Frequentist vs Bayesian data analysis

Shravan Vasishth / Bruno Nicenboim
LOT Winter School 2020

The frequentist procedure

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

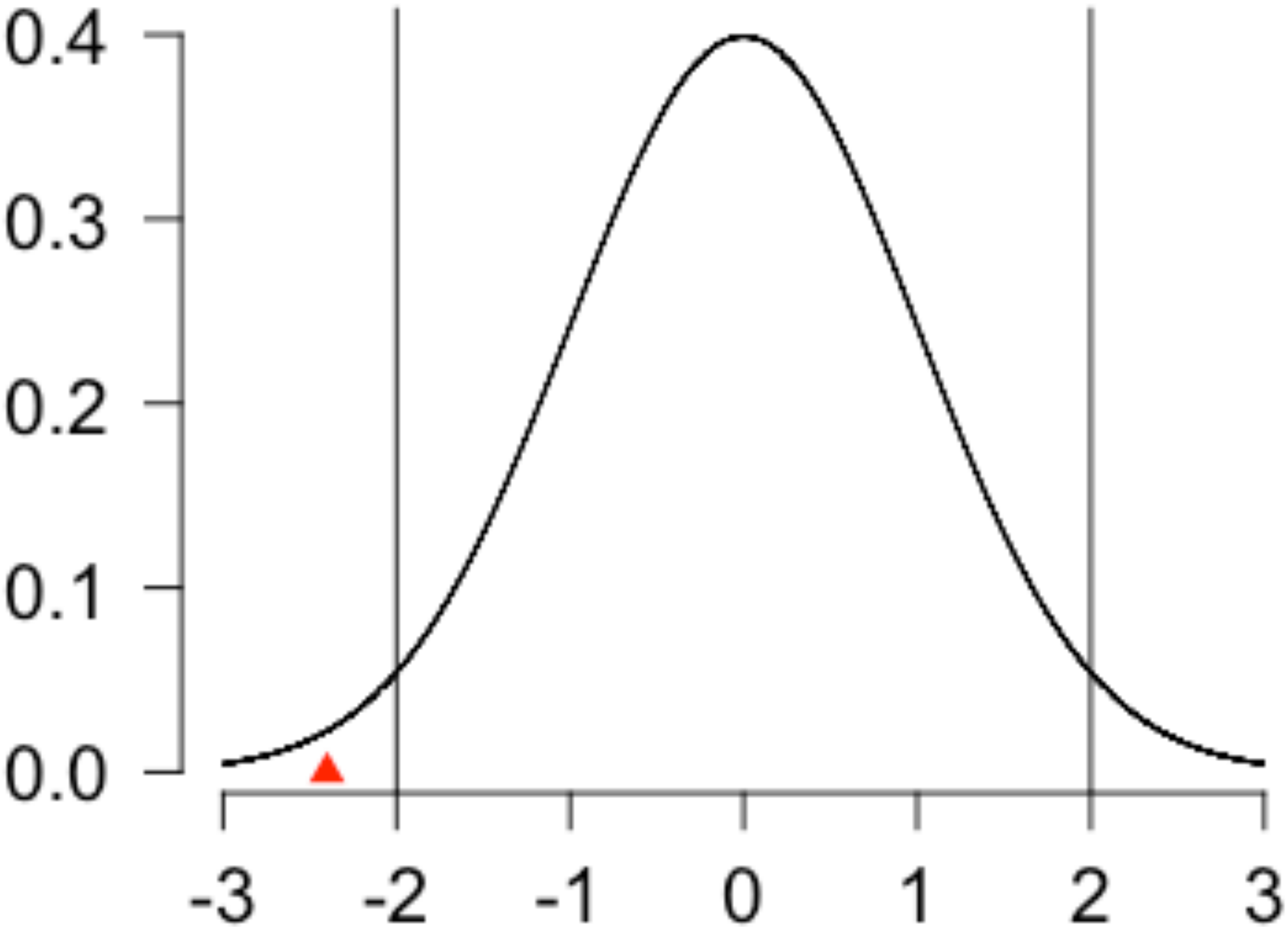
$$X \sim Normal(\mu, \sigma)$$

1. Set up a null hypothesis: $H_0 : \mu = \mu_0$
2. Check how far sample mean $\bar{x}$ is from μ_0 in SE units:
   $t_{observed} \times SE = \bar{x} - \mu_0$
3. If $t_{observed}$ is large enough, reject null hypothesis

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

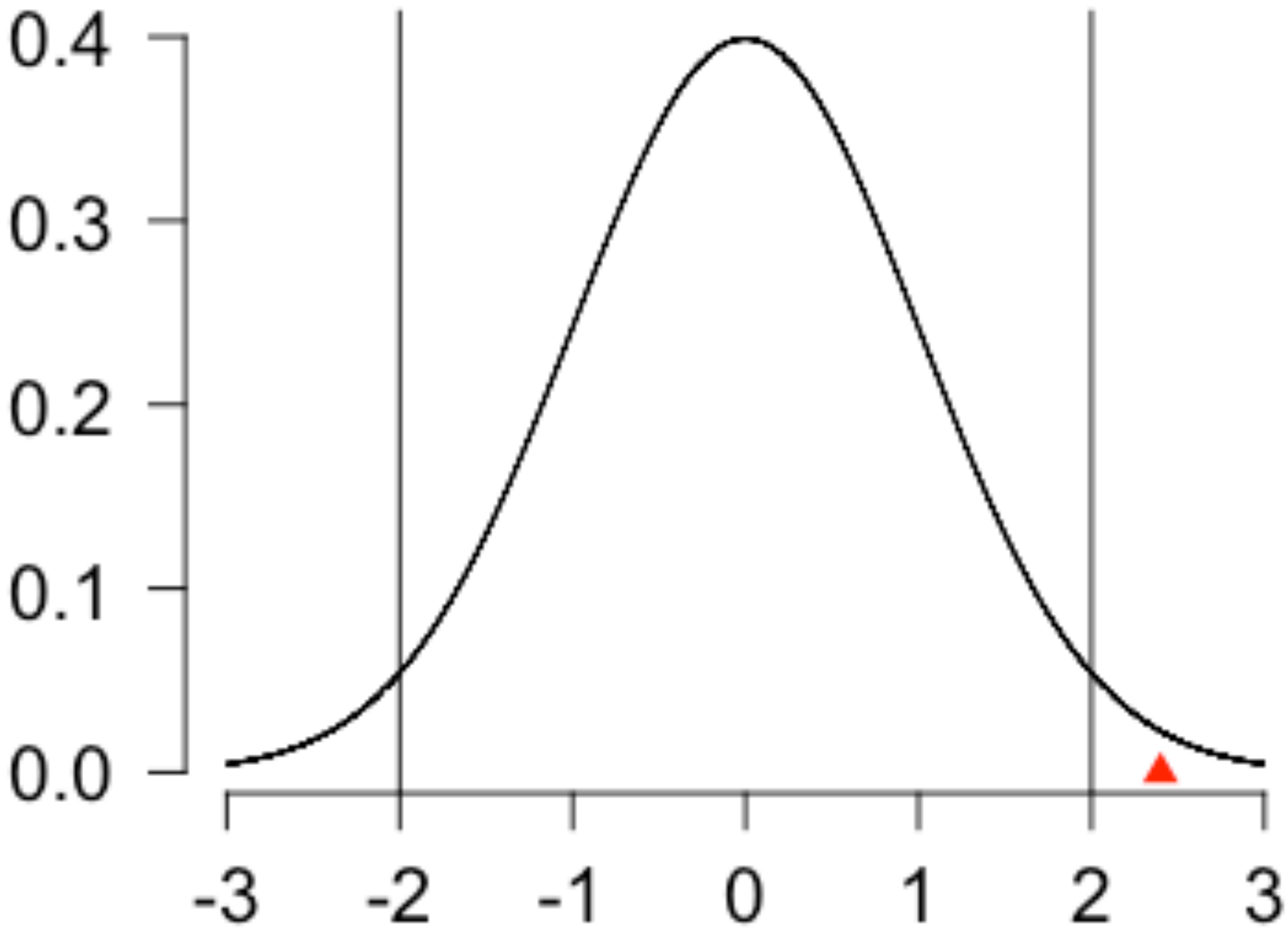
The frequentist procedure

Decision: Reject null and publish



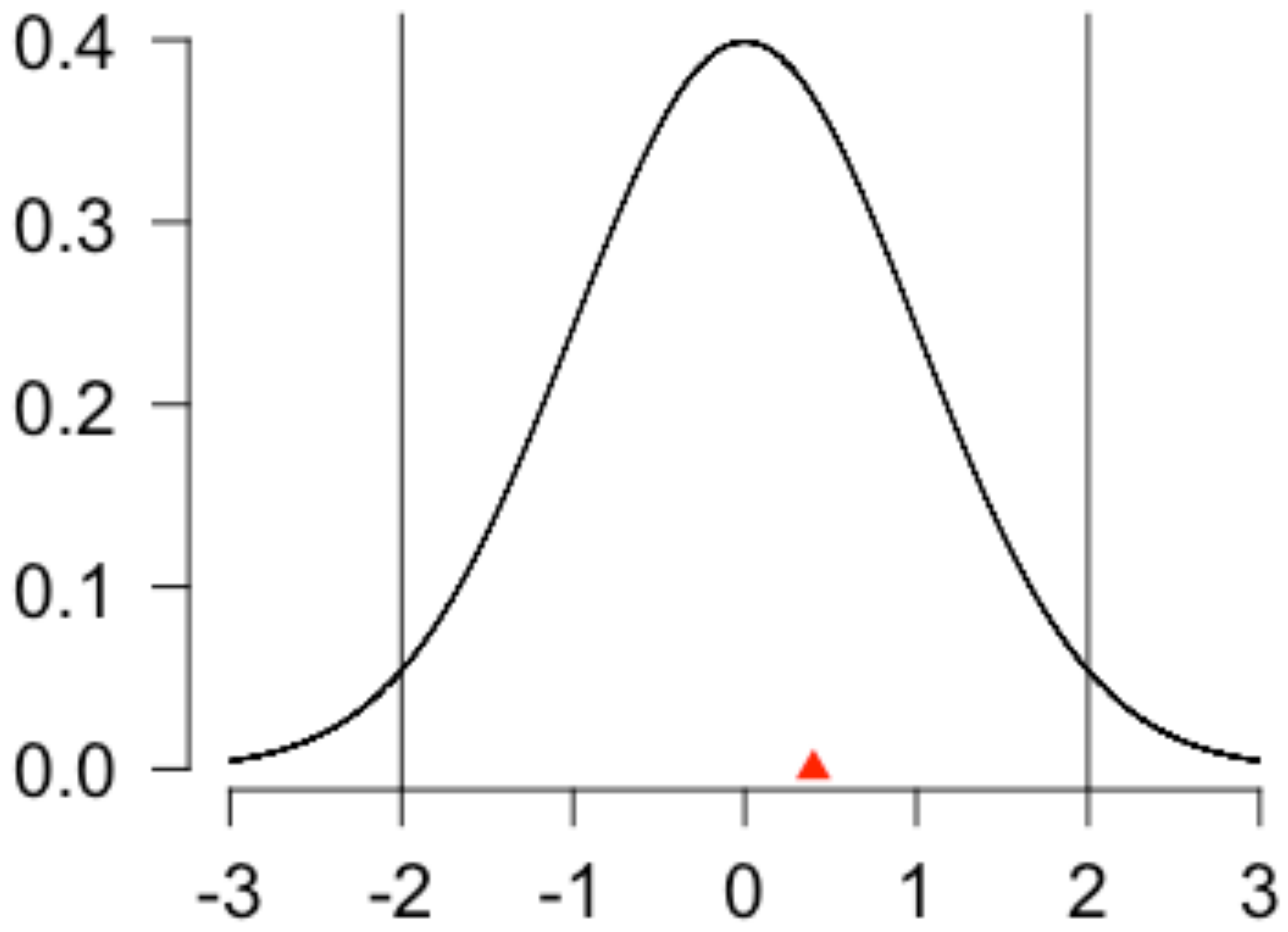
The frequentist procedure

Decision: Reject null and publish



The frequentist procedure

Accept null?



The frequentist procedure

Power: the probability of detecting a particular effect

Power depends on:

- effect size (+ experiment design)
- standard deviation(s)
- sample size

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

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

Example: agreement attraction

Journal of Memory and Language 111 (2020) 104063

Contents lists available at ScienceDirect

Journal of Memory and Language

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

Interference patterns in subject-verb agreement and reflexives revisited: A large-sample study☆

Lena A. Jäger[a,b], Daniela Mertzen[b], Julie A. Van Dyke[c], Shravan Vasishth[b,*]

[a] Department of Computer Science, University of Potsdam, Germany
[b] Department of Linguistics, University of Potsdam, Germany
[c] Haskins Laboratories, New Haven, CT, United States

Jäger, Mertzen, Van Dyke & Vasishth, JML, 2020

Example: agreement attraction

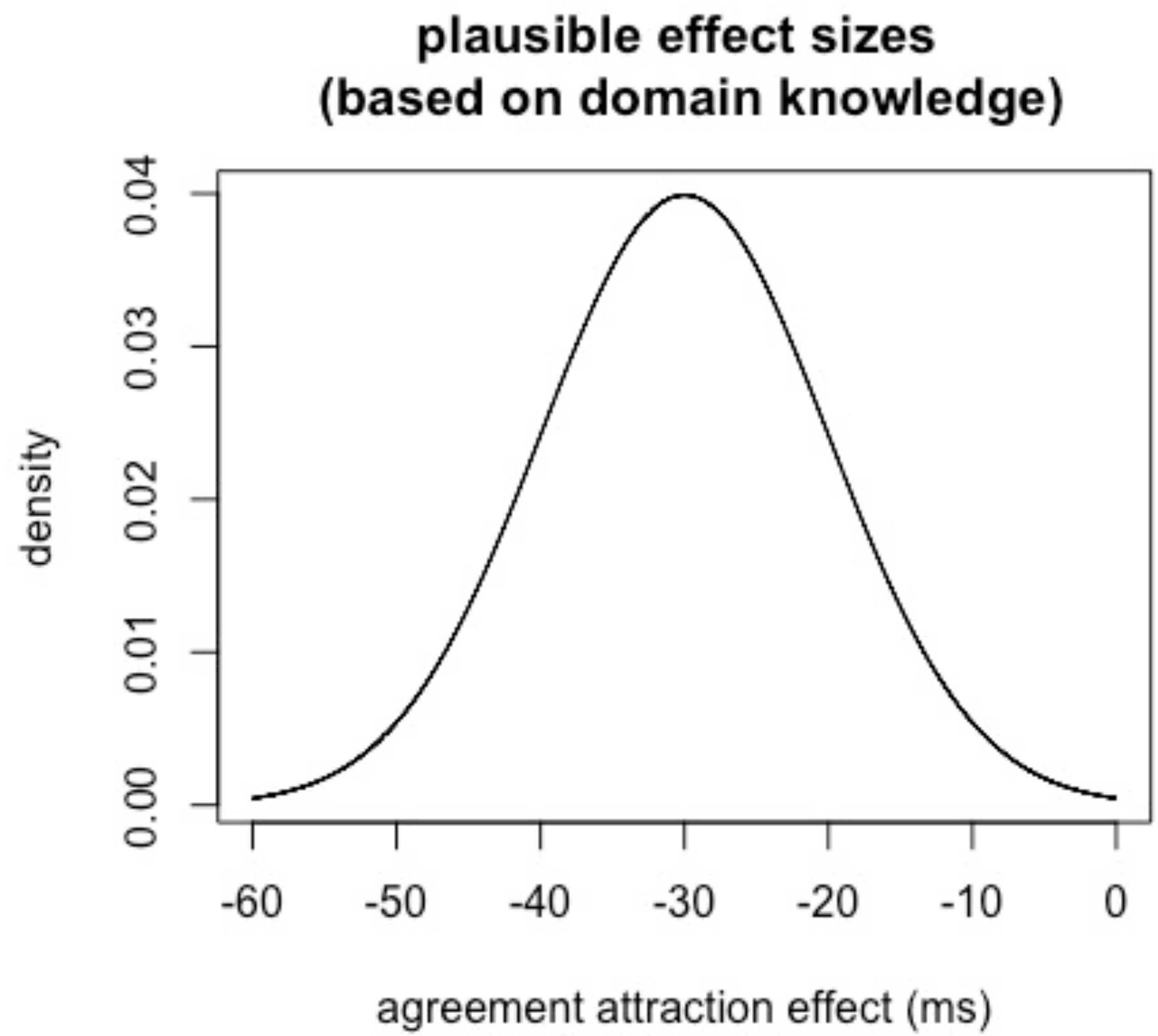
a. *The bodybuilder$^{-plural}_{+subject}$ who met the trainers$^{+plural}_{-subject}$ were$\{^{plural}_{subject}\}$

...

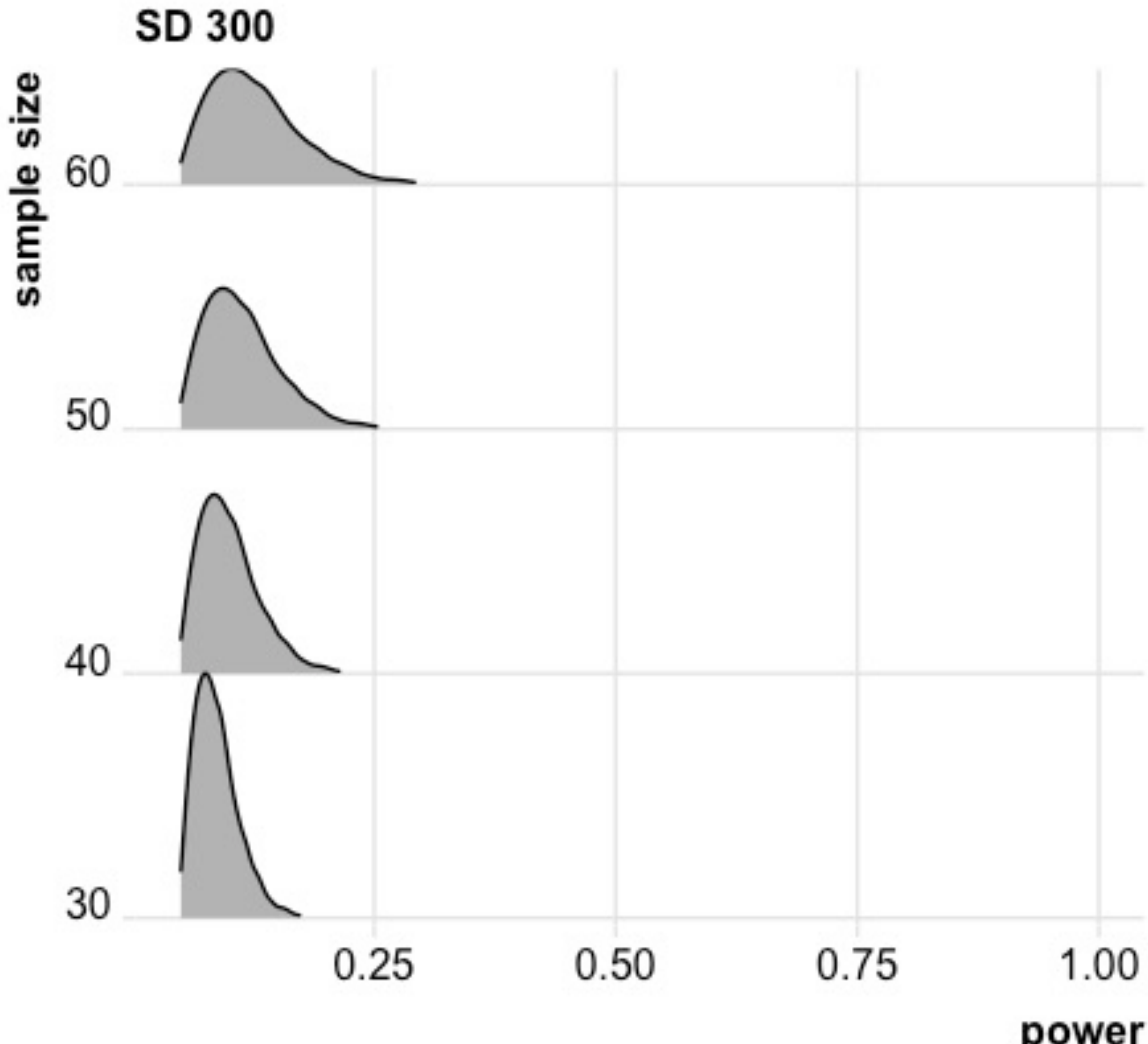
b. *The bodybuilder$^{-plural}_{+subject}$ who met the trainer$^{-plural}_{-subject}$ were$\{^{plural}_{subject}\}$

...

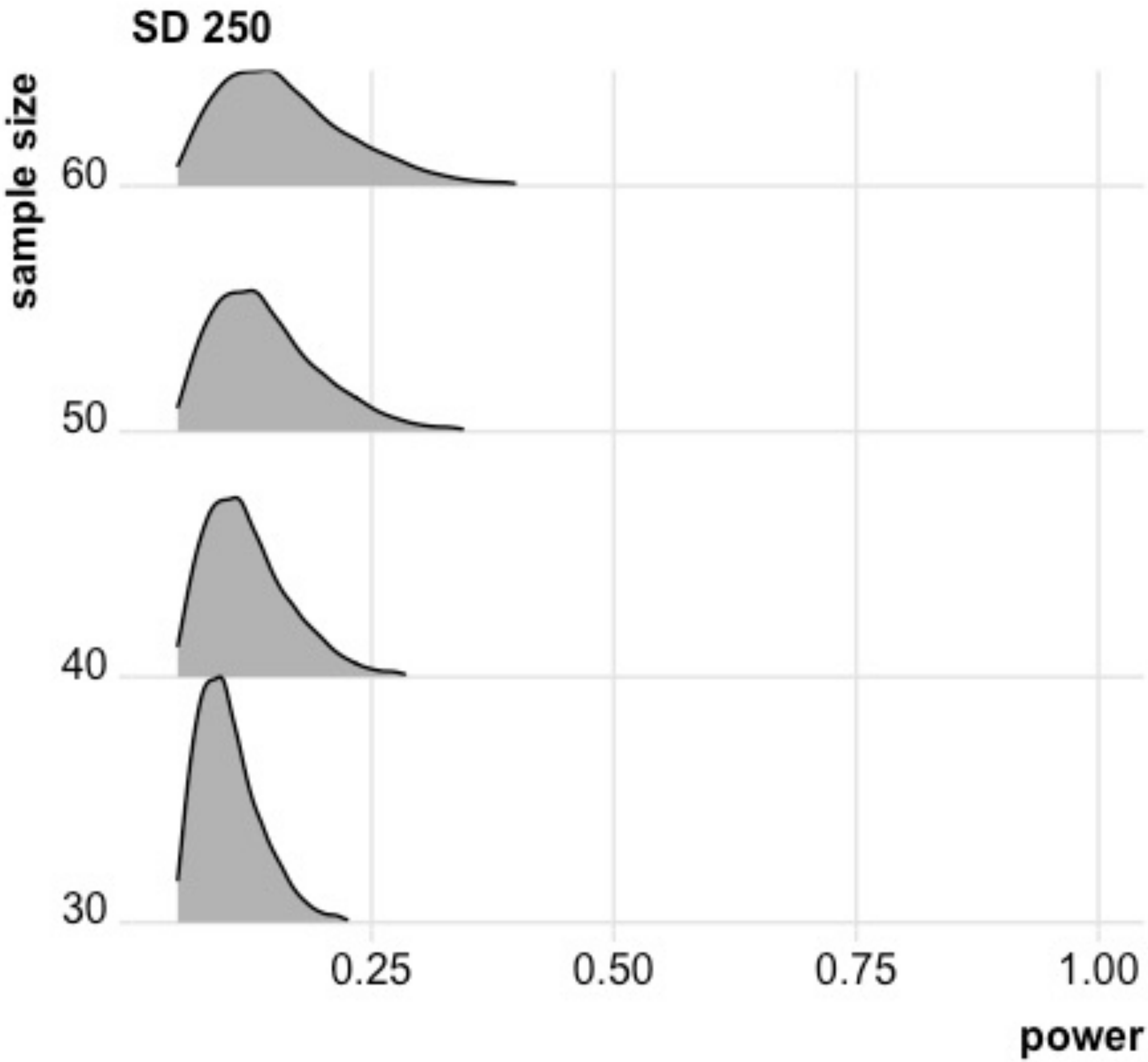
a is read faster than b at *were*



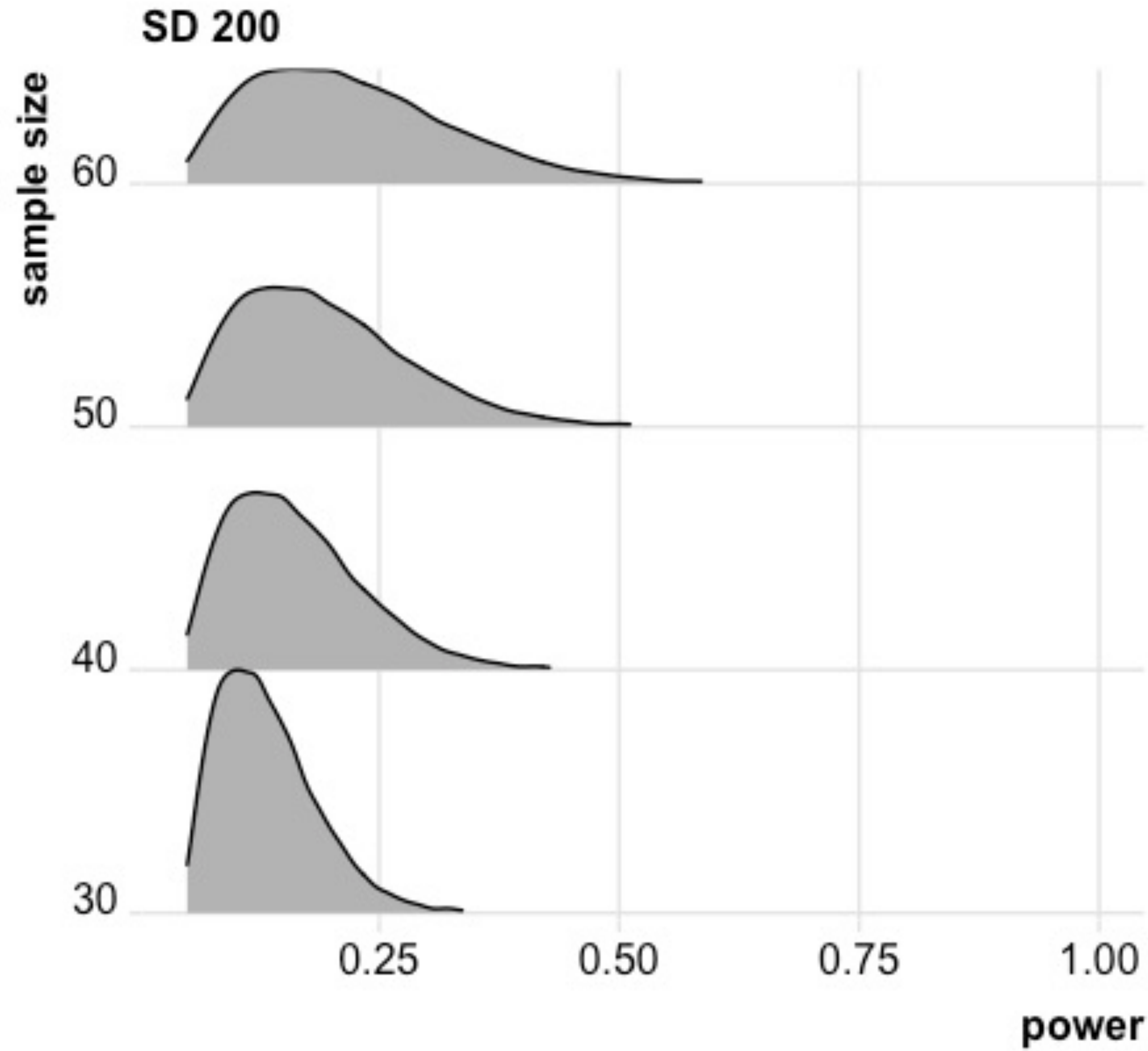
Power in reading studies on agreement attraction



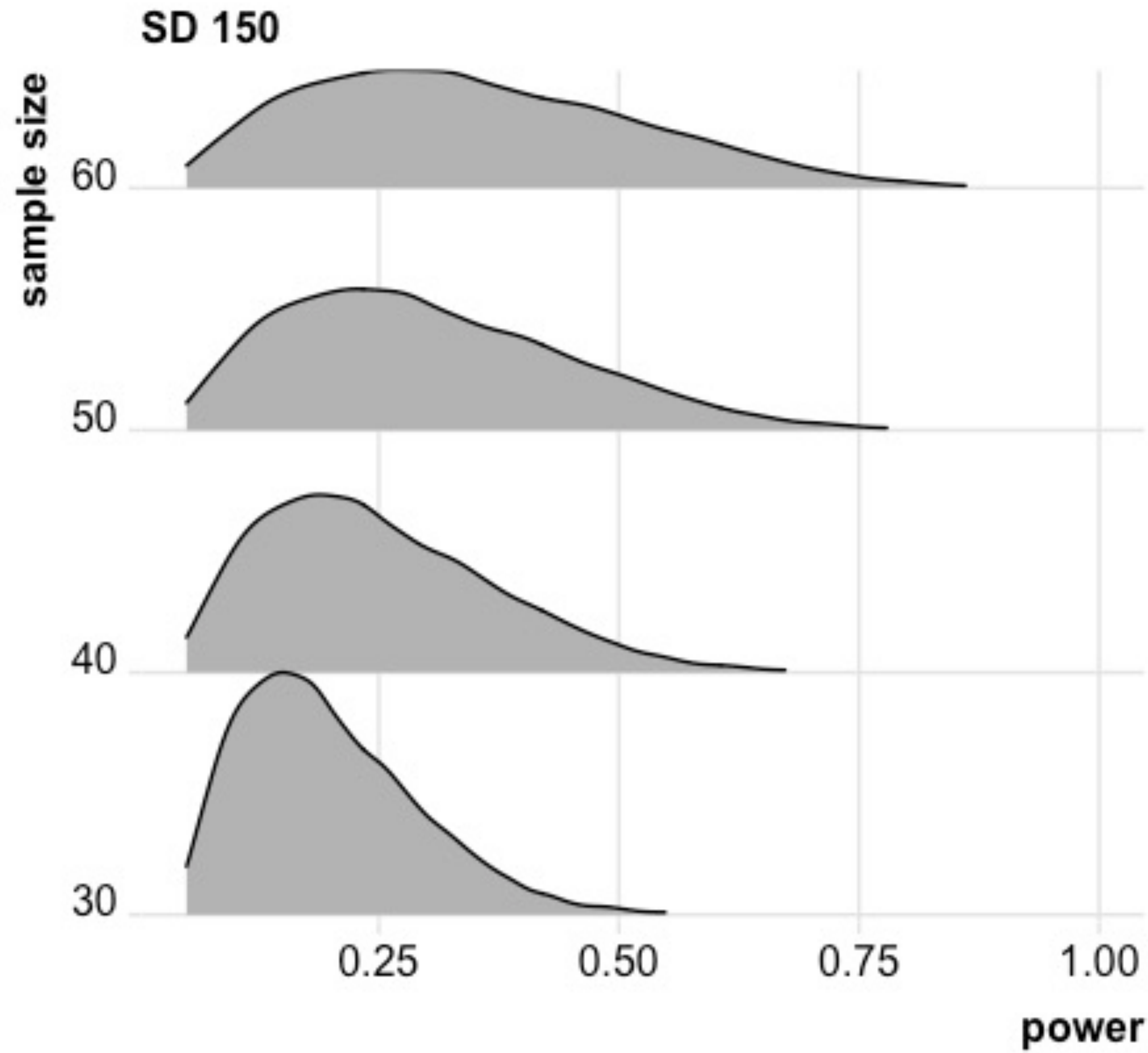
Power in reading studies on agreement attraction



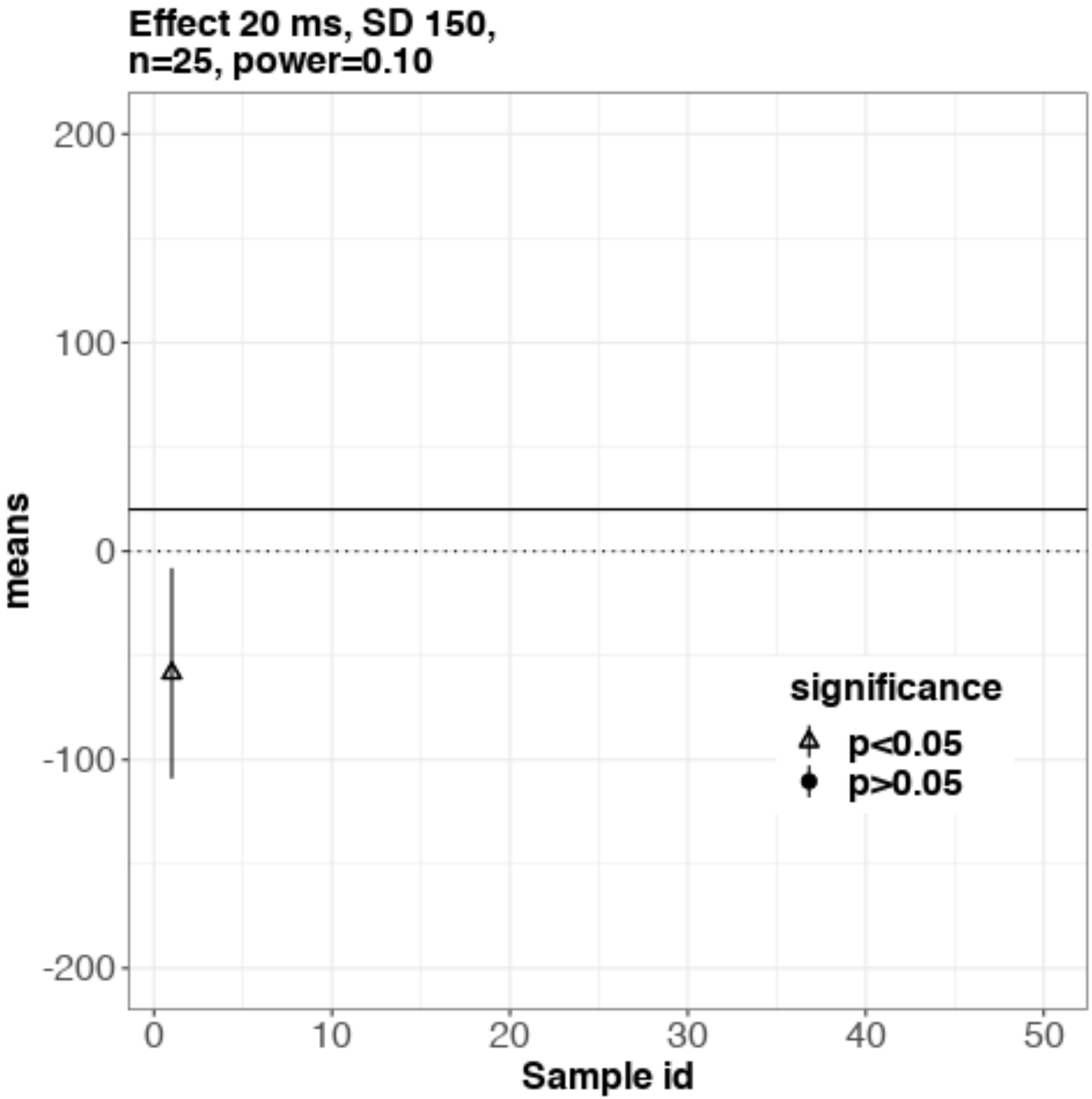
Power in reading studies on agreement attraction



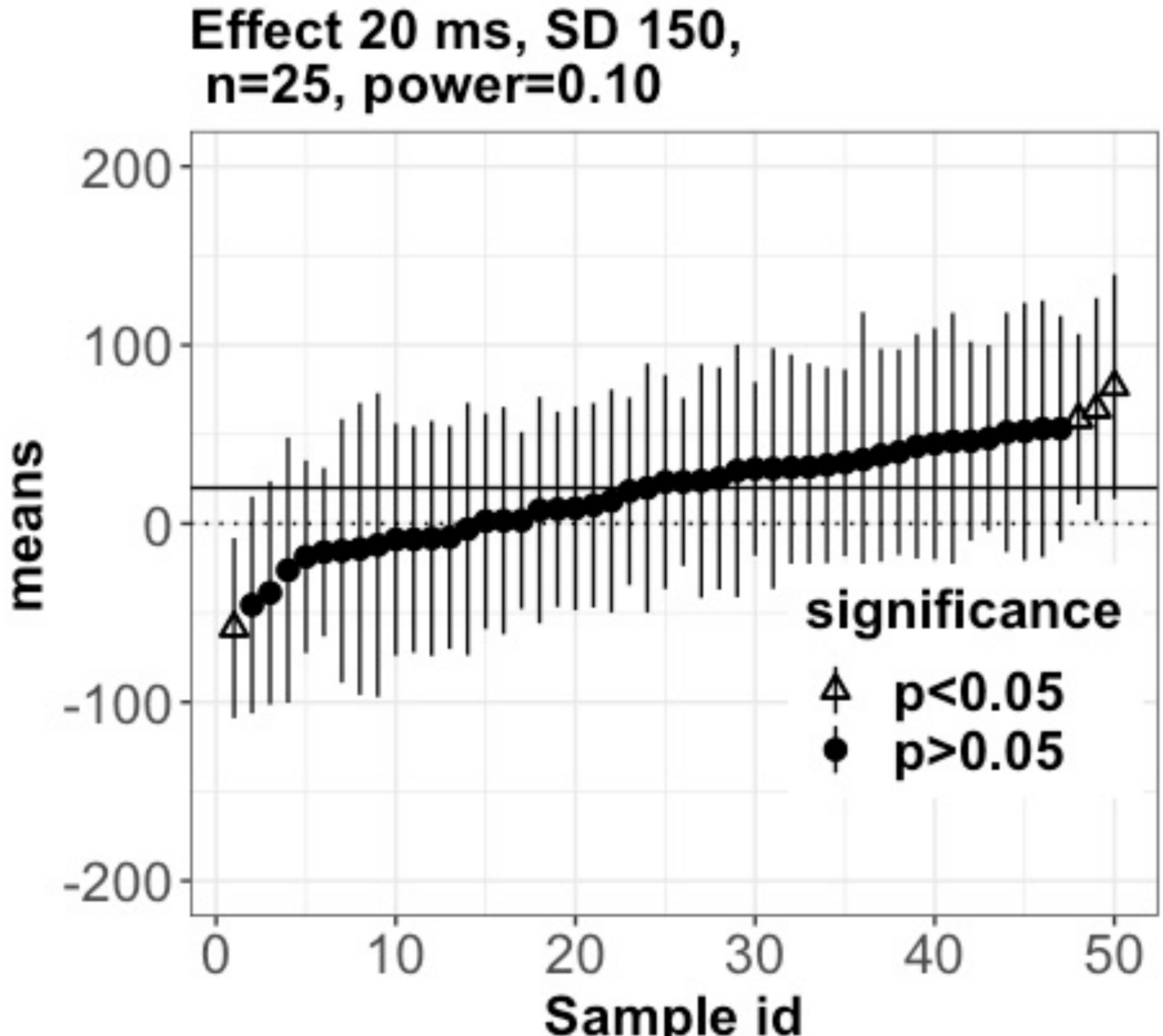
Power in reading studies on agreement attraction



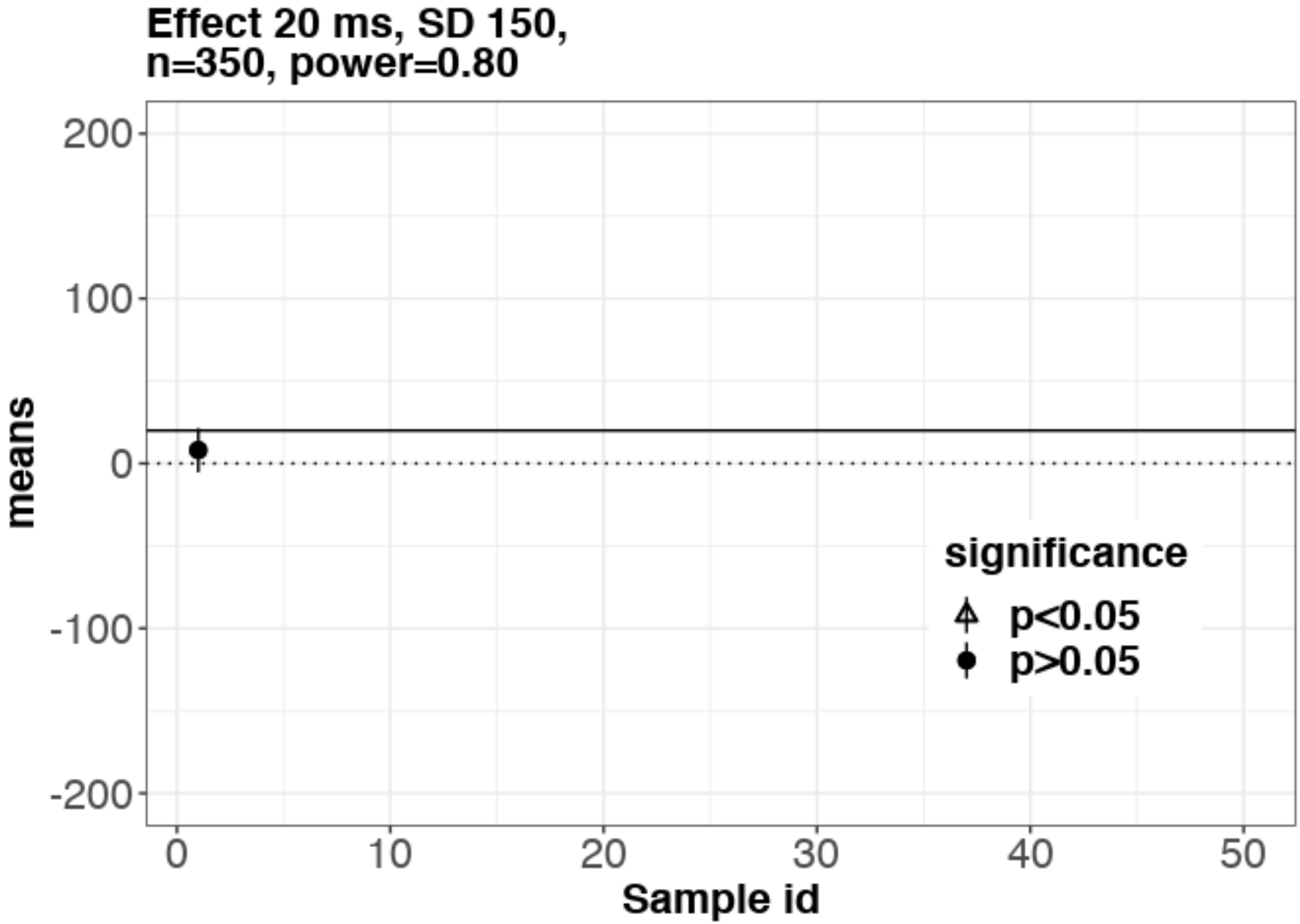
Low power leads to exaggerated estimates: Type M error



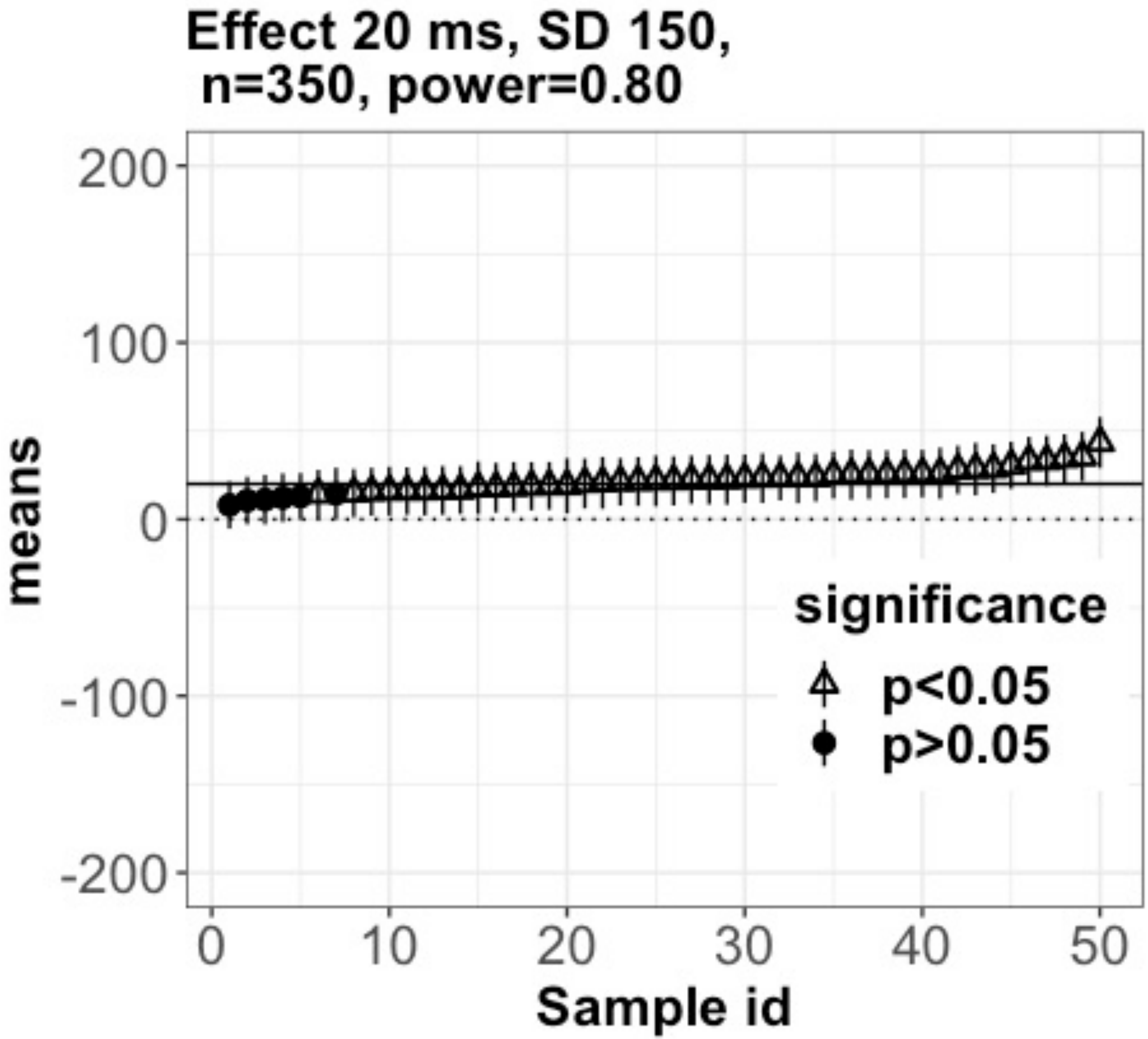
Low power leads to exaggerated estimates: Type M error



Compare with a high power situation



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

The Bayesian approach

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

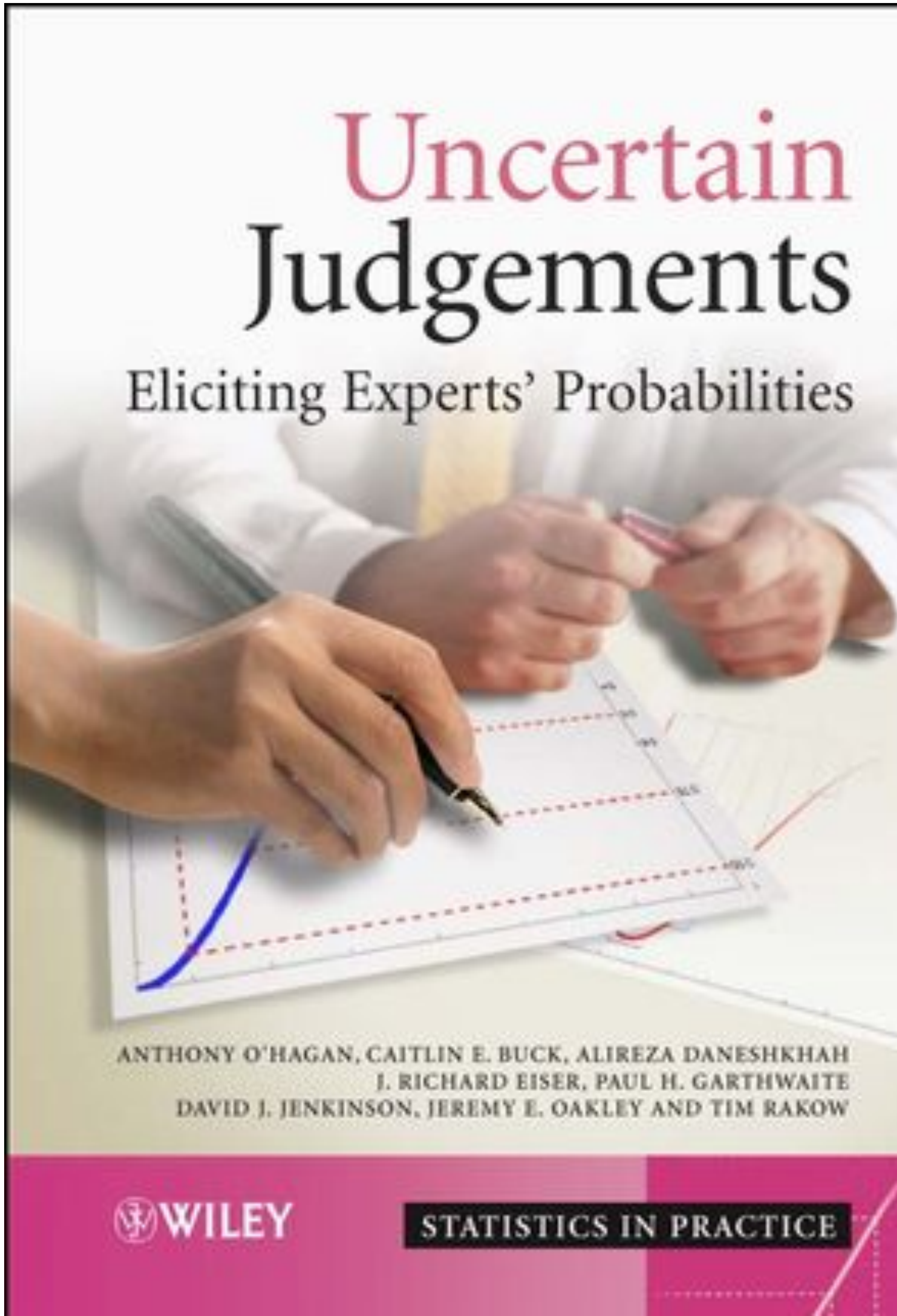
$$X \sim Normal(\mu, \sigma)$$

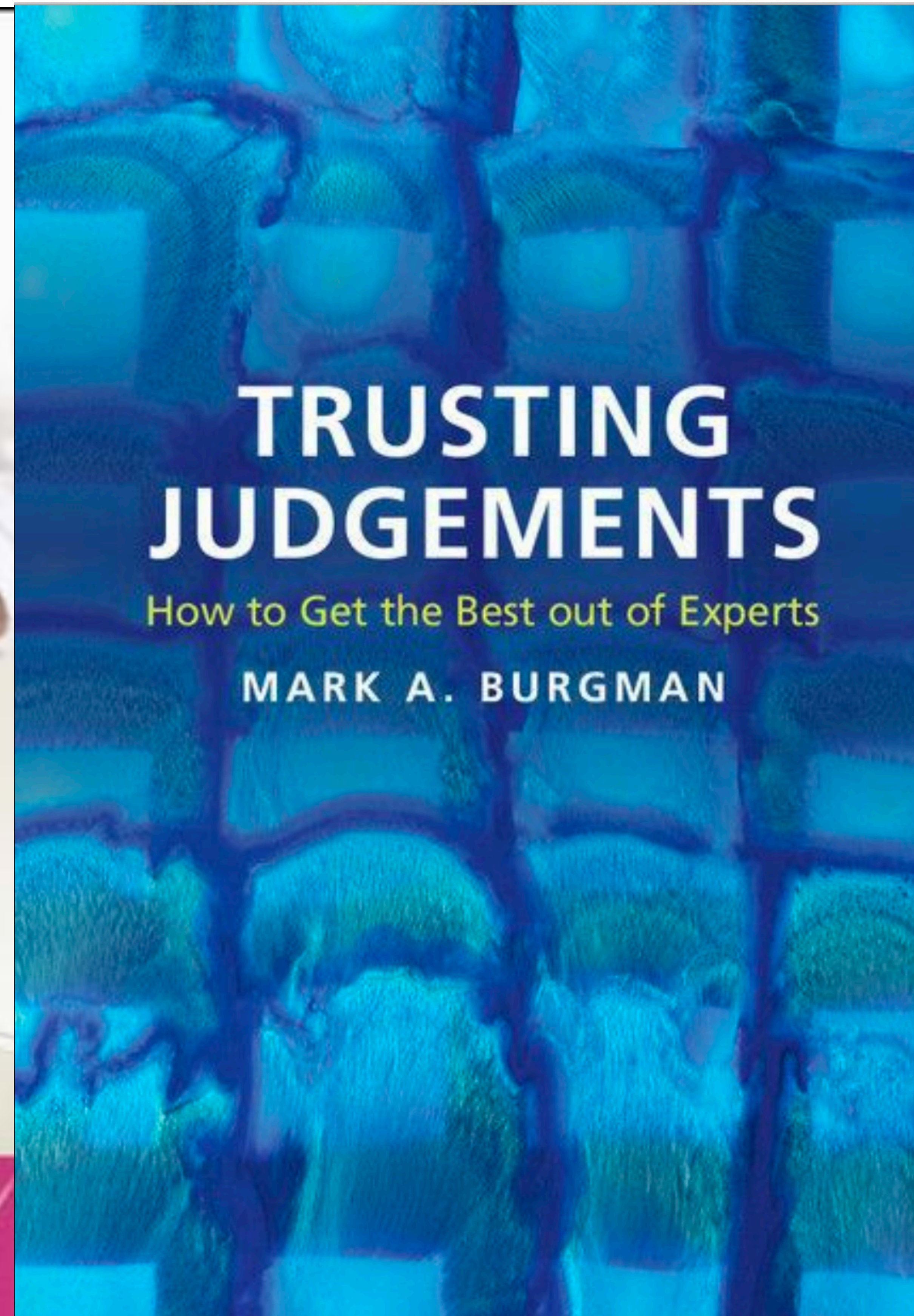
1. Define **prior distributions** for the parameters (here, μ, σ)
2. Derive **posterior distributions** of the parameters of interest using Bayes' rule:

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

3. Carry out inference based on the posterior

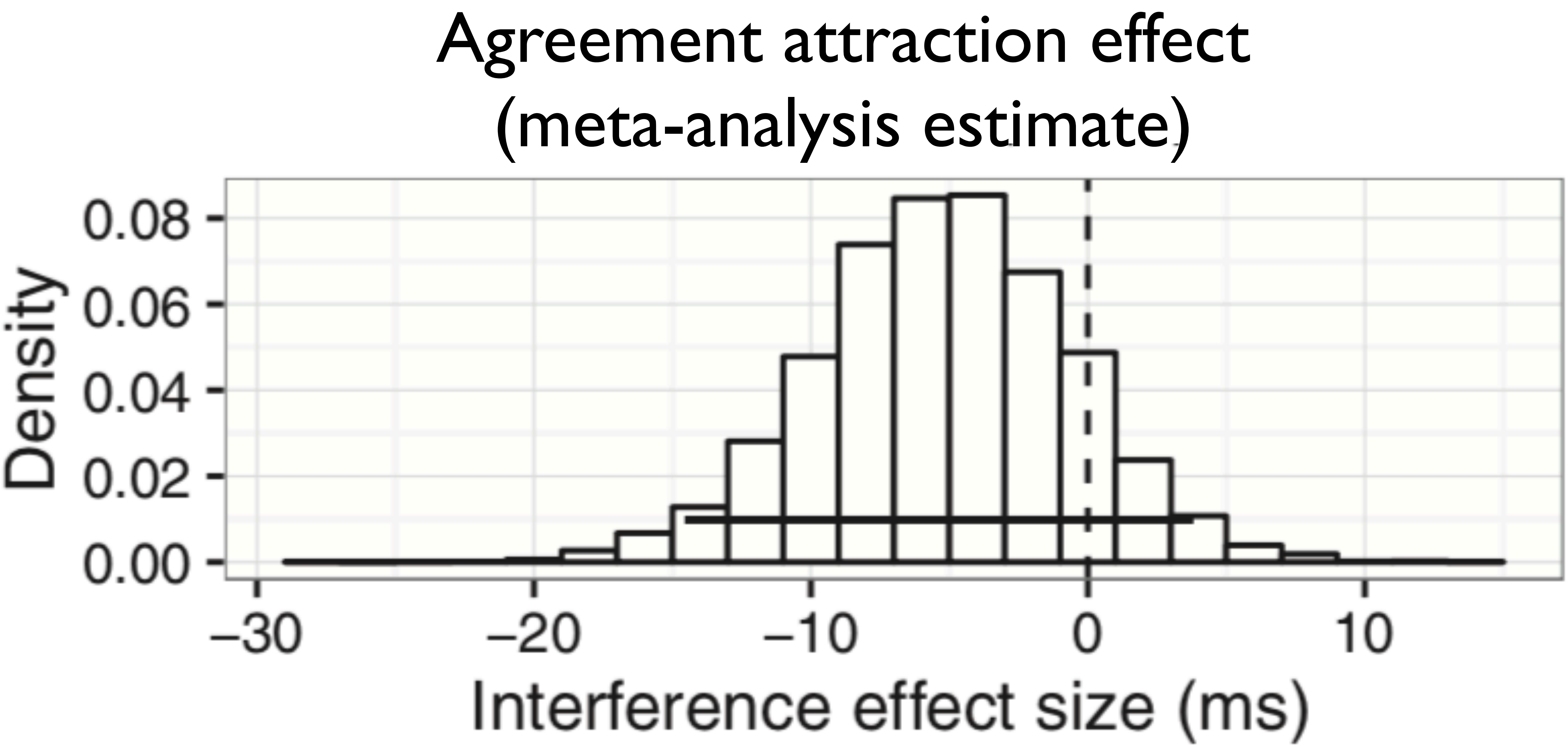
Deriving prior distributions requires domain knowledge





The Bayesian approach

The end result of a Bayesian analysis is a posterior distribution of the parameter of interest



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

* Random variables

** Type I, II error could be seen as relevant for Bayes

*** Confidence intervals can be a proxy

ANOVA vs Bayes factors

$$BF_{12} = \frac{Likelihood(Data \mid Model_1)}{Likelihood(Data \mid Model_2)}$$

The Bayes factor is similar to the frequentist likelihood ratio test (or ANOVA), with the difference that in the Bayes factor, the likelihood is integrated over the parameter space.

The BF can be highly sensitive to the priors.
A Bayes factor analysis must come with a sensitivity analysis.

Why is the Bayesian approach useful?

1. Handles sparse data without any problems
2. Highly customised models can be defined
3. The focus is on **uncertainty quantification**

But Bayes comes with a cost

1. You have to think about your prior knowledge/belief

2. There is no one answer corresponding to $p<0.05$ or $p>0.05$.

3. You have to learn to think about **uncertainty**:
   Compare:
   "50% probability of rain tomorrow"
   "95% sure that probability for rain is between 40-60%"
   "95% sure that probability for rain is between 5-95%"

Summarizing the Bayesian/frequentist divide

“[Bayesian data analysis] is a method for summarizing **uncertainty** and making estimates and predictions using probability statements conditional on observed data and an assumed model.

“Frequentist statistics … is an approach for evaluating statistical **procedures** conditional on some family of posited probability models.”

Gelman, 2008. Rejoinder. *Bayesian Analysis*.