Improving the validity and quality of our research

Daniël Lakens

Eindhoven University of Technology

@Lakens

Technische Universiteit Eindhoven University of Technology

Sample Size Planning

How do you determine the sample size for a new study?

1) It is “known” that an effect exists in the population.

2) You have the following expectation for your study:

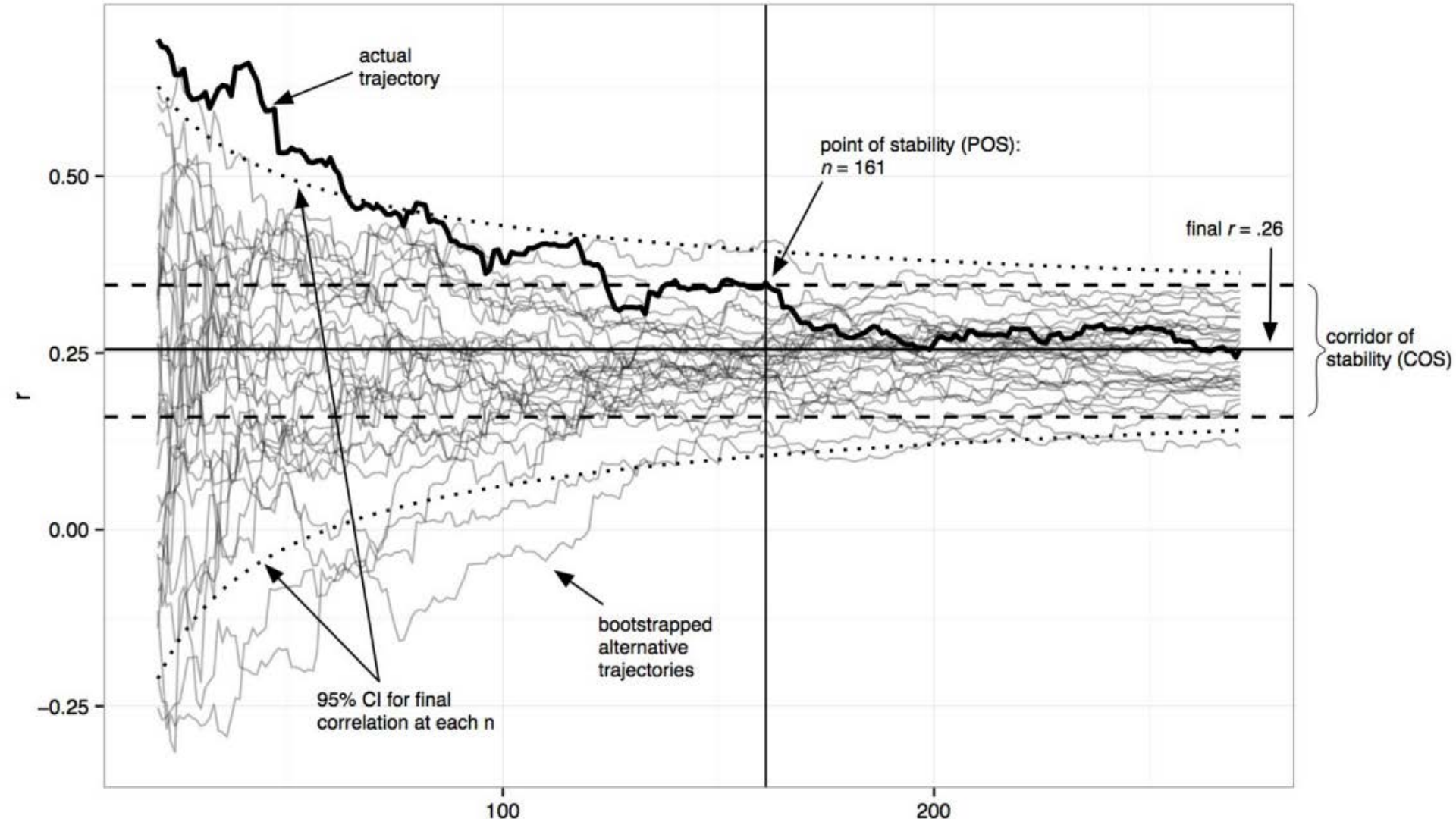
A pilot study revealed a difference between Group 1 (M = 5.68, SD = 0.98) and Group 2 (M = 6.28, SD = 1.11)

$p < .05$ (Hurray!)

You collected 22 people in one group, and 23 people in the other group. Now you set out to repeat this experiment.

What is the chance you will observe a significant effect?

Unless you aim for accuracy…



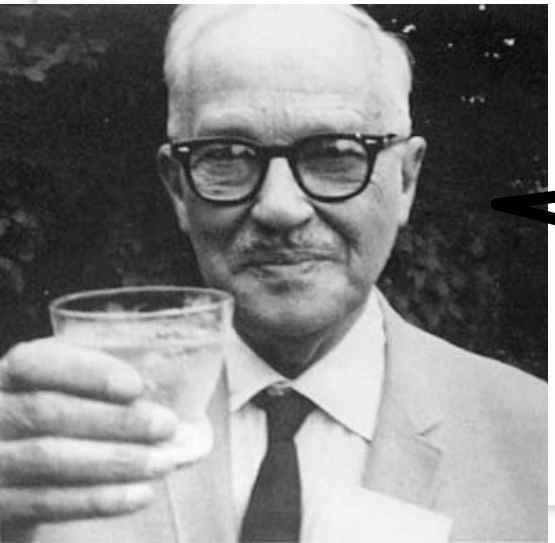
Always perform a power analysis

Main goal:

estimate the feasibility of a study

Prevent studies with low power

Power is 35% if you use 21 ppn/condition and the effect size is *d* = 0.5.



With a 65% probability of observing a False Negative, that's not what I'd call good error control!

Power Analysis

- Step 1: Determine the effect size you expect, or the Smallest Effect Size Of Interest (SESOI)

- A) Look at a meta-analysis
- B) Calculate it from a reported study
- C) Correct for bias (due to publication bias, most published effect sizes are inflated)

Calculate effect size from an article

From_R2D2: Effect size conversion spreadsheet

Indicate whether the effect size is calculated for a within or a between subjects design by choosing the correc provided in the article and press Return (check the tooltips for details). Effect sizes below the boxes that turn g me at D.Lakens@tue.nl or @Lakens. Version 1.1. Check http://osf.io/i

Within or between effect? Click on the cell to change.	Fill in all the information provided in the article									
	r	d_{pop}	n1	n2	N	t	F	df_{effect}	df_{error}	p
Between			43	55		2.34				
Never convert r from a within design to d_{pop} or d_s in a between design. See Lakens (2013) on calculating d_{av} and/or d_{rm} if you have SD's and/or the correlation between means.	Effect sizes from t-value and N for dependent t-test		Effect sizes from t-value and n1 and n2 for independent t-test		Effect sizes from t-value and N for independent t-test		Effect sizes from Cohen's d_{pop} and n1 and n2		Effect sizes from Cohen's d_{pop} (and N if known)	
	F	d_z	d_{pop}	d_s	d_{pop}	d_s	d_s	Hedges's g_s	d_s	Hedges's
			0.481272	0.476336						
	d_{effect}	d_{error}	r	r_{adj}	r	r_{adj}	r	r_{adj}	r	r_{adj}
			0.232292	0.210012						
	r	r_{adj}	η^2	CL	η^2	CL	η^2	CL	η^2	CL
			0.05396	0.633189						
	η^2		Hedges's g_s		Hedges's g_s					
			0.472605							

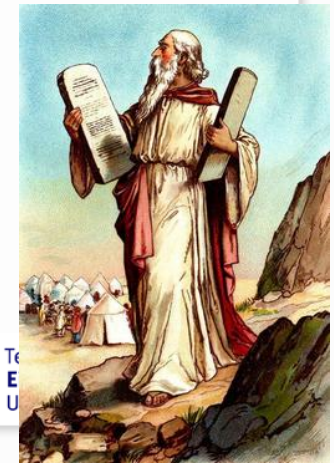
Download from https://osf.io/ixgcd/

Technische Universiteit Eindhoven University of Technology

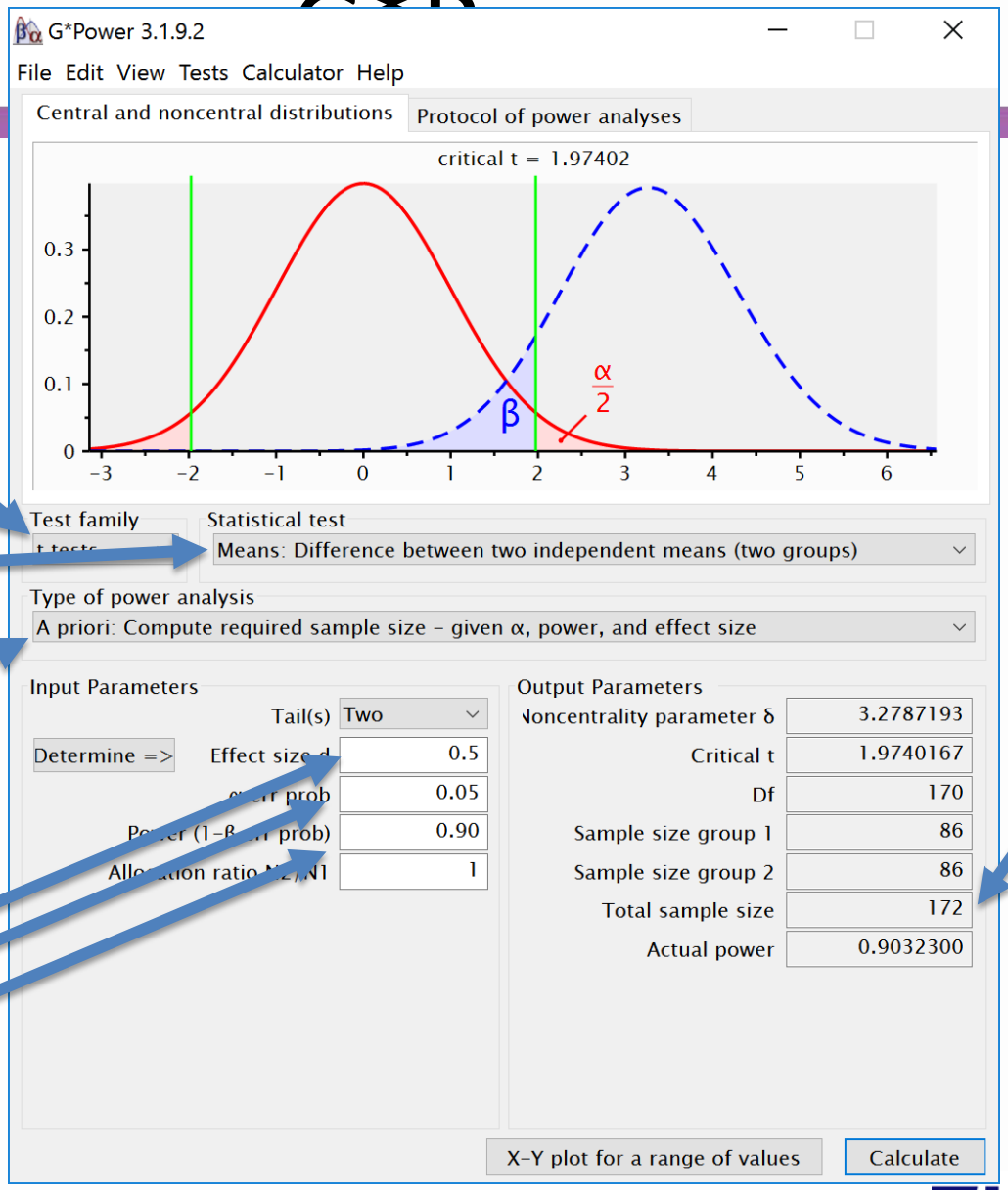
Sample Size Planning

- Power analyses provide an estimated sample size, based on the effect size, desired power, and desired alpha level (typically .05).

- You obviously can't change the alpha of 0.05, since it was one of the 10 commandments brought down from Sinai by Mozes.



G*P



Select test Family

Select specific test

Select power analysis (a-priori, sensitivity

Effect size
Alpha
Desired Power

Sample Size needed, e.g, for a medium effect (d=0.5) and 90% power

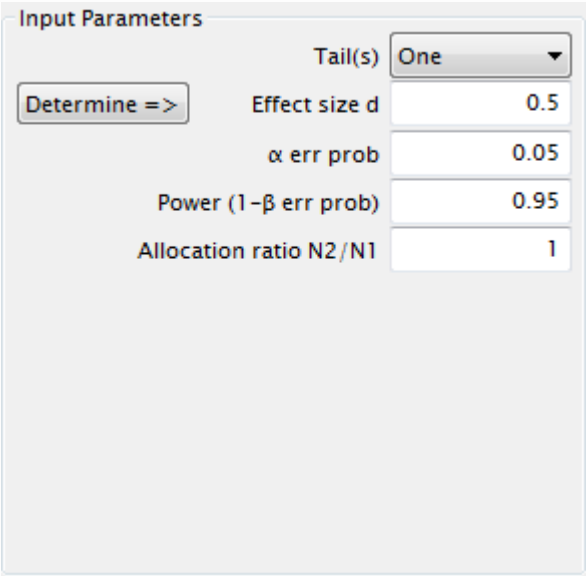
Sample Size Planning

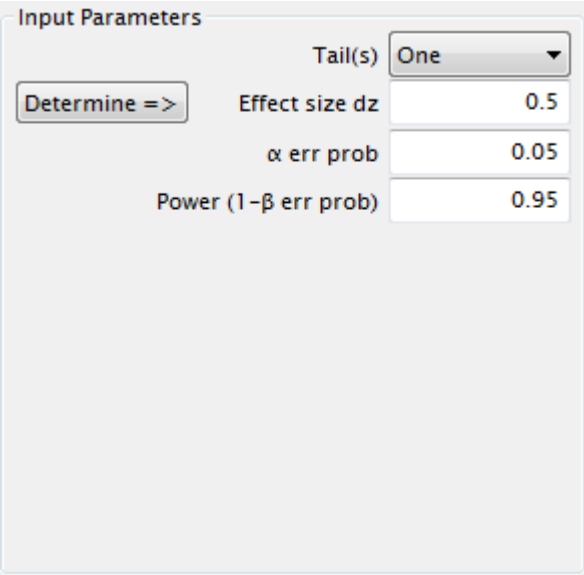
- Got a more difficult design? Learn how to simulate data in R, recreate the data you expect, and run simulations, performing the test you want to do.

- Ask for help – this is a job *real* statisticians do all the time.

Sample Size Planning

- Some things to remember:
  - There are different versions of Cohen's *d*. Subscripts are used to distinguish them.

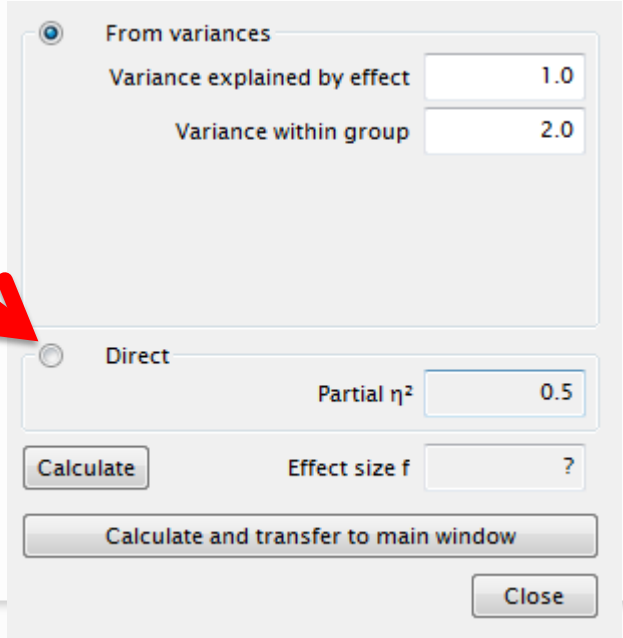




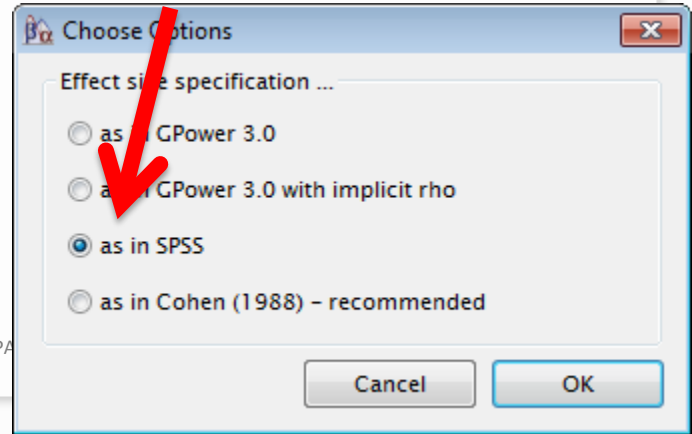
Sample Size Planning

- Some things to remember:
  - If you insert partial eta squared from repeated measure ANOVA's from SPSS directly into G*Power, use the 'AS IN SPSS' option!
    - (Many people make this error)

ONLY insert partial eta squared from SPSS



If you have selected 'As in SPSS' in the options window



Sample Size Planning

- Don't be surprised by what you find. Average effect size in psychology is d = 0.43 (= r = .21).
  - Independent sample t-test, two sided, power = .80
  - Need 86 ppn in each condition (N = 172)

  - **"Often when we statisticians present the results of a sample size calculation, the clinicians with whom we work protest that they have been able to find statistical significance with much smaller sample sizes. Although they do not conceptualize their argument in terms of power, we believe their experience comes from an intuitive feel for 50 percent power."**
  - Proschan, Lan, & Wittes, 2006

- If you perform 100 studies, how many times can you expect to observe a Type 1 error and how many times can you expect to observe a Type 2 error?
- This depends on how many times you will examine an effect where H1 is true, and how many times you will examine an effect where H0 is true, or the **prior probability**.

What will your next study yield?

For your thesis you set out to perform a completely novel study examining a hypothesis that has never been examined before. Let's assume you think it is equally likely that the null-hypothesis is true, as that it is false (both are **50% likely**). You set the **significance level at 0.05**. You design a study to have **80% power** if there is a true effect (assume you succeed perfectly). **Based on your intuition** (we will do the math later – now just answer intuitively) **what is the most likely outcome of this single study**? Choose one of the next four multiple choice answers.

A) It is most likely that you will observe a true positive (i.e., there is an effect, and the observed difference is significant).

B) It is most likely that you will observe a true negative (i.e., there is no effect, and the observed difference is not significant)

C) It is most likely that you will observe a false positive (i.e., there is no effect, but the observed difference is significant).

D) It is most likely that you will observe a false negative (i.e., there is an effect, but the observed difference is not significant)

What will your next study yield?

	H0 True (A-Priori 50% Likely)	H1 True (A-Priori 50% Likely)
Significant Finding	False Positives (Type 1 error) 2.5%	True Positives 40%
Non-Significant Finding	True Negatives 47.5%	False Negatives (Type 2 error) 10%

Power

A generally accepted minimum level of power is .80 (Cohen, 1988)

Why?

Power

This minimum is based on the idea that with a significance criterion of .05 the balance of a Type 2 error (1 – power) to a Type 1 error is .20/.05. (Cohen, 1988).

Concluding there *is* an effect when there is *no* effect in the population is considered four times as serious as concluding there is *no* effect when there *is* an effect in the population.

Power

Cohen (1988, p. 56) offered his recommendation in the hope that ‘it will be ignored whenever an investigator can find a basis in his substantive concerns in his specific research investigation to choose a value *ad hoc*.”

Power

But whatever conclusion is reached the following position must be recognised. If we reject H_0, we may reject it when it is true; if we accept H_0, we may be accepting it when it is false, that is to say, when really some alternative H_t is true. These two sources of error can rarely be eliminated completely; in some cases it will be more important to avoid the first, in others the second. We are reminded of the old problem considered by LAPLACE of the number of votes in a court of judges that should be needed to convict a prisoner. Is it more serious to convict an innocent man or to acquit a guilty? That will depend upon the consequences of the error; is the punishment death or fine; what is the danger to the community of released criminals; what are the current ethical views on punishment? From the point of view of mathematical theory all that we can do is to show how the risk of the errors may be controlled and minimised. The use of these statistical tools in any given case, in determining just how the balance should be struck, must be left to the investigator.

[Neyman & Pearson, 1933]

Power

At our department, the ethical committee requires a justification of the sample size you collect. Journals are starting to ask for this justification as well. Make sure you can justify your sample size.

If our researchers request money from the department, they should aim for 90% power. Exceptions are always possible, but the general rule is clear. We will not waste money on research that is unlikely to be informative.

Are most published findings false?

Researchers degrees of freedom

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding is less likely to be true when the studies conducted in a field are smaller; when effect sizes are smaller; when there is a greater number and lesser preselection of tested relationships; where there is greater flexibility in designs, definitions, outcomes, and analytical modes; when there is greater financial and other interest and prejudice; and when more teams are involved in a scientific field in chase of statistical significance. Simulations show that for most study designs and settings, it is more likely for a research claim to be false than true. Moreover, for many current scientific fields, claimed research findings may often be simply accurate measures of the

factors that influence this problem and some corollaries thereof.

Modeling the Framework for False Positive Findings

Several methodologists have pointed out [9–11] that the high rate of nonreplication (lack of confirmation) of research discoveries is a consequence of the convenient, yet ill-founded strategy of claiming conclusive research findings solely on the basis of a single study assessed by formal statistical significance, typically for a *p*-value less than 0.05. Research is not most appropriately represented and summarized by *p*-values, but, unfortunately, there is a widespread notion that medical research articles

It can be proven that most claimed research findings are false.

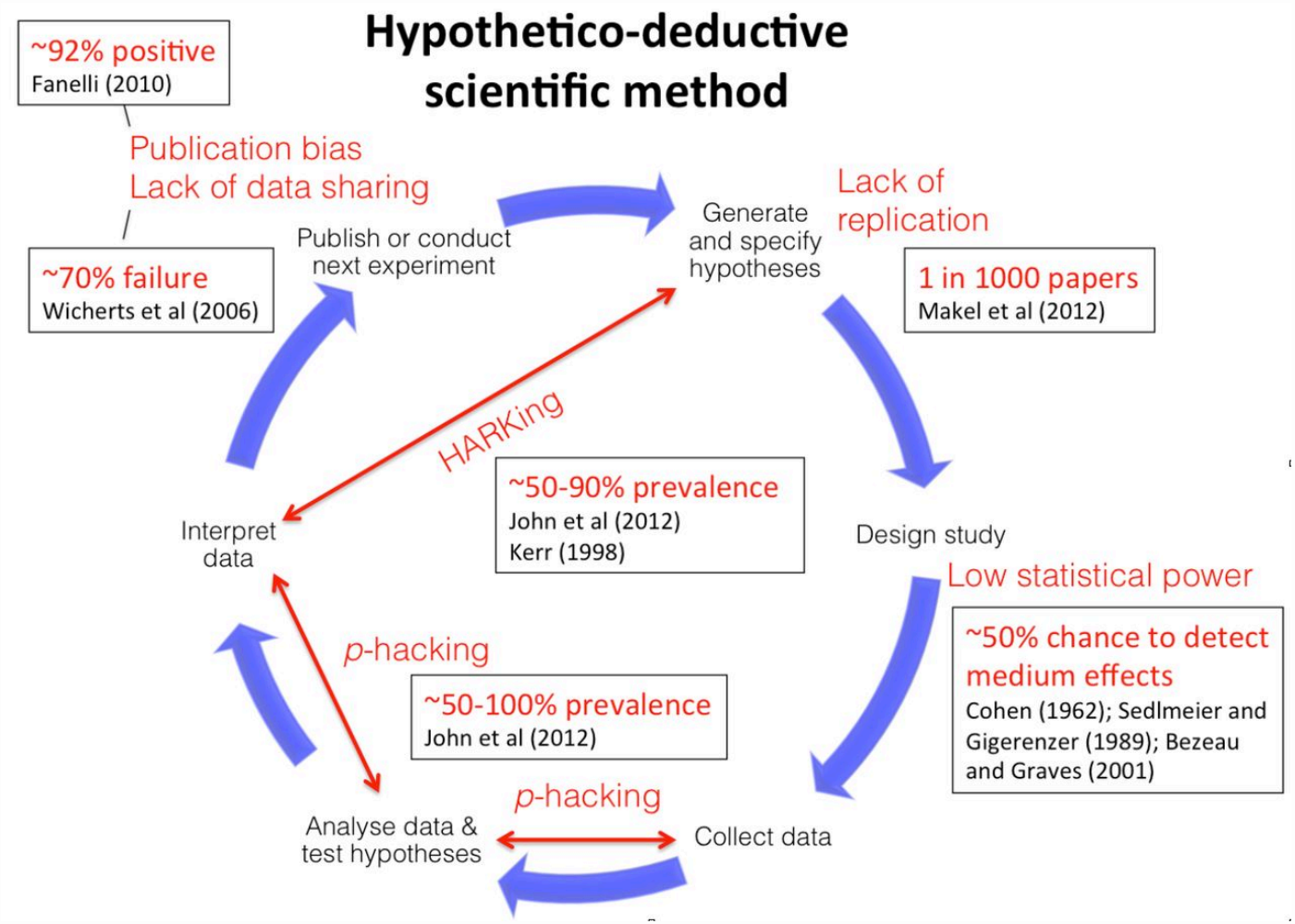
should be interpreted based only on *p*-values. Research findings are defined here as any relationship reaching

is characteristic of the field and can vary a lot depending on whether the field targets highly likely relationships or searches for only one or a few true relationships among thousands and millions of hypotheses that may be postulated. Let us also consider, for computational simplicity, circumscribed fields where either there is only one true relationship (among many that can be hypothesized) or the power is similar to find any of the several existing true relationships. The pre-study probability of a relationship being true is $R/(R + 1)$. The probability of a study finding a true relationship reflects the power $1 - \beta$ (one minus the Type II error rate). The probability of claiming a relationship when none truly exists reflects the Type I error rate, α. Assuming that c relationships are being probed in the field, the expected values of the 2×2 table are given in Table 1. After a research finding has been claimed based on achieving formal statistical significance, the post-study probability that it is true is the positive predictive value, PPV.

What do you think?

- How much published research is false?
- How much published research should be true?

What's the problem?



Technische Universiteit Eindhoven University of Technology

What is *p*-hacking?

- <u>Aiming for $p < \alpha$ by:</u>

- Optional stopping
- Dropping conditions
- Trying out different covariates
- Trying out different outlier criteria
- Combining DV's into sums, difference scores, etc.

- IMPORTANT: Only bad if you only report analyses that give $p < \alpha$, without telling people about the 20 other analyses you did.

The consequences

Table 1. Likelihood of Obtaining a False-Positive Result

	Significance level		
Researcher degrees of freedom	$p < .1$	$p < .05$	$p < .01$
Situation A: two dependent variables ($r = .50$)	17.8%	9.5%	2.2%
Situation B: addition of 10 more observations per cell	14.5%	7.7%	1.6%
Situation C: controlling for gender or interaction of gender with treatment	21.6%	11.7%	2.7%
Situation D: dropping (or not dropping) one of three conditions	23.2%	12.6%	2.8%
Combine Situations A and B	26.0%	14.4%	3.3%
Combine Situations A, B, and C	50.9%	30.9%	8.4%
Combine Situations A, B, C, and D	81.5%	60.7%	21.5%

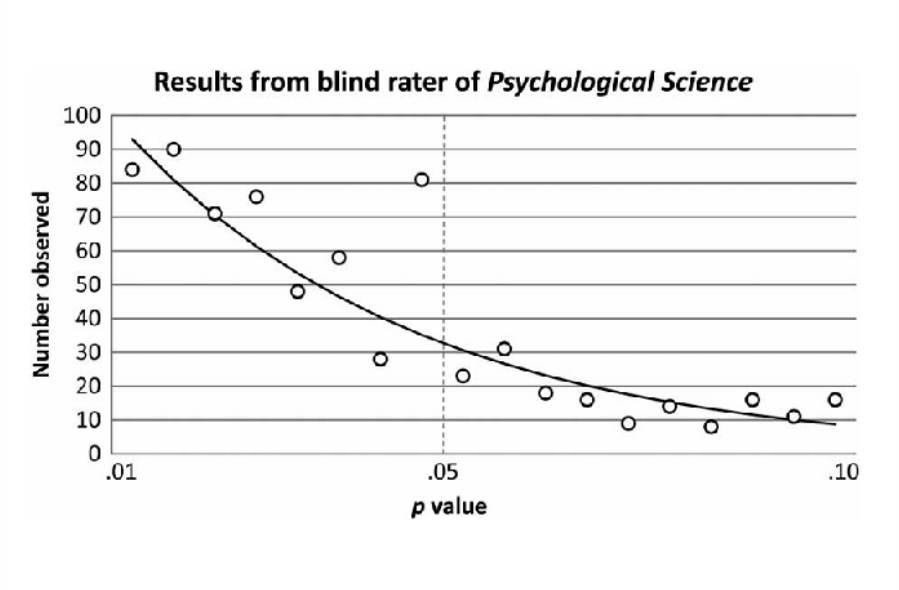
Note: The table reports the percentage of 15,000 simulated samples in which at least one of a set of analyses was significant. Observations were drawn independently from a normal distribution. Baseline is a two-condition design with 20 observations per cell. Results for Situation A were obtained by conducting three *t* tests, one on each of two dependent variables and a third on the average of these two variables. Results for Situation B were obtained by conducting one *t* test after collecting 20 observations per cell and another after collecting an additional 10 observations per cell. Results for Situation C were obtained by conducting a *t* test, an analysis of covariance with a gender main effect, and an analysis of covariance with a gender interaction (each observation was assigned a 50% probability of being female). We report a significant effect if the effect of condition was significant in any of these analyses or if the Gender × Condition interaction was significant. Results for Situation D were obtained by conducting *t* tests for each of the three possible pairings of conditions and an ordinary least squares regression for the linear trend of all three conditions (coding: low = −1, medium = 0, high = 1).

False Positives

Is there a ‘a peculiar prevalence of *p*-values just below 0.05’ (Masicampo & Lalande, 2012), are ”just significant” results on the rise’ (Leggett, Loetscher, & Nichols, 2013), and is there a ‘surge of *p*-values between 0.041-0.049’ (De Winter & Dodou, 2015)?

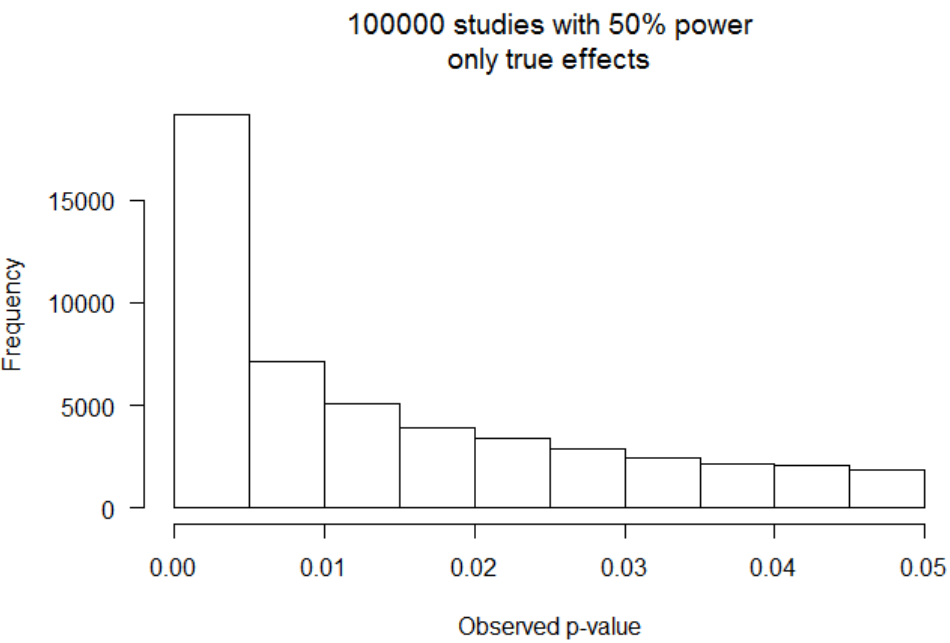
No (Lakens, 2014, 2015) – these claims over huge sets of studies are false. Remember to also be skeptical about the skeptics.

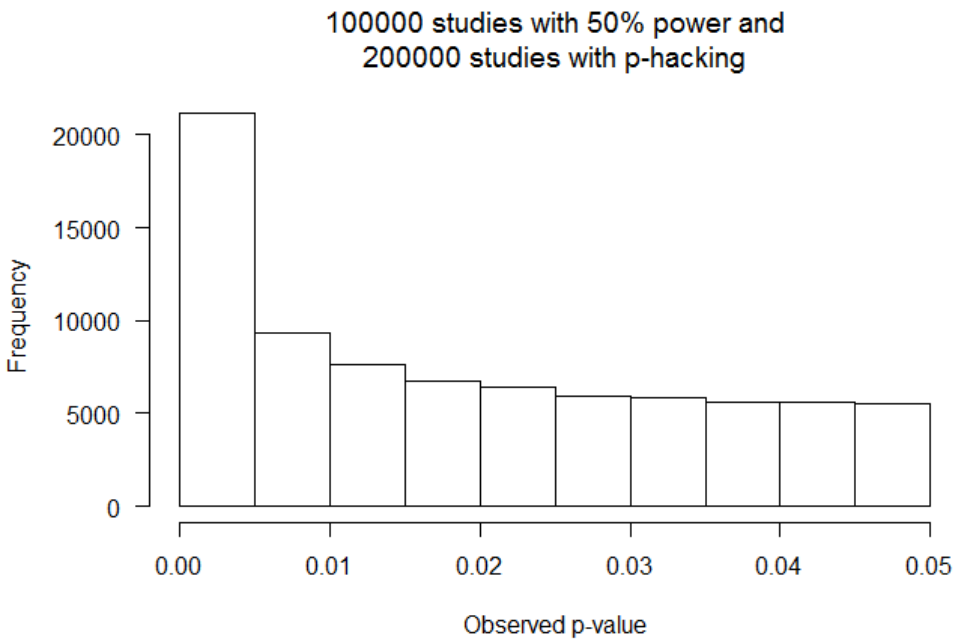
False Positives



Masicampo & Lalande (2012)

False Positives





Lakens, D. (2014). What *p*-hacking really looks like: A comment on Masicampo & LaLande (2012). *Quarterly Journal of Experimental Psychology, 68,* 829-832. doi: 10.1080/17470218.2014.982664.

False Positives

False positives should not be our biggest concern of the Big 3 (Publication Bias, Low Power, and False Positives) that threaten the False Positive Report Probability (Wacholder, Chanock, Garcia-Closas, El ghormli, & Rothman (2004) or Positive Predictive Value (Ioannidis, 2005).

However, it is by far the easiest one to fix, and to *identify*.

P-curve analysis

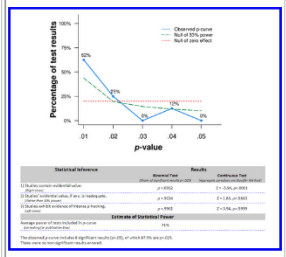
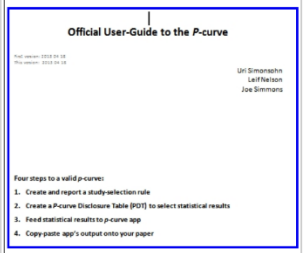
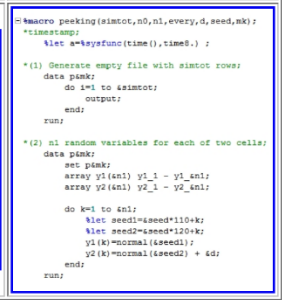
- Determine whether studies have evidential value

- Know what to trust, build on, and cite, and what to ignore (not build on or cite) untill beter evidence is available.

www.p-curve.com

P-curve.com

Paper 1 Evidential Value	Paper 2 Effect size	Paper 3 'Better P-curves' (robustness)	The online app 4.0	The User Guide	Supp Materials

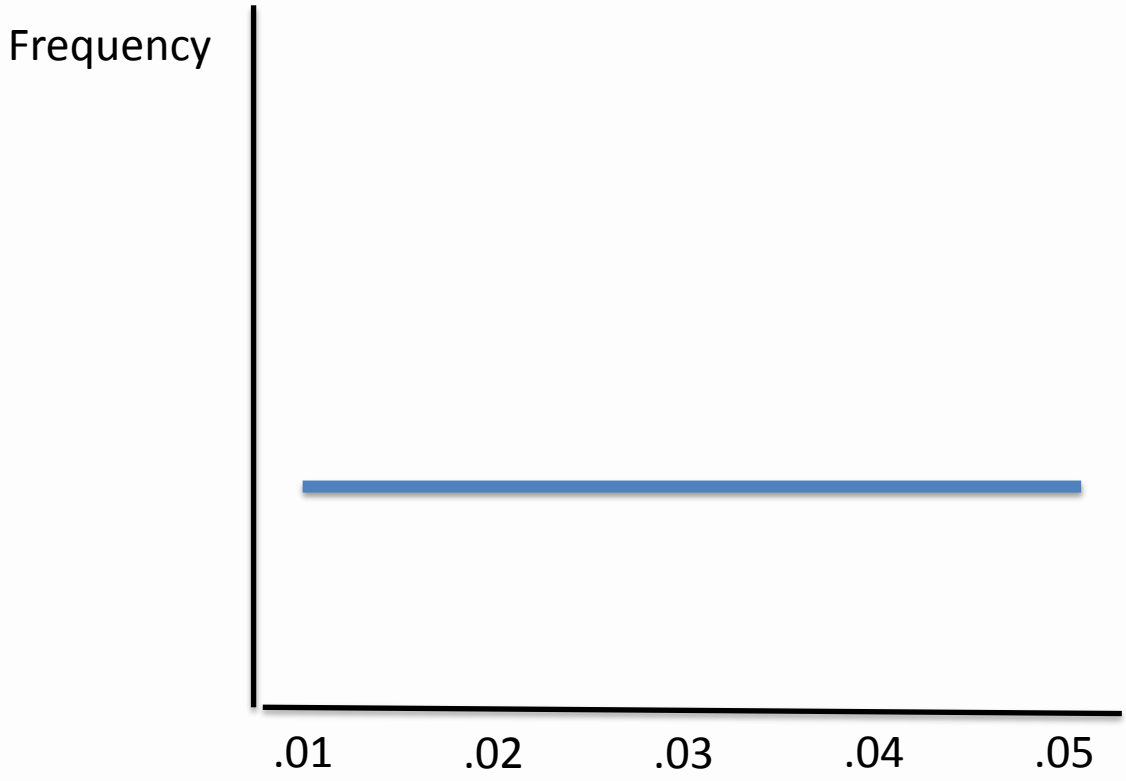

P-Curve: A Key to the File-Drawer
Uri Simonsohn
University of Pennsylvania
Leif D. Nelson
University of California, Berkeley
Joseph P. Simmons
University of Pennsylvania
	
	p-Curve and Effect Size: Correcting for Publication Bias Using Only Significant Results
	Uri Simonsohn[1], Leif D. Nelson[2], and Joseph P. Simmons[1]
		
		Better P-Curves
		Uri Simonsohn
		University of Pennsylvania - The Wharton School
		Joseph P. Simmons
		University of Pennsylvania - The Wharton School
		Leif D. Nelson
		University of California, Berkeley - Haas School of Business
		July 10, 2015
			
			
				
				Official User-Guide to the *P*-curve
				Uri Simonsohn
				Leif Nelson
				Joe Simmons
				Four steps to a valid *p*-curve:
				1. Create and report a study-selection rule
				2. Create a *P*-curve Disclosure Table (PDT) to select statistical results
				3. Feed statistical results to *p*-curve app
				4. Copy-paste app's output onto your paper
					
					%macro peeking(simtot,n0,n1,every,d,seed,mk);
					*timestamp;
					%let a=%sysfunc(time(),time8.) ;
					*(1) Generate empty file with simtot rows;
					data p&mk;
					do i=1 to &simtot;
					output;
					end;
					run;
					*(2) n1 random variables for each of two cells;
					data p&mk;
					set p&mk;
					array y1(&n1) y1_1 - y1_&n1;
					array y2(&n1) y2_1 - y2_&n1;
					do k=1 to &n1;
					%let seed1=&seed*110+k;
					%let seed2=&seed*120+k;
					y1(k)=normal(&seed1);
					y2(k)=normal(&seed2) + &d;
					end;
					run;

Distribution of *p*-values

- Take 100 studies that find a significant effect and plot the frequency of *p*-values.

- What should that distribution look like?

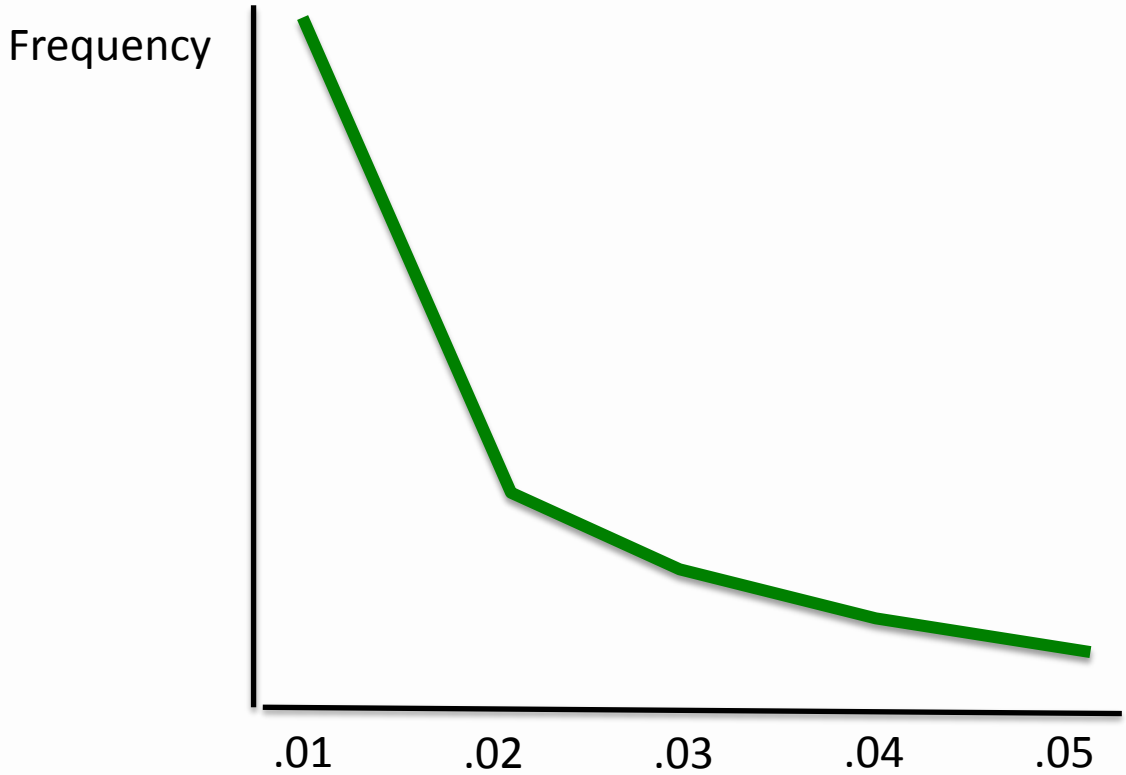
Distribution of *p*-values



No effect
Uniform

Every *p*-value is equally likely

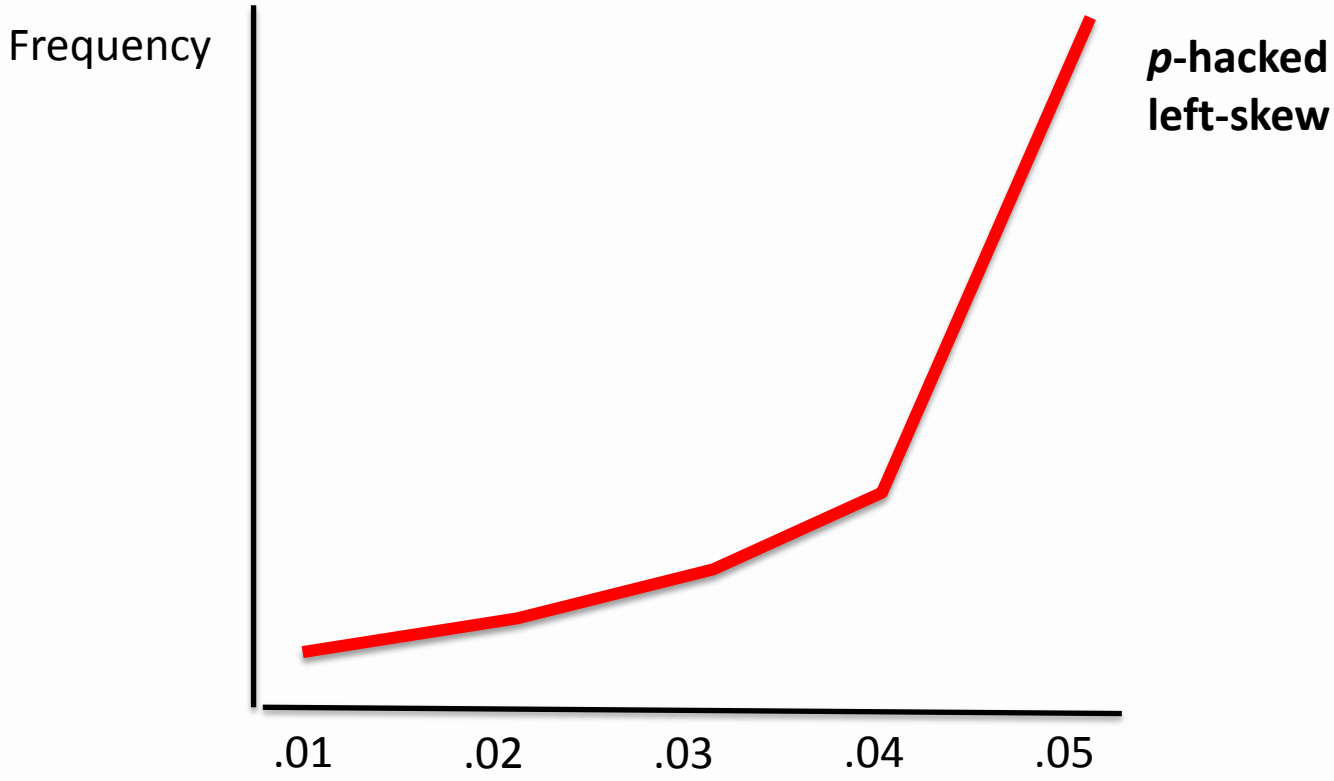
Distribution of *p*-values



True effect
Right-skew

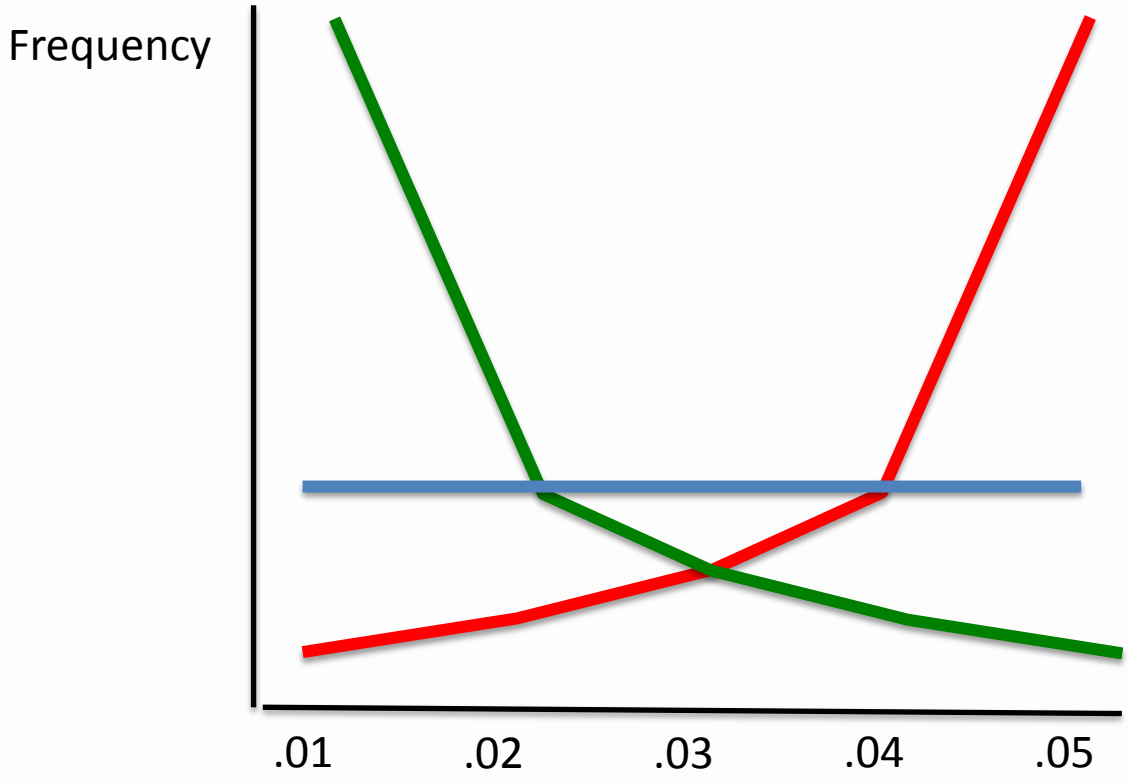
Small *p*-values are more likely

Distribution of *p*-values



Large *p*-values are more likely

Distribution of *p*-values



An example

Professors are Not Elderly: Evaluating the Evidential Value of Two Social Priming Effects Through P-Curve Analyses

Daniel Lakens
Eindhoven University of Technology (TUE)

January 20, 2014

Abstract:
It is possible that the number of false positives in the literature is much greater than is desirable due to a combination of low statistical power, publication bias, and flexibility when analyzing data. Recently, some researchers have argued the replicability crisis social priming research is greatly exaggerated (Dijksterhuis, 2014; Stroebe & Strack, 2014). To quantify the extent to which researcher degrees of freedom are a real problem, I present two p-curve analyses that examine the evidential value of research lines on professor priming and elderly priming. The results indicate studies examining elderly priming are p-hacked, while studies examining professor priming contain evidential value. I believe a polarized discussion about whether social priming is true or not, whether direct replications or conceptual replications are preferable, or whether methodological rigor or theory development is needed is unlikely to lead to scientific progress. Instead, we have to meta-analytically evaluate individual effects based on their evidential value, and collaboratively examine what is likely to be true.

Number of Pages in PDF File: 13

Keywords: P-curve, Social Priming, Statistical Power, Meta-Analysis

working papers series

Figure 1. *P-curve analysis of elderly priming studies*

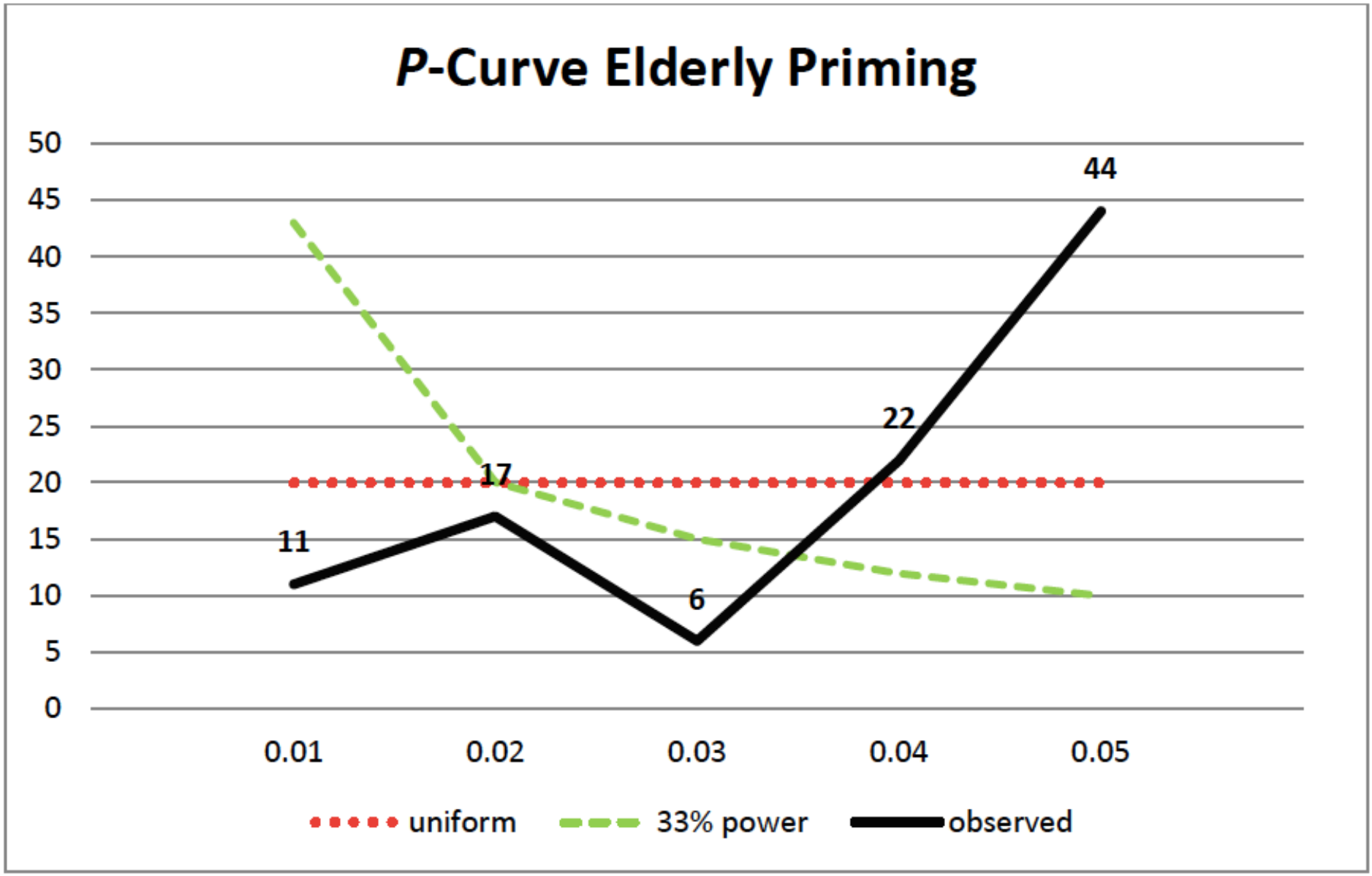
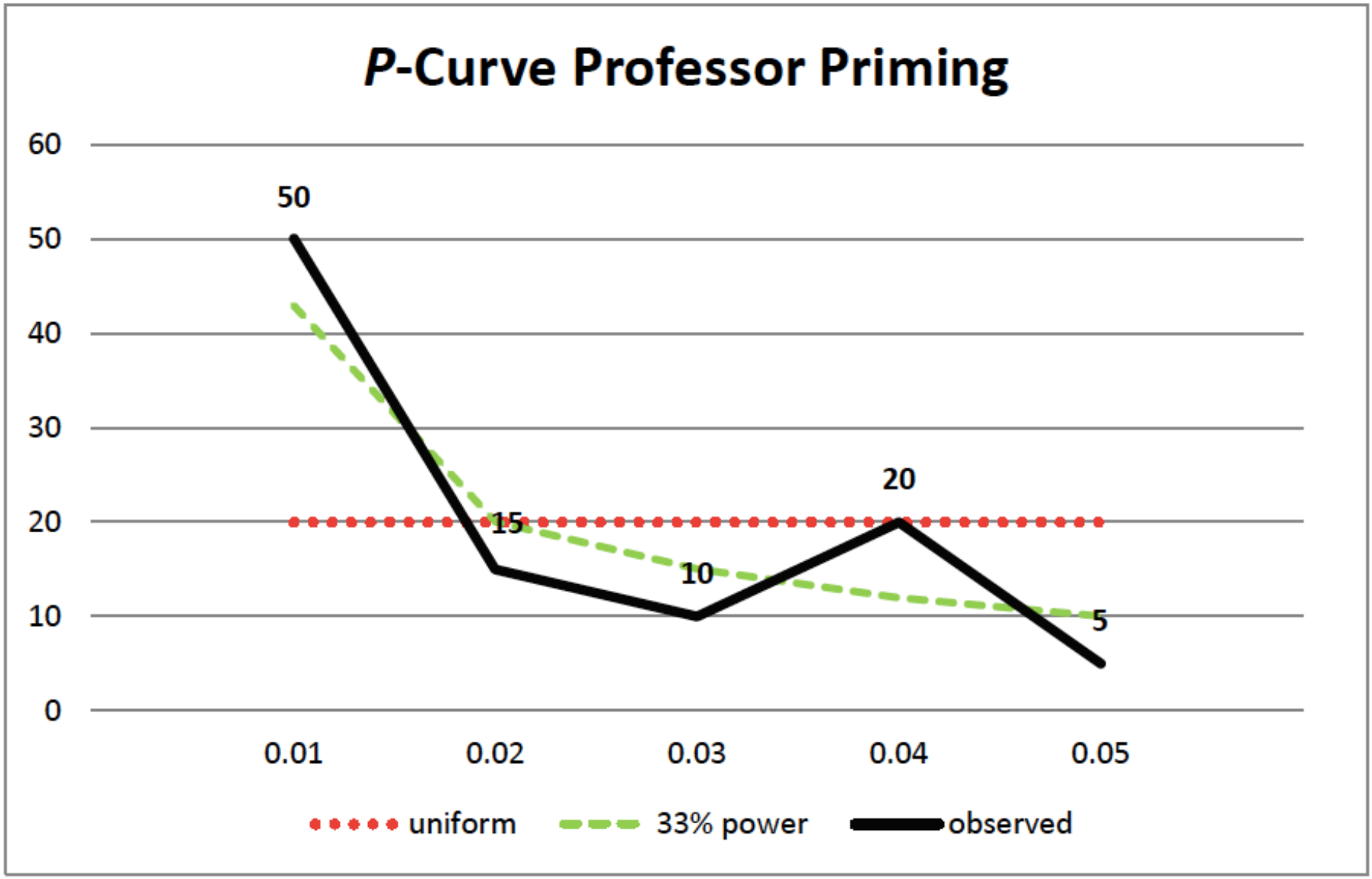


Figure 2. *P-curve analysis of professor priming studies*



What went wrong?

- One problem is that people tended to collect data, look at the data, collect more data, and stop when $p < 0.05$.
  - Called *optional stopping*
- With optional stopping the chance of $p < 0.05$ when H0 is true is 100% (if you are patient).

Ethical Issues in Data Collection

Continuing data collection whenever the desired level of confidence is reached, or whenever it is sufficiently clear the expected effects are not present, is a waste of the time of participants and the money provided by tax-payers.

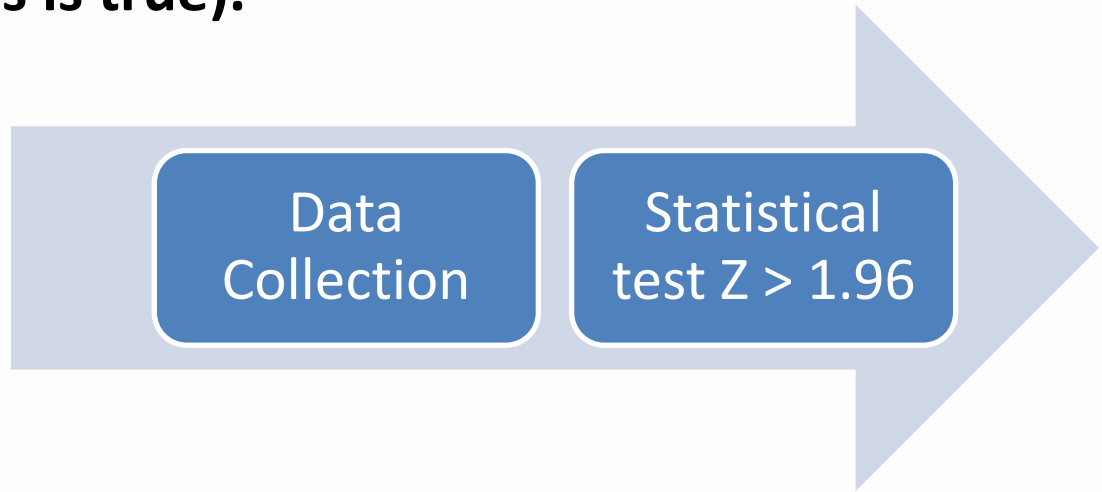
So do optional stopping right.

Sequential analyses

Because of the substantial savings in the expected number of observations effected by the sequential probability ratio test, and because of the simplicity of this test procedure in practical applications, the National Defense Research Committee considered these developments sufficiently useful for the war effort to make it desirable to keep the results out of the reach of the enemy, at least for a certain period of time. The author was, therefore, requested to submit his findings in a restricted report [7] which was dated September, 1943.[3] In this

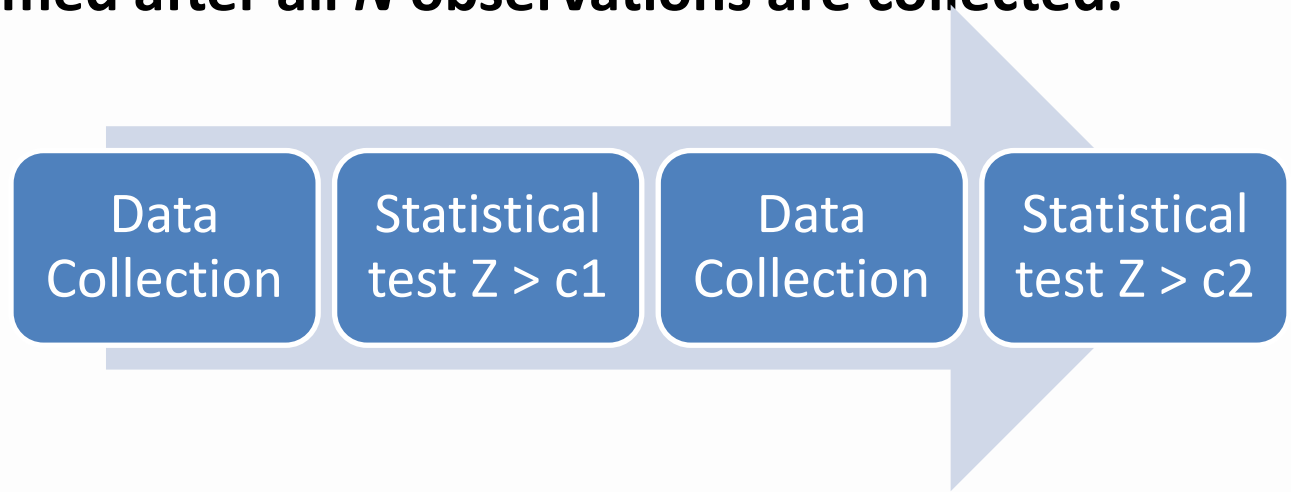
The main idea

- **With a symmetrical two-sided test, and an α = .05, this test should yield a *Z*-value larger than 1.96 (or smaller than -1.96) for the observed effect to be considered significant (which has a probability smaller than .025 for each tail, assuming the null-hypothesis is true).**



The main idea

- **When using sequential analyses with a single planned interim analysis, and a final analysis when all data is collected, one test is performed after *n* (e.g., 80) of the planned *N* (e.g., 160) observations have been collected, and another test is performed after all *N* observations are collected.**



We need to select boundary critical Z-values c1 and c2 (for the first and the second analysis) such that (for the upper boundary) the probability (Pr) that the null-hypothesis is rejected either when in the first analysis $Zn \geq c1$, or (when $Zn < c1$ in the first analysis) $ZN \geq c2$ in the second analysis. In formal terms:

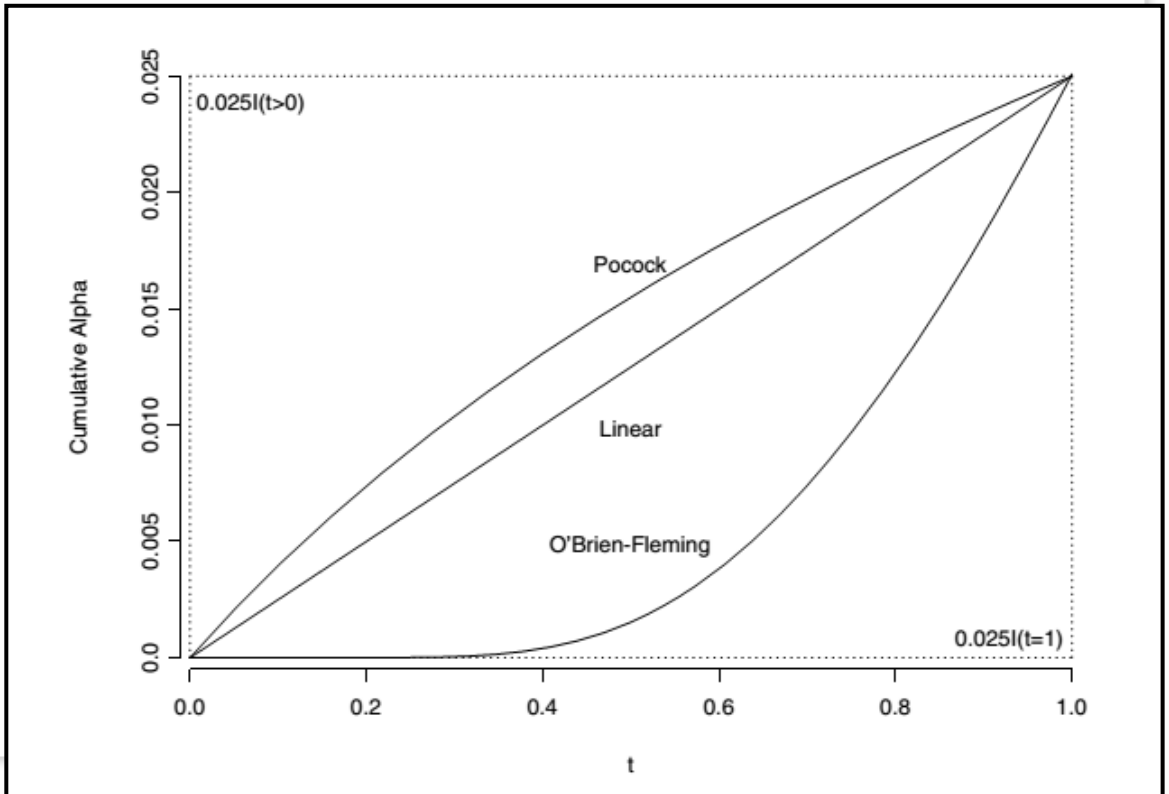
$$Pr\{Zn \geq c1\} + Pr\{Zn < c1, ZN \geq c2\} = 0.025$$

- See Proschan, Gordon-Lan, & Turk Wittes (2006)

(don't worry too much about the math)

$$
\begin{aligned}
0.025 &= \Pr\{Z_n \geq c_1\} + \Pr\{Z_n < c_1, Z_N \geq c_2\} \\
&= \Phi(-c_1) + \Pr\left\{\sqrt{\frac{n}{N}} \cdot Z_n < \sqrt{\frac{n}{N}} \cdot c_1, Z_N - \sqrt{\frac{n}{N}} \cdot Z_n \geq c_2 - \sqrt{\frac{n}{N}} \cdot Z_n\right\} \\
&= \Phi(-c_1) + \int_{-\infty}^{\sqrt{\frac{n}{N}} \cdot c_1} \int_{c_2 - x}^{\infty} \left(1 - \frac{n}{N}\right)^{-1/2} \phi\left(\frac{y}{\sqrt{1 - \frac{n}{N}}}\right) dy \left(\frac{n}{N}\right)^{-1/2} \phi\left(\frac{x}{\sqrt{\frac{n}{N}}}\right) dx \\
&= \Phi(-c_1) + \int_{-\infty}^{\sqrt{\frac{n}{N}} \cdot c_1} \Phi\left(\frac{x - c_2}{\sqrt{1 - \frac{n}{N}}}\right) \left(\frac{n}{N}\right)^{-1/2} \phi\left(\frac{x}{\sqrt{\frac{n}{N}}}\right) dx.
\end{aligned}
$$

So how do we determine the critical values? (and their accompanying nominal α levels) There are different approaches, each with its own rationale.



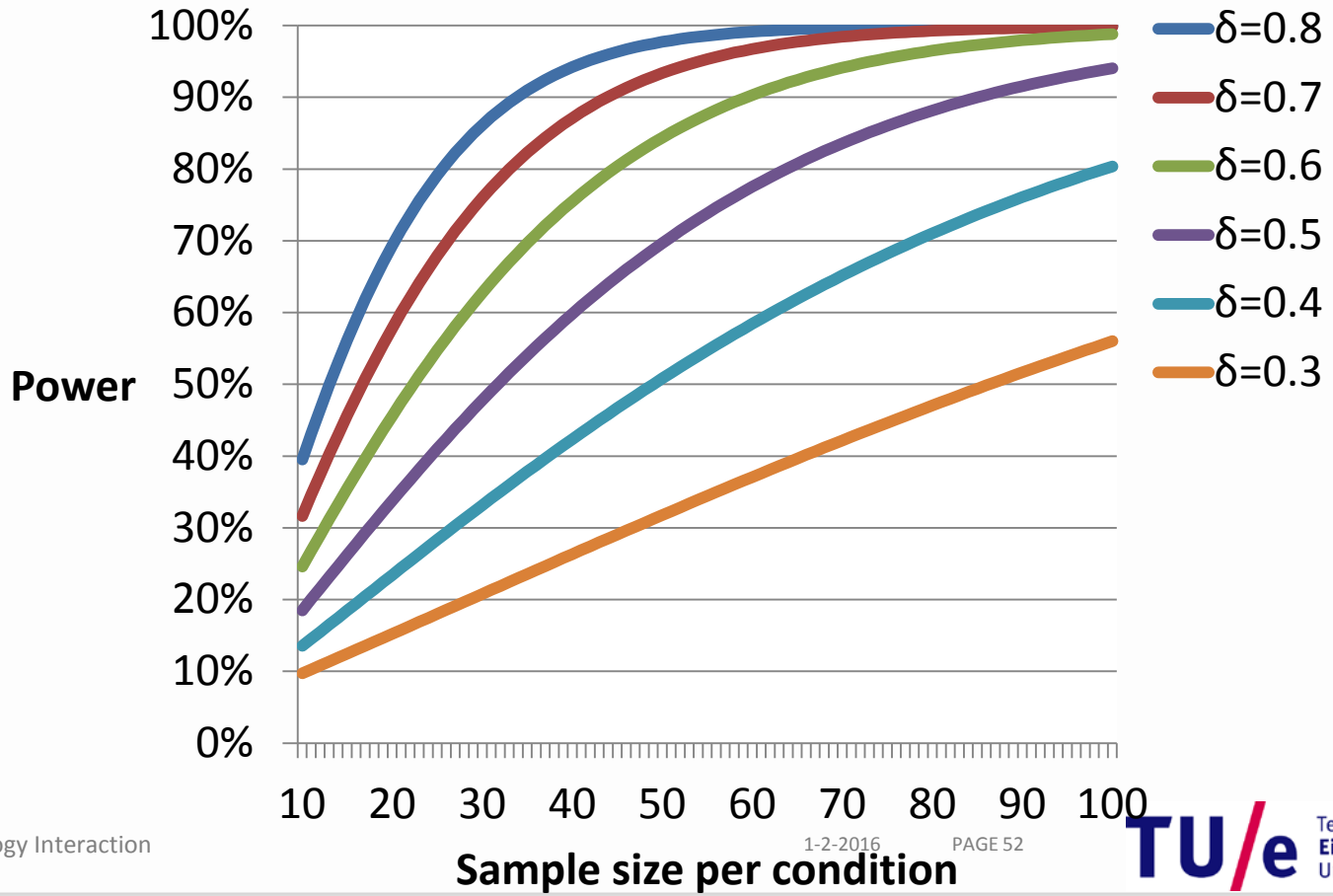
- For example, the Pocock boundary will lower the alpha level for each interim analysis. With 2 looks, the $\alpha = 0.0294$ for each analysis.

- Let's imagine after the first analysis, you find: $t(79) = 2.30$, $p = .024$.

- Because $p < .0294$, you terminate the data collection (and take the rest of the day off!).

The Benefit of Early Stopping

- Remember power is a concave function:



Getting Started

- For a practical introduction with step-by-step instructions, see Lakens (2014), European Journal of Social Psychology.
- **Using sequential analyses when you plan designs based on their power will make you 20/30% move efficient (when H1 is true, and save you even more when H0 is true).**

#OpenScience

	Pro-Self (no sharing, file-drawer, p-hacking)	Pro-Social (data sharing, replication, pre-registration)
Pro-Self (no sharing, file-drawer, p-hacking)	+- / +-	- / +
Pro-Social (data sharing, replication, pre-registration)	+ / -	++ / ++

Sci Eng Ethics (2007) 13:437–461
DOI 10.1007/s11948-007-9042-5

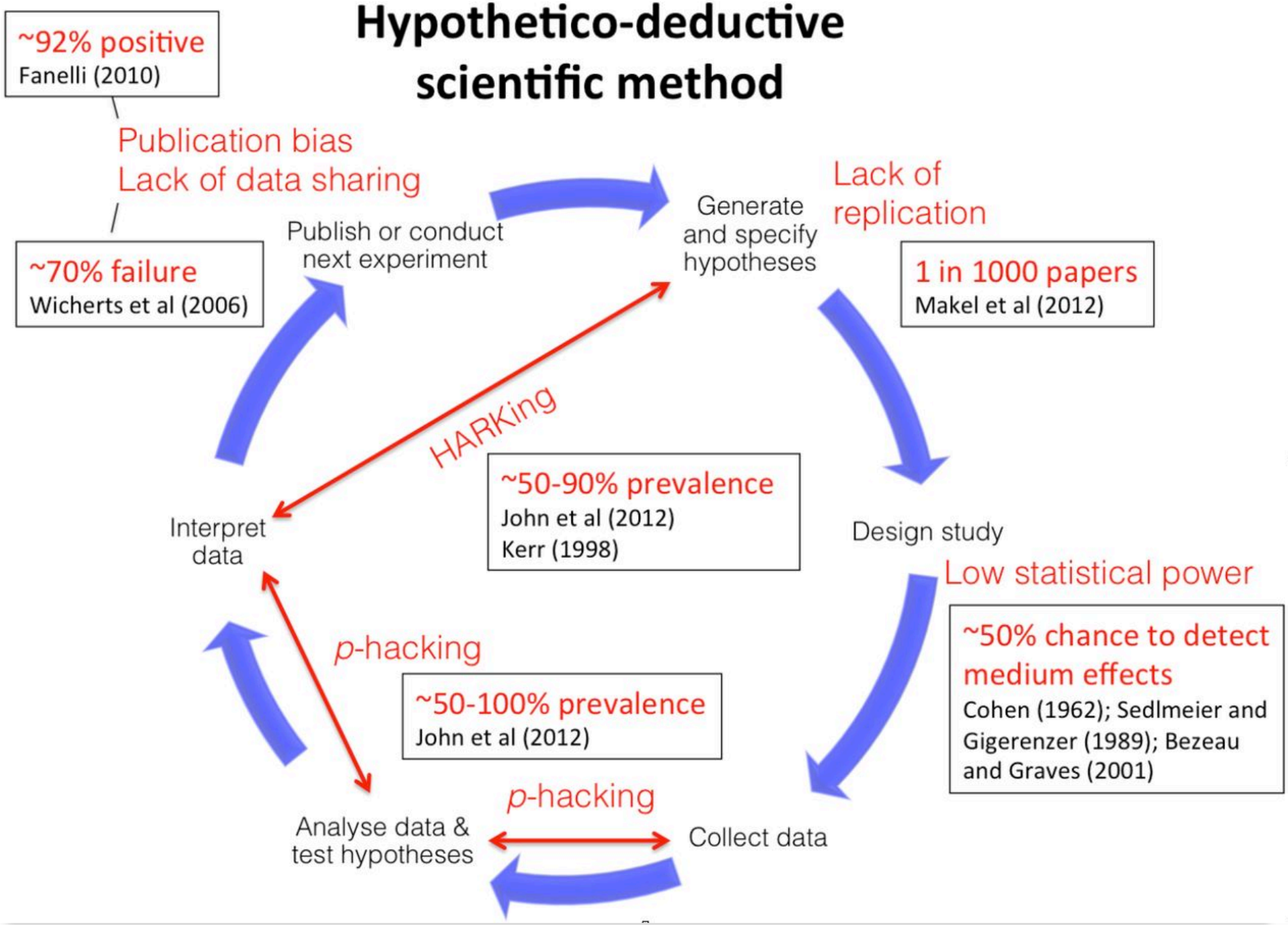
ORIGINAL PAPER

The Perverse Effects of Competition on Scientists' Work and Relationships

Melissa S. Anderson · Emily A. Ronning · Raymond De Vries · Brian C. Martinson

Abstract Competition among scientists for funding, positions and prestige, among other things, is often seen as a salutary driving force in U.S. science. Its effects on scientists, their work and their relationships are seldom considered. Focus-group discussions with 51 mid- and early-career scientists, on which this study is based, reveal a dark side of competition in science. According to these scientists, competition contributes to strategic game-playing in science, a decline in free and open sharing of information and methods, sabotage of others' ability to use one's work, interference with peer-review processes, deformation of relationships, and careless or questionable research conduct. When competition is pervasive, such effects may jeopardize the progress, efficiency and integrity of science.

Hypothetico-deductive scientific method

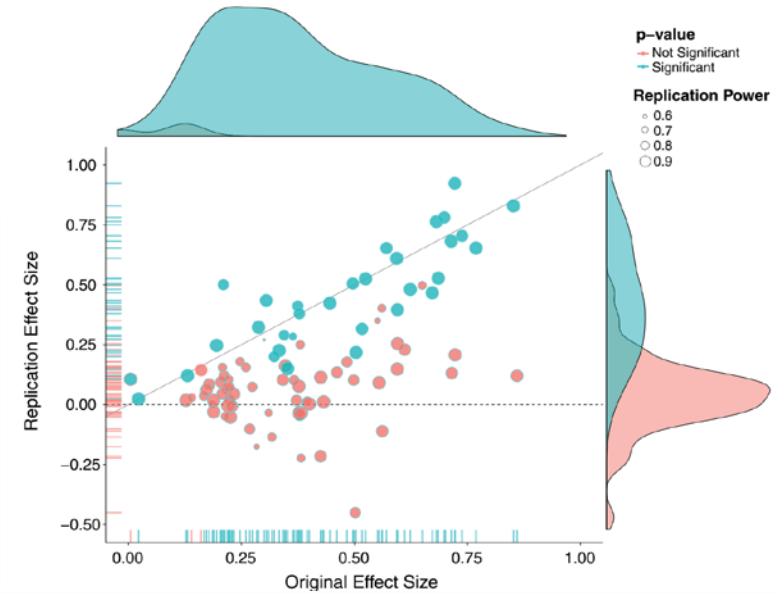


RESEARCH ARTICLE SUMMARY

PSYCHOLOGY

Estimating the reproducibility of psychological science

Open Science Collaboration*



			Effect size comparison					Original and replication combined			
	Replications $P < 0.05$ in original direction	Percent	Mean (SD) original effect size	Median original *df/N*	Mean (SD) replication effect size	Median replication *df/N*	Average replication power	Meta-analytic mean (SD) estimate	Percent meta-analytic ($P < 0.05$)	Percent original effect size within replication 95% CI	Percent subjective "yes" to "Did it replicate?"
Overall	35/97	36	0.403 (0.188)	54	0.197 (0.257)	68	0.92	0.309 (0.223)	68	47	39

RESEARCH ARTICLE SUMMARY

PSYCHOLOGY

Estimating the reproducibility of psychological science

Open Science Collaboration*

			Effect size comparison					Original and replication combined			
	Replications $P < 0.05$ in original direction	Percent	Mean (SD) original effect size	Median original *df/N*	Mean (SD) replication effect size	Median replication *df/N*	Average replication power	Meta-analytic mean (SD) estimate	Percent meta-analytic ($P < 0.05$)	Percent original effect size within replication 95% CI	Percent subjective "yes" to "Did it replicate?"
Overall	35/97	36	0.403 (0.188)	54	0.197 (0.257)	68	0.92	0.309 (0.223)	68	47	39

RESEARCH ARTICLE SUMMARY

PSYC

Es

ps

Open

Reproducibility Project (~60% failure rate)
(Open Science Collaboration, 2015)

Social Psych special issue (~70% failure rate)
(Nosek & Lakens, 2014)

Cancer cell biology (~90% failure rate)
(Begley & Ellis, 2012)

Cardiovascular health (~75% failure rate)
(Prinz, Schlange, & Asadullah, 2011)

	$P < 0.05$ in original direction	Percent	original effect size	original *df/N*	replication effect size	replication *df/N*	replication power	analytic mean (SD) estimate	meta-analytic ($P < 0.05$)	effect size within replication 95% CI	Percent subjective "yes" to "Did it replicate?"
Overall	35/97	36	0.403 (0.188)	54	0.197 (0.257)	68	0.92	0.309 (0.223)	68	47	39

Don’t focus on single *p*-values

Don’t care too much about every individual study having a *p*-value < .05.

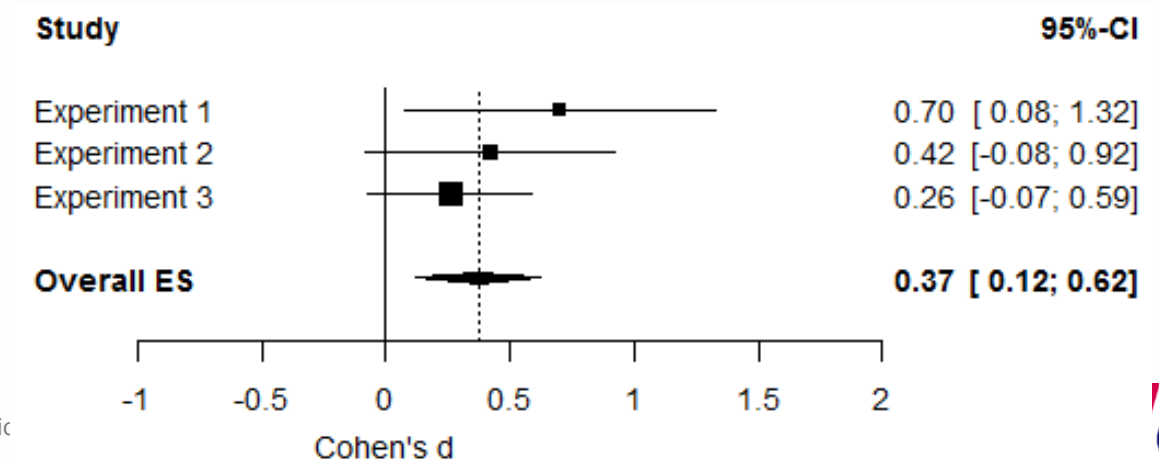
As long as you perform close replications, **report all the data**, and perform a small scale meta-analysis.

Zhang, Lakens, & IJsselsteijn, 2015

In press, Acta Psychologica

3 almost identical studies, study 3 pre-registered, 1/3 with $p<.05$

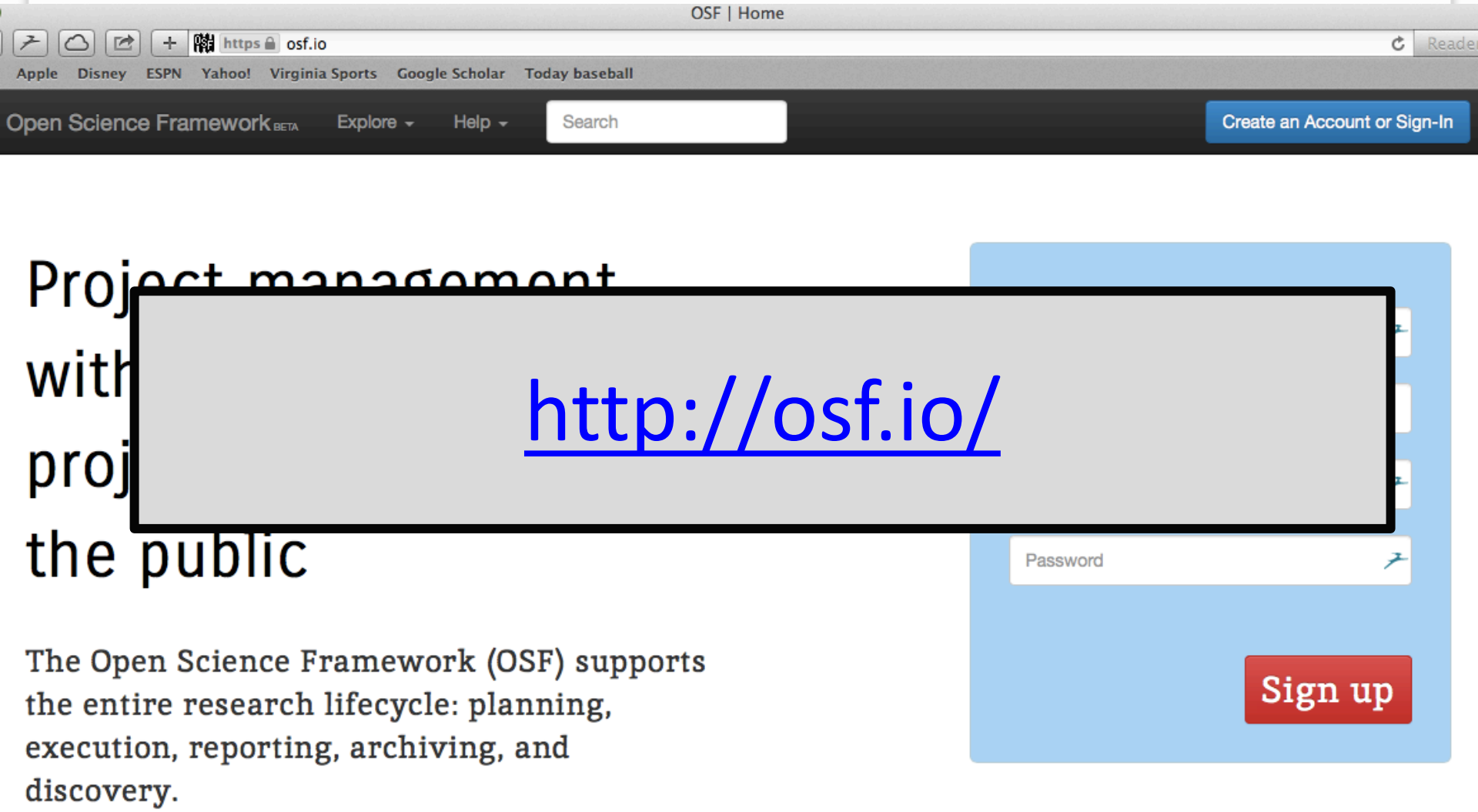
overall Cohen's d = 0.37, 95% CI [0.12, 0.62], t = 2.89, p = .004



35% increase in data sharing over the last 1.5 years by *just asking* for it

Dutch Science funder NWO will make data sharing a requirement for all tax funded research

Open Science Framework



The Spatial Grouping of Response Keys Influences Conceptual Congruency Effects

Make Private | Public

Contributors: Daniel Lakens, Iris Schneider, Sascha Topolinski, Thorsten Erle

Date Created: 2013-07-29 03:42 PM | Last Updated: 2014-04-09 09:22 PM

Description: No description

Overview | Files | Wiki | Statistics | Registrations | Forks | Sharing | Settings

The Spatial Grouping of Response Keys Conceptual Congruency Effects / Preregistration_Study_5.pdf

Preregistration_Study_5.pdf

Version	Date	User
1	2014-04-09 07:21 PM	Daniel Lakens

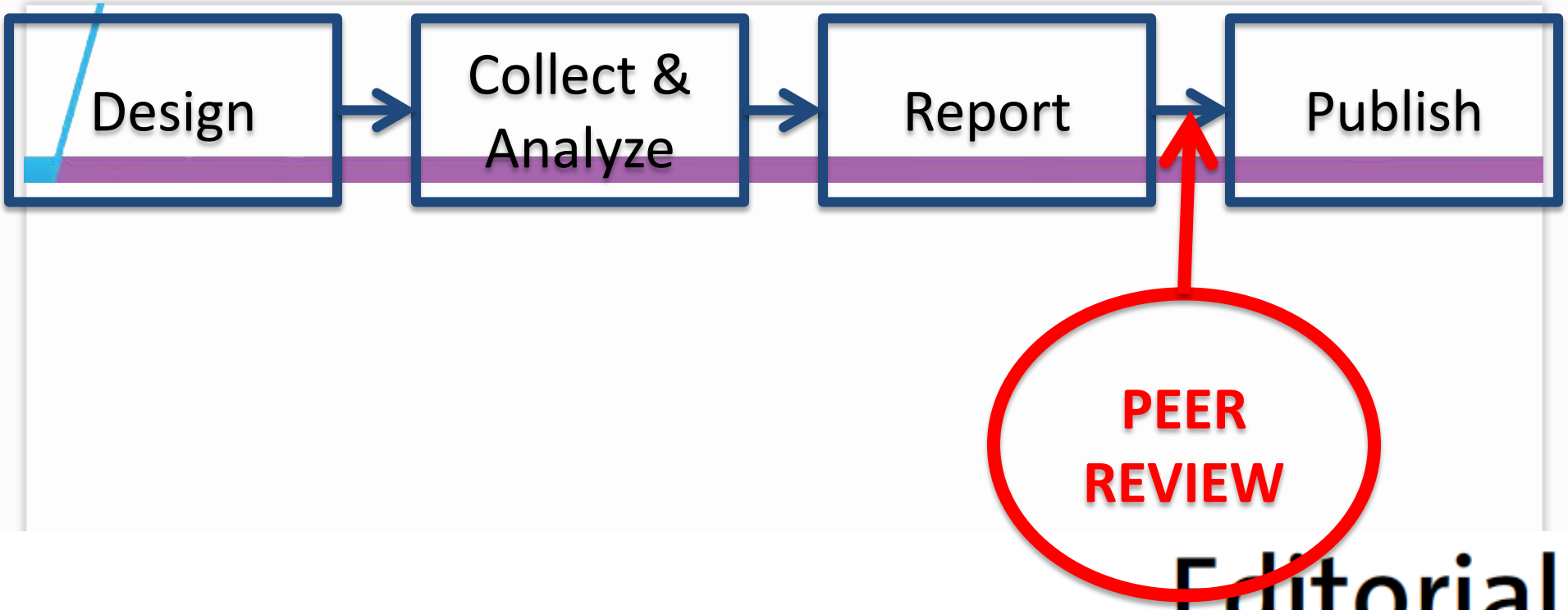
Page: 1 / 2

Pre-registration of Lakens, Erle, Schneider, & Topolinksi, Study 5, for the paper with the working title:' The Spatial Grouping of Response Keys Influences Conceptual Congruency Effects.'

This is a pre-registration of a planned sequential analysis to examine the difference in the congruency effect between two modified versions of the IAT, where participants use four response keys, either close together or far apart. Note that the data collection has started and is currently in progress, but that no data has been analyzed.

Procedure

Participants will perform a modified version of the IAT, and will be randomly assigned to the condition with two spatially differentiated groups of adjacent response keys (AS KL) or the condition where four spatially adjacent response keys did not easily afford a left vs. right subgrouping of the response keys (FGHJ). The A/F and the L/J keys are always paired with targets from the valence dimension, while the S/G and K/H keys are paired with



Editorial

Registered Reports

A Method to Increase the Credibility of Published Results

Brian A. Nosek[1] and Daniël Lakens[2]

SHARE PAPERS, DATA, ANALYSIS SCRIPTS, AND MATERIALS

SHERPA/RoMEO

Publisher copyright policies & self-archiving

Search

Journal titles or ISSNs Publisher names

Journal of Experimental Social Psychology

Exact title starts with contains ISSN

Advanced Search Search Reset

SHERPA/RoMEO

. . . opening access to rese

Home • Search • Journals • Publishers • FAQ • Suggest •

Search - Publisher copyright policies & self-archiving

English | Español | Magyar | Po

One journal found when searched for: **journal of experimental social psychology**

Journal:	Journal of Experimental Social Psychology (ISSN: 0022-1031)
RoMEO:	This is a RoMEO **green** journal
Paid OA:	A paid open access option is **available** for this journal.
Author's Pre-print:	✓ author **can** archive pre-print (ie pre-refereeing)
Author's Post-print:	✓ author **can** archive post-print (ie final draft post-refereeing)
Publisher's Version/PDF:	✗ author **cannot** archive publisher's version/PDF
General Conditions:	• Pre-print allowed on any website or open access repository • Voluntary deposit by author of authors post-print allowed on authors' personal website, arXiv.org or institutions open scholarly website including Institutional Repository, without embargo, where there is not a policy or mandate • Deposit due to Funding Body, Institutional and Governmental policy or mandate only allowed where separate agreement between repository and the publisher exists. • Permitted deposit due to Funding Body, Institutional and Governmental policy or mandate, may be required to comply with embargo periods of 12 months to 48 months . • Set statement to accompany deposit • Published source must be acknowledged • Must link to journal home page or articles' DOI • Publisher's version/PDF cannot be used • Articles in some journals can be made Open Access on payment of additional charge • NIH Authors articles will be submitted to PubMed Central after 12 months
Mandated OA:	(*Awaiting information*)
Paid Open Access:	Open Access Articles
Notes:	• Publisher last contacted on 18/10/2013
Copyright:	Article Posting Policies - Rights & responsibilities - Funding Body Agreements - Green Open Access - Open Access License Policy Green Open Access - Elsevier Journal Specific Embargo Periods
Updated:	16-May-2014 - Suggest an update for this record

OSF | Home

https osf.io

Apple Disney ESPN Yahoo! Virginia Sports Google Scholar Today baseball

Open Science Framework BETA Explore Help Search Create an Account or Sign-In

Project management with collaborators, project sharing with the

Full Name

Contact Email

Confirm Email

http://osf.io/

The O

the e

execution, reporting, archiving, and discovery.

https://osf.io/wx7ck/

https://osf.io/c97pd/

Register

Registration

Registration cannot be undone, and the archived content and files cannot be deleted after registration. Please be sure the project is complete and comprehensive for what you wish to register.

Type "register" if you are sure you want to continue

Open Science Framework BETA | My Dashboard | Explore | Help | Search | Daniel Lakens

Revisiting Tversky's Diagnosticity Principle

Make Private | Public | 1 | 0

Contributors: Daniel Lakens, Ellen Evers
Date Created: 2014-06-07 10:56 AM | Last Updated: 2014-07-22 07:04 AM
Description: Two pre-registered replications of the studies on the diagnosticity effect reported in Tversky (1977).

Overview | Files | Wiki | Statistics | Registrations | Forks | Sharing | Settings

Wiki

No wiki content

Files

Search files...

Name

- Project: Revisiting Tversky's Diagnosticity Principle
  - ALL_DATA.xlsx
  - Classroom_Data_Faces_2012_and_2013.xlsx
  - Pre-registration.pdf
  - Component: Data Study 1a
    - faces_happy_eindhoven_SIMILARITY.xlsx
    - faces_sad_eindhoven_SIMILARITY.xlsx
    - tversky_faces_eindhoven_SIMILARITY.sav
    - tversky_faces_tilburg_CATEGORIZATION.sav
  - Component: Data Study 1b
    - mturk_tversky_faces_CATEGORIZATION_cleaned.sav
    - mturk_tversky_faces_CATEGORIZATION_raw.sav
    - mturk_tversky_faces_SIMILARITY_cleaned.sav
    - mturk_tversky_faces_SIMILARITY_raw.sav
  - Component: Data Study 2a
    - Countries_eindhoven_A_Hongarije_CATEGORIZATION.xlsx
    - Countries_eindhoven_B_India_CATEGORIZATION.xlsx
    - tversky_landen_tilburg_SIMILARITY.sav
  - Component: Data Study 2b
    - mturk_tversky_countries_CATEGORIZATION_cleaned.sav
    - mturk_tversky_countries_CATEGORIZATION_raw.sav
    - mturk_tversky_countries_SIMILARITY_raw.sav
    - mturk_tversky_landen_SIMILARITY_cleaned.sav

Citation: osf.io/e6cr3 more

Components

Add Component | Add Links

Data Study 1a +
Lakens & Evers
7 contributions

Data Study 1b +
Lakens & Evers
7 contributions

Data Study 2a +
Lakens & Evers
6 contributions

Data Study 2b +
Lakens & Evers
7 contributions

Materials Study 1 +
Lakens & Evers
5 contributions

Materials Study 2 +
Lakens & Evers
7 contributions

Meta-Analysis +

Supporting the spread of open research practices

Search …

RECENT POSTS

PRO Initiative media

RECENT COMMENTS

The Peer Reviewers' Openness Initiative – Learn. Understand. Create on The Initiative

Peer Reviewers' Openness Initiative | Carlos Velasco on The Initiative

The Peer Reviewers' Openness Initiative « Mind Hacks on The Initiative

Datatilgængelighed og forskningstroværdighed | Erik Gahner Larsen on The Initiative

Richard Morey on Guidelines

THE INITIATIVE

Read the paper about the PRO initiative

Join the Initiative!

Peer Reviewers' Openness Initiative, version 2.1

Openness and transparency are core values
festation of those values, a minimum requirem
any scientific results must be the public subm
used in generating those results. As reviewers

An der **Fakultät für Psychologie und Pädagogik** der Ludwig-Maximilians-Universität München ist zum Wintersemester 2016/2017 eine

Professur (W3) für Sozialpsychologie (Lehrstuhl)

zu besetzen.

Zu den Aufgaben in der Lehre gehört die Vertretung des Faches Sozialpsychologie in seiner ganzen Breite im Bachelor-Studiengang „Psychologie", in verschiedenen Nebenfachstudiengängen der Psychologie und im Masterstudiengang „M.Sc. in Psychologie: Wirtschafts-, Organisations- und Sozialpsychologie".

Forschungsschwerpunkte mit Anschlussfähigkeit an die Forschungsaktivitäten im Rahmen des „Munich Center of the Learning Sciences" (MCLS, www.mcls.lmu.de ↗), des „Munich Experimental Laboratory for Economic and Social Sciences" (MELESSA, www.melessa.lmu.de ↗) oder der „Graduate School of Systemic Neurosciences" (GSN, www.gsn.lmu.de ↗) sind erwünscht.

Die Ludwig-Maximilians-Universität München (LMU) möchte eine hervorragend ausgewiesene Persönlichkeit gewinnen, die ihre wissenschaftliche Qualifikation im Anschluss an ein abgeschlossenes Hochschulstudium und eine überdurchschnittliche Promotion im Bereich Psychologie durch international sichtbare, exzellente Leistungen in Forschung (z. B. Publikationen in international anerkannten Fachzeitschriften, erfolgreiche Einwerbung von Drittmitteln) und Lehre (u. a. durch Lehrevaluation) nachgewiesen hat. Die Mitwirkung an den Forschungsaktivitäten im Rahmen von MCLS, MELESSA oder GSN ist erwünscht.

Das Department Psychologie legt Wert auf transparente und replizierbare Forschung und unterstützt diese Ziele durch Open Data, Open Material und Präregistrierungen. Bewerber/innen werden daher gebeten, in ihrem Anschreiben darzulegen, auf welche Art und Weise sie diese Ziele bereits verfolgt haben und in Zukunft verfolgen möchten.

Thanks for Your Attention!

Blog on methods & statistics

http://daniellakens.blogspot.nl/

Questions when you start using these techniques?
Contact me on Twitter:

@Lakens