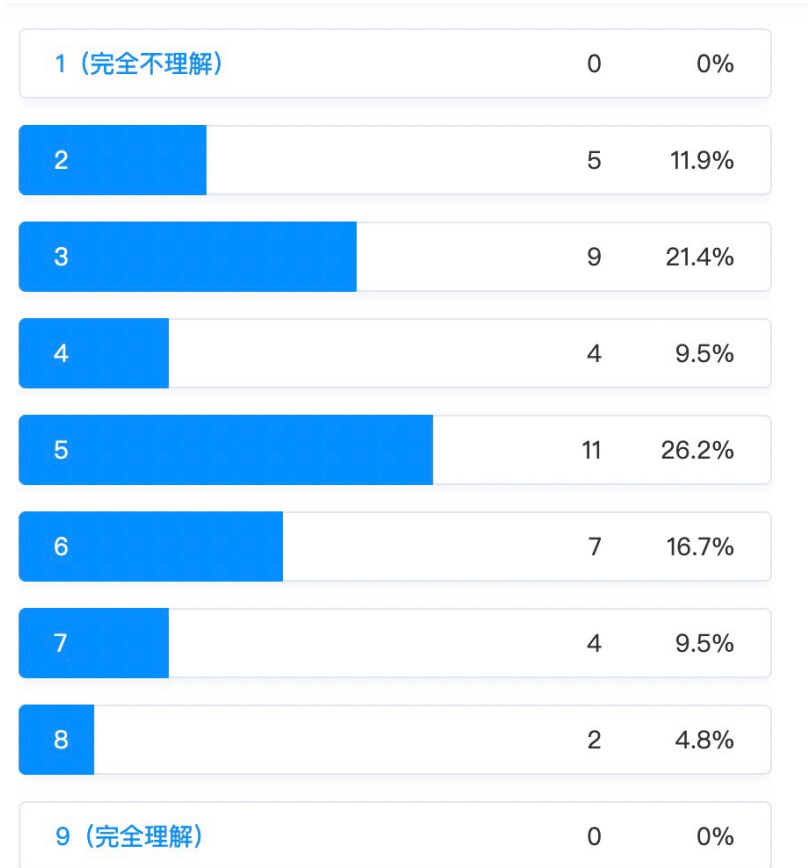
Statistical Power

---- A non-technical view

Hu Chuan-Peng (胡传鹏)
Nanjing Normal University
hcp4715@hotmail.com

Chinese Open Science Network
COSN 1st COSN hackathon
May the POWER be with you!
Aug. 27th-28th 2022, Online



N = 42

- NHST下统计检验力分析中的四个变量
- 模型思维视角下的power analysis
- 历史与现实中的power analysis
- 替代的方案？

1. NHST框架下的Power analysis

https://rpsychologist.com/d3/nhst/

Power contour estimation

Maximum sample size

10 | 200 | 1,000

10 110 210 310 410 510 610 710 810 910 1,000

Maximum number of trials

10 | 100 | 1,000

10 110 210 310 410 510 610 710 810 910 1,000

Alpha level

0.05

Mean difference

1

Within-subject SD

10

Between-subject SD

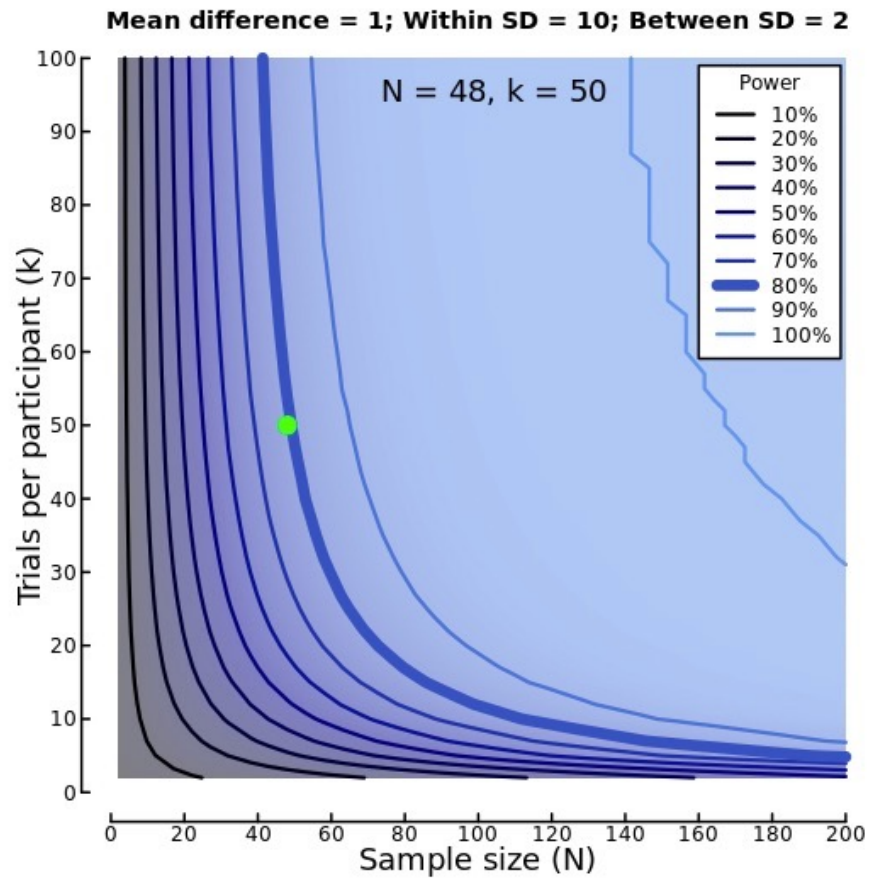
2

☑ Estimate optimal point?

Recruitment cost (trials)

0 | 100 | 1,000

0 100 200 300 400 500 600 700 800 900 1,000



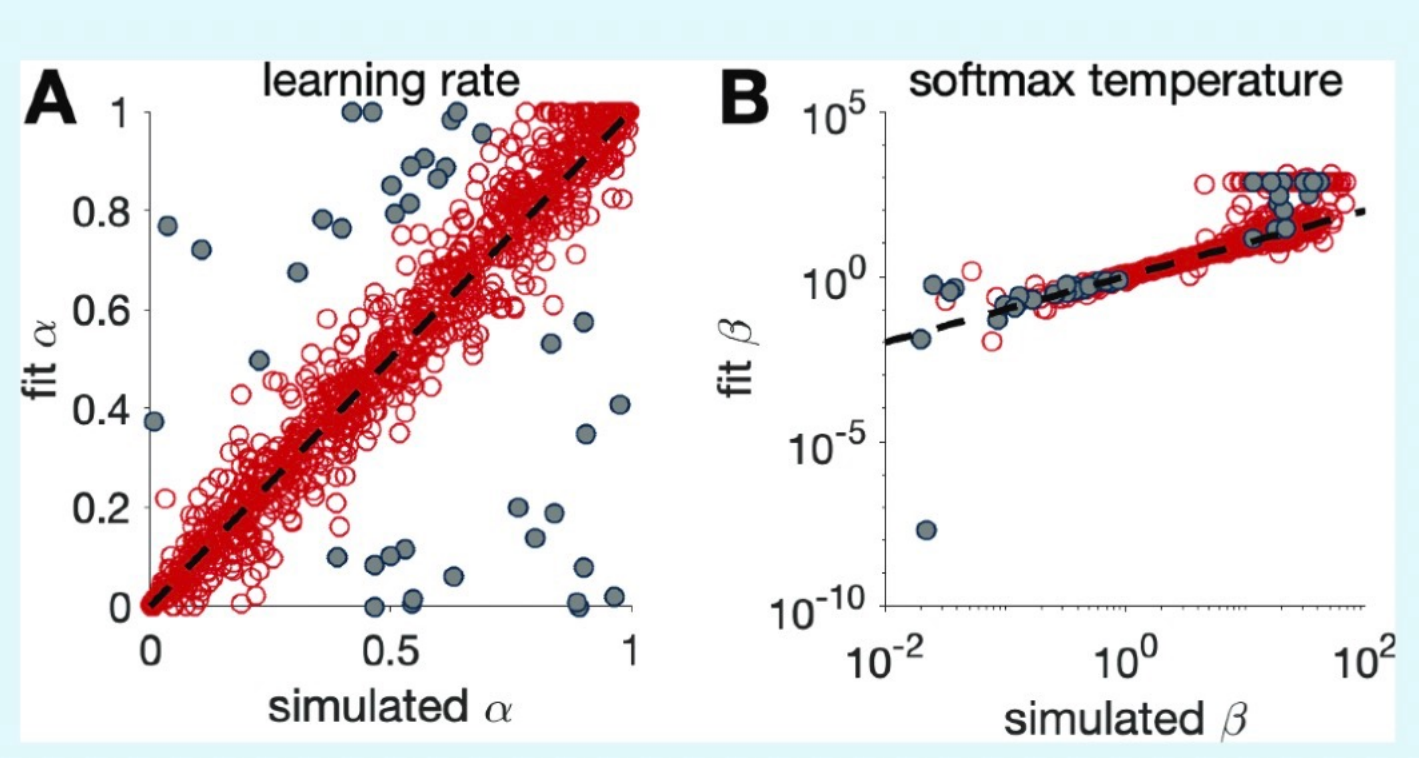
https://shiny.york.ac.uk/powercontours/; Baker et al., 2021, Psych Meth, https://doi.org/10.1037/met0000337

2. 模型思路下的Power analysis

- NHST本身是一个统计模型

2. 模型思路下的Power analysis

- Parameter recovery



Wilson & Collins, 2019, elife. Box 4-figure 1

3. 历史中的Power analysis

Journal of Abnormal and Social Psychology
1962, Vol. 65, No. 3, 145–153

THE STATISTICAL POWER OF ABNORMAL–SOCIAL PSYCHOLOGICAL RESEARCH:

A REVIEW [1]

JACOB COHEN

New York University

TABLE 2

VALUES OF POPULATION PARAMETERS WHICH DEFINE THE LEVELS OF SIZE OF EFFECT FOR THE VARIOUS STATISTICAL TESTS

Test	Population parameter	Values		
		Small	Medium	Large
1. t (two means are equal)	$\lvert M_1 - M_2 \rvert / \sigma$	.25	.50	1.00
2. Normal (two proportions are equal)	$\lvert P_1 - P_2 \rvert$	.10	.20	.30
3. Normal (two r's are equal)	$\lvert r_1 - r_2 \rvert$	.10	.20	.30
4. t ($r = 0$)	$\lvert r \rvert$	.20	.40	.60
5. Sign test	$\lvert P - .50 \rvert$	.10	.20	.30
6. F (k means are equal)	$\sigma_{M_i} / \sigma = f$	.125	.25	.50
7a. χ^2 (k proportions are equal)	Ratio: $\frac{\text{Largest } P}{\text{Smallest } P}$	3:2	2:1	4:1
7b. χ^2 (contingency test)	$\sum_{i=1}^{kr} \frac{(P_{0i} - P_{1i})^2}{P_{0i}} = l$	Varies with table size, but uses criteria equivalent, for equal degrees of freedom, to 7a (see text).		

3. 历史中的Power analysis

Journal of Abnormal and Social Psychology
1962, Vol. 65, No. 3, 145–153

TABLE 2

VALUES OF POPULATION PARAMETERS WHICH DEFINE THE LEVELS OF SIZE OF EFFECT FOR THE VARIOUS STATISTICAL TESTS

Test	Population parameter	Values		
		Small	Medium	Large
1. t (two means are equal)	$\lvert M_1 - M_2 \rvert / \sigma$	.25	.50	1.00
2. Normal (two proportions are equal)	$\lvert P_1 - P_2 \rvert$	.10	.20	.30
3. Normal (two r's are equal)	$\lvert r_1 - r_2 \rvert$	.10	.20	.30
4. t ($r = 0$)	$\lvert r \rvert$			
5. Sign test	$\lvert P - .50 \rvert$			
6. F (k means are equal)	$\sigma_{M_i} / \sigma = f$			
7a. χ^2 (k proportions are equal)	Ratio: $\frac{\text{Largest } P}{\text{Smallest } P}$			
7b. χ^2 (contingency test)	$\sum_{i=1}^{kr} \frac{(P_{0i} - P_{1i})^2}{P_{0i}} = l$			

Trends in Cognitive Sciences

Supports open access

OPINION | VOLUME 24, ISSUE 3, P200-207, MARCH 01, 2020

Avoid Cohen's 'Small', 'Medium', and 'Large' for Power Analysis

Joshua Correll • Christopher Mellinger • Gary H. McClelland • Charles M. Judd

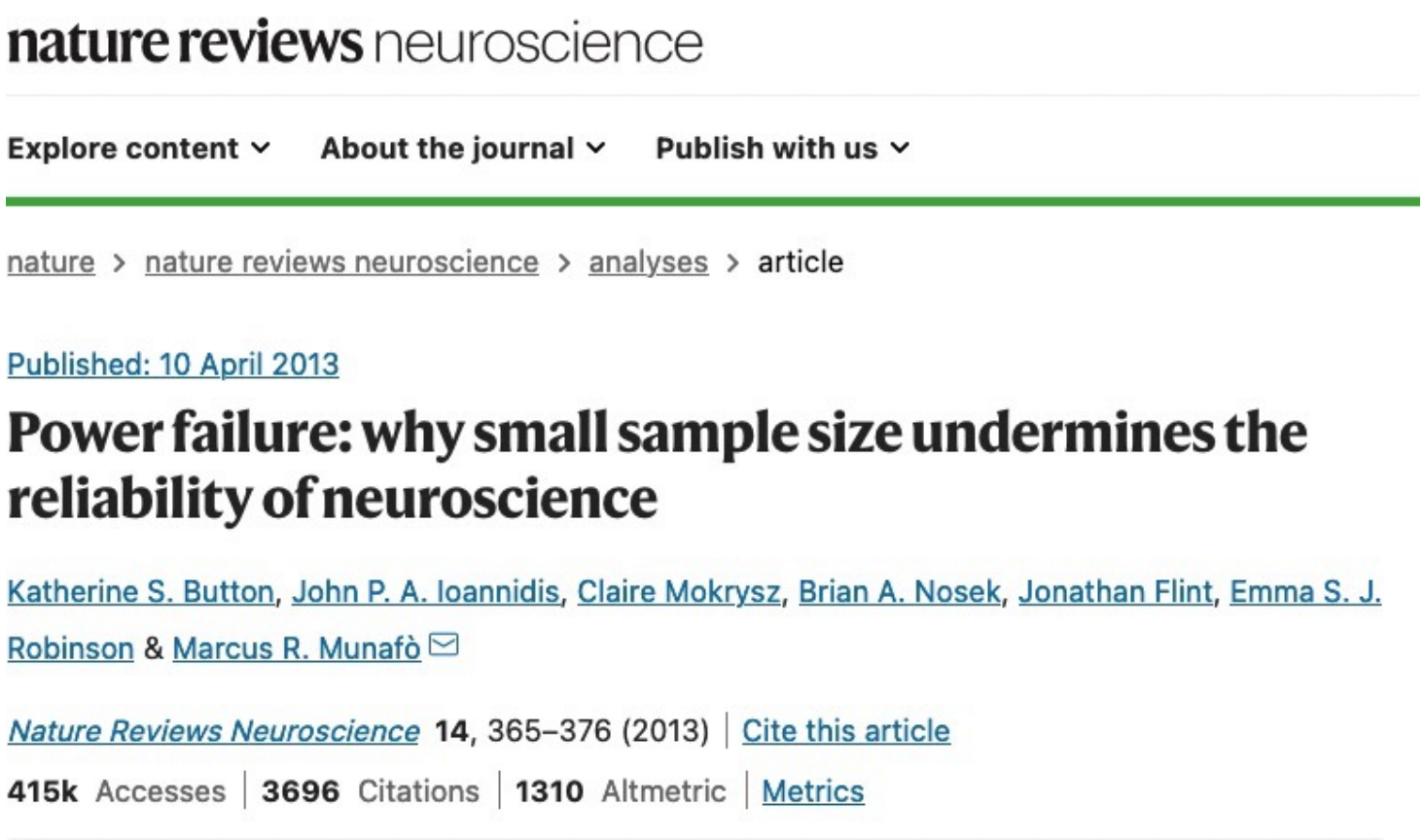
Published: January 15, 2020 • DOI: https://doi.org/10.1016/j.tics.2019.12.009 •

Check for updates

3. 历史中的Power analysis

our main focus is the desirability of achieving accurate parameter estimates, either instead of or in addition to obtaining sufficient power.

3. 历史中的Power analysis



nature reviews neuroscience

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › nature reviews neuroscience › analyses › article

Published: 10 April 2013

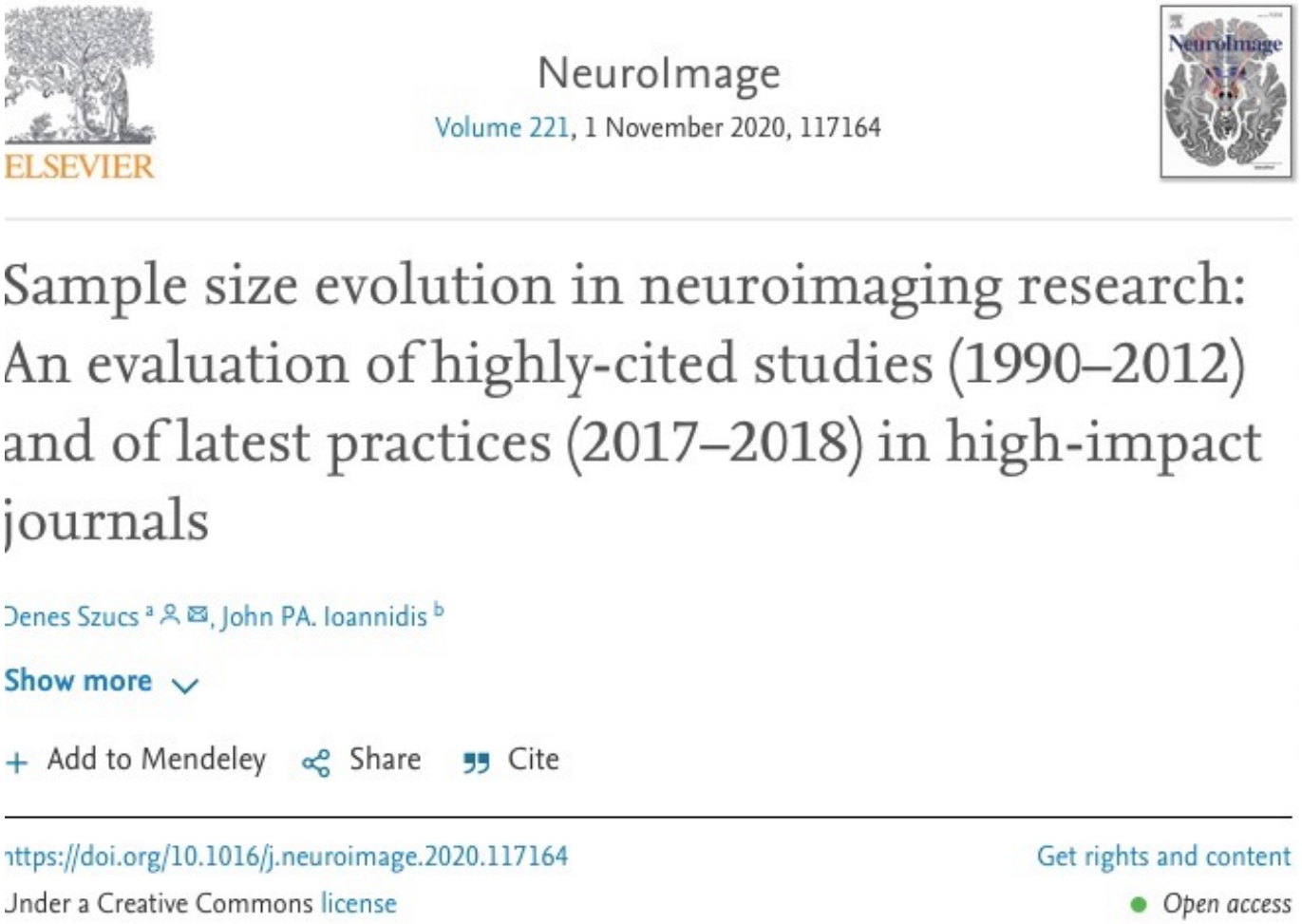
Power failure: why small sample size undermines the reliability of neuroscience

Katherine S. Button, John P. A. Ioannidis, Claire Mokrysz, Brian A. Nosek, Jonathan Flint, Emma S. J. Robinson & Marcus R. Munafò ✉

Nature Reviews Neuroscience **14**, 365–376 (2013) | Cite this article

415k Accesses | **3696** Citations | **1310** Altmetric | Metrics

3. 历史中的Power analysis



NeuroImage

Volume 221, 1 November 2020, 117164

Sample size evolution in neuroimaging research: An evaluation of highly-cited studies (1990–2012) and of latest practices (2017–2018) in high-impact journals

Denes Szucs [a], John PA. Ioannidis [b]

Show more

+ Add to Mendeley Share Cite

https://doi.org/10.1016/j.neuroimage.2020.117164

Under a Creative Commons license

Get rights and content

Open access

nature

Explore content | About the journal | Publish with us

nature > articles > article

Article | Open Access | Published: 16 March 2022

Reproducible brain-wide association studies require thousands of individuals

Scott Marek, Brenden Tervo-Clemmens, ... Nico U. F. Dosenbach + Show authors

Nature **603**, 654–660 (2022) | Cite this article

44k Accesses | **73** Citations | **1441** Altmetric | Metrics

nature

Explore content ⌄ About the journal ⌄

nature > articles > article

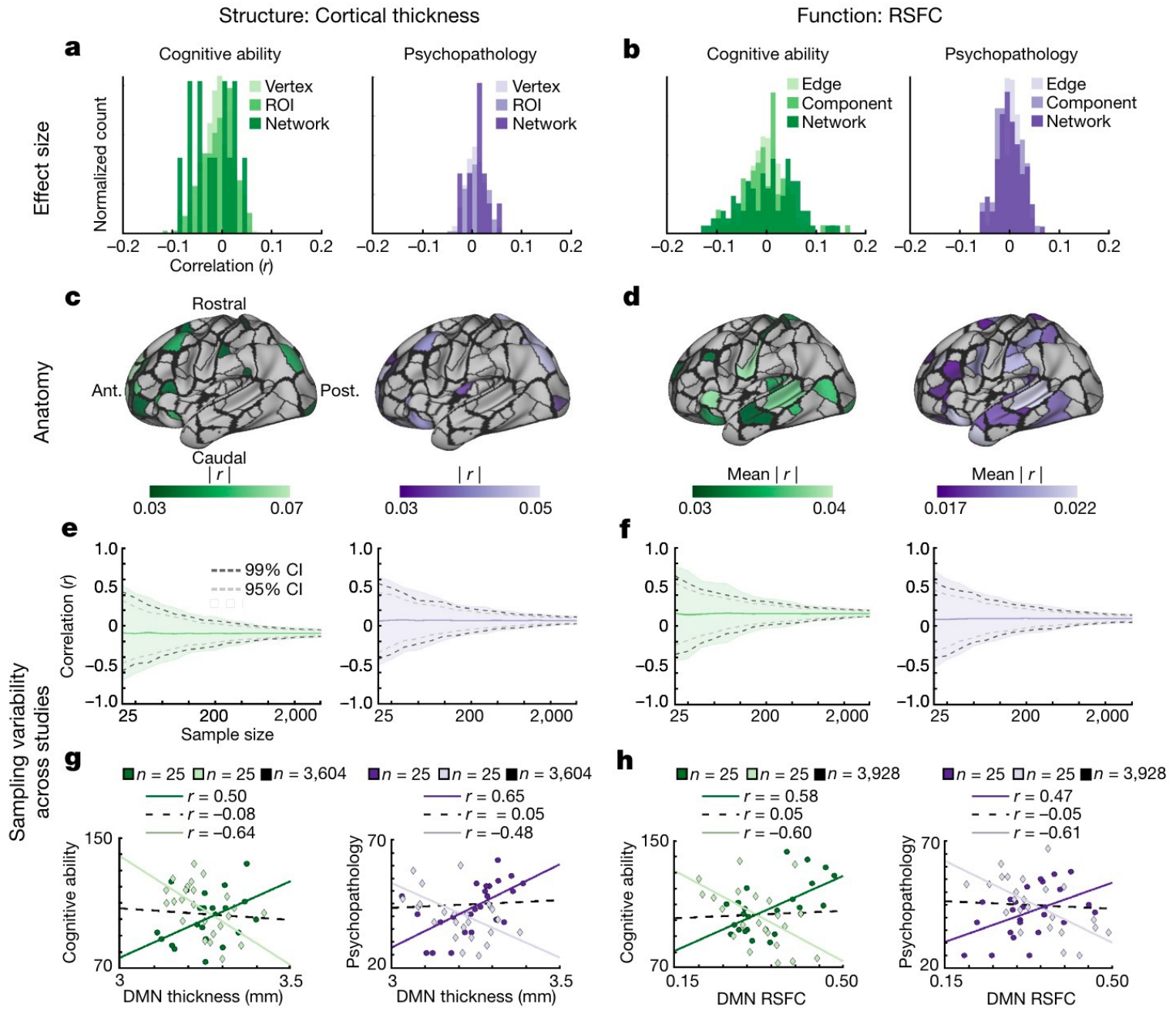
Article | Open Access | Published: 16 March

Reproducible brain-wide thousands of individuals

Scott Marek ✉, Brenden Tervo-Clemmens ✉

Nature **603**, 654–660 (2022) | Cite this arti

44k Accesses | **73** Citations | **1441** Altme



Neuron

Volume 110, Issue 9, 4 May 2022, Pages 1446-1449

Spotlight

Brain-behavior correlations: Two paths toward reliability

Caterina Gratton [1, 2], Steven M. Nelson [3, 4], Evan M. Gordon [5]

Show more

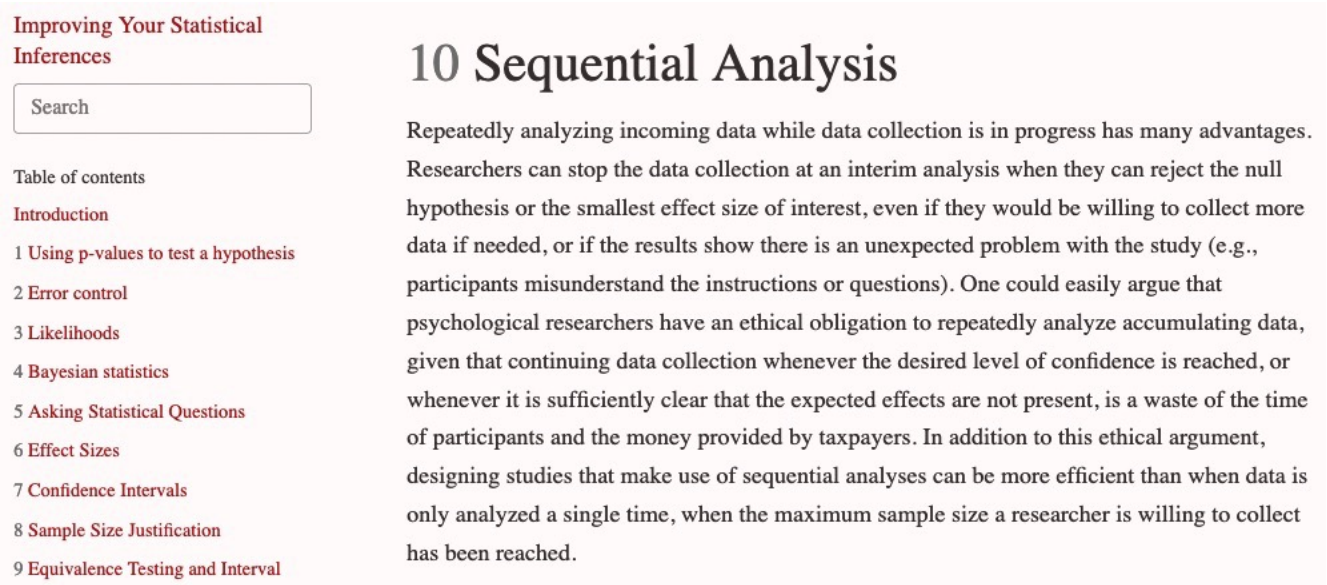
https://doi.org/10.1016/j.neuron.2022.04.018

4. 替代方案

- Sample size justification or power analysis

https://lakens.github.io/statistical_inferences/sequential.html

4. 替代方案



Improving Your Statistical Inferences

Search

Table of contents

Introduction

1 Using p-values to test a hypothesis

2 Error control

3 Likelihoods

4 Bayesian statistics

5 Asking Statistical Questions

6 Effect Sizes

7 Confidence Intervals

8 Sample Size Justification

9 Equivalence Testing and Interval

10 Sequential Analysis

Repeatedly analyzing incoming data while data collection is in progress has many advantages. Researchers can stop the data collection at an interim analysis when they can reject the null hypothesis or the smallest effect size of interest, even if they would be willing to collect more data if needed, or if the results show there is an unexpected problem with the study (e.g., participants misunderstand the instructions or questions). One could easily argue that psychological researchers have an ethical obligation to repeatedly analyze accumulating data, given that continuing data collection whenever the desired level of confidence is reached, or whenever it is sufficiently clear that the expected effects are not present, is a waste of the time of participants and the money provided by taxpayers. In addition to this ethical argument, designing studies that make use of sequential analyses can be more efficient than when data is only analyzed a single time, when the maximum sample size a researcher is willing to collect has been reached.

https://lakens.github.io/statistical_inferences/sequential.html

4. 替代方案

Psychological Methods
2017, Vol. 22, No. 2, 322–339

1082-989X/17/$12.00 http://dx.doi.org/10.1037/met0000061

Sequential Hypothesis Testing With Bayes Factors: Efficiently Testing Mean Differences

Felix D. Schönbrodt
Ludwig-Maximilians-Universität München

Eric-Jan Wagenmakers
University of Amsterdam

Michael Zehetleitner
Ludwig-Maximilians-Universität München

Marco Perugini
University of Milan–Bicocca

https://lakens.github.io/statistical_inferences/sequential.html

谢谢大家！

希望大家hack得开心！