

1 Event History Analyses 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,
10 Writing - 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
13 GLC, 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
19 the timeline of psychological effects, such as their onset time and duration. The ability to
20 reveal finer-grained, “temporal states” of cognitive processes can have important
21 consequences for theory development by qualitatively changing the key inferences that are
22 drawn from psychological data. Moreover, well-established analytical approaches, such as
23 event history analysis, are able to evaluate the detailed shape of time-to-event distributions,
24 and thus characterise the timeline of psychological states. One barrier to wider use of event
25 history analysis, however, is that the analytical workflow is typically more time-consuming
26 and complex than orthodox approaches. To help achieve broader uptake, in this paper we
27 outline 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. We touch upon several key aspects of the
31 workflow, such as how to specify regression models, the implications for experimental
32 design, as well as how to manage inter-individual differences. We finish the article by
33 considering the benefits of the approach for understanding psychological states, as well as
34 the limitations and future directions of this work. Finally, the project is written in R and
35 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 Event History Analyses for psychological time-to-event data: A tutorial in R with examples
41 in Bayesian and frequentist workflows

42 **1. Introduction**

43 **1.1 Motivation and background context: Comparing means versus
44 distributional shapes**

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

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

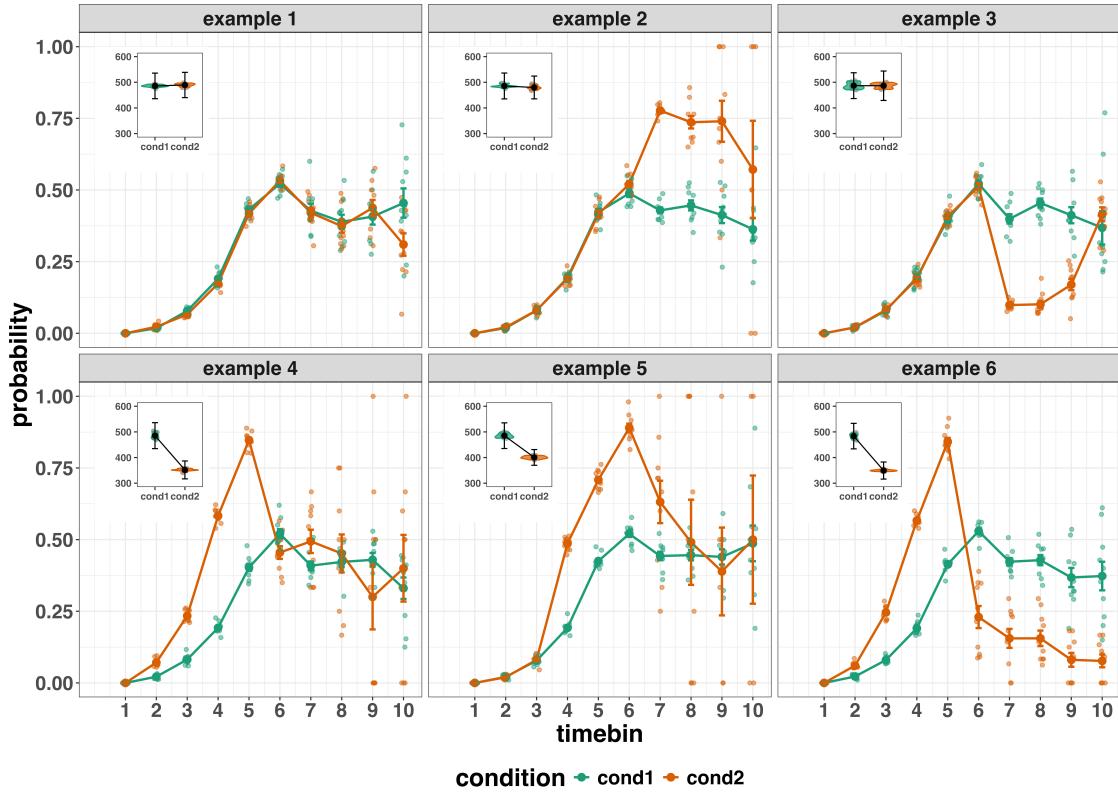


Figure 1. Means versus distributional shapes for six different simulated dataset examples. For our purposes here, it is enough to know that the distributions plotted represent the probability of an event occurring in that timebin, given that it has not yet occurred. Insets show mean reaction time per condition.

65 Why does this matter for research in psychology? Compared to the aggregation of
 66 data across trials, a distributional approach offers the possibility to reveal the timecourse of
 67 psychological states. As such, the approach permits different kinds of questions to be
 68 asked, different inferences to be made, and it holds the potential to discriminate between
 69 different theoretical accounts of psychological and/or brain-based processes. For example,
 70 the distributions in Example 4 show that the effect starts around 200 ms and is gone by
 71 600 ms. In contrast, in Example 5, the effect starts around 400 ms and is gone by 800 ms.
 72 And in the Example 6, the effect reverses around between 500 and 600 ms. What kind of
 73 theory or set of theories could account for such effects? Are there new auxiliary

74 assumptions that theories need to adopt? And are there new experiments that need to be
75 run to test the novel predictions that follow from these analyses? As we show later using
76 concrete examples from past experimental data, for many psychological questions this
77 “temporal states” information can be theoretically meaningful by leading to more
78 fine-grained understanding of psychological processes as well as adding a relatively
79 under-used dimension – the passage of waiting time – to our theory building toolkit.

80 From a historical perspective, it is worth noting that the development of analytical
81 tools that can estimate or predict whether and when events will occur is not a new
82 endeavour. Indeed, hundreds of years ago, analytical methods were developed to predict
83 time to death (REFs). The same logic has been applied to psychological time-to-event
84 data, as previously demonstrated (Panis et al., 2020). Here, in the paper, we hope to show
85 the value of EHA for knowledge and theory building in cognitive psychology and related
86 areas of research, such as cognitive neuroscience, as well as provide practical tutorials that
87 provide step-by-step code and instructions in the hope that we can enable others to use
88 EHA in a more 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
92 analysis, survival analysis, and transition analysis. We first provide a brief overview of
93 hazard analysis to orient the reader to the basic concepts that we will use throughout the
94 paper. However, this will remain relatively short, as this has been covered in detail before
95 (Allison, 1982, 2010; Singer & Willett, 2003), and our primary aim here is to introduce a
96 set of tutorials, which explain **how** to do such analyses, rather than repeat in any detail
97 **why** you should do them.

98 We then provide six different tutorials, each of which is written in the R

99 programming language and publicly available on our Github and the Open Science
100 Framework (OSF) pages, along with all of the other code and material associated with the
101 project. The tutorials provide hands-on, concrete examples of key parts of the analytical
102 process, so that others can apply the analyses to their own time-to-event data sets. Each
103 tutorial is provided as an RMarkdown file, so that others can download and adapt the code
104 to fit their own purposes. Additionally, each tutorial is made available as .html file, so that
105 it can be viewed by any web browser, and thus available to those that do not use R.

106 In Tutorial 1a, we illustrate how to process or “wrangle” a previously published RT +
107 accuracy data set to calculate descriptive statistics when there is one independent variable.
108 The descriptive statistics are plotted, and we comment on their interpretation. In Tutorial
109 1b we provide a generalisation of this approach to illustrate how one can calculate the
110 descriptive statistics when using a more complex design, such as when there are two
111 independent variables. In Tutorial 2a, we illustrate how one can fit Bayesian multi-level
112 regression models to RT data using the R package brms. We discuss possible link
113 functions, and plot the model-based effects of our predictors of interest. In Tutorial 2b we
114 fit Bayesian multi-level regression models to *timed* accuracy data to perform a micro-level
115 speed-accuracy tradeoff (SAT) analysis, which complements the hazard analysis. In
116 Tutorial 3a, we illustrate how to fit the same type of hazard regression models in a
117 frequentist framework using the R package lme4. We then briefly compare and contrast
118 these inferential frameworks when applied to EHA. In Tutorial 3b, we illustrate how to
119 perform the SAT analysis in a frequentist framework.

120 In summary, even though event history analyses is a widely used statistical tool and
121 there already exist many excellent reviews (REFs) and tutorials (Allison, 2010) on its
122 general use-cases, we are not aware of any tutorials that are aimed at psychological
123 time-to-event data, and which provide worked examples of the key data processing and
124 multi-level regression modelling steps. Therefore, our ultimate goal is twofold: first, we
125 want to convince readers of the many benefits of using hazard analysis when dealing with

126 time-to-event data with a focus on psychological time-to-event data, and second, we want
127 to provide a set of practical tutorials, which provide step-by-step instructions on how you
128 actually perform a discrete-time hazard analysis on time-to-event data, as well as a
129 complementary discrete-time SAT analysis on timed accuracy data.

130 **2. A brief introduction to hazard analysis**

131 For a comprehensive background context to hazard analysis, we recommend several
132 excellent textbooks (Singer & Willett, 2003). Likewise, for general introduction to
133 understanding regression equations, we recommend several introductory level textbooks
134 (REFs). Our focus here is not on providing a detailed account of the underlying regression
135 equations, since this topics has been comprehensively covered many times before. Instead,
136 we want to provide an intuition to how EHA works in general as well as in the context of
137 experimental psychology. As such, we only supply regression equations in the
138 supplementary material and then refer to them in the text whenever relevant.

139 **2.1 Basic features of hazard analysis**

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

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

146 The definition of hazard and the type of models employed depend on whether one is
147 using continuous or discrete time units. Since our focus here is on hazard models that use

148 discrete time units, we describe that approach. After dividing time in discrete, contiguous
149 time bins indexed by t (e.g., $t = 1:10$ timebins), let RT be a discrete random variable
150 denoting the rank of the time bin in which a particular person's response occurs in a
151 particular experimental condition. For example, the first response might occur at 546 ms
152 and it would be in timebin 6 (any RTs from 501 ms to 600). Continuous RT data is treated
153 here as interval-censored data.

154 Discrete-time EHA focuses on the discrete-time hazard function and the discrete-time
155 survivor function (Figure 2). The equations that define both of these functions are reported
156 in the supplementary material (equations 1 and 2 in part A). The discrete-time hazard
157 probability gives you the probability that the event occurs (sometime) in bin t , given that
158 the event does not occur in previous bins. In other words, it reflects the instantaneous
159 likelihood that the event occurs in the current bin, given that it has not yet occurred in the
160 past, i.e., in one of the prior bins. This conditionality in the definition of hazard is what
161 makes the hazard function so diagnostic for studying event occurrence, as an event can
162 physically not occur when it has already occurred before. In contrast, the discrete-time
163 survivor function cumulates the bin-by-bin risks of event *nonoccurrence* to obtain the
164 probability that the event occurs after bin t . In other words, the survivor function reflects
165 the likelihood that the event occurs in the future, i.e., in one of the subsequent timebins.

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

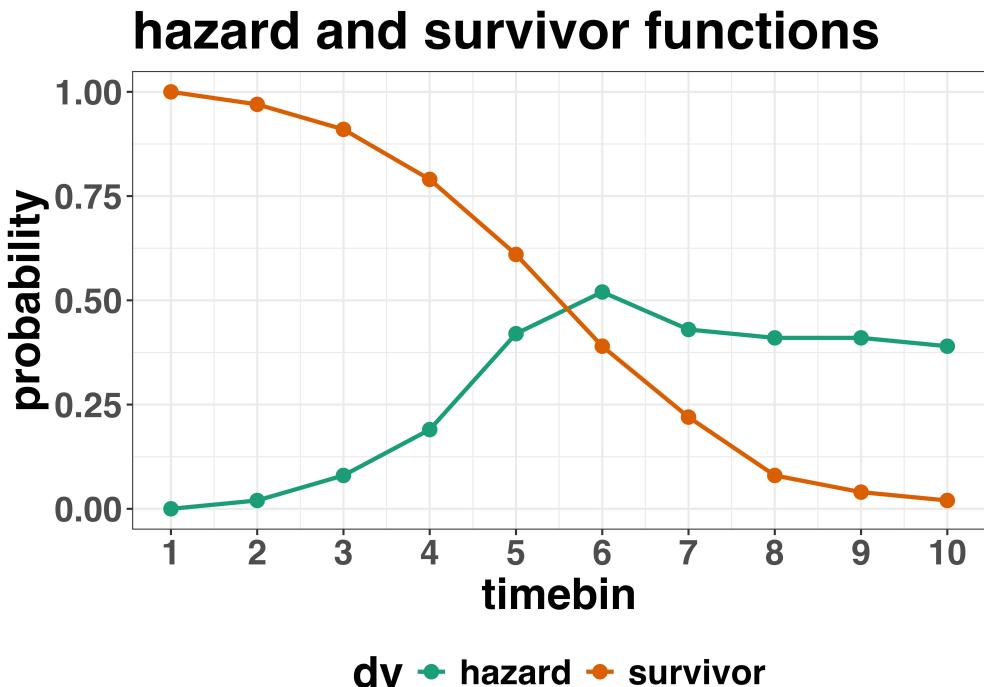


Figure 2. Hazard and survivor functions

¹⁷³ **2.2 Hazard analysis in the context of experimental psychology**

¹⁷⁴ **2.2.1 A worked example.** In the context of experimental psychology, it is
¹⁷⁵ common for participants to be presented with a task that has a right and a wrong answer.
¹⁷⁶ For example, a task may involve choosing between two response options with only one of
¹⁷⁷ them being correct. For such two-choice RT data, the discrete-time hazard analysis can be
¹⁷⁸ extended with a discrete-time SAT analysis. Specifically, the hazard function of event
¹⁷⁹ occurrence can be extended with the discrete-time conditional accuracy function (see
¹⁸⁰ equation 5 in part A of the supplementary material), which gives you the probability that a
¹⁸¹ response is correct given that it is emitted in time bin t (Allison, 2010; Kantowitz &
¹⁸² Pachella, 2021; Wickelgren, 1977).

¹⁸³ Integrating results between hazard and conditional accuracy functions can be
¹⁸⁴ informative for understanding psychological processes. To illustrate, we consider a
¹⁸⁵ hypothetical example that is inspired by real data (Panis et al., 2016), but simplified to

186 make the main point clearer (Figure 3). In a standard response priming paradigm, there is
187 a prime stimulus (e.g., an arrow pointing left or right) followed by a target stimulus
188 (another arrow pointing left or right). The prime can then be congruent or incongruent
189 with the target. Figure 3 shows that the early upswing in hazard is equal for both prime
190 conditions, and that early responses are always correct in the congruent condition and
191 always incorrect in the incongruent condition. These results show that for early responses
192 (< bin 6), responses always follow the prime (and not the target, as instructed). And then
193 for later bins, response hazard is lower in incongruent compared to congruent trials, and all
194 responses emitted in these later bins are correct. This is interesting because mean-average
195 RT would only represent the overall ability of cognition to overcome interference, on
196 average, across trials. And such a conclusion is not supported when the effects are explored
197 over a timeline. Instead, the psychological conclusion is much more nuanced and suggests
198 that multiple states start, stop and possibly interact over a particular temporal window.

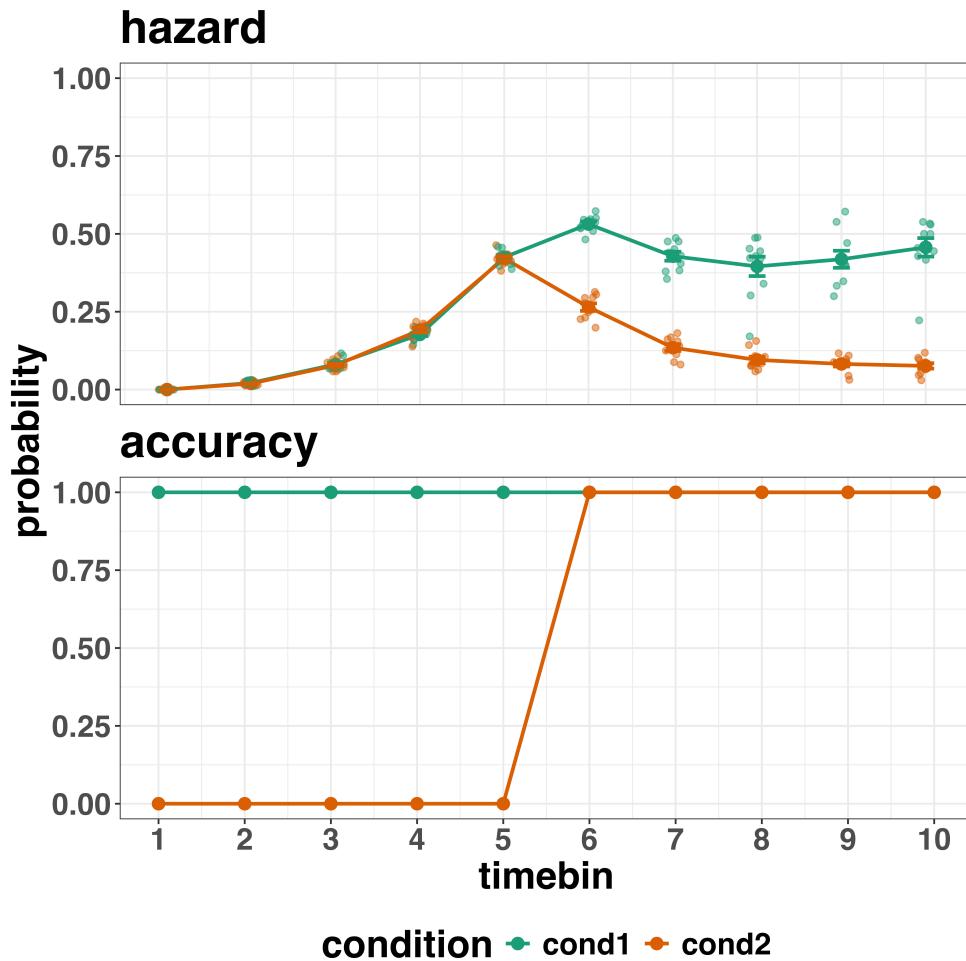


Figure 3. Hazard and conditional accuracy

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

2.2.2 Implications for designing experiments.

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

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

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

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

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

We used R (Version 4.4.0; R Core Team, 2024)¹ for all reported analyses. The

content of the tutorials is mainly based on Allison (2010), Singer and Willett (2003),

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*

²³³ McElreath (2018), Kurz (2023a), and Kurz (2023b).

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

²³⁵ Although the focus is on EHA, we also want to briefly comment on broader aspects of
²³⁶ our general analytical workflow, which relate more to data science and data analysis
²³⁷ workflows.

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

²³⁹ Descriptive, data science workflow. Data wrangling via tidyverse principles and a
²⁴⁰ functional programming approach (cite R4DS textbook here). Functional programming
²⁴¹ basically means you don't write your own loops but instead use functions that have been
²⁴² built and tested by others. [[more here, as necessary]].

²⁴³ **3.2 Inferential statistical approach**

²⁴⁴ Our lab adopts a estimation approach to multi-level regression (Kruschke & Liddel,
²⁴⁵ 2018; Winter, 2019), which is heavily influenced by Bayesian approach as suggested by
²⁴⁶ Richard McElreath (McElreath, 2020; Kurz, 202?). We also use a “keep it maximal”

(Version 0.8.1.9000; Gabry, Češnovar, Johnson, & Broder, 2024), *dplyr* (Version 1.1.4; Wickham, François, Henry, Müller, & Vaughan, 2023), *forcats* (Version 1.0.0; Wickham, 2023a), *ggplot2* (Version 3.5.1; Wickham, 2016), *lme4* (Version 1.1.35.5; Bates, Mächler, Bolker, & Walker, 2015), *lubridate* (Version 1.9.3; Grolemund & Wickham, 2011), *Matrix* (Version 1.7.0; Bates, Maechler, & Jagan, 2024), *nlme* (Version 3.1.166; Pinheiro & Bates, 2000), *papaja* (Version 0.1.2.9000; Aust & Barth, 2023), *patchwork* (Version 1.2.0; Pedersen, 2024), *purrr* (Version 1.0.2; Wickham & Henry, 2023), *RColorBrewer* (Version 1.1.3; Neuwirth, 2022), *Rcpp* (Eddelbuettel & Balamuta, 2018; Version 1.0.12; Eddelbuettel & François, 2011), *readr* (Version 2.1.5; Wickham, Hester, & Bryan, 2024), *RJ-2021-048* (Bengtsson, 2021), *standist* (Version 0.0.0.9000; Girard, 2024), *stringr* (Version 1.5.1; Wickham, 2023b), *tibble* (Version 3.2.1; Müller & Wickham, 2023), *tidybayes* (Version 3.0.6; Kay, 2023), *tidyxr* (Version 1.3.1; Wickham, Vaughan, & Girlich, 2024), *tidyverse* (Version 2.0.0; Wickham et al., 2019), and *tinylabels* (Version 0.2.4; Barth, 2023).

247 approach to specifying varying (or random) effects (Barr et al., 2013). This means that
248 wherever possible we include varying intercepts and slopes per participant To make
249 inferences, we use two main approaches. We compare models of different complexity, using
250 information criteria, such as WAIC or LOO, to evaluate out-of-sample predictive accuracy.
251 We also take the most complex model and evaluate key parameters of interest using point
252 and interval estimates.

253 **4. Tutorials**

254 Tutorials 1a and 1b show how to calculate and plot the descriptive statistics when
255 there are one and two independent variables, respectively. Tutorials 2a and 2b illustrate
256 how to use Bayesian multilevel modeling to fit hazard and conditional accuracy models,
257 respectively. Tutorials 3a and 3b show how to implement, respectively, multilevel models
258 for hazard and conditional accuracy in the frequentist framework. Additionally, to further
259 simplify the process for other users, the tutorials rely on a set of our own custom functions
260 that make sub-processes easier to automate, such as data wrangling and plotting functions
261 (see part B in the supplemental material for a list of the custom functions).

262 Our list of tutorials is as follows:

263 1a. Wrangle raw data and descriptive stats for one independent variable. 1b.
264 Wrangle raw data and descriptive stats for two independent variables. 2a. Bayesian
265 multilevel modeling for $h(t)$ 2b. Bayesian multilevel modeling for $ca(t)$ 3a. Frequentist
266 multilevel modeling for $h(t)$ 3b. Frequentist multilevel modeling for $ca(t)$

267 Planning (T4) - if we get a simulation and power analysis script working, which we
268 are happy with then we could include it here.

269 **4.1 Tutorial 1a: Calculating descriptive statistics using a life table**

270 **4.1.1 Data wrangling aims.** Our data wrangling procedures serve two related
271 purposes. First, we want to summarise and visualise descriptive statistics that relate to our
272 main research questions. Second, we want to produce two different datasets that can each
273 be submitted to different types of inferential modelling approaches. The two types of data
274 structure we label as ‘person-trial’ data (Table 1) and ‘person-trial-bin’ data (Table 2).
275 The ‘person-trial’ data will be familiar to most researchers who record behavioural
276 responses from participants, as it represents the measured RT and accuracy per trial within
277 an experiment. In contrast, the ‘person-trial-bin’ data has a different, more extended
278 structure, which indicates in which bin a response occurred, if at all, in each trial.
279 Therefore, the ‘person-trial-bin’ dataset generates a 0 in each bin until an event occurs and
280 then it generates a 1 to signal an event has occurred in that bin. It is worth pointing out
281 that there is no requirement for an event to occur at all (in any bin), as maybe there was
282 no response on that trial or the event occurred after the time window of interest. Likewise,
283 when the event occurs in bin 1 there would only be one row of data for that trial in the
284 person-trial-bin data set.

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.

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 are shown. These data are simulated and for illustrative purposes only.

285 **4.1.2 A real data wrangling example.** To illustrate how to quickly set up life
 286 tables for calculating the descriptive statistics (functions of discrete time), we use a
 287 published data set on masked response priming from Panis and Schmidt (2016). In their
 288 first experiment, Panis and Schmidt (2016) presented a double arrow for 94 ms that
 289 pointed left or right as the target stimulus with an onset at time point zero in each trial.
 290 Participants had to indicate the direction in which the double arrow pointed using their
 291 corresponding index finger, within 800 ms after target onset. Response time and accuracy
 292 were recorded on each trial. Prime type (blank, congruent, incongruent) and mask type
 293 were manipulated. Here we focus on the subset of trials in which no mask was presented.
 294 The 13-ms prime stimulus was a double arrow with onset at -187 ms for the congruent

295 (same direction as target) and incongruent (opposite direction as target) prime conditions.

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

299 The required column names are as follows:

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

306 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")))
```

307 Next, we can set up the life tables and plots of the discrete-time functions $h(t)$, $S(t)$,
 308 $ca(t)$, and $P(t)$ – see part A of the supplementary material. To do so using a functional
 309 programming approach, one has to nest the data within participants using the
 310 `group_nest()` function, and supply a user-defined censoring time and bin width to our
 311 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 dataset
  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

```

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

319 Removing such trials before calculating the mean event time can introduce a sampling bias
 320 (REFs). The person-trial-bin oriented dataset has one row for each time bin of each trial
 321 that is at risk for event occurrence. The variable “event” in the person-trial-bin oriented
 322 data set indicates whether a response occurs (1) or not (0) for each bin.

323 The next step is to plot the data using our custom function `plot_eha()`. When
 324 creating the plots, some warning messages will likely be generated, like these:

- 325 • Removed 2 rows containing missing values or values outside the scale range
 326 (`geom_line()`).
- 327 • Removed 2 rows containing missing values or values outside the scale range
 328 (`geom_point()`).
- 329 • Removed 2 rows containing missing values or values outside the scale range
 330 (`geom_segment()`).

331 The warning messages are generated because some bins have no hazard and $ca(t)$
 332 estimates, and no error bars. They can thus safely be ignored. One can now inspect
 333 different aspects, including the life table for a particular condition of a particular subject,
 334 and a plot of the different functions for a particular participant. It is important to visually

335 inspect the functions first for each participant, in order to identify possible cheaters (e.g., a
336 flat conditional accuracy function at .5 indicates (s)he was only guessing), outlying
337 individuals, and/or different groups with qualitatively different behavior.

338 Table 3 shows the life table for condition “blank” (no prime stimulus presented) for
339 participant 6 – compare to Figure 1. A life table includes for each time bin, the risk set
340 (number of trials that are event-free at the start of the bin), the number of observed
341 events, and the estimates of $h(t)$, $S(t)$, possibly $ca(t)$, and their estimated standard errors
342 (se). At time point zero, no events can occur and therefore $h(t)$ and $ca(t)$ are undefined.

343 Figure 4 displays the discrete-time hazard, survivor, conditional accuracy, and
344 probability mass functions for each prime condition for participant 6. By using
345 discrete-time $h(t)$ functions of event occurrence – in combination with $ca(t)$ functions for
346 two-choice tasks – one can provide an unbiased, time-varying, and probabilistic description
347 of the latency and accuracy of responses based on all trials of any data set.

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

353 Furthermore, the estimated conditional accuracy values in bin (240,280] are 0.29, 1,
354 and 0 for the blank, congruent, and incongruent prime conditions, respectively. In other
355 words, if a response is emitted in bin (240,280], then the probability that it is correct is
356 estimated to be 0.29, 1, and 0 for the blank, congruent, and incongruent prime conditions,
357 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
0.00	220.00	NA	NA	NA	1.00	0.00	NA	NA
40.00	220.00	0.00	0.00	0.00	1.00	0.00	NA	NA
80.00	220.00	0.00	0.00	0.00	1.00	0.00	NA	NA
120.00	220.00	0.00	0.00	0.00	1.00	0.00	NA	NA
160.00	220.00	0.00	0.00	0.00	1.00	0.00	NA	NA
200.00	220.00	0.00	0.00	0.00	1.00	0.00	NA	NA
240.00	220.00	0.00	0.00	0.00	1.00	0.00	NA	NA
280.00	220.00	7.00	0.03	0.01	0.97	0.01	0.29	0.17
320.00	213.00	13.00	0.06	0.02	0.91	0.02	0.77	0.12
360.00	200.00	26.00	0.13	0.02	0.79	0.03	0.92	0.05
400.00	174.00	40.00	0.23	0.03	0.61	0.03	1.00	0.00
440.00	134.00	48.00	0.36	0.04	0.39	0.03	0.98	0.02
480.00	86.00	37.00	0.43	0.05	0.22	0.03	1.00	0.00
520.00	49.00	32.00	0.65	0.07	0.08	0.02	1.00	0.00
560.00	17.00	9.00	0.53	0.12	0.04	0.01	1.00	0.00
600.00	8.00	4.00	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.

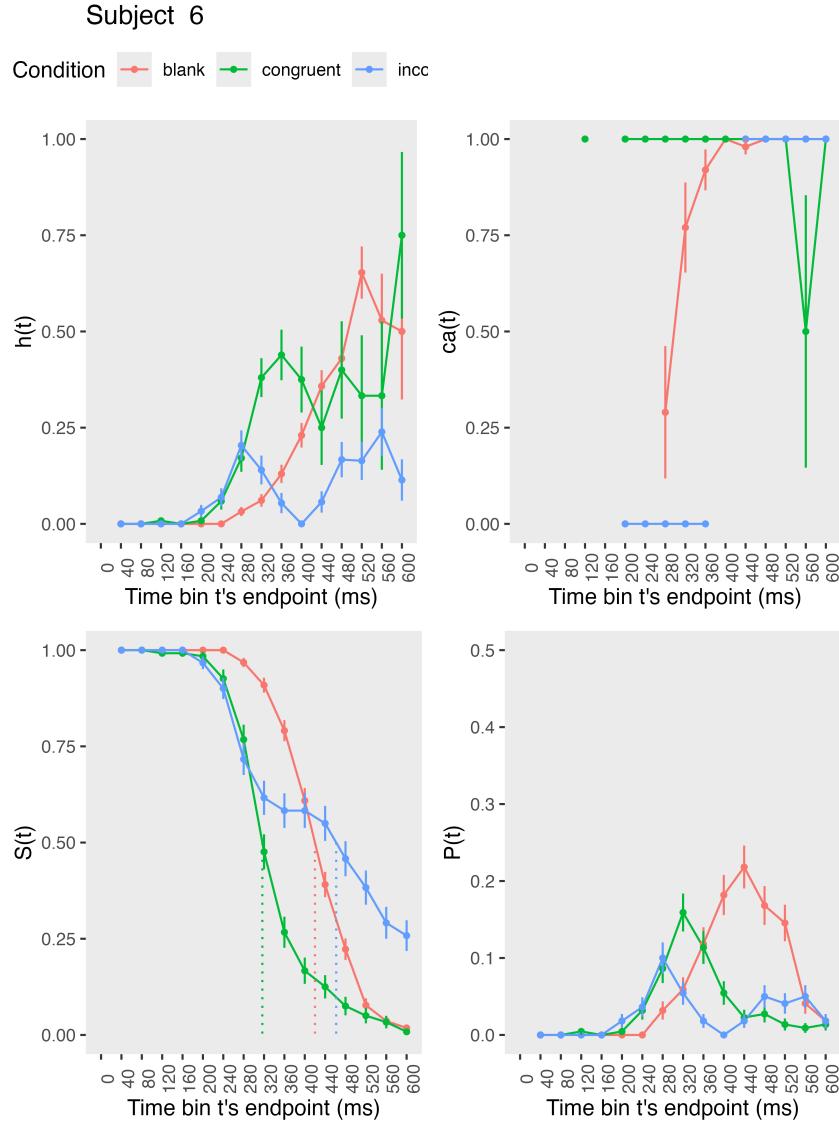


Figure 4. Estimated discrete-time hazard, survivor, and conditional accuracy functions for participant 6, as a function of the passage of discrete waiting time.

358 However, when the waiting time has increased until 400 ms after target onset, then
 359 the conditional probability of response occurrence in the next 40 ms is estimated to be
 360 0.36, 0.25, and 0.06 for the blank, congruent, and incongruent prime conditions,
 361 respectively. And when a response does occur in bin (400,440], then the probability that it
 362 is correct is estimated to be 0.98, 1, and 1 for the blank, congruent, and incongruent prime

363 conditions, respectively.

364 When participants show qualitatively the same distributional patterns, one might
365 consider to aggregate their data and make one plot (see Tutorial_1a.Rmd).

366 These results suggest that the participant is initially responding to the prime even
367 though (s)he was instructed to only respond to the target, that response competition
368 emerges in the incongruent prime condition around 300 ms, and that only later response
369 are fully controlled by the target stimulus. Qualitatively similar results were obtained for
370 the other five participants.

371 These results go against the (often implicit) assumption that all observed responses
372 are primed responses to the target stimulus. Instead, the distributional data show that
373 early responses are triggered exclusively by the prime stimulus, while only later responses
374 reflect primed responses to the target stimulus.

375 At this point, we have calculated, summarised and plotted descriptive statistics for
376 the key variables in EHA. As we will show in Tutorials 2 and 3, statistical models for $h(t)$
377 and $ca(t)$ can be implemented as generalized linear mixed regression models predicting
378 event occurrence (1/0) and accuracy (1/0) in each bin of a selected time window for
379 analysis. As such, multi-level regression is what we turn to in those tutorials. But first we
380 consider calculating the descriptive statistics for two independent variables.

381 **4.2 Tutorial 1b: Generalising to a more complex design**

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

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

389 for the full data set of Experiment 1 of Panis and Schmidt (2016), which includes two

390 independent variables: mask type and prime type. As we use the same functional

391 programming approach as in Tutorial 1a, we simply present the sample-based functions for

392 participant 6 in Figure 5.

393 Compared to the no-mask condition (column 1 in Figure 5), there is a negative

394 compatibility effect in the hazard and conditional accuracy functions when a (relevant,

395 irrelevant, or lines) mask is present.

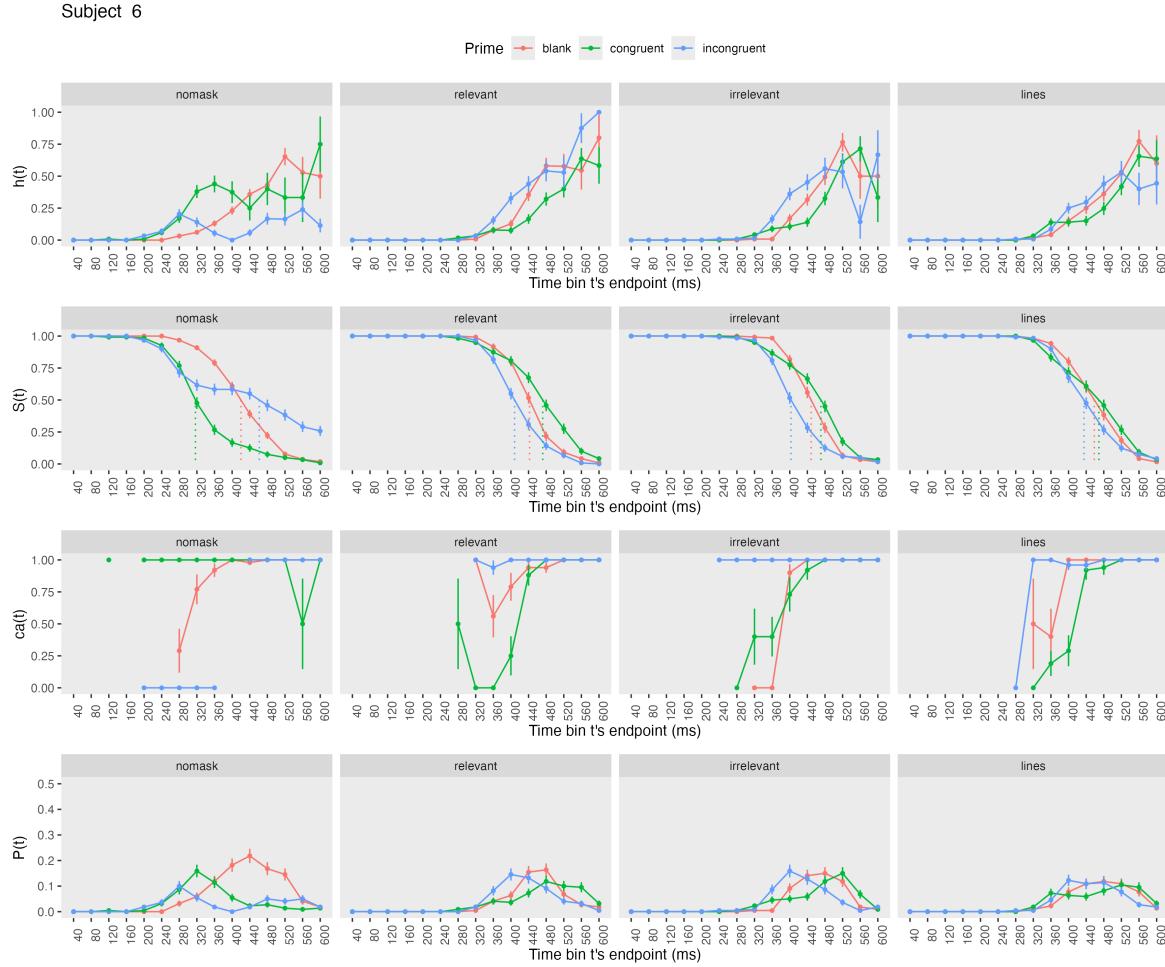


Figure 5. Sample-based discrete-time hazard, survivor, and conditional accuracy functions for participant 6, as a function of prime type (blank, congruent, incongruent) and mask type (no mask, relevant, irrelevant, lines).

396 4.3 Tutorial 2a: Fitting Bayesian hazard models to time-to-event data

397 In this third tutorial, we illustrate how to fit Bayesian hazard regression models to
 398 the masked response priming data set used in Tutorial 1a. Fitting (Bayesian or
 399 non-Bayesian) regression models to the data is important when you want to study how the
 400 shape of the hazard function depends on various predictors (Singer & Willett, 2003).

4.3.1 Hazard model considerations. There are several analytic decisions one

401 has to make when fitting a hazard model. First, one has to select an analysis time window,
402 i.e., a contiguous set of bins for which there is enough data for each participant. Second,
403 given that the dependent variable is binary, one has to select a link function (see part C in
404 the supplementary material). The cloglog link is preferred over the logit link when events
405 can occur in principle at any time point within a bin, which is the case for RT data (Singer
406 & Willett, 2003). Third, one has to choose a specification of the effect of discrete TIME
407 (i.e., the time bin index t). One can choose a general specification (one intercept per bin)
408 or a functional specification, such as a polynomial one (compare model 1 with models 2, 3,
409 and 4 below). We provide relevant example regression formulas in part D of the
410 supplementary material.

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

413 discrete-time hazard model can be estimated using standard logistic regression software
414 after expanding the typical person-trial-oriented data set into a person-trial-bin-oriented
415 data set (Allison, 2010). When there is clustering in the data, as in the case of a small- N
416 design with repeated measurements, the parameters of a discrete-time hazard model can be
417 estimated using population-averaged methods (e.g., Generalized Estimating Equations),
418 and Bayesian or frequentist generalized linear mixed models (Allison, 2010).

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

420 experimental predictor variables and other covariates: The linearity assumption for
421 continuous predictors (the effect of a 1 unit change is the same anywhere on the scale), the
422 additivity assumption (predictors do not interact), and the proportionality assumption
423 (predictors do not interact with TIME).

424 In this tutorial we will fit four Bayesian multilevel models (i.e., generalized linear

425 mixed models) that differ in complexity to the person-trial-bin oriented data set that we
426 created in Tutorial 1a. We select the analysis range (200,600] and the cloglog link. The

427 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 range: (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 (period) on bin 9, and trial on trial 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))
```

428 **4.3.2 Prior distributions.** To get the posterior distribution of each model

429 parameter given the data, we need to specify a prior distribution for each parameter. The
 430 middle column of Figure 12 in part E of the supplementary material shows seven examples
 431 of prior distributions on the logit and/or cloglog scales.

432 While a normal distribution with relatively large variance is often used as a weakly
 433 informative prior for continuous dependent variables, rows A and B in Figure 12 show that
 434 specifying such distributions on the logit and cloglog scales leads to rather informative
 435 distributions on the original probability scale, as most mass is pushed to probabilities of 0
 436 and 1. The other rows in Figure 9 show prior distributions on the logit and cloglog scale
 437 that we use instead.

438 **4.3.3 Model 1: A general specification of TIME, and main effects of**

439 **congruency and trial number.** When you do not want to make assumptions about the
 440 shape of the hazard function, or its shape is not smooth but irregular, you can use a
 441 general specification of TIME, i.e., one intercept per time bin. In this first model, we use a
 442 general specification of TIME for the selected baseline condition (blank prime), and assume
 443 that the effects of prime-target congruency and trial number are proportional and additive,
 444 and that the effect of trial number is linear. Before we fit model 1, we remove unnecessary
 445 columns from the data, and specify our priors. In the code of Tutorial 2a, model M1 is
 446 specified as follows.

```
plan(multicore)

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")
```

447 After selecting the binomial family and the cloglog link, the model formula is

448 specified. The fixed effects include 9 dummy variables, the explicit Intercept variable
 449 (which represents bin 9 in this example), and the main effects of priming condition and
 450 centered trial number. Each of these effects is allowed to vary across individuals.

451 Estimating model M1 took about 70 minutes on a MacBook Pro (Sonoma 14.6.1 OS, 18GB
 452 Memory, M3 Pro Chip).

453 **4.3.4 Model 2: A polynomial specification of TIME, and main effects of**

454 **congruency and trial number.** When the shape of the hazard function is rather

smooth, one can fit a more parsimonious model by using a polynomial specification of TIME. For our second example model, we thus use a third-order polynomial specification of TIME for the baseline condition (blank prime), and again assume that the effects of prime-target congruency and centered trial number are proportional and additive, and that the effect of trial number is linear. The model formula for model M2 looks 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),
```

Because TIME is centered on bin 9, and trial number on trial 1000, the Intercept represents the cloglog-hazard in bin 9 for the blank prime condition in trial 1000. Estimating model M2 took about 144 minutes.

4.3.5 Model 3: A polynomial specification of TIME, and relaxing the proportionality assumption. So far, we assumed that the effect of our predictors condition and centered trial number are the same in each time bin. However, the descriptive plots suggest that the effect of condition can vary across time bins. For the third model, we thus use a third-order polynomial specification of TIME for the baseline condition (blank prime), and relax the proportionality assumption for the predictor variables prime-target congruency (variable “condition”) and 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),

```

471 Note that duplicate terms in the model formula are ignored. Estimating model M3
472 took about 268 minutes.

473 **4.3.6 Model 4: A polynomial specification of TIME, and relaxing all three
474 assumptions.** Based on previous work (Panis, 2020; Panis, Moran, et al., 2020; Panis &
475 Schmidt, 2022; Panis et al., 2017; Panis & Wagemans, 2009), we relax all three
476 assumptions in model 4. We thus add a squared term for the continuous predictor centered
477 trial number – I(trial_c^2) – and include interaction terms.

```

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)

```

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

480 **4.3.7 Compare the models.** We can compare the four models using the Widely
481 Applicable Information Criterion (WAIC) and Leave-One-Out (LOO) cross-validation, and
482 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)
```

```
483 ## model_M1 model_M2 model_M3 model_M4
484 ##      0      0      0      1
```

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

```
485 ## model_M1 model_M2 model_M3 model_M4
486 ##      0      0      0      1
```

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

489 **4.3.8 Evaluate parameter estimates.** To make inferences from the parameter
 490 estimates in model M4, we summarize the draws from the posterior distributions of the
 491 effect of congruent and incongruent primes relative to the blank prime condition, in each
 492 time bin for trial numbers 500, 1000, and 1500, in terms of point and interval estimates.

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

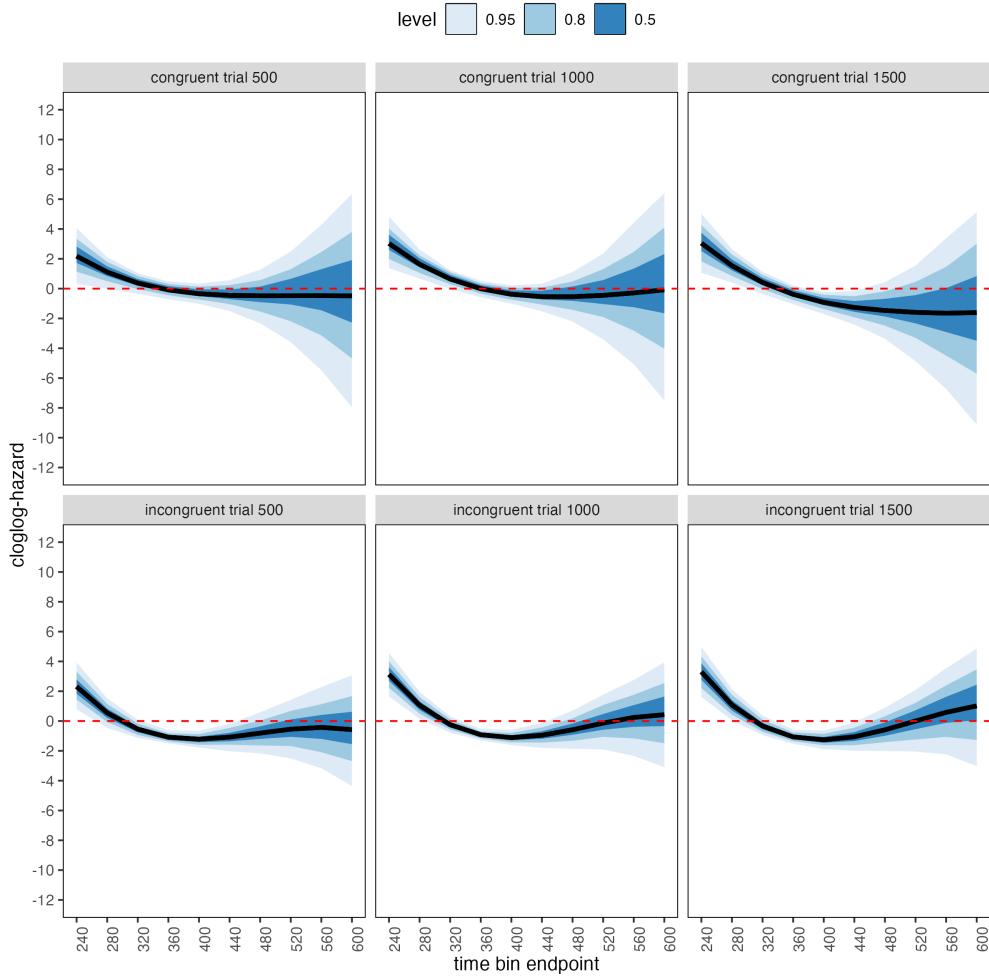


Figure 6. Means and 50/80/95% highest density intervals of the draws from the posterior distributions representing the effect of congruent and incongruent primes relative to neutral primes in each bin for trial numbers 500, 1000, and 1500.

496 Table 4 shows the summaries of the draws from the posterior distributions of the
 497 effects of congruent and incongruent primes relative to the blank prime condition in trials
 498 500, 1000, and 1500, in terms of a point estimate (the mean) and the upper and lower
 499 bounds of the 95% highest density interval. Also, by exponentiating the mean we obtain an
 500 effect size in terms of a hazard ratio.

Table 4

Point and interval estimates, and hazard ratios.

bin_endpoint	condition	mean	.lower	.upper	.width	hazard ratio
240.00	c500	2.18	0.33	4.05	0.95	8.82
280.00	c500	1.11	-0.02	2.11	0.95	3.03
320.00	c500	0.37	-0.34	1.04	0.95	1.45
360.00	c500	-0.09	-0.70	0.48	0.95	0.91
400.00	c500	-0.35	-1.02	0.34	0.95	0.71
440.00	c500	-0.45	-1.50	0.56	0.95	0.64
480.00	c500	-0.48	-2.32	1.27	0.95	0.62
520.00	c500	-0.48	-3.57	2.52	0.95	0.62
560.00	c500	-0.52	-5.69	4.27	0.95	0.60
600.00	c500	-0.66	-8.56	6.99	0.95	0.52
240.00	c1000	3.03	1.37	4.82	0.95	20.63
280.00	c1000	1.63	0.68	2.63	0.95	5.13
320.00	c1000	0.64	-0.02	1.24	0.95	1.90
360.00	c1000	-0.01	-0.57	0.52	0.95	0.99
400.00	c1000	-0.38	-1.01	0.22	0.95	0.68
440.00	c1000	-0.54	-1.52	0.32	0.95	0.58
480.00	c1000	-0.54	-2.20	1.11	0.95	0.58
520.00	c1000	-0.45	-3.40	2.35	0.95	0.64
560.00	c1000	-0.34	-5.78	3.90	0.95	0.71
600.00	c1000	-0.25	-8.34	6.73	0.95	0.78
240.00	c1500	3.05	1.07	5.02	0.95	21.02
280.00	c1500	1.54	0.40	2.65	0.95	4.66
320.00	c1500	0.42	-0.36	1.13	0.95	1.52
360.00	c1500	-0.38	-1.05	0.21	0.95	0.68
400.00	c1500	-0.92	-1.70	-0.24	0.95	0.40
440.00	c1500	-1.26	-2.41	-0.18	0.95	0.28
480.00	c1500	-1.47	-3.36	0.43	0.95	0.23
520.00	c1500	-1.60	-4.86	1.58	0.95	0.20
560.00	c1500	-1.71	-7.01	3.37	0.95	0.18
600.00	c1500	-1.88	-10.07	5.98	0.95	0.15
240.00	i500	2.31	0.79	3.93	0.95	10.10
280.00	i500	0.55	-0.46	1.52	0.95	1.72
320.00	i500	-0.54	-1.13	0.08	0.95	0.58
360.00	i500	-1.08	-1.50	-0.61	0.95	0.34
400.00	i500	-1.22	-1.78	-0.65	0.95	0.30
440.00	i500	-1.08	-2.03	-0.19	0.95	0.34
480.00	i500	-0.81	-2.16	0.59	0.95	0.44

Table 4 continued

bin_endpoint	condition	mean	.lower	.upper	.width	hazard ratio
520.00	i500	-0.55	-2.50	1.42	0.95	0.58
560.00	i500	-0.42	-3.16	2.28	0.95	0.65
600.00	i500	-0.58	-4.35	3.10	0.95	0.56
240.00	i1000	3.12	1.66	4.58	0.95	22.68
280.00	i1000	1.06	0.15	1.95	0.95	2.88
320.00	i1000	-0.24	-0.78	0.31	0.95	0.78
360.00	i1000	-0.92	-1.30	-0.52	0.95	0.40
400.00	i1000	-1.11	-1.61	-0.59	0.95	0.33
440.00	i1000	-0.95	-1.80	-0.12	0.95	0.39
480.00	i1000	-0.58	-1.86	0.70	0.95	0.56
520.00	i1000	-0.14	-1.90	1.77	0.95	0.87
560.00	i1000	0.24	-2.33	2.75	0.95	1.27
600.00	i1000	0.42	-3.17	3.85	0.95	1.52
240.00	i1500	3.30	1.63	4.98	0.95	27.07
280.00	i1500	1.08	0.05	2.14	0.95	2.94
320.00	i1500	-0.33	-0.94	0.36	0.95	0.72
360.00	i1500	-1.06	-1.52	-0.57	0.95	0.35
400.00	i1500	-1.26	-1.88	-0.65	0.95	0.28
440.00	i1500	-1.06	-1.99	-0.09	0.95	0.35
480.00	i1500	-0.59	-2.01	0.88	0.95	0.55
520.00	i1500	0.00	-2.05	2.09	0.95	1.00
560.00	i1500	0.58	-2.23	3.54	0.95	1.79
600.00	i1500	1.01	-3.02	4.86	0.95	2.75

Note. c500 = effect of congruent primes in trial 500; c1000 = effect of congruent primes in trial 1000; c1500 = effect of congruent primes in trial 1500; i500 = effect of incongruent primes in trial 500; i1000 = effect of incongruent primes in trial 1000; i1500 = effect of incongruent primes in trial 1500.

503 To conclude this Tutorial 2a, Figure 7 shows the model-based hazard functions for

504 each prime type for participant 6, in trial 500, 1000, and 1500.

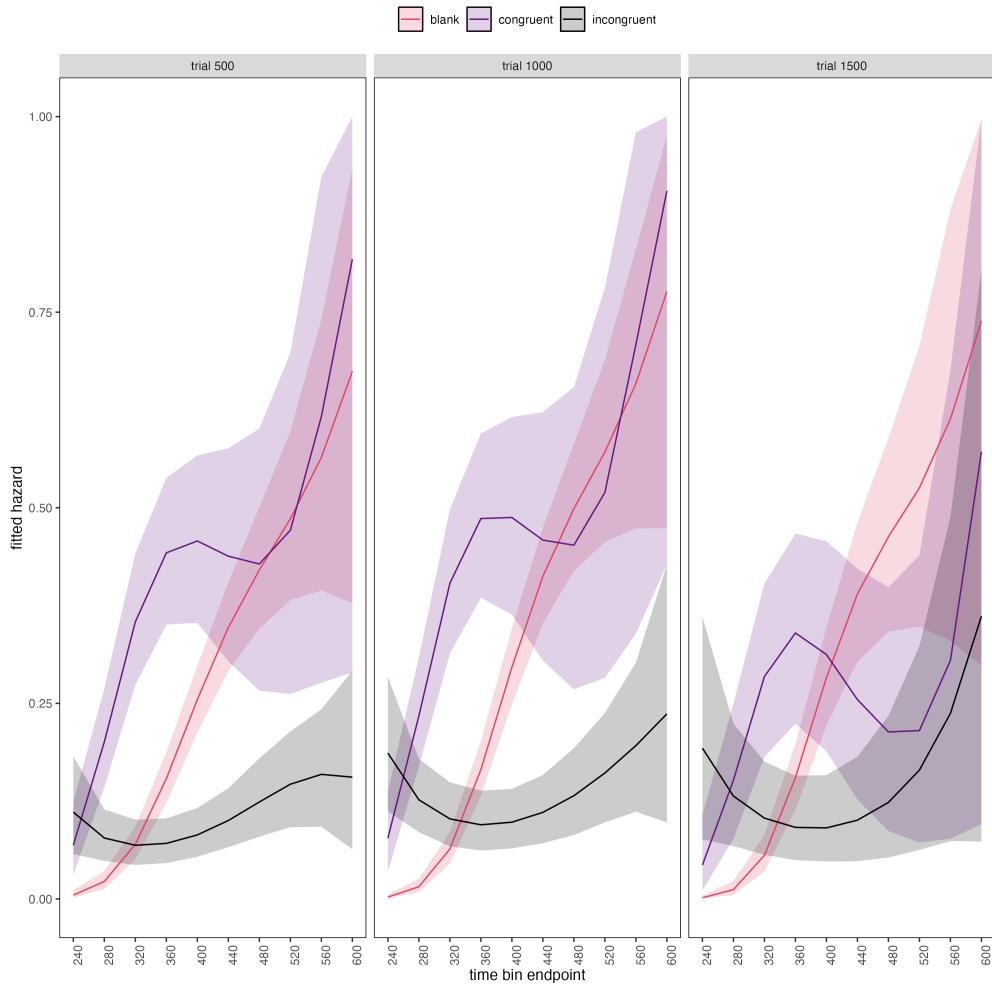


Figure 7. Model-based hazard functions for participant 6 in trial 500, 1000, and 1500.

505 **4.4 Tutorial 2b: Fitting Bayesian conditional accuracy models**

506 In this fourth tutorial, we illustrate how to fit a Bayesian regression model to the
 507 timed accuracy data from the masked response priming data set used in Tutorial 1a. For
 508 illustration purposes, we only fitted the effects of model M4 (see Tutorial 2a).

509 To make inferences from the parameter estimates in model M4_ca, we summarize the
 510 draws from the posterior distributions of the effect of congruent and incongruent primes

511 relative to the blank prime condition, in each time bin for trial numbers 500, 1000, and
 512 1500, in terms of point and interval estimates.

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

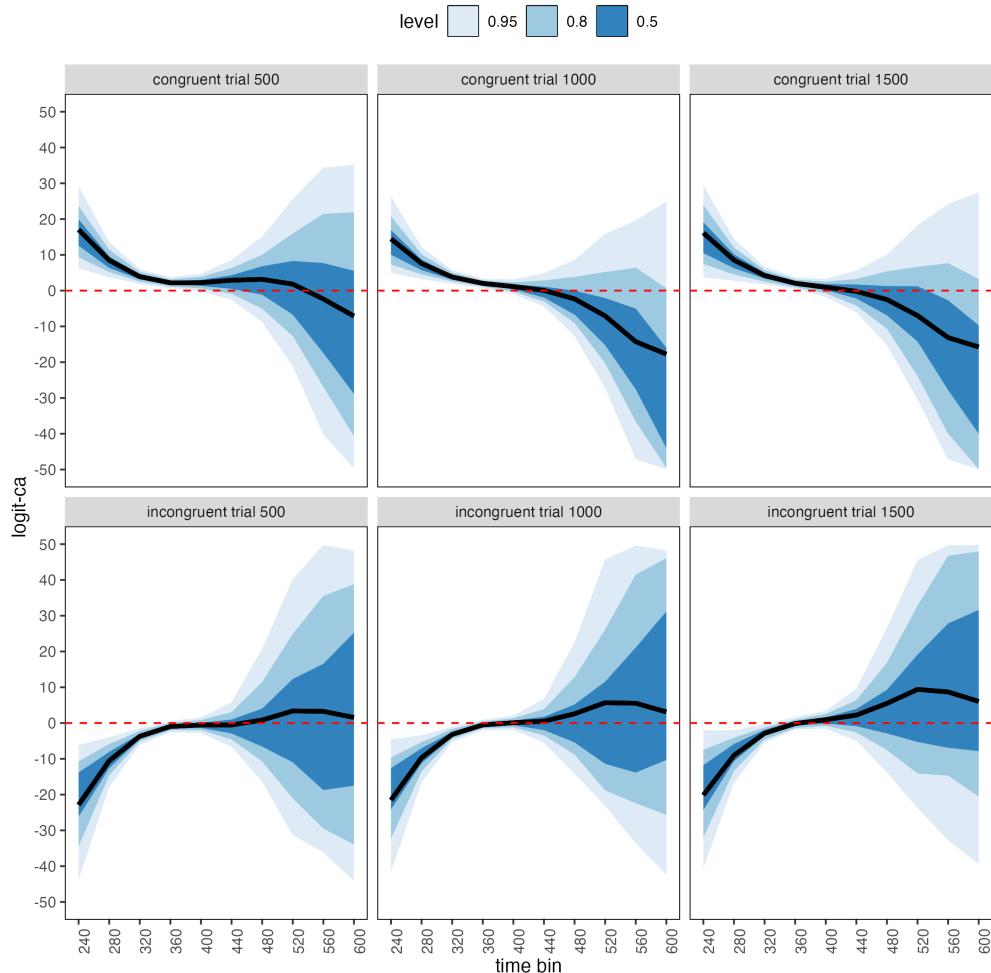


Figure 8. Means and 50/80/95% highest density intervals of the draws from the posterior distributions representing the effect of congruent and incongruent primes relative to neutral primes in each bin for trial numbers 500, 1000, and 1500.

516 Table 5 shows the summaries of the draws from the posterior distributions of the
 517 effects of congruent and incongruent primes relative to the blank prime condition in trials

⁵¹⁸ 500, 1000, and 1500, in terms of a point estimate (the mean) and the upper and lower
⁵¹⁹ limits of the 95% highest density interval. Also, by exponentiating the mean we obtain an
⁵²⁰ effect size in terms of an odds ratio.

Table 5

Point and interval estimates, and odds ratios.

bin_endpoint	condition	mean	.lower	.upper	.width	odds ratio
240.00	c500	17.02	6.26	29.22	0.95	24,618,458.61
280.00	c500	8.49	3.71	13.54	0.95	4,846.09
320.00	c500	3.91	1.88	6.05	0.95	49.79
360.00	c500	2.19	0.69	3.75	0.95	8.89
400.00	c500	2.22	-0.25	4.75	0.95	9.19
440.00	c500	2.91	-2.56	8.54	0.95	18.31
480.00	c500	3.15	-8.77	15.10	0.95	23.41
520.00	c500	1.86	-20.13	26.73	0.95	6.40
560.00	c500	-2.08	-39.94	42.41	0.95	0.12
600.00	c500	-9.77	-73.17	61.54	0.95	0.00
240.00	c1000	14.46	4.94	26.35	0.95	1,899,836.02
280.00	c1000	7.58	3.21	12.18	0.95	1,961.83
320.00	c1000	3.80	1.90	5.71	0.95	44.87
360.00	c1000	2.02	0.72	3.35	0.95	7.57
400.00	c1000	1.14	-0.99	3.11	0.95	3.14
440.00	c1000	0.06	-4.41	4.87	0.95	1.06
480.00	c1000	-2.32	-12.62	8.61	0.95	0.10
520.00	c1000	-7.10	-27.24	15.97	0.95	0.00
560.00	c1000	-15.39	-54.71	23.54	0.95	0.00
600.00	c1000	-28.27	-92.96	35.54	0.95	0.00
240.00	c1500	16.12	3.74	29.48	0.95	10,001,085.39
280.00	c1500	8.54	2.78	14.43	0.95	5,124.44
320.00	c1500	4.22	1.75	6.70	0.95	68.12
360.00	c1500	2.06	0.48	3.71	0.95	7.82
400.00	c1500	0.95	-1.75	3.26	0.95	2.58
440.00	c1500	-0.20	-6.03	5.65	0.95	0.82
480.00	c1500	-2.49	-15.23	10.07	0.95	0.08
520.00	c1500	-7.03	-30.41	18.55	0.95	0.00
560.00	c1500	-14.91	-58.81	27.21	0.95	0.00
600.00	c1500	-27.22	-95.59	43.50	0.95	0.00
240.00	i500	-23.34	-44.42	-4.87	0.95	0.00

Table 5 continued

bin_endpoint	condition	mean	.lower	.upper	.width	odds ratio
280.00	i500	-10.55	-17.93	-3.94	0.95	0.00
320.00	i500	-3.71	-6.06	-1.51	0.95	0.02
360.00	i500	-0.97	-2.57	0.56	0.95	0.38
400.00	i500	-0.52	-2.75	1.55	0.95	0.59
440.00	i500	-0.53	-6.67	5.86	0.95	0.59
480.00	i500	0.83	-16.41	20.71	0.95	2.30
520.00	i500	5.40	-32.44	52.48	0.95	222.04
560.00	i500	15.00	-58.75	104.35	0.95	3,282,435.93
600.00	i500	31.47	-90.20	190.08	0.95	46,319,712,352,328.76
240.00	i1000	-21.85	-43.05	-4.10	0.95	0.00
280.00	i1000	-9.67	-16.56	-3.46	0.95	0.00
320.00	i1000	-3.17	-5.23	-0.99	0.95	0.04
360.00	i1000	-0.53	-2.03	0.89	0.95	0.59
400.00	i1000	0.09	-1.88	2.11	0.95	1.10
440.00	i1000	0.52	-5.54	6.73	0.95	1.68
480.00	i1000	2.58	-14.16	22.53	0.95	13.20
520.00	i1000	8.10	-28.51	55.88	0.95	3,307.44
560.00	i1000	18.92	-51.75	111.96	0.95	164,758,701.84
600.00	i1000	36.86	-89.39	191.12	0.95	10,165,856,639,901,592.00
240.00	i1500	-20.51	-42.95	-2.49	0.95	0.00
280.00	i1500	-9.04	-16.80	-2.03	0.95	0.00
320.00	i1500	-2.86	-5.47	-0.25	0.95	0.06
360.00	i1500	-0.14	-1.81	1.67	0.95	0.87
400.00	i1500	0.94	-1.61	3.40	0.95	2.57
440.00	i1500	2.22	-4.97	9.63	0.95	9.21
480.00	i1500	5.52	-13.87	26.57	0.95	249.51
520.00	i1500	12.67	-31.46	58.41	0.95	318,500.40
560.00	i1500	25.50	-53.08	115.21	0.95	119,299,568,240.94
600.00	i1500	45.85	-86.60	200.06	0.95	81,670,189,671,651,033,088.00

Note. c500 = effect of congruent primes in trial 500; c1000 = effect of congruent primes in trial 1000; c1500 = effect of congruent primes in trial 1500; i500 = effect of incongruent primes in trial 500; i1000 = effect of incongruent primes in trial 1000; i1500 = effect of incongruent primes in trial 1500.

521

522 Interpretation . . .

523 plot 9 ca(t)

524 **4.5 Tutorial 3a: Fitting Frequentist hazard models**

525 In this fifth tutorial we illustrate how to fit a multilevel hazard regression model in
526 the frequentist framework, for the data set used in Tutorial 1a. For illustration purposes,
527 we only fitted the effects from model M3 (see Tutorial 2a) using the function `glmer()` from
528 the R package `lme4`. Alternatively, one could also use the function `glmmPQL()` from the R
529 package `MASS`. The resulting hazard model is called `M3_f`.

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

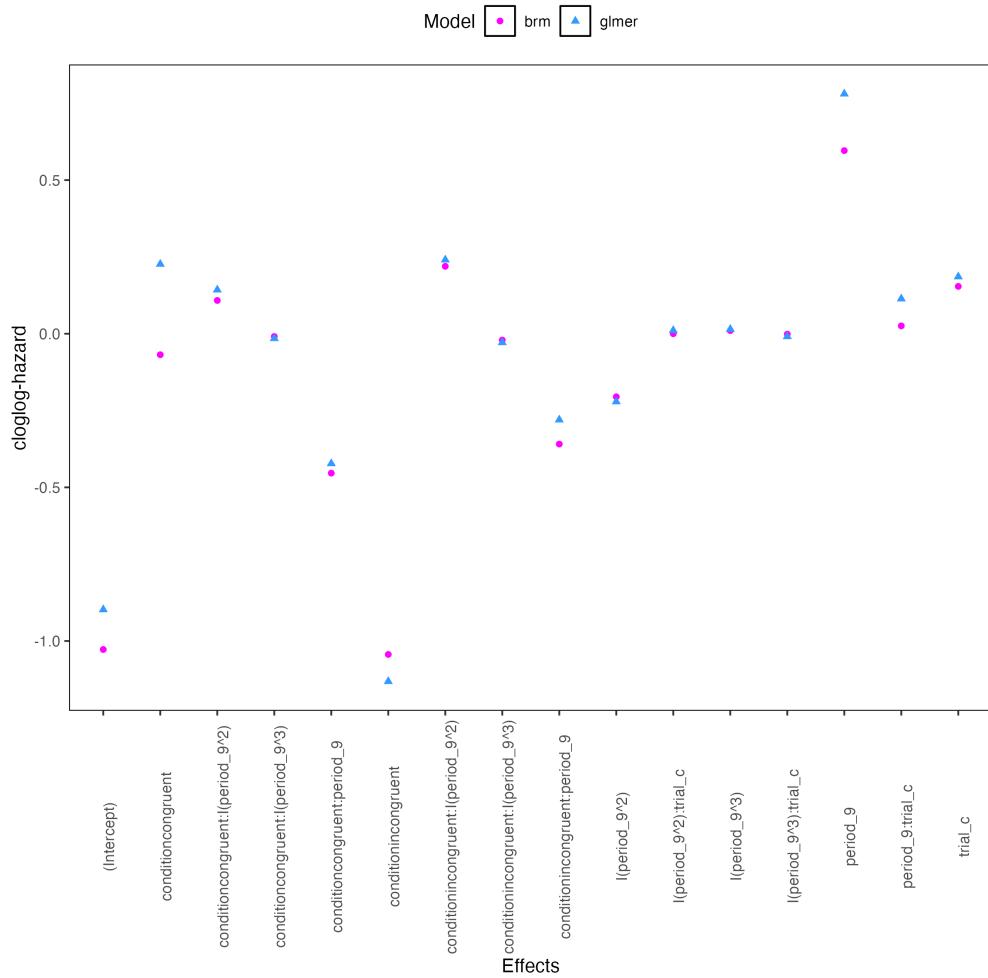


Figure 9. Parameter estimates for model M3 from `brm()` and model M3_f from `glmer()`.

532 Figure 10 confirms that the parameter estimates from both Bayesian and frequentist

533 models are pretty similar. However, the random effects structure of model M3 was already
534 too complex for the frequentist model as it did not converge and resulted in a singular fit.

535 This is of course one of the reasons why Bayesian modeling has become so popular in

536 recent years. But the price you pay for being able to fit more complex models in a

537 Bayesian framework is computation time. In other words, as we have noted throughout,

538 some of the Bayesian models in Tutorials 2a and 2b took several hours to build.

539 **4.6 Tutorial 3b: Fitting Frequentist conditional accuracy models**

540 In this sixth tutorial we illustrate how to fit a multilevel regression model to the
541 timed accuracy data in the frequentist framework, for the data set used in Tutorial 1a. For
542 illustration purposes, we only fitted the effects from model M3 (see Tutorial 2a) using the
543 function glmer() from the R package lme4. Alternatively, one could also use the function
544 glmmPQL() from the R package MASS. Again, the resulting conditional accuracy model
545 M3_ca_f did not converge and resulted in a singular fit.

546 **5. Discussion**

547 This main motivation for writing this paper is the observation that event history
548 analysis remains under-used in psychological research, which means the field of research is
549 not taking full advantage of the many benefits EHA provides compared to more
550 conventional analyses. By providing a freely available set of tutorials, which provide
551 step-by-step guidelines and ready-to-use R code, we hope that researchers will feel more
552 comfortable using EHA in the future. Indeed, we hope that our tutorials may help to
553 overcome a barrier to entry with EHA, which is the increase in analytical complexity
554 compared to mean-average comparisons. While we have focused here on within-subject,
555 factorial, small- N designs, it is important to realize that event history analysis can be
556 applied to other designs as well (large- N designs with only one measurement per subject,
557 between-subject designs, etc.). As such, the general workflow and associated code can be
558 modified and applied more broadly to other contexts and research questions. In the
559 following, we discuss issues relating to model complexity versus interpretability, individual
560 differences, limitations of the approach, and future extensions.

561 **5.1 Advantages of hazard analysis**

562 Statisticians and mathematical psychologists recommend focusing on the hazard
563 function when analyzing time-to-event data for various reasons. First, as discussed by
564 Holden, Van Orden, and Turvey (2009), “probability density functions can appear nearly
565 identical, both statistically and to the naked eye, and yet are clearly different on the basis
566 of their hazard functions (but not vice versa). Hazard functions are thus more diagnostic
567 than density functions” (p. 331) when one is interested in studying the detailed shape of a
568 RT distribution (see also Figure 1 in Panis, Schmidt, et al., 2020). Therefore, [[why should
569 people care? What is the functional relevance for exp psych and researchers?]]

570 [[This para needs to be way shorter and easier to read or we get rid of it]] Second,
571 because RT distributions may differ from one another in multiple ways, Townsend (1990)
572 developed a dominance hierarchy of statistical differences between two arbitrary
573 distributions A and B. For example, if $F_A(t) > F_B(t)$ for all t , then both cumulative
574 distribution functions are said to show a complete ordering. Townsend (1990) showed that
575 a complete ordering on the hazard functions — $\lambda_A(t) > \lambda_B(t)$ for all t — implies a complete
576 ordering on both the cumulative distribution and survivor functions — $F_A(t) > F_B(t)$ and
577 $S_A(t) < S_B(t)$ — which in turn implies an ordering on the mean latencies —mean A <
578 mean B. In contrast, an ordering on two means does *not* imply a complete ordering on the
579 corresponding $F(t)$ and $S(t)$ functions, and a complete ordering on these latter functions
580 does *not* imply a complete ordering on the corresponding hazard functions. This means
581 that stronger conclusions can be drawn from data when comparing the hazard functions
582 using EHA. For example, when mean A < mean B, the hazard functions might show a
583 complete ordering (i.e., for all t), a partial ordering (e.g., only for $t > 300$ ms, or only for t
584 < 500 ms), or they may cross each other one or more times. As a result, instead of using
585 delta-plots for RT – differences in quantiles from $F(t)^{-1}$ – one can simply plot delta-h(t)
586 functions (see Panis, 2020).

587 Third, EHA does not discard right-censored observations when estimating hazard
588 functions, that is, trials for which we do not observe a response during the data collection
589 period in a trial so that we only know that the RT must be larger than some value (i.e., the
590 response deadline). This is important because although a few right-censored observations
591 are inevitable in most RT tasks, a lot of right-censored observations are expected in
592 experiments on masking, the attentional blink, and so forth. In other words, by using EHA
593 you can analyze RT data from experiments that typically do not measure response times.
594 As a result, EHA can also deal with long RTs in experiments without a response deadline,
595 which are typically treated as outliers and are discarded before calculating a mean. This
596 orthodox procedure can lead to a sampling bias, however, which results in underestimation
597 of the mean. By introducing a fixed censoring time for all trials at the end of the analysis
598 time window, trials with long RTs are not discarded but contribute to the risk set of each
599 bin.

600 Fourth, hazard modeling allows incorporating time-varying explanatory covariates
601 such as heart rate, electroencephalogram (EEG) signal amplitude, gaze location, etc.
602 (Allison, 2010). This is useful for linking physiological effects with behavioral effects when
603 performing cognitive psychophysiology (Meyer, Osman, Irwin, & Yantis, 1988).

604 Finally, as explained by Kelso, Dumas, and Tognoli (2013), it is crucial to first have a
605 precise description of the macroscopic behavior of a system (here: $h(t)$ and $ca(t)$ functions)
606 in order to know what to derive on the microscopic level. EHA can thus solve the problem
607 of model mimicry, i.e., the fact that different computational models can often predict the
608 same mean RTs as observed in the empirical data, but not necessarily the detailed shapes
609 of the empirical RT hazard distributions. Also, fitting parametric functions or
610 computational models to data without studying the shape of the empirical discrete-time
611 $h(t)$ and $ca(t)$ functions can miss important features in the data (Panis, Moran, et al.,
612 2020; Panis & Schmidt, 2016).

613 5.2 Model complexity versus interpretability

614 Models for discrete-time $h(t)$ and $ca(t)$ can quickly become very complex when
615 adding more than 1 time scale, due to many higher-order interactions. For example, model
616 M4 contains two time scales as covariates: the passage of time on the across-bin or
617 within-trial time scale (variable period_9), and the passage of time on the across-trial or
618 within-experiment time scale (variable trial_c). However, when trials are presented in
619 blocks, and blocks of trials within sessions, and when the experiment comprises three
620 sessions, then four time scales can be defined (across-bin or within-trial, across-trial or
621 within-block, across-block or within-session, and across-session or within-experiment).
622 From a theoretical perspective, adding more than 1 time scale is important to capture
623 plasticity (e.g., proactive control) and other learning effects that play out on such longer
624 time scales (across-trials, across-blocks, across-sessions), and that are probably present in
625 each experiment in general. From a practical perspective, therefore, it might be interesting
626 for interpretational purposes to limit the number of experimental variables, because adding
627 time scales quickly increases model complexity.

628 5.3 Individual differences

629 One important issue is that of possible individual differences in the overall location of
630 the distribution, and the time course of psychological effects. For example, when you wait
631 for a response of the participant on each trial, you allow the participant to have control
632 over the trial duration, and some participants might respond only when they are confident
633 that their emitted response will be correct. These issues can be avoided by introducing a
634 (relatively short) response deadline in each trial, e.g., 600 ms for simple detection tasks,
635 1000 ms for more difficult discrimination tasks, or 2 s for tasks requiring extended
636 high-level processing. Because EHA can deal in a straightforward fashion with
637 right-censored observations (i.e., trials without an observed response), introducing a

638 response deadline is recommended when designing RT experiments. Furthermore,
639 introducing a response deadline and asking participants to respond before the deadline as
640 much as possible, will also lead to individual distributions that overlap in time, which is
641 important when selecting a common analysis time window when fitting hazard models.

642 But even when using a response deadline, participants can differ qualitatively in the
643 effects they display (see Panis, 2020). One way to deal with this is to describe and
644 interpret the different patterns. Another way is to run a clustering algorithm on the
645 individual hazard estimates across all conditions. The obtained dendrogram can then be
646 used to identify a (hopefully big) cluster of participants that behave similarly, and to
647 identify a (hopefully small) cluster of participants with outlying behavioral patterns. One
648 might then exclude the outlying participants before fitting a hazard model.

649 Another approach: fit models to individual subjects and describe prevalence... REF

650 5.4 Limitation(s)

651 Compared to the orthodox method – comparing mean-averages between conditions –
652 the most important limitation of multilevel hazard modeling is that it might take a long
653 time to estimate the parameters using Bayesian methods or the model might have to be
654 simplified significantly to use frequentist methods.

655 Another issue is that you need a relatively large number of trials per condition to
656 estimate the hazard function with high temporal resolution. Indeed, in general, there is a
657 trade-off between the number of trials per condition and the temporal resolution (i.e., bin
658 width) of the hazard function. Therefore, we recommend researchers to collect as many
659 trials as possible per experimental condition, given the available resources and considering
660 the participant experience (e.g., fatigue and boredom). For instance, if the maximum
661 session length deemed reasonable is between 1 and 2 hours, what is the maximum number
662 of trials per condition that you could reasonably collect? After consideration, it might be

663 worth conducting multiple testing sessions per participant and/or reducing the number of
664 experimental conditions. Finally, there is a user-friendly online tool for calculating
665 statistical power as a function of the number of trials as well as the number of participants,
666 and this might be worth consulting to guide the research design process (Baker et al., 2021).

667 We did not discuss continuous-time hazard analysis. As indicated by Allison (2010),
668 learning discrete-time methods first will help in learning continuous-time methods. Given
669 that RT is typically treated as a continuous variable, it is possible that continuous-time
670 methods will ultimately prevail. However, they require much more data to estimate the
671 continuous-time hazard (rate) function well. Thus, by trading a bit of temporal resolution
672 for a lower number of trials, discrete-time methods seem ideal for dealing with typical
673 psychological time-to-event data sets for which there are less than ~200 trials per condition
674 per experiment.

675 **5.5 Extensions**

676 The hazard models in this tutorial assume that there is one event of interest. For RT
677 data, this event constitutes a single transition between an “idle” state and a “responded”
678 state. However, in certain situations, more than one event of interest might exist. For
679 example, in a medical or health-related context, an individual might transition back and
680 forth between a “healthy” state and a “depressed” state, before a final “death” state.
681 When you have data on the timing of these transitions, one can apply multi-state models,
682 which generalize survival analysis to transitions between three or more states (Steele,
683 Goldstein, & Browne, 2004). Also, the predictor variables in this tutorial are
684 time-invariant, i.e., their value did not change over the course of a trial. Thus, another
685 extension is to include time-varying predictors, i.e., predictors whose value can change
686 across the time bins within a trial (@ Allison, 2010). For example, when gaze position is
687 tracked during a visual search trial, the gaze-target distance will vary during a trial when
688 the eyes move around before a manual response is given; shorter gaze-target distances

689 should be associated with a higher hazard of response occurrence.

690 **6. Conclusions**

691 RT and accuracy distributions are a rich source of information on the time course of
692 cognitive processing, which have been largely undervalued in the history of experimental
693 psychology and cognitive neuroscience. We hope that by providing a set of hands-on,
694 step-by-step tutorials, which come with custom-built and freely available code, researchers
695 will feel more comfortable embracing event history analysis and investigating the temporal
696 profile of cognitive states. On a broader level, we think that wider adoption of such
697 approaches will have a meaningful impact on the inferences drawn from data, as well as the
698 development of theories regarding the structure of cognition.

699

References

- 700 Allison, P. D. (1982). Discrete-Time Methods for the Analysis of Event Histories.
701 *Sociological Methodology*, 13, 61. <https://doi.org/10.2307/270718>
- 702 Allison, P. D. (2010). *Survival analysis using SAS: A practical guide* (2. ed). Cary, NC:
703 SAS Press.
- 704 Aust, F. (2019). *Citr: 'RStudio' add-in to insert markdown citations*. Retrieved from
705 <https://github.com/crsh/citr>
- 706 Aust, F., & Barth, M. (2023). *papaja: Prepare reproducible APA journal articles with R*
707 *Markdown*. Retrieved from <https://github.com/crsh/papaja>
- 708 Baker, D. H., Vilidaite, G., Lygo, F. A., Smith, A. K., Flack, T. R., Gouws, A. D., &
709 Andrews, T. J. (2021). Power contours: Optimising sample size and precision in
710 experimental psychology and human neuroscience. *Psychological Methods*, 26(3),
711 295–314. <https://doi.org/10.1037/met0000337>
- 712 Barth, M. (2023). *tinylabes: Lightweight variable labels*. Retrieved from
713 <https://cran.r-project.org/package=tinylabes>
- 714 Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects
715 models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
716 <https://doi.org/10.18637/jss.v067.i01>
- 717 Bates, D., Maechler, M., & Jagan, M. (2024). *Matrix: Sparse and dense matrix classes and*
718 *methods*. Retrieved from <https://CRAN.R-project.org/package=Matrix>
- 719 Bengtsson, H. (2021). A unifying framework for parallel and distributed processing in r
720 using futures. *The R Journal*, 13(2), 208–227. <https://doi.org/10.32614/RJ-2021-048>
- 721 Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan.
722 *Journal of Statistical Software*, 80(1), 1–28. <https://doi.org/10.18637/jss.v080.i01>
- 723 Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms.
724 *The R Journal*, 10(1), 395–411. <https://doi.org/10.32614/RJ-2018-017>
- 725 Bürkner, P.-C. (2021). Bayesian item response modeling in R with brms and Stan. *Journal*

- 726 *of Statistical Software*, 100(5), 1–54. <https://doi.org/10.18637/jss.v100.i05>
- 727 Eddelbuettel, D., & Balamuta, J. J. (2018). Extending R with C++: A Brief Introduction
728 to Rcpp. *The American Statistician*, 72(1), 28–36.
729 <https://doi.org/10.1080/00031305.2017.1375990>
- 730 Eddelbuettel, D., & François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal
731 of Statistical Software*, 40(8), 1–18. <https://doi.org/10.18637/jss.v040.i08>
- 732 Gabry, J., Češnovar, R., Johnson, A., & Broder, S. (2024). *Cmdstanr: R interface to
733 'CmdStan'*. Retrieved from <https://github.com/stan-dev/cmdstanr>
- 734 Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2019). Visualization
735 in bayesian workflow. *J. R. Stat. Soc. A*, 182, 389–402.
736 <https://doi.org/10.1111/rssa.12378>
- 737 Girard, J. (2024). *Standist: What the package does (one line, title case)*. Retrieved from
738 <https://github.com/jmgirard/standist>
- 739 Grolemund, G., & Wickham, H. (2011). Dates and times made easy with lubridate.
740 *Journal of Statistical Software*, 40(3), 1–25. Retrieved from
741 <https://www.jstatsoft.org/v40/i03/>
- 742 Holden, J. G., Van Orden, G. C., & Turvey, M. T. (2009). Dispersion of response times
743 reveals cognitive dynamics. *Psychological Review*, 116(2), 318–342.
744 <https://doi.org/10.1037/a0014849>
- 745 Kantowitz, B. H., & Pachella, R. G. (2021). The Interpretation of Reaction Time in
746 Information-Processing Research 1. *Human Information Processing*, 41–82.
747 <https://doi.org/10.4324/9781003176688-2>
- 748 Kay, M. (2023). *tidybayes: Tidy data and geoms for Bayesian models*.
749 <https://doi.org/10.5281/zenodo.1308151>
- 750 Kelso, J. A. S., Dumas, G., & Tognoli, E. (2013). Outline of a general theory of behavior
751 and brain coordination. *Neural Networks: The Official Journal of the International
752 Neural Network Society*, 37, 120–131. <https://doi.org/10.1016/j.neunet.2012.09.003>

- 753 Kurz, A. S. (2023a). *Applied longitudinal data analysis in brms and the tidyverse* (version
754 0.0.3). Retrieved from <https://bookdown.org/content/4253/>
- 755 Kurz, A. S. (2023b). *Statistical rethinking with brms, ggplot2, and the tidyverse: Second*
756 *edition* (version 0.4.0). Retrieved from <https://bookdown.org/content/4857/>
- 757 Luce, R. D. (1991). *Response times: Their role in inferring elementary mental organization*
758 (1. issued as paperback). Oxford: Univ. Press.
- 759 McElreath, R. (2018). *Statistical Rethinking: A Bayesian Course with Examples in R and*
760 *Stan* (1st ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9781315372495>
- 761 Meyer, D. E., Osman, A. M., Irwin, D. E., & Yantis, S. (1988). Modern mental
762 chronometry. *Biological Psychology*, 26(1-3), 3–67.
763 [https://doi.org/10.1016/0301-0511\(88\)90013-0](https://doi.org/10.1016/0301-0511(88)90013-0)
- 764 Müller, K., & Wickham, H. (2023). *Tibble: Simple data frames*. Retrieved from
765 <https://CRAN.R-project.org/package=tibble>
- 766 Neuwirth, E. (2022). *RColorBrewer: ColorBrewer palettes*. Retrieved from
767 <https://CRAN.R-project.org/package=RColorBrewer>
- 768 Panis, S. (2020). How can we learn what attention is? Response gating via multiple direct
769 routes kept in check by inhibitory control processes. *Open Psychology*, 2(1), 238–279.
770 <https://doi.org/10.1515/psych-2020-0107>
- 771 Panis, S., Moran, R., Wolkersdorfer, M. P., & Schmidt, T. (2020). Studying the dynamics
772 of visual search behavior using RT hazard and micro-level speed–accuracy tradeoff
773 functions: A role for recurrent object recognition and cognitive control processes.
774 *Attention, Perception, & Psychophysics*, 82(2), 689–714.
775 <https://doi.org/10.3758/s13414-019-01897-z>
- 776 Panis, S., Schmidt, F., Wolkersdorfer, M. P., & Schmidt, T. (2020). Analyzing Response
777 Times and Other Types of Time-to-Event Data Using Event History Analysis: A Tool
778 for Mental Chronometry and Cognitive Psychophysiology. *I-Perception*, 11(6),
779 2041669520978673. <https://doi.org/10.1177/2041669520978673>

- 780 Panis, S., & Schmidt, T. (2016). What Is Shaping RT and Accuracy Distributions? Active
781 and Selective Response Inhibition Causes the Negative Compatibility Effect. *Journal of*
782 *Cognitive Neuroscience*, 28(11), 1651–1671. https://doi.org/10.1162/jocn_a_00998
- 783 Panis, S., & Schmidt, T. (2022). When does “inhibition of return” occur in spatial cueing
784 tasks? Temporally disentangling multiple cue-triggered effects using response history
785 and conditional accuracy analyses. *Open Psychology*, 4(1), 84–114.
786 <https://doi.org/10.1515/psych-2022-0005>
- 787 Panis, S., Torfs, K., Gillebert, C. R., Wagemans, J., & Humphreys, G. W. (2017).
788 Neuropsychological evidence for the temporal dynamics of category-specific naming.
789 *Visual Cognition*, 25(1-3), 79–99. <https://doi.org/10.1080/13506285.2017.1330790>
- 790 Panis, S., & Wagemans, J. (2009). Time-course contingencies in perceptual organization
791 and identification of fragmented object outlines. *Journal of Experimental Psychology:*
792 *Human Perception and Performance*, 35(3), 661–687.
793 <https://doi.org/10.1037/a0013547>
- 794 Pedersen, T. L. (2024). *Patchwork: The composer of plots*. Retrieved from
795 <https://CRAN.R-project.org/package=patchwork>
- 796 Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in s and s-PLUS*. New York:
797 Springer. <https://doi.org/10.1007/b98882>
- 798 R Core Team. (2024). *R: A language and environment for statistical computing*. Vienna,
799 Austria: R Foundation for Statistical Computing. Retrieved from
800 <https://www.R-project.org/>
- 801 Singer, J. D., & Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling
802 Change and Event Occurrence*. Oxford, New York: Oxford University Press.
- 803 Smith, P. L., & Little, D. R. (2018). Small is beautiful: In defense of the small-N design.
804 *Psychonomic Bulletin & Review*, 25(6), 2083–2101.
805 <https://doi.org/10.3758/s13423-018-1451-8>
- 806 Steele, F., Goldstein, H., & Browne, W. (2004). A general multilevel multistate competing

- 807 risks model for event history data, with an application to a study of contraceptive use
808 dynamics. *Statistical Modelling*, 4(2), 145–159.
- 809 <https://doi.org/10.1191/1471082X04st069oa>
- 810 Townsend, J. T. (1990). Truth and consequences of ordinal differences in statistical
811 distributions: Toward a theory of hierarchical inference. *Psychological Bulletin*, 108(3),
812 551–567. <https://doi.org/10.1037/0033-2909.108.3.551>
- 813 Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics.
814 *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)
- 815 Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New
816 York. Retrieved from <https://ggplot2.tidyverse.org>
- 817 Wickham, H. (2023a). *Forcats: Tools for working with categorical variables (factors)*.
818 Retrieved from <https://forcats.tidyverse.org/>
- 819 Wickham, H. (2023b). *Stringr: Simple, consistent wrappers for common string operations*.
820 Retrieved from <https://stringr.tidyverse.org>
- 821 Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., ...
822 Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43),
823 1686. <https://doi.org/10.21105/joss.01686>
- 824 Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *Dplyr: A
825 grammar of data manipulation*. Retrieved from <https://dplyr.tidyverse.org>
- 826 Wickham, H., & Henry, L. (2023). *Purrr: Functional programming tools*. Retrieved from
827 [https://purrr.tidyverse.org/](https://purrr.tidyverse.org)
- 828 Wickham, H., Hester, J., & Bryan, J. (2024). *Readr: Read rectangular text data*. Retrieved
829 from <https://readr.tidyverse.org>
- 830 Wickham, H., Vaughan, D., & Girlich, M. (2024). *Tidyr: Tidy messy data*. Retrieved from
831 <https://tidyr.tidyverse.org>

832

Supplementary material833 **A. Definitions of discrete-time hazard, survivor, and conditional accuracy**834 **functions**

835 The shape of a distribution of waiting times can be described in multiple ways (Luce,
 836 1991). After dividing time in discrete, contiguous time bins indexed by t , let RT be a
 837 discrete random variable denoting the rank of the time bin in which a particular person's
 838 response occurs in a particular experimental condition. Discrete-time EHA focuses on the
 839 discrete-time hazard function

$$840 \quad h(t) = P(RT = t | RT \geq t) \quad (1)$$

841 and the discrete-time survivor function

$$842 \quad S(t) = P(RT > t) = [1-h(t)].[1-h(t-1)].[1-h(t-2)] \dots [1-h(1)] \quad (2)$$

843 and not on the probability mass function

$$844 \quad P(t) = P(RT = t) = h(t).S(t-1) \quad (3)$$

845 nor the cumulative distribution function

$$846 \quad F(t) = P(RT \leq t) = 1-S(t) \quad (4)$$

847 The discrete-time hazard function of event occurrence gives you the probability that
 848 the event occurs (sometime) in bin t , given that the event has not occurred yet in previous
 849 bins. While the discrete-time hazard function assesses the unique risk of event occurrence
 850 associated with each time bin, the discrete-time survivor function cumulates the bin-by-bin
 851 risks of event *nonoccurrence* to obtain the probability that the event occurs after bin t . The
 852 probability mass function cumulates the risk of event occurrence in bin t with the risks of
 853 event nonoccurrence in bins 1 to $t-1$. From equation 3 we find that hazard in bin t is equal
 854 to $P(t)/S(t-1)$.

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

856 discrete-time conditional accuracy function

857 $ca(t) = P(\text{correct} \mid RT = t)$ (5)

858 which gives you the probability that a response is correct given that it is emitted in time

859 bin t (Allison, 2010; Kantowitz & Pachella, 2021; Wickelgren, 1977). This latter function is

860 also known as the micro-level speed-accuracy tradeoff (SAT) function.

861 The survivor function provides a context for the hazard function, as $S(t-1) = P(RT >$

862 $t-1) = P(RT \geq t)$ tells you on how many percent of the trials the estimate $h(t) = P(RT =$

863 $t \mid RT \geq t$) is based. The probability mass function provides a context for the conditional

864 accuracy function, as $P(t) = P(RT = t)$ tells you on how many percent of the trials the

865 estimate $ca(t) = P(\text{correct} \mid RT = t)$ is based.

866 When time is treated as a continuous variable, let RT be a continuous random variable

867 denoting a particular person's response time in a particular experimental condition.

868 Because waiting times can only increase, continuous-time EHA does not focus on the

869 cumulative distribution function $F(t) = P(RT \leq t)$ and its derivative, the probability

870 density function $f(t) = F(t)'$, but on the survivor function $S(t) = P(RT > t)$ and the

871 hazard rate function $\lambda(t) = f(t)/S(t)$. The hazard rate function gives you the instantaneous

872 *rate* of event occurrence at time point t , given that the event has not occurred yet.

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

874 We defined 13 custom functions that we list here.

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

- 879 • setup_lt_2IV(ptb) : set up a life table for each combination of levels of 2
880 independent variables
- 881 • calc_ca(df) : estimate the conditinal accuracies when there is 1 independent variable
- 882 • calc_ca_2IV(df) : estimate the conditional accuraiies when there are 2 independent
883 variables
- 884 • join_lt_ca(df1,df2) : add the ca(t) estimates to the life tables (1 independent
885 variable)
- 886 • join_lt_ca_2IV(df1,df2) : add the ca(t) estimates to the life tables (2 independent
887 variables)
- 888 • extract_median(df) : estimate quantiles S(t).50 (1 independent variable)
- 889 • extract_median_2IV(df) : estimate quantiles S(t).50 (2 independent variables)
- 890 • plot_eha(df,subj,haz_yaxis) : create plots of the discrete-time functions (1
891 independent variable)
- 892 • plot_eha_2IV(df,subj,haz_yaxis) : create plots of the discrete-time functions (2
893 independent variables)
- 894 • plot_eha_agg(df,subj,haz_yaxis) : create 1 plot for aggregated data (1 independent
895 variable)

896 When you want to analyse simple RT data from a detection experiment with one
897 independent variable, the functions calc_ca() and join_lt_ca() should not be used, and
898 the code to plot the conditional accuracy functions should be removed from the function
899 plot_eha(). When you want to analyse simple RT data from a detection experiment with
900 two independent variables, the functions calc_ca_2IV() and join_lt_ca_2IV() should not
901 be used, and the code to plot the conditional accuracy functions should be removed from
902 the function plot_eha_2IV().

903 **C. Link functions**

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

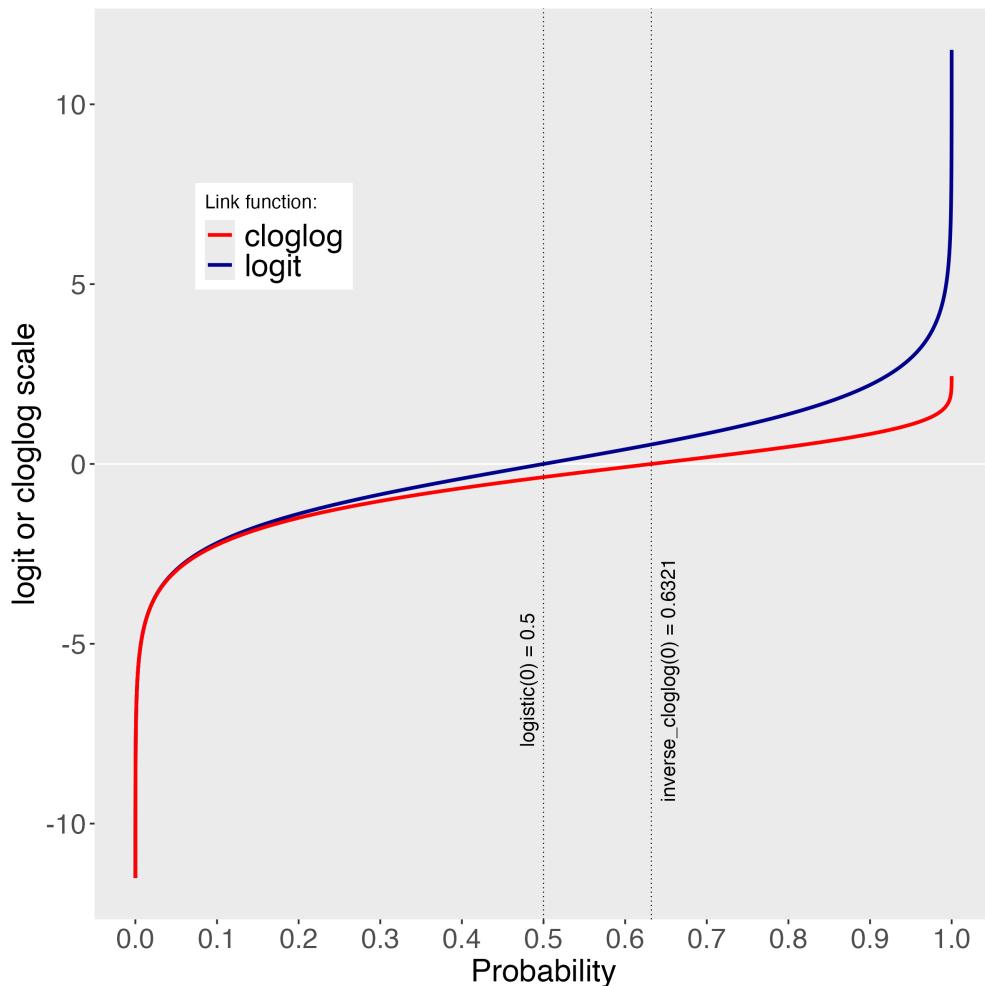


Figure 10. The logit and cloglog link functions.

906 **D. Regression equations**

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

$$910 \quad \text{cloglog}[h(t)] = \ln(-\ln[1-h(t)]) = [\alpha_1 \text{ONE} + \alpha_2(\text{TIME}-9) + \alpha_3(\text{TIME}-9)^2] + [\beta_1 X_1 +$$

$$911 \quad \beta_2 X_2 + \beta_3 X_2(\text{TIME}-9)]$$

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

913 in this example. The first set of terms within brackets, the alpha parameters multiplied by

914 their polynomial specifications of (centered) time, represents the shape of the baseline

915 cloglog-hazard function (i.e., when all predictors X_i take on a value of zero). The second set

916 of terms (the beta parameters) represents the vertical shift in the baseline cloglog-hazard

917 for a 1 unit increase in the respective predictor variable. Predictors can be discrete,

918 continuous, and time-varying or time-invariant. For example, the effect of a 1 unit increase

919 in X_1 is to vertically shift the whole baseline cloglog-hazard function by β_1 cloglog-hazard

920 units. However, if the predictor interacts linearly with TIME (see X_2 in the example), then

921 the effect of a 1 unit increase in X_2 is to vertically shift the predicted cloglog-hazard in bin

922 9 by β_2 cloglog-hazard units (when $\text{TIME}-9 = 0$), in bin 10 by $\beta_2 + \beta_3$ cloglog-hazard

923 units (when $\text{TIME}-9 = 1$), and so forth. To interpret the effects of a predictor, its β

924 parameter is exponentiated, resulting in a hazard ratio (due to the use of the cloglog link).

925 When using the logit link, exponentiating a β parameter results in an odds ratio.

926 An example (single-level) discrete-time hazard model with a general specification for

927 TIME (separate intercepts for each of six bins, where D1 to D6 are binary variables

928 identifying each bin) and a single predictor (X_1) can be written as follows:

$$929 \quad \text{cloglog}[h(t)] = \ln(-\ln[1-h(t)]) = [\alpha_1 D1 + \alpha_2 D2 + \alpha_3 D3 + \alpha_4 D4 + \alpha_5 D5 + \alpha_6 D6] +$$

$$930 \quad [\beta_1 X_1]$$

931 E. Prior distributions

932 To gain a sense of what prior *logit* values would approximate a uniform distribution

933 on the probability scale, Kurz (2023a) simulated a large number of draws from the

934 Uniform(0,1) distribution, converted those draws to the log-odds metric, and fitted a

935 Student's t distribution. Row C in Figure 4 shows that using a t-distribution with 7.61
 936 degrees of freedom and a scale parameter of 1.57 as a prior on the logit scale, approximates
 937 a uniform distribution on the probability scale. According to Kurz (2023a), such a prior
 938 might be a good prior for the intercept(s) in a logit-hazard model, while the $N(0,1)$ prior in
 939 row D might be a good prior for the non-intercept parameters in a logit-hazard model, as it
 940 gently regularizes p towards .5 (i.e., a zero effect on the logit scale).

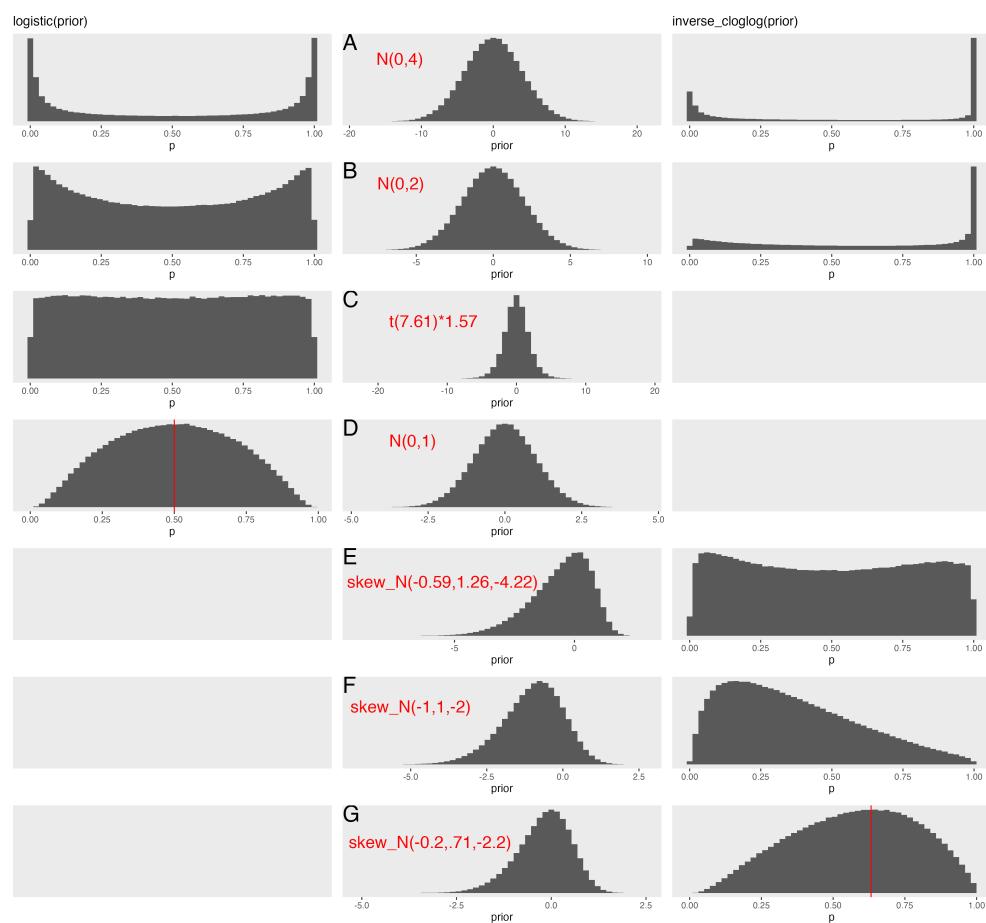


Figure 11. Prior distributions on the logit and/or cloglog scales (middle column), and their implications on the probability scale after applying the inverse-logit (or logistic) transformation (left column), and the inverse-cloglog transformation (right column).

941 To gain a sense of what prior *cloglog* values would approximate a uniform distribution
 942 on the hazard probability scale, we followed Kurz's approach and simulated a large number

943 of draws from the Uniform(0,1) distribution, converted them to the cloglog metric, and
944 fitted a skew-normal model (due to the asymmetry of the cloglog link function). Row E
945 shows that using a skew-normal distribution with a mean of -0.59, a standard deviation of
946 1.26, and a skewness of -4.22 as a prior on the cloglog scale, approximates a uniform
947 distribution on the probability scale. However, because hazard values below .5 are more
948 likely in RT studies, using a skew-normal distribution with a mean of -1, a standard
949 deviation of 1, and a skewness of -2 as a prior on the cloglog scale (row F), might be a good
950 weakly informative prior for the intercept(s) in a cloglog-hazard model. A skew-normal
951 distribution with a mean of -0.2, a standard deviation of 0.71, and a skewness of -2.2 might
952 be a good weakly informative prior for the non-intercept parameters in a cloglog-hazard
953 model as it gently regularizes p towards .6321 (i.e., a zero effect on the cloglog scale).