

1 A tutorial on Bayesian and Frequentist Event History Analyses for psychological  
2 time-to-event data

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).

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

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

13

## Abstract

14 Time-to-event data such as response times, saccade latencies, and fixation durations are  
15 ubiquitous in experimental psychology. The orthodox method for analysing such data –  
16 comparing means with analysis-of-variance – is actually hiding a lot of information about  
17 psychological effects, such as their onset time and duration, and how they evolve with  
18 increasing waiting time. Such information can change key conclusions about psychological  
19 processes and can be revealed by using distributional measures that portray the detailed  
20 shape of time-to-event distributions.

21 Here we provide a set of tutorials on how to implement one particular distributional  
22 method known as discrete-time event history analysis, a.k.a. hazard analysis, duration  
23 analysis, failure-time analysis, survival analysis, and transition analysis. We illustrate how  
24 one can calculate the descriptive statistics, and how one can implement Bayesian and  
25 frequentist regression models, using the R packages tidyverse, brms, and lme4. The R code  
26 is publicly available on Github and OSF, and can easily be adapted for other data sets. We  
27 discuss possible link functions and prior distributions, how to manage inter-individual  
28 differences, implications for experimental design, and the advantages of a hazard analysis  
29 over other distributional methods available in the literature. Our ultimate goal is to  
30 convince readers to start using hazard analysis more often when dealing with time-to-event  
31 data.

32 *Keywords:* response times, event history analysis, Bayesian regression models

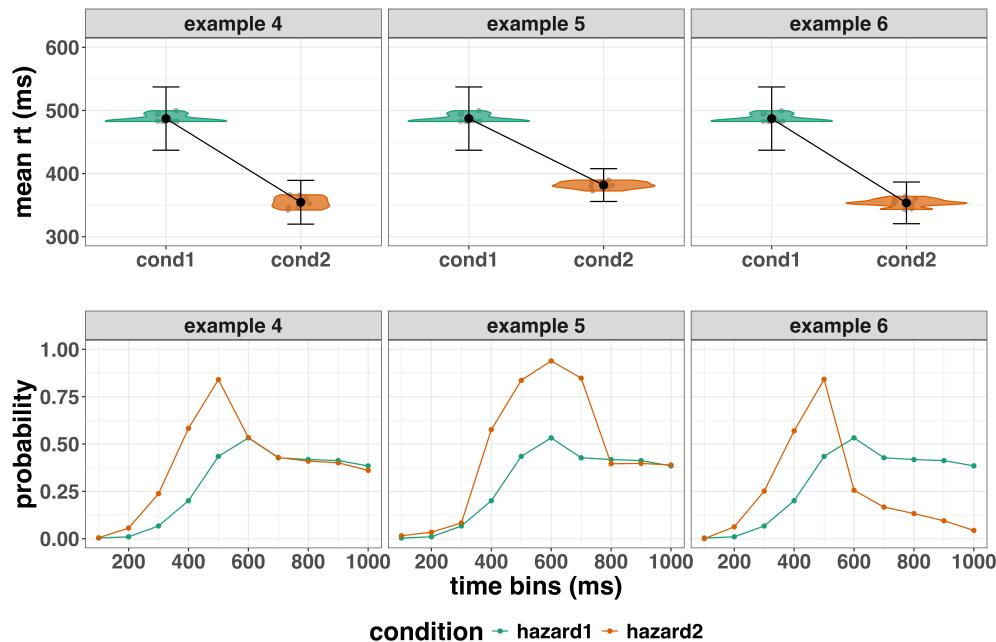
33 Word count: X

34 A tutorial on Bayesian and Frequentist Event History Analyses for psychological  
 35 time-to-event data

36 **Introduction**

37 **Means versus distributional shapes**

38 In experimental psychology, it is still standard practice to analyse response times  
 39 (RTs), saccade latencies, and fixation durations by calculating average performance across  
 40 a series of trials. However, differences in means conceal when an experimental effect starts,  
 41 how long it lasts, how it evolves over increasing waiting time, and whether its onset is  
 42 time-locked to other events. Such information is useful not only for interpretation, but also  
 43 for cognitive psychophysiology and computational model selection (Panis, Schmidt,  
 44 Wolkersdorfer, & Schmidt, 2020). As a simple illustration, Figure 1 shows three examples  
 45 of how an observed difference in mean response times (RTs) between two experimental  
 46 conditions conceals differences in the shapes of the underlying RT distributions.



*Figure 1.* Means versus distributional shapes.

47 In each example, the mean RT is lower in condition 2 compared to condition 1.  
48 However, the distributions in the first example show that the effect starts around 200 ms  
49 and is gone by 600 ms. In the second example, the distributional effect starts around 400  
50 ms and is gone by 800 ms. And in the third example, the distributional effect reverses  
51 around 550 ms.

## 52 Outline of the paper

53 In this paper we focus on a distributional method known as discrete-time event  
54 history analysis, a.k.a. hazard analysis, duration analysis, failure-time analysis, survival  
55 analysis, and transition analysis. We first provide a brief overview of hazard analysis to  
56 orient the reader to the basic concepts and ideas that we will use throughout the paper.  
57 However, this will remain relatively short, and for detailed treatment, see Singer and  
58 Willett (2003), Allison (1982), and Allison (2010).

59 We then provide four different tutorials, each of which is written in R code and  
60 publicly available on Github and the Open Science Framework (OSF). The tutorials  
61 provide hands-on, concrete examples of key parts of the analytical process, so that others  
62 can apply the analyses to their own time-to-event data sets. In Tutorial 1 we illustrate how  
63 to calculate the descriptive statistics for a published data set when there is one  
64 independent variable. The descriptive statistics are plotted, and we comment on their  
65 interpretation. In Tutorial 2 we illustrate how one can fit Bayesian hazard models to the  
66 data. After selecting the best of four models, we plot the model-based effects and the  
67 model fits for a few subjects. In Tutorial 3 we illustrate how to fit hazard models in a  
68 frequentist framework. We compare the model-based effects between Bayesian and  
69 frequentist approaches. In Tutorial 4 we illustrate how to calculate the descriptive  
70 statistics when there are two independent variables.

71                   **Overview of hazard analysis**

72         To apply event history analysis (EHA), one must be able to define the event of  
73     interest (any qualitative change that can be situated in time, e.g., a button press, a saccade  
74     onset, a fixation offset, etc.), time point zero (e.g., target stimulus onset, fixation onset),  
75     and measure the passage of time between time point zero and event occurrence in discrete  
76     or continuous time units. Both the definition of hazard and the type of models employed  
77     depend on whether one is using continuous or discrete time units.

78         The shape of a distribution of waiting times can be described in multiple ways (Luce,  
79     1991). Let  $RT$  be a continuous random variable denoting a particular person's response time  
80     in a particular experimental condition. Because waiting times can only increase,  
81     continuous-time EHA does not focus on the cumulative distribution function  $F(t) = P(RT$   
82      $\leq t)$  and its derivative, the probability density function  $f(t) = F(t)'$ , but on the survivor  
83     function  $S(t) = P(RT > t)$  and the hazard rate function  $\lambda(t) = f(t)/S(t)$ . The hazard rate  
84     function gives you the instantaneous rate of event occurrence at time point  $t$ , given that  
85     the event has not occurred yet.

86         Similarly, after dividing time in discrete, contiguous time bins indexed by  $t$ , let  $RT$  be  
87     a discrete random variable denoting the rank of the time bin in which a particular person's  
88     response occurs in a particular experimental condition. Discrete-time EHA focuses on the  
89     discrete-time hazard function  $h(t) = P(RT = t | RT \geq t)$  and the discrete-time survivor  
90     function  $S(t) = P(RT > t) = [1-h(t)].[1-h(t-1)].[1-h(t-2)] \dots [1-h(1)]$ , and not on the  
91     probability mass function  $p(t) = h(t).S(t-1)$  and the cumulative distribution function  $F(t)$   
92      $= 1-S(t)$ . The discrete-time hazard probability function gives you the probability that the  
93     event occurs (sometime) in bin  $t$ , given that the event has not occurred yet in previous  
94     bins. Unlike the discrete-time hazard function, which assesses the unique risk associated  
95     with each time bin, the discrete-time survivor function cumulates the bin-by-bin risks of  
96     event *nonoccurrence*.

97 For two-choice RT data, the discrete-time hazard function can be extended with the  
98 discrete-time conditional accuracy function  $ca(t) = P(\text{correct} \mid RT = t)$ , which gives you  
99 the probability that a response is correct given that it has been emitted in time bin  $t$   
100 (Allison, 2010; Kantowitz & Pachella, 2021; Wickelgren, 1977). This latter function is also  
101 known as the micro-level speed-accuracy tradeoff function.

102 We focus on factorial within-subject designs in which a large number of observations  
103 are made on a relatively small number of participants (small- $N$  designs). This approach  
104 emphasizes the precision and reproducibility of data patterns at the individual participant  
105 level to increase the inferential validity of the design (Baker et al., 2021; Smith & Little,  
106 2018). In contrast to the large- $N$  design that averages across many participants without  
107 being able to scrutinize individual data patterns, small- $N$  designs retain crucial  
108 information about the data patterns of individual observers. This is of great advantage  
109 whenever participants differ systematically in their strategies or in the time-courses of their  
110 effects, so that blindly averaging them would lead to misleading data patterns. Indeed,  
111 Smith and Little (2018) argue that, “if psychology is to be a mature quantitative science,  
112 then its primary theoretical aim should be to investigate systematic functional  
113 relationships as they are manifested at the individual participant level” (p. 2083). Note  
114 that because statistical power derives both from the number of participants and from the  
115 number of repeated measures per participant and condition, small- $N$  designs can have  
116 excellent power (Baker et al., 2021; Smith & Little, 2018).

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

---

<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* (Version 0.8.1.9000; Gabry, Češnovar, Johnson, & Brodner, 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),

118 are printed in bold. The content of the tutorials is mainly based on Allison (2010), Singer  
119 and Willett (2003), McElreath (2018), Kurz (2023a), and Kurz (2023b).

120 **Tutorial 1: Calculating descriptive statistics using a life table**

121 To illustrate how to quickly set up life tables for calculating the descriptive statistics  
122 (functions of discrete time), we use a published data set on masked response priming from  
123 Panis and Schmidt (2016), available on **ResearchGate**. In their first experiment, Panis  
124 and Schmidt (2016) presented a double arrow for 94 ms that pointed left or right as the  
125 target stimulus with an onset at time point zero in each trial. Participants had to indicate  
126 the direction in which the double arrow pointed using their corresponding index finger,  
127 within 800 ms after target onset. Response time and accuracy were recorded on each trial.  
128 Prime type (blank, congruent, incongruent) and mask type were manipulated. Here we  
129 focus on the subset of trials in which no mask was presented. The 13-ms prime stimulus  
130 was a double arrow with onset at -187 ms for the congruent (same direction as target) and  
131 incongruent (opposite direction as target) prime conditions.

132 After loading in the data file, one has to (a) supply required column names, and (b)  
133 specify the factor condition with the correct levels and labels. The required column names  
134 are as follows:

---

*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 &  
Fran ois, 2011), *readr* (Version 2.1.5; Wickham, Hester, & Bryan, 2024), *RJ-2021-048* (Bengtsson, 2021,  
2021, 2021, 2021, 2021, 2021, 2021, 2021, 2021, 2021, 2021, 2021, 2021, 2021, 2021, 2021, 2021,  
2021, 2021, 2021, 2021, 2021, 2021, 2021, 2021, 2021, 2021, 2021, 2021, 2021, 2021, 2021), *standist* (Version  
0.0.0.9000; Girard, 2024), *stringr* (Version 1.5.1; Wickham, 2023b), *tibble* (Version 3.2.1; M ller &  
Wickham, 2023), *tidybayes* (Version 3.0.6; Kay, 2023), *tidyr* (Version 1.3.1; Wickham, Vaughan, & Girlich,  
2024), *tidyverse* (Version 2.0.0; Wickham et al., 2019), and *tinylabels* (Version 0.2.4; Barth, 2023).

- 135 • “pid”, indicating unique participant IDs;
- 136 • “trial”, indicating each unique trial per participant;
- 137 • “condition”, a factor indicating the levels of the independent variable (1, 2, . . . ) and  
the corresponding labels;
- 138 • “rt”, indicating the response times in ms;
- 139 • “acc”, indicating the accuracies (1/0).

141 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")))
```

142 To set up the life tables and plots of the discrete-time functions  $h(t)$ ,  $S(t)$  and  $ca(t)$

143 using functional programming, one has to nest the data within participants using the  
144 `group_nest()` function, and supply a user-defined censoring time and bin width to our  
145 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
```

146 Note that the censoring time should be a multiple of the bin width (both in ms). The  
147 censoring time should be a time point after which no informative responses are expected  
148 anymore. In experiments that implement a response deadline in each trial the censoring  
149 time can equal that deadline time point. Trials with a RT larger than the censoring time,  
150 or trials in which no response is emitted during the data collection period, are treated as

151 right-censored observations in EHA. In other words, these trials are not discarded, because  
152 they contain the information that the event did not occur before the censoring time.  
153 Removing such trials before calculating the mean event time can introduce a sampling bias.  
154 The person-trial-bin oriented dataset has one row for each time bin of each trial that is at  
155 risk for event occurrence. The variable “event” in the person-trial-bin oriented data set  
156 indicates whether a response occurs (1) or not (0) for each bin. When creating the plots  
157 using our function `plot_eha()`, some warning messages will likely be generated, like these:

- 158 • Removed 2 rows containing missing values or values outside the scale range  
159     (`geom_line()`).  
160 • Removed 2 rows containing missing values or values outside the scale range  
161     (`geom_point()`).  
162 • Removed 2 rows containing missing values or values outside the scale range  
163     (`geom_segment()`).

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

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

173 Figure 2 displays the discrete-time hazard, survivor, and conditional accuracy  
174 functions for each prime condition for participant 6. By using discrete-time  $h(t)$  functions  
175 of event occurrence - in combination with  $ca(t)$  functions for two-choice tasks - one can

<sup>176</sup> provide an unbiased, time-varying, and probabilistic description of the latency and  
<sup>177</sup> accuracy of responses based on all trials of any data set.

<sup>178</sup> For example, for participant 6, the estimated hazard values in bin (240,280] are 0.03,  
<sup>179</sup> 0.17, and 0.20 for the blank, congruent, and incongruent prime conditions, respectively. In  
<sup>180</sup> other words, when the waiting time has increased until *240 ms* after target onset, then the  
<sup>181</sup> conditional probability of response occurrence in the next 40 ms is more than five times  
<sup>182</sup> larger for both prime-present conditions, compared to the blank prime condition.

<sup>183</sup> Furthermore, the estimated conditional accuracy values in bin (240,280] are 0.29, 1,  
<sup>184</sup> and 0 for the blank, congruent, and incongruent prime conditions, respectively. In other  
<sup>185</sup> words, if a response is emitted in bin (240,280], then the probability that it is correct is  
<sup>186</sup> estimated to be 0.29, 1, and 0 for the blank, congruent, and incongruent prime conditions,  
<sup>187</sup> respectively.

Table 1

*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

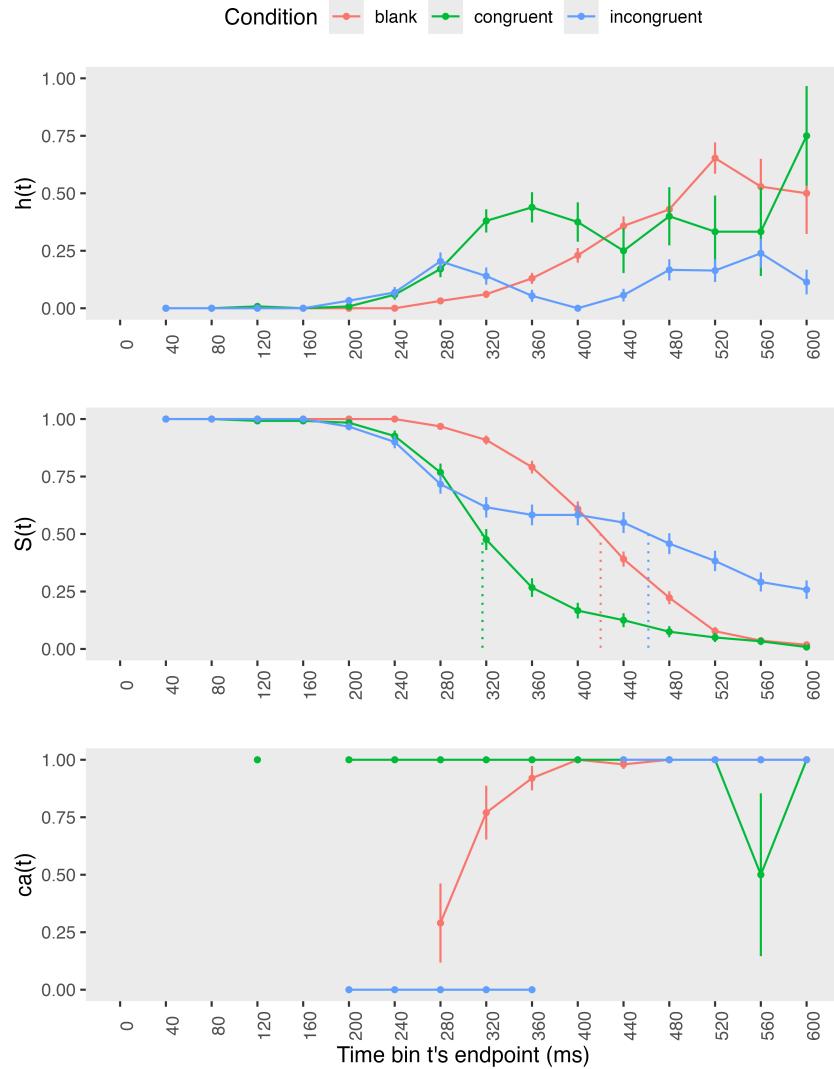


Figure 2. Estimated discrete-time hazard, survivor, and conditional accuracy functions for participant 6, as a function of the passage of discrete waiting time.

188 However, when the waiting time has increased until 400 ms after target onset, then  
 189 the conditional probability of response occurrence in the next 40 ms is estimated to be  
 190 0.36, 0.25, and 0.06 for the blank, congruent, and incongruent prime conditions,  
 191 respectively. And when a response does occur in bin (400,440], then the probability that it  
 192 is correct is estimated to be 0.98, 1, and 1 for the blank, congruent, and incongruent prime

193 conditions, respectively.

194 These results show that this participant is initially responding to the prime even  
195 though (s)he was instructed to only respond to the target, that response competition  
196 emerges in the incongruent prime condition around 300 ms, and that only later response  
197 are fully controlled by the target stimulus. Qualitatively similar results were obtained for  
198 the other five participants. Also, in their second Experiment, Panis and Schmidt (2016)  
199 showed that the negative compatibility effect in the mask-present conditions is time-locked  
200 to mask onset. This example shows that a simple difference between two means fails to  
201 reveal the dynamic behavior people display in many experimental paradigms (Panis, 2020;  
202 Panis, Moran, Wolkersdorfer, & Schmidt, 2020; Panis & Schmidt, 2022; Panis, Torfs,  
203 Gillebert, Wagemans, & Humphreys, 2017; Panis & Wagemans, 2009; Schmidt, Panis,  
204 Wolkersdorfer, & Vorberg, 2022). In other words, statistically controlling for the passage of  
205 time during data analysis is equally important as experimental control during the design of  
206 an experiment, to better understand human behavior in experimental paradigms. As we  
207 will show in Tutorials 2 and 3, statistical models for  $h(t)$  can be implemented as  
208 generalized linear mixed regression models predicting event occurrence (1/0) in each bin of  
209 a selected time range.

210 **Tutorial 2: Fitting Bayesian hazard models**

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

215 There are several analytic decisions one has to make. First, one has to select an  
216 analysis time window, i.e., a contiguous set of bins for which there is enough data for each  
217 participant. Second, given that the dependent variable is binary, one has to select a link

<sup>218</sup> function (see Figure 3).

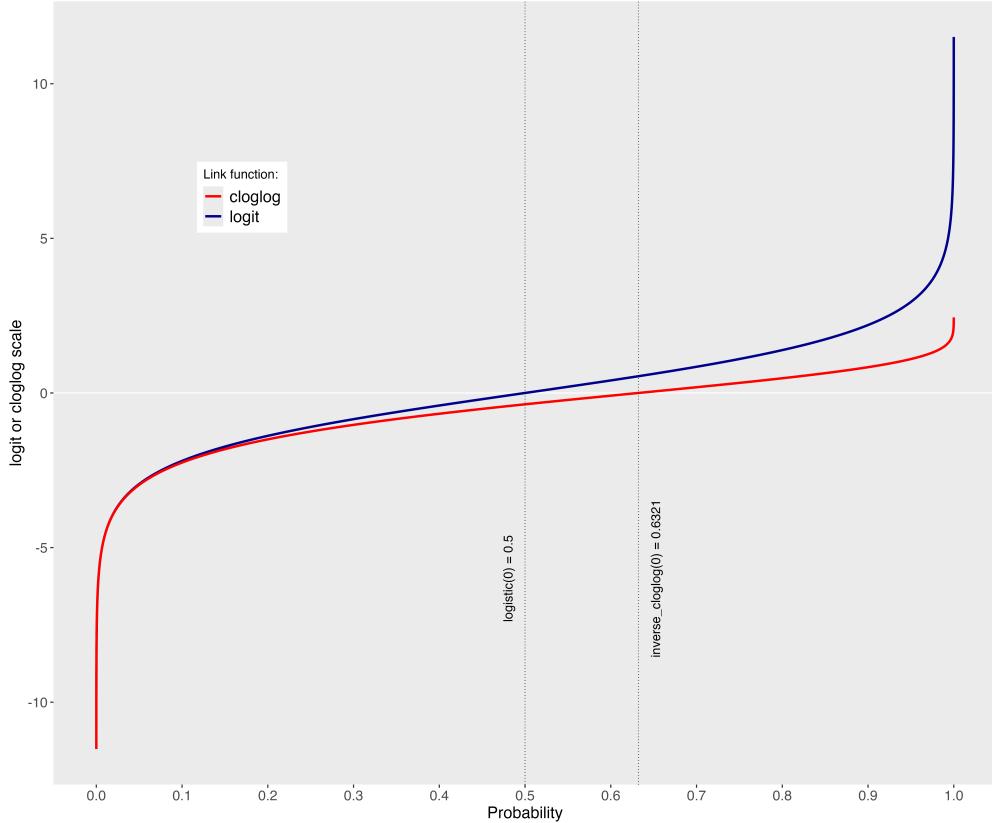


Figure 3. The logit and cloglog link functions.

<sup>219</sup> The symmetric logit link function transforms a (hazard) probability into the log of  
<sup>220</sup> the odds ratio. The asymmetric complementary log-log (cloglog) link function transforms  
<sup>221</sup> hazard into the logarithm of the negated logarithm of the probability of event  
<sup>222</sup> nonoccurrence. An important difference between these two link functions is that cloglog  
<sup>223</sup> provides a discrete-time hazard model that has a built-in proportional hazards assumption,  
<sup>224</sup> while logit provides a proportional odds assumption (see below). The cloglog link is  
<sup>225</sup> preferred over the logit link when events can occur in principle at any time point within a  
<sup>226</sup> bin, which is the case for RT data (Singer & Willett, 2003). Third, one has to choose a  
<sup>227</sup> specification of the effect of discrete TIME (i.e., the time bin index  $t$ ). One can choose a  
<sup>228</sup> general specification (one intercept per bin) or a functional specification, such as a

<sup>229</sup> polynomial one (compare model 1 with models 2, 3, and 4 below).

<sup>230</sup> An example (single-level) discrete-time hazard model with three predictors (TIME,  
<sup>231</sup> X1, X2) and the cloglog link function can be written as follows:

$$\begin{aligned} \text{cloglog}[h(t)] &= \ln(-\ln[1-h(t)]) = [\alpha_1 \text{ONE} + \alpha_2(\text{TIME}-9) + \alpha_3(\text{TIME}-9)^2] + [\beta_1 X_1 + \\ \text{cloglog}[h(t)] &= \beta_2 X_2 + \beta_3 X_2(\text{TIME}-9)]. \end{aligned}$$

<sup>234</sup> The main predictor variable TIME is the time bin index t that is centered on value 9  
<sup>235</sup> in this example. The first set of terms within brackets, the alpha parameters multiplied by  
<sup>236</sup> their polynomial specifications of (centered) time, represents the shape of the baseline  
<sup>237</sup> cloglog-hazard function (i.e., when all predictors  $X_i$  take on a value of zero). The second set  
<sup>238</sup> of terms (the beta parameters) represents the vertical shift in the baseline cloglog-hazard  
<sup>239</sup> for a 1 unit increase in the respective predictor variable. Predictors can be discrete,  
<sup>240</sup> continuous, and time-varying or time-invariant. For example, the effect of a 1 unit increase  
<sup>241</sup> in  $X_1$  is to vertically shift the whole baseline cloglog-hazard function by  $\beta_1$  cloglog-hazard  
<sup>242</sup> units. However, if the predictor interacts linearly with TIME (see  $X_2$  in the example), then  
<sup>243</sup> the effect of a 1 unit increase in  $X_2$  is to vertically shift the predicted cloglog-hazard in bin  
<sup>244</sup> 9 by  $\beta_2$  cloglog-hazard units (when  $\text{TIME}-9 = 0$ ), in bin 10 by  $\beta_2 + \beta_3$  cloglog-hazard  
<sup>245</sup> units (when  $\text{TIME}-9 = 1$ ), and so forth. To interpret the effects of a predictor, its  $\beta$   
<sup>246</sup> parameter is exponentiated, resulting in a hazard ratio (due to the use of the cloglog link).  
<sup>247</sup> When using the logit link, exponentiating a  $\beta$  parameter results in an odds ratio.

<sup>248</sup> In the case of a large- $N$  design without repeated measurements, the parameters of a  
<sup>249</sup> discrete-time hazard model can be estimated using standard logistic regression software  
<sup>250</sup> (after expanding the typical person-trial-oriented data set into a person-trial-bin-oriented  
<sup>251</sup> data set (see Tutorial 1); Allison (2010)). When there is clustering in the data, as in the  
<sup>252</sup> case of a small- $N$  design with repeated measurements, the parameters of a discrete-time  
<sup>253</sup> hazard model can be estimated using population-averaged methods (e.g., Generalized  
<sup>254</sup> Estimating Equations), and Bayesian or frequentist generalized linear mixed models

255 (Allison, 2010).

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

261 In this tutorial we will fit four Bayesian multilevel models (i.e., generalized linear  
 262 mixed models) to the data. We select the analysis range (200,600] and the cloglog link.  
 263 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 t=6 to t=15) of 40 ms
#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 discrete TIME (period) on bin 9, and trial on trial 1000
#ptb_data <- ptb_data %>% mutate(period_9 = period - 9,
#                                trial_c = (trial - 1000)/1000)
# remove unnecessary columns before fitting a model
#ptb_data <- ptb_data %>% select(-c(bl,tr,trial,period)) # 12840 obs. of 5 variables
```

264 **Model 1: A general specification of TIME, and main effects of congruency and  
 265 trial number**

266 For the first model, we use a general specification of TIME (i.e., one intercept per  
 267 time bin) for the baseline condition (blank prime), and assume that the effects of  
 268 prime-target congruency and trial number are proportional and additive, and that the  
 269 effect of trial number is linear.

270 **Model 2: A polynomial specification of TIME, and main effects of congruency**  
271 **and trial number**

272 , and use a polynomial to specify the effect of TIME in the “blank” prime condition.

273 **Model 3: A polynomial specification of TIME, and relaxing the proportionality**  
274 **assumption**

275 **Model 4: A polynomial specification of TIME, and relaxing all three**  
276 **assumptions**

277 Based on previous work (Panis, 2020; Panis, Moran, et al., 2020; Panis & Schmidt,  
278 2022; Panis et al., 2017; Panis & Wagemans, 2009) and because cognition is likely the  
279 behavior of a non-linear dynamical system [ref], we relax all three assumptions in model 4.

280 **Compare the models.**

281 Use WAIC to compare models (ref McElreath, Kurz)

282 **Plot congruency effects and subject-specific fits for the selected model**

283 Plot the effects of congruent and incongruent for each time bin for the selected model.

284 Plot the model-based hazard and survivor functions.

285 **Tutorial 3: Fitting Frequentist hazard models**

286 In this third tutorial we illustrate how to fit a multilevel hazard regression model in  
287 the frequentist framework, for the data set used in the first tutorial.

## Tutorial 4: Calculating descriptive statistics when there are two independent variables

In this final tutorial we illustrate how to calculate and plot the descriptive statistics for the full data set of Experiment 1 of Panis and Schmidt (2016), which includes two dependent variables: mask type and prime type.

## Discussion

We noticed that many researchers are still reluctant to abandon analysis-of-variance switch to event history analysis when analyzing time-to-event data. By providing this material, we hope that researchers will start using hazard analysis more often, due to the many advantages described below. While we focus here on within-subject, factorial,  $N$ - $N$  designs, it is important to realize that event history analysis can be applied to  $N$ - $1$  designs as well (large- $N$  designs with only one measurement per subject, between-subject designs, etc.).

### **Advantages of hazard functions**

Statisticians and mathematical psychologists recommend focusing on the hazard function when analyzing time-to-event data for various reasons. First, as discussed by Den, Van Orden, and Turvey (2009), “probability density functions can appear nearly identical, both statistically and to the naked eye, and yet are clearly different on the basis of their hazard functions (but not vice versa). Hazard functions are thus more diagnostic than density functions” (p. 331; see also Figure 1 in . . .) when one is interested in examining the detailed shape of a RT distribution.

Second, because RT distributions may differ from one another in multiple ways, *Asend* (1990) developed a dominance hierarchy of statistical differences between two arbitrary distributions A and B. For example, if  $F_A(t) > F_B(t)$  for all  $t$ , then both

312 cumulative distribution functions are said to show a complete ordering. Townsend (1990)  
313 showed that a complete ordering on the hazard functions — $\lambda_A(t) > \lambda_B(t)$  for all  $t$ —  
314 implies a complete ordering on both the cumulative distribution and survivor functions  
315 — $F_A(t) > F_B(t)$  and  $S_A(t) < S_B(t)$ — which in turn implies an ordering on the mean  
316 latencies —mean A < mean B. In contrast, an ordering on two means does *not* imply a  
317 complete ordering on the corresponding  $F(t)$  and  $S(t)$  functions, and a complete ordering  
318 on these latter functions does *not* imply a complete ordering on the corresponding hazard  
319 functions. This means that stronger conclusions can be drawn from data when comparing  
320 the hazard functions using EHA. For example, when mean A < mean B, the hazard  
321 functions might show a complete ordering (i.e., for all  $t$ ), a partial ordering (e.g., only for  $t$   
322  $> 300$  ms, or only for  $t < 500$  ms), or they may cross each other one or more times. Thus,  
323 because the discrete-time hazard function identifies unique information about event  
324 occurrence in each bin, while  $F(t)$  accumulates the complements of the hazard estimates  
325 from the current and previous bins — $S(t) = P(RT > t) =$   
326  $[1-h(t)].[1-h(t-1)].[1-h(t-2)]\dots[1-h(1)]$ —, instead of using delta-plots —differences in  
327 quantiles from  $F(t)^{-1}$ — one can simply plot delta- $h(t)$  functions (see ref).

328 Third, EHA does not discard right-censored observations when estimating hazard  
329 functions, that is, trials for which we do not observe a response during the data collection  
330 period in a trial so that we only know that the RT must be larger than some value (i.e., the  
331 response deadline). This is important because although a few right-censored observations  
332 are inevitable in most RT tasks, a lot of right-censored observations are expected in  
333 experiments on masking, the attentional blink, and so forth. In other words, by using EHA  
334 you can analyze RT data from experiments that typically do not measure response times.  
335 As a result, EHA can also deal with long RTs in experiments without a response deadline,  
336 which are typically treated as outliers and are discarded before calculating a mean. This  
337 orthodox procedure can lead to a sampling bias, however, which results in underestimation  
338 of the mean. By introducing a fixed censoring time for all trials at the end of the analysis

339 time window, trials with long RTs are not discarded but contribute to the risk set of each  
340 bin.

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

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

### 353 Individual differences

354 One important issue is that of possible individual differences in the overall location of  
355 the distribution, and the time course of psychological effects. For example, when you wait  
356 for a response of the participant on each trial, you allow the participant to have control  
357 over the trial duration, and some participants might respond only when they are confident  
358 that their emitted response will be correct. These issues can be avoided by introducing a  
359 (relatively short) response deadline in each trial, e.g., 600 ms for simple detection tasks,  
360 1000 ms for more difficult discrimination tasks, or 2 s for tasks requiring extended  
361 high-level processing. Because EHA can deal in a straightforward fashion with  
362 right-censored observations (i.e., trials without an observed response), introducing a  
363 response deadline is recommended when designing RT experiments. Furthermore,

364 introducing a response deadline and asking participants to respond before the deadline as  
365 much as possible, will also lead to individual distributions that overlap in time, which is  
366 important when selecting a common analysis time window when fitting hazard models.  
367 But even when using a response deadline, participants can differ qualitatively in the effects  
368 they display (refs). One way to deal with this is to run a clustering algorithm on the  
369 individual hazard estimates across all conditions. The obtained dendrogram can then be  
370 used to identify (hopefully big) clusters of participants that behave similarly, and to  
371 identify (hopefully small) clusters of participants with outlying behavioral patterns. One  
372 might then exclude the outlying participants before fitting a hazard model.

### 373 **Limitations**

374 One important limitation is that you need a relatively large number of trials per  
375 condition to estimate the hazard function with high temporal resolution. There is a  
376 tradeoff between the number of trials per condition and the temporal resolution (i.e., bin  
377 width) of the hazard function. We therefore recommend to design as many trials as  
378 possible per experimental condition given the available resources. Another limitation of  
379 multilevel hazard modeling is that it might take some time to fit the model.

### 380 **Extensions**

381 The hazard models in this tutorial assume that there is one event of interest.  
382 However, in certain situations, more than one event of interest might exist. For example,  
383 ... - multi-state hazard models refs

### 384 **Conclusions**

385 see paper...

386

## References

- 387 Allison, P. D. (1982). Discrete-Time Methods for the Analysis of Event Histories.  
388     *Sociological Methodology*, 13, 61. <https://doi.org/10.2307/270718>
- 389 Allison, P. D. (2010). *Survival analysis using SAS: A practical guide* (2. ed). Cary, NC:  
390     SAS Press.
- 391 Aust, F. (2019). *Citr: 'RStudio' add-in to insert markdown citations*. Retrieved from  
392     <https://github.com/crsh/citr>
- 393 Aust, F., & Barth, M. (2023). *papaja: Prepare reproducible APA journal articles with R*  
394     *Markdown*. Retrieved from <https://github.com/crsh/papaja>
- 395 Baker, D. H., Vilidaite, G., Lygo, F. A., Smith, A. K., Flack, T. R., Gouws, A. D., &  
396     Andrews, T. J. (2021). Power contours: Optimising sample size and precision in  
397     experimental psychology and human neuroscience. *Psychological Methods*, 26(3),  
398     295–314. <https://doi.org/10.1037/met0000337>
- 399 Barth, M. (2023). *tinylabes: Lightweight variable labels*. Retrieved from  
400     <https://cran.r-project.org/package=tinylabes>
- 401 Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects  
402     models using lme4. *Journal of Statistical Software*, 67(1), 1–48.  
403     <https://doi.org/10.18637/jss.v067.i01>
- 404 Bates, D., Maechler, M., & Jagan, M. (2024). *Matrix: Sparse and dense matrix classes and*  
405     *methods*. Retrieved from <https://CRAN.R-project.org/package=Matrix>
- 406 Bengtsson, H. (2021). A unifying framework for parallel and distributed processing in r  
407     using futures. *The R Journal*, 13(2), 208–227. <https://doi.org/10.32614/RJ-2021-048>
- 408 Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan.  
409     *Journal of Statistical Software*, 80(1), 1–28. <https://doi.org/10.18637/jss.v080.i01>
- 410 Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms.  
411     *The R Journal*, 10(1), 395–411. <https://doi.org/10.32614/RJ-2018-017>
- 412 Bürkner, P.-C. (2021). Bayesian item response modeling in R with brms and Stan. *Journal*

- 413       *of Statistical Software*, 100(5), 1–54. <https://doi.org/10.18637/jss.v100.i05>
- 414   Eddelbuettel, D., & Balamuta, J. J. (2018). Extending R with C++: A Brief Introduction  
415   to Rcpp. *The American Statistician*, 72(1), 28–36.  
416           <https://doi.org/10.1080/00031305.2017.1375990>
- 417   Eddelbuettel, D., & François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal  
418   of Statistical Software*, 40(8), 1–18. <https://doi.org/10.18637/jss.v040.i08>
- 419   Gabry, J., Češnovar, R., Johnson, A., & Brodner, S. (2024). *Cmdstanr: R interface to  
420   'CmdStan'*. Retrieved from <https://github.com/stan-dev/cmdstanr>
- 421   Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2019). Visualization  
422   in bayesian workflow. *J. R. Stat. Soc. A*, 182, 389–402.  
423           <https://doi.org/10.1111/rssa.12378>
- 424   Girard, J. (2024). *Standist: What the package does (one line, title case)*. Retrieved from  
425           <https://github.com/jmgirard/standist>
- 426   Grolemund, G., & Wickham, H. (2011). Dates and times made easy with lubridate.  
427           *Journal of Statistical Software*, 40(3), 1–25. Retrieved from  
428           <https://www.jstatsoft.org/v40/i03/>
- 429   Holden, J. G., Van Orden, G. C., & Turvey, M. T. (2009). Dispersion of response times  
430   reveals cognitive dynamics. *Psychological Review*, 116(2), 318–342.  
431           <https://doi.org/10.1037/a0014849>
- 432   Kantowitz, B. H., & Pachella, R. G. (2021). The Interpretation of Reaction Time in  
433   Information-Processing Research 1. *Human Information Processing*, 41–82.  
434           <https://doi.org/10.4324/9781003176688-2>
- 435   Kay, M. (2023). *tidybayes: Tidy data and geoms for Bayesian models*.  
436           <https://doi.org/10.5281/zenodo.1308151>
- 437   Kelso, J. A. S., Dumas, G., & Tognoli, E. (2013). Outline of a general theory of behavior  
438   and brain coordination. *Neural Networks: The Official Journal of the International  
439   Neural Network Society*, 37, 120–131. <https://doi.org/10.1016/j.neunet.2012.09.003>

- 440 Kurz, A. S. (2023a). *Applied longitudinal data analysis in brms and the tidyverse* (version  
441 0.0.3). Retrieved from <https://bookdown.org/content/4253/>
- 442 Kurz, A. S. (2023b). *Statistical rethinking with brms, ggplot2, and the tidyverse: Second*  
443 *edition* (version 0.4.0). Retrieved from <https://bookdown.org/content/4857/>
- 444 Luce, R. D. (1991). *Response times: Their role in inferring elementary mental organization*  
445 (1. issued as paperback). Oxford: Univ. Press.
- 446 McElreath, R. (2018). *Statistical Rethinking: A Bayesian Course with Examples in R and*  
447 *Stan* (1st ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9781315372495>
- 448 Meyer, D. E., Osman, A. M., Irwin, D. E., & Yantis, S. (1988). Modern mental  
449 chronometry. *Biological Psychology*, 26(1-3), 3–67.  
450 [https://doi.org/10.1016/0301-0511\(88\)90013-0](https://doi.org/10.1016/0301-0511(88)90013-0)
- 451 Müller, K., & Wickham, H. (2023). *Tibble: Simple data frames*. Retrieved from  
452 <https://CRAN.R-project.org/package=tibble>
- 453 Neuwirth, E. (2022). *RColorBrewer: ColorBrewer palettes*. Retrieved from  
454 <https://CRAN.R-project.org/package=RColorBrewer>
- 455 Panis, S. (2020). How can we learn what attention is? Response gating via multiple direct  
456 routes kept in check by inhibitory control processes. *Open Psychology*, 2(1), 238–279.  
457 <https://doi.org/10.1515/psych-2020-0107>
- 458 Panis, S., Moran, R., Wolkersdorfer, M. P., & Schmidt, T. (2020). Studying the dynamics  
459 of visual search behavior using RT hazard and micro-level speed–accuracy tradeoff  
460 functions: A role for recurrent object recognition and cognitive control processes.  
461 *Attention, Perception, & Psychophysics*, 82(2), 689–714.  
462 <https://doi.org/10.3758/s13414-019-01897-z>
- 463 Panis, S., Schmidt, F., Wolkersdorfer, M. P., & Schmidt, T. (2020). Analyzing Response  
464 Times and Other Types of Time-to-Event Data Using Event History Analysis: A Tool  
465 for Mental Chronometry and Cognitive Psychophysiology. *I-Perception*, 11(6),  
466 2041669520978673. <https://doi.org/10.1177/2041669520978673>

- 467 Panis, S., & Schmidt, T. (2016). What Is Shaping RT and Accuracy Distributions? Active  
468 and Selective Response Inhibition Causes the Negative Compatibility Effect. *Journal of*  
469 *Cognitive Neuroscience*, 28(11), 1651–1671. [https://doi.org/10.1162/jocn\\_a\\_00998](https://doi.org/10.1162/jocn_a_00998)
- 470 Panis, S., & Schmidt, T. (2022). When does “inhibition of return” occur in spatial cueing  
471 tasks? Temporally disentangling multiple cue-triggered effects using response history  
472 and conditional accuracy analyses. *Open Psychology*, 4(1), 84–114.  
473 <https://doi.org/10.1515/psych-2022-0005>
- 474 Panis, S., Torfs, K., Gillebert, C. R., Wagemans, J., & Humphreys, G. W. (2017).  
475 Neuropsychological evidence for the temporal dynamics of category-specific naming.  
476 *Visual Cognition*, 25(1-3), 79–99. <https://doi.org/10.1080/13506285.2017.1330790>
- 477 Panis, S., & Wagemans, J. (2009). Time-course contingencies in perceptual organization  
478 and identification of fragmented object outlines. *Journal of Experimental Psychology:*  
479 *Human Perception and Performance*, 35(3), 661–687.  
480 <https://doi.org/10.1037/a0013547>
- 481 Pedersen, T. L. (2024). *Patchwork: The composer of plots*. Retrieved from  
482 <https://CRAN.R-project.org/package=patchwork>
- 483 Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in s and s-PLUS*. New York:  
484 Springer. <https://doi.org/10.1007/b98882>
- 485 R Core Team. (2024). *R: A language and environment for statistical computing*. Vienna,  
486 Austria: R Foundation for Statistical Computing. Retrieved from  
487 <https://www.R-project.org/>
- 488 Schmidt, T., Panis, S., Wolkersdorfer, M. P., & Vorberg, D. (2022). Response inhibition in  
489 the Negative Compatibility Effect in the absence of inhibitory stimulus features. *Open*  
490 *Psychology*, 4(1), 219–230. <https://doi.org/10.1515/psych-2022-0012>
- 491 Singer, J. D., & Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling*  
492 *Change and Event Occurrence*. Oxford, New York: Oxford University Press.
- 493 Smith, P. L., & Little, D. R. (2018). Small is beautiful: In defense of the small-N design.

- 494        *Psychonomic Bulletin & Review*, 25(6), 2083–2101.
- 495        <https://doi.org/10.3758/s13423-018-1451-8>
- 496        Townsend, J. T. (1990). Truth and consequences of ordinal differences in statistical  
497        distributions: Toward a theory of hierarchical inference. *Psychological Bulletin*, 108(3),  
498        551–567. <https://doi.org/10.1037/0033-2909.108.3.551>
- 499        Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics.  
500        *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)
- 501        Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New  
502        York. Retrieved from <https://ggplot2.tidyverse.org>
- 503        Wickham, H. (2023a). *Forcats: Tools for working with categorical variables (factors)*.  
504        Retrieved from <https://forcats.tidyverse.org/>
- 505        Wickham, H. (2023b). *Stringr: Simple, consistent wrappers for common string operations*.  
506        Retrieved from <https://stringr.tidyverse.org>
- 507        Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., ...  
508        Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43),  
509        1686. <https://doi.org/10.21105/joss.01686>
- 510        Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *Dplyr: A  
511        grammar of data manipulation*. Retrieved from <https://dplyr.tidyverse.org>
- 512        Wickham, H., & Henry, L. (2023). *Purrr: Functional programming tools*. Retrieved from  
513        <https://purrr.tidyverse.org/>
- 514        Wickham, H., Hester, J., & Bryan, J. (2024). *Readr: Read rectangular text data*. Retrieved  
515        from <https://readr.tidyverse.org>
- 516        Wickham, H., Vaughan, D., & Girlich, M. (2024). *Tidyr: Tidy messy data*. Retrieved from  
517        <https://tidyr.tidyverse.org>