

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

3 Sven Panis<sup>1</sup> & Richard Ramsey<sup>1</sup>

4 <sup>1</sup> 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

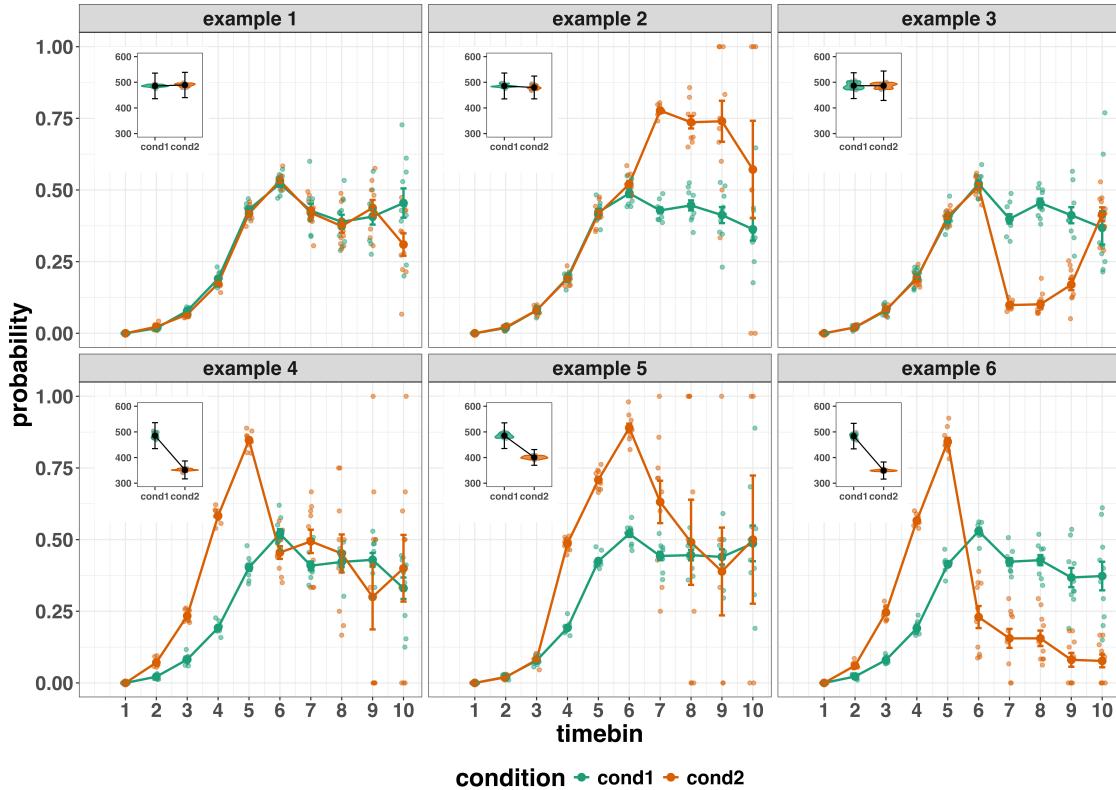
- 40 Event History Analyses for psychological time-to-event data: A tutorial in R with examples  
41 in Bayesian and frequentist workflows

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 (insert REF). Such information is useful not only for interpretation of the effects, but also  
53 for cognitive psychophysiology and computational model selection (Panis, Schmidt,  
54 Wolkersdorfer, & Schmidt, 2020).

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

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

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

78 From a historical perspective, it is worth noting that the development of analytical  
79 tools that can estimate or predict when events will occur is not a new endeavour. Indeed,  
80 hundreds of years ago, analytical methods were developed to predict time to death (REFs).  
81 The same logic has been applied to psychological time-to-event data, as previously  
82 demonstrated (Panis et al., 2020). Here, in the paper, we hope to show the value of EHA  
83 for knowledge and theory building in cognitive psychology and related areas of research,  
84 such as cognitive neuroscience, as well as provide practical tutorials that provide  
85 step-by-step code and instructions in the hope that we can enable others to use EHA in a  
86 more routine, efficient and effective manner.

## 87 1.2 Aims and structure of the paper

88 In this paper, we focus on a distributional method known as discrete-time event  
89 history analysis, a.k.a. hazard analysis, duration analysis, failure-time analysis, survival  
90 analysis, and transition analysis. We first provide a brief overview of hazard analysis to  
91 orient the reader to the basic concepts that we will use throughout the paper. However,  
92 this will remain relatively short, as this has been covered in detail before Singer and  
93 Willett (2003), Allison (1982), and Allison (2010), and our primary aim here is to  
94 introduce a set of tutorials, which explain **how** to do such analyses, rather than repeat in  
95 any detail **why** you should do them.

96 We then provide four different tutorials, each of which is written in the R

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

104 In Tutorial 1, we illustrate how to process or “wrangle” a previously published RT  
105 dataset to calculate descriptive statistics when there is one independent variable. The  
106 descriptive statistics are plotted, and we comment on their interpretation. In Tutorial 2, we  
107 illustrate how one can fit Bayesian multi-level regression hazard models to the data using  
108 the R package brms. We discuss possible link functions, and plot the model-based effects of  
109 our predictors of interest. In Tutorial 3, we illustrate how to fit the same type of regression  
110 hazard models in a frequentist framework using the R package lme4. We then briefly  
111 compare and contrast these inferential frameworks when applied to EHA. In Tutorial 4, we  
112 provide a generalisation of the approach to illustrate one might describe statistics when  
113 using a more complex design, such as when there are two independent variables.

114 In summary, even though event history analyses is a widely used statistical tool and  
115 there already exist many excellent reviews (REFs) and tutorials (REFs) on its general  
116 use-cases, we are not aware of any tutorials that are aimed at psychological time-to-event  
117 data, and which provide worked examples of the key data processing and multi-level  
118 regression modelling steps. Therefore, our ultimate goal is twofold: first, we want to  
119 convince readers of the many benefits of using hazard analysis when dealing with  
120 time-to-event data with a focus on psychological time-to-event data, and second, we want  
121 to provide a set of practical tutorials, which provide step-by-step instructions on how you  
122 actually perform hazard analysis.

## 2. A brief introduction to hazard analysis

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

### 2.1 Basic features of hazard analysis

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

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

The definition of hazard and the type of models employed depend on whether one is using continuous or discrete time units. Since our focus here is on hazard models that use discrete time units, we describe that approach. After dividing time in discrete, contiguous time bins indexed by  $t$  (e.g.,  $t = 1:10$  timebins), let  $RT$  be a discrete random variable denoting the rank of the time bin in which a particular person's response occurs in a particular experimental condition. For example, the first response could occur in 550 ms and it would be in timebin 6 (any RTs from 501 ms to 600).

146 Discrete-time EHA focuses on the discrete-time hazard function and the discrete-time  
 147 survivor function (Figure X). The equations that define both of these functions are  
 148 reported in supplementary materials (Supp XX). The discrete-time hazard probability  
 149 gives you the probability that the event occurs (sometime) in bin t, given that the event  
 150 has not occurred yet in previous bins. In contrast, the discrete-time survivor function  
 151 cumulates the bin-by-bin risks of event *nonoccurrence* to obtain the probability that the  
 152 event occurs after bin t. In other words, the survivor function reflects the likelihood that  
 153 the event occurs in a subsequent timebin.

154 The survivor function can help to qualify or provide context to the interpretation of  
 155 the hazard function. For example, it can give a sense of how many trials may contribute to  
 156 that part of the distribution. If each participant completes 100 trials in an experiment, and  
 157 the survivor function prob of 0.03, then only 3% of trials remain beyond this point, which  
 158 in this case would amount to 3 trials. Therefore, the error bars in this part of the  
 159 distribution would be wider and less precise compared to other parts.

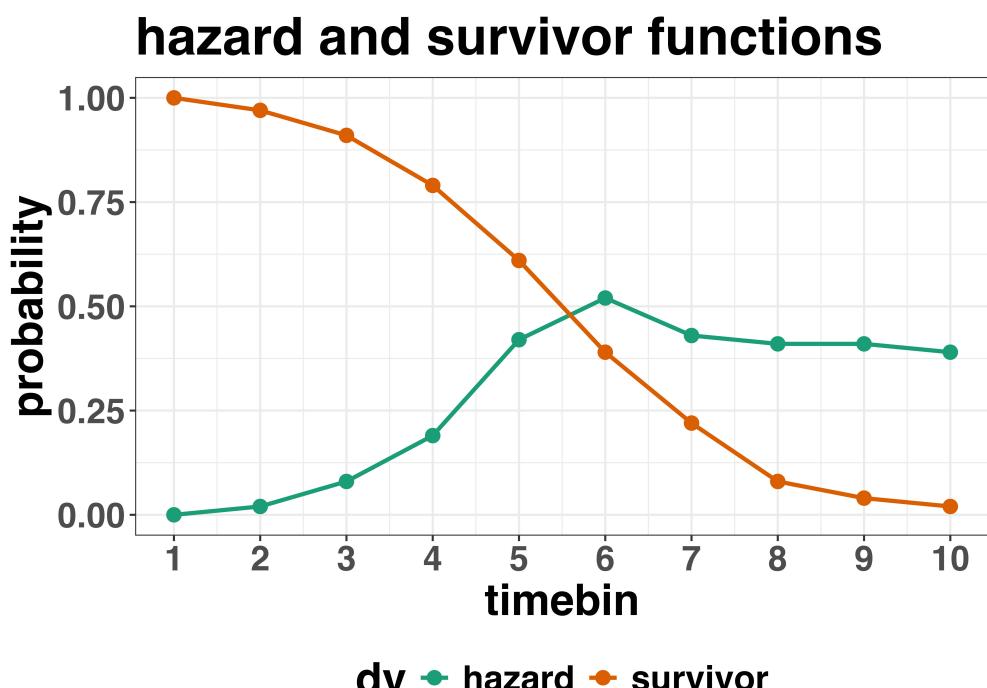


Figure 2. Hazard and survivor functions

160 **2.2 Hazard analysis in the context of experimental psychology**

161       **2.2.1 A worked example.** In the context of experimental psychology, it is  
162 common for participants to be presented with a task that has a right and a wrong answer.  
163 For example, a task may involve choosing between two response options with only one of  
164 them being correct. For such two-choice RT data, the discrete-time hazard function can be  
165 extended with the discrete-time conditional accuracy function (see equ. X in Supps), which  
166 gives you the probability that a response is correct given that it has been emitted in time  
167 bin  $t$  (Allison, 2010; Kantowitz & Pachella, 2021; Wickelgren, 1977).

168       Integrating results between hazard and conditional accuracy functions can be  
169 informative for understanding psychological processes. To illustrate, we consider a  
170 hypothetical example that is inspired by real data (Panis et al., 2016), but simplified to  
171 make the main point clearer (Figure 3). In a standard stimulus-response-compatibility  
172 paradigm, there is a prime stimulus (e.g., a double arrow) followed by a target stimulus  
173 (single arrow). The prime can then be congruent or incongruent with the target. Taken  
174 together, the results show that for early responses (< bin 6), responses always follow the  
175 prime (and not the target, as instructed). And then for later responses, the prime can be  
176 overridden, as both conditions are now always correct. This is interesting because  
177 mean-average RT would only represent the overall ability of cognition to overcome  
178 interference, on average, across trials. And such a conclusion is not supporting when the  
179 effects are explored over a timeline. Instead, the psychological conclusion is much more  
180 nuanced and suggests that multiple states start, stop and possibly interact over a  
181 particular temporal window.

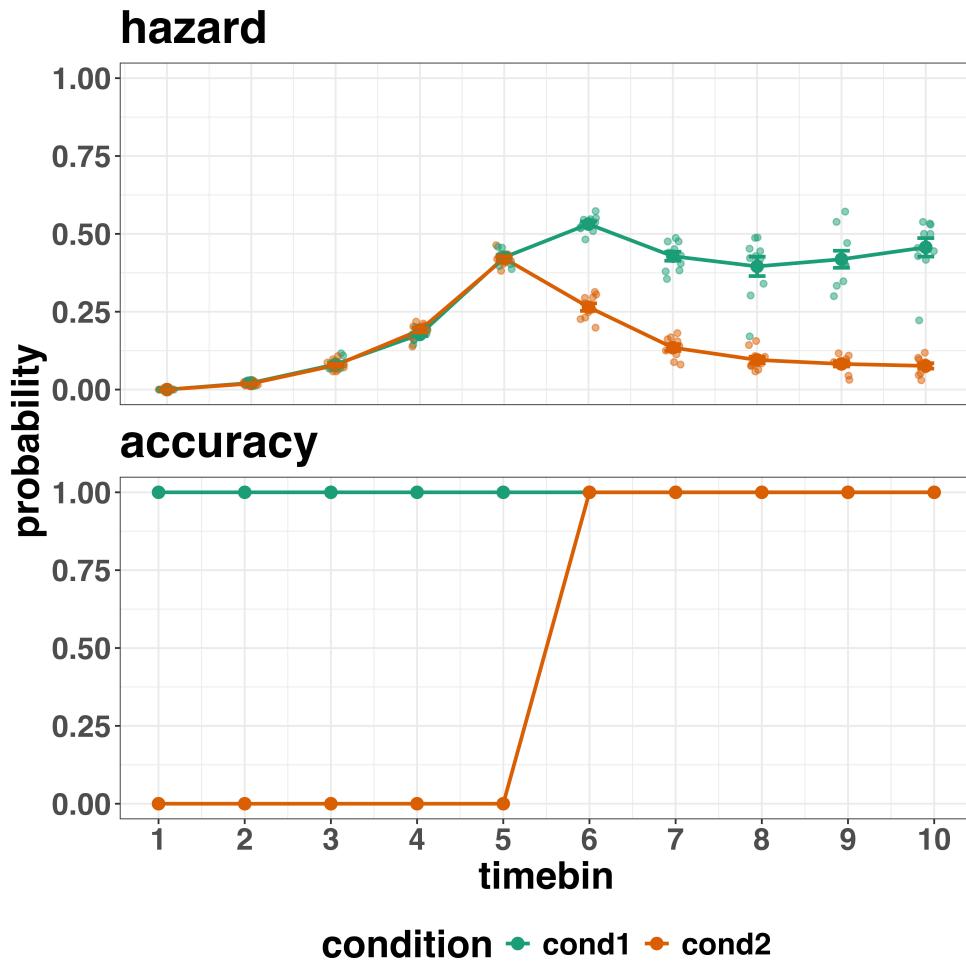


Figure 3. Hazard and conditional accuracy

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

191        **2.2.2 Implications for designing experiments.** Performing hazard analyses in

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

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

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

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

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

214        We used R (Version 4.4.0; R Core Team, 2024)<sup>1</sup> for all reported analyses. Web links

215 are printed in bold. The content of the tutorials is mainly based on Allison (2010), Singer

---

<sup>1</sup> 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*

<sup>216</sup> and Willett (2003), McElreath (2018), Kurz (2023a), and Kurz (2023b).

<sup>217</sup> **3. An overview of the general analytical workflow**

<sup>218</sup> Although the focus is on EHA, we also want to briefly comment on broader aspects of  
<sup>219</sup> our general analytical workflow, which relate more to data science and data analysis  
<sup>220</sup> workflows.

<sup>221</sup> **3.1 Data science workflow and descriptive statistics**

<sup>222</sup> Descriptive, data science workflow. Data wrangling via tidyverse principles and a  
<sup>223</sup> functional programming approach (cite R4DS textbook here). Functional programming  
<sup>224</sup> basically means you don't write your own loops but instead use functions that have been  
<sup>225</sup> built and tested by others. [[more here, as necessary]].

<sup>226</sup> **3.2 Inferential statistical approach**

<sup>227</sup> Our lab adopts a estimation approach to multi-level regression (Kruschke & Liddel,  
<sup>228</sup> 2018; Winter, 2019), which is heavily influenced by Bayesian approach as suggested by  
<sup>229</sup> 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).

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

236

#### 4. Tutorials

237 [[provide a short summary of the aims and scope of each tutorial, as well as the links  
238 between them]]. Additionally, to further simplify the process for other users, the tutorials  
239 rely on a set of our own user-defined functions that make sub-processes easier to automate,  
240 such as data wrangling and plotting functions.

241 Then a list of tutorials:

242 1a. Wrangle raw data and descriptive stats (T1). 1b.

243 2a. 2b.

244 3a. 3b.

245 Inferential stats (T2 and T3).

246 Generalisation (T4). Should this be online in Supps?? It would make the main text  
247 shorter and simpler, but make it no less available. We could just have a sentence at the end  
248 of T1, which says that we provide a generalisation and extension in T4, which is in Supps.

249 Planning (T5) - if we get a simulation and power analysis script working, which we  
250 are happy with then we could include it here.

251 **4.1 Tutorial 1: Calculating descriptive statistics using a life table**

252       **4.1.1 Data wrangling aims.** Our data wrangling procedures serve two related  
253      purposes. First, we want to summarise and visualise descriptive statistics that relate to our  
254      main research questions. Second, we want to produce two different datasets that can each  
255      be submitted to different types of inferential modelling approaches. The two types of data  
256      structure we label as ‘person-trial’ data (Table 1) and ‘person-trial-bin’ data (Table 2).

257      The ‘person-trial’ data will be familiar to most researchers who record behavioural  
258      responses from participants, as it represents the measured RT and accuracy per trial within  
259      an experiment. In contrast, the ‘person-trial-bin’ data has a different, more extended  
260      structure, which indicates in which bin a response occurred, if at all, in each trial.

261      Therefore, the ‘person-trial-bin’ dataset generates a 0 in each bin until an event occurs and  
262      then it generates a 1 to signal an event has occurred. It is worth pointing out that there is  
263      no requirement for an event to occur at all (in any bin), as maybe there was no response on  
264      that trial or the event occurred after the timewindow of interest. Likewise, the event could  
265      occur in bin 1 there would only be 1 row of data for that trial.

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.

<sup>266</sup>        **4.1.2 A real data wrangling example.** To illustrate how to quickly set up life  
<sup>267</sup> tables for calculating the descriptive statistics (functions of discrete time), we use a  
<sup>268</sup> published data set on masked response priming from Panis and Schmidt (2016). In their  
<sup>269</sup> first experiment, Panis and Schmidt (2016) presented a double arrow for 94 ms that  
<sup>270</sup> pointed left or right as the target stimulus with an onset at time point zero in each trial.  
<sup>271</sup> Participants had to indicate the direction in which the double arrow pointed using their  
<sup>272</sup> corresponding index finger, within 800 ms after target onset. Response time and accuracy  
<sup>273</sup> were recorded on each trial. Prime type (blank, congruent, incongruent) and mask type  
<sup>274</sup> were manipulated. Here we focus on the subset of trials in which no mask was presented.  
<sup>275</sup> The 13-ms prime stimulus was a double arrow with onset at -187 ms for the congruent

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

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

280 The required column names are as follows:

- 281 • “pid”, indicating unique participant IDs;
- 282 • “trial”, indicating each unique trial per participant;
- 283 • “condition”, a factor indicating the levels of the independent variable (1, 2, ...) and  
 284 the corresponding labels;
- 285 • “rt”, indicating the response times in ms;
- 286 • “acc”, indicating the accuracies (1/0).

287 In the code of Tutorial 1, 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")))
```

288 Next, we can set up the life tables and plots of the discrete-time functions  $h(t)$ ,  $S(t)$   
 289 and  $ca(t)$ . To do so using a functional programming approach, one has to nest the data  
 290 within participants using the `group_nest()` function, and supply a user-defined censoring  
 291 time and bin width to our 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
```

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

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

303 The next step is to plot the data using our custom made plotting tool `plot_eha()`.

304 When creating the plots, some warning messages will likely be generated, like these:

- 305 • Removed 2 rows containing missing values or values outside the scale range  
306     (`geom_line()`).
- 307 • Removed 2 rows containing missing values or values outside the scale range  
308     (`geom_point()`).
- 309 • Removed 2 rows containing missing values or values outside the scale range  
310     (`geom_segment()`).

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

312 estimates, and no error bars. They can thus safely be ignored. One can now inspect  
313 different aspects, including the life table for a particular condition of a particular subject,  
314 and a plot of the different functions for a particular participant.

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

316 compare to Figure 1. A life table includes for each time bin, the risk set (number of trials  
317 that are event-free at the start of the bin), the number of observed events, and the

318 estimates of  $h(t)$ ,  $S(t)$ ,  $ca(t)$  and their estimated standard errors (se). At time point zero,  
319 no events can occur and therefore  $h(t)$  and  $ca(t)$  are undefined.

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

325 For example, for participant 6, the estimated hazard values in bin (240,280] are 0.03,  
326 0.17, and 0.20 for the blank, congruent, and incongruent prime conditions, respectively. In  
327 other words, when the waiting time has increased until *240 ms* after target onset, then the  
328 conditional probability of response occurrence in the next 40 ms is more than five times  
329 larger for both prime-present conditions, compared to the blank prime condition.

330 Furthermore, the estimated conditional accuracy values in bin (240,280] are 0.29, 1,  
331 and 0 for the blank, congruent, and incongruent prime conditions, respectively. In other  
332 words, if a response is emitted in bin (240,280], then the probability that it is correct is  
333 estimated to be 0.29, 1, and 0 for the blank, congruent, and incongruent prime conditions,  
334 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.

Subject 6

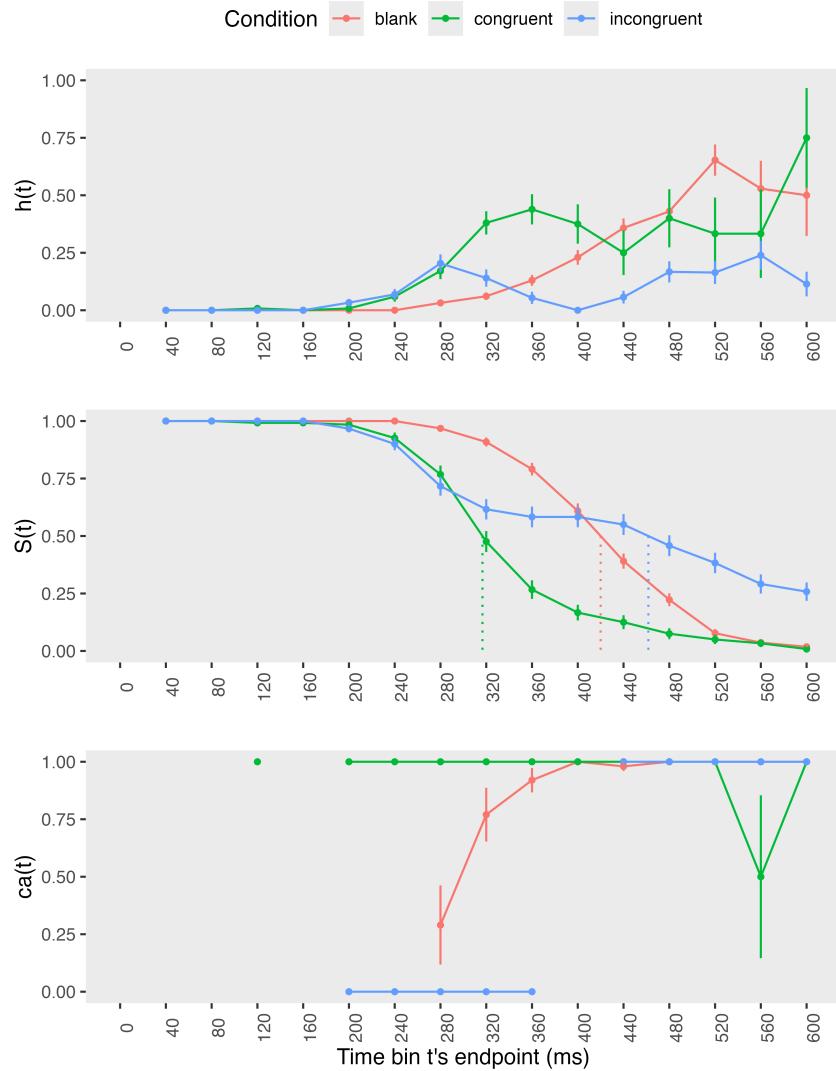


Figure 4. Estimated discrete-time hazard, survivor, and conditional accuracy functions for participant 6, as a function of the passage of discrete waiting time.

335 However, when the waiting time has increased until 400 ms after target onset, then  
 336 the conditional probability of response occurrence in the next 40 ms is estimated to be  
 337 0.36, 0.25, and 0.06 for the blank, congruent, and incongruent prime conditions,  
 338 respectively. And when a response does occur in bin (400,440], then the probability that it  
 339 is correct is estimated to be 0.98, 1, and 1 for the blank, congruent, and incongruent prime

340 conditions, respectively.

341 These results suggest that the participant is initially responding to the prime even  
342 though (s)he was instructed to only respond to the target, that response competition  
343 emerges in the incongruent prime condition around 300 ms, and that only later response  
344 are fully controlled by the target stimulus. Qualitatively similar results were obtained for  
345 the other five participants. These results go against the (often implicit) assumption that all  
346 observed responses are primed responses to the target stimulus.

347 At this point, we have calculated, summarised and plotted descriptive statistics for  
348 the key variables in EHA. As we will show in Tutorials 2 and 3, statistical models for  $h(t)$   
349 can be implemented as generalized linear mixed regression models predicting event  
350 occurrence (1/0) in each bin of a selected time range. As such multi-level regression is  
351 what we turn to in the next tutorials.

## 352 4.2 Tutorial 2: Fitting Bayesian hazard models

353 In this second tutorial, we illustrate how to fit Bayesian hazard regression models to  
354 the masked response priming data set used in the first tutorial. Fitting (Bayesian or  
355 non-Bayesian) regression models to the data is important when you want to study how the  
356 shape of the hazard function depends on various predictors (Singer & Willett, 2003).

357 **4.2.1 Hazard model considerations.** There are several analytic decisions one  
358 has to make when fitting a hazard model. First, one has to select an analysis time window,  
359 i.e., a contiguous set of bins for which there is enough data for each participant. Second,  
360 given that the dependent variable is binary, one has to select a link function (see Supps).  
361 The cloglog link is preferred over the logit link when events can occur in principle at any  
362 time point within a bin, which is the case for RT data (Singer & Willett, 2003). Third, one  
363 has to choose a specification of the effect of discrete TIME (i.e., the time bin index  $t$ ). One  
364 can choose a general specification (one intercept per bin) or a functional specification, such

365 as a polynomial one (compare model 1 with models 2, 3, and 4 below). We provide relevant  
 366 example regression formulas in supplementary materials.

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

374 In general, there are three assumptions one can make or relax when adding  
 375 experimental predictor variables: The linearity assumption for continuous predictors (the  
 376 effect of a 1 unit change is the same anywhere on the scale), the additivity assumption  
 377 (predictors do not interact), and the proportionality assumption (predictors do not interact  
 378 with TIME).

379 In this tutorial we will fit four Bayesian multilevel models (i.e., generalized linear  
 380 mixed models) to the person-trial-bin oriented data set that we created in Tutorial 1. We  
 381 select the analysis range (200,600] and the cloglog link. The data is prepared as follows.

```
# load person-trial-bin oriented data set
ptb_data <- read_csv("../Tutorial_1_descriptive_stats/data/inputfile_hazard_modeling.csv")

# select analysis time 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))

```

#### 382        4.2.2 Prior distributions. To get the posterior distribution of each parameters

383        given the data, we need to specify a prior distribution for each parameter. The middle  
 384        column of Supplementary Figure 4 shows seven examples of prior distributions on the logit  
 385        and/or cloglog scales.

386        While a normal distribution with relatively large variance is often used as a weakly  
 387        informative prior for continuous dependent variables, rows A and B in Figure 3 show that  
 388        specifying such distributions on the logit and cloglog scales leads to rather informative  
 389        distributions on the original probability (i.e., discrete-time hazard) scale, as most mass is  
 390        pushed to probabilities of 0 and 1.

#### 391        4.2.3 Model 1: A general specification of TIME, and main effects of

392        congruency and trial number. [[Here let's give some intuition on why we would want  
 393        to setup the model like this]]

394        For the first model, we use a general specification of TIME (i.e., one intercept per  
 395        time bin) for the baseline condition (blank prime), and assume that the effects of  
 396        prime-target congruency and trial number are proportional and additive, and that the  
 397        effect of trial number is linear. Before we fit model 1, we remove unnecessary columns from  
 398        the data, and specify our priors. In the code of Tutorial 2, this is accomplished as follows.

```

# remove unnecessary columns before fitting a model
M1_data <- ptb_data %>% select(-c(bl,tr,trial,period, period_9,d9))

```

```
# Specify priors
priors_M1 <- c(
  set_prior("skew_normal(-0.2,0.71,-2.2)", class = "b"),
  set_prior("skew_normal(-1,1,-2)", class = "b", coef = "d6"),
  set_prior("skew_normal(-1,1,-2)", class = "b", coef = "d7"),
  set_prior("skew_normal(-1,1,-2)", class = "b", coef = "d8"),
  set_prior("skew_normal(-1,1,-2)", class = "b", coef = "d10"),
  set_prior("skew_normal(-1,1,-2)", class = "b", coef = "d11"),
  set_prior("skew_normal(-1,1,-2)", class = "b", coef = "d12"),
  set_prior("skew_normal(-1,1,-2)", class = "b", coef = "d13"),
  set_prior("skew_normal(-1,1,-2)", class = "b", coef = "d14"),
  set_prior("skew_normal(-1,1,-2)", class = "b", coef = "d15"),
  set_prior("skew_normal(-1,1,-2)", class = "b", coef = "Intercept"),
  set_prior("normal(0, 1)", class = "sd"),
  set_prior("lkj(2)", class = "cor")
)
```

399 We can now estimate our first Bayesion regression model, 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")
```

400 Estimating model M1 took about 70 minutes on a MacBook Pro (Sonoma 14.6.1 OS,

401 18GB Memory, M3 Pro Chip).

402 **4.2.4 Model 2: A polynomial specification of TIME, and main effects of**  
 403 **congruency and trial number.** [[Here let's give some intuition on why we would want

404 to modify the formula and model features]]

405 For the second model, we use a third-order polynomial specification of TIME for the  
406 baseline condition (blank prime), and again assume that the effects of prime-target  
407 congruency and trial number are proportional and additive, and that the effect of trial  
408 number is linear. We first remove unnecessary columns and specify our priors.

409 Estimating model M2 took about 144 minutes.

410 **4.2.5 Model 3: A polynomial specification of TIME, and relaxing the**  
411 **proportionality assumption.** [[Here let's give some intuition on why we would want to  
412 modify the formula and model features]]

413 For the third model, we use a third-order polynomial specification of TIME for the  
414 baseline condition (blank prime), and relax the proportionality assumption for the  
415 predictor variables congruency (variable "condition") and trial number (variable "trial\_c").  
416 We use the same data set and priors as for model 2.

417 Estimating model M3 took about 268 minutes.

418 **4.2.6 Model 4: A polynomial specification of TIME, and relaxing all three**  
419 **assumptions.** Based on previous work (Panis, 2020; Panis, Moran, Wolkersdorfer, &  
420 Schmidt, 2020; Panis & Schmidt, 2022; Panis, Torfs, Gillebert, Wagemans, & Humphreys,  
421 2017; Panis & Wagemans, 2009), we relax all three assumptions in model 4. We use the  
422 same data set and priors as for model 2.

423 Estimating model M4 took about 8 hours.

424 **4.2.7 Compare the models.** We can compare the four models using the Widely  
425 Applicable Information Criterion (WAIC) and Leave-One-Out (LOO) cross-validation, and  
426 look at model weights (Kurz, 2023a; McElreath, 2018).

427 Clearly, both weighting schemes prefer model M4.

428       **4.2.8 Evaluate parameter estimates.** Figure 5 shows the effects of congruent  
 429 and incongruent primes relative to neutral primes, for each time bin in trial number 1000  
 430 for the selected model.

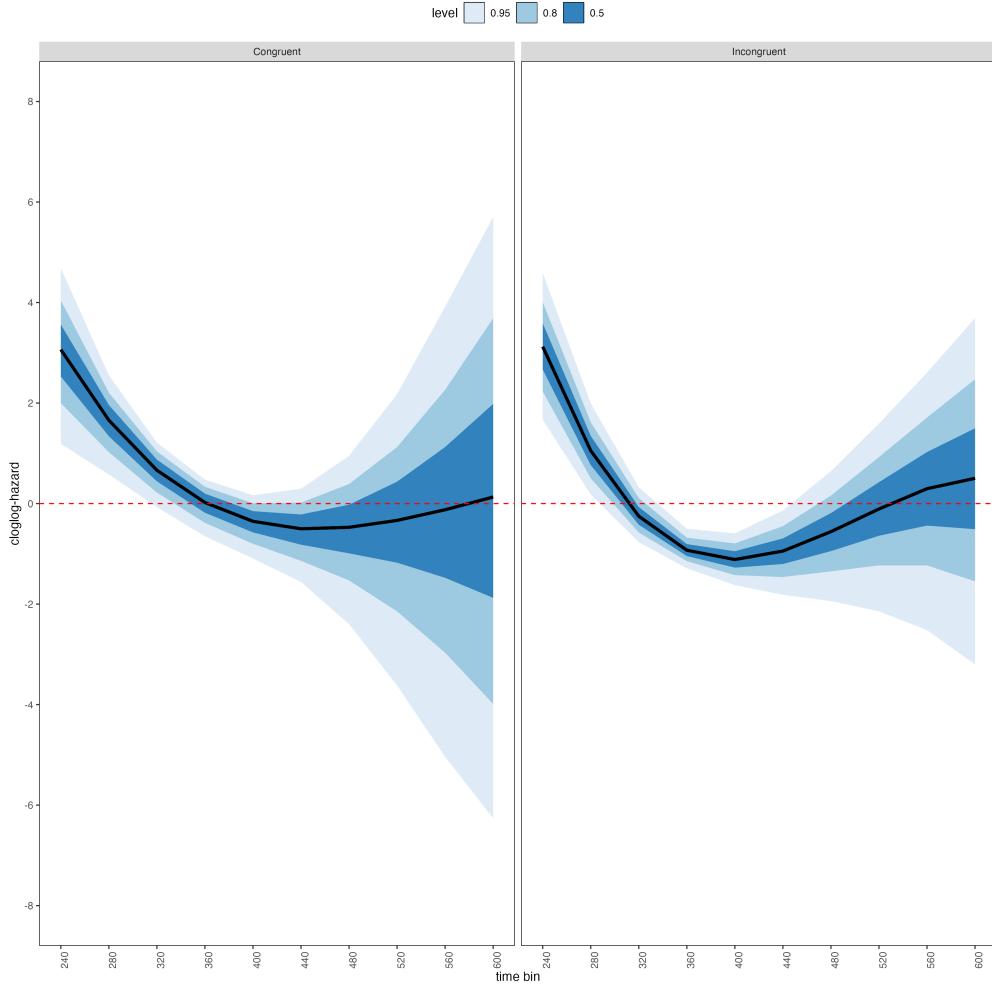
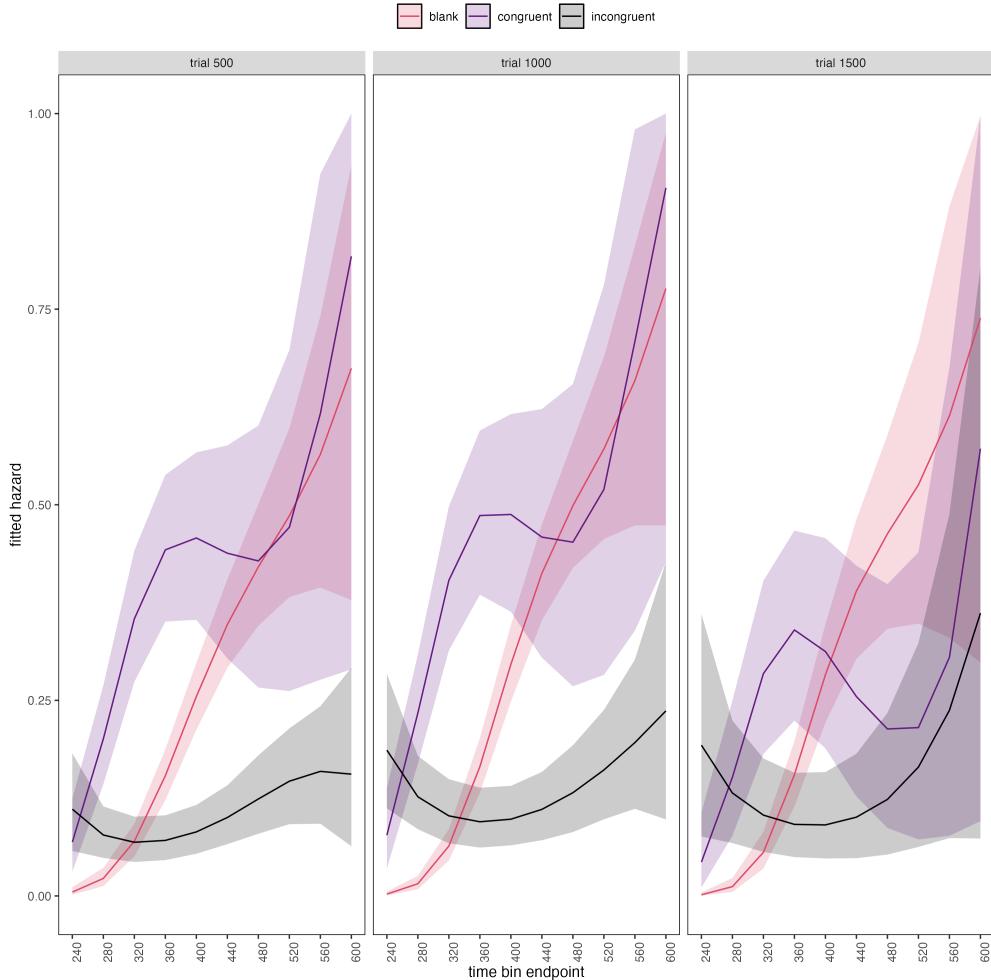


Figure 5. 50/80/95 percentile intervals of the draws from the posterior distributions representing the effect of congruent and incongruent primes relative to neutral primes in trial number 1000.

431       Figure 6 shows the model-based hazard functions for each prime type for participant  
 432 6, in trial 500, 1000, and 1500.



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

433

[[let's have a paragraph on how we might interpret these plots.]]

434

### 4.3 Tutorial 3: Fitting Frequentist hazard models

435

In this third tutorial we illustrate how to fit a multilevel hazard regression model in

436

the frequentist framework, for the data set used in the first tutorial. For illustration

437

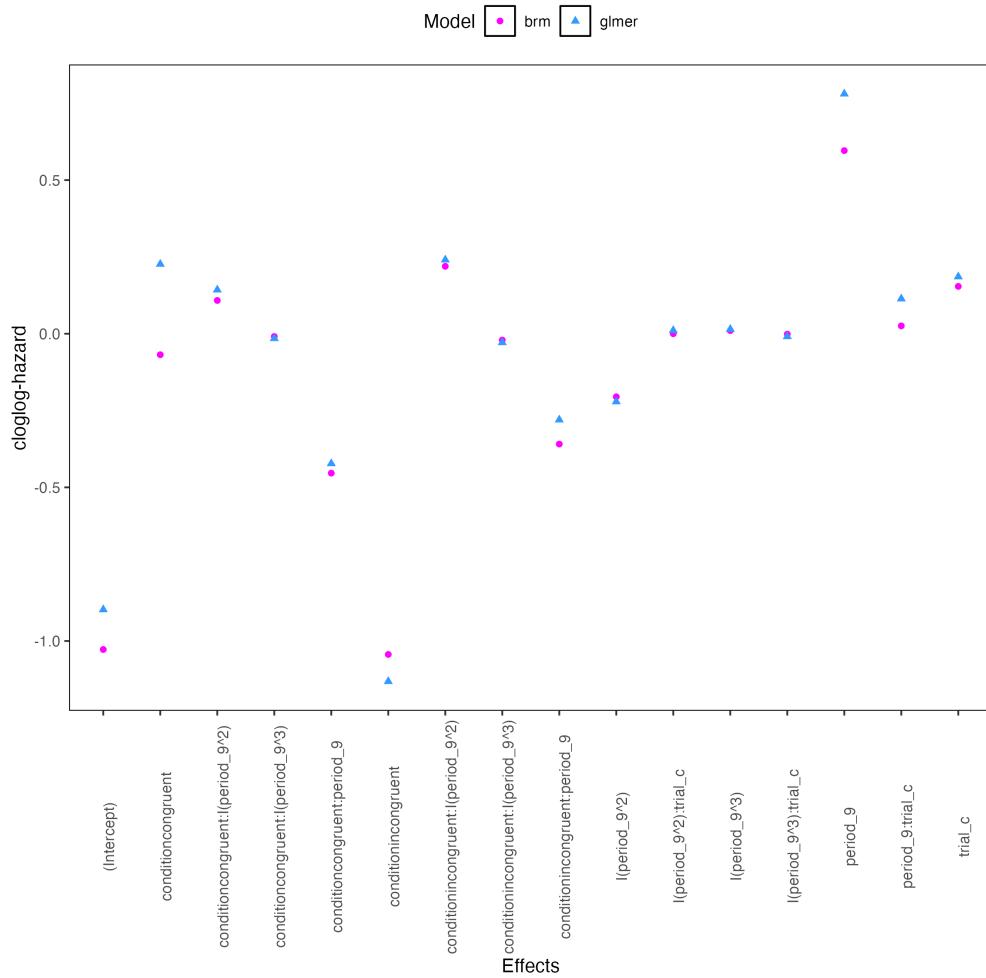
purposes, we only fitted model M3 using the function `glmer()` from the package `lme4`.

438

In Figure 7 we compare the parameter estimates of model M3 from `brm()` with those

439

of `glmer()`.



*Figure 7.* Parameter estimates for model M3 from brm() and glmer().

440       Figure 7 confirms that the parameter estimates from both Bayesian and frequentist  
 441       models are pretty similar. However, the random effects structure of model M3 was already  
 442       too complex for the frequentist model as it did not converge and resulted in a singular fit.  
 443       This is of course one of the reasons why Bayesian modeling has become so popular in  
 444       recent years. But the price you pay for being able to fit more complex models in a  
 445       Bayesian framework is computation time. In other words, as we have noted throughout,  
 446       some of the Bayesian models in Tutorial 2 took several hours to build.

**447 4.4 Tutorial 4: Generalising to a more complex design**

448 So far in this paper, we have used a simple experimental design, which involved one  
449 condition with two levels. But psychological experiments are often more complex, with  
450 crossed factorial designs with more conditions and more than two levels. The purpose of  
451 Tutorial 4, therefore, is to provide a generalisation of the basic approach, which extends to  
452 a more complicated design. We felt that this might be useful for researchers in  
453 experimental psychology that typically use crossed factorial designs.

454 To this end, Tutorial 4 illustrates how to calculate and plot the descriptive statistics  
455 for the full data set of Experiment 1 of Panis and Schmidt (2016), which includes two  
456 independent variables: mask type and prime type. As we use the same functional  
457 programming approach as in Tutorial 1, we simply present the sample-based functions for  
458 participant 6 in Figure 8. Note the negative compatibility effect in the hazard and  
459 conditional accuracy functions when a (relevant, irrelevant, or lines) mask is present.



*Figure 8.* 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).

## 5. Discussion

This main motivation for writing this paper is the observation that event history analysis remains under-used in psychological research, which means the field of research is not taking full advantage of the many benefits EHA provides compared to more conventional analyses. By providing a freely available set of tutorials, which provide step-by-step guidelines and ready-to-use R code, we hope that researchers will feel more comfortable using EHA in the future. Indeed, we hope that our tutorials may help to overcome a barrier to entry with EHA, which is the increase in analytical complexity compared to mean-average comparisons. While we have focused here on within-subject, factorial, small-*N* designs, it is important to realize that event history analysis can be

470 applied to other designs as well (large- $N$  designs with only one measurement per subject,  
471 between-subject designs, etc.). As such, the general workflow and associated code can be  
472 modified and applied more broadly to other contexts and research questions. In the  
473 following, we discuss issues relating to individual differences, limitations of the approach,  
474 and future extensions.

475 **5.1 Advantages of hazard analysis**

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

484 [[This para needs to be way shorter and easier to read or we get rid of it]] Second,  
485 because RT distributions may differ from one another in multiple ways, Townsend (1990)  
486 developed a dominance hierarchy of statistical differences between two arbitrary  
487 distributions A and B. For example, if  $F_A(t) > F_B(t)$  for all  $t$ , then both cumulative  
488 distribution functions are said to show a complete ordering. Townsend (1990) showed that  
489 a complete ordering on the hazard functions — $\lambda_A(t) > \lambda_B(t)$  for all  $t$ — implies a complete  
490 ordering on both the cumulative distribution and survivor functions — $F_A(t) > F_B(t)$  and  
491  $S_A(t) < S_B(t)$ — which in turn implies an ordering on the mean latencies —mean A <  
492 mean B. In contrast, an ordering on two means does *not* imply a complete ordering on the  
493 corresponding  $F(t)$  and  $S(t)$  functions, and a complete ordering on these latter functions  
494 does *not* imply a complete ordering on the corresponding hazard functions. This means  
495 that stronger conclusions can be drawn from data when comparing the hazard functions

496 using EHA. For example, when mean A < mean B, the hazard functions might show a  
497 complete ordering (i.e., for all t), a partial ordering (e.g., only for  $t > 300$  ms, or only for  $t$   
498 < 500 ms), or they may cross each other one or more times. As a result, instead of using  
499 delta-plots for RT – differences in quantiles from  $F(t)^{-1}$  – one can simply plot delta-h(t)  
500 functions (see Panis, 2020).

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

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

518 Finally, as explained by Kelso, Dumas, and Tognoli (2013), it is crucial to first have a  
519 precise description of the macroscopic behavior of a system (here:  $h(t)$  and  $ca(t)$  functions)  
520 in order to know what to derive on the microscopic level. EHA can thus solve the problem  
521 of model mimicry, i.e., the fact that different computational models can often predict the

522 same mean RTs as observed in the empirical data, but not necessarily the detailed shapes  
523 of the empirical RT hazard distributions. Also, fitting parametric functions or  
524 computational models to data without studying the shape of the empirical discrete-time  
525  $h(t)$  and  $ca(t)$  functions can miss important features in the data (Panis, Moran, et al.,  
526 2020; Panis & Schmidt, 2016).

## 527 5.2 Individual differences

528 One important issue is that of possible individual differences in the overall location of  
529 the distribution, and the time course of psychological effects. For example, when you wait  
530 for a response of the participant on each trial, you allow the participant to have control  
531 over the trial duration, and some participants might respond only when they are confident  
532 that their emitted response will be correct. These issues can be avoided by introducing a  
533 (relatively short) response deadline in each trial, e.g., 600 ms for simple detection tasks,  
534 1000 ms for more difficult discrimination tasks, or 2 s for tasks requiring extended  
535 high-level processing. Because EHA can deal in a straightforward fashion with  
536 right-censored observations (i.e., trials without an observed response), introducing a  
537 response deadline is recommended when designing RT experiments. Furthermore,  
538 introducing a response deadline and asking participants to respond before the deadline as  
539 much as possible, will also lead to individual distributions that overlap in time, which is  
540 important when selecting a common analysis time window when fitting hazard models.

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

548 **5.3 Limitation(s)**

549       Compared to the orthodox method – comparing mean-averages between conditions –  
550   the most important limitation of multilevel hazard modeling is that it might take a long  
551   time to estimate the parameters using Bayesian methods or the model might have to be  
552   simplified significantly to use frequentist methods. Another issue is that you need a  
553   relatively large number of trials per condition to estimate the hazard function with high  
554   temporal resolution. Indeed, in general, there is a trade-off between the number of trials  
555   per condition and the temporal resolution (i.e., bin width) of the hazard function.  
556   Therefore, we recommend researchers to collect as many trials as possible per experimental  
557   condition, given the available resources and considering the participant experience (e.g.,  
558   fatigue and boredom). For instance, if the maximum session length deemed reasonable is  
559   between 1 and 2 hours, what is the maximum number of trials per condition that you could  
560   reasonably collect? After consideration, it might be worth conducting multiple testing  
561   sessions per participant and/or reducing the number of experimental conditions. Finally,  
562   there is a user-friendly online tool for calculating statistical power as a function of the  
563   number of trials as well as the number of participants, and this might be worth consulting  
564   to guide the research design process (Baker et al., 2021).

565 **5.4 Extensions**

566       The hazard models in this tutorial assume that there is one event of interest. For RT  
567   data, this event constitutes a single transition between an “idle” state and a “responded”  
568   state. However, in certain situations, more than one event of interest might exist. For  
569   example, in a medical or health-related context, an individual might transition back and  
570   forth between a “healthy” state and a “depressed” state, before a final “death” state.  
571   When you have data on the timing of these transitions, one can apply multi-state models,  
572   which generalize survival analysis to transitions between three or more states (Steele,

573 Goldstein, & Browne, 2004). Also, the predictor variables in this tutorial are  
574 time-invariant, i.e., their value did not change over the course of a trial. Thus, another  
575 extension is to include time-varying predictors, i.e., predictors whose value can change  
576 across the time bins within a trial (REF). [[give a concrete example for this latter point]]

577 **6. Conclusions**

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

586

## References

- 587 Allison, P. D. (1982). Discrete-Time Methods for the Analysis of Event Histories.  
588     *Sociological Methodology*, 13, 61. <https://doi.org/10.2307/270718>
- 589 Allison, P. D. (2010). *Survival analysis using SAS: A practical guide* (2. ed). Cary, NC:  
590     SAS Press.
- 591 Aust, F. (2019). *Citr: 'RStudio' add-in to insert markdown citations*. Retrieved from  
592     <https://github.com/crsh/citr>
- 593 Aust, F., & Barth, M. (2023). *papaja: Prepare reproducible APA journal articles with R*  
594     *Markdown*. Retrieved from <https://github.com/crsh/papaja>
- 595 Baker, D. H., Vilidaite, G., Lygo, F. A., Smith, A. K., Flack, T. R., Gouws, A. D., &  
596     Andrews, T. J. (2021). Power contours: Optimising sample size and precision in  
597     experimental psychology and human neuroscience. *Psychological Methods*, 26(3),  
598     295–314. <https://doi.org/10.1037/met0000337>
- 599 Barth, M. (2023). *tinylabes: Lightweight variable labels*. Retrieved from  
600     <https://cran.r-project.org/package=tinylabes>
- 601 Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects  
602     models using lme4. *Journal of Statistical Software*, 67(1), 1–48.  
603     <https://doi.org/10.18637/jss.v067.i01>
- 604 Bates, D., Maechler, M., & Jagan, M. (2024). *Matrix: Sparse and dense matrix classes and*  
605     *methods*. Retrieved from <https://CRAN.R-project.org/package=Matrix>
- 606 Bengtsson, H. (2021). A unifying framework for parallel and distributed processing in r  
607     using futures. *The R Journal*, 13(2), 208–227. <https://doi.org/10.32614/RJ-2021-048>
- 608 Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan.  
609     *Journal of Statistical Software*, 80(1), 1–28. <https://doi.org/10.18637/jss.v080.i01>
- 610 Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms.  
611     *The R Journal*, 10(1), 395–411. <https://doi.org/10.32614/RJ-2018-017>
- 612 Bürkner, P.-C. (2021). Bayesian item response modeling in R with brms and Stan. *Journal*

- 613       *of Statistical Software*, 100(5), 1–54. <https://doi.org/10.18637/jss.v100.i05>
- 614   Eddelbuettel, D., & Balamuta, J. J. (2018). Extending R with C++: A Brief Introduction  
615       to Rcpp. *The American Statistician*, 72(1), 28–36.  
616       <https://doi.org/10.1080/00031305.2017.1375990>
- 617   Eddelbuettel, D., & François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal  
618       of Statistical Software*, 40(8), 1–18. <https://doi.org/10.18637/jss.v040.i08>
- 619   Gabry, J., Češnovar, R., Johnson, A., & Broder, S. (2024). *Cmdstanr: R interface to  
620       'CmdStan'*. Retrieved from <https://github.com/stan-dev/cmdstanr>
- 621   Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2019). Visualization  
622       in bayesian workflow. *J. R. Stat. Soc. A*, 182, 389–402.  
623       <https://doi.org/10.1111/rssa.12378>
- 624   Girard, J. (2024). *Standist: What the package does (one line, title case)*. Retrieved from  
625       <https://github.com/jmgirard/standist>
- 626   Grolemund, G., & Wickham, H. (2011). Dates and times made easy with lubridate.  
627       *Journal of Statistical Software*, 40(3), 1–25. Retrieved from  
628       <https://www.jstatsoft.org/v40/i03/>
- 629   Holden, J. G., Van Orden, G. C., & Turvey, M. T. (2009). Dispersion of response times  
630       reveals cognitive dynamics. *Psychological Review*, 116(2), 318–342.  
631       <https://doi.org/10.1037/a0014849>
- 632   Kantowitz, B. H., & Pachella, R. G. (2021). The Interpretation of Reaction Time in  
633       Information-Processing Research 1. *Human Information Processing*, 41–82.  
634       <https://doi.org/10.4324/9781003176688-2>
- 635   Kay, M. (2023). *tidybayes: Tidy data and geoms for Bayesian models*.  
636       <https://doi.org/10.5281/zenodo.1308151>
- 637   Kelso, J. A. S., Dumas, G., & Tognoli, E. (2013). Outline of a general theory of behavior  
638       and brain coordination. *Neural Networks: The Official Journal of the International  
639       Neural Network Society*, 37, 120–131. <https://doi.org/10.1016/j.neunet.2012.09.003>

- 640 Kurz, A. S. (2023a). *Applied longitudinal data analysis in brms and the tidyverse* (version  
641 0.0.3). Retrieved from <https://bookdown.org/content/4253/>
- 642 Kurz, A. S. (2023b). *Statistical rethinking with brms, ggplot2, and the tidyverse: Second*  
643 *edition* (version 0.4.0). Retrieved from <https://bookdown.org/content/4857/>
- 644 McElreath, R. (2018). *Statistical Rethinking: A Bayesian Course with Examples in R and*  
645 *Stan* (1st ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9781315372495>
- 646 Meyer, D. E., Osman, A. M., Irwin, D. E., & Yantis, S. (1988). Modern mental  
647 chronometry. *Biological Psychology*, 26(1-3), 3–67.  
648 [https://doi.org/10.1016/0301-0511\(88\)90013-0](https://doi.org/10.1016/0301-0511(88)90013-0)
- 649 Müller, K., & Wickham, H. (2023). *Tibble: Simple data frames*. Retrieved from  
650 <https://CRAN.R-project.org/package=tibble>
- 651 Neuwirth, E. (2022). *RColorBrewer: ColorBrewer palettes*. Retrieved from  
652 <https://CRAN.R-project.org/package=RColorBrewer>
- 653 Panis, S. (2020). How can we learn what attention is? Response gating via multiple direct  
654 routes kept in check by inhibitory control processes. *Open Psychology*, 2(1), 238–279.  
655 <https://doi.org/10.1515/psych-2020-0107>
- 656 Panis, S., Moran, R., Wolkersdorfer, M. P., & Schmidt, T. (2020). Studying the dynamics  
657 of visual search behavior using RT hazard and micro-level speed–accuracy tradeoff  
658 functions: A role for recurrent object recognition and cognitive control processes.  
659 *Attention, Perception, & Psychophysics*, 82(2), 689–714.  
660 <https://doi.org/10.3758/s13414-019-01897-z>
- 661 Panis, S., Schmidt, F., Wolkersdorfer, M. P., & Schmidt, T. (2020). Analyzing Response  
662 Times and Other Types of Time-to-Event Data Using Event History Analysis: A Tool  
663 for Mental Chronometry and Cognitive Psychophysiology. *I-Perception*, 11(6),  
664 2041669520978673. <https://doi.org/10.1177/2041669520978673>
- 665 Panis, S., & Schmidt, T. (2016). What Is Shaping RT and Accuracy Distributions? Active  
666 and Selective Response Inhibition Causes the Negative Compatibility Effect. *Journal of*

- 667        *Cognitive Neuroscience*, 28(11), 1651–1671. [https://doi.org/10.1162/jocn\\_a\\_00998](https://doi.org/10.1162/jocn_a_00998)
- 668        Panis, S., & Schmidt, T. (2022). When does “inhibition of return” occur in spatial cueing  
669        tasks? Temporally disentangling multiple cue-triggered effects using response history  
670        and conditional accuracy analyses. *Open Psychology*, 4(1), 84–114.  
671        <https://doi.org/10.1515/psych-2022-0005>
- 672        Panis, S., Torfs, K., Gillebert, C. R., Wagemans, J., & Humphreys, G. W. (2017).  
673        Neuropsychological evidence for the temporal dynamics of category-specific naming.  
674        *Visual Cognition*, 25(1-3), 79–99. <https://doi.org/10.1080/13506285.2017.1330790>
- 675        Panis, S., & Wagemans, J. (2009). Time-course contingencies in perceptual organization  
676        and identification of fragmented object outlines. *Journal of Experimental Psychology:  
677        Human Perception and Performance*, 35(3), 661–687.  
678        <https://doi.org/10.1037/a0013547>
- 679        Pedersen, T. L. (2024). *Patchwork: The composer of plots*. Retrieved from  
680        <https://CRAN.R-project.org/package=patchwork>
- 681        Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in s and s-PLUS*. New York:  
682        Springer. <https://doi.org/10.1007/b98882>
- 683        R Core Team. (2024). *R: A language and environment for statistical computing*. Vienna,  
684        Austria: R Foundation for Statistical Computing. Retrieved from  
685        <https://www.R-project.org/>
- 686        Singer, J. D., & Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling  
687        Change and Event Occurrence*. Oxford, New York: Oxford University Press.
- 688        Smith, P. L., & Little, D. R. (2018). Small is beautiful: In defense of the small-N design.  
689        *Psychonomic Bulletin & Review*, 25(6), 2083–2101.  
690        <https://doi.org/10.3758/s13423-018-1451-8>
- 691        Steele, F., Goldstein, H., & Browne, W. (2004). A general multilevel multistate competing  
692        risks model for event history data, with an application to a study of contraceptive use  
693        dynamics. *Statistical Modelling*, 4(2), 145–159.

- 694 https://doi.org/10.1191/1471082X04st069oa
- 695 Townsend, J. T. (1990). Truth and consequences of ordinal differences in statistical  
696 distributions: Toward a theory of hierarchical inference. *Psychological Bulletin*, 108(3),  
697 551–567. https://doi.org/10.1037/0033-2909.108.3.551
- 698 Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics.  
699 *Acta Psychologica*, 41(1), 67–85. https://doi.org/10.1016/0001-6918(77)90012-9
- 700 Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New  
701 York. Retrieved from https://ggplot2.tidyverse.org
- 702 Wickham, H. (2023a). *Forcats: Tools for working with categorical variables (factors)*.  
703 Retrieved from https://forcats.tidyverse.org/
- 704 Wickham, H. (2023b). *Stringr: Simple, consistent wrappers for common string operations*.  
705 Retrieved from https://stringr.tidyverse.org
- 706 Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., ...  
707 Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43),  
708 1686. https://doi.org/10.21105/joss.01686
- 709 Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *Dplyr: A  
710 grammar of data manipulation*. Retrieved from https://dplyr.tidyverse.org
- 711 Wickham, H., & Henry, L. (2023). *Purrr: Functional programming tools*. Retrieved from  
712 https://purrr.tidyverse.org/
- 713 Wickham, H., Hester, J., & Bryan, J. (2024). *Readr: Read rectangular text data*. Retrieved  
714 from https://readr.tidyverse.org
- 715 Wickham, H., Vaughan, D., & Girlich, M. (2024). *Tidyr: Tidy messy data*. Retrieved from  
716 https://tidyr.tidyverse.org