functions (1 independent variable), and specify the
upper limit of the y-axis in the hazard plot, with which bin to start plotting, whether
the data is aggregated across participants, and across how many participants
- 1111 • plot_eha_2IV(df, subj, haz_yaxis=1, first_bin_shown=1, aggregated_data=F,
Nsubj=6) : create plots of the discrete-time functions (2 independent variables), and
specify the upper limit of the y-axis in the hazard plot, with which bin to start plotting,
whether the data is aggregated across participants, and across how many participants

1122 When you want to analyse simple RT data from a detection experiment with one
1123 independent variable, the functions calc_ca() and join_lt_ca() should not be used, and the

1124 code to plot the conditional accuracy functions should be removed from the function
 1125 `plot_eha()`. When you want to analyse simple RT data from a detection experiment with
 1126 two independent variables, the functions `calc_ca_2IV()` and `join_lt_ca_2IV()` should not
 1127 be used, and the code to plot the conditional accuracy functions should be removed from the
 1128 function `plot_eha_2IV()`.

1129 **C. Link functions**

1130 Popular link functions include the logit link and the complementary log-log link, as
 1131 shown in Figure 15.

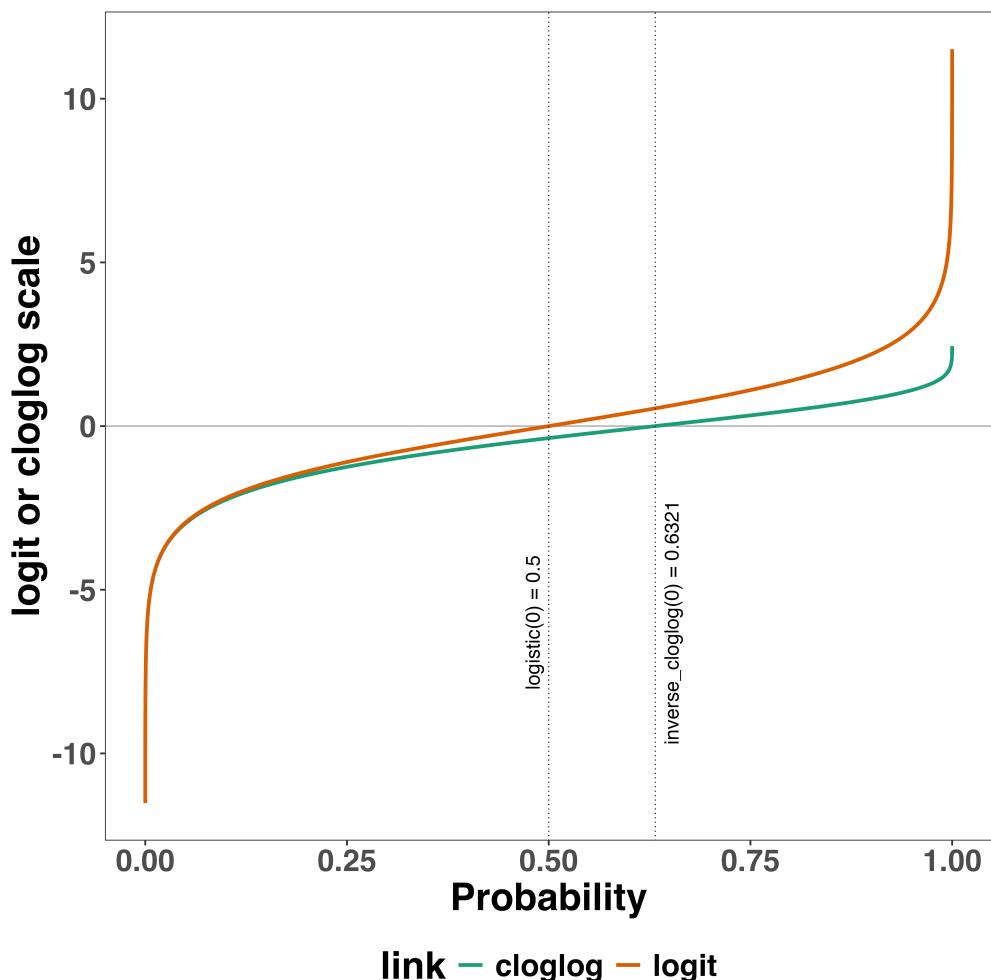


Figure 17. The logit and cloglog link functions.

₁₁₃₂ **D. Regression equations**

₁₁₃₃ An example (single-level) discrete-time hazard model with three predictors (TIME, X₁,
₁₁₃₄ X₂), the cloglog link function, and a second-order polynomial specification for TIME can be
₁₁₃₅ written as follows:

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

₁₁₃₈ The main predictor variable TIME is the time bin index t that is centered on value 9
₁₁₃₉ in this example. The first set of terms within brackets, the parameters β_0 to β_2 multiplied by
₁₁₄₀ their polynomial specifications of (centered) time, represents the shape of the baseline
₁₁₄₁ cloglog-hazard function (i.e., when all predictors X_i take on a value of zero). The second set
₁₁₄₂ of terms (the beta parameters β_3 to β_5) represents the vertical shift in the baseline
₁₁₄₃ cloglog-hazard for a 1 unit increase in the respective predictor variable. Predictors can be
₁₁₄₄ discrete, continuous, and time-varying or time-invariant. For example, the effect of a 1 unit
₁₁₄₅ increase in X₁ is to vertically shift the whole baseline cloglog-hazard function by β_3
₁₁₄₆ cloglog-hazard units. However, if the predictor interacts linearly with TIME (see X₂ in the
₁₁₄₇ example), then the effect of a 1 unit increase in X₂ is to vertically shift the predicted
₁₁₄₈ cloglog-hazard in bin 9 by β_4 cloglog-hazard units (when TIME-9 = 0), in bin 10 by $\beta_4 + \beta_5$
₁₁₄₉ cloglog-hazard units (when TIME-9 = 1), and so forth. To interpret the effects of a predictor,
₁₁₅₀ its β parameter is exponentiated, resulting in a hazard ratio (due to the use of the cloglog
₁₁₅₁ link). When using the logit link, exponentiating a β parameter results in an odds ratio.

₁₁₅₂ An example (single-level) discrete-time hazard model with a general specification for
₁₁₅₃ TIME (separate intercepts for each of six bins, where D1 to D6 are binary variables
₁₁₅₄ identifying each bin) and a single predictor (X₁) can be written as follows:

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

1156 E. Prior distributions

1157 To gain a sense of what prior *logit* values would approximate a uniform distribution on
1158 the probability scale, Kurz (2023a) simulated a large number of draws from the Uniform(0,1)
1159 distribution, converted those draws to the log-odds metric, and fitted a Student's t
1160 distribution. Row C in Figure 16 shows that using a t-distribution with 7.61 degrees of
1161 freedom and a scale parameter of 1.57 as a prior on the logit scale, approximates a uniform
1162 distribution on the probability scale. According to Kurz (2023a), such a prior might be a
1163 good prior for the intercept(s) in a logit-hazard model, while the N(0,1) prior in row D might
1164 be a good prior for the non-intercept parameters in a logit-hazard model, as it gently
1165 regularizes p towards .5 (i.e., a zero effect on the logit scale).

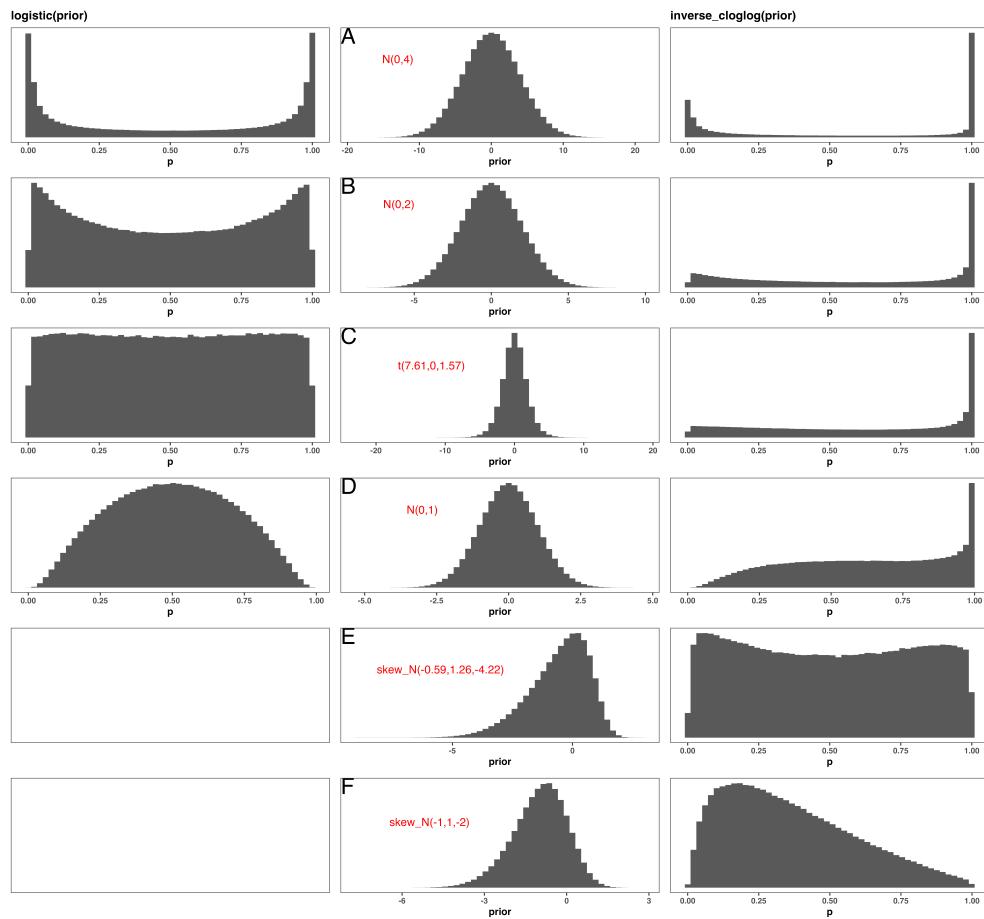


Figure 18. Prior distributions for the Intercept on the logit and/or cloglog scales (middle column), and their implications on the probability scale after applying the inverse-logit (or logistic) transformation (left column), and the inverse-cloglog transformation (right column).

1166 To gain a sense of what prior *cloglog* values would approximate a uniform distribution
 1167 on the hazard probability scale, we followed Kurz's approach and simulated a large number
 1168 of draws from the Uniform(0,1) distribution, converted them to the cloglog metric, and fitted
 1169 a skew-normal model (due to the asymmetry of the cloglog link function). Row E shows that
 1170 using a skew-normal distribution with a mean of -0.59, a standard deviation of 1.26, and a
 1171 skewness of -4.22 as a prior on the cloglog scale, approximates a uniform distribution on the
 1172 probability scale. However, because hazard values below .5 are more likely in RT studies,
 1173 using a skew-normal distribution with a mean of -1, a standard deviation of 1, and a

₁₁₇₄ skewness of -2 as a prior on the cloglog scale (row F), might be a good weakly informative
₁₁₇₅ prior for the intercept(s) in a cloglog-hazard model.

₁₁₇₆ **F. Advantages of hazard analysis**

₁₁₇₇ Statisticians and mathematical psychologists recommend focusing on the hazard
₁₁₇₈ function when analyzing time-to-event data for various reasons. First, as discussed by
₁₁₇₉ Holden, Van Orden, and Turvey (2009), “probability density [and mass] functions can appear
₁₁₈₀ nearly identical, both statistically and to the naked eye, and yet are clearly different on the
₁₁₈₁ basis of their hazard functions (but not vice versa). Hazard functions are thus more
₁₁₈₂ diagnostic than density functions” (p. 331) when one is interested in studying the detailed
₁₁₈₃ shape of a RT distribution (see also Figure 1 in Panis, Schmidt, et al., 2020). Therefore,
₁₁₈₄ when the goal is to study how psychological effects change over time, hazard and conditional
₁₁₈₅ accuracy functions are the preferred ways to describe the RT + accuracy data.

₁₁₈₆ Second, because RT distributions may differ from one another in multiple ways,
₁₁₈₇ Townsend (1990) developed a dominance hierarchy of statistical differences between two
₁₁₈₈ arbitrary distributions A and B. For example, if $h_A(t) > h_B(t)$ for all t, then both hazard
₁₁₈₉ functions are said to show a complete ordering. Townsend (1990) concluded that stronger
₁₁₉₀ conclusions can be drawn from data when comparing the hazard functions using EHA. For
₁₁₉₁ example, when $\text{mean } A < \text{mean } B$, the hazard functions might show a complete ordering
₁₁₉₂ (i.e., for all t), a partial ordering (e.g., only for $t > 300$ ms, or only for $t < 500$ ms), or they
₁₁₉₃ may cross each other one or more times.

₁₁₉₄ Third, EHA does not discard right-censored observations when estimating hazard
₁₁₉₅ functions, that is, trials for which we do not observe a response during the data collection
₁₁₉₆ period in a trial so that we only know that the RT must be larger than some value (e.g., the
₁₁₉₇ response deadline). This is important because although a few right-censored observations are
₁₁₉₈ inevitable in most RT tasks, a lot of right-censored observations are expected in experiments

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

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

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