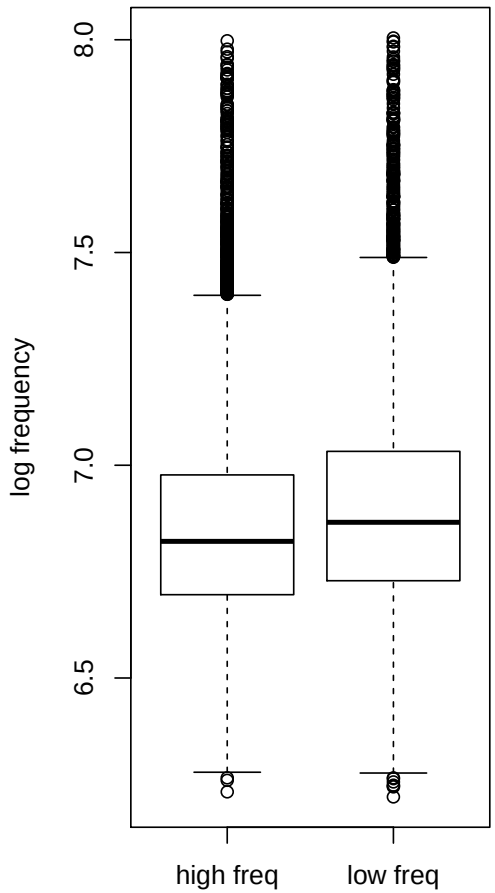
Survival analysis:

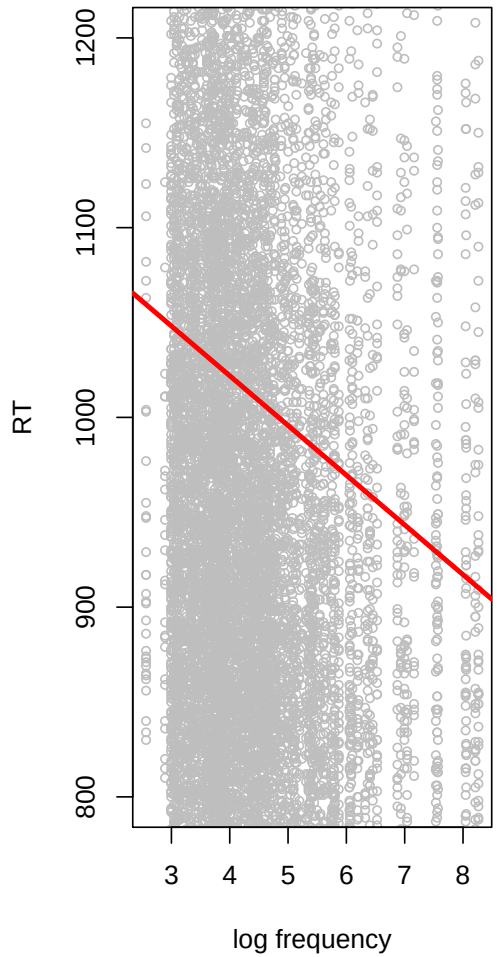
a tool for timing semantic and formal effects on derived and compound word recognition

Victor Kuperman, Daniel Schmidtke, Kazunaga Matsuki

MOPROC: June 23rd, 2017

Typical analyses of the mean RTs



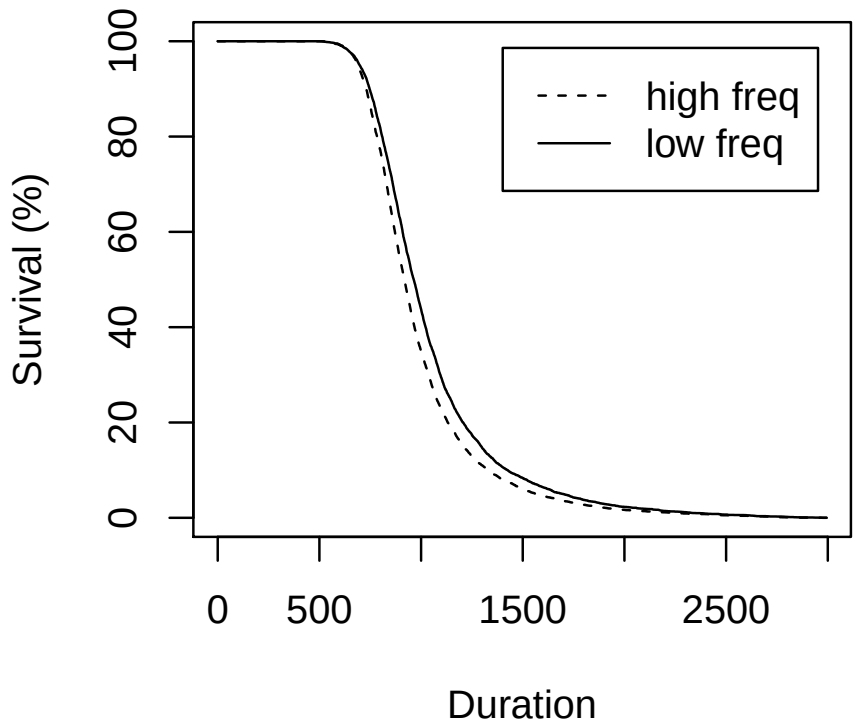


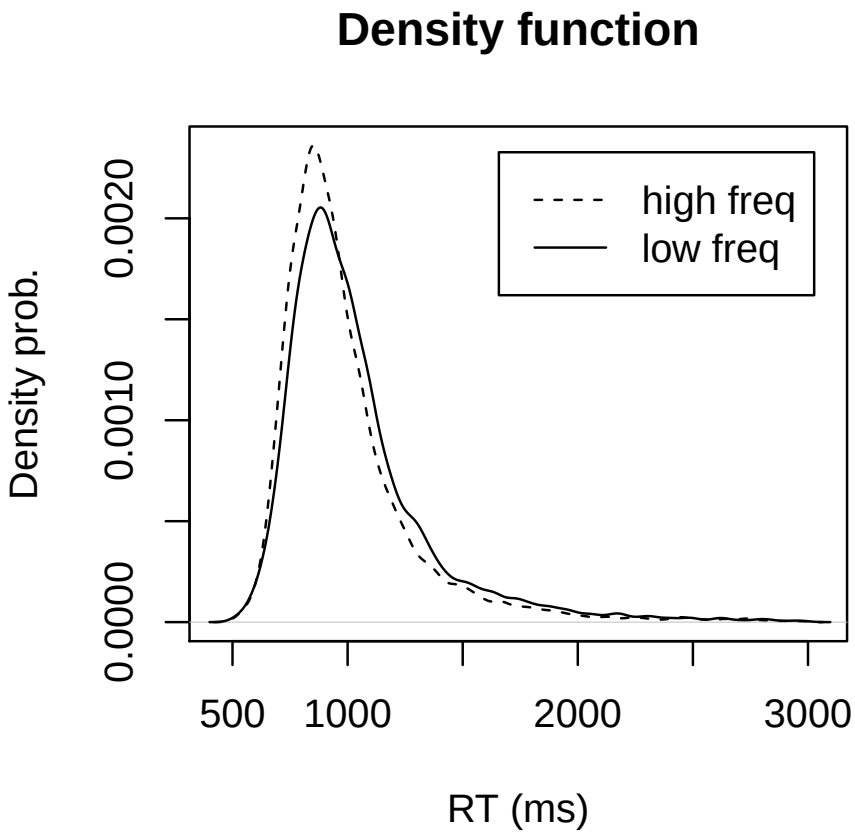
Distributional analyses

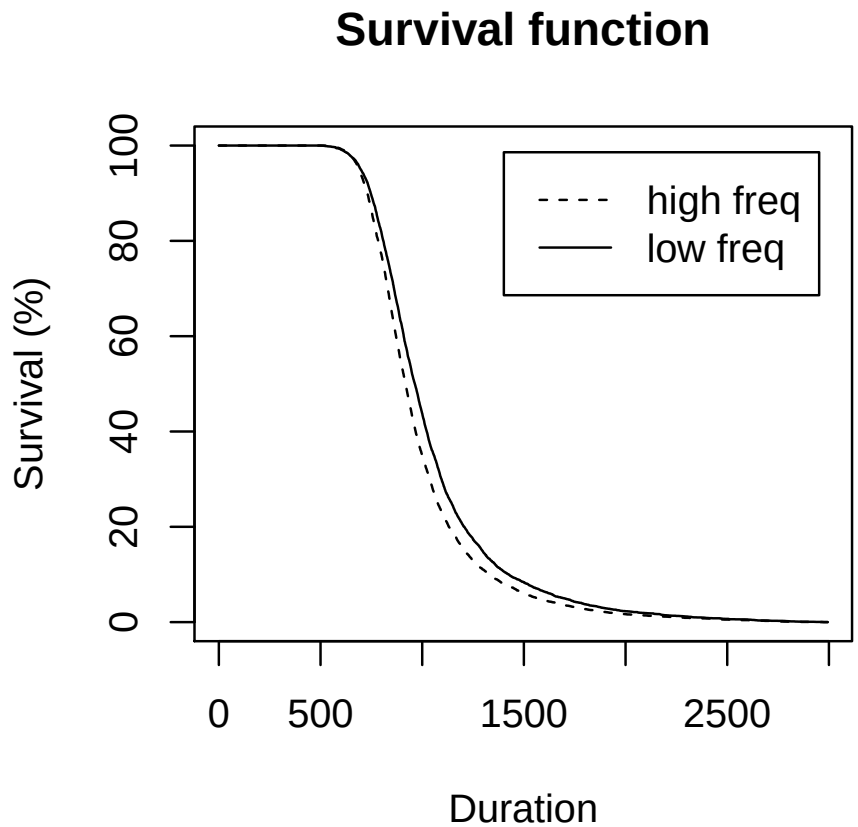
- How do distributions of RTs change across the timecourse?
- How early do RT distributions diverge across conditions?
- Vincentile analysis: Vincent (1910), Heathcote (1995), Balota and Abrams (1995), etc.
- Survival analysis: Rueckl and Galantucci (2005); Reingold et al. (2012); Reingold and Sheridan (2014; Frontiers)

Survival analysis

Beginning of split between two distributions indicates the onset of an effect.




Density function
Density prob.
0.0000
0.0010
0.0020
high freq
low freq
500
1000
2000
3000
RT (ms)

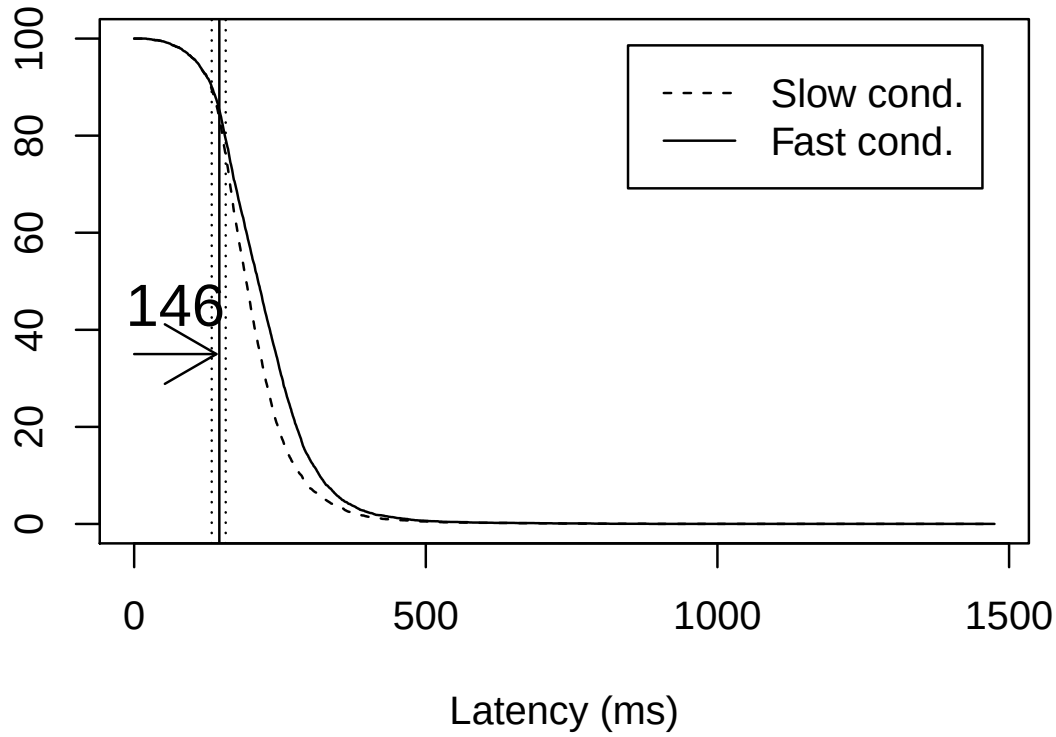

Survival function
Survival (%)
0
20
40
60
80
100
high freq
low freq
0
500
1500
2500
Duration

Reference

- Present analyses reported in Schmidtke, Matsuki, & Kuperman (2017). Surviving blind decomposition: A distributional analysis of the time-course of complex word recognition. *JEP:LMC*.
- A minimal demo of survival analysis in **R** user the **RTsurvival package** (Matsuki, 2016). The original MATLAB routines are found in Reingold & Sheridan (2014).

Confidence interval divergence point analysis (CI DPA procedure)

- The goal is to establish at what point in time the two survival curves diverge.
- Divergence Point Estimate (DPE).


Slow cond.
Fast cond.
146
0
20
40
60
80
100
0
500
1000
1500
Latency (ms)

Confidence interval divergence point analysis (CI DPA procedure)

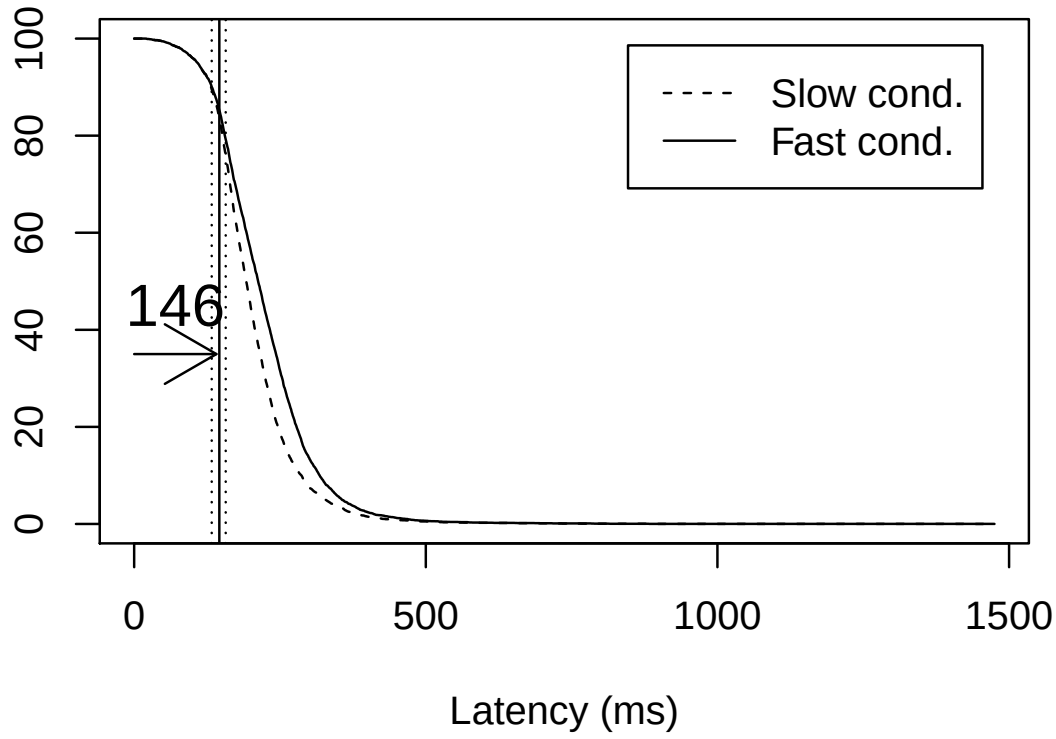
- Divergence point: **first 1-ms bin in a run of five consecutive bins** of a substantial difference in survival percentages.
- Reingold and Sheridan (2014) recommend **1.5%**: good even for low-powered designs.
- Schmidtke et al. (in press) chose **3%** for greater reliability.

CI DPA continued

- For 1000 iterations, the CI DPA procedure:
  1. randomly draws data for an individual participant (bootstrapping with replacement)
  2. generates individual survival curves per condition for each participant
  3. computes average survival % across participants for each condition and each bin.
  4. estimates the divergence point (as defined above)

CI DPA continued

- A distribution of 1000 divergence points is generated.
- We take the median as the divergence point
- Also, the 95% confidence interval is available.


Slow cond.
Fast cond.
146
0
20
40
60
80
100
0
500
1000
1500
Latency (ms)

Output

```
# Method of Estimation:
# [1] "CI"
# Divergence Point Estimate:
# [1] 146
# Confidence Interval:
#     2.5%   97.5%
#  133.975 157.000
```

Application to morphology

- Obligatory meaning-blind decomposition versus form-and-meaning accounts.
- Relative time-course of formal versus semantic effects: early vs late, or roughly simultaneous.
- Relative timing of the surface frequency effect: very late (Solomyak & Marantz, 2010) or not.
- Absolute timeframe of processing: neurophysiological estimates (post-300 ms) against behavioral data.

Procedure:

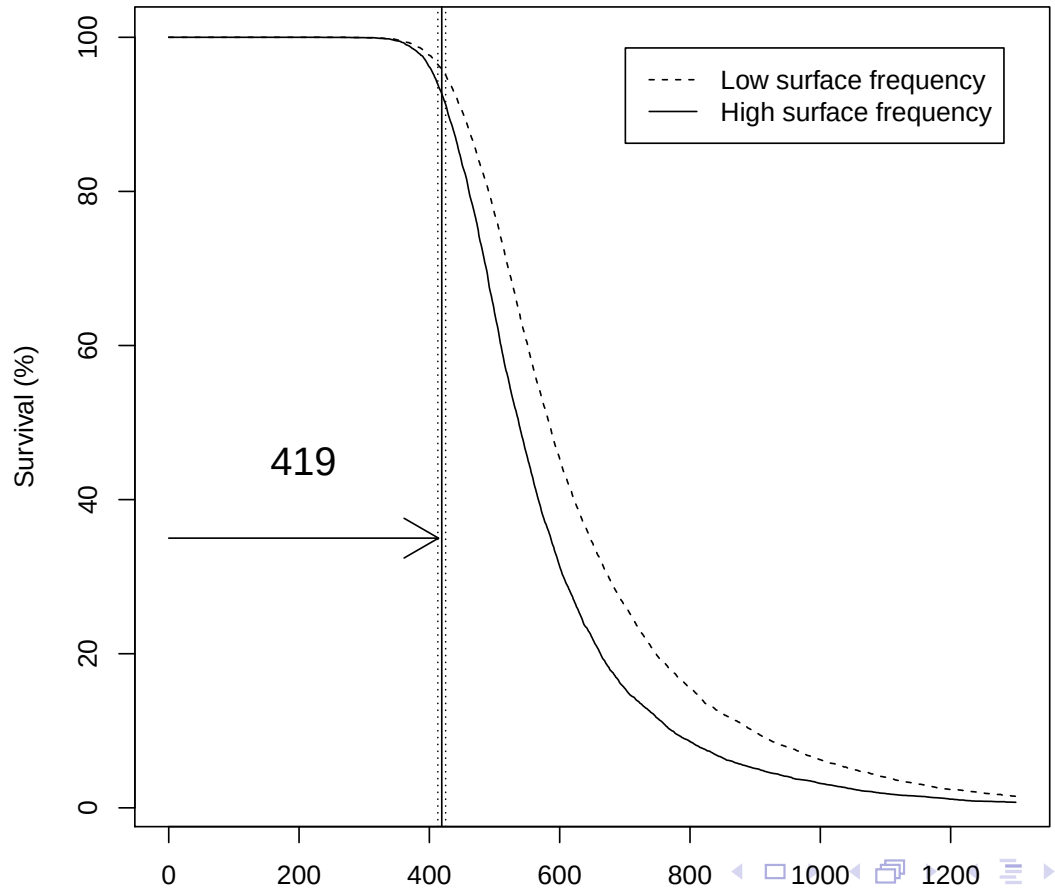
- Visual processing only; no priming here; but same method applies to any RT data.
- Identified two behavioural response types (lexical decision latencies and eye-movements) for
- LD: derived words in English and Dutch (Studies 1 and 2; BLP and DLP); pseudo-derived words and orthographic controls in English (Studies 3 and 4, BLP).

Procedure:

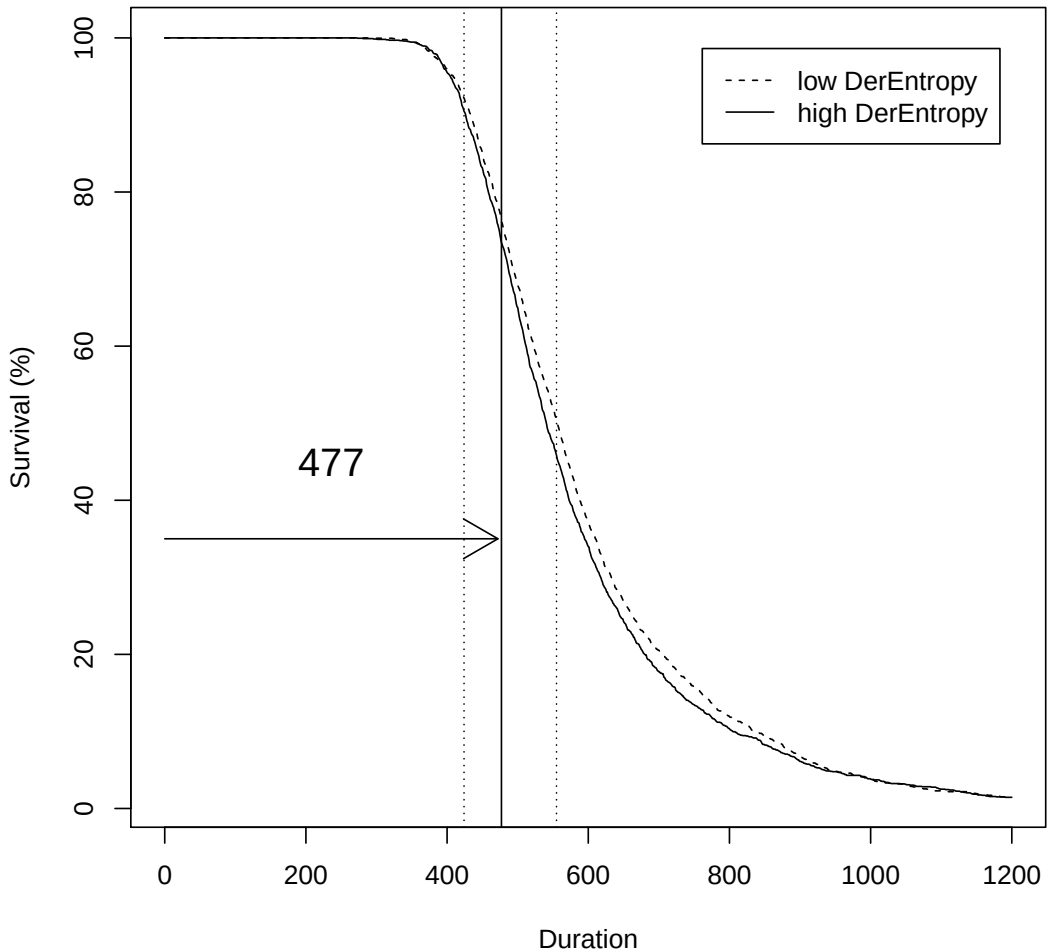
- Lexical variables proposed as diagnostic of orthography, semantics, and morphology.
- Continuous variables: split at the median to form conditions.
- Survival analysis: calculate divergence points for each lexical variable that is diagnostic of a recognition stage.

Divergence point estimate for surface frequency

English derived word recognition: British Lexicon Project (Keuleers et al., 2012)

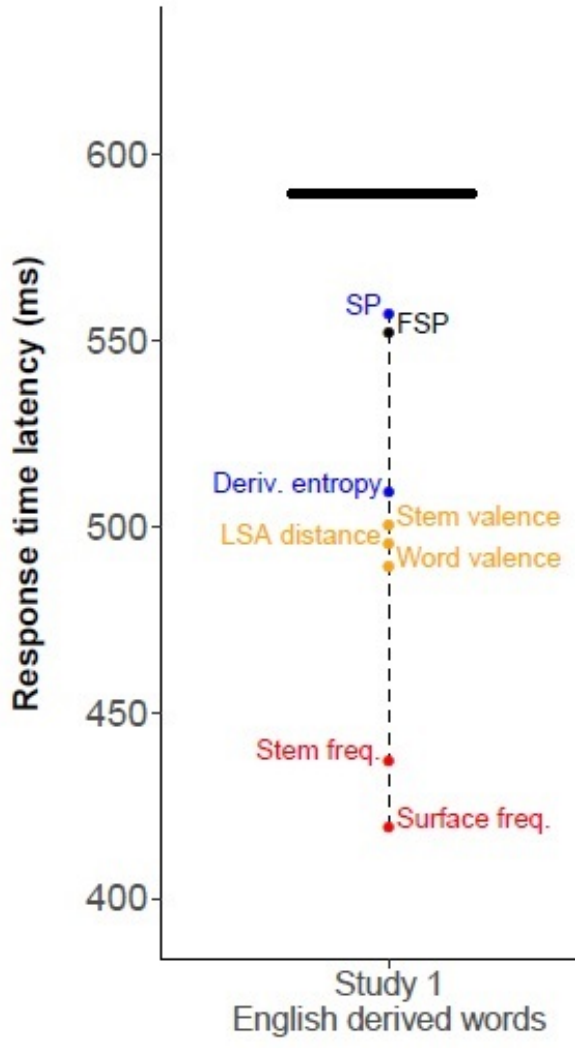


Divergence point estimate for derivational entropy



Divergence point estimate for all predictors in Study 1

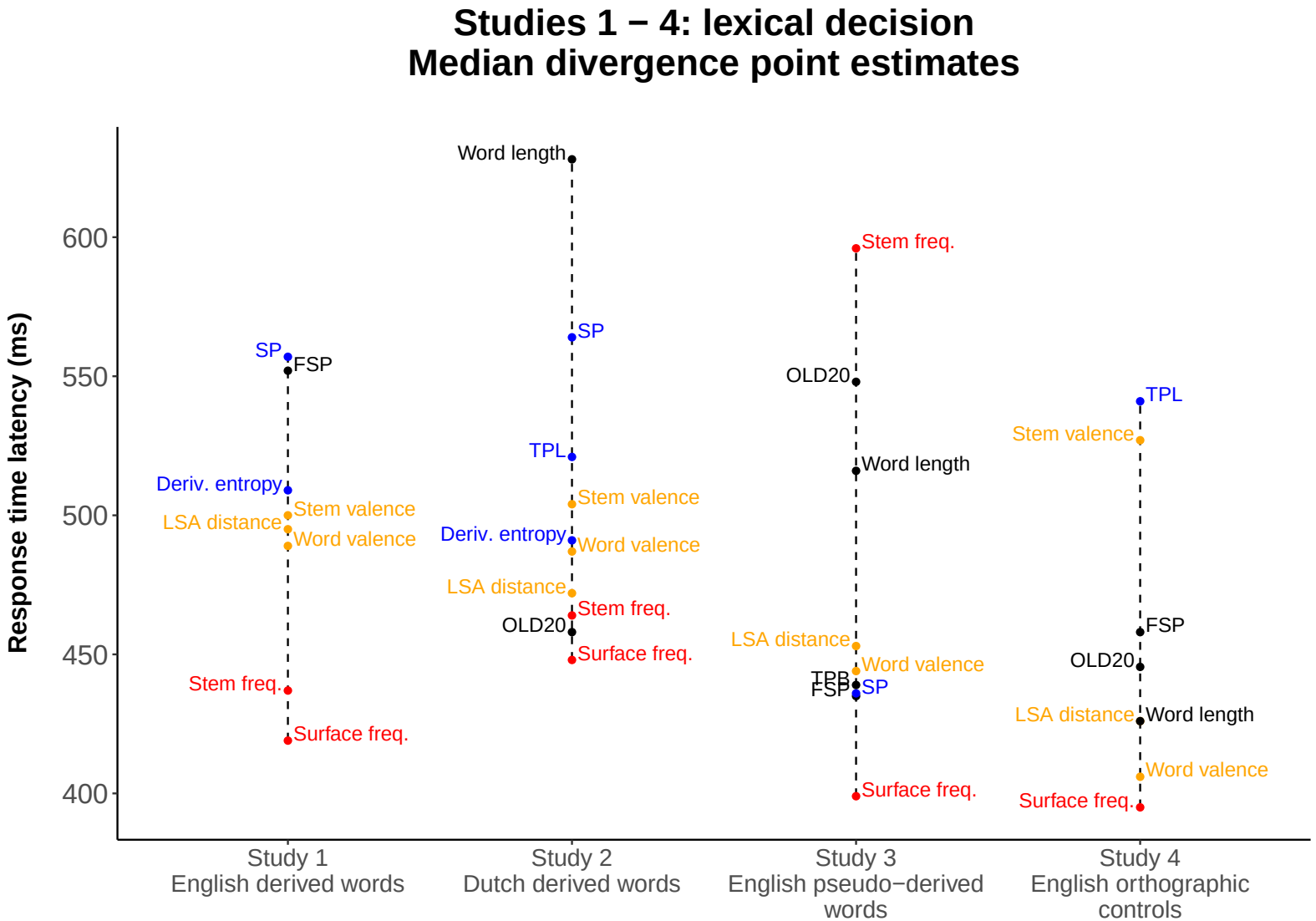
English derived word recognition, study 1:



Controls and checks

- We applied a conservative estimate of divergence point
- No correlation between effect size and relative order of divergence points
- No evidence that the divergence point for variable A was reflective of changes of an underlying variable B

Results: lexical decision timecourse



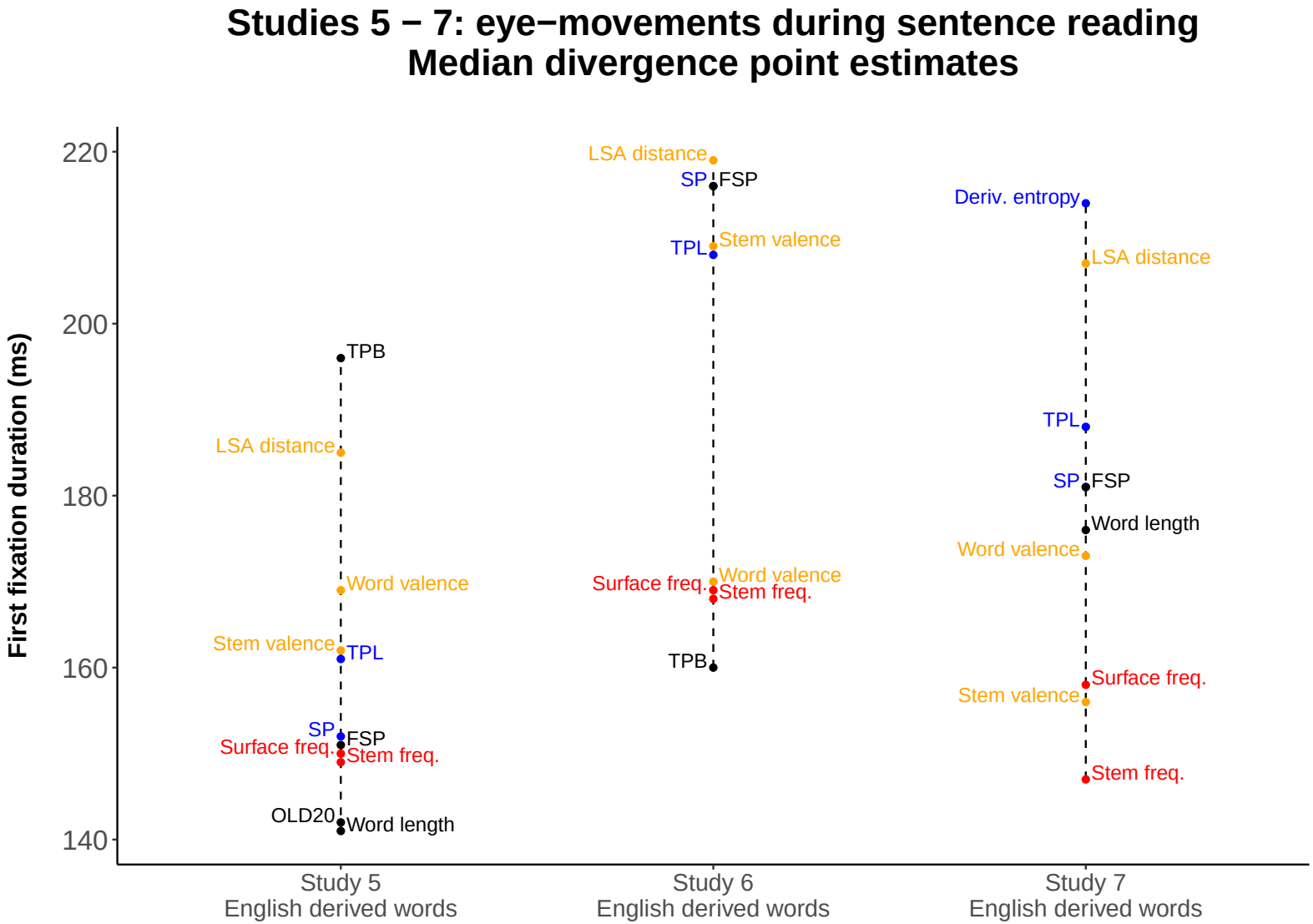
Summary: Relative order

- Surface frequency first
- Onset of semantics as early as or earlier than morphology
- No major difference across derived, pseudo- and control words.

Procedure: Eye movements

- EM: first fixations on derived words in English sentences (3 studies).

Results: Eye-movement timecourse



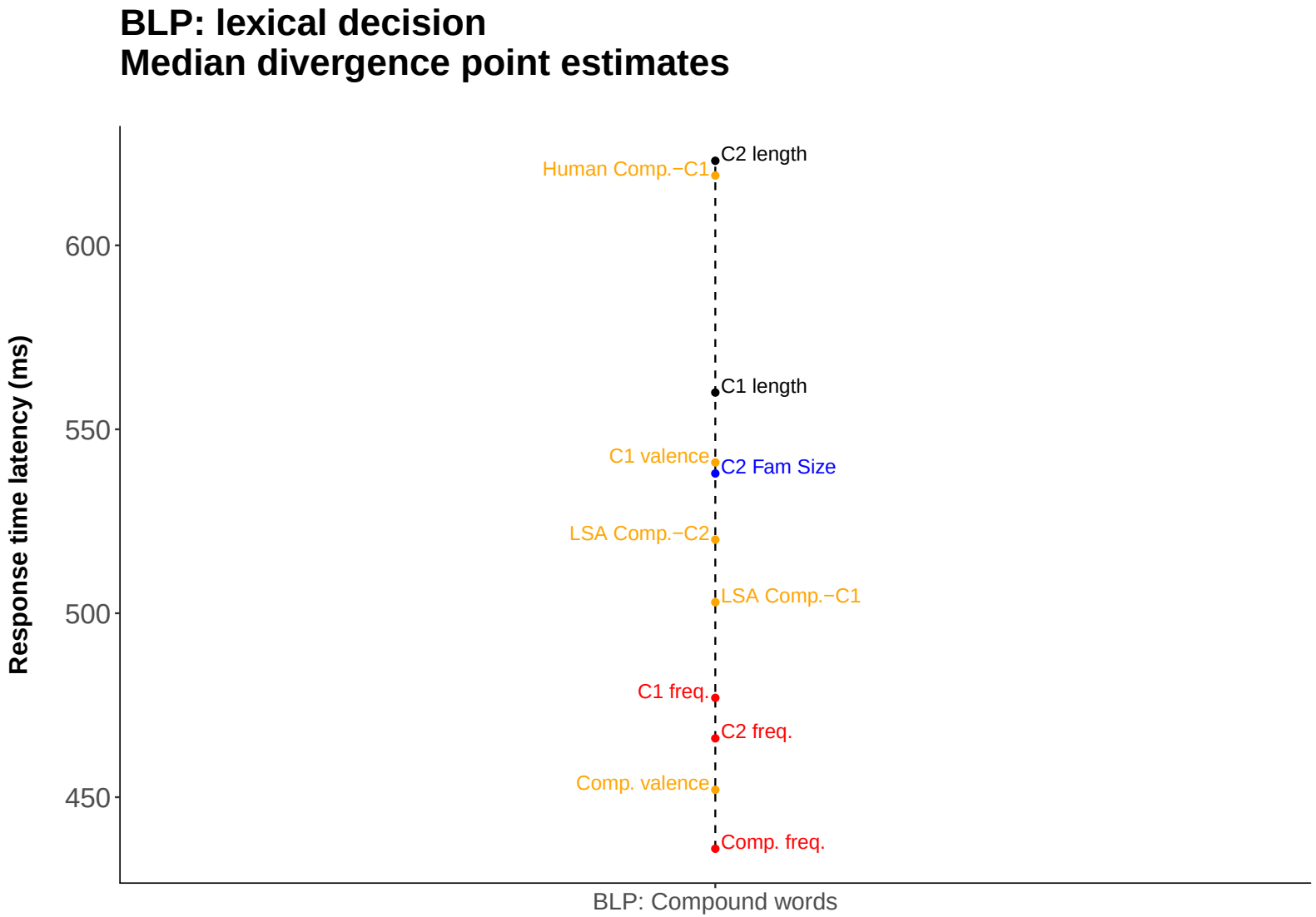
Summary: Eye movements

- 140 - 220 ms onset of effects.
- Surface and stem frequency simulataneous.
- Onset of semantics as early as or earlier than morphology

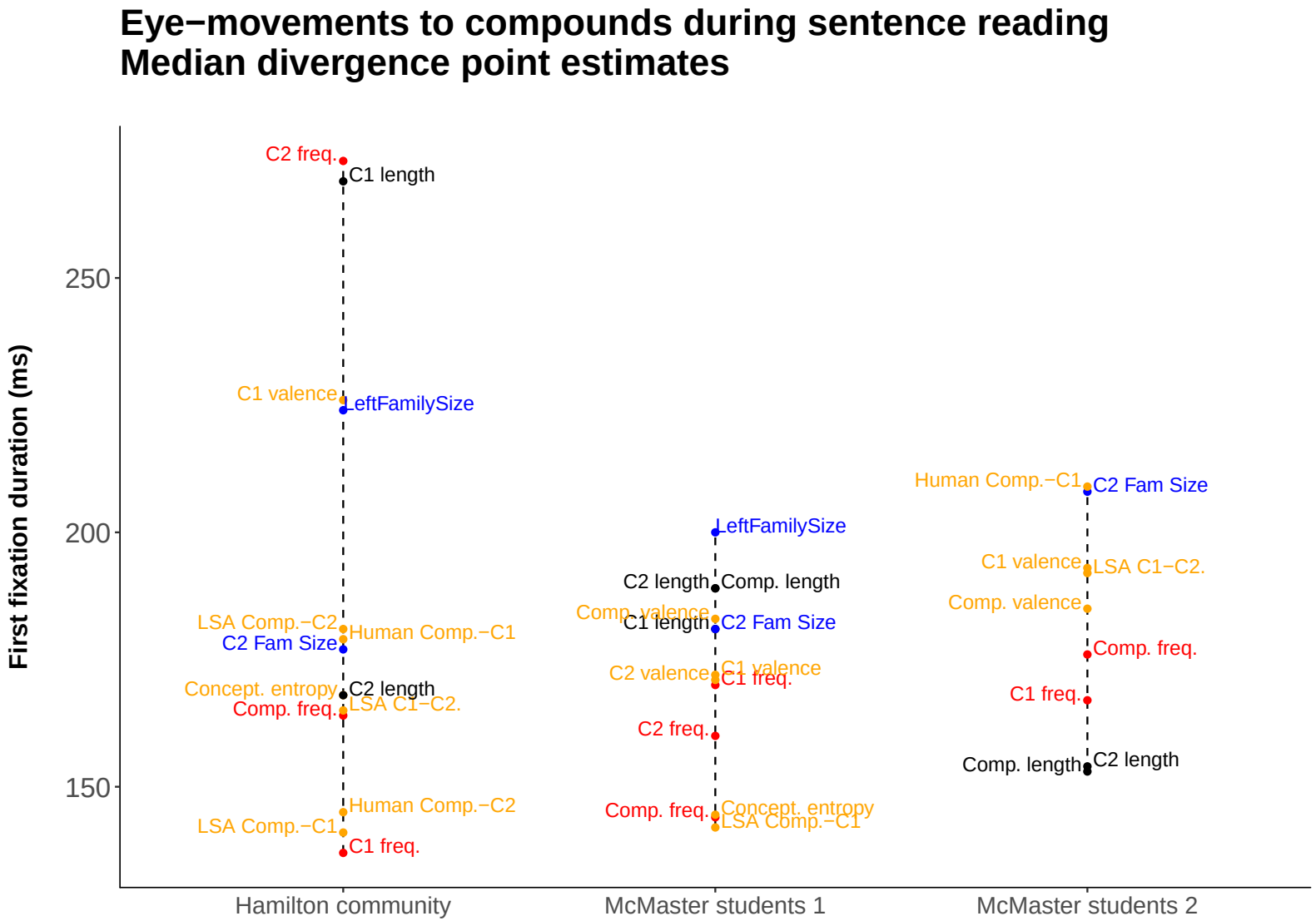
Compounds

- Lexical decision: 269 compounds
- Eye movements: 400 compounds

Compounds: lexical decision



Compounds: eye-movements



Survival analysis: Advantages and drawbacks

- Pros:
  - Great for estimating absolute onset of an effect: 140-220 ms for *all* effects. Constrains for brain studies.
  - Great for mapping relative onsets of effects.
- Cons:
  - Less great for figuring out how long an effect goes for, or when it ends.
  - Robust but not as optimal for multiple continuous predictors

Thank you!