

Review

> Nat Rev Neurosci. 2013 May;14(5):365–76. doi: 10.1038/nrn3475. Epub 2013 Apr 10.

# Power failure: why small sample size undermines the reliability of neuroscience

Katherine S Button <sup>1</sup>, John P A Ioannidis, Claire Mokrysz, Brian A Nosek, Jonathan Flint, Emma S J Robinson, Marcus R Munafò

Affiliations + expand

PMID: 23571845 DOI: [10.1038/nrn3475](https://doi.org/10.1038/nrn3475)

## Abstract

A study with low statistical power has a reduced chance of detecting a true effect, but it is less well appreciated that low power also reduces the likelihood that a statistically significant result reflects a true effect. Here, we show that the average statistical power of studies in the

<https://www.nature.com/articles/nrn3475>

- Low statistical power undermines the purpose of scientific research; it reduces the chance of detecting a true effect.
- Perhaps less intuitively, low power also reduces the likelihood that a statistically significant result reflects a true effect.
- Empirically, we estimate the median statistical power of studies in the neurosciences is between ~8% and ~31%.

**\*data from 49 meta-analyses and 730 individual primary studies**

<https://www.nature.com/articles/nrn3475>



► [Front Psychol.](#) 2019 Apr 11;10:813. doi: [10.3389/fpsyg.2019.00813](https://doi.org/10.3389/fpsyg.2019.00813) [↗](#)

## **The Meaningfulness of Effect Sizes in Psychological Research: Differences Between Sub-Disciplines and the Impact of Potential Biases**

[Thomas Schäfer](#)<sup>1,\*</sup>, [Marcus A Schwarz](#)<sup>1</sup>

► [Author information](#) ► [Article notes](#) ► [Copyright and License information](#)

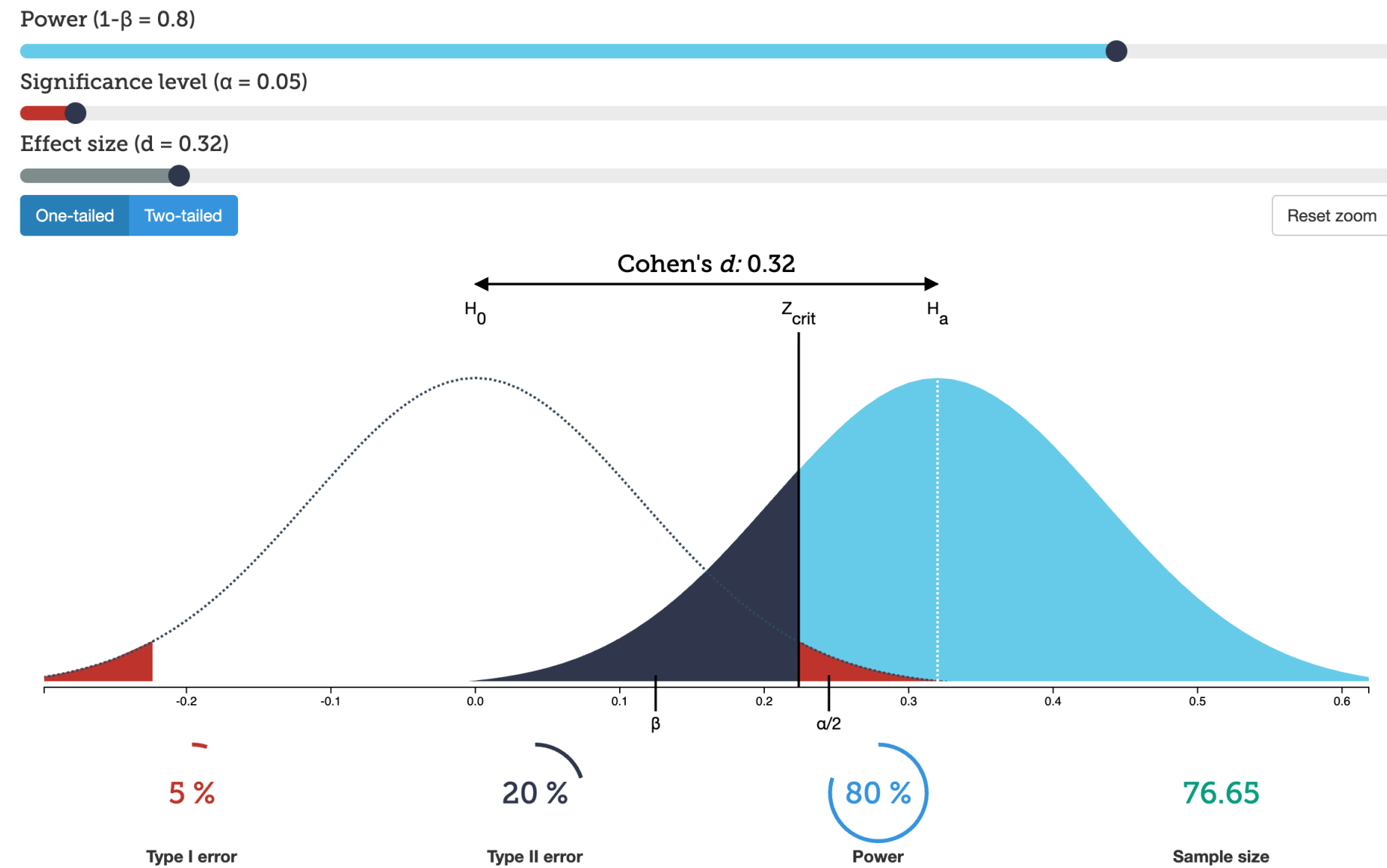
PMCID: PMC6470248 PMID: [31031679](#)

<https://pmc.ncbi.nlm.nih.gov/articles/PMC6470248/>

From past publications without pre-registration, 900 effects were randomly drawn and compared with 93 effects from publications with pre-registration, revealing a large difference: Effects from the former (median  $r = 0.36$ ) were much larger than effects from the latter (median  $r = 0.16$ ).

**\*  $r = 0.16$  entspricht einem  $d$  von ca. 0.324**

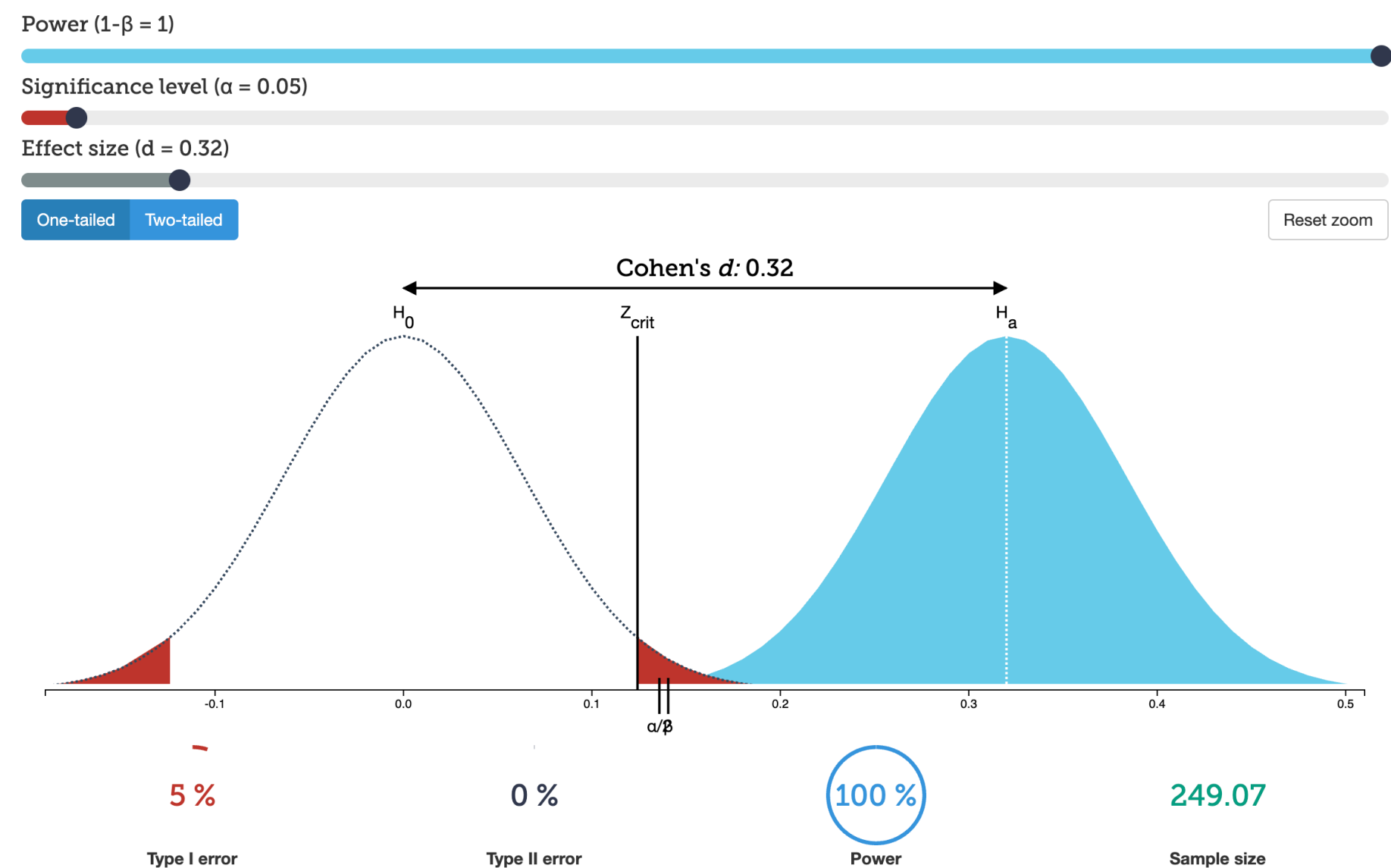
<https://www.psychometrica.de/effektstaerke.html>



## Power ( $1-\beta = 0.8$ )

Wir akzeptieren eine Wahrscheinlichkeit von 20% für einen Typ 2 Fehler (wir lehnen eine wahre Alternativhypothese ab und nehmen die  $H_0$  an)

Bei  $d=0.32$  und  $\alpha = 0.05$  benötigen wir **77 Versuchspersonen**



## Power ( $1-\beta = 1$ )

Wir wollen uns „sicher“ sein einen echten Effekt zu erkennen, wenn dieser vorliegt.

Bei  $d=0.32$  und  $\alpha = 0.05$  benötigen wir **249 Versuchspersonen**

Original Contribution

FREE

## Characteristics of Clinical Trials Registered in ClinicalTrials.gov, 2007-2010

Robert M. Califf, MD; Deborah A. Zarin, MD; Judith M. Kramer, MD, MS ; [et al](#)

» [Author Affiliations](#) | [Article Information](#)

☰ [RELATED ARTICLES](#)    [FIGURES](#)    [SUPPLEMENTAL CONTENT](#)

Most interventional trials registered between 2007 and 2010 were small, with 62% enrolling 100 or fewer participants.

The median number of participants per trial was 58 (IQR, 27-161) for completed trials and 70 (IQR, 35-190) for trials that have been registered but not completed.

<https://jamanetwork.com/journals/jama/fullarticle/1150093>





# Forschungsfrage(n)

1. Wie häufig werden in **psychologischen** Forschungsartikeln Poweranalysen berichtet, und wie oft sind diese nachvollziehbar dokumentiert (Effektgröße,  $\alpha$ , Power, Testtyp)?
2. Wie oft sind veröffentlichte Studien „underpowered“ in Bezug auf ihre Hauptanalysen?
3. In präregistrierten Studien: Wie stark weichen tatsächlich erhobene Stichproben von der geplanten Poweranalyse ab?



# Skizzierte Vorgehensweise

1. Ausarbeitung & Preregistration (z.B. bei [aspredicted.org](https://aspredicted.org))
2. Bereitstellung von Analyse Skripten und erhobenen Daten via freizugänglichem Repo (inkl. Beschreibung & Anleitung)
3. Feingliedrige Dokumentation bei der Auswahl der Studien und gut strukturiertes Script mit nachvollziehbaren Kommentaren (ggf. Verwendung von RMarkdown/Notebook o.ä.)
4. Veröffentlichung von Preprint z.B. bei OSF Preprints (durch doi / link einfach shareable)
5. Open Peer Review z.B. durch Seminarteilnehmer\*innen oder Kommilitonen mittels Plattform wie z.B. <https://openreview.net/>
6. Ggf. Einreichung bei geeignetem Journal

# Aktuell offene Fragen

- Stichprobe (Auswahl der Paper)
  - Größe / Anzahl
  - Manuell selektiert oder bspw. automatisch (data scraping)
  - Bereich (übergeordnet Psychologie oder spezifischer Bereich(e))
- Recherche: Was wurde bisher hierzu publiziert? Gibt es Lücken?
- Welche Variablen sollen erhoben werden
  - z.B. Poweranalyse vorhanden (ja/nein), Parameter (Effektgröße,  $\alpha$ , Power, Testtyp) angegeben (ja/nein), Geplante Stichprobengröße vs. tatsächliche N, Hauptbefund signifikant (ja/nein), Präregistrierung vorhanden (ja/nein)
- Ggf. Spezifizierung Forschungsfragen
- Formulierung von Hypothesen
- ....