

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, Gloriastrasse 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. Moreover, well-established analytical approaches, such as event history
23 analysis, are able to evaluate the detailed shape of time-to-event distributions, and thus
24 characterise the timeline of psychological states. One barrier to wider use of event history
25 analysis, however, is that the analytical workflow is typically more time-consuming and
26 complex than orthodox approaches. To help achieve broader uptake, in this paper we outline
27 a set of tutorials that detail how to implement one distributional method known as
28 discrete-time event history analysis. We illustrate how to wrangle raw data files and
29 calculate descriptive statistics, as well as how to calculate inferential statistics via Bayesian
30 and frequentist multi-level regression modelling. Along the way, we touch upon several key
31 aspects of the workflow, such as how to specify regression models, the implications for
32 experimental design, as well as how to manage inter-individual differences. We finish the
33 article by considering the benefits of the approach for understanding psychological states, as
34 well as the limitations and future directions of this work. Finally, the project is written in R
35 and freely available, which means the general approach can easily be adapted to other data
36 sets, and all of the tutorials are available as .html files to widen access beyond R-users.

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

39 Word count: X

40

1. Introduction

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

43 In experimental psychology, it is standard practice to analyse reaction 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, top row). 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, bottom row). Therefore, a comparison of
62 means 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.

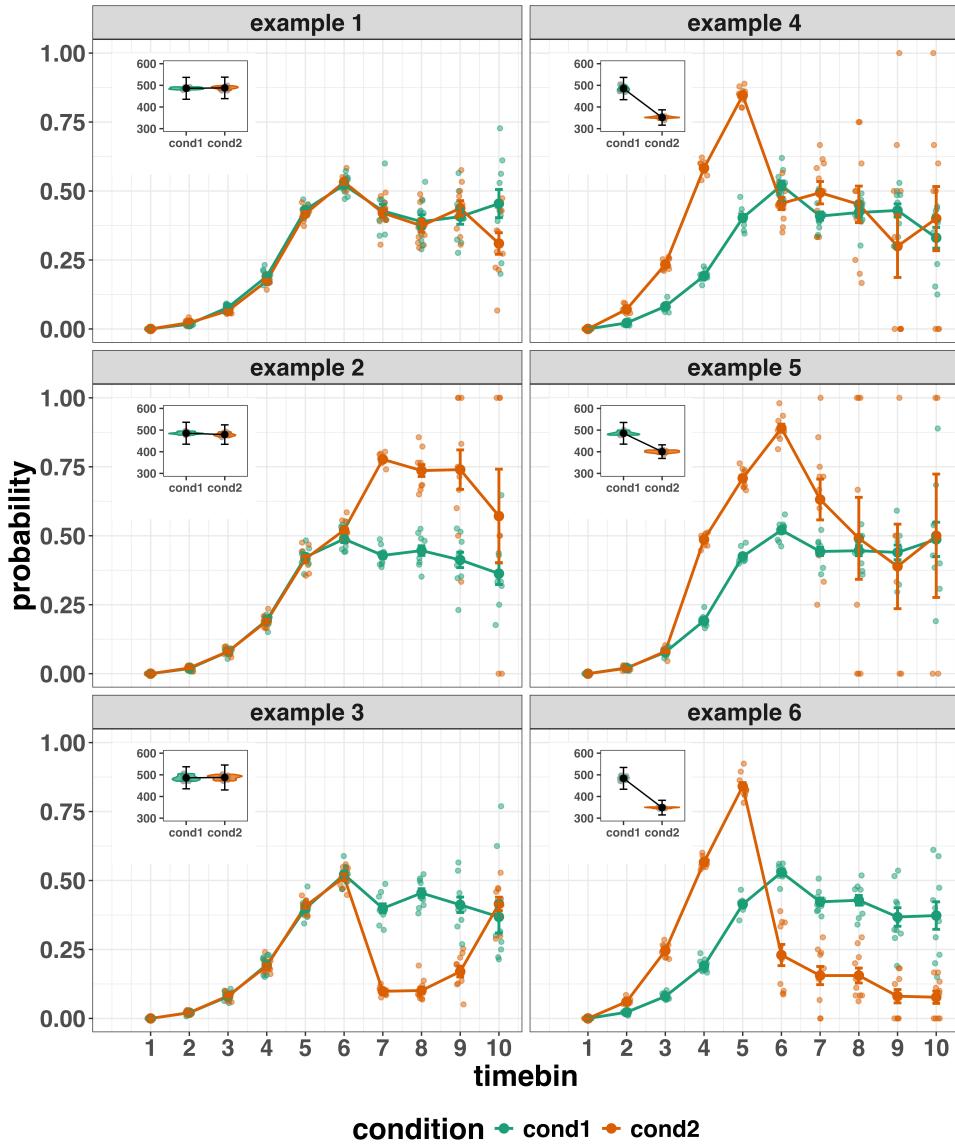


Figure 1. Means versus distributional shapes for six different simulated data set examples.

The first second after stimulus onset is divided in ten bins of 100 ms. 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 reaction time per condition.

64 Why does this matter for research in psychology? Compared to the aggregation of data
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 around 200 ms and is gone by 600 ms.
70 In contrast, in Example 5, the effect starts around 400 ms and is gone by 800 ms. And in the
71 Example 6, the effect reverses between 500 and 600 ms. What kind of theory or theories
72 could account for such effects? Are there new auxiliary assumptions that theories need to
73 adopt? And are there new experiments that need to be performed to test the novel
74 predictions that follow from these analyses? As we show later using published examples, for
75 many psychological questions, such “temporal states” information can be theoretically
76 meaningful by leading to more fine-grained understanding of psychological processes, as well
77 as adding a relatively under-used dimension – the passage of time – to the theory building
78 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., William Matthew Makeham, 1860). The same logic
83 has been applied to psychological time-to-event data, as previously demonstrated (Panis et
84 al., 2020). Here, in the current paper, we hope to show the value of event history analysis for
85 knowledge and theory building in cognitive psychology and related areas of research, such as
86 cognitive neuroscience. Moreover, we provide tutorials that provide step-by-step code and
87 instructions in the hope that we can enable others to use event history analysis in a more
88 routine, efficient and effective manner.

89 **1.2 Aims and structure of the paper**

90 In this paper, we focus on a distributional method for time-to-event data known as
91 *discrete-time event history analysis*, a.k.a. hazard analysis, duration analysis, failure-time

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

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

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

In Tutorial 2a, we illustrate how one can fit Bayesian multi-level regression models to RT data using the R package `brms`. We discuss possible link functions, and plot the model-based effects of our predictors of interest. In Tutorial 2b we fit Bayesian multi-level regression models to *timed* accuracy data to perform a micro-level speed-accuracy tradeoff

118 (SAT) analysis, which complements the event history analysis of RT data for choice RT data.
119 In Tutorial 3a, we illustrate how to fit the same type of multilevel regression model for RT
120 data in a frequentist framework using the R package lme4. We then briefly compare and
121 contrast these inferential frameworks when applied to event history analysis. In Tutorial 3b,
122 we illustrate how to perform the SAT analysis in a frequentist framework.

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

127 In summary, even though event history analysis is a widely used statistical tool and
128 there already exist many excellent reviews (e.g., Blossfeld & Rohwer, 2002;
129 Box-Steffensmeier, 2004; Hosmer, Lemeshow, & May, 2011; Teachman, 1983) and tutorials
130 (e.g., Allison, 2010; Landes, Engelhardt, & Pelletier, 2020) on its general use-cases, we are
131 not aware of any tutorials that are aimed at psychological time-to-event data, and which
132 provide worked examples of the key data processing and multi-level regression modelling
133 steps. Therefore, our ultimate goal is twofold: first, we want to convince readers of the
134 many benefits of using event history analysis when dealing with time-to-event data with a
135 focus on psychological time-to-event data, and second, we want to provide a set of practical
136 tutorials, which provide step-by-step instructions on how you actually perform a
137 discrete-time event history analysis on time-to-event data such as RT data, as well as a
138 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 event history analysis, we recommend
141 several excellent textbooks (Allison, 2010; Singer & Willett, 2003). Likewise, for a general
142 introduction to understanding regression equations, we recommend several excellent

143 textbooks (Sven - See my comments in the notes file for relevant references here). Our focus
144 here is not on providing a detailed account of the underlying regression equations, since this
145 topic has been comprehensively covered many times before. Instead, we want to provide an
146 intuition regarding how event history analysis works in general, as well as in the context of
147 experimental psychology. As such, we only supply regression equations in supplementary
148 material and then refer to them in the text whenever relevant.

149 **2.1 Basic features of event history analysis**

150 To apply event history analysis (EHA), one must be able to:

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

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

164 Discrete-time EHA focuses on the discrete-time hazard function of event occurrence
165 and the discrete-time survivor function (Figure 2). The equations that define both of these
166 functions are reported in the supplementary material (equations 1 and 2 in part A). The

167 discrete-time hazard function gives you, for each time bin, the probability that the event
168 occurs (sometime) in bin t , given that the event does not occur in previous bins. In other
169 words, it reflects the instantaneous likelihood that the event occurs in the current bin, given
170 that it has not yet occurred in the past, i.e., in one of the prior bins. This conditionality in
171 the definition of hazard is what makes the hazard function so diagnostic for studying event
172 occurrence, as an event can physically not occur when it has already occurred before. In
173 contrast, the discrete-time survivor function cumulates the bin-by-bin risks of event
174 nonoccurrence to obtain the probability that the event occurs after bin t . In other words, the
175 survivor function gives you for each time bin the likelihood that the event occurs in the
176 future, i.e., in one of the subsequent timebins.

177 The survivor function can help to qualify or provide context to the interpretation of
178 the hazard function. For example, it can give a sense of how many trials contribute to each
179 part of the hazard distribution. If a participant completes 100 trials in an experiment, and
180 the survivor function reaches a probability of 0.03 at the end of timebin (400,500], then only
181 3% of trials remain beyond this point, which in this case would amount to 3 trials. Therefore,
182 the error bars in later parts of the hazard function would typically be wider and less precise
183 compared to earlier parts.

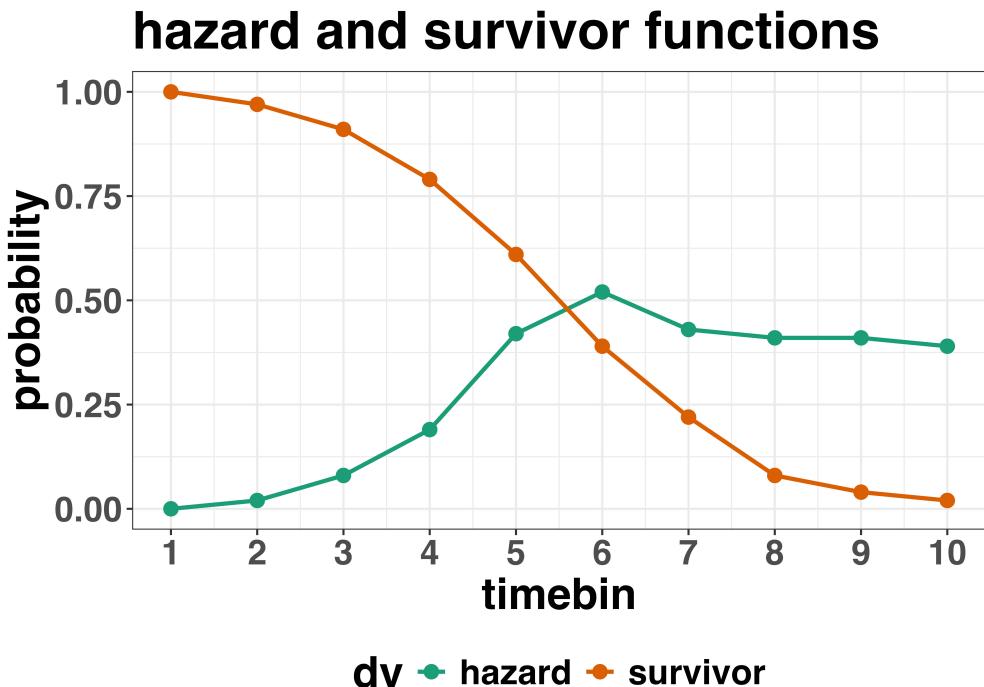


Figure 2. Discrete-time hazard and survivor functions.

¹⁸⁴ **2.2 Benefits of event history analysis**

¹⁸⁵ Statisticians and mathematical psychologists recommend focusing on the hazard
¹⁸⁶ function when analyzing time-to-event data for various reasons. We do not cover these
¹⁸⁷ benefits in detail here, as these are more general topics that have been covered elsewhere in
¹⁸⁸ textbooks. Instead, we briefly summarise list the benefits below, and refer the reader to
¹⁸⁹ section F of Supplementary Materials for more detailed coverage of the benefits. A summary
¹⁹⁰ of the benefits are as follows:

- ¹⁹¹ 1. Hazard functions are more diagnostic than density functions when one is interested in
¹⁹² studying the detailed shape of a RT distribution (Holden et al., 2009).
- ¹⁹³ 2. RT distributions may differ from each other in multiple ways, and hazard functions
¹⁹⁴ allow one to capture these differences than mean-average comparisons may conceal
¹⁹⁵ (Townsend, 1990).

- 196 3. EHA takes account of more of the data collected in a typical speeded response
197 experiment, by virtue of not discarding right-censored observations. Trials with longer
198 RTs are not discarded, but instead contribute to the risk set in each time bin.
- 199 4. Hazard modeling allows one to incorporate time-varying explanatory covariates, such
200 as heart rate, electroencephalogram (EEG) signal amplitude, gaze location, etc.
201 (Allison, 2010). This is useful for linking physiological effects to behavioral effects when
202 performing cognitive psychophysiology (Meyer, Osman, Irwin, & Yantis, 1988).
- 203 5. EHA can help to solve the problem of model mimicry, i.e., the fact that different
204 computational models can often predict the same mean RTs as observed in the
205 empirical data, but not necessarily the detailed shapes of the empirical RT hazard
206 distributions. As such, EHA can be a tool to help distinguish between competing
207 theories of cognition and brain function.

208 **2.3 Event history analysis in the context of experimental psychology**

209 To make EHA more relevant to researchers studying cognitive psychology and
210 cognitive neuroscience, in this section we provide a relevant worked example and consider
211 implications that are relevant to that domain of research.

212 **2.3.1 A worked example.** In the context of experimental psychology, it is common
213 for participants to be presented with either a 1-button detection task or a 2-button
214 discrimination task, i.e., a task that has a right and a wrong answer. For example, a task
215 may involve choosing between two response options with only one of them being correct. For
216 such two-choice RT data, the discrete-time EHA can be extended with a discrete-time SAT
217 analysis. Specifically, the hazard function of event occurrence can be extended with the
218 discrete-time conditional accuracy function, which gives you the probability that a response
219 is correct given that it is emitted in time bin t (Allison, 2010; Kantowitz & Pachella, 2021;
220 Wickelgren, 1977). We refer to this extended analysis for choice RT data as EHA/SAT.

221 Integrating results between hazard and conditional accuracy functions for choice RT

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

228 Figure 3 shows that the early upswing in hazard is equal for both prime conditions,

229 and that early responses are always correct in the congruent condition and always incorrect
230 in the incongruent condition. These results show that for short waiting times (< bin 6),
231 responses always follow the prime (and not the target, as instructed). And then for longer
232 waiting times, the response hazard is lower in incongruent compared to congruent trials, and
233 all responses emitted in these later bins are correct.

234 This joint pattern of results is interesting because it can provide meaningfully different

235 conclusions about psychological processes compared to conventional analyses, such as
236 computing mean-average RT across trials. Mean-average RT would only represent the overall
237 ability of cognition to overcome interference, on average, across trials. For instance, if
238 mean-average RT was higher in incongruent than congruent trials, one may conclude that
239 cognitive mechanisms that support interference control are working as expected across trials.

240 But such a conclusion is not supported when the effects are explored over a timeline.

241 Instead, the psychological conclusion is much more nuanced and suggests that multiple
242 states start, stop and possibly interact over a particular temporal window.

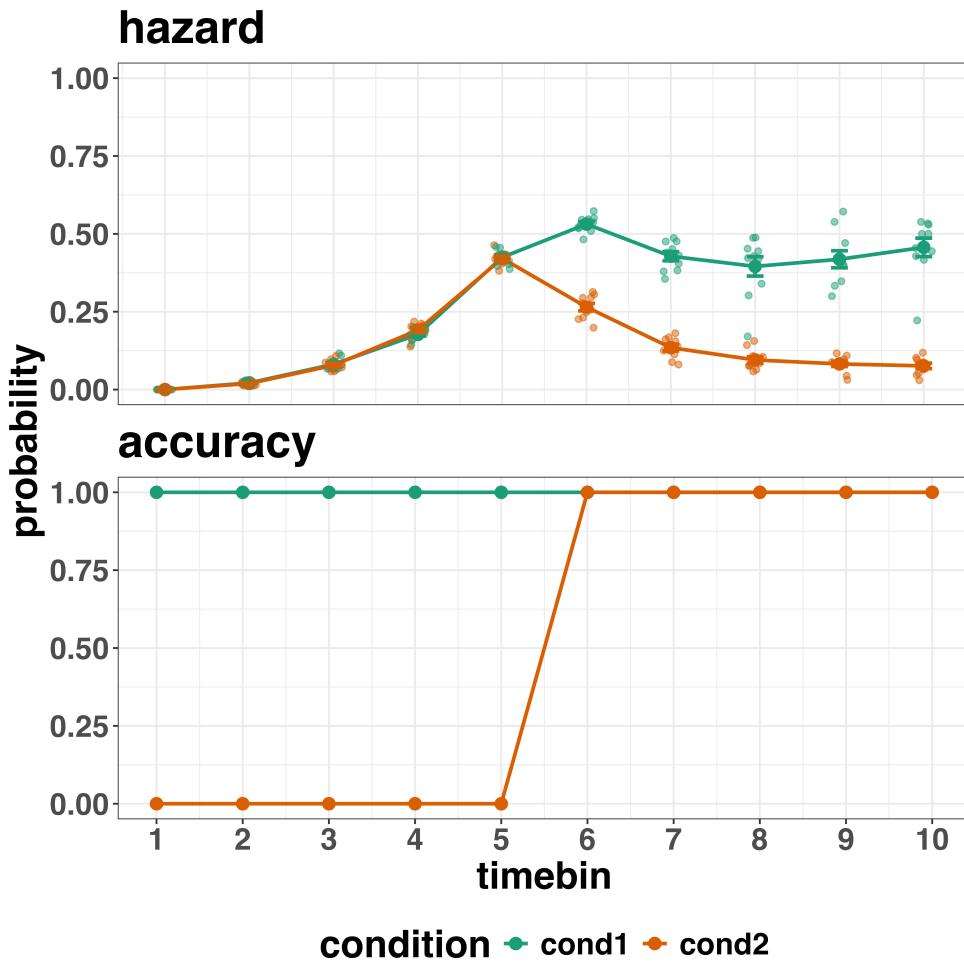


Figure 3. Discrete-time hazard and conditional accuracy functions.

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

252 **2.3.2 Implications for designing experiments.** Performing event history analyses

253 in experimental psychology has implications for how experiments are designed. Indeed, if
254 trials are categorised as a function of when responses occur, then each timebin will only
255 include a subset of the total number of trials. For example, let's consider an experiment
256 where each participant performs 2 conditions and there are 100 trial repetitions per condition.
257 Those 100 trials must be distributed in some manner across the chosen number of bins.

258 In such experimental designs, since the number of trials per condition are spread across

259 bins, it is important to have a relatively large number of trial repetitions per participant and
260 per condition. Accordingly, experimental designs using this approach typically focus on
261 factorial, within-subject designs, in which a large number of observations are made on a
262 relatively small number of participants (so-called small- N designs). This approach
263 emphasizes the precision and reproducibility of data patterns at the individual participant
264 level to increase the inferential validity of the design (Baker et al., 2021; Smith & Little,
265 2018).

266 In contrast to the large- N design that typically average across many participants

267 without being able to scrutinize individual data patterns, small- N designs retain crucial
268 information about the data patterns of individual observers. This can be advantageous
269 whenever participants differ systematically in their strategies or in the time courses of their
270 effects, so that averaging them would lead to misleading data patterns. Note that because
271 statistical power derives both from the number of participants and from the number of
272 repeated measures per participant and condition, small- N designs can still achieve what are
273 generally considered acceptable levels of statistical power, if they have have a sufficient
274 amount of data overall (Baker et al., 2021; Smith & Little, 2018).

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

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

279 **3.1 Data science workflow and descriptive statistics**

280 Descriptive, data science workflow. We perform data wrangling following tidyverse
281 principles and a functional programming approach (Wickham, Çetinkaya-Rundel, &
282 Gromelund, 2023). Functional programming basically means you don't write your own loops
283 but instead use functions that have been built and tested by others. [[more here, as
284 necessary]].

285 **3.2 Inferential statistical approach**

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

²⁹⁵ **3.3 Implementation**

²⁹⁶ We used R (Version 4.4.1; R Core Team, 2024)¹ for all reported analyses. The content
²⁹⁷ of the tutorials, in terms of EHA and multi-level regression modelling, is mainly based on
²⁹⁸ Allison (2010), Singer and Willett (2003), McElreath (2018), Kurz (2023a), and Kurz
²⁹⁹ (2023b).

³⁰⁰ **4. Tutorials**

³⁰¹ Tutorials 1a and 1b show how to calculate and plot the descriptive statistics of
³⁰² EHA/SAT when there are one and two independent variables, respectively. Tutorials 2a and
³⁰³ 2b illustrate how to use Bayesian multilevel modeling to fit hazard and conditional accuracy
³⁰⁴ models, respectively. Tutorials 3a and 3b show how to implement, respectively, multilevel
³⁰⁵ models for hazard and conditional accuracy in the frequentist framework. Additionally, to
³⁰⁶ further simplify the process for other users, the tutorials rely on a set of our own custom
³⁰⁷ functions that make sub-processes easier to automate, such as data wrangling and plotting
³⁰⁸ 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, & Brander, 2024), *dplyr* (Version 1.1.4; Wickham, François, Henry, Müller, & Vaughan, 2023), *forcats* (Version 1.0.0; Wickham, 2023a), *ggplot2* (Version 3.5.1; Wickham, 2016), *lme4* (Version 1.1.35.5; Bates, Mächler, Bolker, & Walker, 2015), *lubridate* (Version 1.9.3; Grolemund & Wickham, 2011), *Matrix* (Version 1.7.0; Bates, Maechler, & Jagan, 2024), *nlme* (Version 3.1.166; Pinheiro & Bates, 2000), *papaja* (Version 0.1.2.9000; Aust & Barth, 2023), *patchwork* (Version 1.2.0; Pedersen, 2024), *purrr* (Version 1.0.2; Wickham & Henry, 2023), *RColorBrewer* (Version 1.1.3; Neuwirth, 2022), *Rcpp* (Eddelbuettel & Balamuta, 2018; Version 1.0.12; Eddelbuettel & François, 2011), *readr* (Version 2.1.5; Wickham, Hester, & Bryan, 2024), *RJ-2021-048* (Bengtsson, 2021), *standist* (Version 0.0.0.9000; Girard, 2024), *stringr* (Version 1.5.1; Wickham, 2023b), *tibble* (Version 3.2.1; Müller & Wickham, 2023), *tidybayes* (Version 3.0.6; Kay, 2023), *tidyverse* (Version 2.0.0; Wickham et al., 2019), and *tinylabels* (Version 0.2.4; Barth, 2023).

309 Our list of tutorials is as follows:

- 310 • 1a. Wrangle raw data and calculate descriptive stats for one independent variable.
- 311 • 1b. Wrangle raw data and calculate descriptive stats for two independent variables.
- 312 • 2a. Bayesian multilevel modeling for $h(t)$
- 313 • 2b. Bayesian multilevel modeling for $ca(t)$
- 314 • 3a. Frequentist multilevel modeling for $h(t)$
- 315 • 3b. Frequentist multilevel modeling for $ca(t)$

316 Planning (T4) - if we get a simulation and power analysis script working, which we are

317 happy with then we could include it here.

318 4.1 Tutorial 1a: Calculating descriptive statistics using a life table

319 **4.1.1 Data wrangling aims.** Our data wrangling procedures serve two related

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

Table 1

Data structure for ‘person-trial’ data

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

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

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

³³⁶ **4.1.2 A real data wrangling example.** To illustrate how to quickly set up life
³³⁷ tables for calculating the descriptive statistics (functions of discrete time), we use a
³³⁸ published data set on masked response priming from Panis and Schmidt (2016). In their first
³³⁹ experiment, Panis and Schmidt (2016) presented a double arrow for 94 ms that pointed left
³⁴⁰ or right as the target stimulus with an onset at time point zero in each trial. Participants
³⁴¹ had to indicate the direction in which the double arrow pointed using their corresponding
³⁴² index finger, within 800 ms after target onset. Response time and accuracy were recorded on
³⁴³ each trial. Prime type (blank, congruent, incongruent) and mask type were manipulated.
³⁴⁴ Here we focus on the subset of trials in which no mask was presented. The 13-ms prime
³⁴⁵ stimulus was a double arrow with onset at -187 ms for the congruent (same direction as

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

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

350 The required column names are as follows:

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

357 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)))
```

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

```
data_nested <- data_wr %>% group_nest(pid)
data_final <- data_nested %>%
  mutate(censored = map(data, censor, 600, 40)) %>% # ! user input: censoring time, and bin width
  mutate(ptb_data = map(censored, ptb)) %>% # create person-trial-bin data set
  mutate(lifetable = map(ptb_data, setup_lt)) %>% # create life tables without ca(t)
  mutate(condacc = map(censored, calc_ca)) %>% # calculate ca(t)
```

```
mutate(lifetable_ca = map2(lifetable, condacc, join_lt_ca)) %>%    # create life tables with ca(t)
mutate(plot      = map2(.x = lifetable_ca, .y = pid, plot_eha,1))  # create plots
```

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

370 Removing such trials before calculating the mean event time will result in underestimation of
 371 the true mean.

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

376 The next step is to set up the life table using our custom function setup_lt(), calculate
 377 the conditional accuracies using our custom function calc_ca(), add the ca(t) estimates to
 378 the life table using our custom function join_lt_ca(), and then plot the descriptive statistics
 379 using our custom function plot_eha(). When creating the plots, some warning messages will
 380 likely be generated, like these:

- 381 • Removed 2 rows containing missing values or values outside the scale range
 382 (geom_line()).
- 383 • Removed 2 rows containing missing values or values outside the scale range
 384 (geom_point()).
- 385 • Removed 2 rows containing missing values or values outside the scale range

386 (geom_segment()).

387 The warning messages are generated because some bins have no hazard and $ca(t)$
388 estimates, and no error bars. They can thus safely be ignored. One can now inspect different
389 aspects, including the life table for a particular condition of a particular subject, and a plot
390 of the different functions for a particular participant.

391 In general, it is important to visually inspect the functions first for each participant, in
392 order to identify individuals that may be guessing (e.g., a flat conditional accuracy function
393 at .5 indicates that someone may be guessing), outlying individuals, and/or different groups
394 with qualitatively different behavior.

395 Table 3 shows the life table for condition “blank” (no prime stimulus presented) for
396 participant 6. A life table includes for each time bin, the risk set (i.e., the number of trials
397 that are event-free at the start of the bin), the number of observed events, and the estimates
398 of $h(t)$, $S(t)$, $P(t)$, possibly $ca(t)$, and their estimated standard errors (se). At time point
399 zero, no events can occur and therefore $h(t)$ and $ca(t)$ are undefined.

400 Figure 4 displays the discrete-time hazard, survivor, conditional accuracy, and
401 probability mass functions for each prime condition for participant 6. By using discrete-time
402 hazard functions of event occurrence – in combination with conditional accuracy functions
403 for two-choice tasks – one can provide an unbiased, time-varying, and probabilistic
404 description of the latency and accuracy of responses based on all trials of any data set.

405 For example, for participant 6, the estimated hazard values in bin (240,280] are 0.03,
406 0.17, and 0.20 for the blank, congruent, and incongruent prime conditions, respectively. In
407 other words, when the waiting time has increased until 240 ms after target onset, then the
408 conditional probability of response occurrence in the next 40 ms is more than five times
409 larger for both prime-present conditions, compared to the blank prime condition.

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

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

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

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

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. se = standard error. ca = conditional accuracy.

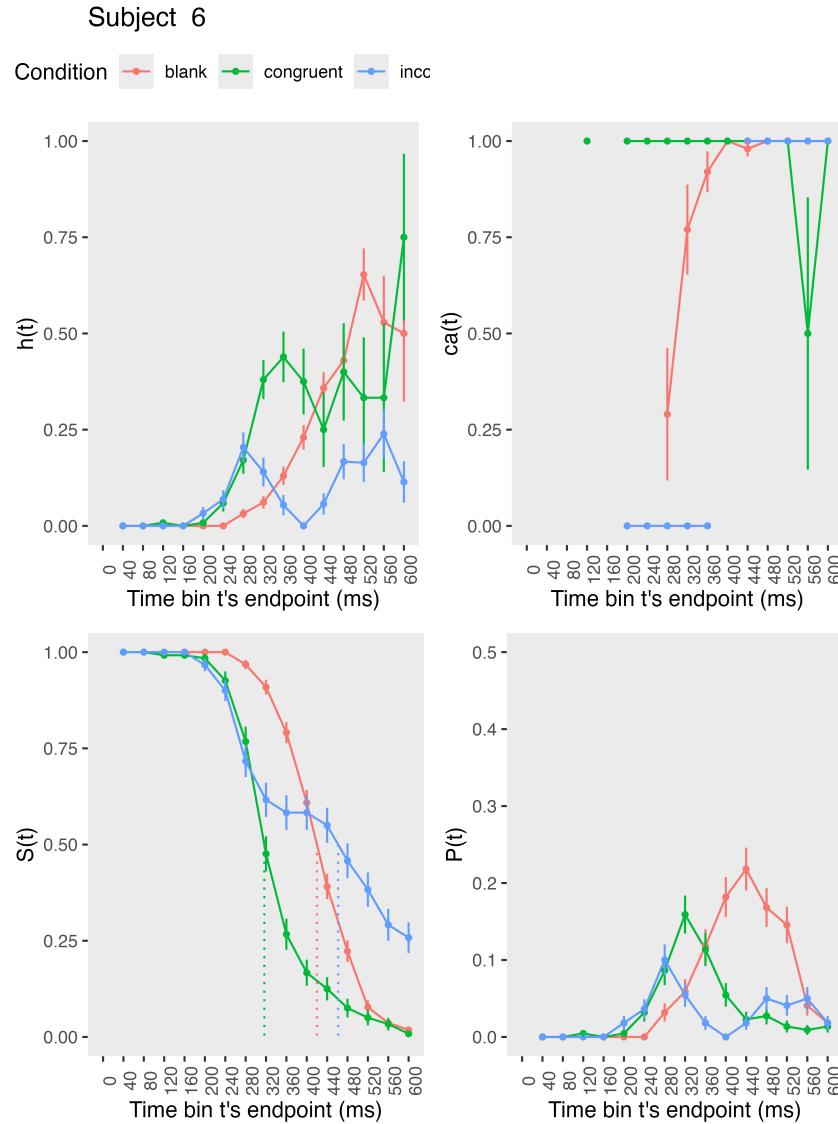


Figure 4. Estimated discrete-time hazard, survivor, probability mass, and conditional accuracy functions for participant 6, as a function of the passage of discrete waiting time.

414 However, when the waiting time has increased until 400 ms after target onset, then the
 415 conditional probability of response occurrence in the next 40 ms is estimated to be 0.36, 0.25,
 416 and 0.06 for the blank, congruent, and incongruent prime conditions, respectively. And when
 417 a response does occur in bin (400,440], then the probability that it is correct is estimated to
 418 be 0.98, 1, and 1 for the blank, congruent, and incongruent prime conditions, respectively.

419 These distributional results suggest that the participant 6 is initially responding to the
420 prime even though (s)he was instructed to only respond to the target, that response
421 competition emerges in the incongruent prime condition around 300 ms, and that only slower
422 responses are fully controlled by the target stimulus. Qualitatively similar results were
423 obtained for the other five participants. When participants show qualitatively the same
424 distributional patterns, one might consider to aggregate their data and make one plot (see
425 Tutorial_1a.Rmd).

426 In general, these results go against the (often implicit) assumption in research on
427 priming that all observed responses are primed responses to the target stimulus. Instead, the
428 distributional data show that early responses are triggered exclusively by the prime stimulus,
429 while only later responses reflect primed responses to the target stimulus.

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

436 4.2 Tutorial 1b: Generalising to a more complex design

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

443 To this end, Tutorial 1b illustrates how to calculate and plot the descriptive statistics

444 for the full data set of Experiment 1 of Panis and Schmidt (2016), which includes two
445 independent variables: mask type and prime type. As we use the same functional
446 programming approach as in Tutorial 1a, we simply present the sample-based functions for
447 participant 6 as part of Tutorial 1b for those that are interested.

448 **4.3 Tutorial 2a: Fitting Bayesian hazard models to time-to-event data**

449 In this third tutorial, we illustrate how to fit Bayesian multi-level regression models to
450 the RT data of the masked response priming data set used in Tutorial 1a. Fitting (Bayesian
451 or non-Bayesian) regression models to time-to-event data is important when you want to
452 study how the shape of the hazard function depends on various predictors (Singer & Willett,
453 2003).

454 **4.3.1 Hazard model considerations.** There are several analytic decisions one has
455 to make when fitting a hazard model. First, one has to select an analysis time window, i.e., a
456 contiguous set of bins for which there is enough data for each participant. Second, given that
457 the dependent variable (event occurrence) is binary, one has to select a link function (see
458 part C in the supplementary material). The cloglog link is preferred over the logit link when
459 events can occur in principle at any time point within a bin, which is the case for RT data
460 (Singer & Willett, 2003). Third, one has to choose a specification of the effect of discrete
461 TIME (i.e., the time bin index t) in a selected baseline condition. One can choose a general
462 specification (one intercept per bin) or a functional specification, such as a polynomial one
463 (compare model 1 with models 2, 3, and 4 below). Relevant example regression formulas are
464 provided in equations 6 and 7 in part D of the supplementary material.

465 In the case of a large- N design without repeated measurements, the parameters of a
466 discrete-time hazard model can be estimated using standard logistic regression software after
467 expanding the typical person-trial data set into a person-trial-bin data set (Allison, 2010).
468 When there is clustering in the data, as in the case of a small- N design with repeated
469 measurements, the parameters of a discrete-time hazard model can be estimated using

⁴⁷⁰ population-averaged methods (e.g., Generalized Estimating Equations), and Bayesian or
⁴⁷¹ frequentist generalized linear mixed models (Allison, 2010).

⁴⁷² In general, there are three assumptions one can make or relax when adding
⁴⁷³ experimental predictor variables and other covariates: The linearity assumption for
⁴⁷⁴ continuous predictors (the effect of a 1 unit change is the same anywhere on the scale), the
⁴⁷⁵ additivity assumption (predictors do not interact), and the proportionality assumption
⁴⁷⁶ (predictors do not interact with TIME).

⁴⁷⁷ In this tutorial we will fit four Bayesian multilevel models (i.e., generalized linear
⁴⁷⁸ mixed models) that differ in complexity to the person-trial-bin oriented data set that we
⁴⁷⁹ created in Tutorial 1a. We decided to select the analysis time window (200,600] and the
⁴⁸⁰ cloglog link. The data is prepared as follows.

```
# load person-trial-bin oriented data set
ptb_data <- read_csv("../Tutorial_1_descriptive_stats/data/inputfile_hazard_modeling.csv")
# select analysis time window: (200,600] with 10 bins (time bin ranks 6 to 15)
ptb_data <- ptb_data %>% filter(period > 5)
# create factor condition, with "blank" as the reference level
ptb_data <- ptb_data %>% mutate(condition = factor(condition, labels = c("blank", "congruent", "incongruent")))
# center TIME (variable period) on bin 9, and variable trial on number 1000 and rescale; Add dummy variables for each bin.
ptb_data <- ptb_data %>%
  mutate(period_9 = period - 9,
         trial_c = (trial - 1000)/1000,
         d6 = if_else(period == 6, 1, 0),
         d7 = if_else(period == 7, 1, 0),
         d8 = if_else(period == 8, 1, 0),
         d9 = if_else(period == 9, 1, 0),
         d10 = if_else(period == 10, 1, 0),
         d11 = if_else(period == 11, 1, 0),
         d12 = if_else(period == 12, 1, 0),
         d13 = if_else(period == 13, 1, 0),
         d14 = if_else(period == 14, 1, 0),
         d15 = if_else(period == 15, 1, 0))
```

⁴⁸¹ **4.3.2 Prior distributions.** To get the posterior distribution of each model
⁴⁸² parameter given the data, we need to specify a prior distribution for each parameter. The

483 middle column of Figure 12 in part E of the supplementary material shows seven examples of
 484 prior distributions on the logit and/or cloglog scales.

485 While a normal distribution with relatively large variance is often used as a weakly
 486 informative prior for continuous dependent variables, rows A and B in Figure 12 show that
 487 specifying such distributions on the logit and cloglog scales leads to rather informative
 488 distributions on the original probability scale, as most mass is pushed to probabilities of 0
 489 and 1. The other rows in Figure 12 show prior distributions on the logit and cloglog scale
 490 that we use instead.

491 **4.3.3 Model 1: A general specification of TIME, and main effects of
 492 congruency and trial number.** When you do not want to make assumptions about the
 493 shape of the hazard function in the selected baseline condition, or its shape is not smooth
 494 but irregular, then you can use a general specification of TIME, i.e., fit one intercept per
 495 time bin. In this first model, we use a general specification of TIME for the selected baseline
 496 condition (blank prime), and assume that the effects of prime-target congruency and trial
 497 number are proportional and additive, and that the effect of trial number is linear. Before we
 498 fit model 1, we remove unnecessary columns from the data, and specify our priors. In the
 499 code of Tutorial 2a, model M1 is specified as follows.

```
model_M1 <-
  brm(data = M1_data,
       family = binomial(link="cloglog"),
       event | trials(1) ~ 0 + d6 + d7 + d8 + Intercept + d10 + d11 + d12 + d13 + d14 + d15 +
              condition + trial_c +
              (d6 + d7 + d8 + 1 + d10 + d11 + d12 + d13 + d14 + d15 +
               condition + trial_c | pid),
       prior = priors_M1,
       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_M1")
```

500 After selecting the binomial family and the cloglog link, the model formula is specified.

501 The fixed effects include 9 dummy variables, the explicit Intercept variable (which represents
 502 bin 9 in this example), and the main effects of prime-target congruency (variable condition)
 503 and centered trial number (variable trial_c). Each of these effects is allowed to vary across
 504 individuals (variable pid). Estimating model M1 took about 70 minutes on a MacBook Pro
 505 (Sonoma 14.6.1 OS, 18GB Memory, M3 Pro Chip).

506 **4.3.4 Model 2: A polynomial specification of TIME, and main effects of
 507 congruency and trial number.** When the shape of the hazard function is rather smooth,
 508 as it is for behavioral RT data, one can fit a more parsimonious model by using a polynomial
 509 specification of TIME. For our second example model, we thus use a third-order polynomial
 510 specification of TIME for the selected baseline condition (blank prime), and again assume
 511 that the effects of prime-target congruency and centered trial number are proportional and
 512 additive, and that the effect of trial number is linear. The model formula for model M2 looks
 513 as follows.

```
event | trials(1) ~ 0 + Intercept + period_9 + I(period_9^2) + I(period_9^3) +
       condition + trial_c +
       (1 + period_9 + I(period_9^2) + I(period_9^3) +
       condition + trial_c | pid),
```

514 Because TIME is centered on bin 9, and trial number on trial 1000, the Intercept
 515 represents the cloglog-hazard in bin 9 for the blank prime condition in trial 1000. Estimating
 516 model M2 took about 2.5 hours.

517 **4.3.5 Model 3: A polynomial specification of TIME, and relaxing the
 518 proportionality assumption.** So far, we assumed that the effect of our predictors
 519 prime-target congruency and centered trial number are the same in each time bin. However,
 520 the descriptive plots (e.g., Figure 4) suggest that the effect of prime-target congruency varies
 521 across time bins. Previous research has shown that psychological effects typically change
 522 over time (Panis, 2020; Panis, Moran, et al., 2020; Panis & Schmidt, 2022; Panis et al., 2017;
 523 Panis & Wagemans, 2009). For the third model, we thus use a third-order polynomial
 524 specification of TIME for the baseline condition (blank prime), and relax the proportionality

525 assumption for the predictor variables prime-target congruency (variable condition) and
 526 centered trial number (variable trial_c).

```
event | trials(1) ~ 0 + Intercept +
      condition*period_9 +
      condition*I(period_9^2) +
      condition*I(period_9^3) +
      trial_c*period_9 +
      trial_c*I(period_9^2) +
      trial_c*I(period_9^3) +
      (1 + condition*period_9 +
      condition*I(period_9^2) +
      condition*I(period_9^3) +
      trial_c*period_9 +
      trial_c*I(period_9^2) +
      trial_c*I(period_9^3) | pid),
```

527 Note that duplicate terms in the model formula are ignored. Estimating model M3
 528 took about 4.5 hours.

529 **4.3.6 Model 4: A polynomial specification of TIME, and relaxing all three
 530 assumptions.** Based on previous work (e.g., Panis, 2020) we expect nonlinear effects of
 531 trial number on the hazard of response occurrence. We thus relax all three assumptions in
 532 model 4. We add a squared term for the continuous predictor centered trial number –
 533 $I(trial_c^2)$ – and include interaction terms. For example, how the effect of congruent primes
 534 changes across time bins within a trial might change across the trials within an experiment.

```
event | trials(1) ~ 0 + Intercept +
      condition*period_9*trial_c +
      condition*period_9*I(trial_c^2) +
      condition*I(period_9^2)*trial_c +
      condition*I(period_9^2)*I(trial_c^2) +
      condition*I(period_9^3) +
      trial_c*I(period_9^3) +
      (1 + condition*period_9*trial_c +
      condition*period_9*I(trial_c^2) +
      condition*I(period_9^2)*trial_c +
      condition*I(period_9^2)*I(trial_c^2) +
      condition*I(period_9^3) +
```

```
trial_c*I(period_9^3) | pid)
```

535 Again, duplicate terms in the model formula are ignored. Estimating model M4 took
 536 about 8 hours.

537 **4.3.7 Compare the models.** We can compare the four models using the Widely
 538 Applicable Information Criterion (WAIC) and Leave-One-Out (LOO) cross-validation, and
 539 look at model weights for both criteria (Kurz, 2023a; McElreath, 2018).

```
model_weights(model_M1, model_M2, model_M3, model_M4, weights = "loo") %>% round(digits = 3)
```

```
540 ## model_M1 model_M2 model_M3 model_M4
541 ##      0      0      0      1
542 ## model_M1 model_M2 model_M3 model_M4
543 ##      0      0      0      1
```

544 Clearly, both the loo and waic weighting schemes assign a weight of 1 to model M4,
 545 and a weight of 0 to the other three simpler models.

546 **4.3.8 Evaluate parameter estimates.** To make inferences from the parameter
 547 estimates in model M4, we summarize the draws from the posterior distributions of the
 548 effects of congruent and incongruent primes relative to the blank prime condition, in each
 549 time bin for trial numbers 500, 1000, and 1500, in terms of point and interval estimates.

550 Figure 6 shows one point (mean) and three highest posterior density interval
 551 (50/80/95%) estimates for the effects of congruent and incongruent primes relative to neutral
 552 primes, for each time bin in trial numbers 500, 1000, and 1500.

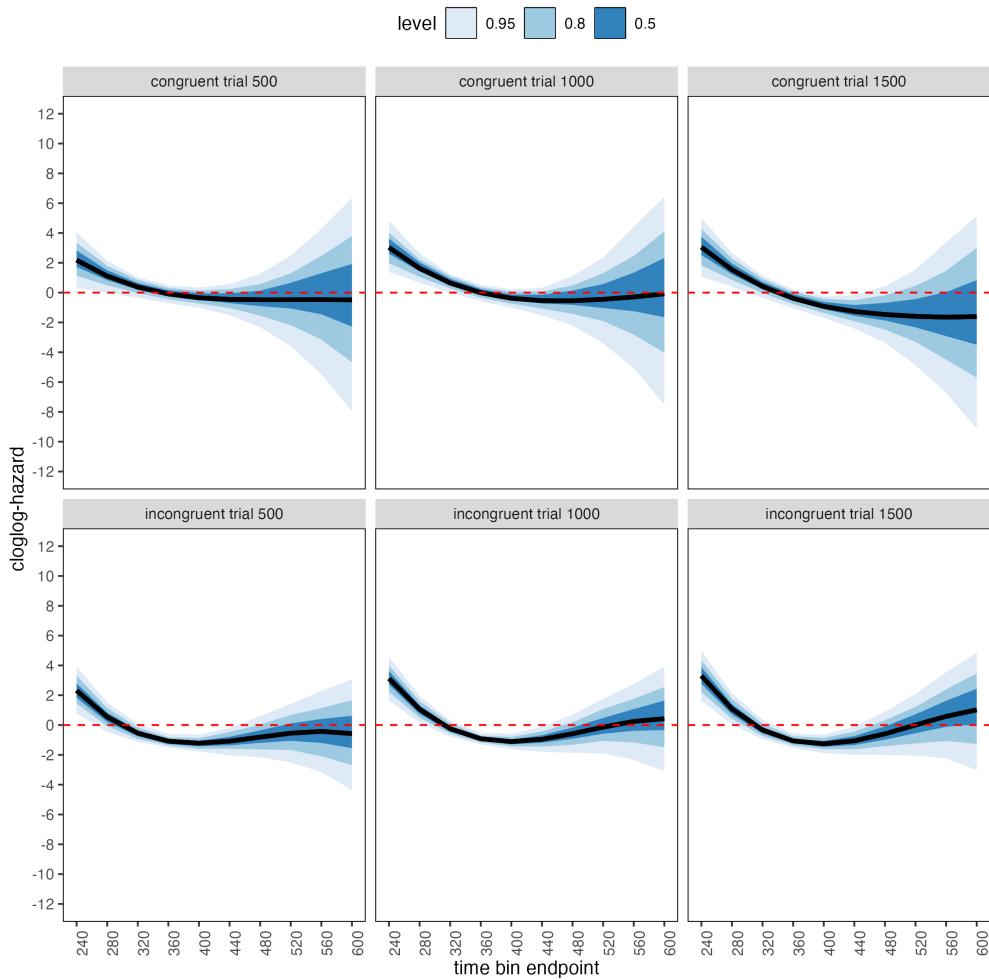


Figure 5. Means and 50/80/95% highest posterior density intervals of the draws from the posterior distributions representing the effect of congruent and incongruent primes relative to neutral primes in each bin for trial numbers 500, 1000, and 1500.

Table 4 shows the summaries of the draws from the posterior distributions of the effects of congruent and incongruent primes relative to the blank prime condition in trials 500, 1000, and 1500, in terms of a point estimate (the mean) and the upper and lower bounds of the 95% highest posterior density interval. Also, by exponentiating the mean we obtain an effect size in terms of a hazard ratio. For simplicity and ease of presentation, we only tabulate data for a subset of the design in the main text (trial 500 for congruent and incongruent conditions). For the full table, see Supplementary materials or Tutorial X.

560 Based on Figure 6 and Supplementary Table XX, we see that at the beginning of the
561 experiment (trial 500), congruent and incongruent primes have a positive effect in time bin
562 (200,240] on cloglog-hazard, relative to the cloglog-hazard estimate in the baseline condition
563 (no prime; red striped lines in Figure 6). For example, the hazard ratio shows that the
564 hazard of response occurrence for congruent primes is estimated to be 8.82 times higher than
565 that for no-prime trials in bin (200,240] of trial 500. Incongruent primes also have a negative
566 effect on cloglog-hazard in bins (320,360], (360,400], and (400,440]. For example, in bin
567 (320,360], the hazard ratio shows that the hazard of response occurrence for incongruent
568 prime is estimated to be .34 times smaller than that for no-prime trials. While the early
569 positive effects reflect responses to the prime stimulus, the later negative effect for
570 incongruent primes likely reflects response competition between the prime-triggered response
571 (e.g., left) and the target-triggered response (e.g., right)

572 In the middle of the experiment (trial 1000), both congruent and incongruent primes
573 have positive effects in bins (200,240] and (240,280], while incongruent primes again have
574 negative effects in bins (320,360], (360,400], and (400,440]. Probably due to practicing
575 stimulus-response associations, the primes generate a higher hazard of response occurrence
576 for 80 ms early in a trial (compared to 40 ms at the beginning of the experiment) compared
577 to the blank prime condition.

578 Towards the end of the experiment (trial 1500), both congruent and incongruent primes
579 have positive and negative effects. Positive effects are present in bins (200,240] and (240,280].
580 Incongruent primes again have negative effects in bins (320,360], (360,400], and (400,440],
581 and congruent primes now also have negative effects in bins (360,400] and (400,440].

582 These results show that the effect of prime-target congruency changes not only on the
583 across-bin/within-trial time scale (variable period_9), but also on the
584 across-trial/within-experiment time scale (variable trial_c). The fact that congruent primes
585 generate negative effects for 80 ms (compared to no-prime trials) towards the end of the

586 experiment, while incongruent primes generate negative effects for 120 ms throughout the
587 experiment, suggests the involvement of separate cognitive processes.

588 Panis and Schmidt (2016) distinguished between automatic response competition
589 (bottom-up lateral inhibition between response channels), active and global inhibition
590 (top-down nonselective response inhibition), and active and selective inhibition (top-down
591 selective response inhibition). While automatic response competition can be expected to be
592 present in the incongruent trials throughout the experiment, active and global response
593 inhibition effects might be present in both congruent and incongruent (unmasked) prime
594 trials. In other words, people learn that the prime-triggered response is premature and that
595 they have to temporarily slow down (increase the global response threshold) in order to allow
596 gating of the correct response to the target stimulus. Thus, it seems that this global
597 inhibitory effect becomes visible in the congruent (compared to no-prime) trials towards the
598 end of the experiment, while it might be masked by the automatic inhibitory effect of
599 response competition in the incongruent trials. Interestingly, while Panis and Schmidt (2016)
600 did not test interactions between prime-target congruency and trial number, they concluded
601 that active (i.e., top-down) response inhibition starts around 360 ms after the onset of the
602 second stimulus (the target stimulus in no-mask trials), which nicely coincides with the onset
603 of the negative effect of congruent primes observed here in trial 1500.

604 To conclude this Tutorial 2a, Figure 7 shows the model-based hazard functions for each
605 prime type for participant 6, in trials 500, 1000, and 1500.

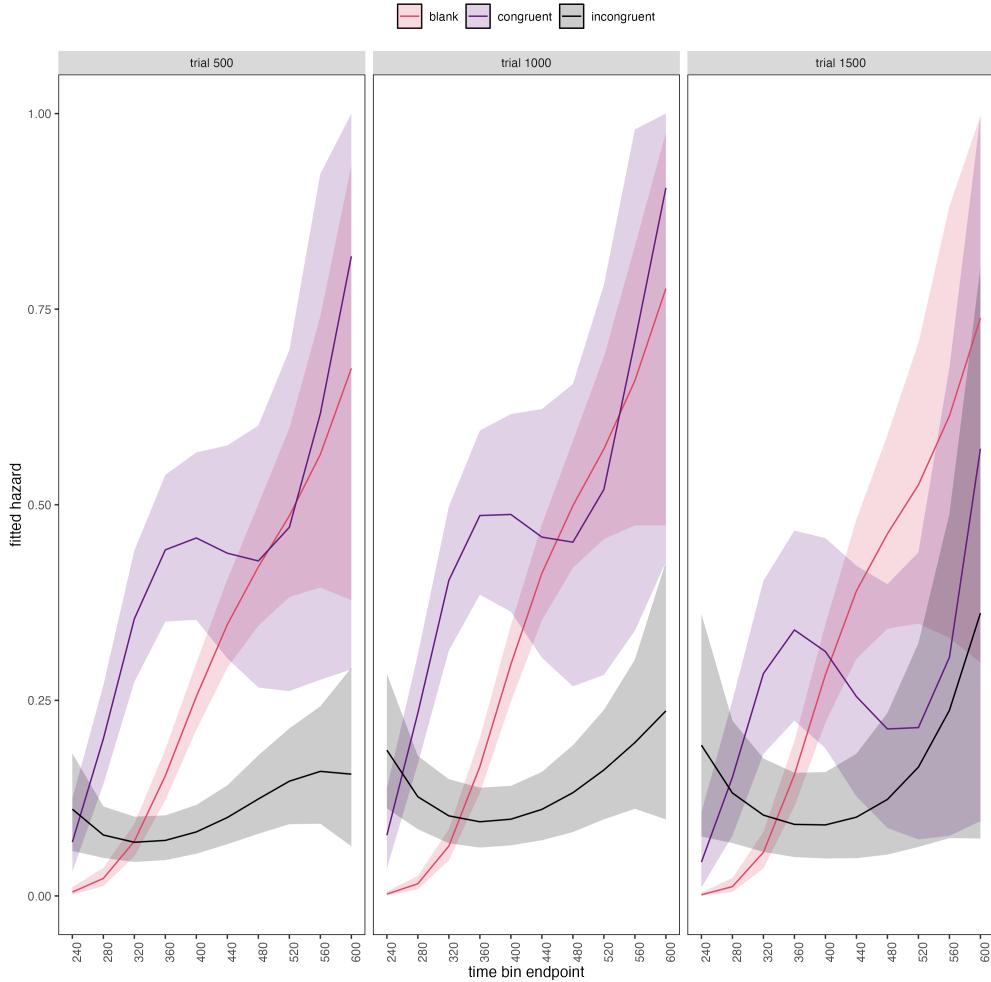


Figure 6. Model-based hazard functions for each prime type for participant 6 in trials 500, 1000, and 1500.

606 4.4 Tutorial 2b: Fitting Bayesian conditional accuracy models

607 In this fourth tutorial, we illustrate how to fit a Bayesian multi-level regression model
 608 to the timed accuracy data from the masked response priming data set used in Tutorial 1a.
 609 The general process is similar to Tutorial 2a, except that (a) we use the person-trial data set,
 610 (b) we use the logit link function, and (c) we change the priors. For illustration purposes, we
 611 only fitted the effects of model M4 (see Tutorial 2a) in the conditional accuracy model called
 612 M4_ca.

613 To make inferences from the parameter estimates in model M4_ca, we summarize the
 614 draws from the posterior distributions of the effects of congruent and incongruent primes on
 615 logit-ca relative to the blank prime condition, in each time bin for trial numbers 500, 1000,
 616 and 1500, in terms of point and interval estimates.

617 Figure 8 shows one point (mean) and three highest posterior density interval
 618 (50/80/95%) estimates for the effects of congruent and incongruent primes relative to neutral
 619 primes on logit-ca, for each time bin in trial numbers 500, 1000, and 1500.

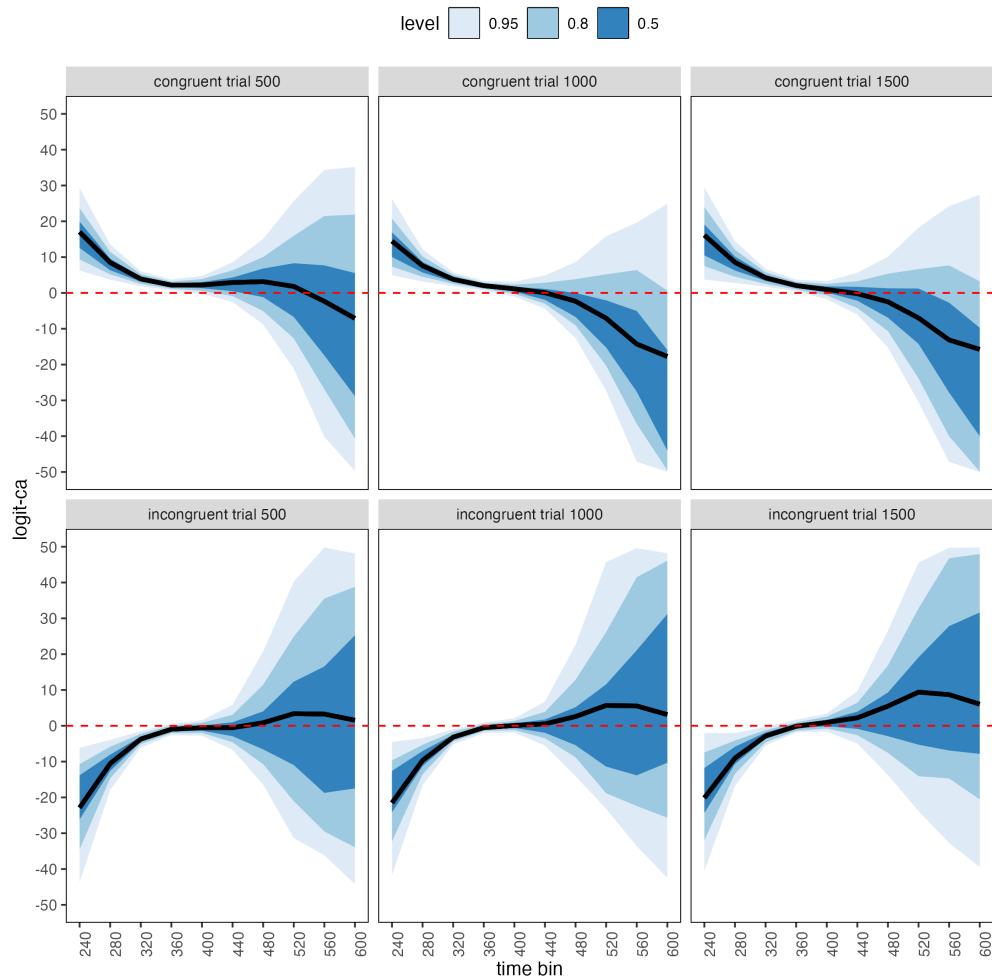


Figure 7. Means and 50/80/95% highest posterior density intervals of the draws from the posterior distributions representing the effect of congruent and incongruent primes relative to neutral primes in each bin for trial numbers 500, 1000, and 1500.

620 Supplementary Table XX shows the summaries of the draws from the posterior
621 distributions of the effects of congruent and incongruent primes relative to the blank prime
622 condition in trials 500, 1000, and 1500, in terms of a point estimate (the mean) and the
623 upper and lower bounds of the 95% highest posterior density interval. Also, by
624 exponentiating the mean we obtain an effect size in terms of an odds ratio.

625 Based on Figure 8 and Supplementary Table XX, we see that throughout the
626 experiment (trials 500, 1000, and 1500), congruent primes have a positive effect on logit-ca(t)
627 in time bins (200,240], (240,280], (280,320], and (320,360], relative to the logit-ca(t)
628 estimates in the baseline condition (blank prime; red dashed lines in Figure 8). For example,
629 the odds ratio for congruent primes in bin (320,360] in trial 500 shows that the odds of a
630 correct response are estimated to be 8.89 times higher than the odds of a correct response
631 when there is no prime. Incongruent primes have a negative effect on logit-ca(t) in time bins
632 (200,240], (240,280], and (280,320] throughout the experiment, relative to the logit-ca(t)
633 estimates in the baseline condition (no prime; red striped lines).

634 To conclude this Tutorial 2b, Figure 9 shows the model-based ca(t) functions for each
635 prime type for participant 6, in trials 500, 1000, and 1500.

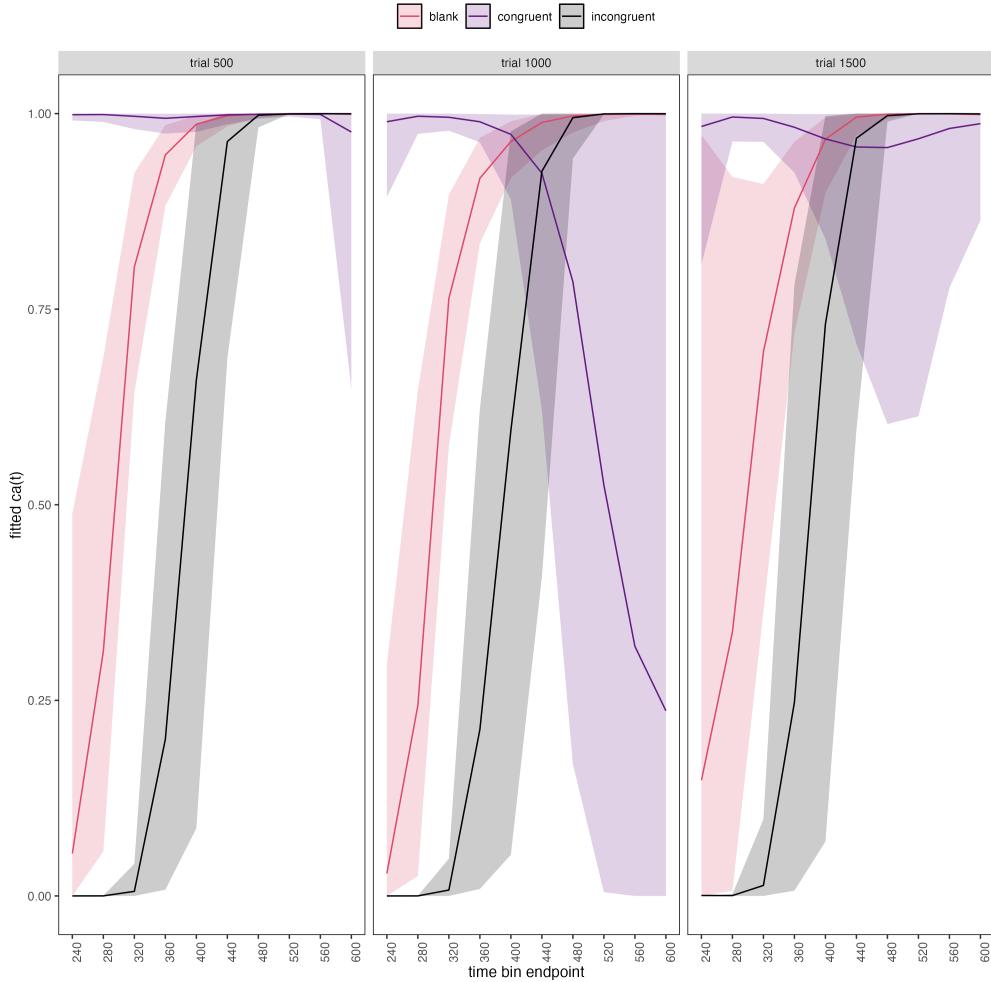


Figure 8. Model-based $ca(t)$ functions for each prime type for participant 6 in trials 500, 1000, and 1500.

636 4.5 Tutorial 3a: Fitting Frequentist hazard models

637 In this fifth tutorial we illustrate how to fit a multilevel regression model to RT data in
 638 the frequentist framework, for the data set used in Tutorial 1a. The general process is similar
 639 to that in Tutorial 2a, except that there are no priors to set. For illustration purposes, we
 640 only fitted the effects from model M3 (see Tutorial 2a) using the function `glmer()` from the R
 641 package `lme4`. Alternatively, one could also use the function `glmmPQL()` from the R package
 642 `MASS` (Ripley et al., 2024). The resulting hazard model is called `M3_f`.

643 In Figure 10 we compare the parameter estimates of model M3 from `brm()` with those
 644 of model M3_f from `glmer()`.

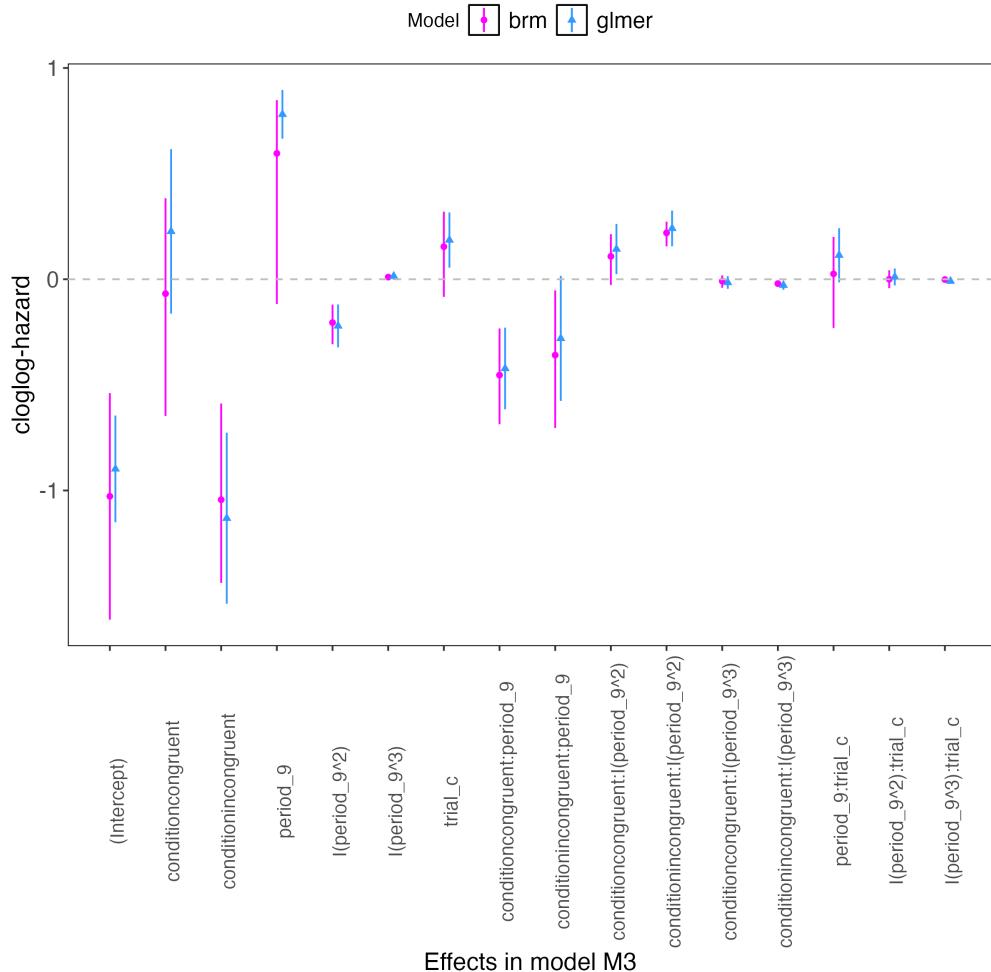


Figure 9. Parameter estimates for model M3 from `brm()` – means and 95% quantile intervals – and model M3_f from `glmer()` – maximum likelihood estimates and 95% confidence intervals.

645 Figure 10 confirms that the parameter estimates from both Bayesian and frequentist
 646 models are pretty similar,, which makes sense given the close similarity in model structure.
 647 However, the random effects structure of model M3_f was already too complex for the
 648 frequentist model as it did not converge and resulted in a singular fit. This is of course one
 649 of the reasons why Bayesian modeling has become so popular in recent years. But the price
 650 you pay for being able to fit more complex random effects models in a Bayesian framework is

651 computation time. In other words, as we have noted throughout, some of the Bayesian
652 models in Tutorials 2a and 2b took several hours to build.

653 **4.6 Tutorial 3b: Fitting Frequentist conditional accuracy models**

654 In this sixth tutorial we illustrate how to fit a multilevel regression model to the timed
655 accuracy data in the frequentist framework, for the data set used in Tutorial 1a. For
656 illustration purposes, we only fitted the effects from model M3 (see Tutorial 2a) using the
657 function `glmer()` from the R package `lme4`. Alternatively, one could also use the function
658 `glmmPQL()` from the R package `MASS` (Ripley et al., 2024). Again, the resulting
659 conditional accuracy model `M3_ca_f` did not converge and resulted in a singular fit.

660 **4.4 Tutorial 4: Planning**

661 This needs adding... RR to do this.

662 **5. Discussion**

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

675 following, we discuss issues relating to model complexity and interpretability, individual
676 differences, as well as limitations of the approach and future extensions.

677 **5.1 What are the main use-cases of EHA for understanding cognition and brain
678 function?**

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

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

695 **5.2 Model complexity versus interpretability**

696 EHA can quickly become very complex when adding more than 1 time scale, due to
697 the many possible higher-order interactions. For example, model M4 contains two time
698 scales as covariates: the passage of time on the within-trial time scale, and the passage of
699 time on the across-trial (or within-experiment) time scale. However, when trials are

700 presented in blocks, and blocks of trials within sessions, and when the experiment comprises
701 three sessions, then four time scales can be defined (within-trial, within-block, within-session,
702 and within-experiment). From a theoretical perspective, adding more than 1 time scale –
703 and their interactions – can be important to capture plasticity and other learning effects that
704 may play out on such longer time scales, and that are probably present in each experiment
705 in general. From a practical perspective, therefore, some choices need to be made to balance
706 the amount of data that is being collected per participant, condition and across the varying
707 timescales. As one example, if there are several timescales of relevance, then it might be
708 prudent for interpretational purposes to limit the number of experimental predictor variables
709 (conditions). This is of course where planning and data simulation efforts would be
710 important to provide a guide to experimental design choices.

711 **5.3 Individual differences**

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

725 But even when using a response deadline, participants can differ qualitatively in the

726 effects they display (see Panis, 2020). One way to deal with this is to describe and interpret
727 the different patterns. Another way is to run a clustering algorithm on the individual hazard
728 estimates across all conditions. The obtained dendrogram can then be used to identify a
729 (hopefully big) cluster of participants that behave similarly, and to identify a (hopefully
730 small) cluster of participants with different behavioral patterns. One might then exclude the
731 smaller sub-group of participants before fitting a hazard model or consider the possibility
732 that different cognitive processes may be at play during task performance across the different
733 sub-groups.

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

740 5.4 Limitations

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

745 Another issue is that you need a relatively large number of trials per condition to
746 estimate the hazard function with high temporal resolution, which is required when testing
747 predictions of process models of cognition. Indeed, in general, there is a trade-off between
748 the number of trials per condition and the temporal resolution (i.e., bin width) of the hazard
749 function. Therefore, we recommend researchers to collect as many trials as possible per
750 experimental condition, given the available resources and considering the participant

751 experience (e.g., fatigue and boredom). For instance, if the maximum session length deemed
752 reasonable is between 1 and 2 hours, what is the maximum number of trials per condition
753 that you could reasonably collect? After consideration, it might be worth conducting
754 multiple testing sessions per participant and/or reducing the number of experimental
755 conditions. Finally, there is a user-friendly online tool for calculating statistical power as a
756 function of the number of trials as well as the number of participants, and this might be
757 worth consulting to guide the research design process (Baker et al., 2021).

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

766 **5.5 Extensions**

767 The hazard models in this tutorial assume that there is one event of interest. For RT
768 data, this event constitutes a single transition between an “idle” state and a “responded”
769 state. However, in certain situations, more than one event of interest might exist. For
770 example, in a medical or health-related context, an individual might transition back and
771 forth between a “healthy” state and a “depressed” state, before being absorbed into a final
772 “death” state. When you have data on the timing of these transitions, one can apply
773 multi-state hazard models, which generalize event history analysis to transitions between
774 three or more states (Steele, Goldstein, & Browne, 2004). Also, the predictor variables in
775 this tutorial are time-invariant, i.e., their value did not change over the course of a trial.
776 Thus, another extension is to include time-varying predictors, i.e., predictors whose value can

777 change across the time bins within a trial (Allison, 2010). For example, when gaze position is
778 tracked during a visual search trial, the gaze-target distance will vary during a trial when the
779 eyes move around before a manual response is given; shorter gaze-target distances should be
780 associated with a higher hazard of response occurrence. Note that the effect of a
781 time-varying predictor (e.g., an occipital EEG signal) can itself vary over time.

782 **6. Conclusions**

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

794

References

- 795 Allison, P. D. (1982). Discrete-Time Methods for the Analysis of Event Histories.
796 *Sociological Methodology*, 13, 61. <https://doi.org/10.2307/270718>
- 797 Allison, P. D. (2010). *Survival analysis using SAS: A practical guide* (2. ed.). Cary, NC: SAS
798 Press.
- 799 Aust, F. (2019). *Citr: 'RStudio' add-in to insert markdown citations*. Retrieved from
800 <https://github.com/crsh/citr>
- 801 Aust, F., & Barth, M. (2023). *papaja: Prepare reproducible APA journal articles with R
802 Markdown*. Retrieved from <https://github.com/crsh/papaja>
- 803 Baker, D. H., Vilidaite, G., Lygo, F. A., Smith, A. K., Flack, T. R., Gouws, A. D., &
804 Andrews, T. J. (2021). Power contours: Optimising sample size and precision in
805 experimental psychology and human neuroscience. *Psychological Methods*, 26(3),
806 295–314. <https://doi.org/10.1037/met0000337>
- 807 Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for
808 confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*,
809 68(3), 10.1016/j.jml.2012.11.001. <https://doi.org/10.1016/j.jml.2012.11.001>
- 810 Barth, M. (2023). *tinylabes: Lightweight variable labels*. Retrieved from
811 <https://cran.r-project.org/package=tinylabes>
- 812 Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models
813 using lme4. *Journal of Statistical Software*, 67(1), 1–48.
814 <https://doi.org/10.18637/jss.v067.i01>
- 815 Bates, D., Maechler, M., & Jagan, M. (2024). *Matrix: Sparse and dense matrix classes and
816 methods*. Retrieved from <https://CRAN.R-project.org/package=Matrix>
- 817 Bengtsson, H. (2021). A unifying framework for parallel and distributed processing in r using
818 futures. *The R Journal*, 13(2), 208–227. <https://doi.org/10.32614/RJ-2021-048>
- 819 Blossfeld, H.-P., & Rohwer, G. (2002). *Techniques of event history modeling: New
820 approaches to causal analysis*, 2nd ed (pp. x, 310). Mahwah, NJ, US: Lawrence Erlbaum

- 821 Associates Publishers.
- 822 Box-Steffensmeier, J. M. (2004). Event history modeling: A guide for social scientists.
- 823 Cambridge: University Press.
- 824 Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan.
- 825 *Journal of Statistical Software*, 80(1), 1–28. <https://doi.org/10.18637/jss.v080.i01>
- 826 Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms.
- 827 *The R Journal*, 10(1), 395–411. <https://doi.org/10.32614/RJ-2018-017>
- 828 Bürkner, P.-C. (2021). Bayesian item response modeling in R with brms and Stan. *Journal*
- 829 *of Statistical Software*, 100(5), 1–54. <https://doi.org/10.18637/jss.v100.i05>
- 830 Eddelbuettel, D., & Balamuta, J. J. (2018). Extending R with C++: A Brief Introduction
- 831 to Rcpp. *The American Statistician*, 72(1), 28–36.
- 832 <https://doi.org/10.1080/00031305.2017.1375990>
- 833 Eddelbuettel, D., & François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal of*
- 834 *Statistical Software*, 40(8), 1–18. <https://doi.org/10.18637/jss.v040.i08>
- 835 Gabry, J., Češnovar, R., Johnson, A., & Broder, S. (2024). *Cmdstanr: R interface to*
- 836 '*CmdStan*'. Retrieved from <https://github.com/stan-dev/cmdstanr>
- 837 Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2019). Visualization in
- 838 bayesian workflow. *J. R. Stat. Soc. A*, 182, 389–402. <https://doi.org/10.1111/rssa.12378>
- 839 Girard, J. (2024). *Standist: What the package does (one line, title case)*. Retrieved from
- 840 <https://github.com/jmgirard/standist>
- 841 Graelund, G., & Wickham, H. (2011). Dates and times made easy with lubridate. *Journal*
- 842 *of Statistical Software*, 40(3), 1–25. Retrieved from <https://www.jstatsoft.org/v40/i03/>
- 843 Holden, J. G., Van Orden, G. C., & Turvey, M. T. (2009). Dispersion of response times
- 844 reveals cognitive dynamics. *Psychological Review*, 116(2), 318–342.
- 845 <https://doi.org/10.1037/a0014849>
- 846 Hosmer, D. W., Lemeshow, S., & May, S. (2011). *Applied Survival Analysis: Regression*
- 847 *Modeling of Time to Event Data* (2nd ed). Hoboken: John Wiley & Sons.

- 848 Ince, R. A., Paton, A. T., Kay, J. W., & Schyns, P. G. (2021). Bayesian inference of
849 population prevalence. *eLife*, 10, e62461. <https://doi.org/10.7554/eLife.62461>
- 850 Kantowitz, B. H., & Pachella, R. G. (2021). The Interpretation of Reaction Time in
851 Information-Processing Research 1. *Human Information Processing*, 41–82.
852 <https://doi.org/10.4324/9781003176688-2>
- 853 Kay, M. (2023). *tidybayes: Tidy data and geoms for Bayesian models*.
854 <https://doi.org/10.5281/zenodo.1308151>
- 855 Kelso, J. A. S., Dumas, G., & Tognoli, E. (2013). Outline of a general theory of behavior
856 and brain coordination. *Neural Networks: The Official Journal of the International
857 Neural Network Society*, 37, 120–131. <https://doi.org/10.1016/j.neunet.2012.09.003>
- 858 Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing,
859 estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic
860 Bulletin & Review*, 25(1), 178–206. <https://doi.org/10.3758/s13423-016-1221-4>
- 861 Kurz, A. S. (2023a). *Applied longitudinal data analysis in brms and the tidyverse* (version
862 0.0.3). Retrieved from <https://bookdown.org/content/4253/>
- 863 Kurz, A. S. (2023b). *Statistical rethinking with brms, ggplot2, and the tidyverse: Second
864 edition* (version 0.4.0). Retrieved from <https://bookdown.org/content/4857/>
- 865 Landes, J., Engelhardt, S. C., & Pelletier, F. (2020). An introduction to event history
866 analyses for ecologists. *Ecosphere*, 11(10), e03238. <https://doi.org/10.1002/ecs2.3238>
- 867 Luce, R. D. (1991). *Response times: Their role in inferring elementary mental organization*
868 (1. issued as paperback). Oxford: Univ. Press.
- 869 McElreath, R. (2018). *Statistical Rethinking: A Bayesian Course with Examples in R and
870 Stan* (1st ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9781315372495>
- 871 Meyer, D. E., Osman, A. M., Irwin, D. E., & Yantis, S. (1988). Modern mental chronometry.
872 *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)
- 873 Müller, K., & Wickham, H. (2023). *Tibble: Simple data frames*. Retrieved from
874 <https://CRAN.R-project.org/package=tibble>

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

- 902 Pedersen, T. L. (2024). *Patchwork: The composer of plots*. Retrieved from
903 <https://CRAN.R-project.org/package=patchwork>
- 904 Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in s and s-PLUS*. New York:
905 Springer. <https://doi.org/10.1007/b98882>
- 906 R Core Team. (2024). *R: A language and environment for statistical computing*. Vienna,
907 Austria: R Foundation for Statistical Computing. Retrieved from
908 <https://www.R-project.org/>
- 909 Ripley, B., Venables, B., Bates, D. M., ca 1998), K. H. (partial. port, ca 1998), A. G.
910 (partial. port, & polr), D. F. (support. functions for. (2024). *MASS: Support Functions*
911 *and Datasets for Venables and Ripley's MASS*.
- 912 Singer, J. D., & Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling Change*
913 *and Event Occurrence*. Oxford, New York: Oxford University Press.
- 914 Smith, P. L., & Little, D. R. (2018). Small is beautiful: In defense of the small-N design.
915 *Psychonomic Bulletin & Review*, 25(6), 2083–2101.
916 <https://doi.org/10.3758/s13423-018-1451-8>
- 917 Steele, F., Goldstein, H., & Browne, W. (2004). A general multilevel multistate competing
918 risks model for event history data, with an application to a study of contraceptive use
919 dynamics. *Statistical Modelling*, 4(2), 145–159.
920 <https://doi.org/10.1191/1471082X04st069oa>
- 921 Teachman, J. D. (1983). Analyzing social processes: Life tables and proportional hazards
922 models. *Social Science Research*, 12(3), 263–301.
923 [https://doi.org/10.1016/0049-089X\(83\)90015-7](https://doi.org/10.1016/0049-089X(83)90015-7)
- 924 Townsend, J. T. (1990). Truth and consequences of ordinal differences in statistical
925 distributions: Toward a theory of hierarchical inference. *Psychological Bulletin*, 108(3),
926 551–567. <https://doi.org/10.1037/0033-2909.108.3.551>
- 927 Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics.
928 *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)

- 929 Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York.
930 Retrieved from <https://ggplot2.tidyverse.org>
- 931 Wickham, H. (2023a). *Forcats: Tools for working with categorical variables (factors)*.
932 Retrieved from <https://CRAN.R-project.org/package=forcats>
- 933 Wickham, H. (2023b). *Stringr: Simple, consistent wrappers for common string operations*.
934 Retrieved from <https://CRAN.R-project.org/package=stringr>
- 935 Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., ... Yutani,
936 H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686.
937 <https://doi.org/10.21105/joss.01686>
- 938 Wickham, H., Çetinkaya-Rundel, M., & Grolemund, G. (2023). *R for data science: Import,
939 tidy, transform, visualize, and model data* (2nd edition). Beijing Boston Farnham
940 Sebastopol Tokyo: O'Reilly.
- 941 Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *Dplyr: A grammar
942 of data manipulation*. Retrieved from <https://CRAN.R-project.org/package=dplyr>
- 943 Wickham, H., & Henry, L. (2023). *Purrr: Functional programming tools*. Retrieved from
944 <https://CRAN.R-project.org/package=purrr>
- 945 Wickham, H., Hester, J., & Bryan, J. (2024). *Readr: Read rectangular text data*. Retrieved
946 from <https://CRAN.R-project.org/package=readr>
- 947 Wickham, H., Vaughan, D., & Girlich, M. (2024). *Tidyr: Tidy messy data*. Retrieved from
948 <https://CRAN.R-project.org/package=tidyr>
- 949 William Matthew Makeham. (1860). *On the Law of Mortality and the Construction of
950 Annuity Tables*. The Assurance Magazine, and Journal of the Institute of Actuaries.
- 951 Wolkersdorfer, M. P., Panis, S., & Schmidt, T. (2020). Temporal dynamics of sequential
952 motor activation in a dual-prime paradigm: Insights from conditional accuracy and
953 hazard functions. *Attention, Perception, & Psychophysics*, 82(5), 2581–2602.
954 <https://doi.org/10.3758/s13414-020-02010-5>

955 **Supplementary material**

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

958 The shape of a distribution of waiting times can be described in multiple ways (Luce,
959 1991). After dividing time in discrete, contiguous time bins indexed by t , let RT be a
960 discrete random variable denoting the rank of the time bin in which a particular person's
961 response occurs in a particular experimental condition. Because waiting times can only
962 increase, discrete-time EHA focuses on the discrete-time hazard function

963
$$h(t) = P(RT = t \mid RT \geq t) \quad (1)$$

964 and the discrete-time survivor function

965
$$S(t) = P(RT > t) = [1-h(t)].[1-h(t-1)].[1-h(t-2)] \dots [1-h(1)] \quad (2)$$

966 and not on the probability mass function

967
$$P(t) = P(RT = t) = h(t).S(t-1) \quad (3)$$

968 nor the cumulative distribution function

969
$$F(t) = P(RT \leq t) = 1-S(t) \quad (4)$$

970 The discrete-time hazard function of event occurrence gives you for each bin the
971 probability that the event occurs (sometime) in that bin, given that the event has not
972 occurred yet in previous bins. While the discrete-time hazard function assesses the unique
973 risk of event occurrence associated with each time bin, the discrete-time survivor function
974 cumulates the bin-by-bin risks of event *non*occurrence to obtain the probability that the
975 event occurs after bin t . The probability mass function cumulates the risk of event occurrence
976 in bin t with the risks of event nonoccurrence in bins 1 to $t-1$. From equation 3 we find that
977 hazard in bin t is equal to $P(t)/S(t-1)$.

978 For two-choice RT data, the discrete-time hazard function can be extended with the

979 discrete-time conditional accuracy function

980 $ca(t) = P(\text{correct} \mid RT = t)$ (5)

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

984 The survivor function provides a context for the hazard function, as $S(t-1) = P(RT >$
 985 $t-1) = P(RT \geq t)$ tells you on how many percent of the trials the estimate $h(t) = P(RT = t \mid$
 986 $RT \geq t)$ is based. The probability mass function provides a context for the conditional
 987 accuracy function, as $P(t) = P(RT = t)$ tells you on how many percent of the trials the
 988 estimate $ca(t) = P(\text{correct} \mid RT = t)$ is based.

989 When time is treated as a continuous variable, let RT be a continuous random variable
 990 denoting a particular person's response time in a particular experimental condition.

991 Continuous-time EHA does not focus on the cumulative distribution function $F(t) = P(RT$
 992 $\leq t)$ and its derivative, the probability density function $f(t) = F(t)'$, but on the survivor
 993 function $S(t) = P(RT > t)$ and the hazard rate function $\lambda(t) = f(t)/S(t)$. The hazard rate
 994 function gives you the instantaneous *rate* of event occurrence at time point t , given that the
 995 event has not occurred yet.

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

997 We defined 13 custom functions that we list here.

- 998 • `censor(df,timeout,bin_width)` : divide the time segment $(0,timeout]$ in bins, identify
 999 any right-censored observations, and determine the discrete RT (time bin rank)
- 1000 • `ptb(df)` : transform the person-trial data set to the person-trial-bin data set
- 1001 • `setup_lt(ptb)` : set up a life table for each level of 1 independent variable

- 1002 • setup_lt_2IV(ptb) : set up a life table for each combination of levels of 2 independent variables
- 1003
- 1004 • calc_ca(df) : estimate the conditinal accuracies when there is 1 independent variable
- 1005 • calc_ca_2IV(df) : estimate the conditional accuraiies when there are 2 independent variables
- 1006
- 1007 • join_lt_ca(df1,df2) : add the ca(t) estimates to the life tables (1 independent variable)
- 1008 • join_lt_ca_2IV(df1,df2) : add the ca(t) estimates to the life tables (2 independent variables)
- 1009
- 1010 • extract_median(df) : estimate quantiles $S(t)._{50}$ (1 independent variable)
- 1011 • extract_median_2IV(df) : estimate quantiles $S(t)._{50}$ (2 independent variables)
- 1012 • plot_oha(df,subj,haz_yaxis) : create plots of the discrete-time functions (1 independent variable)
- 1013
- 1014 • plot_oha_2IV(df,subj,haz_yaxis) : create plots of the discrete-time functions (2 independent variables)
- 1015
- 1016 • plot_oha_agg(df,subj,haz_yaxis) : create 1 plot for data aggregated across participants (1 independent variable)
- 1017

1018 When you want to analyse simple RT data from a detection experiment with one
 1019 independent variable, the functions calc_ca() and join_lt_ca() should not be used, and the
 1020 code to plot the conditional accuracy functions should be removed from the function
 1021 plot_oha(). When you want to analyse simple RT data from a detection experiment with
 1022 two independent variables, the functions calc_ca_2IV() and join_lt_ca_2IV() should not
 1023 be used, and the code to plot the conditional accuracy functions should be removed from the
 1024 function plot_oha_2IV().

1025 C. Link functions

1026 Popular link functions include the logit link and the complementary log-log link, as
 1027 shown in Figure 11.

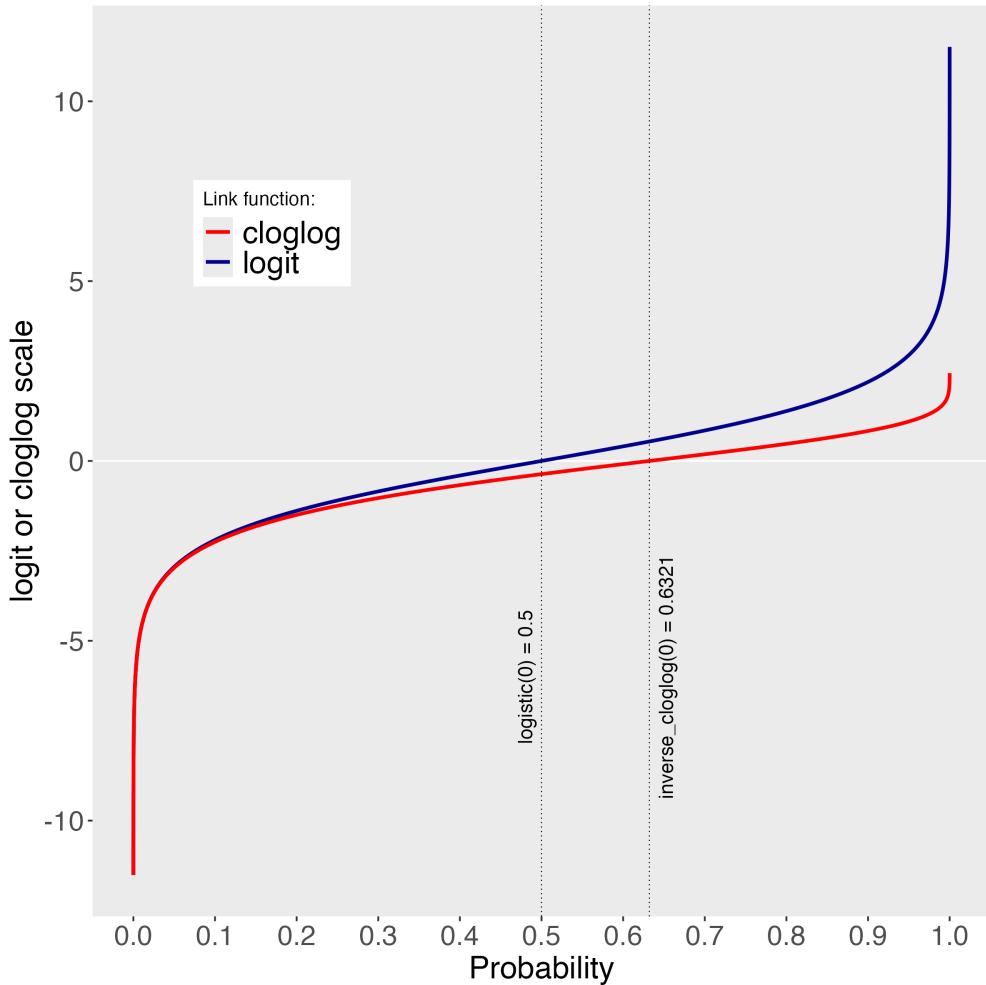


Figure 10. The logit and cloglog link functions.

1028 D. Regression equations

1029 An example (single-level) discrete-time hazard model with three predictors (TIME, X_1 ,
 1030 X_2), the cloglog link function, and a third-order polynomial specification for TIME can be
 1031 written as follows:

1032 $\text{cloglog}[h(t)] = \ln(-\ln[1-h(t)]) = [\beta_0 \text{ONE} + \beta_1(\text{TIME}-9) + \beta_2(\text{TIME}-9)^2] + [\beta_3 X_1 + \beta_4 X_2 +$
 1033 $\beta_5 X_2(\text{TIME}-9)]$ (6)

1034 The main predictor variable TIME is the time bin index t that is centered on value 9

in this example. The first set of terms within brackets, the parameters β_0 to β_2 multiplied by their polynomial specifications of (centered) time, represents the shape of the baseline cloglog-hazard function (i.e., when all predictors X_i take on a value of zero). The second set of terms (the beta parameters β_3 to β_5) represents the vertical shift in the baseline cloglog-hazard for a 1 unit increase in the respective predictor variable. Predictors can be discrete, continuous, and time-varying or time-invariant. For example, the effect of a 1 unit increase in X_1 is to vertically shift the whole baseline cloglog-hazard function by β_1 cloglog-hazard units. However, if the predictor interacts linearly with TIME (see X_2 in the example), then the effect of a 1 unit increase in X_2 is to vertically shift the predicted cloglog-hazard in bin 9 by β_2 cloglog-hazard units (when $TIME-9 = 0$), in bin 10 by $\beta_2 + \beta_3$ cloglog-hazard units (when $TIME-9 = 1$), and so forth. To interpret the effects of a predictor, its β parameter is exponentiated, resulting in a hazard ratio (due to the use of the cloglog link). When using the logit link, exponentiating a β parameter results in an odds ratio.

An example (single-level) discrete-time hazard model with a general specification for TIME (separate intercepts for each of six bins, where D1 to D6 are binary variables identifying each bin) and a single predictor (X_1) can be written as follows:

$$\text{cloglog}[h(t)] = [\beta_0D1 + \beta_1D2 + \beta_2D3 + \beta_3D4 + \beta_4D5 + \beta_5D6] + [\beta_6X_1] \quad (7)$$

1052 E. Prior distributions

To gain a sense of what prior *logit* values would approximate a uniform distribution on the probability scale, Kurz (2023a) simulated a large number of draws from the Uniform(0,1) distribution, converted those draws to the log-odds metric, and fitted a Student's t distribution. Row C in Figure 12 shows that using a t-distribution with 7.61 degrees of freedom and a scale parameter of 1.57 as a prior on the logit scale, approximates a uniform distribution on the probability scale. According to Kurz (2023a), such a prior might be a good prior for the intercept(s) in a logit-hazard model, while the N(0,1) prior in row D might

1060 be a good prior for the non-intercept parameters in a logit-hazard model, as it gently
 1061 regularizes p towards .5 (i.e., a zero effect on the logit scale).

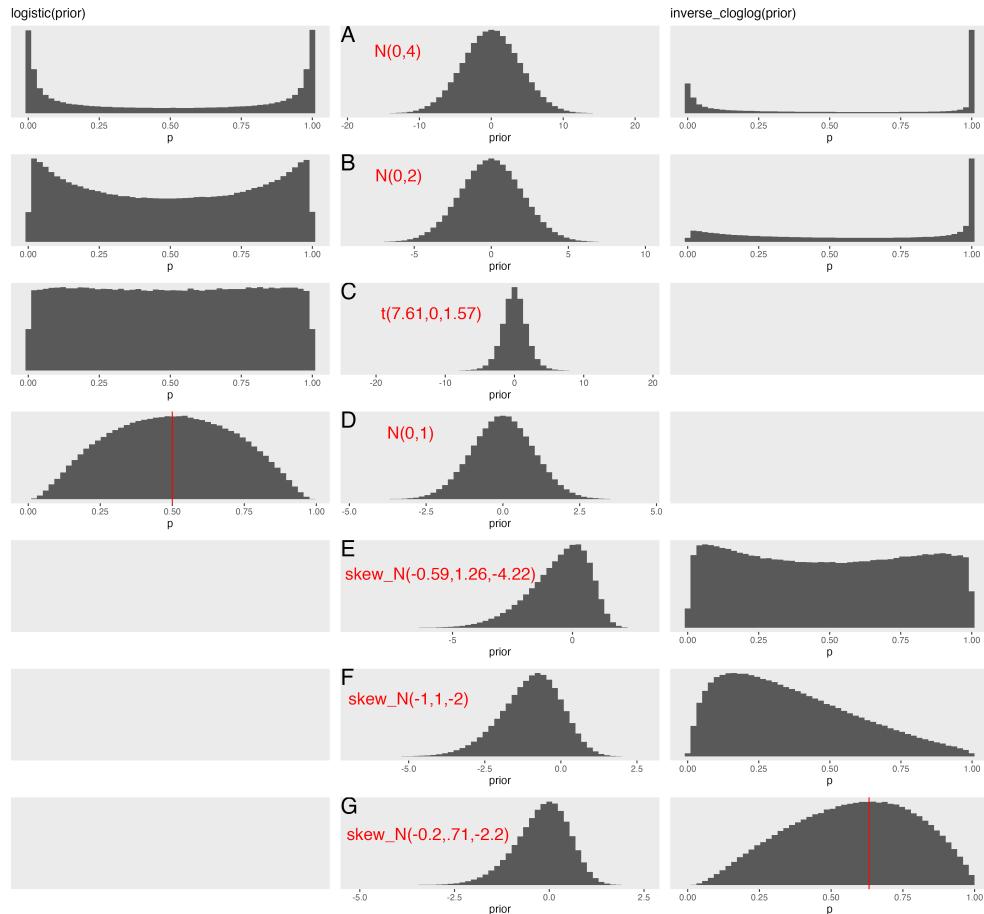


Figure 11. Prior distributions on the logit and/or cloglog scales (middle column), and their implications on the probability scale after applying the inverse-logit (or logistic) transformation (left column), and the inverse-cloglog transformation (right column).

1062 To gain a sense of what prior *cloglog* values would approximate a uniform distribution
 1063 on the hazard probability scale, we followed Kurz's approach and simulated a large number
 1064 of draws from the Uniform(0,1) distribution, converted them to the cloglog metric, and fitted
 1065 a skew-normal model (due to the asymmetry of the cloglog link function). Row E shows that
 1066 using a skew-normal distribution with a mean of -0.59, a standard deviation of 1.26, and a
 1067 skewness of -4.22 as a prior on the cloglog scale, approximates a uniform distribution on the

1068 probability scale. However, because hazard values below .5 are more likely in RT studies,
1069 using a skew-normal distribution with a mean of -1, a standard deviation of 1, and a
1070 skewness of -2 as a prior on the cloglog scale (row F), might be a good weakly informative
1071 prior for the intercept(s) in a cloglog-hazard model. A skew-normal distribution with a mean
1072 of -0.2, a standard deviation of 0.71, and a skewness of -2.2 might be a good weakly
1073 informative prior for the non-intercept parameters in a cloglog-hazard model as it gently
1074 regularizes p towards .6321 (i.e., a zero effect on the cloglog scale).

1075 F. Advantages of hazard analysis

1076 Statisticians and mathematical psychologists recommend focusing on the hazard
1077 function when analyzing time-to-event data for various reasons. First, as discussed by
1078 Holden, Van Orden, and Turvey (2009), “probability density [and mass] functions can appear
1079 nearly identical, both statistically and to the naked eye, and yet are clearly different on the
1080 basis of their hazard functions (but not vice versa). Hazard functions are thus more
1081 diagnostic than density functions” (p. 331) when one is interested in studying the detailed
1082 shape of a RT distribution (see also Figure 1 in Panis, Schmidt, et al., 2020). Therefore,
1083 when the goal is to study how psychological effects change over time, hazard and conditional
1084 accuracy functions are preferred.

1085 Second, because RT distributions may differ from one another in multiple ways,
1086 Townsend (1990) developed a dominance hierarchy of statistical differences between two
1087 arbitrary distributions A and B. For example, if $h_A(t) > h_B(t)$ for all t, then both hazard
1088 functions are said to show a complete ordering. Townsend (1990) concluded that stronger
1089 conclusions can be drawn from data when comparing the hazard functions using EHA. For
1090 example, when $\text{mean } A < \text{mean } B$, the hazard functions might show a complete ordering
1091 (i.e., for all t), a partial ordering (e.g., only for $t > 300$ ms, or only for $t < 500$ ms), or they
1092 may cross each other one or more times.

1093 Third, EHA does not discard right-censored observations when estimating hazard
1094 functions, that is, trials for which we do not observe a response during the data collection
1095 period in a trial so that we only know that the RT must be larger than some value (e.g., the
1096 response deadline). This is important because although a few right-censored observations are
1097 inevitable in most RT tasks, a lot of right-censored observations are expected in experiments
1098 on masking, the attentional blink, and so forth. In other words, by using EHA you can
1099 analyze RT data from experiments that typically do not measure response times. As a result,
1100 EHA can also deal with long RTs in experiments without a response deadline, which are
1101 typically treated as outliers and are discarded before calculating a mean. This orthodox
1102 procedure leads to underestimation of the true mean. By introducing a fixed censoring time
1103 for all trials at the end of the analysis time window, trials with long RTs are not discarded
1104 but contribute to the risk set of each bin.

1105 Fourth, hazard modeling allows incorporating time-varying explanatory covariates such
1106 as heart rate, electroencephalogram (EEG) signal amplitude, gaze location, etc. (Allison,
1107 2010). This is useful for linking physiological effects to behavioral effects when performing
1108 cognitive psychophysiology (Meyer et al., 1988).

1109 Finally, as explained by Kelso, Dumas, and Tognoli (2013), it is crucial to first have a
1110 precise description of the macroscopic behavior of a system (here: $h(t)$ and possibly $ca(t)$
1111 functions) in order to know what to derive on the microscopic level. EHA can thus solve the
1112 problem of model mimicry, i.e., the fact that different computational models can often
1113 predict the same mean RTs as observed in the empirical data, but not necessarily the
1114 detailed shapes of the empirical RT hazard distributions. Also, fitting parametric functions
1115 or computational models to data without studying the shape of the empirical discrete-time
1116 $h(t)$ and $ca(t)$ functions can miss important features in the data (Panis, Moran, et al., 2020;
1117 Panis & Schmidt, 2016).