Tutorial on Bayesian Statistics

Shravan Vasishth
Cognitive Science / Linguistics
University of Potsdam, Germany
www.ling.uni-potsdam.de/~vasishth

The frequentist procedure

Imagine that you have some independent and identically distributed data: $x_1, x_2, \ldots, x_n$

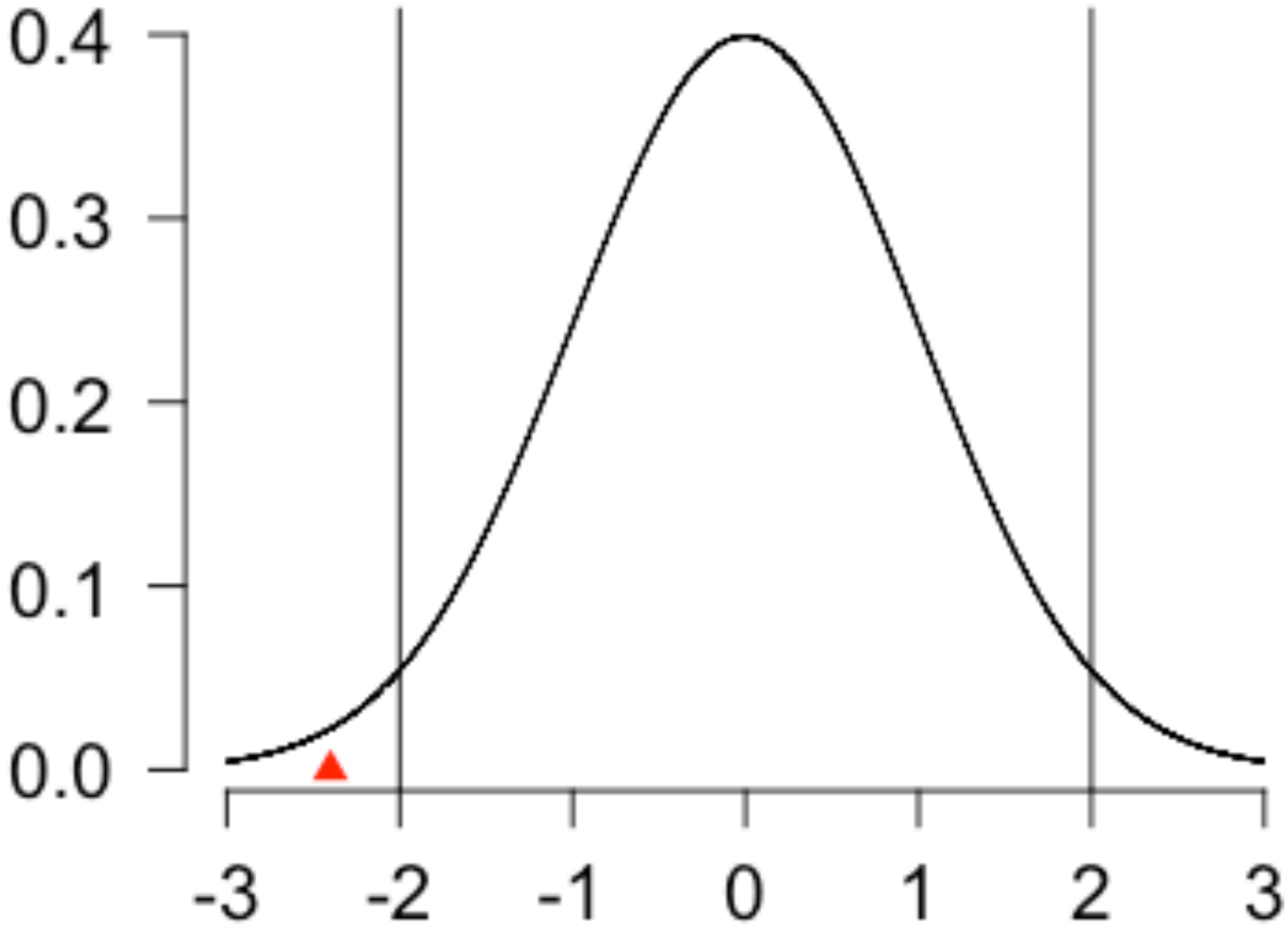
$$X \sim Normal(\mu, \sigma)$$

1. Set up a null hypothesis: $H_0 : \mu = 0$
2. Check if sample mean $\bar{x}$ is consistent with null
3. If inconsistent with null, **accept specific alternative**

Statistical data analysis is reduced to checking for significance (is $p<0.05$?)

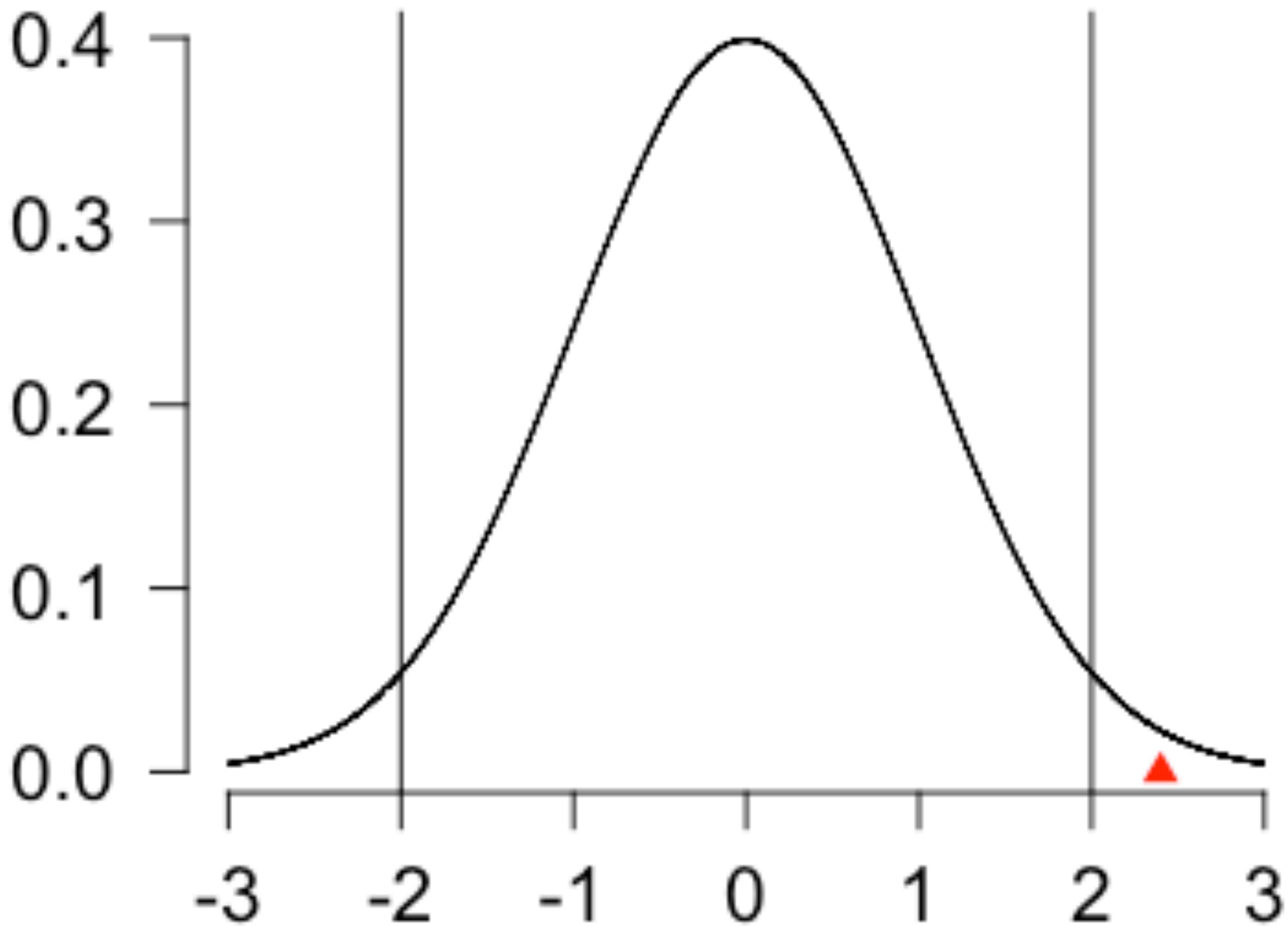
The frequentist procedure

Decision: Reject null and publish



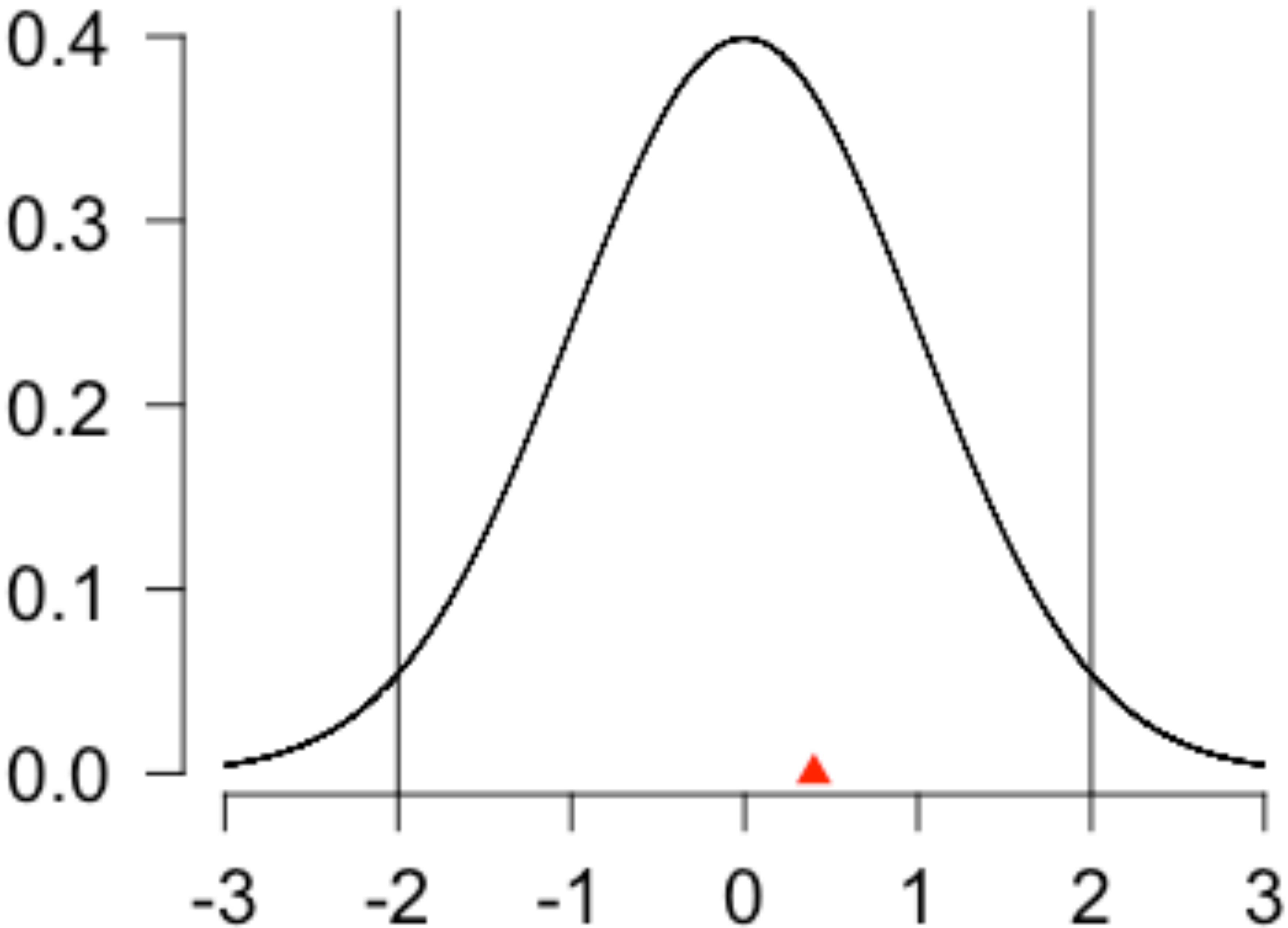
The frequentist procedure

Decision: Reject null and publish



The frequentist procedure

Accept null?



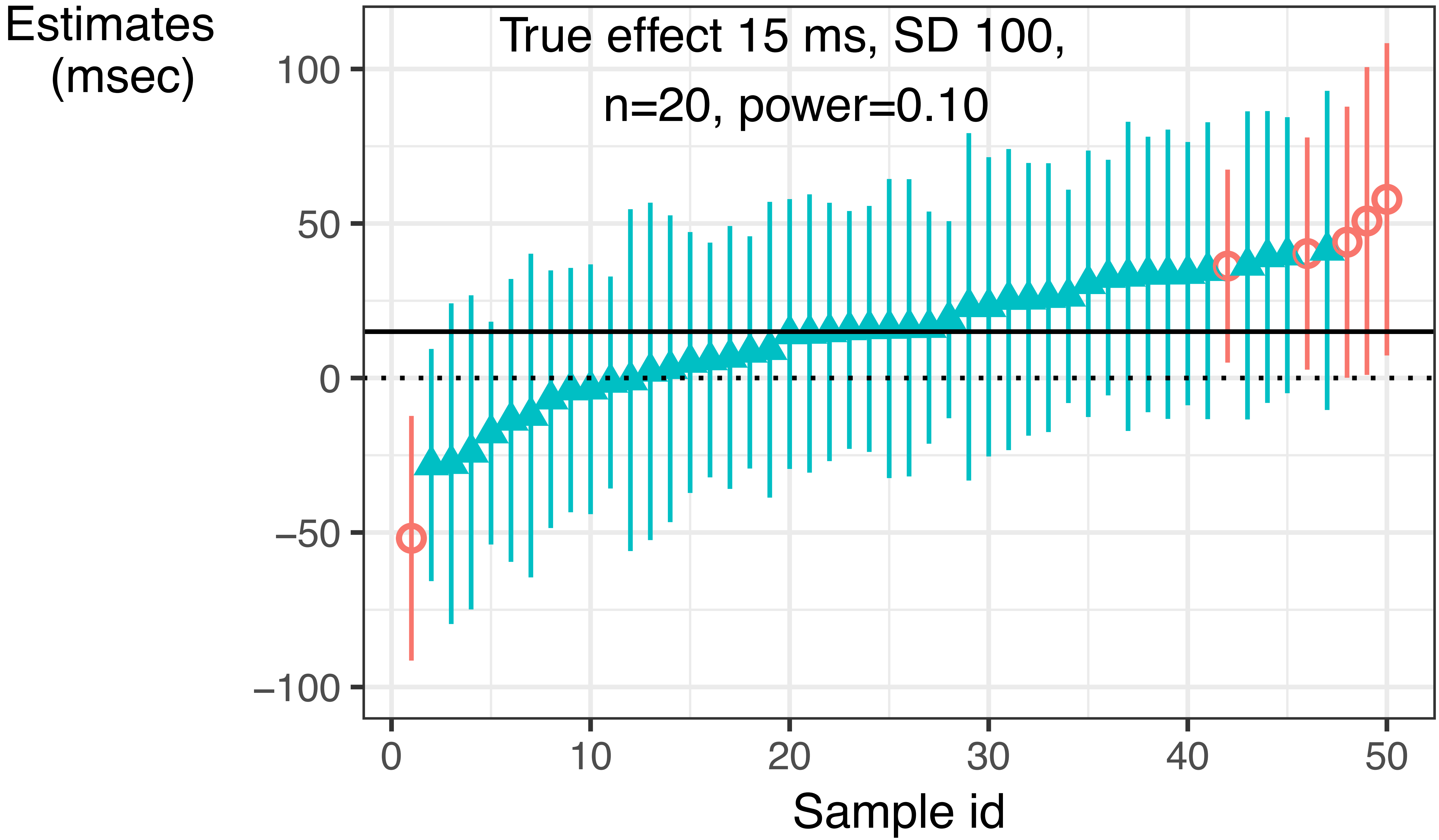
The frequentist procedure

Power: the probability of detecting a particular effect

The frequentist paradigm works when power is high (80% or higher).

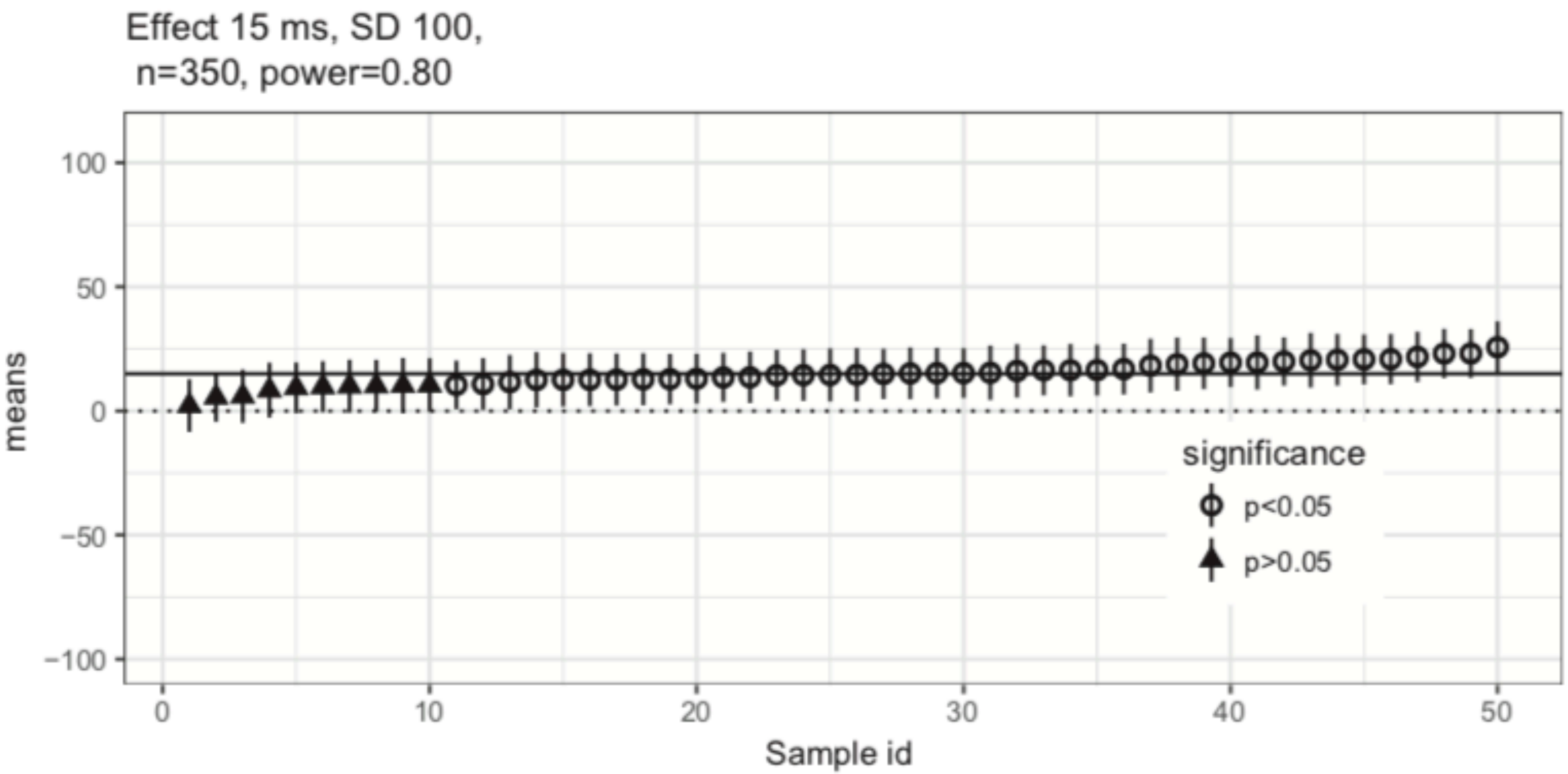
The frequentist paradigm is not designed to be used in low power situations.

Low power leads to exaggerated estimates: Type M error (simulated data)



Gelman & Carlin, 2014

Compare with a high power situation



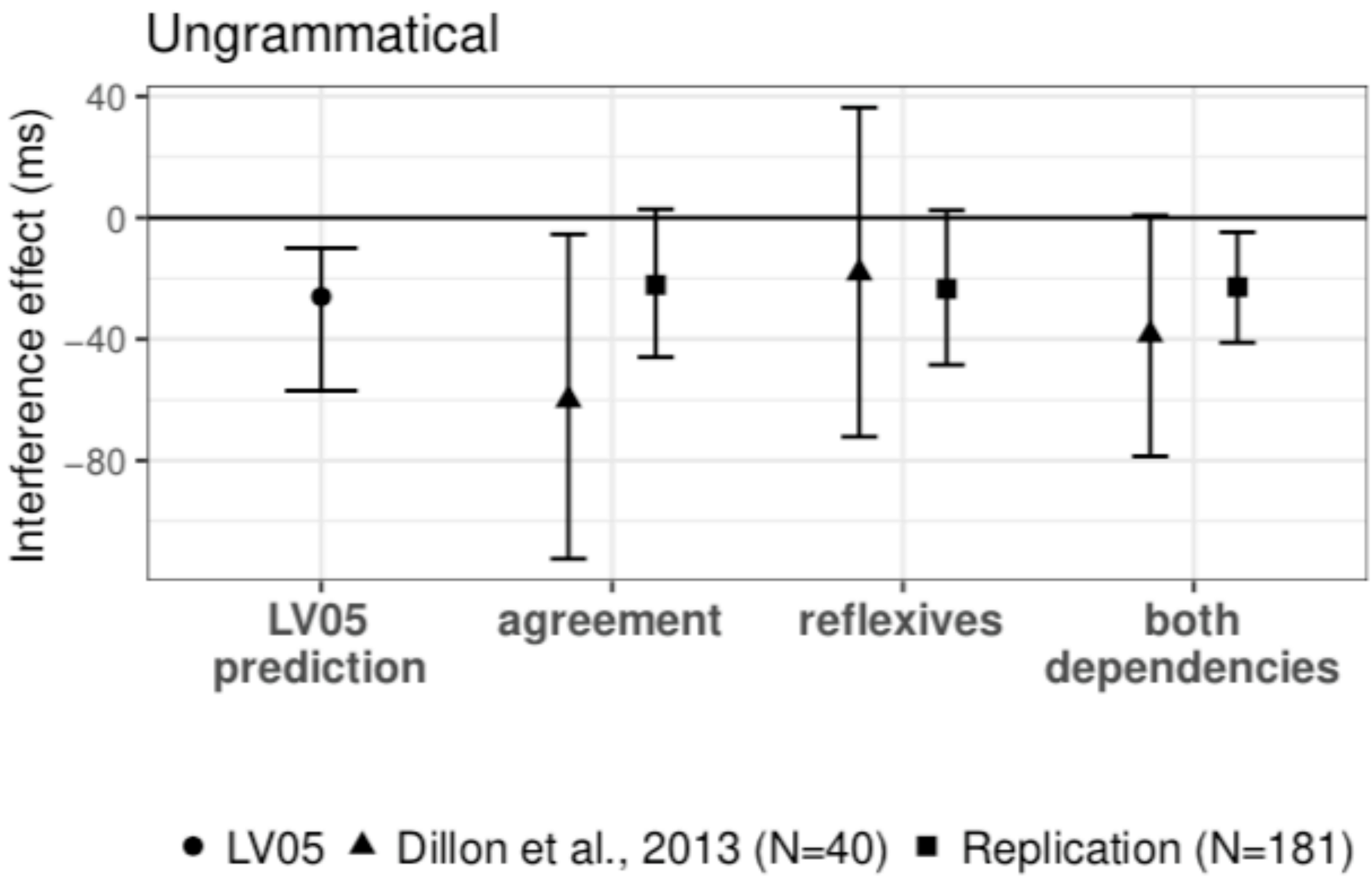
The frequentist paradigm breaks down when power is low

1. Null results are inconclusive
2. Significant results are based on biased estimates (Type M error)

Consequences:

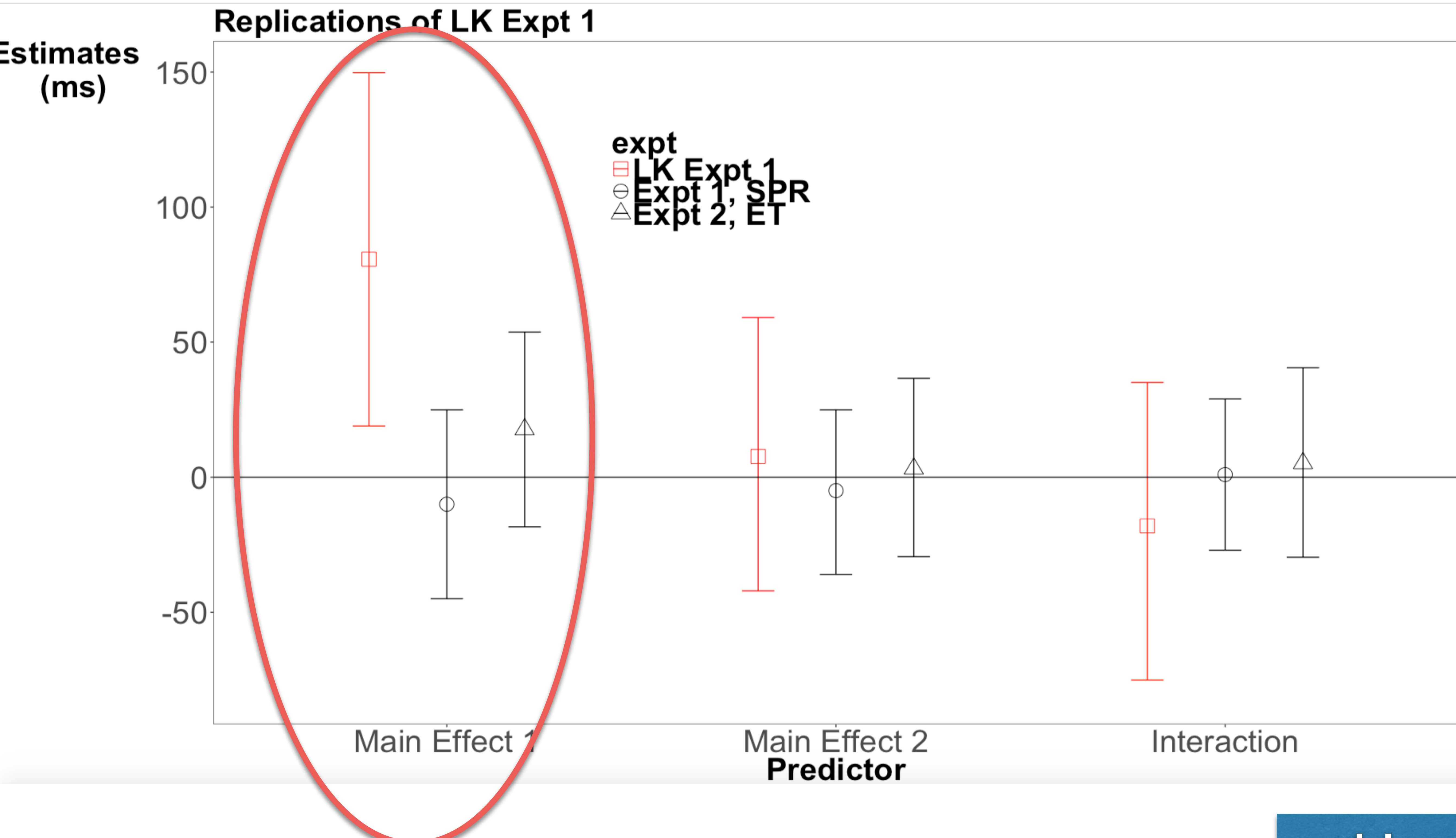
1. Non-replicable results
2. Incorrect inferences

Example 1 of a replication of a low-powered study



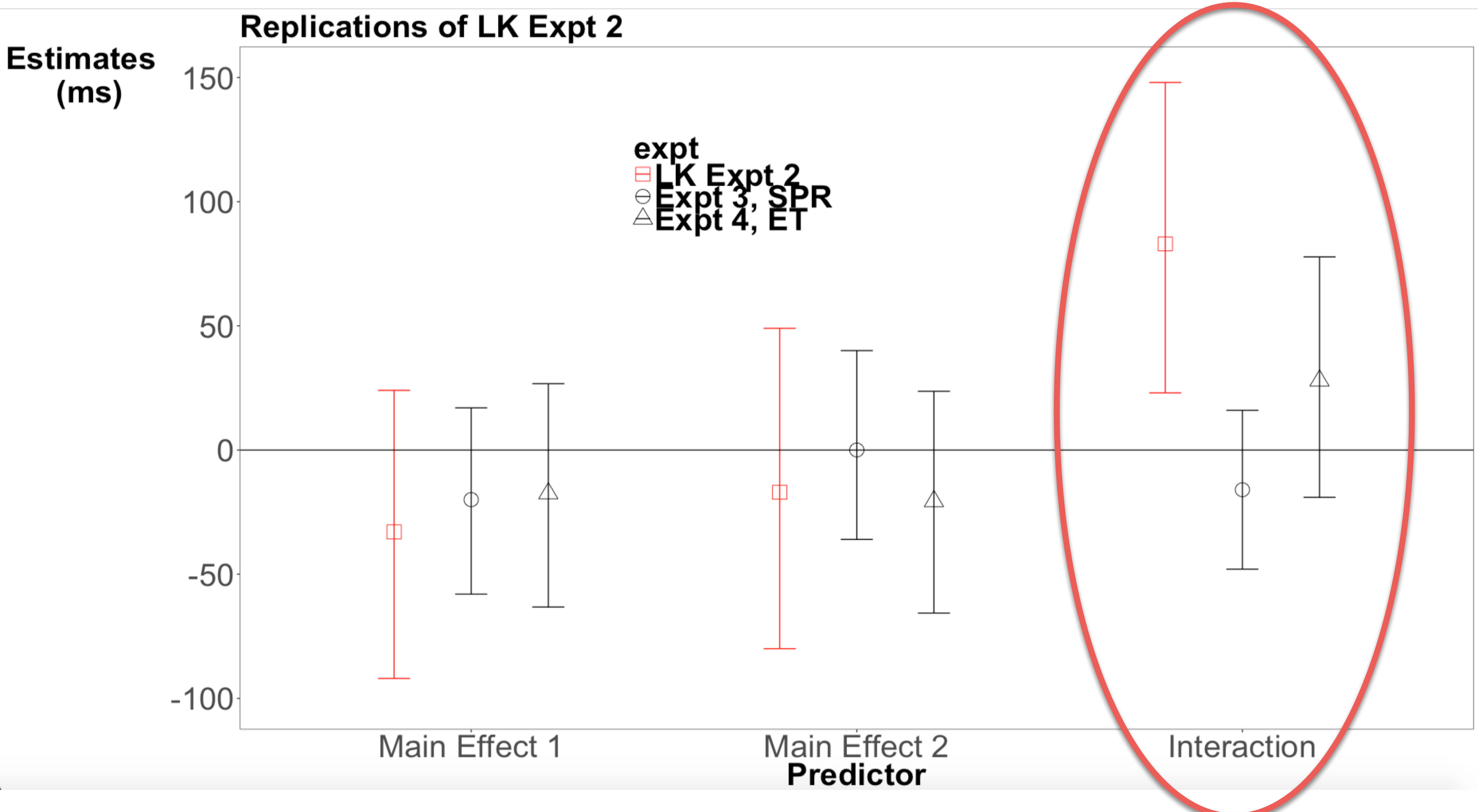
Jäger, Mertzen, Van Dyke & Vasishth, 2020

Example 2 of a replication of a low-powered study



Vasishth, Mertzen, Jäger, Gelman, JML 2018

Example 3 of a replication attempt of a low-powered study



Vasishth, Mertzen, Jäger, Gelman, JML 2018

The Bayesian approach

Imagine that you have some independent and identically distributed data: $x_1, x_2, \ldots, x_n$

$$X \sim Normal(\mu, \sigma)$$

1. Define **priors** for the parameters (here, μ, σ)
2. Derive **posterior distribution** of the parameter(s) of interest using Bayes' rule:

$$\underbrace{f(\mu \mid data)}_{\text{posterior}} \propto \underbrace{f(data \mid \mu)}_{\text{likelihood}} \times \underbrace{f(\mu)}_{\text{prior}}$$

3. Carry out inference based on the posterior

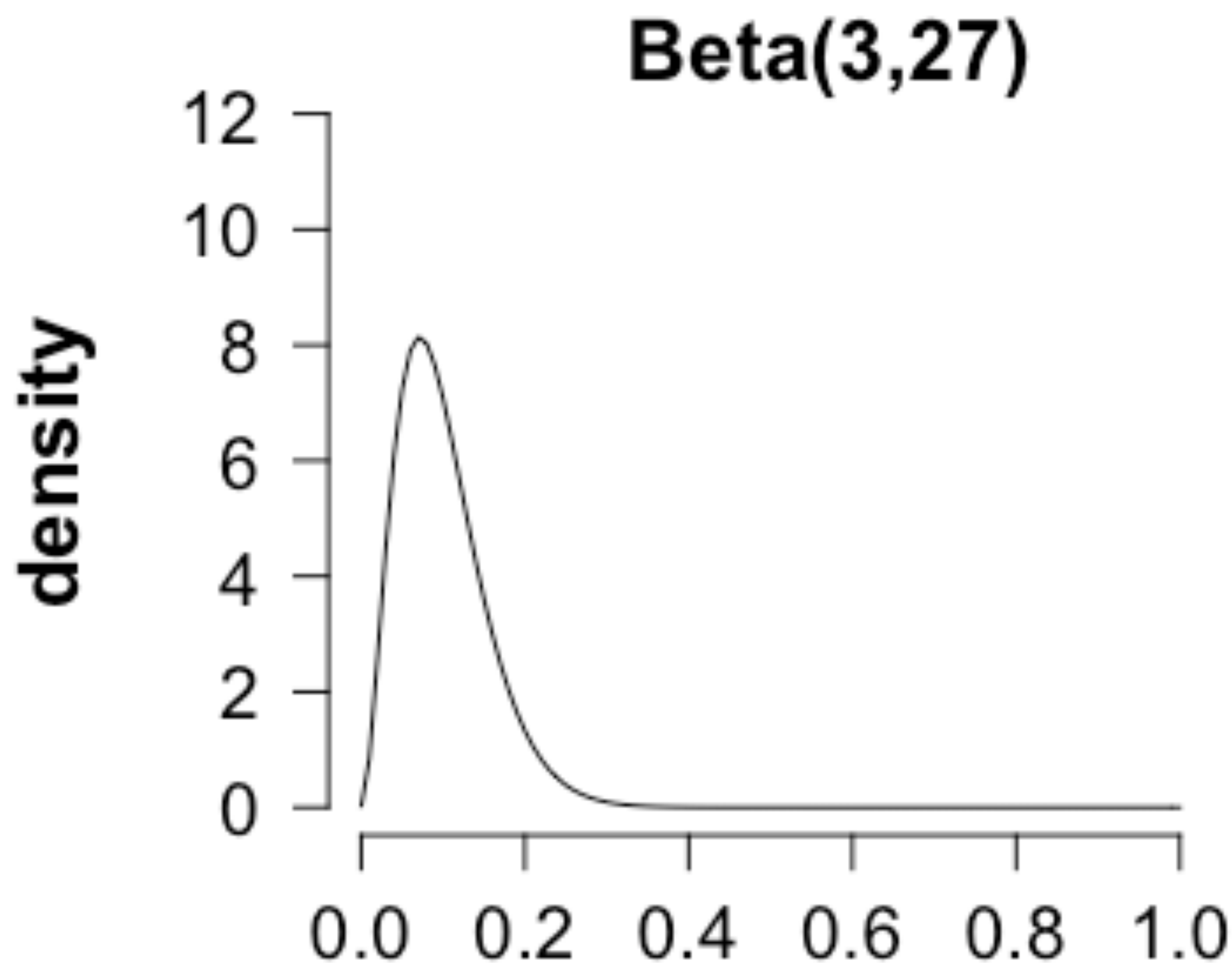
Example: Modeling complications after surgery

Modeling prior knowledge:

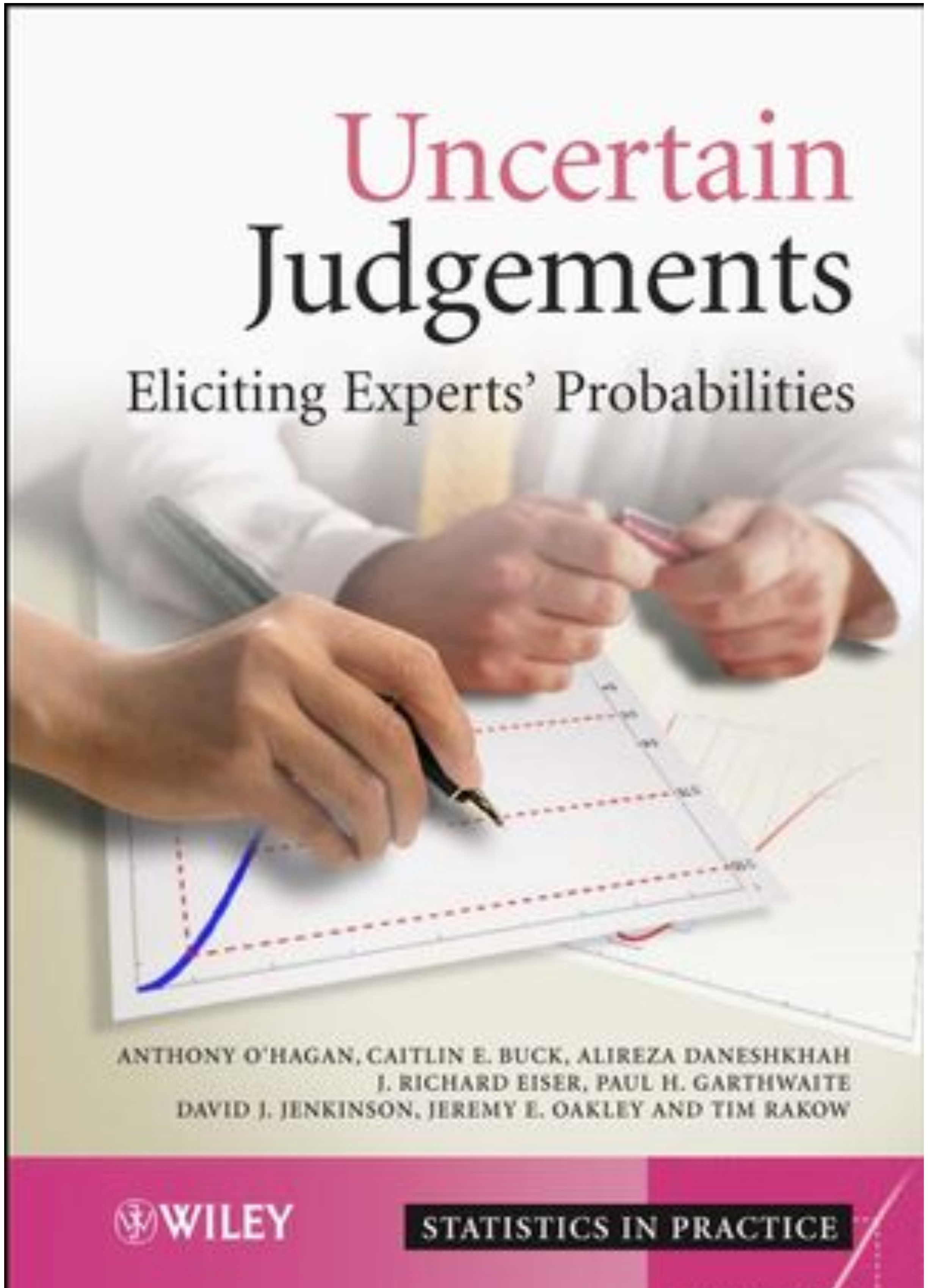
- Suppose we know that 3 out of 30 patients will experience complications after a particular operation
- This prior knowledge can be represented as a Beta(3,27) distribution

Example: Modeling complications after surgery

Modeling prior knowledge:



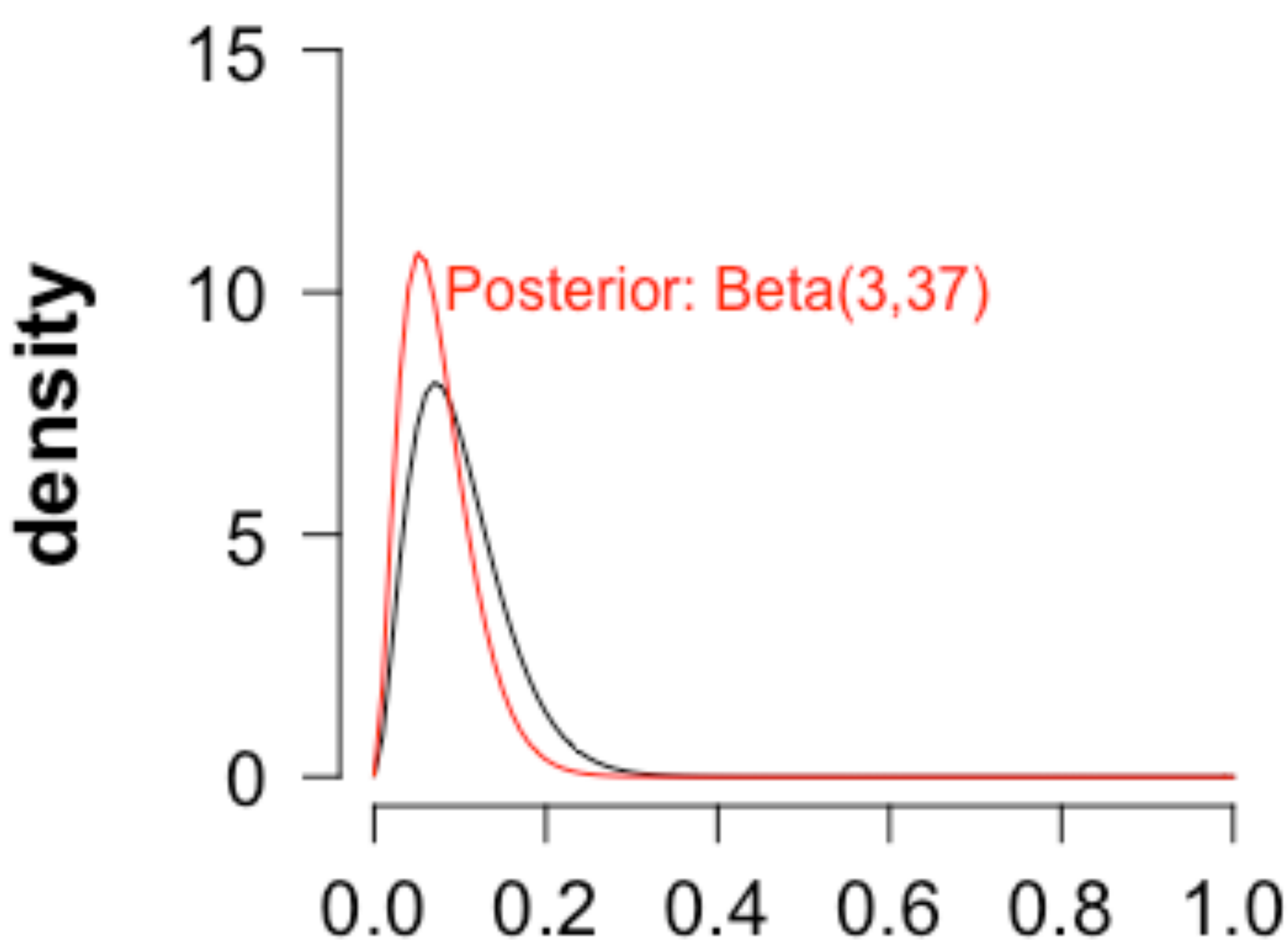
Example: Modeling complications after surgery



Example: Modeling complications after surgery

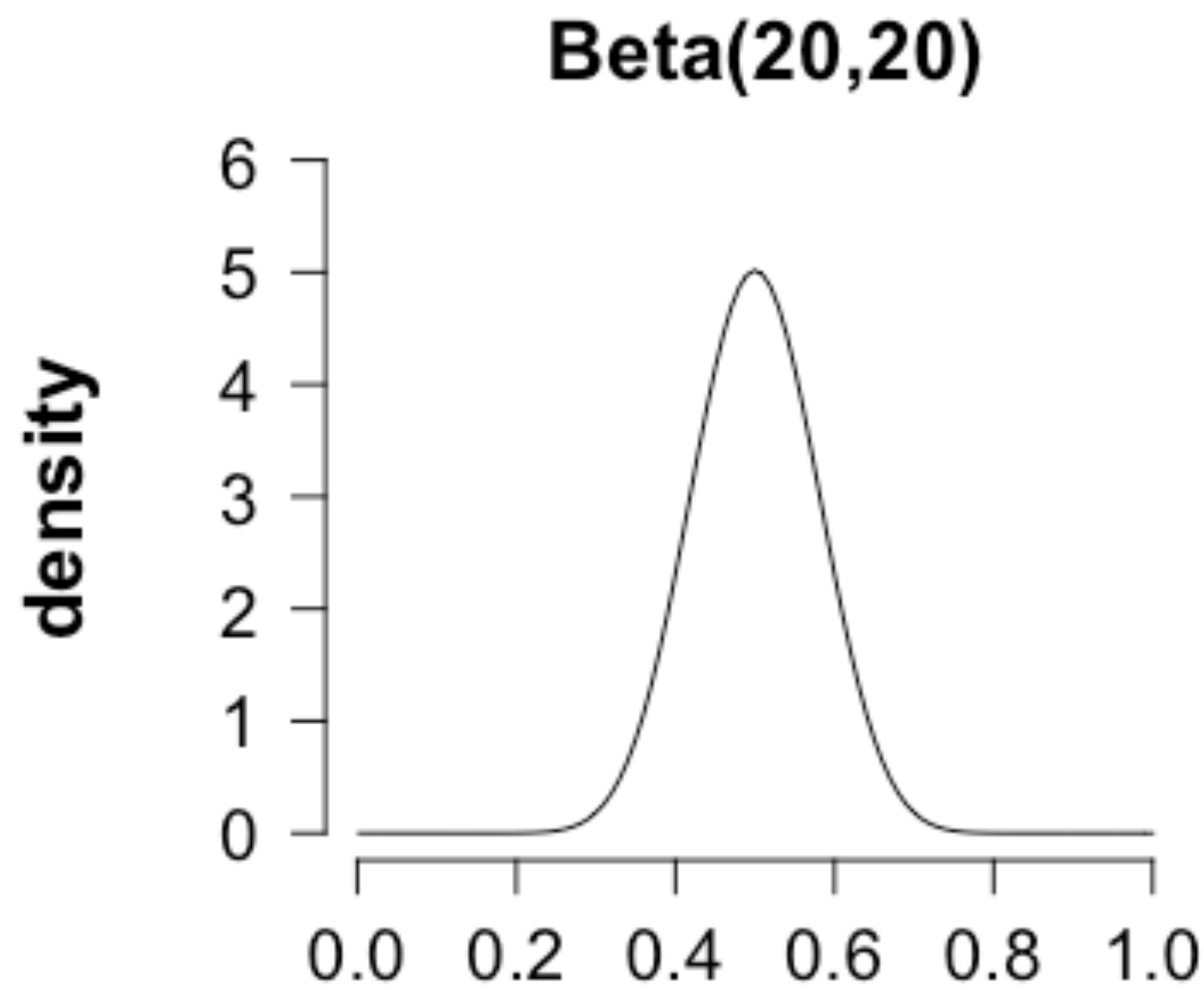
The data: 0 complications in the next 10 operations.
The **posterior distribution** of the probability of complications:

$$Posterior \propto Likelihood \times Prior$$



Example: Modeling complications after surgery

Suppose that Prior probability of complications higher:

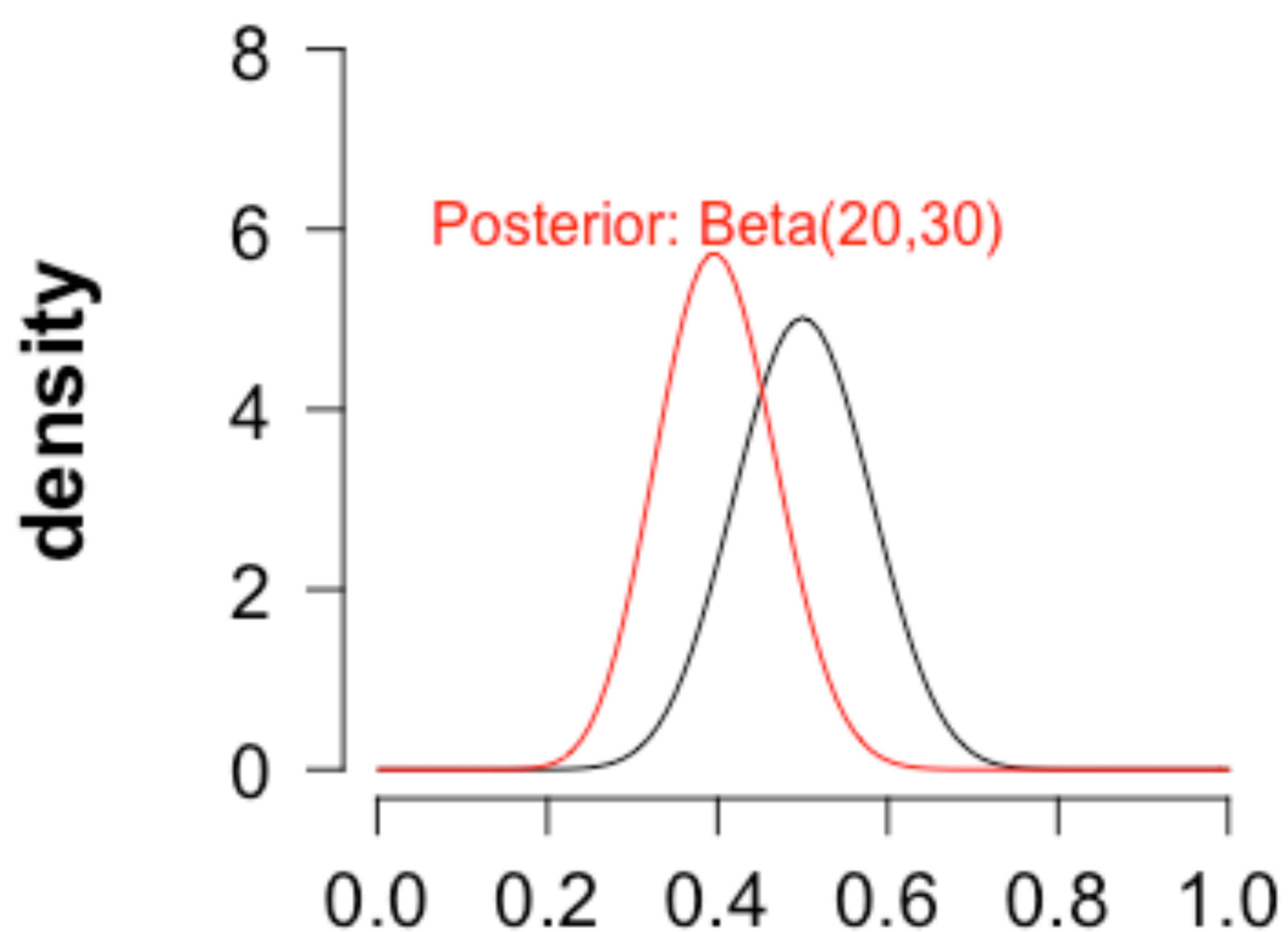


Example: Modeling complications after surgery

The data: 0 complications in the next 10 operations.

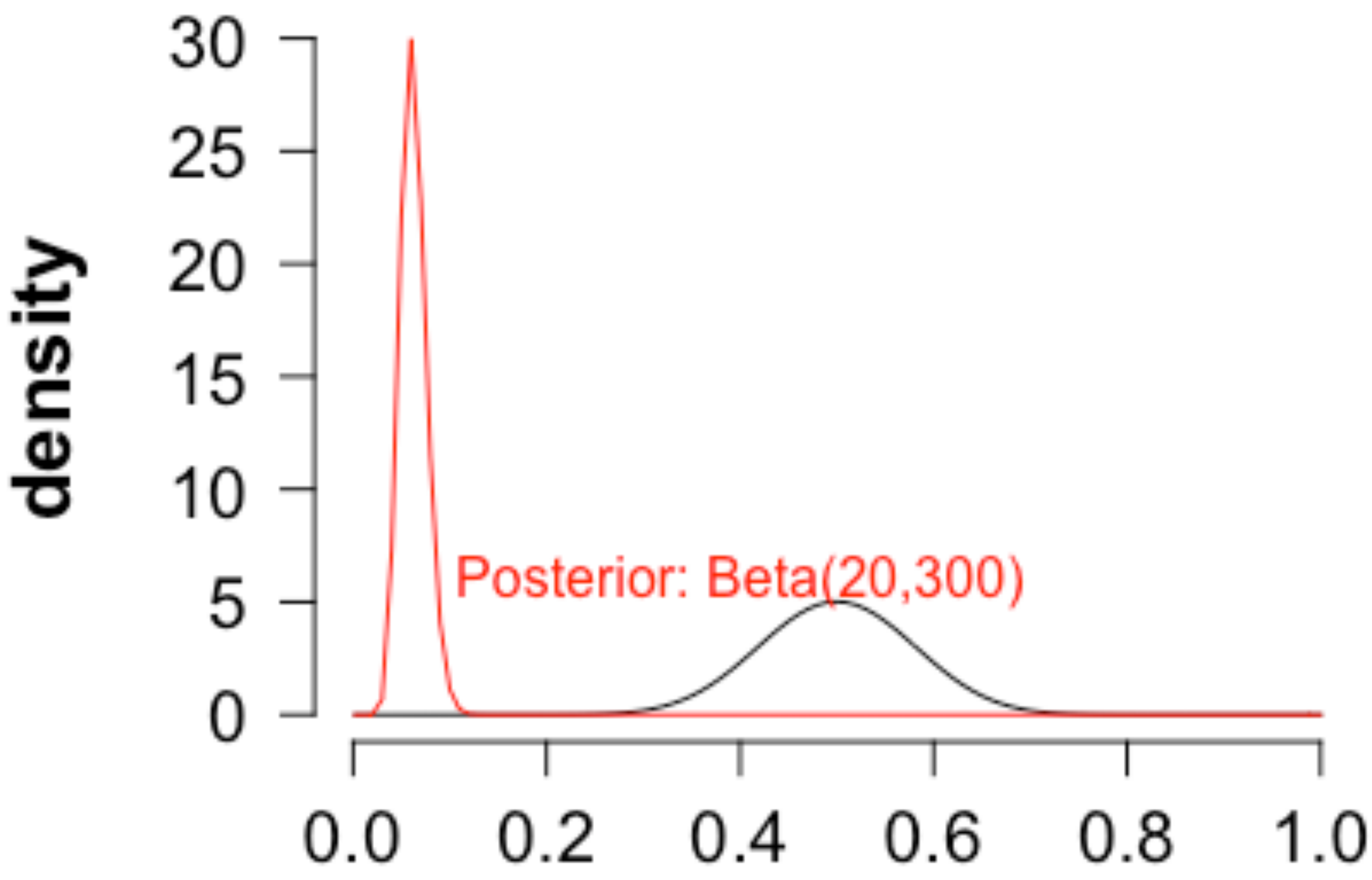
The **posterior distribution** of the probability of complications:

$$Posterior \propto Likelihood \times Prior$$



Example: Modeling complications after surgery

The data: 0 complications in the next **300** operations. The **posterior distribution** of the probability of complications:



Summary

The posterior is a compromise between
the prior and the data

When data are sparse, the posterior reflects the prior

When a lot of data is available, the posterior reflects
the likelihood

Hypothesis testing using the Bayes factor

We may want to compare two alternative models:

$$Model\ 1 : \text{Probability of complications} = 0.5$$

$$Model\ 2 : \text{Probability of complications} \sim Beta(1,1)$$

Bayes factor:

$$BF_{12} = \frac{Prob(Data \mid Model\ 1)}{Prob(Data \mid Model\ 2)}$$

Hypothesis testing using the Bayes factor

$$Model\ 1: \text{Probability of complications} = 0.5$$

$$\binom{n}{k}\theta^0(1-\theta)^{10} = \binom{10}{0}0.5^{10} = 0.000977$$

$$Model\ 2: \text{Probability of complications} \sim Beta(1,1)$$

(Some calculus needed here) $\dfrac{1}{11}$

$$BF_{12} = \frac{Prob(Data \mid Model\ 1)}{Prob(Data \mid Model\ 2)} = \frac{0.000977}{1/11} = 0.01$$

Model 2 is 10 times more likely than Model 1

Comparison of Frequentist vs Bayesian approaches

	Frequentist	Bayesian
Parameters	Fixed	Random
Data	Random	Fixed
Prior knowledge used	No	Yes
Type I, II error	relevant	irrelevant
Hypothesis testing	reject null	Bayes factor
Uncertainty quantification	No	Yes

Some advantages of the Bayesian approach

1. Handles sparse data robustly given priors
2. Highly customised models can be defined
3. The focus is on uncertainty quantification
4. Answers the research question directly

Some "disadvantages" of the Bayesian approach

1. **You have to understand what you are doing**
   - Distribution theory
   - Random variable theory
   - Maximum likelihood estimation
   - Linear modeling theory
2. **Requires programming ability**
   - Statistical computing using Stan (mc-stan.org)
3. **Computational cost**
   - Cluster computing is sometimes needed
   - GPU based computing
4. **Priors require thought**
   - Eliciting priors from experts
   - Adversarial/sensitivity analyses
5. **Low power problems remain!**

A complete example of a Bayesian analysis

Self-paced reading data:
subject vs. object relatives in Chinese (Gibson & Wu 2013)

row	subj	item	so	rt
1	1	13	o	1561
2	1	6	s	959
3	1	5	o	582
4	1	9	o	294
5	1	14	s	438
6	1	4	s	286
⋮	⋮	⋮	⋮	
547	9	11	o	350

A complete example of a Bayesian analysis

Self-paced reading data:
subject vs. object relatives in Chinese (Gibson & Wu 2013)

Standard lmer syntax:

```
lmer(log(rt) ~ so + (1+so|subj),dat)
```

A complete example of a Bayesian analysis

Self-paced reading data:
subject vs. object relatives in Chinese (Gibson & Wu 2013)

The underlying mathematical model:

lmer(log(rt) ~ so + (1+so|subj),dat)

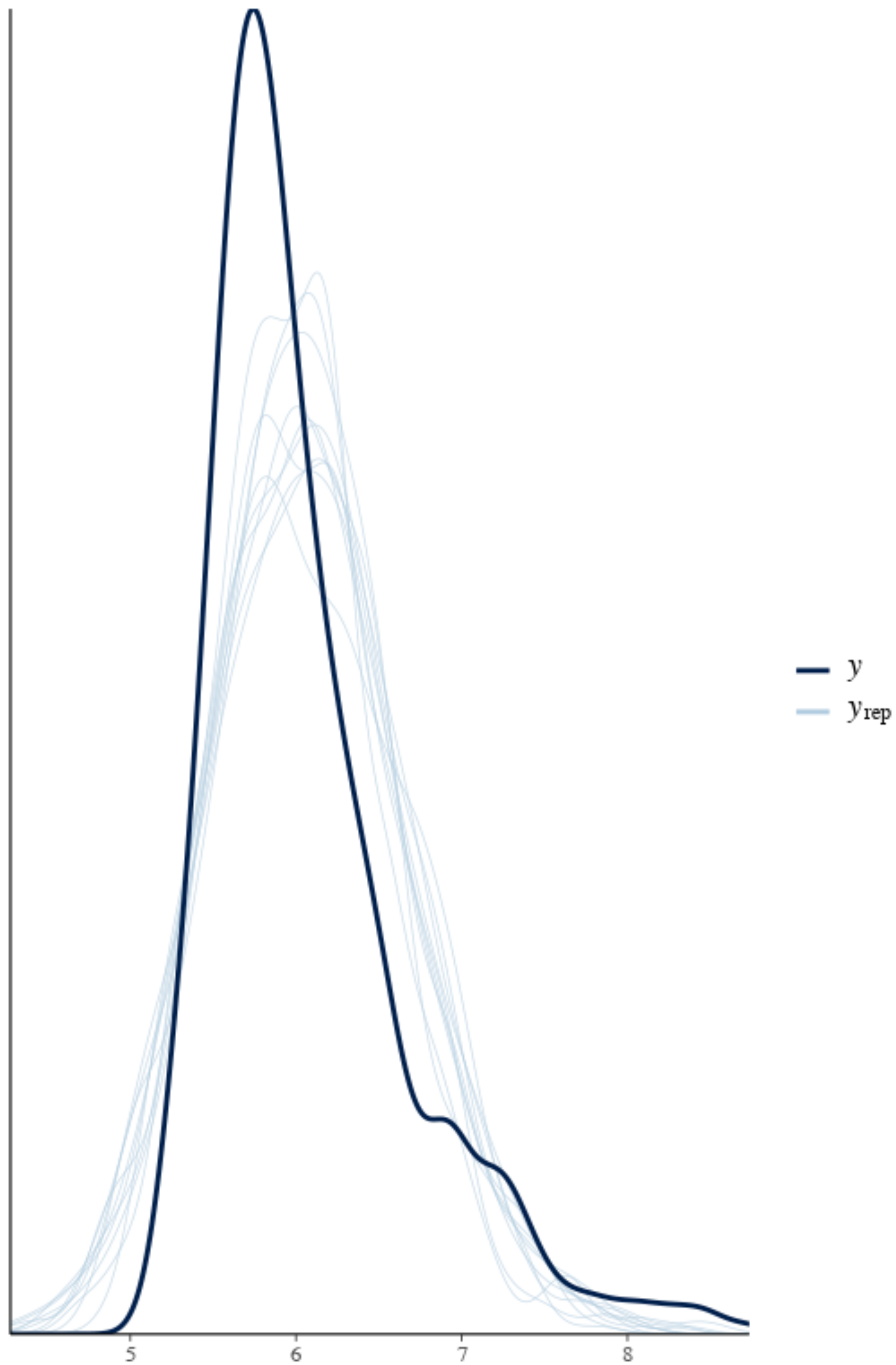
$$rt_n \sim LogNormal(\alpha + u_{1,subj[n]} + so_n \cdot (\beta + u_{2,subj[n]}), \sigma)$$

$$\begin{pmatrix} u_{i,1} \\ u_{i,2} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \boldsymbol{\Sigma_u} \right) \qquad \boldsymbol{\Sigma_u} = \begin{pmatrix} \tau_{u_1}^2 & \rho_u \tau_{u_1} \tau_{u_2} \\ \rho_u \tau_{u_1} \tau_{u_2} & \tau_{u_2}^2 \end{pmatrix}$$

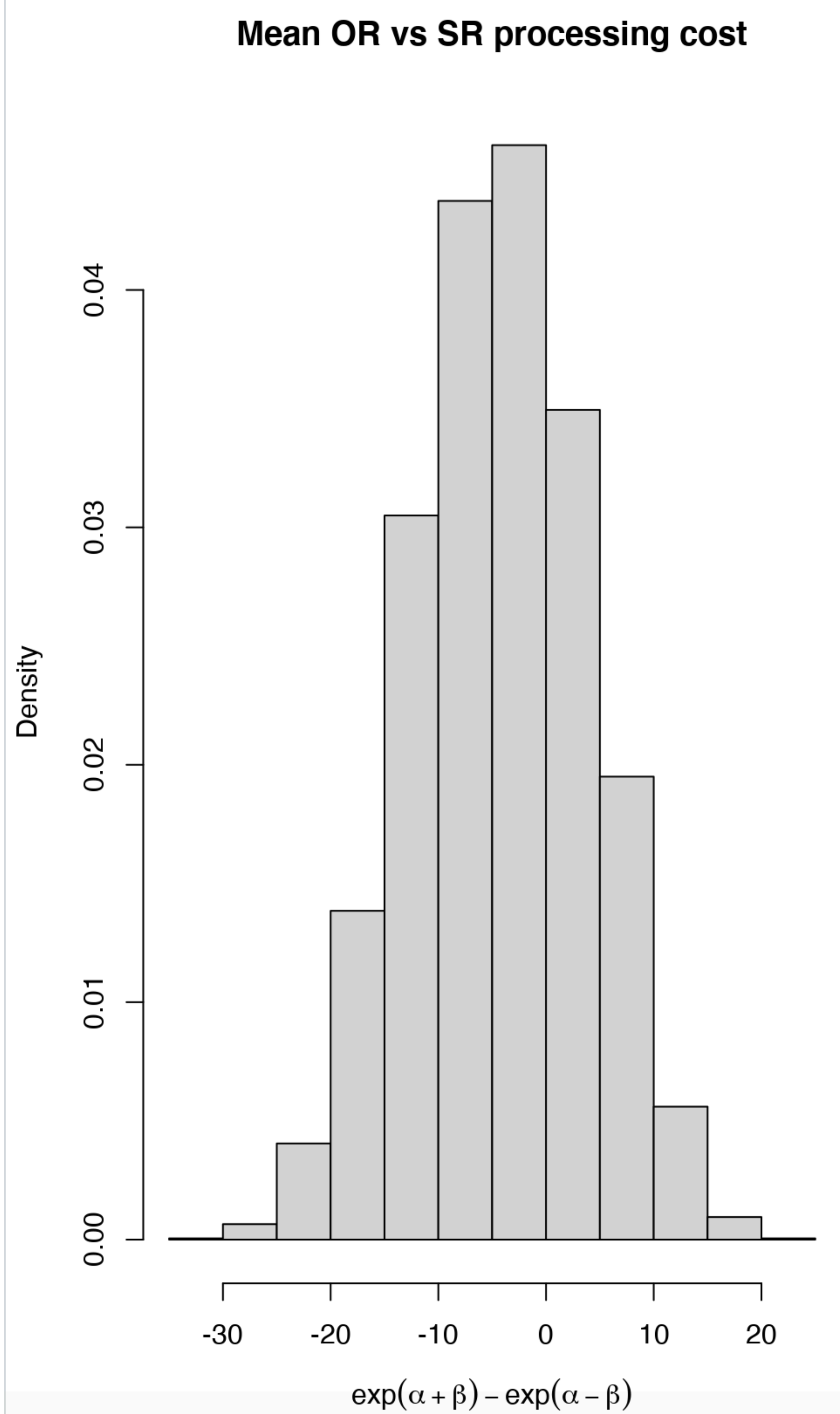
A complete example of a Bayesian analysis

```
library(brms)
fit <- brm(rt ~ 1+so + (1+so | subj),
           family = lognormal(),
           prior =
             c(prior(normal(6, 1.5), class = Intercept),
               prior(normal(0, .01), class = b),
               prior(normal(0, 1), class = sigma),
               prior(normal(0, 1), class = sd),
               prior(lkj(2), class = cor)),
           iter = 4000,
           data = dat)
```

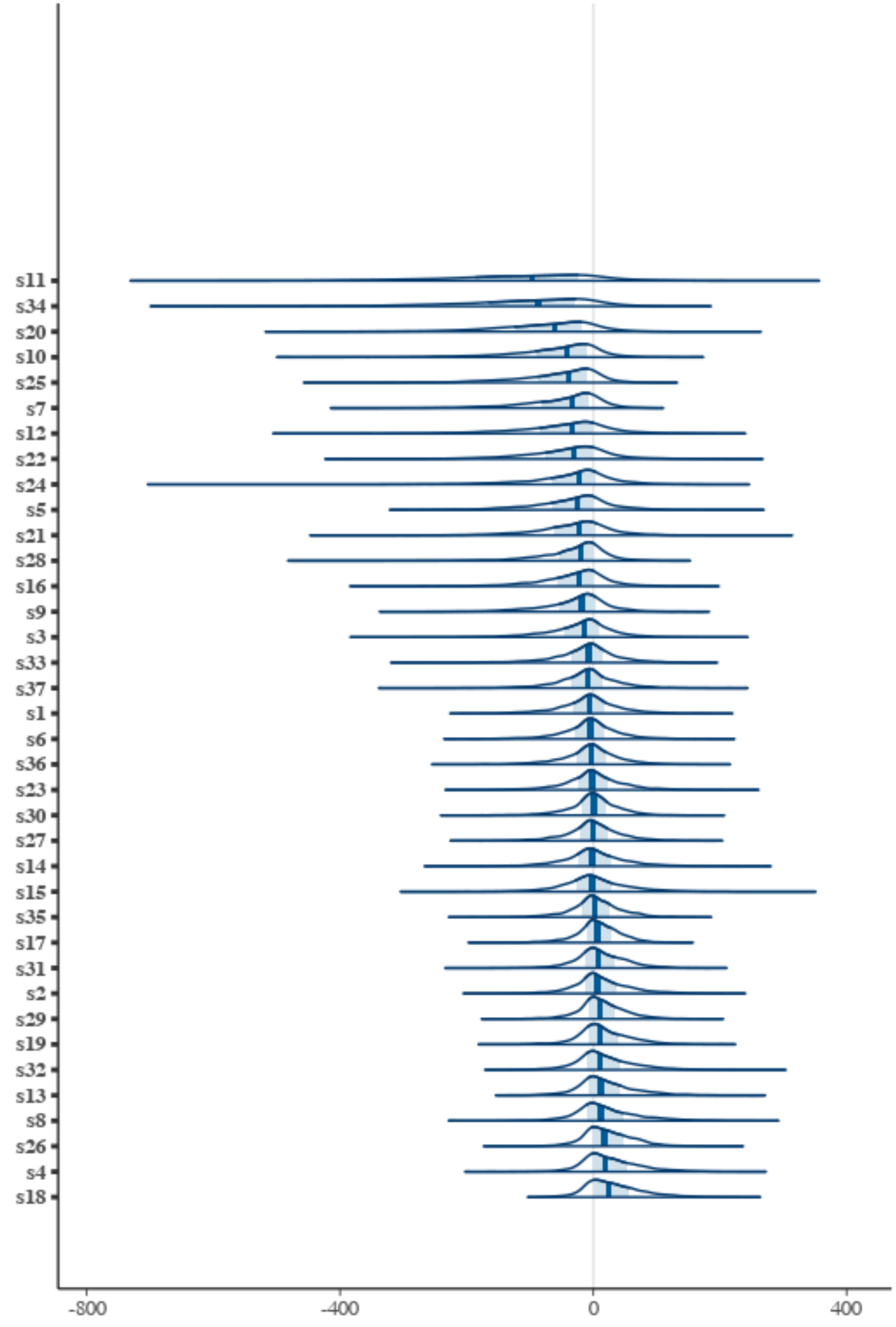
A complete example of a Bayesian analysis



A complete example of a Bayesian analysis



A complete example of a Bayesian analysis



A complete example of a Bayesian analysis

Bayes factor analysis:

```
m_gw<-brm(rt~1+so + (1+so|
subj),gw1,family=lognormal(),
        prior=priors,warmup=5000,
iter=20000,
        save_all_pars = TRUE)

m_gw0<-brm(rt~1 + (1+so|
subj),gw1,family=lognormal(),
         prior=priors0,,warmup=5000,iter=20000,
         save_all_pars = TRUE)
```

A complete example of a Bayesian analysis

Bayes factor analysis:

bayes_factor(m_gw,m_gw0)

Estimated Bayes factor in favor of m_gw over m_gw0:
1.08125

Conclusion: No evidence for difference between RC types

Exercise

Carry out a Bayes factor analysis for the second data-set provided: gw2.

This data-set is intended to be an exact replication of the original study.

Common concerns about Bayesian methods

A post on twitter:

> Doing statistics should be like going to the bathroom. Yes, you have to do it. Yes, when you do it, you want to do it right. But don't make a big deal out of it, be careful about telling other people how to do it, and if your whole life is centered on it, there's something wrong.

I think that the main problem here is that we have been taught in school that statistics should be as easy as:

```
t.test(diff)
```

It's not.
It's hard to unlearn this lesson.

Common concerns about Bayesian methods

A post on twitter:

Doing statistics should be like going to the bathroom. Yes, you have to do it. Yes, when you do it, you want to do it right. But don't make a big deal out of it, be careful about telling other people how to do it, and if your whole life is centered on it, there's something wrong.

We can make data analysis as easy as going to the bathroom, but we should not be surprised if what comes out is crap.

Examples of quick analyses that went wrong

Gibson and Wu 2013:

41,285, $p < .05$; $F_2(1, 14) = 2.23$, $MS_{within} = 22{,}120$, $p = .16$]. The next word consisted of the head noun for the RC, N2. This region was read more slowly in the SRC condition [$F_1(1, 36) = 6.92$, $MS_{within} = 280{,}810$, $p = .01$; $F_2(1, 14) = 4.62$, $MS_{within} = 110{,}132$, $p < .05$]. When these two regions – the RC

Levy and Keller 2013:

ELSEVIER

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

The statistical significance filter leads to overoptimistic expectations of replicability

Shravan Vasishth[a,*], Daniela Mertzen[a], Lena A. Jäger[a], Andrew Gelman[b]

[a] Department of Linguistics, University of Potsdam, Potsdam, Germany
[b] Department of Statistics, Columbia University, New York, USA

Recommended readings

Textbooks

- Kruschke, J. (2014). Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan. Elsevier.
- Lambert, B. (2018). A student's guide to Bayesian statistics. Sage.
- McElreath, R. (2020). Statistical rethinking: A Bayesian course with examples in R and Stan. CRC Press.

Further suggested readings

https://vasishth.github.io/IntroductionBayes/

Almost anything written by:
- Wagenmakers
- Rouder
- Morey

Further suggested readings

Bayes factors tutorial

Statistics > Methodology

[Submitted on 15 Mar 2021 (v1), last revised 18 Mar 2021 (this version, v2)]

Workflow Techniques for the Robust Use of Bayes Factors

Daniel J. Schad, Bruno Nicenboim, Paul-Christian Bürkner, Michael Betancourt, Shravan Vasishth

Bayesian workflow tutorial

Statistics > Methodology

[Submitted on 29 Apr 2019 (v1), last revised 28 Feb 2020 (this version, v3)]

Toward a principled Bayesian workflow in cognitive science

Daniel J. Schad, Michael Betancourt, Shravan Vasishth

Annual summer school in Statistical Methods for Linguistics and Psychology at Potsdam

https://vasishth.github.io/smlp2021/