Learn to *p*-hack like the pros!

J.J. at the English language Wikipedia

Prof. Dr. Felix Schönbrodt
Ludwig-Maximilians-Universität München

RESEARCH TRANSPARENCY

OSC
LMU Open Science Center

www.nicebread.de
www.researchtransparency.org
@nicebread303

SCIENCE

The **smartest heads** in the world immerse themselves into a research topic for years.

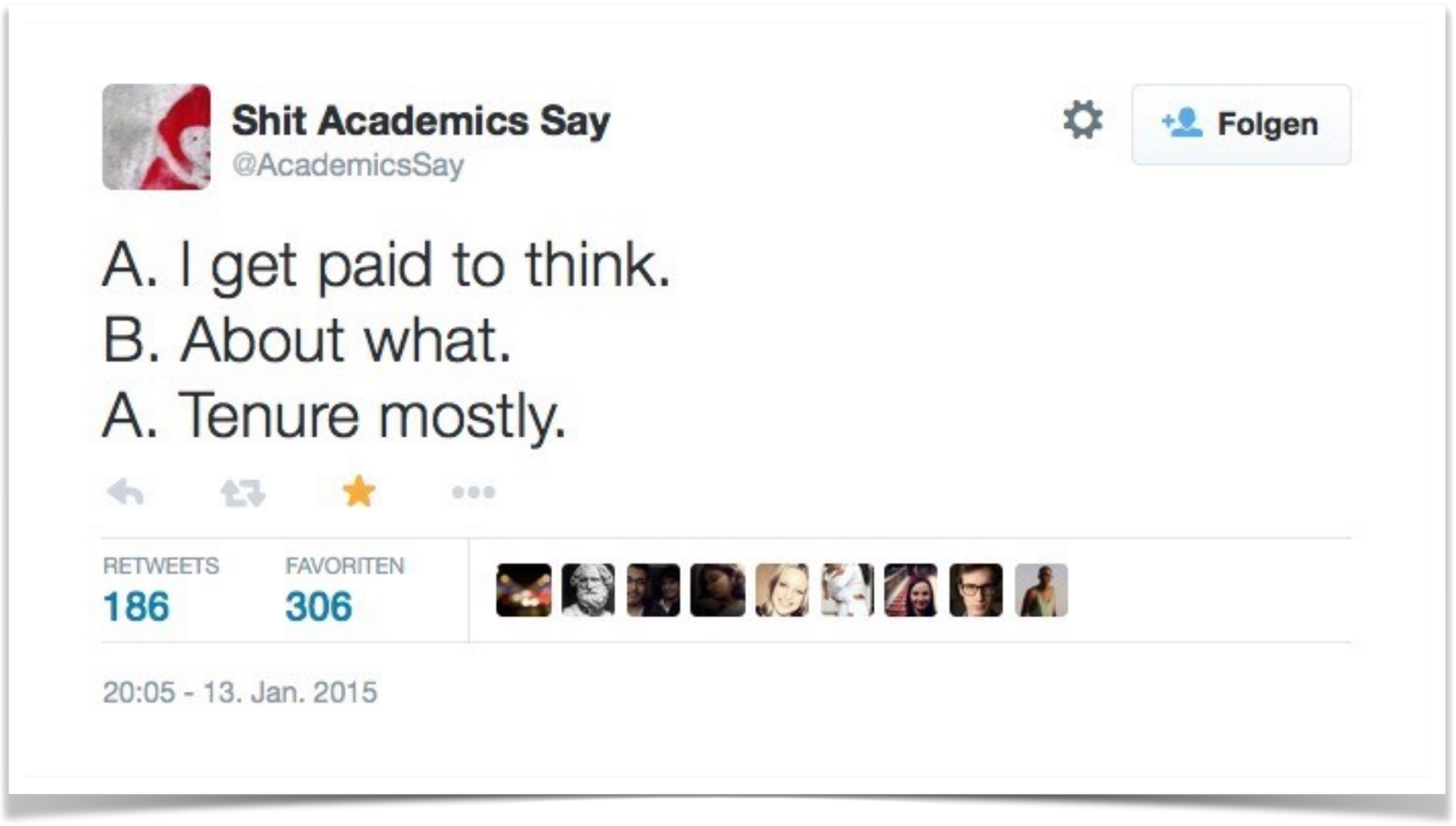
In that process, they become the experts – nobody knows more about that topic. The boundaries of knowledge have been **pushed forward**.

When the scientist are confident in their findings, they publish them in the best scientific journals, with the highest standards of **quality, rigor, and integrity**.

HOW MUCH OF THAT LITERATURE DO YOU THINK IS TRUE?

Researchers are not rewarded for being right,
but rather for publishing a lot.

Nelson, Simmons, & Simonsohn (2012); Nosek, Spies, Motyl (2012); Munafo (2016)



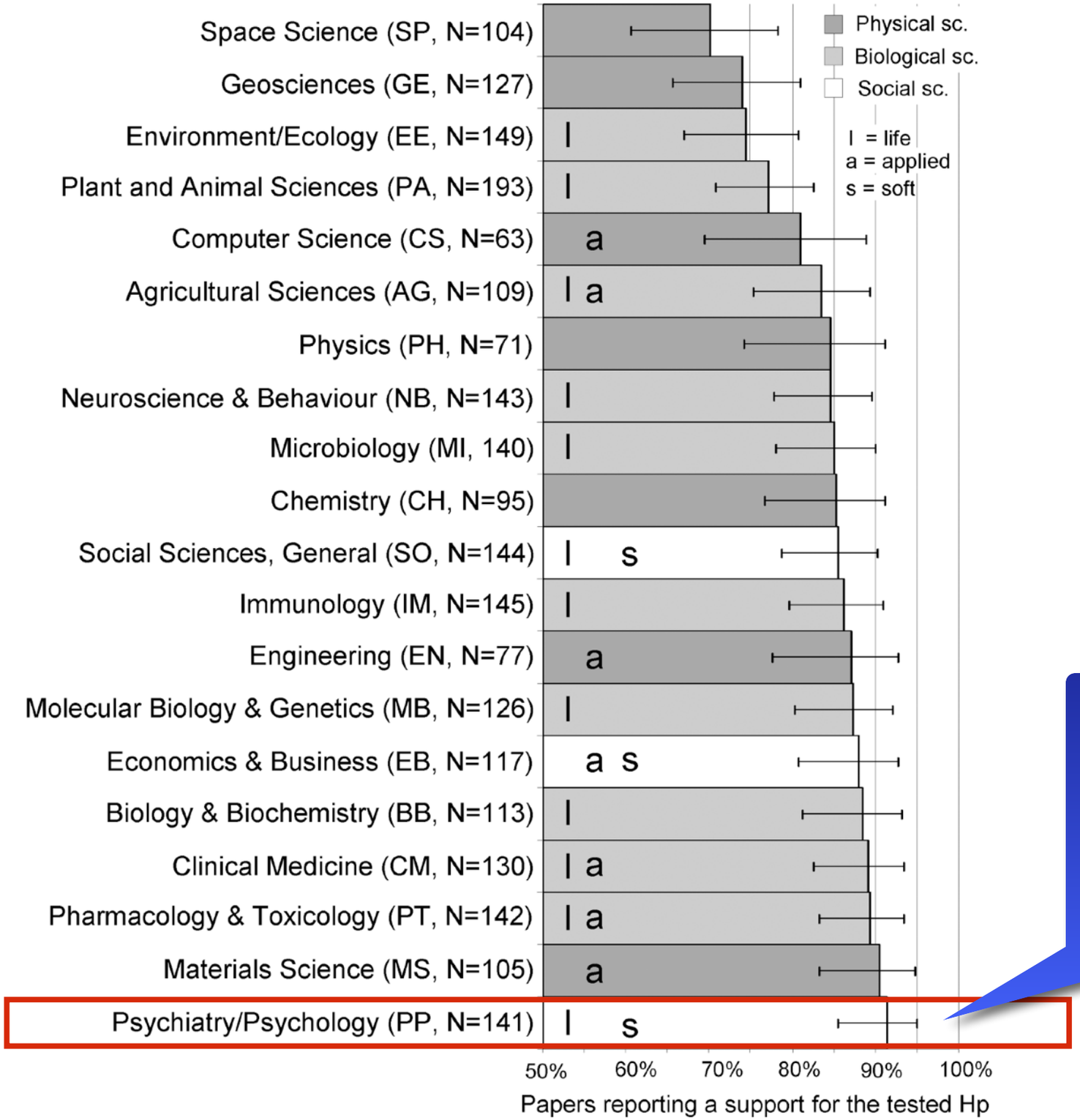
How to become a Professor?

Actual (not desired) relevance in professorship hiring committees	Rank
Number of peer-reviewed publications	1
Fit of research profile to the hiring department	2
Quality of research talk	3
Number of publications	4
Volume of acquired third-party funding	5
Number of first authorships	6
…	…

N = 1453 psychology researchers, 66% were actually members of a professorship hiring committee.

Abele-Brehm, A. E., & Bühner, M. (2016). Wer soll die Professur bekommen? Psychologische Rundschau, 67(4), 250–261. http://doi.org/10.1026/0033-3042/a000335

How to get lots of publications?



92% of published papers have significant, positive results

Fanelli, D. (2010). “Positive” Results Increase Down the Hierarchy of the Sciences. *PLOS ONE*, *5*, e10068. doi:10.1371/journal.pone.0010068

p-hack your way to scientific glory!

***p*-hacking** (*n.*). Tune your data analysis in a way that you achieve a significant *p*-value in situations where it would have been non-significant.

Questionable research practices (**QRPs**) (*n.*). Practices of data collection and data analysis that are not outright fraud, but also not really kosher.

Tool 1: Outcome switching

COMPARE

TRACKING SWITCHED OUTCOMES IN CLINICAL TRIALS

PROJECT RESULTS TEAM BLOG FAQ

Tracking switched outcomes in clinical trials

Here's what we found.

67	9	300	357
TRIALS CHECKED	TRIALS WERE PERFECT	OUTCOMES NOT REPORTED	NEW OUTCOMES SILENTLY ADDED

On average, each trial reported just 62.1% of its specifie
average, each trial silently added 5.3 new outcomes.

For [redacted], *"the authors conducted two additional money priming studies that showed no effects, the details of which were shared with us."* and *"reported nine dependent measures that were statistically affected by the manipulation in the predicted direction (one in each experiment) but did not report 19 additional measures that were statistically unchanged"*.

http://compare-trials.org/ http://blogs.discovermagazine.com/neuroskeptic/2015/07/23/social-priming-money-for-nothing/#.VuKRSRi5KJM

Tool 1: Outcome switching

- 2 outcome variables:

  false positive rate **5% → 9.5%**

- 5 outcome variables with one-sided testing:

  false positive rate **5% → 41%**

- How prevalent is it?
  - John, Loewenstein and Prelec (2012):
    66% of researchers admit having done this.

Tool 2: Many conditions, report only those that worked

- Assess more than two conditions (and leave out conditions that are not significantly different).
- E.g., testing "high", "medium" and "low" conditions and reporting only the results of a "high" versus "medium" comparison.
- Gives you more than one chance to find an effect. Can increases the false positive rate to **12.6%**.
- How prevalent is it?
  - 27% of researchers admit having done this (John et al., 2012).

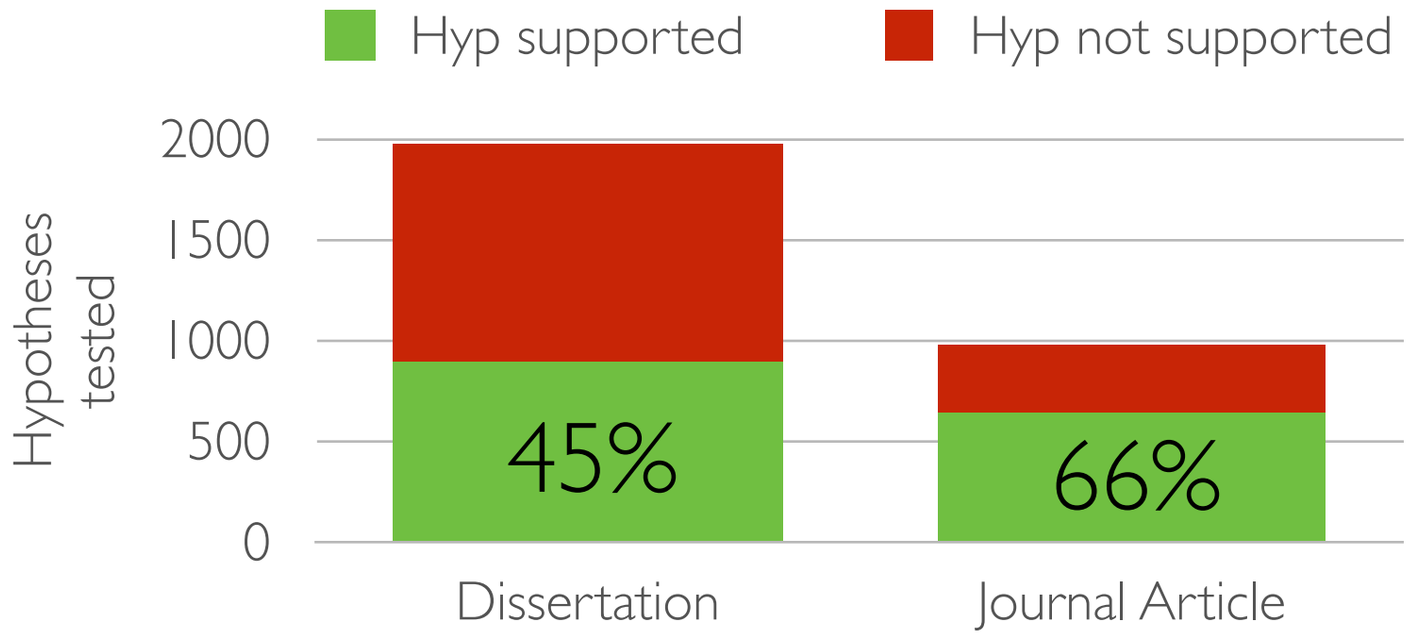
Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychological Science, 22, 1359–1366. doi:10.1177/0956797611417632

Tool 2: Many conditions, report only those that worked

<u>Best-practice example:</u>
Transforming a boring dissertation into a groundbreaking publication

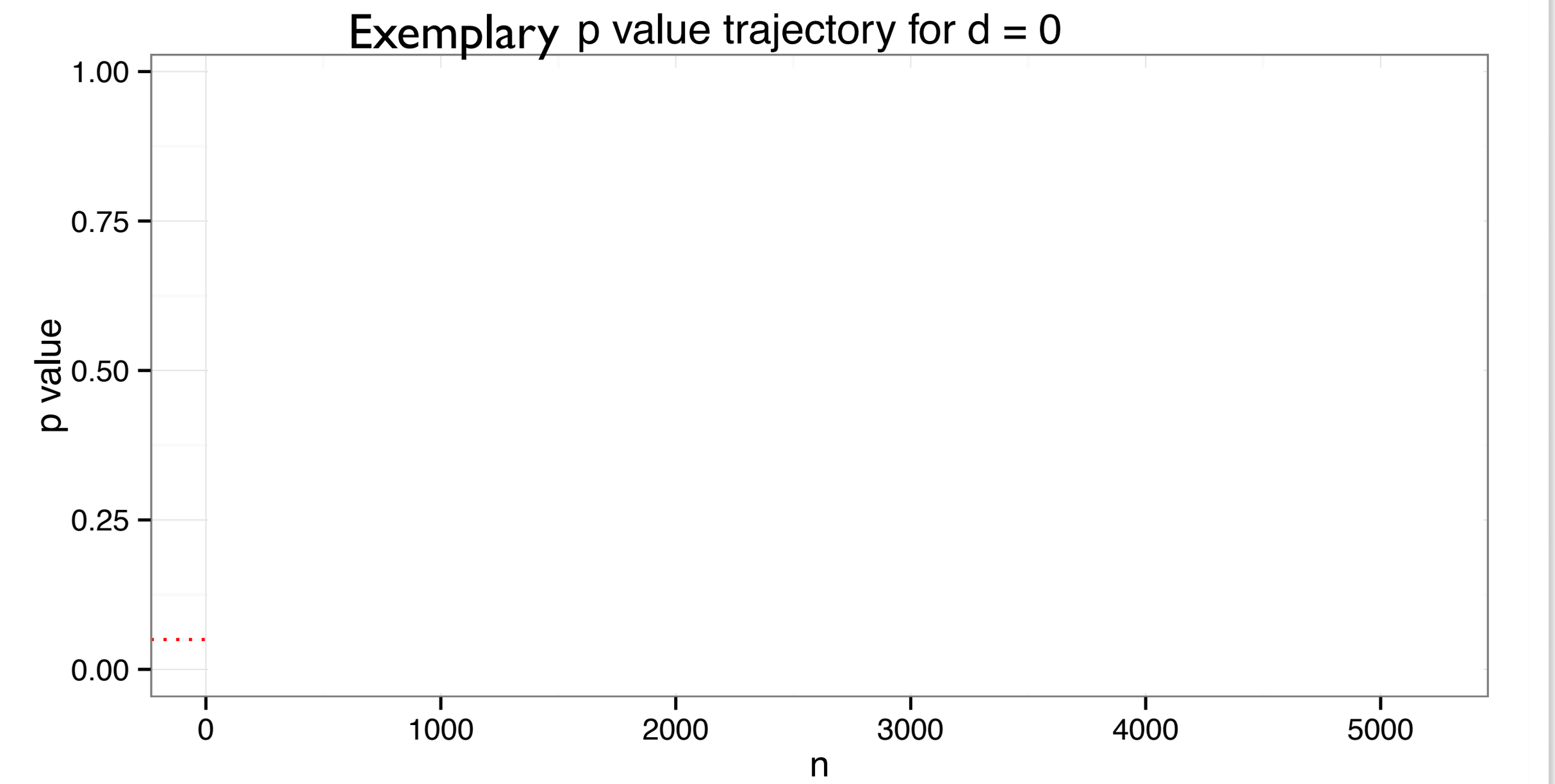
The „Chrysalis effect“ (O’Boyle et al, 2017)

- 142 dissertations that were subsequently published in a refereed journal (149 field studies and 26 experiments)

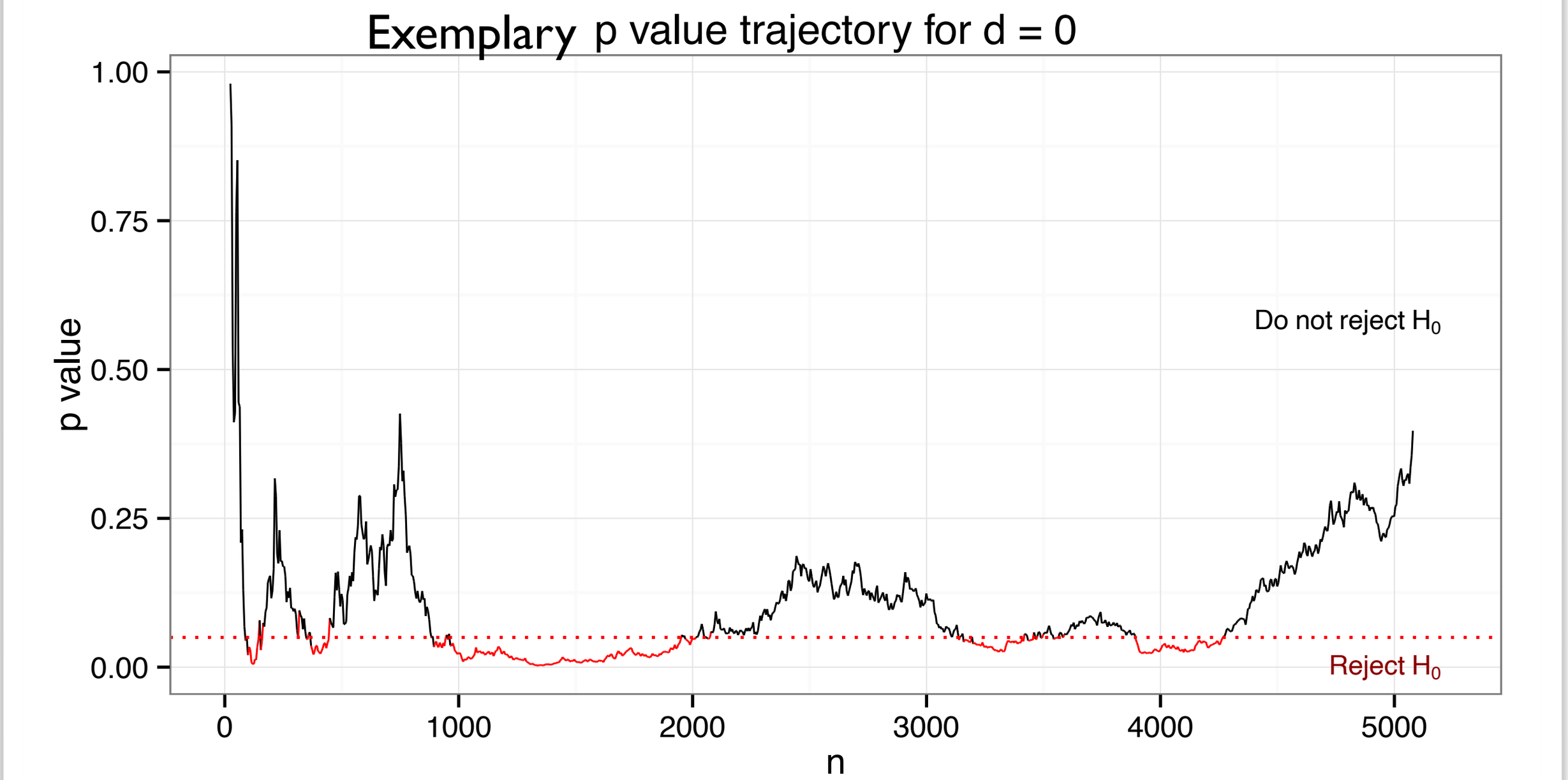


O’Boyle, E. H., Banks, G. C., & Gonzalez-Mulé, E. (2017). The Chrysalis Effect: How Ugly Initial Results Metamorphosize Into Beautiful Articles. Journal of Management, 43(2), 376–399. https://doi.org/10.1177/0149206314527133
https://twitter.com/JoeHilgard/status/699693258386051072

p-values under H_0



Under H_0, p values meander infinitely



Repeated Significance Tests on Accumulating Data

By P. ARMITAGE, C. K. MCPHERSON and B. C. ROWE

Department of Medical Statistics and Epidemiology,
London School of Hygiene and Tropical Medicine

TABLE 2

The probability of being absorbed at or before the nth observation in sampling from a normal distribution with known variance, with repeated tests at a nominal two-sided significance level 2α (i.e. standardized normal deviate k)†

	2α	0·10		0·05		0·02		0·01	
	k	1·645		1·960		2·326		2·576	
n		Q	S	Q	S	Q	S	Q	S
1		0·10000	0·0970	0·05000	0·0545	0·02000	0·0230	0·01000	0·0135
2		0·16015	0·1650	0·08312	0·0885	0·0345			
3		0·20207	0·1980	0·10726	0·1115	0·0456			
4		0·23399	0·2295	0·12617	0·1260	0·0545			
5		0·25963	0·2590	0·14169	0·1420	0·0620			
160		0·63315		0·40829		0·2083			
180		0·64301		0·41677		0·2135			
200		0·65165		0·42429		0·2182			
250		0·670		0·440		0·228			
500		0·720		0·487		0·259			
750		0·746		0·513		0·276			
1,000		0·763		0·53				0·172	

With long enough sampling and optional stopping, it is **guaranteed** to get a significant result!

100%

Armitage, P., McPherson, C. K., & Rowe, B. C. (1969). Repeated significance tests on accumulating data. Journal of the Royal Statistical Society. Series A (General), 132, 235–244.

Tool 3: Optional stopping

- Collect an initial sample, analyze the results, add additional participants if not significant, stop when significance is found
- Increase twice: α = **11%**
- But with enough looks can be pushed to **100%!**

- How prevalent is it?
  - 70% of researchers admit having continued or stopped data collection based on looking at the interim results (John et al., 2012).

Tool 4: Multiple comparisons in ANOVA

- ANOVA, 3 factors, full model
  - 3 main effects, 3 two-way interactions, 1 three-way interaction
- Type I error rate for at least 1 significant term?
- Well-Known: Corrections for post-hoc comparisons of levels within one factor
- Less-known: The need for correcting multiple interactions.

Cramer et al. (2013), Smith, Levine, & Lachlan (2002)

Tool 5: Subgroup analyses

Research question: Do aggressive primes trigger aggressive behavior?

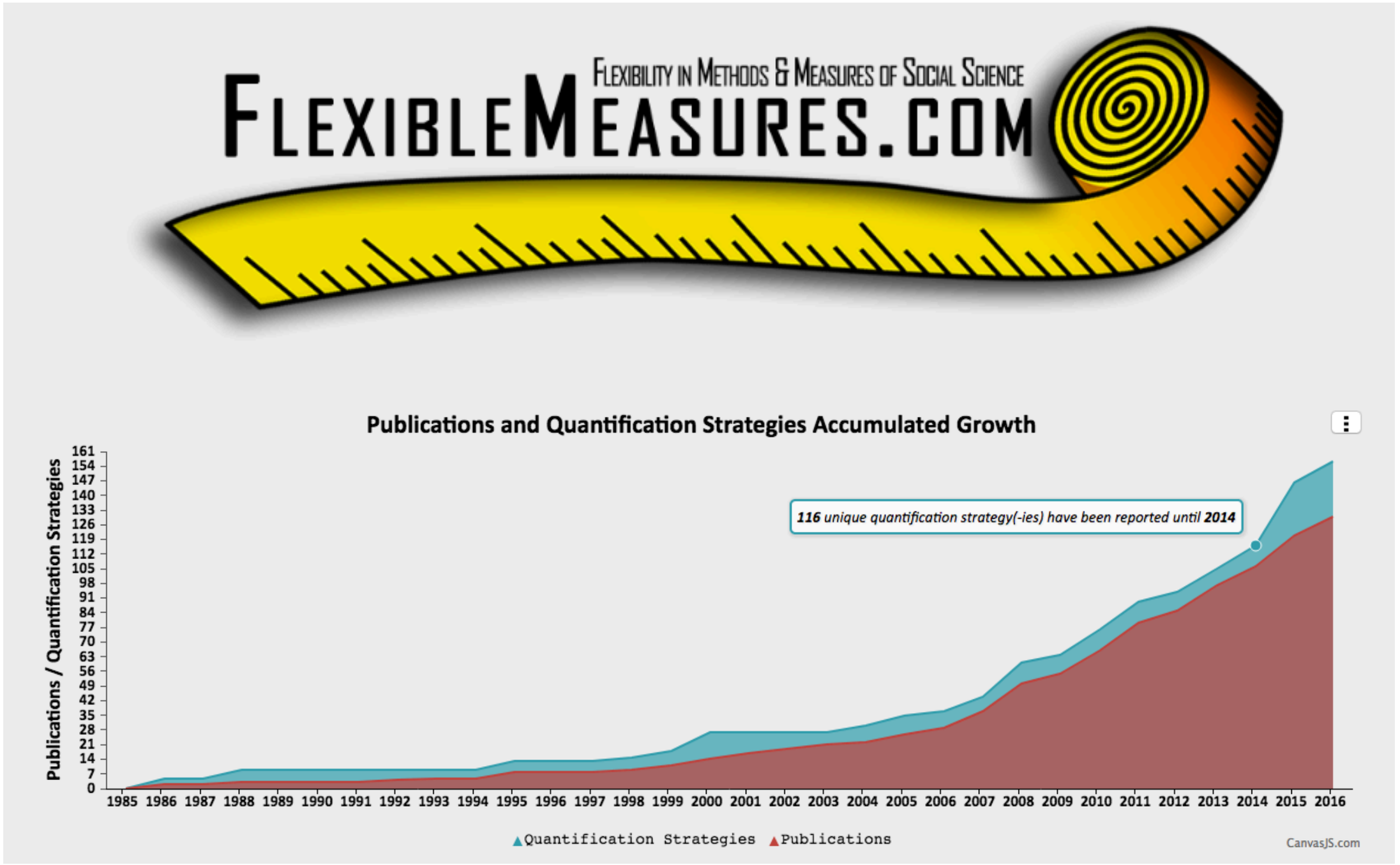
A second study in Turner, Layton, and Simons (1975) collects a larger sample of men and women driving vehicles of all years. **The design was a 2 (Rifle: present, absent) × 2 (Bumper Sticker: "Vengeance", absent) design with 200 subjects.**

➞ presumably, no effect … (yet! Do not give up so easily)

They **divide this further by driver's sex** and by a **median split on vehicle year**. They find that the Rifle/Vengeance condition increased honking relative to the other three, but only among newer-vehicle male drivers, $F(1, 129) = 4.03$, $\boldsymbol{p = .047}$. But then they report that the Rifle/Vengeance condition decreased honking among older-vehicle male drivers, $F(1, 129) = 5.23$, $\boldsymbol{p = .024}$! No results were found among female drivers.

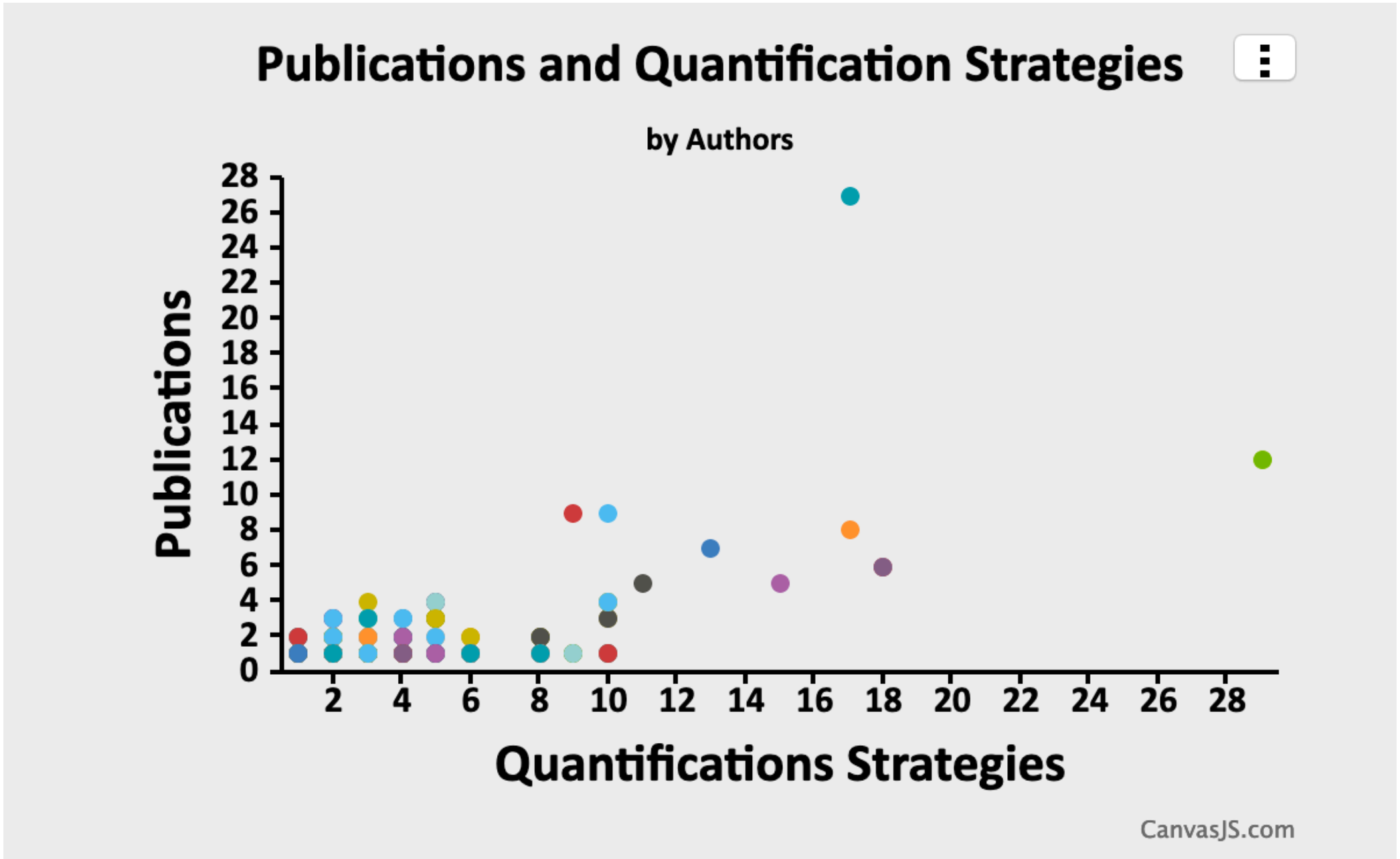
Cited from Joe Hilgard's excellent blog post on the Weapon Priming Effect: http://crystalprisonzone.blogspot.com/2016/03/the-weapons-priming-effect.html; Turner, C. W., Layton, J. F., & Simons, L. S. (1975). Naturalistic studies of aggressive behavior: aggressive stimuli, victim visibility, and horn honking. Journal of Personality and Social Psychology, 31(6), 1098–1107.

Tool 6: Flexible measures



http://www.flexiblemeasures.com/ by Malte Elson

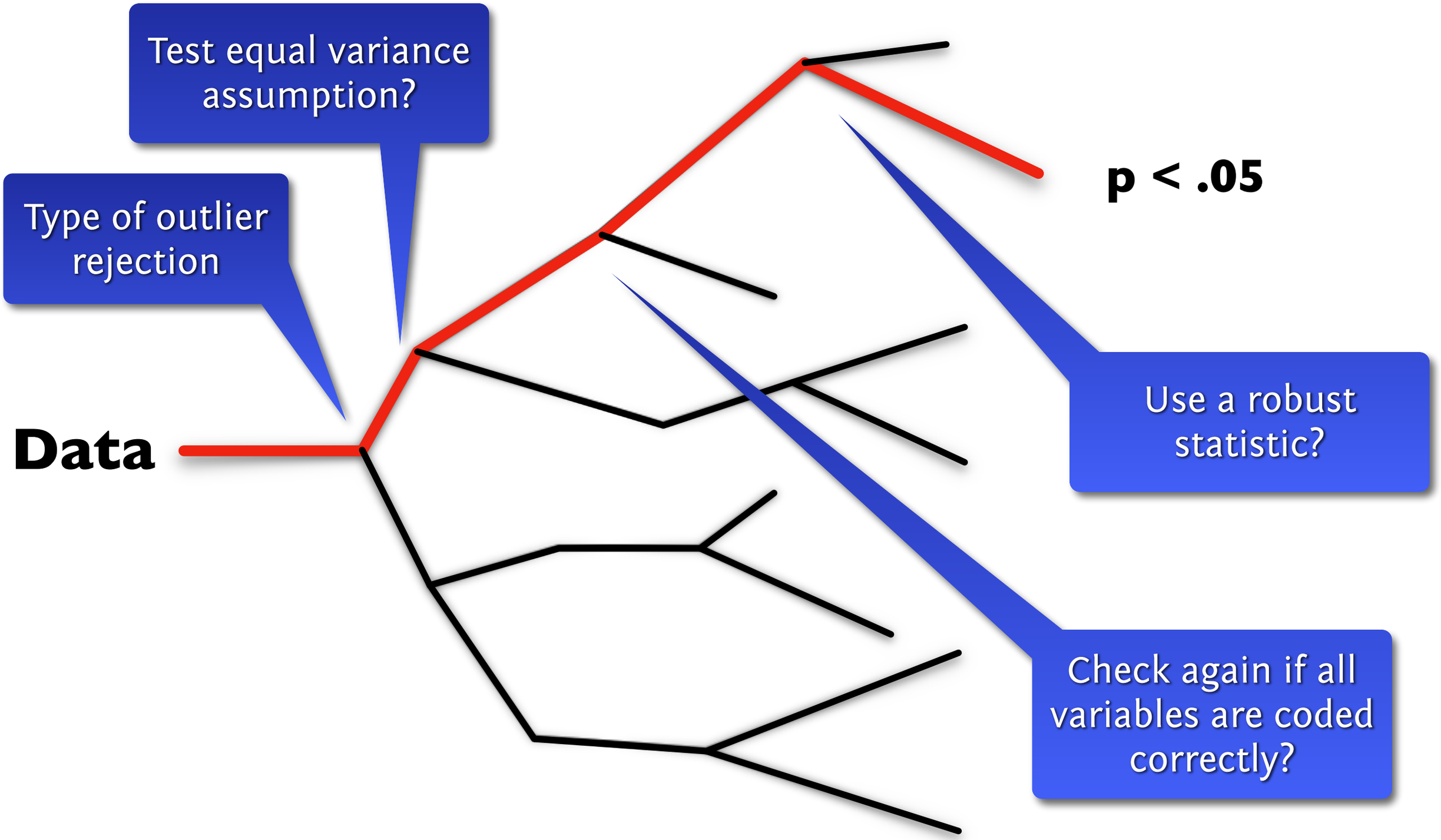
Tool 6: Flexible measures



http://www.flexiblemeasures.com/ by Malte Elson

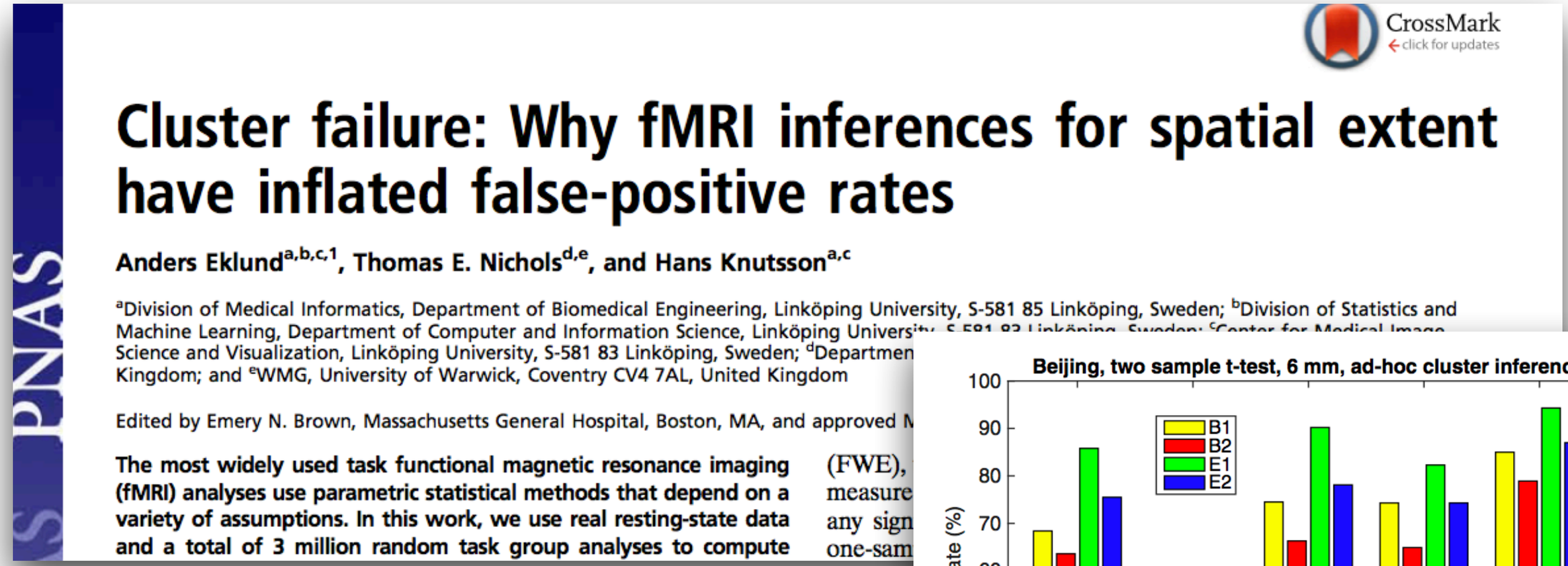
Tool 7: Explore the garden of forking paths

Andrew Gelman & Eric Loken, 2013



Inspired by Neurosceptic's blog: http://blogs.discovermagazine.com/neuroskeptic/2015/05/18/p-hacking-a-talk-and-further-thoughts/#.VV2TiOePKsN

Tool 8: Build the *p*-hacking into the software!



CrossMark click for updates

Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates

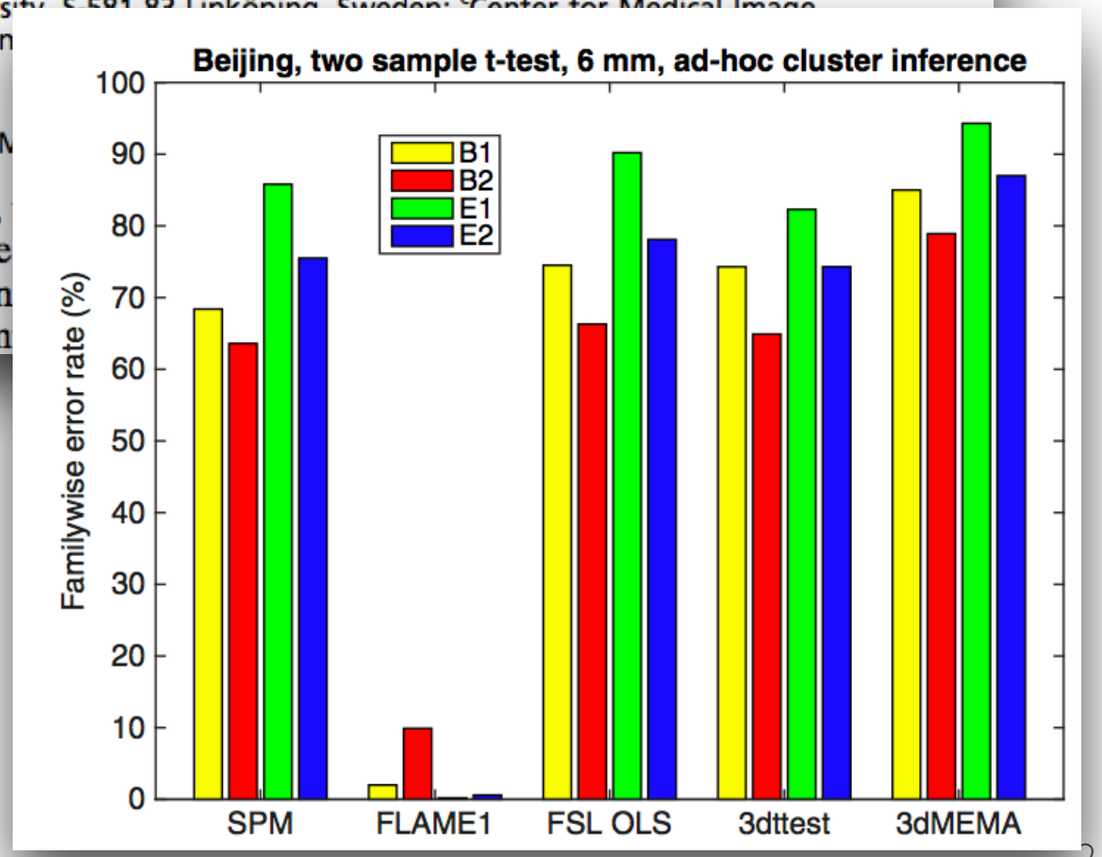
Anders Eklund[a,b,c,1], Thomas E. Nichols[d,e], and Hans Knutsson[a,c]

[a]Division of Medical Informatics, Department of Biomedical Engineering, Linköping University, S-581 85 Linköping, Sweden; [b]Division of Statistics and Machine Learning, Department of Computer and Information Science, Linköping University, S-581 83 Linköping, Sweden; [c]Center for Medical Image Science and Visualization, Linköping University, S-581 83 Linköping, Sweden; [d]Departmen Kingdom; and [e]WMG, University of Warwick, Coventry CV4 7AL, United Kingdom

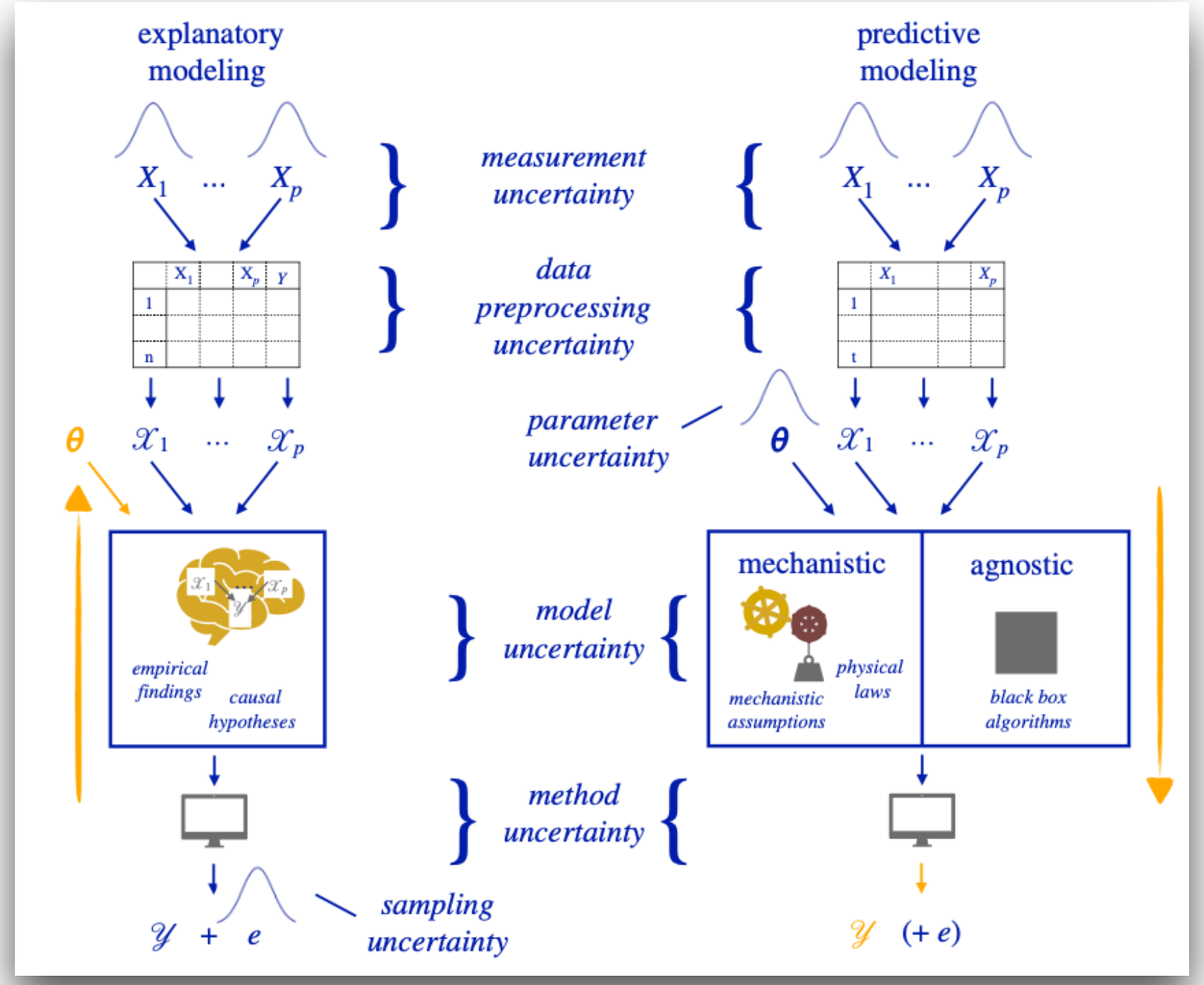
Edited by Emery N. Brown, Massachusetts General Hospital, Boston, MA, and approved M

The most widely used task functional magnetic resonance imaging (fMRI) analyses use parametric statistical methods that depend on a variety of assumptions. In this work, we use real resting-state data and a total of 3 million random task group analyses to compute

(FWE), measure any sign one-sam



The multiplicity of analysis strategies jeopardizes replicability: Lessons learned across disciplines

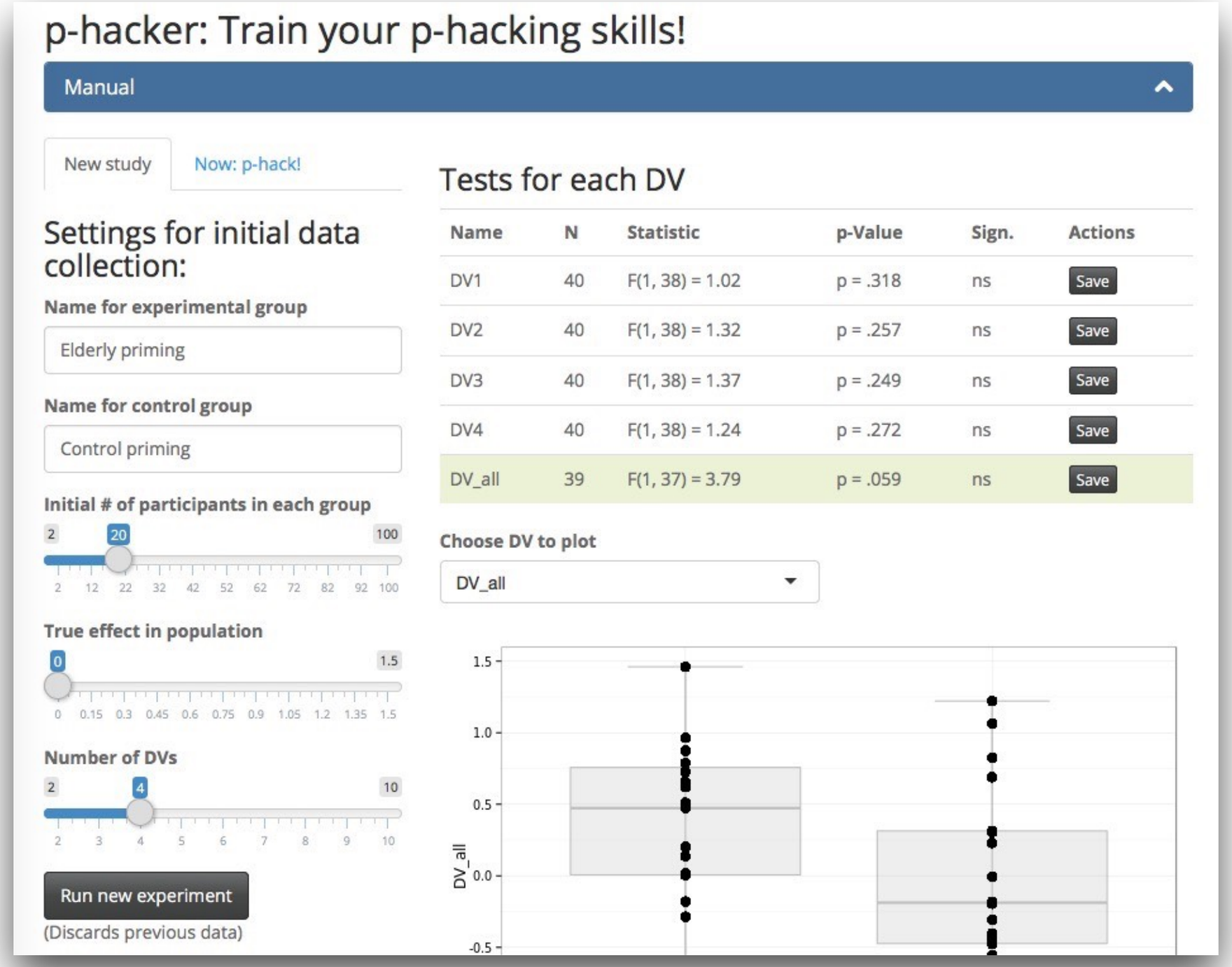


Hoffmann, S., Schönbrodt, F., Elsas, R., Wilson, R., Strasser, U., & Boulesteix, A.-L. (2021). The multiplicity of analysis strategies jeopardizes replicability: Lessons learned across disciplines. *Royal Society Open Science*, *8*(4), 201925. https://doi.org/10.1098/rsos.201925

Freedom is nothing else but a chance to be better.

Albert Camus (1913 - 1960)

Train your skills!



http://shinyapps.org/apps/p-hacker/

The impact of *p*-hacking on the rate of significant results

It is done …

Table 1. Biostatician-Reported Frequency and Severity Rating of Requests for Inappropriate Analysis and Reporting (n = 390)*

Violation Request	Respondents Rating the Item as "Most Severe," %†	Reported Requests During the Past 5 Years, %		
		0	1–9	≥10
Falsify the statistical significance (such as the *P* value) to support a desired result	84	97	2	1
Change data to achieve the desired outcome (such as the prevalence rate of cancer or another disease)	84	93	7	–
Remove or alter some data records (observations) to better support the research hypothesis	80	76	22	2
Interpret the statistical findings on the basis of expectations, not the actual results	68	70	28	2
Do not fully describe the treatment under study because protocol was not exactly followed	62	85	15	–
Do not report the presence of key missing data that could bias the results	68	76	23	1
Ignore violations of assumptions because results may change to negative	64	71	28	1
Modify a measurement scale to achieve some desired results rather than adhering to the original scale as validated	55	79	20	1
Report power on the basis of a post hoc calculation, but make it seem like an a priori statement	54	76	23	2
Request to not properly adjust for multiple testing when "a priori, originally planned secondary outcomes" are shifted to an "a posteriori primary outcome status"	56	80	18	2
Conduct too many post hoc tests, but purposefully do not adjust α levels to make results look more impressive than they really are	54	60	36	4
Remove categories of a variable to report more favorable results	48	68	31	1
Do not mention interim analyses to avoid "too much testing"	50	81	18	1
Report results before data have been cleaned and validated	48	56	39	5
Do not discuss the duration of follow-up because it was inconsistent	45	84	15	1
Stress only the significant findings, but underreport nonsignificant ones	42	45	48	7
Do not report the model statistics (including effect size in ANOVA or R^2 in linear regression) because they seemed too small to indicate any meaningful changes	42	76	23	1
Do not show plot because it did not show as strong an effect as you had hoped	33	58	39	3

ANOVA = analysis of variance.
* Based on findings from questions 1–18 of the Bioethical Issues in Biostatistical Consulting Questionnaire, which asked biostatisticians "to estimate the number of times—during the past 5 years—that you, personally, have been DIRECTLY asked to do this." Data are presented in decreasing order by the percentage of respondents with a perceived severity score of 4 or 5.
† Items were defined as "most severe" if respondents ranked the severity as 4 or 5 on a scale of 0–5.

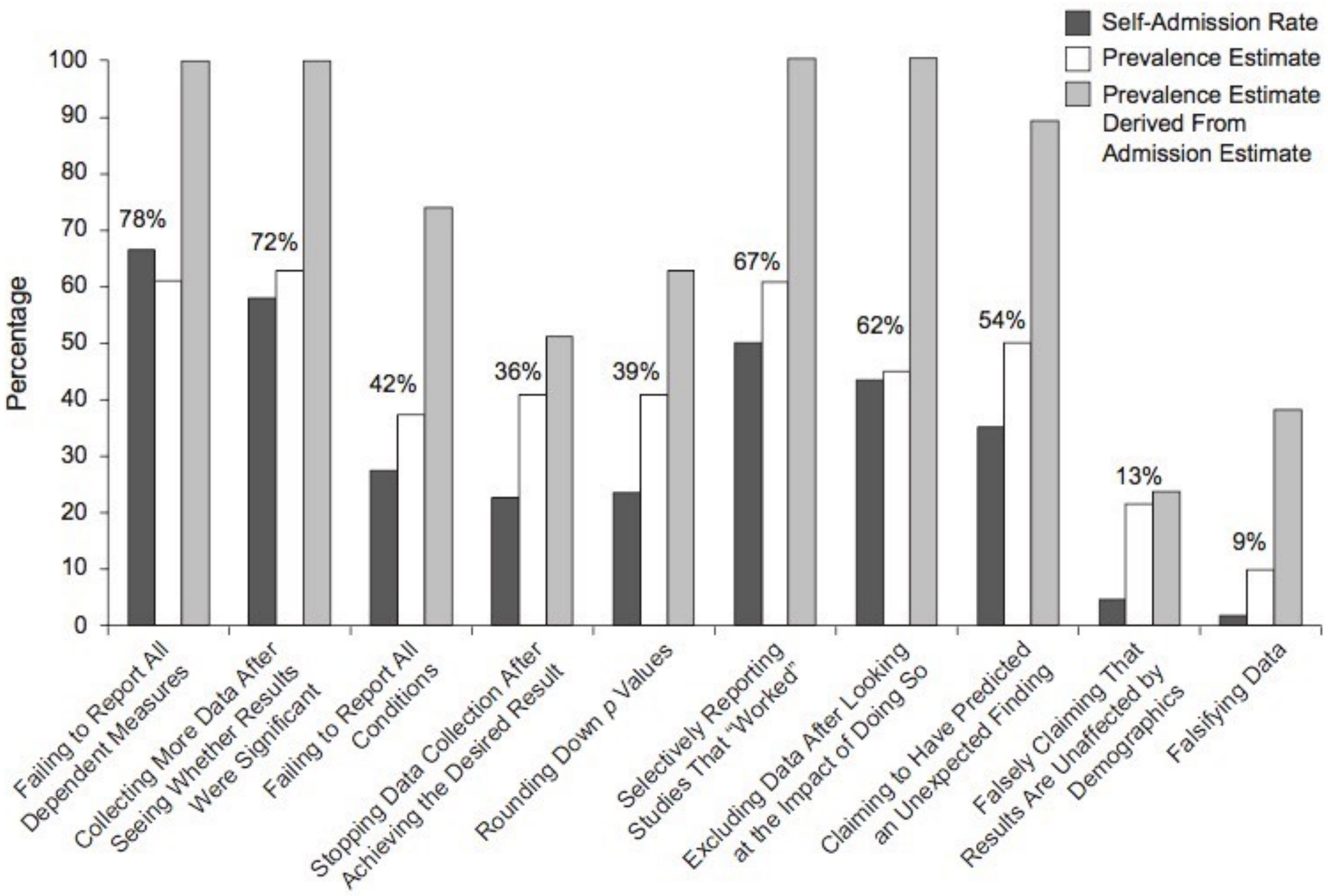
Wang M, Yan AF, & Katz RV. (2018). Researcher requests for inappropriate analysis and reporting: A u.s. survey of consulting biostatisticians. *Annals of Internal Medicine*. doi:10.7326/M18-1230

Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling *in psychology*

Psychological Science
23(5) 524–532

DOI: 10.1177/0956797611430953
http://pss.sagepub.com
SAGE

Leslie K. John[1], George Loewenstein[2], and Drazen Prelec[3]
[1]Marketing Unit, Harvard Business School; [2]Department of Social & Decision Sciences, Carnegie Mellon University; and [3]Sloan School of Management and Departments of Economics and Brain & Cognitive Sciences, Massachusetts Institute of Technology

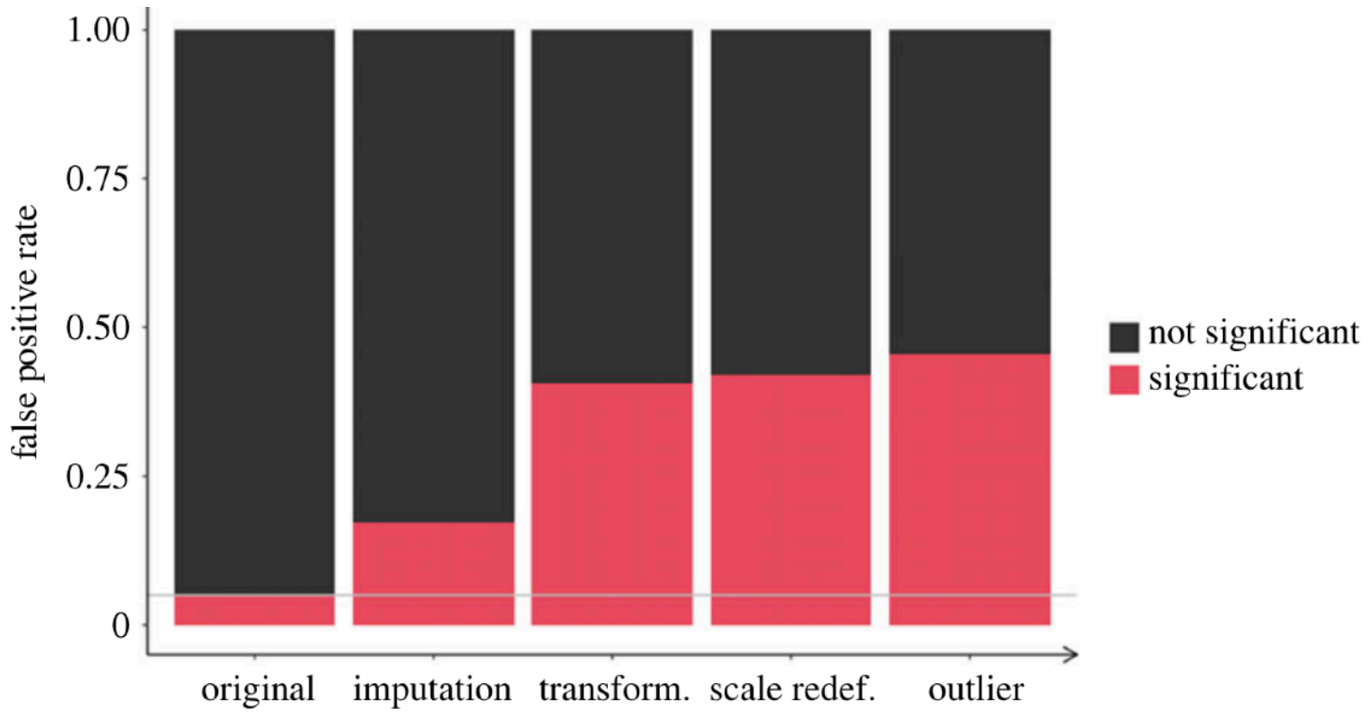


Survey among 6,813 academic researchers in The Netherlands: Self-reported prevalence of fabrication and falsification in the last 3 years

		Disciplinary field			
QRP	**Description (In the last three years.)**	**Life and medical sciences**	**Social and behavioural sciences**	**Natural and engineering sciences**	**Arts and humanities**
Fabrication	Making up of data or results				
Falsification	Manipulating research materials, data or results				
Any FF	Fabrication and/or Falsification				

Gopalakrishna, G., ter Riet, G., Vink, G., Stoop, I., Wicherts, J. M., & Bouter, L. M. (2022). Prevalence of questionable research practices, research misconduct and their potential explanatory factors: A survey among academic researchers in The Netherlands. *PLOS ONE*, *17*(2), e0263023. https://doi.org/10.1371/journal.pone.0263023

How bad can it be?



- Doing some of these *questionable research practices* (QRPs) in combination can raise false positive rate from 5% to **> 50%**!
- QRPs corrupt the logic of the *p*-value and "renders the reported *p*-values essentially uninterpretable."

Stefan, A. M., & Schönbrodt, F. D. (2023). Big little lies: A compendium and simulation of p -hacking strategies. Royal Society Open Science, 10(2), 220346. https://doi.org/10.1098/rsos.220346
Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on p-values: context, process, and purpose. American Statistician, 00–00. http://doi.org/10.1080/00031305.2016.1154108

P-hacking under H_1

- From a statistical point of view, *p*-hacking increases your statistical power

$$Pr(p < .05 \mid H_1, phack) > Pr(p < .05 \mid H_1)$$

- For example:
  - Meta-analysis with $k = 10$ studies, true effect is $\delta = 0.2$, typical sample sizes
  - Power **without** *p*-hacking in primary studies: **53%**
  - Power **with** *p*-hacking in primary studies: **76%**

Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (2019). Correcting for bias in psychology: A comparison of meta-analytic methods. *Advances in Methods and Practices in Psychological Science*. https://doi.org/10.31234/osf.io/9h3nu

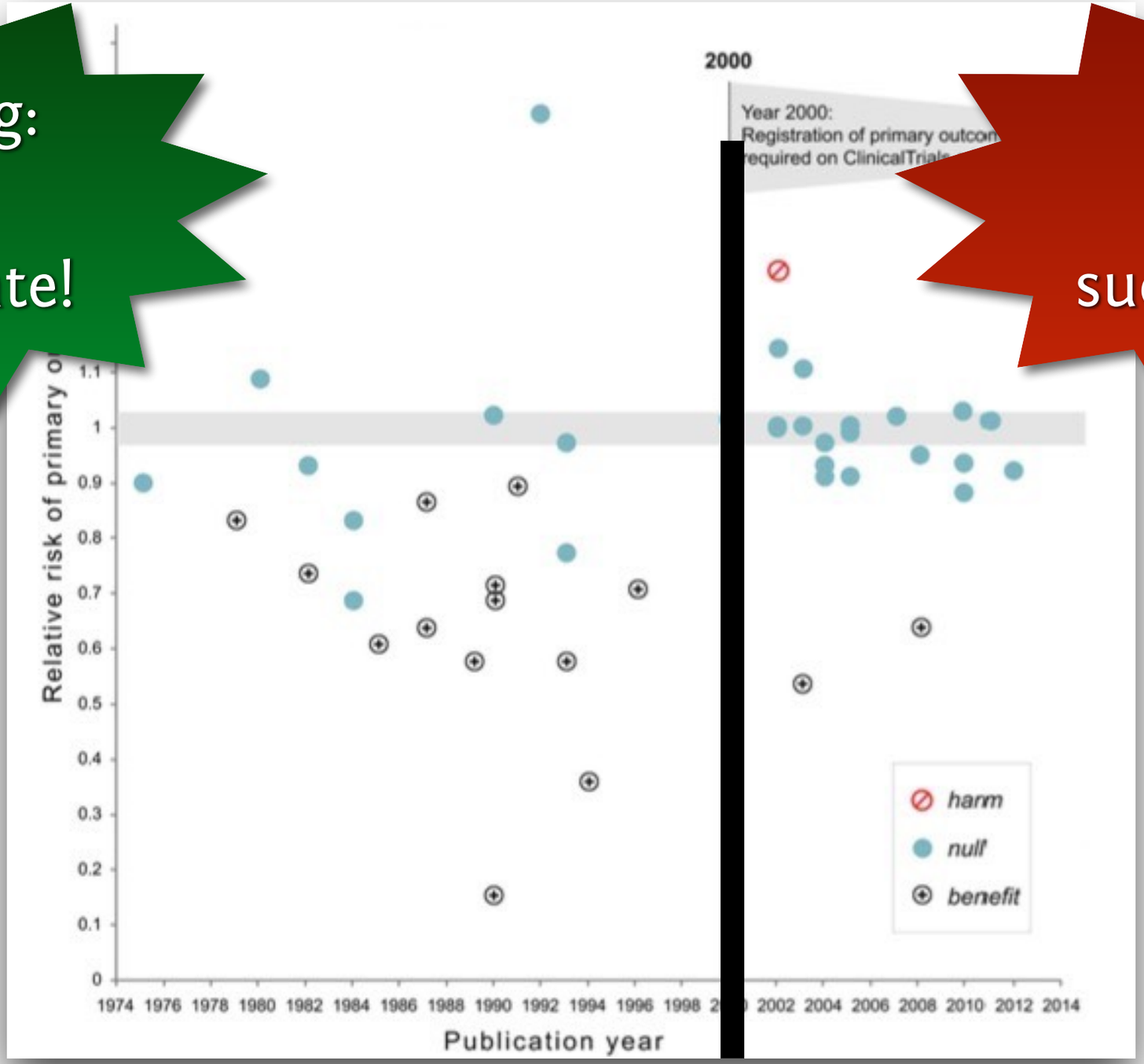
Things to avoid

Anti-tool:
Pre-registration stops *p*-hacking

http://chrisblattman.com/2016/03/01/13719/
Kaplan, R. M., & Irvin, V. L. (2015). Likelihood of Null Effects of Large NHLBI Clinical Trials Has Increased over Time. *PLoS ONE*, *10*(8), e0132382–12. http://doi.org/10.1371/journal.pone.0132382

Tool 9: Do **not** pre-register!

no prereg:
57%
success rate!

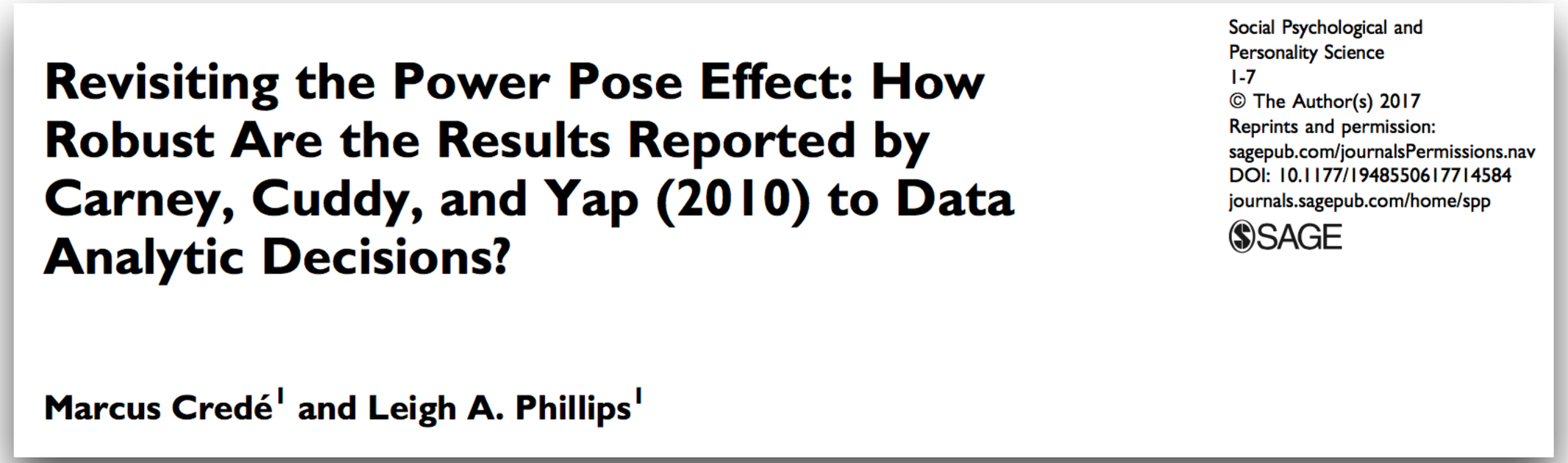


prereg:
8%
success rate...

http://chrisblattman.com/2016/03/01/13719/

Kaplan, R. M., & Irvin, V. L. (2015). Likelihood of Null Effects of Large NHLBI Clinical Trials Has Increased over Time. *PLoS ONE*, *10*(8), e0132382–12. http://doi.org/10.1371/journal.pone.0132382

Tool 10: Do **not** share open data



Revisiting the Power Pose Effect: How Robust Are the Results Reported by Carney, Cuddy, and Yap (2010) to Data Analytic Decisions?

Marcus Credé[1] and Leigh A. Phillips[1]

Social Psychological and Personality Science
1-7

DOI: 10.1177/1948550617714584
journals.sagepub.com/home/spp
SAGE

- A "multiverse analysis" (Steegen, Tuerlinchx, Gelman, & Vanpaemel, 2016): Report results for all plausible analytical decisions
- Check robustness of results: Do several analytical paths lead to comparable conclusions?
- Based on open data by Carney et al. (2010)

Table 1. Multiverse Analysis for the Effect of Power Posing on Testosterone.

		Outlier Identification: Entire Sample ($N = 39$)		Outlier Identification: Test. Conditioned on Gender ($N = 41$)		Outlier Identification: Multivariate or No Exclusion ($N = 42$)	
Gender Effect	Control Variables	DV: T2 Test.	DV: Δ in Test.	DV: T2 Test.	DV: Δ in Test.	DV: T2 Test.	DV: Δ in Test.
Combined	Gender		.047 ($p = .19$)		.019 ($p = .39$)		.036 ($p = .23$)
Combined	Gender and T1 test.	.029 ($p = .31$)		.042 ($p = .21$)		.055 ($p = .15$)	
Combined	Gender and T1 cort.		.045 ($p = .21$)		.017 ($p = .43$)		.018 ($p = .42$)
Combined	Gender, T1 test., and T1 cort.	.037 ($p = .26$)		.040 ($p = .23$)		.043 ($p = .21$)	
Combined	T1 cort. and T2 cort.		.089 ($p = .07$)		.038 ($p = .23$)		.037 ($p = .24$)
Combined	Gender, T1 test., T1 cort., and T2 cort.	**.123 ($p = .04$)**		.099 ($p = .06$)		.102 ($p = .051$)	
Men only	No controls		.192 ($p = .13$)		.047 ($p = .44$)		.096 ($p = .24$)
Men only	T1 test.	.000 ($p = .96$)		.073 ($p = .35$)		.101 ($p = .25$)	
Men only	T1 cort.		.184 ($p = .17$)		.121 ($p = .22$)		.063 ($p = .37$)
Men only	T1 test. and T1 cort.						
Men only	T1 cort. and T2 cort.						.41)
Men only	T1 test., T1 cort., and T2 cort.						
Women only	No controls						.73)
Women only	T1 test.						
Women only	T1 cort.						.75)
Women only	T1 test. and T1 cort.	.023 ($p = .48$)		.023 ($p = .48$)		.023 ($p = .48$)	
Women only	T1 cort. and T2 cort.		.077 ($p = .19$)		.077 ($p = .19$)		.077 ($p = .19$)
Women only	T1 test., T1 cort., and T2 cort.	.167 ($p = .053$)		.167 ($p = .053$)		.167 ($p = .053$)	

cessful efforts to replicate these findings. That is, our results suggest that the data described by Carney et al. (2010), like the data from various unsuccessful replication attempts, are not supportive of a robust effect for power poses. It should, of

Note. Entries are partial η^2 values and (in parentheses) the associated p value. The entry in boldface is the effect for the analyses originally reported in the Carney, Cuddy, and Yap (2010) paper. Blank entries mean that the analyses would not be recommended for reasons described in the text. The number of women was constant across the three outlier strategies. DV = dependent variable; Test. = testosterone; cort. = cortisol; T1 = premanipulation; T2 = postmanipulation.

Of 54 plausible analyses exactly ***one*** was significant.
Guess which has been reported in the original paper?

Open Letter by Dana Carney (2016)

5. Initially, the primary DV of interest was risk-taking. We ran subjects in chunks and checked the effect along the way. It was something like 25 subjects run, then 10, then 7, then 5. Back then this did not seem like p-hacking. It seemed like saving money (assuming your effect size was big enough and p-value was the only issue).
6. Some subjects were excluded on bases such as "didn't follow directions." The total number of exclusions was 5. The final sample size was *N* = 42.
7. The cortisol and testosterone data (in saliva at that point) were sent to Salimetrics (which was in State College, PA at that time). The hormone results came back and data were analyzed.
8. For the risk-taking DV: One p-value for a Pearson chi square was .052 and for the Likelihood ratio it was .05. The smaller of the two was reported despite the Pearson being the more ubiquitously used test of significance for a

10. The self-report DV was p-hacked in that many different power questions were asked and those chosen were the ones that "worked."

http://faculty.haas.berkeley.edu/dana_carney/pdf_my%20position%20on%20power%20poses.pdf

https://www.youtube.com/watch?v=ZaNtz76dNSI&sns=em

(Un)Intentional?

- Intentional?
  - Evil researcher who only cares about his/her career and not at all about truth-seeking?
- Unintentional?
  - Wrong education?
  - Wrong/uncritical standards of the field?
  - Pushed by supervisors, reviewers, or editors?
    ➙ http://bulliedintobadscience.org/
- My guess is that really most *p*-hacking occurs unintentional / in best faith.
- Distorting effects on the published record are probably comparable, but the ethical evaluations differs strongly.

Lower the threshold?


comment

Redefine statistical significance

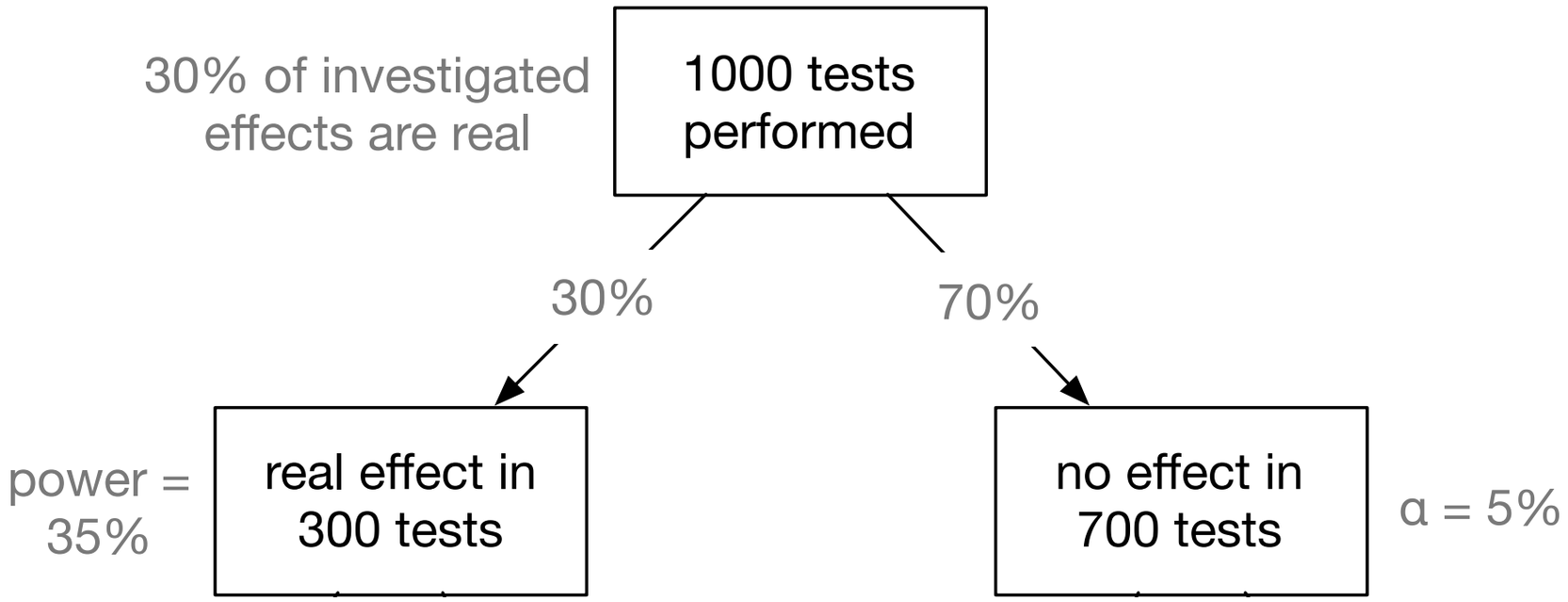
We propose to change the default *P*-value threshold for statistical significance from 0.05 to 0.005 for claims of new discoveries.

- Suggestion:
  Lower the default threshold from .05 to .005
- $p < .005$ ➞ „significant“
- $.005 < p < .05$ ➞ „suggestive“
- $p > .05$ ➞ „not significant“

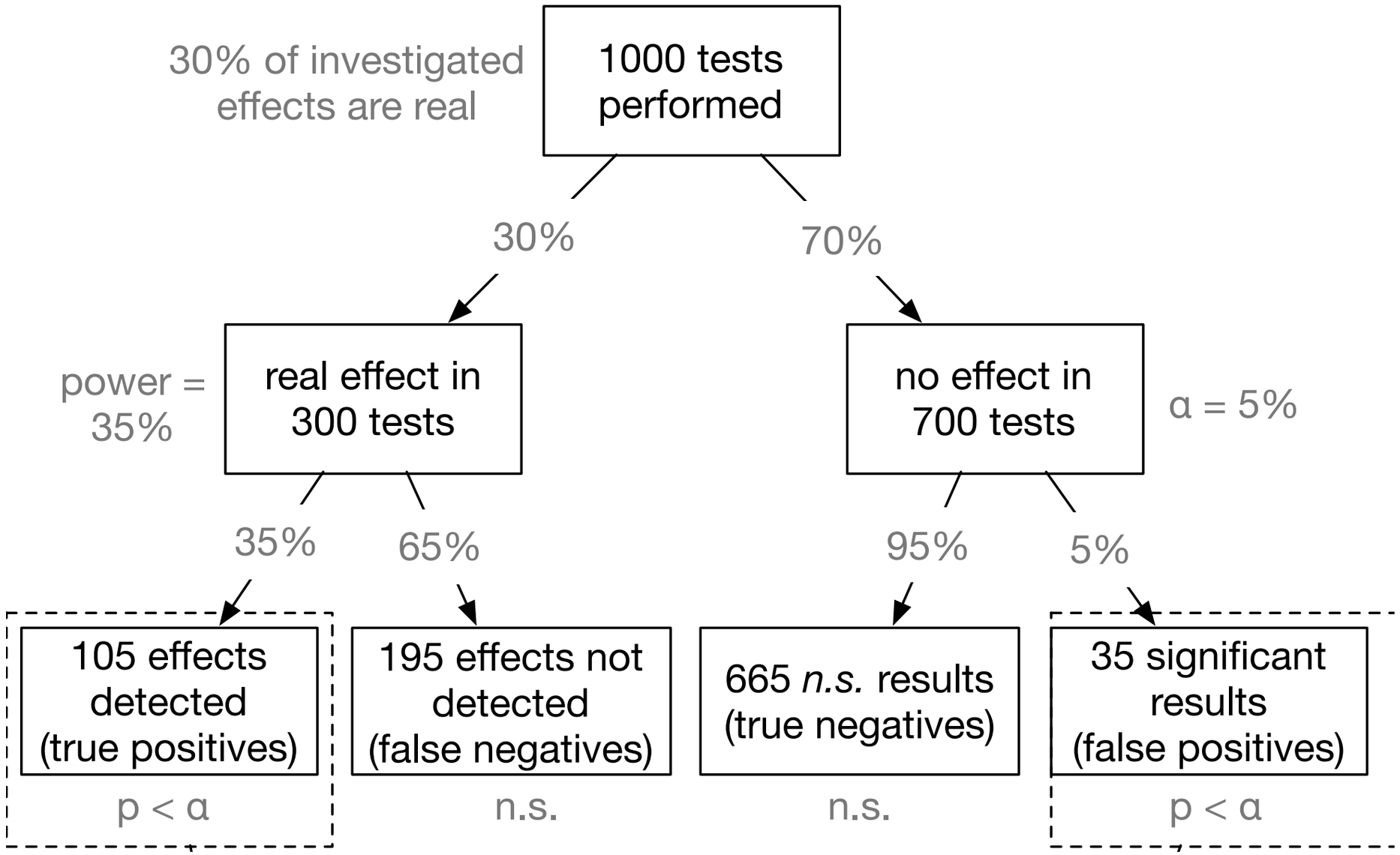
Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E. J., Berk, R., et al. (2017). Redefine statistical significance. Nature Human Behaviour, 1–0. http://doi.org/10.1038/s41562-017-0189-z

Congratulations:
You hacked your results to significance!

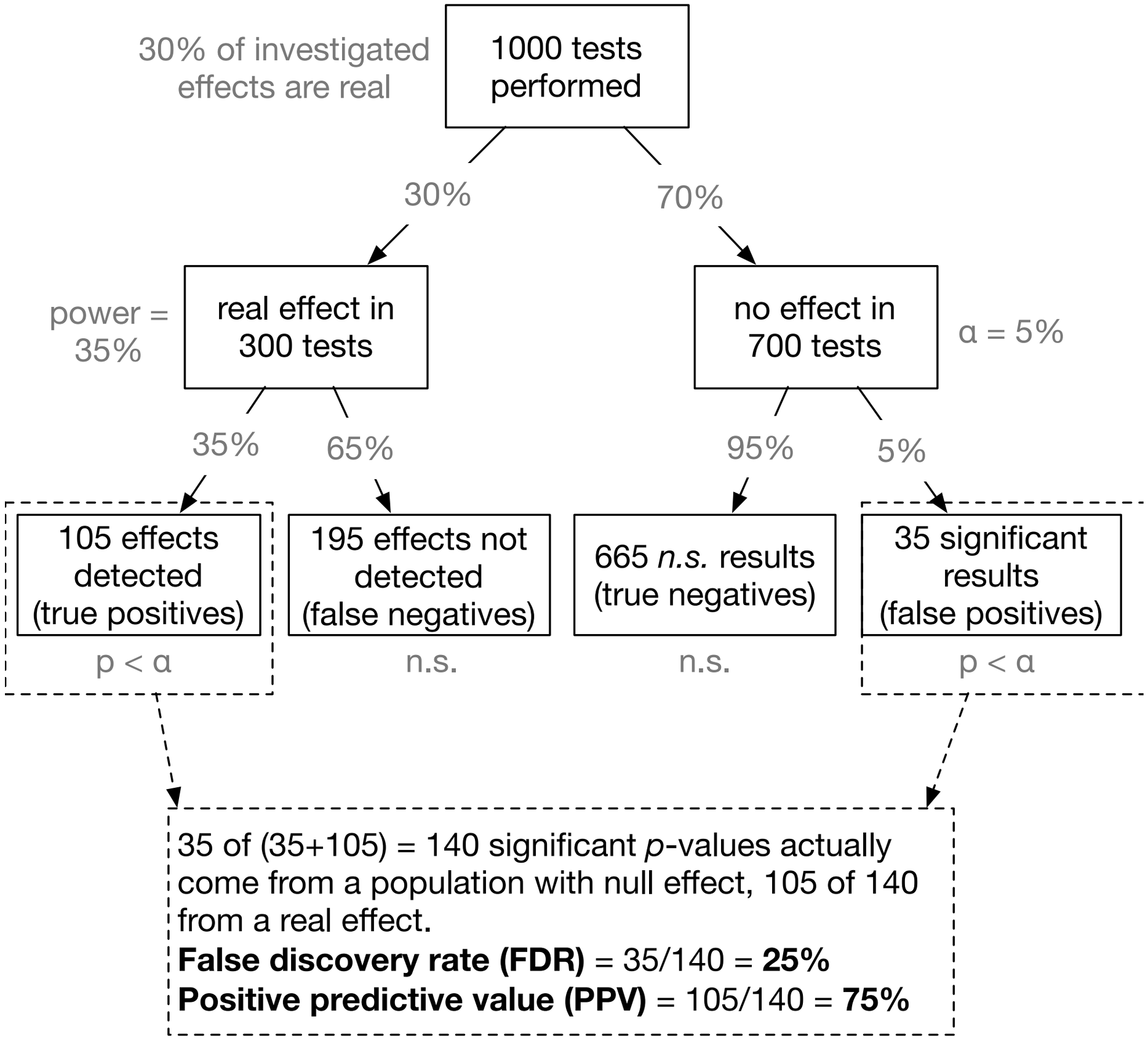
Now: What's the probability that a significant *p*-value indicates a true effect?



Nuzzo, R. (2014). Statistical errors. *Nature*.
Colquhoun, D. (2014). An investigation of the false discovery rate and the misinterpretation of p-values. Royal Society Open Science, 1(3), 140216–140216. http://doi.org/10.1073/pnas.1313476110



Nuzzo, R. (2014). Statistical errors. *Nature*.
Colquhoun, D. (2014). An investigation of the false discovery rate and the misinterpretation of p-values. Royal Society Open Science, 1(3), 140216–140216. http://doi.org/10.1073/pnas.1313476110



Nuzzo, R. (2014). Statistical errors. *Nature*.
Colquhoun, D. (2014). An investigation of the false discovery rate and the misinterpretation of p-values. Royal Society Open Science, 1(3), 140216–140216. http://doi.org/10.1073/pnas.1313476110

When does a significant *p*-value indicate a true effect?

Understanding the Positive Predictive Value (PPV) of a *p*-value

Across all investigated hypotheses: What % of them is actually true?

% of a priori true hypotheses:

0 | 10 | 100

0 10 20 30 40 50 60 70 80 90 100

What is your Type I error (α; typically 5%)?

α level

0.01 | 0.05 | 0.1

0.01 0.02 0.03 0.04 0.05 0.06 0.07 0.08 0.09 0.1

On what power level are the studies conducted?

Power

0.01 | 0.8 | 0.99

0.01 0.11 0.21 0.31 0.41 0.51 0.61 0.71 0.81 0.91 0.99

% of studies that report a significant result, although it's not ?

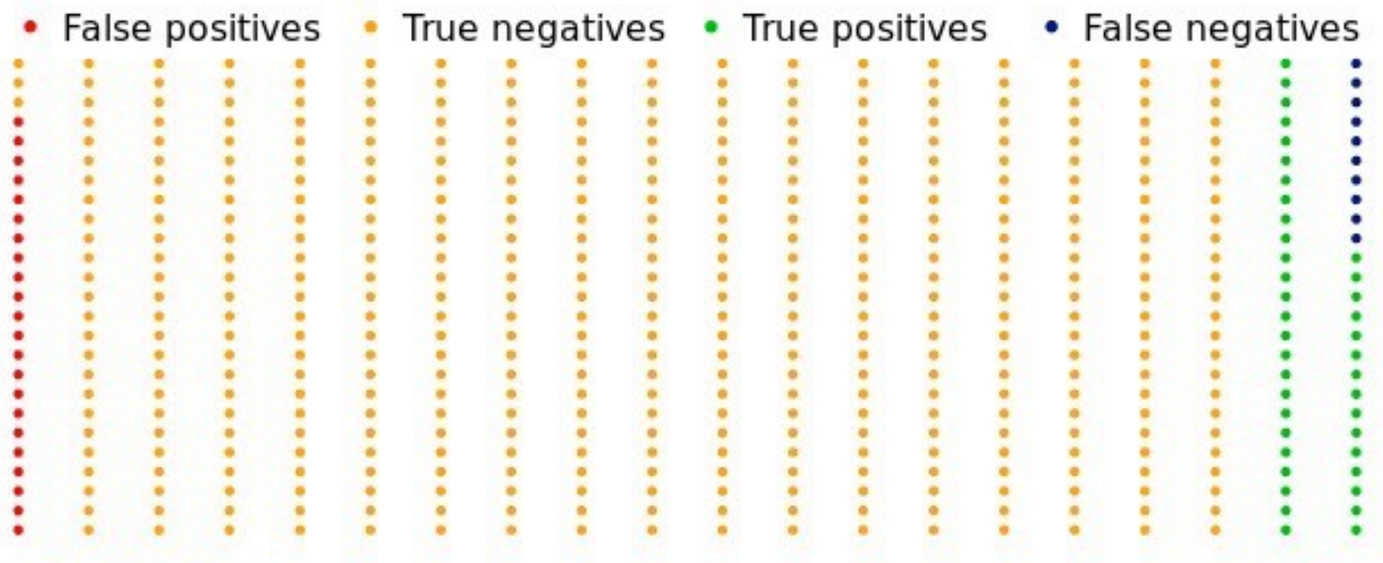
% of p-hacked studies

0 | 100

0 10 20 30 40 50 60 70 80 90 100

Positive predictive rate (PPV): 64% of claimed findings are true
False discovery rate (FDR): 36% of claimed findings are false

If we consider all findings, it looks like this (each point is one study):



If we consider **only the significant** findings, the ratio of true to false positives looks like this:

Practice with the PPV app! http://shinyapps.org/apps/PPV/

** My lawyer told me to show that.*

Disclaimer*

- *p*-hacking increases the false positive rate
- *p*-hacking „renders the reported *p*-values essentially uninterpretable“ (ASA statement)
- *p*-hacking is ethically wrong and violates rules of good scientific practice
- If you *p*-hack systematically:
  - many of your research results will simply be wrong (depending on the prior probability of your hypotheses)
  - consequentially, your research won't replicate
- Every time you *p*-hack, you waste public money, you waste participants' time, you bias the literature, and **a kitten dies****.

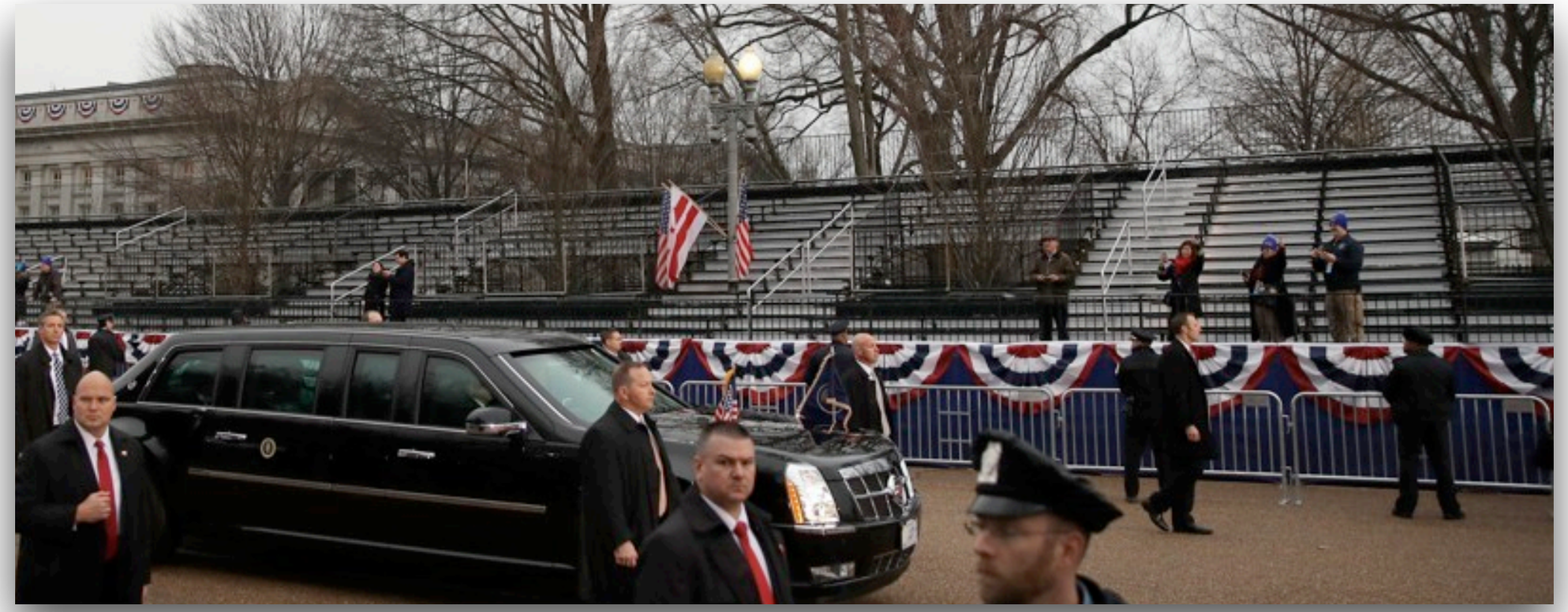
** *If your research is about feline drug development*

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on p-values: context, process, and purpose. American Statistician, 00–00. http://doi.org/10.1080/00031305.2016.1154108

Gaming the system

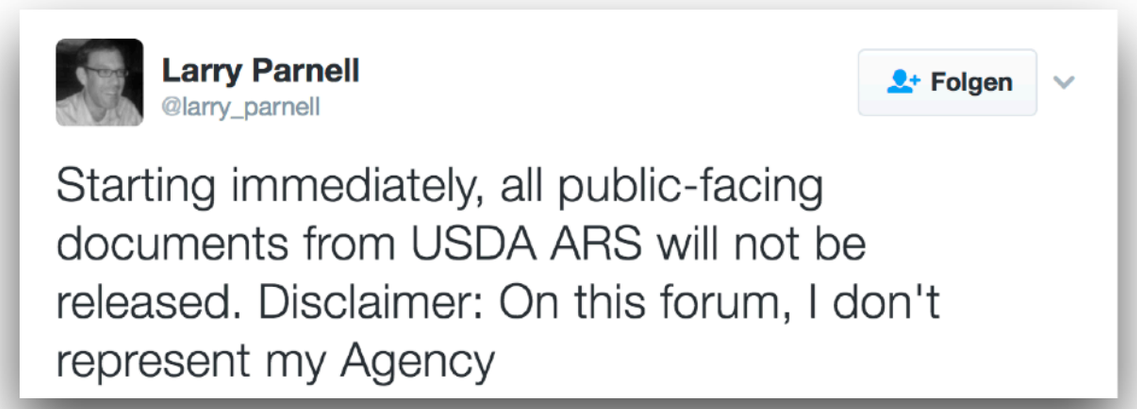
or

Producing real knowledge?



„This was the largest audience ever to witness an inauguration, period.“

Press Secretary Sean Spicer



post-truth!

Fake-News!

https://www.theatlantic.com/politics/archive/2017/01/the-pointless-needless-lies-of-the-trump-administration/514061/

Trump's budget director pick: “Do we really need government-funded research at all”

Mick Mulvaney suggested Zika science is uncertain, so we shouldn't bother to fund it.

Updated by Julia Belluz | @juliaoftoronto | julia.belluz@voxmedia.com | Jan 24, 2017, 11:07am EST

TWEET SHARE

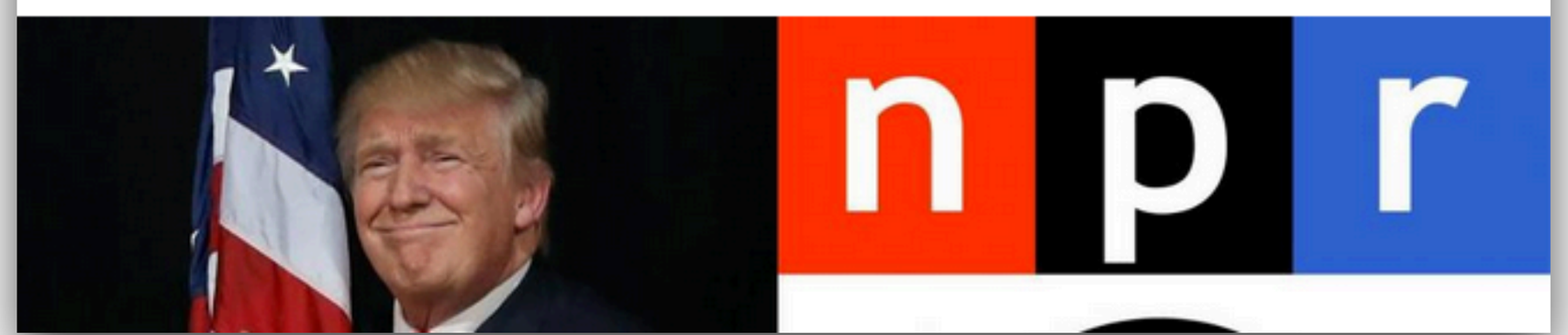
http://www.vox.com/science-and-health/2016/12/
budget-director-research-science-mulvaney

Culture

Donald Trump set to ‘eliminate arts funding programs’, cutting off NPR and PBS

He is notoriously not fond of broadcast media

Christopher Hooton | @christophhooton | 2 months ago | 319 Gefällt Click to follow The Independent Culture



http://www.independent.co.uk/arts-entertainment/donald-trump-budget-cuts-arts-humanities-nea-neh-npr-cbs-president-a7536741.html

Norms

Communality

Open sharing

Universalism

Evaluate research on own merit

Disinterestedness

Motivated by knowledge and discovery

Organized skepticism

Consider all new evidence, even against one’s prior work

Quality

Counternorms

Secrecy

Closed

Particularlism

Evaluate research by reputation

Self-interestedness

Treat science as a competition

Organized dogmatism

Invest career promoting one’s own theories, findings

Quantity

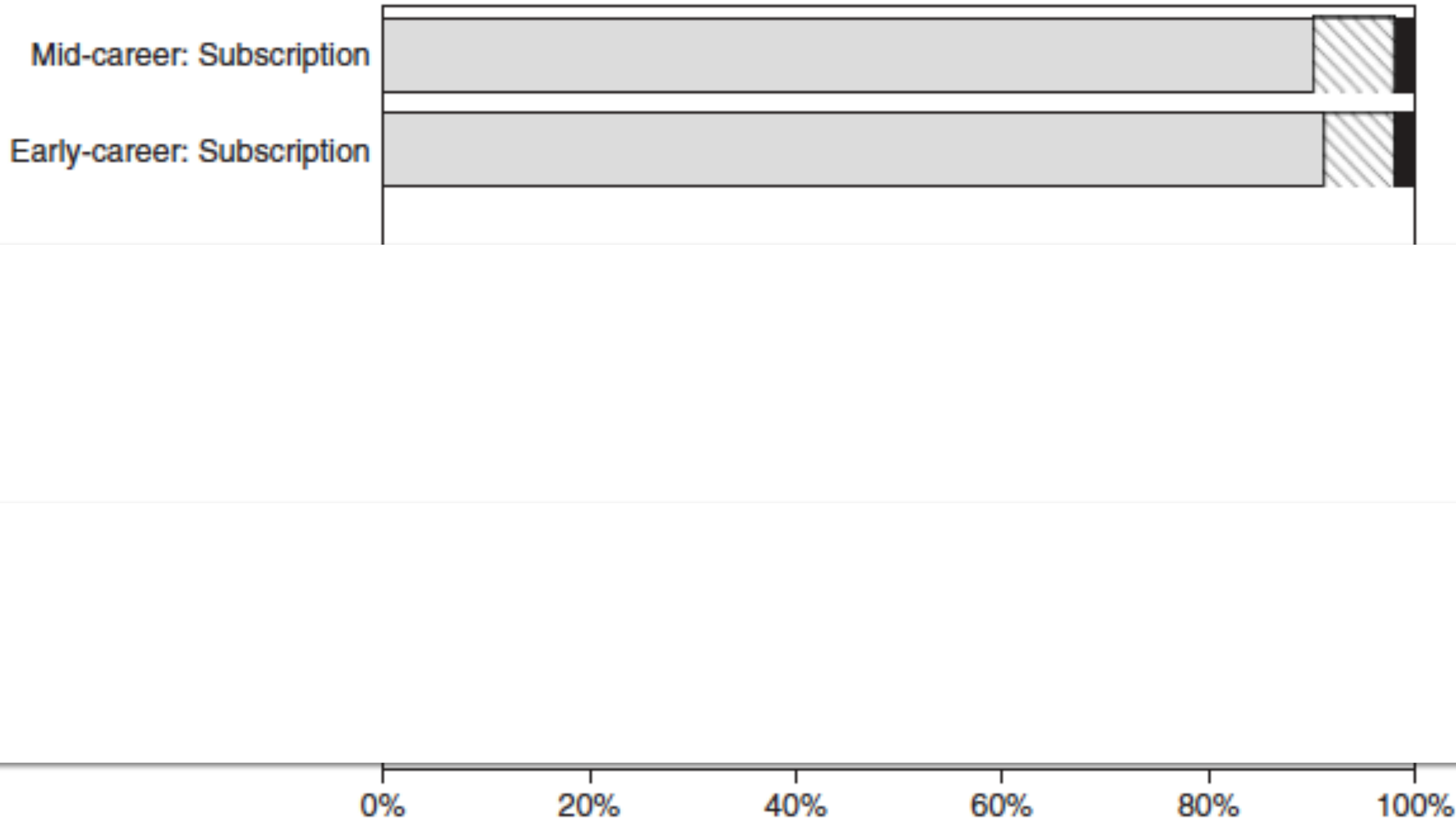


FIG. 3. Norm versus Counternorm Scores: Percent with Norm > Counternorm (dotted), Norm = Counternorm (striped), Norm < Counternorm (solid).

Anderson, Martinson, & DeVries, 2007

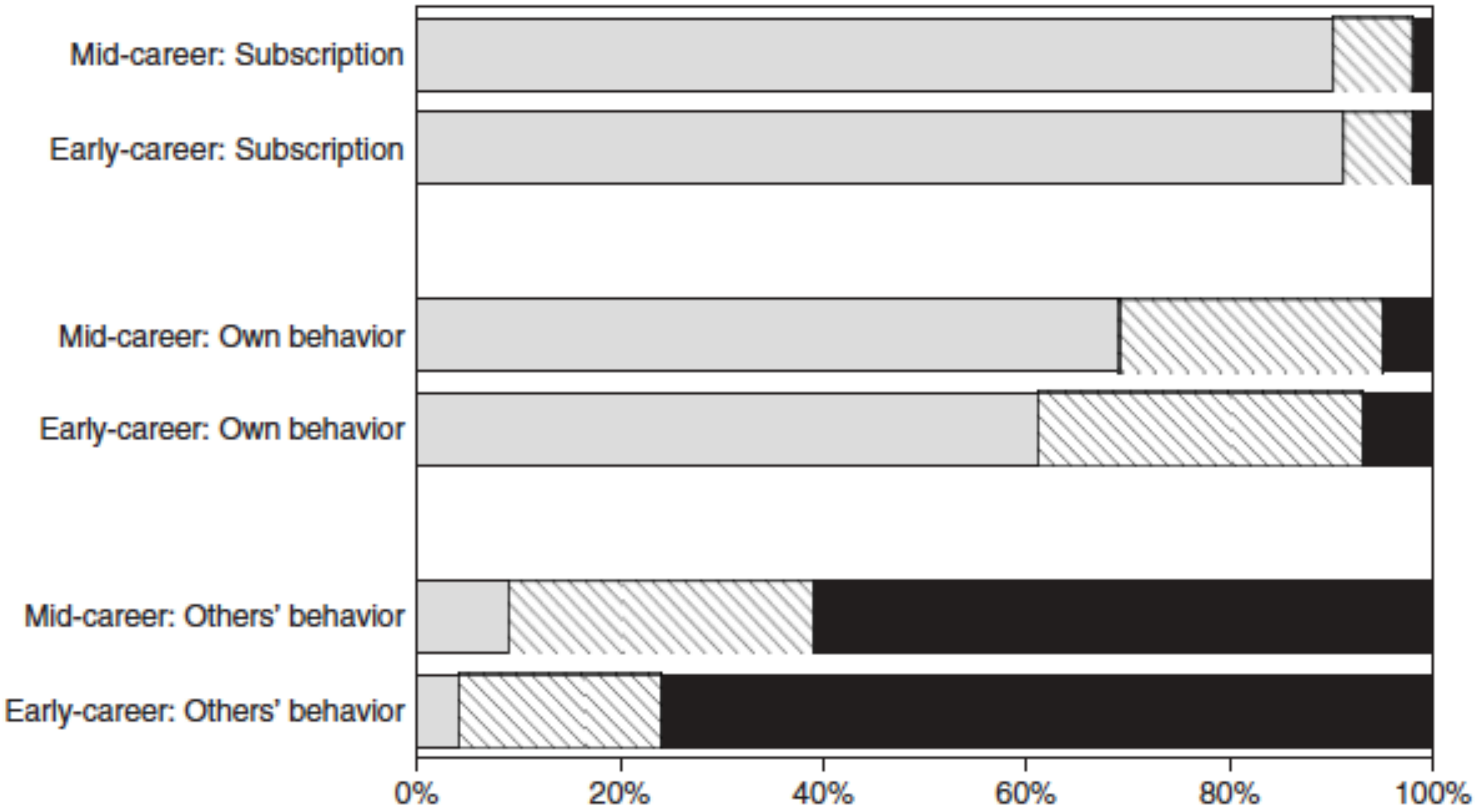


FIG. 3. Norm versus Counternorm Scores: Percent with Norm > Counternorm (dotted), Norm = Counternorm (striped), Norm < Counternorm (solid).

Anderson, Martinson, & DeVries, 2007