Anne Gärtner
Faculty of Psychology

Workshop Open Science Practices
Part 1
P-hacking

Overview

Time	Topic
13:00 – 13:10	Welcome
13:10 – 13:40	Power (Alex)
13:40 – 13:55	Discussion
13:55 – 14:40	Data Collection (Anne)
14:40 – 14:55	Break
14:55 – 15:40	*P*-hacking (Anne)
15:40 – 16:10	Publication Bias (Alex)
16:10 – 16:25	Discussion
16:25 – 16:40	Break
16:40 – 17:25	Preregistration (Anne)
17:25 – 17:40	Discussion
17:40 – 18:00	Wrap Up, Evaluation

Workshop material

Zoom Poll

Outline

What is *p*–hacking?

How to *p*-hack: Tools

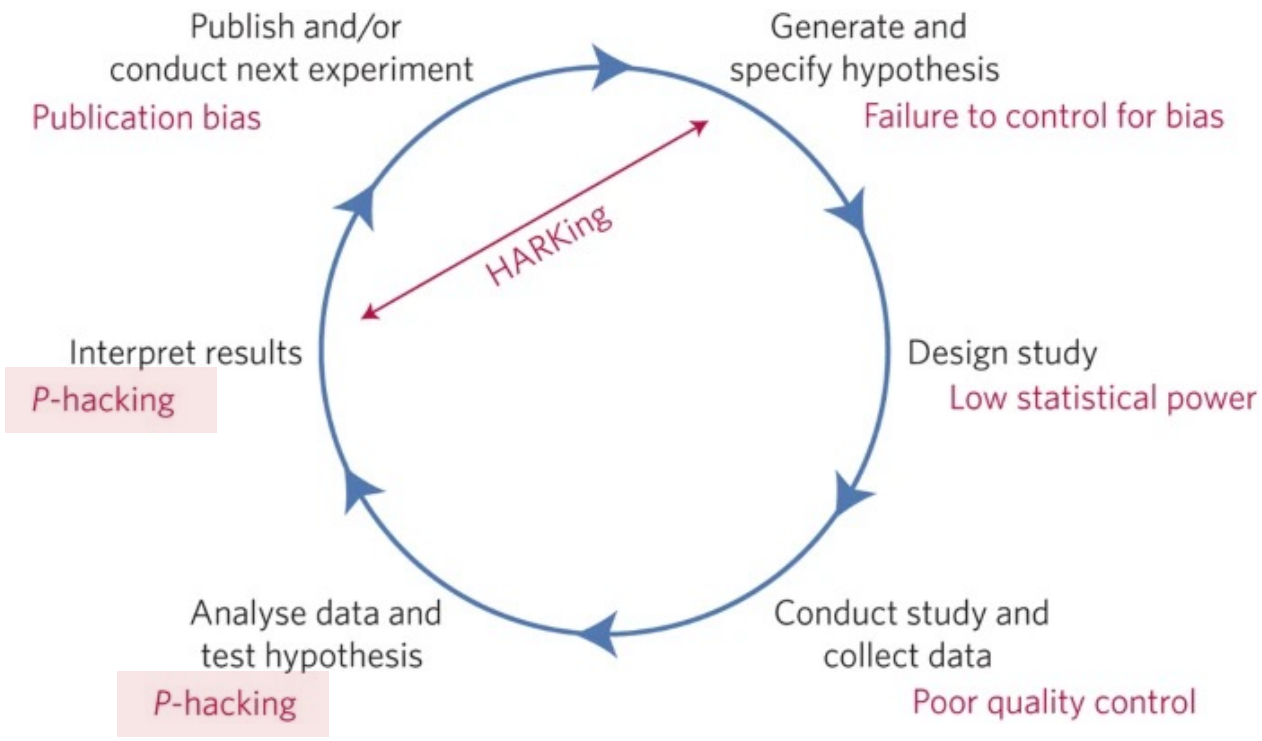
— Outcome switching, selective reporting, optional stopping, subgroup analysis, flexible measures, no preregistration

Exercise

— Train with *p*-hacker

Discussion

Solutions and further recommendations



What is *p*-hacking?

Introduction

A tale of two papers (by Michael Inzlicht)

Original version

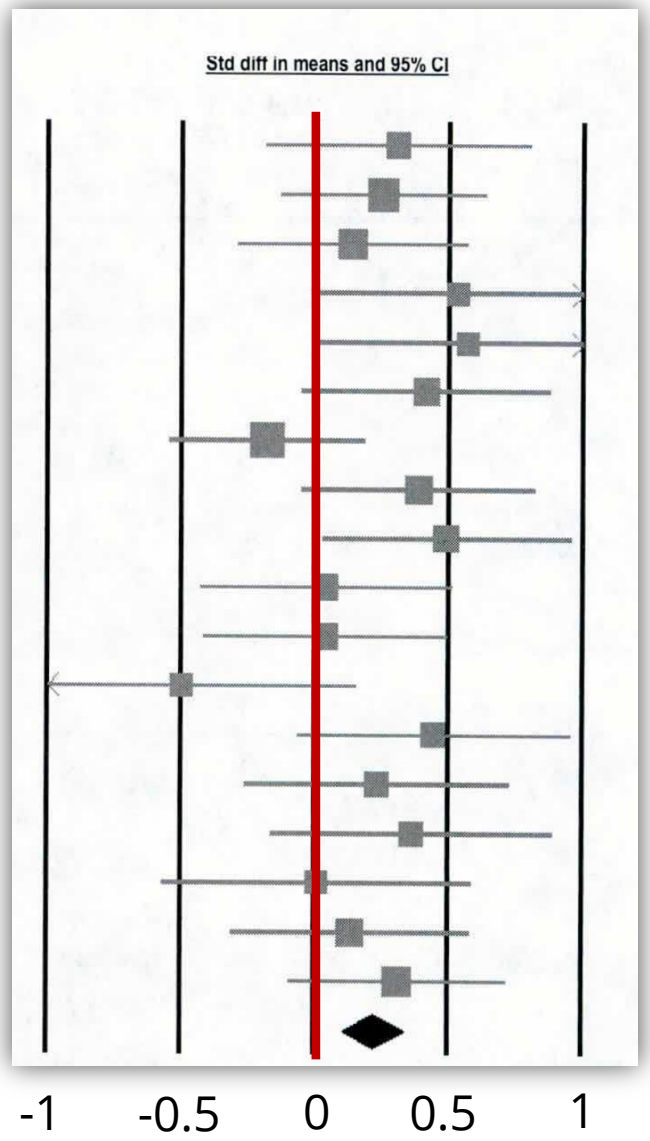
- 7 experiments
- 7/7 significant
- effect sizes in the medium to large range
- ad-hoc covariates
- "Excessive significance"

"The first [paper] was emblematic of the old way of doing business, with 7 studies that were scrubbed clean to be near-perfect."

Revised version

- 18 experiments
- 2/18 significant
- Some studies with reversed direction

"This is what real data look like. The data are not always pretty, they have warts, but they are real."



Blog Post | Tuk, Zhang, & Sweldens (2016). *J Exp Psychol: Gen, 144*, 639–654.

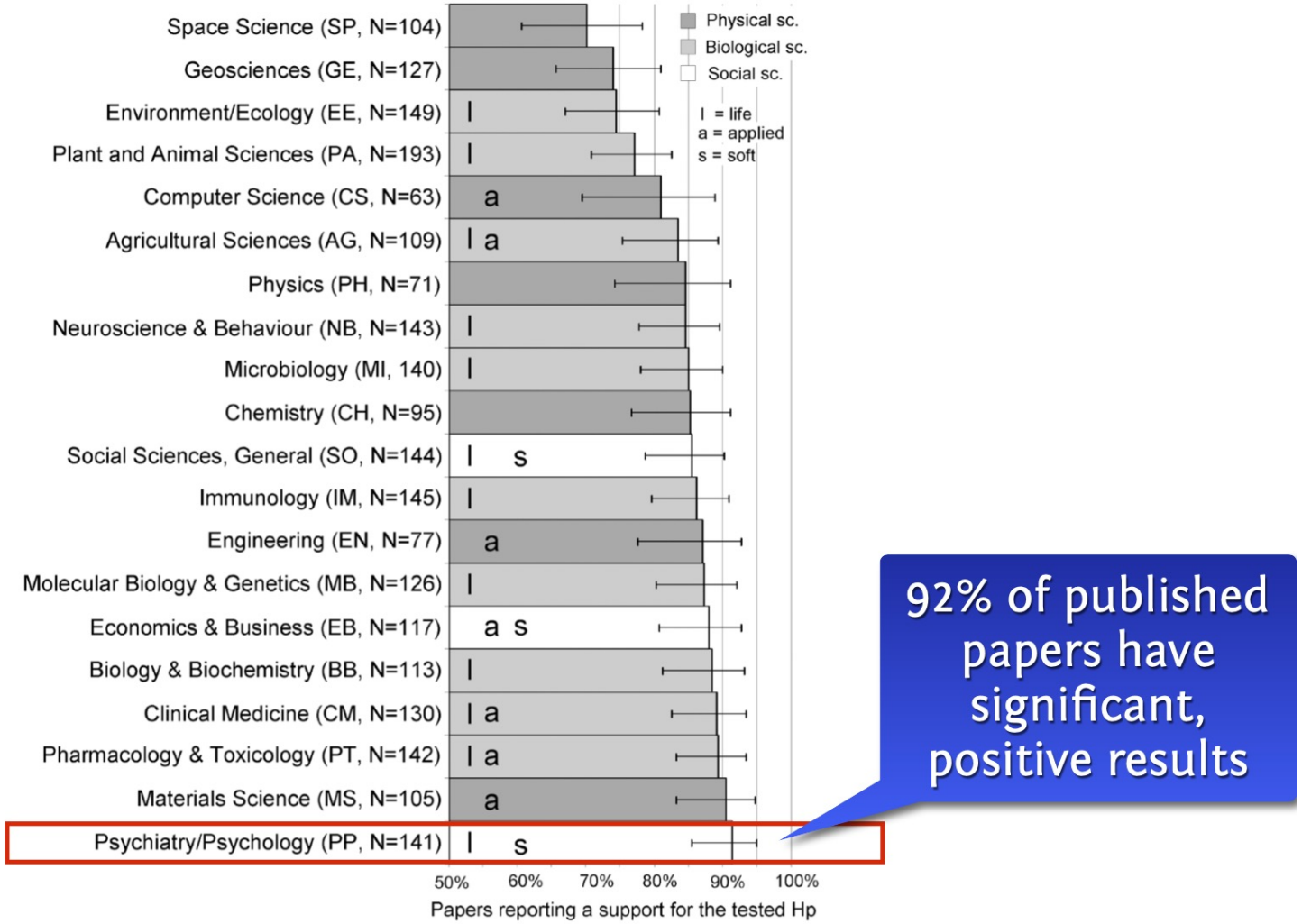
Introduction
How to become a Professor

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

Introduction
How to get lots of publications

"Positive" results increase down the hierarchy of the sciences



Fanelli (2010). *PLOS ONE*, 5, e10068.

What is *p*-hacking?

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

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

What is *p*-hacking?

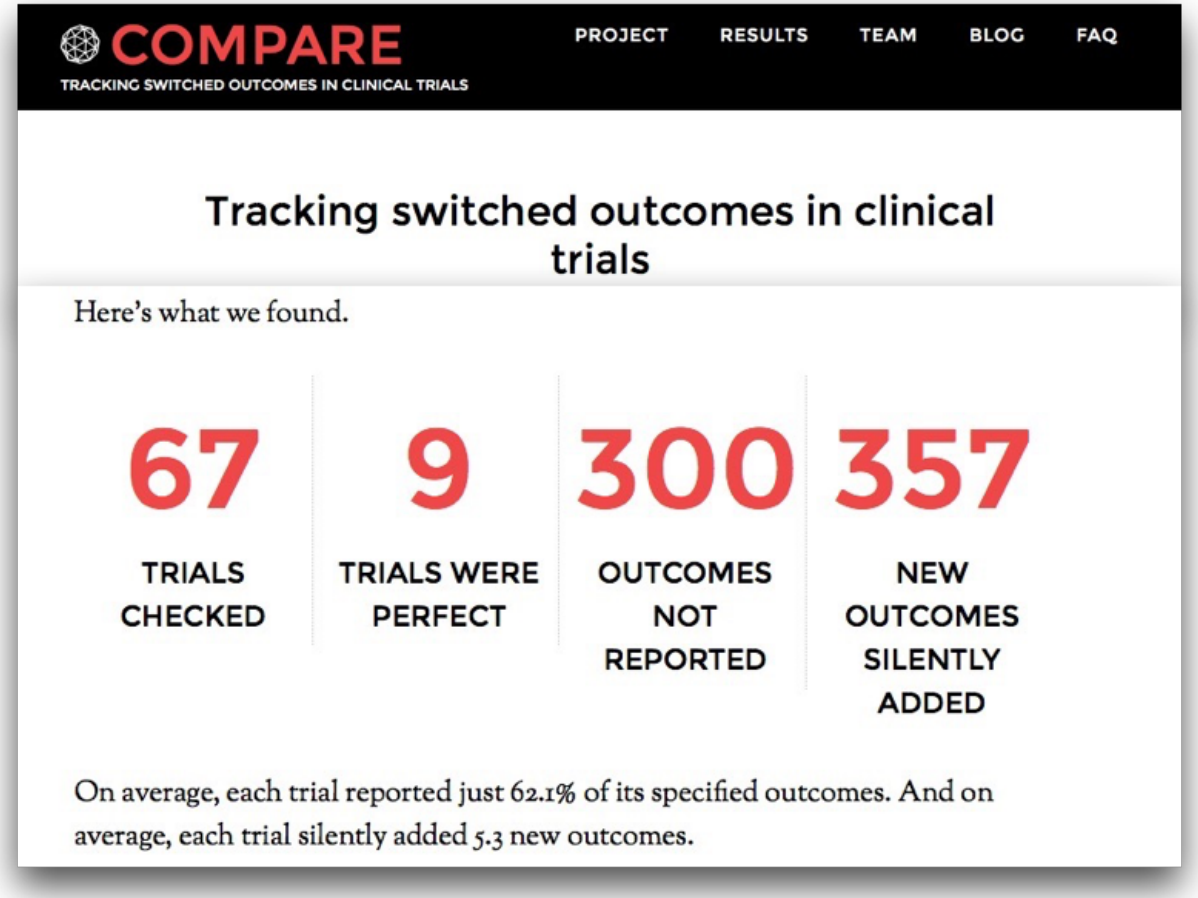
***p*-hacking** (*n.*). Tune your data analysis in a way that you achieve a significant *p*-value in situations where it would have been non-significant.

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

p-hack your way to scientific glory!

How to *p*-hack: Tools
Tool 1: Outcome switching

- Assess more than one dependent variable (DV), but report only those which "worked"
- 2 outcome variables: false positive rate increases from **5%** → **9.5%**
- 5 (uncorrelated) DVs with one-sided testing: false positive rate increases from **5%** → **41%**
- How prevalent is it?
  - **66% of researchers** admit having done this (John, Loewenstein, & Prelec; 2012)

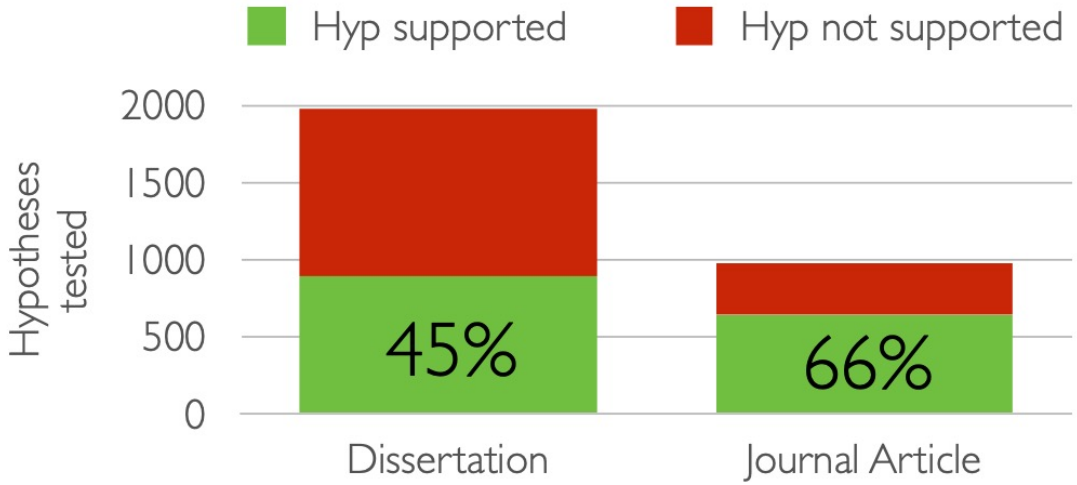


COMPARE | John, Loewenstein, & Prelec (2012). *Psychological Science, 23*(5), 524–532.

How to *p*-hack: Tools
Tool 2: Many conditions

- Assess more than two conditions but report only those with $p <. 05$
- *Example 1*: testing "high", "medium", and "low" conditions and reporting only the results of a "high" versus "medium" comparison
- *Example 2*: Thesis vs. manuscript (O'Boyle et al., 2017)
- Gives you more than one chance to find an effect, increases false positive rate up to → **12.6%**
- How prevalent is it?
  - **27% of researchers** admit having done this (John et al., 2012)

The Chrysalis effect: How ugly initial results metamorphosize into beautiful articles

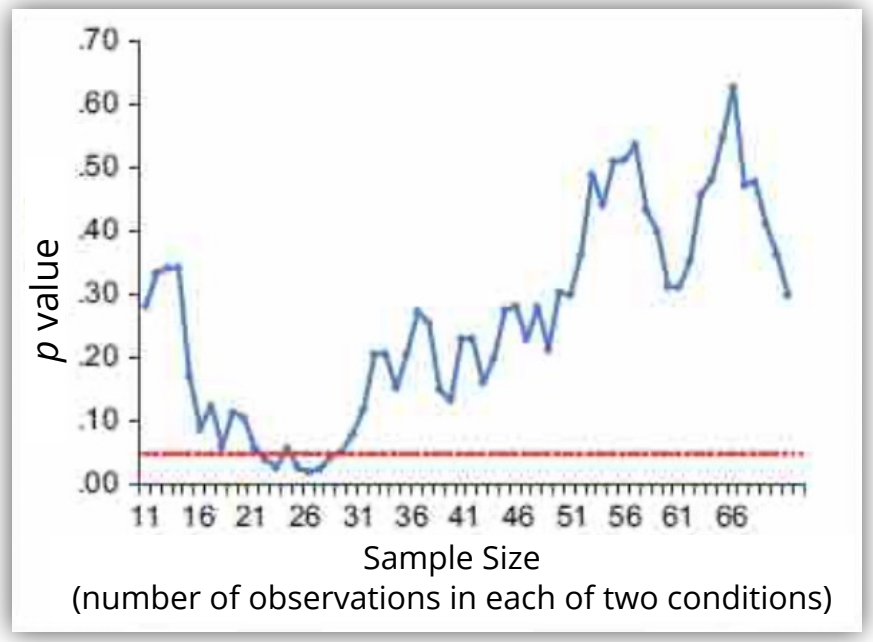


142 dissertations that were subsequently published in a refereed journal (149 field studies, 26 experiments)

Twitter Status | O'Boyle et al. (2017). *Journal of Management, 43*(2). 376–399.

How to *p*-hack: Tools
Tool 3: Optional stopping

- Collect an initial sample, analyze results, add additional participants if not significant, repeat until significance is found
- Increase once: false positive rate = **7.7%**, twice: false positive rate = **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)



John, Loewenstein, & Prelec (2012). *Psychological Science, 23*(5), 524–532.

How to *p*-hack: Tools
Tool 4: Subgroup analysis

- Research question: Do aggressive primers trigger aggressive behavior?

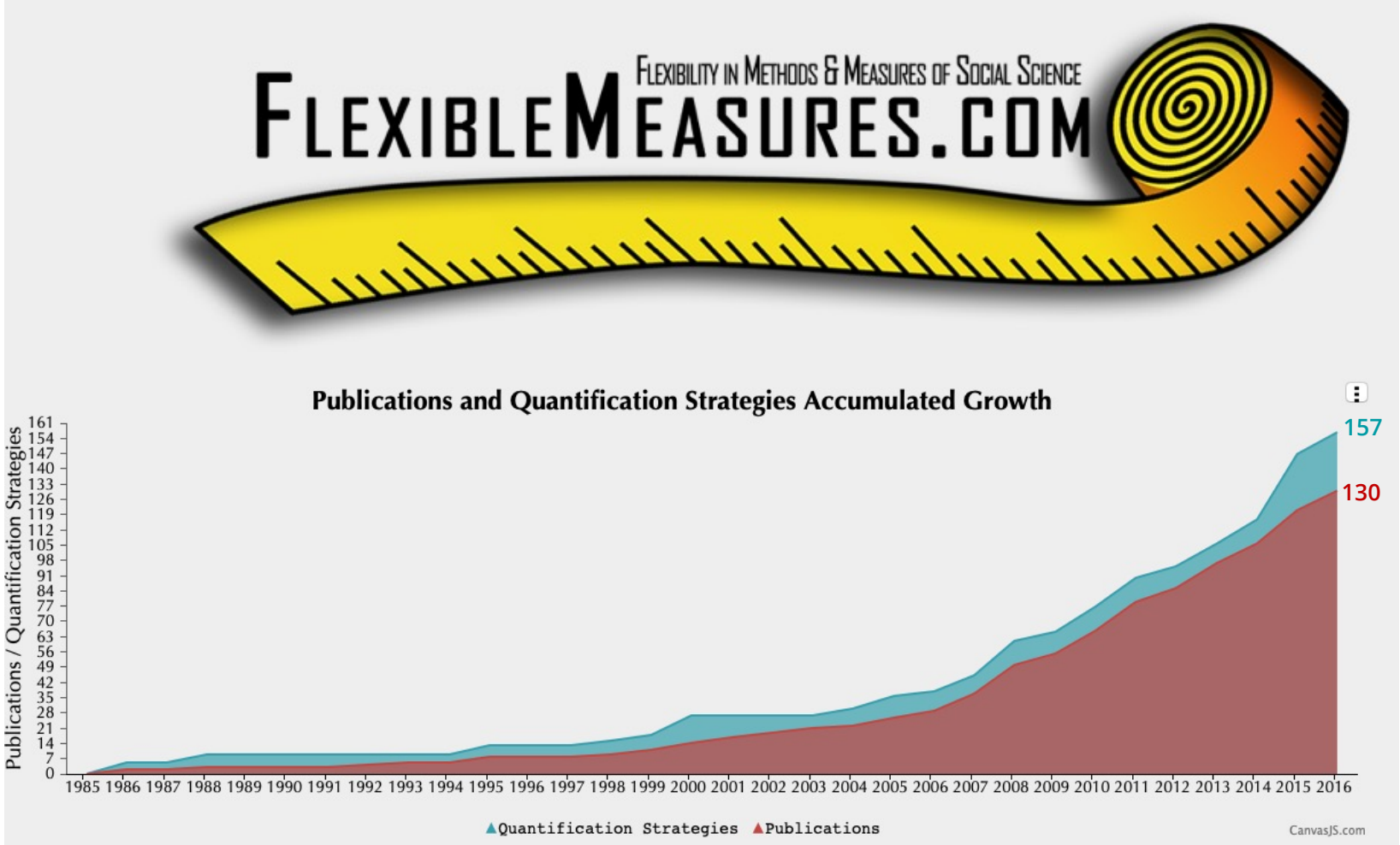
> A second study in Turner, Layton, and Simons (1975) collects a larger sample of men and women driving vehicles of all years. **The design was a 2 (Rifle: present, absent) × 2 (Bumper Sticker: "Vengeance", absent) design with 200 subjects.**

- Presumably no effect, but wait…

> 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, $F(1, 129) = 4.03$, $p = .047$. But then they report that the Rifle/Vengeance condition decreased honking among older-vehicle male drivers, $F(1, 129) = 5.23$, $p = .024$! No results were found among female drivers.

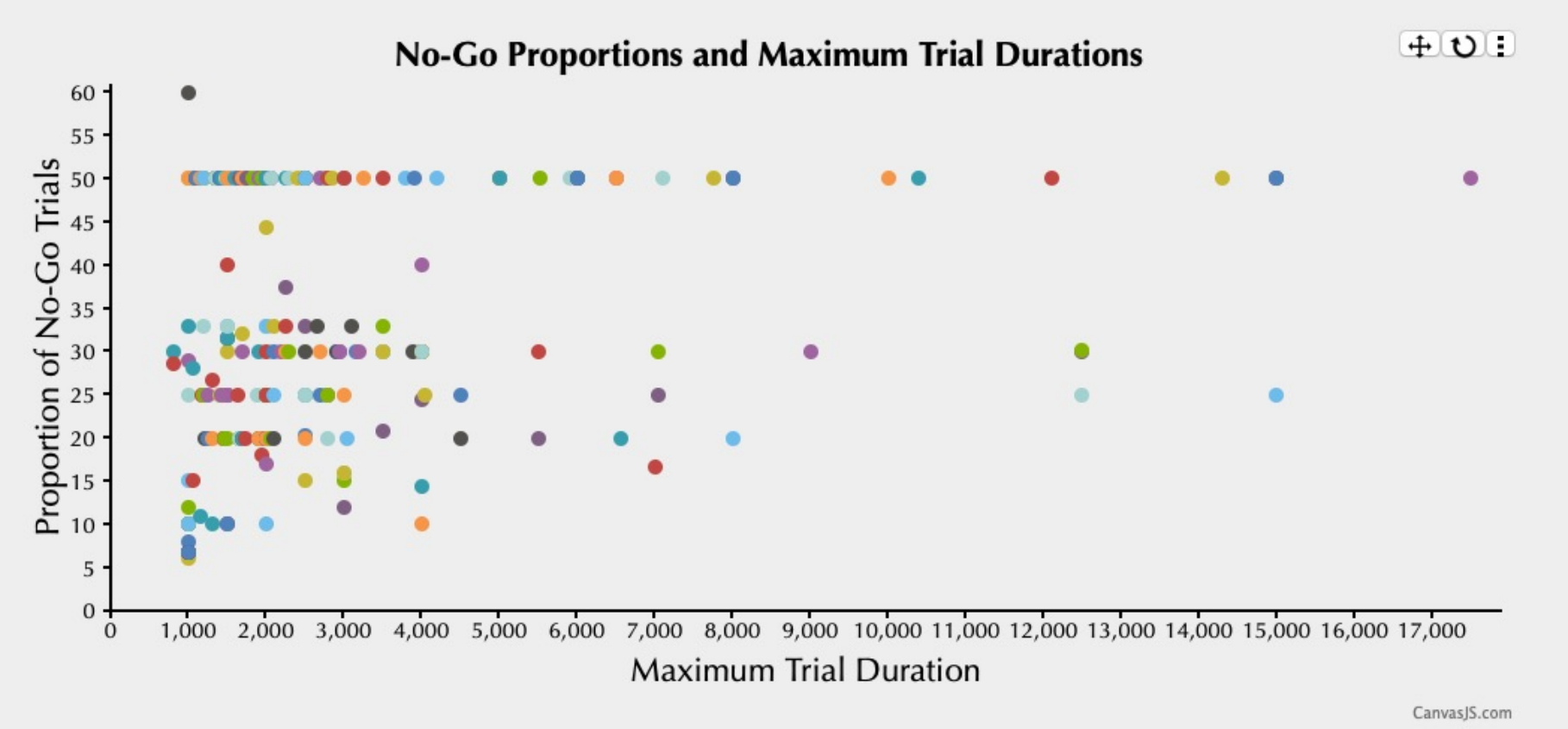
Blog Post | Turner et al. (1975). *JPSP, 31*(6), 1098–1107.

How to *p*-hack: Tools
Tool 5: Flexible measures



How to *p*-hack: Tools

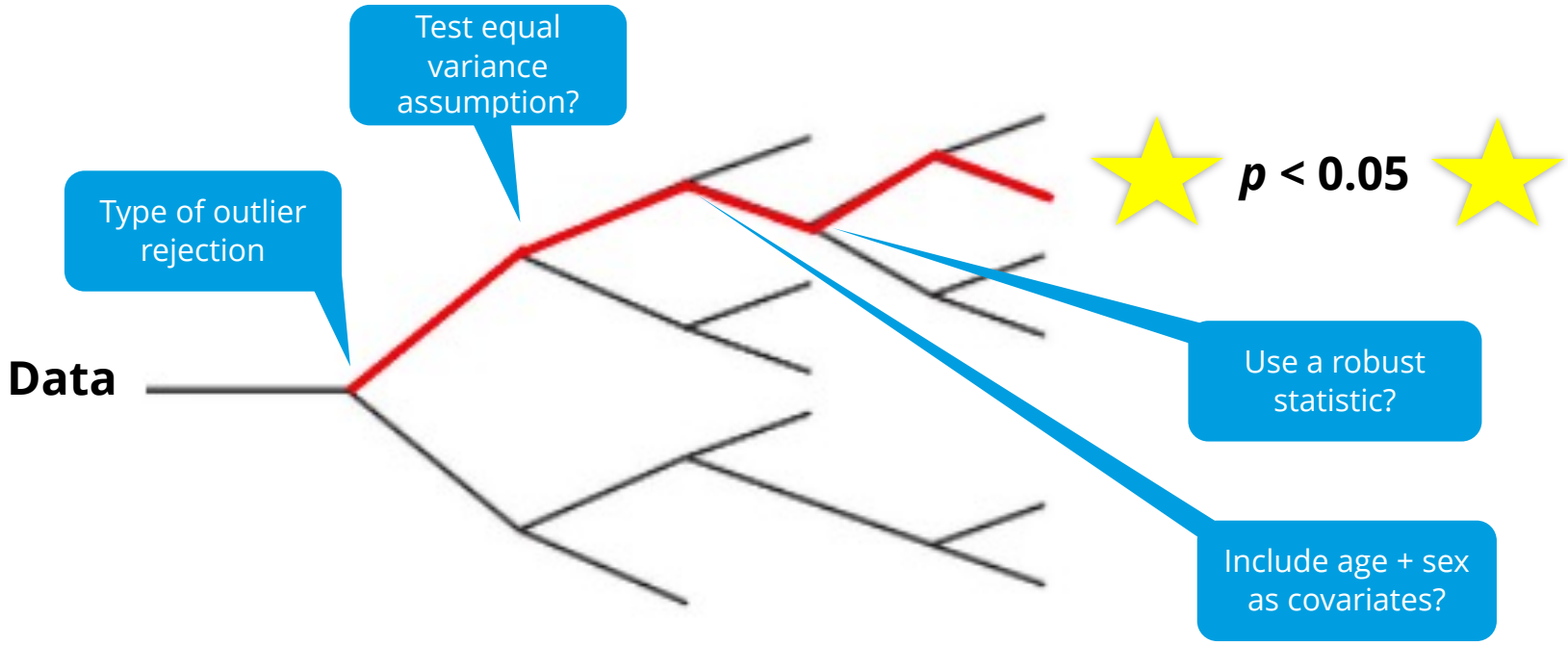
Tool 5: Flexible measures



How to *p*-hack: Tools

Tool 6: Many researcher degrees of freedom

The garden of forking paths

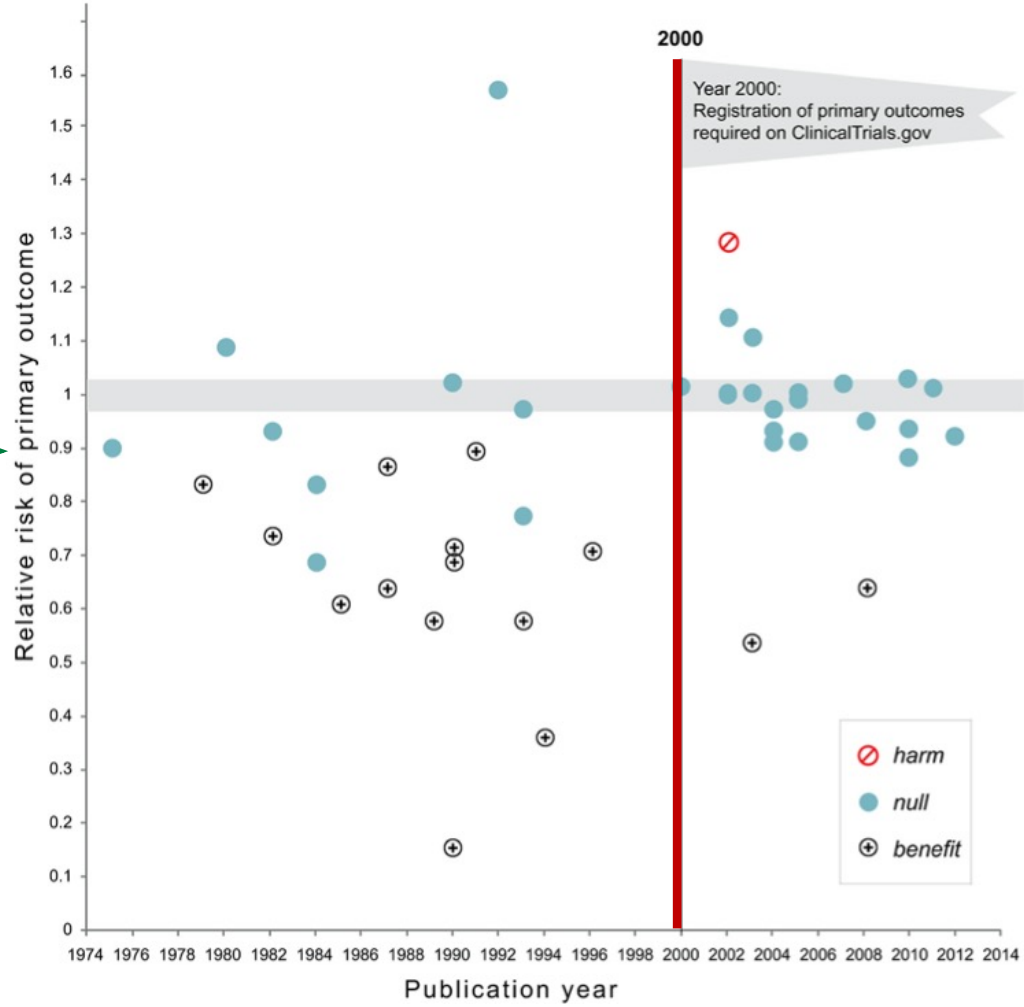


Blog Post Neuroskeptic | Gelman & Loken (2013). *Department of Statistics, Columbia University, 348.*

How to *p*-hack: Tools

Tool 7: No preregistration

No prereg:
57%
success rate!

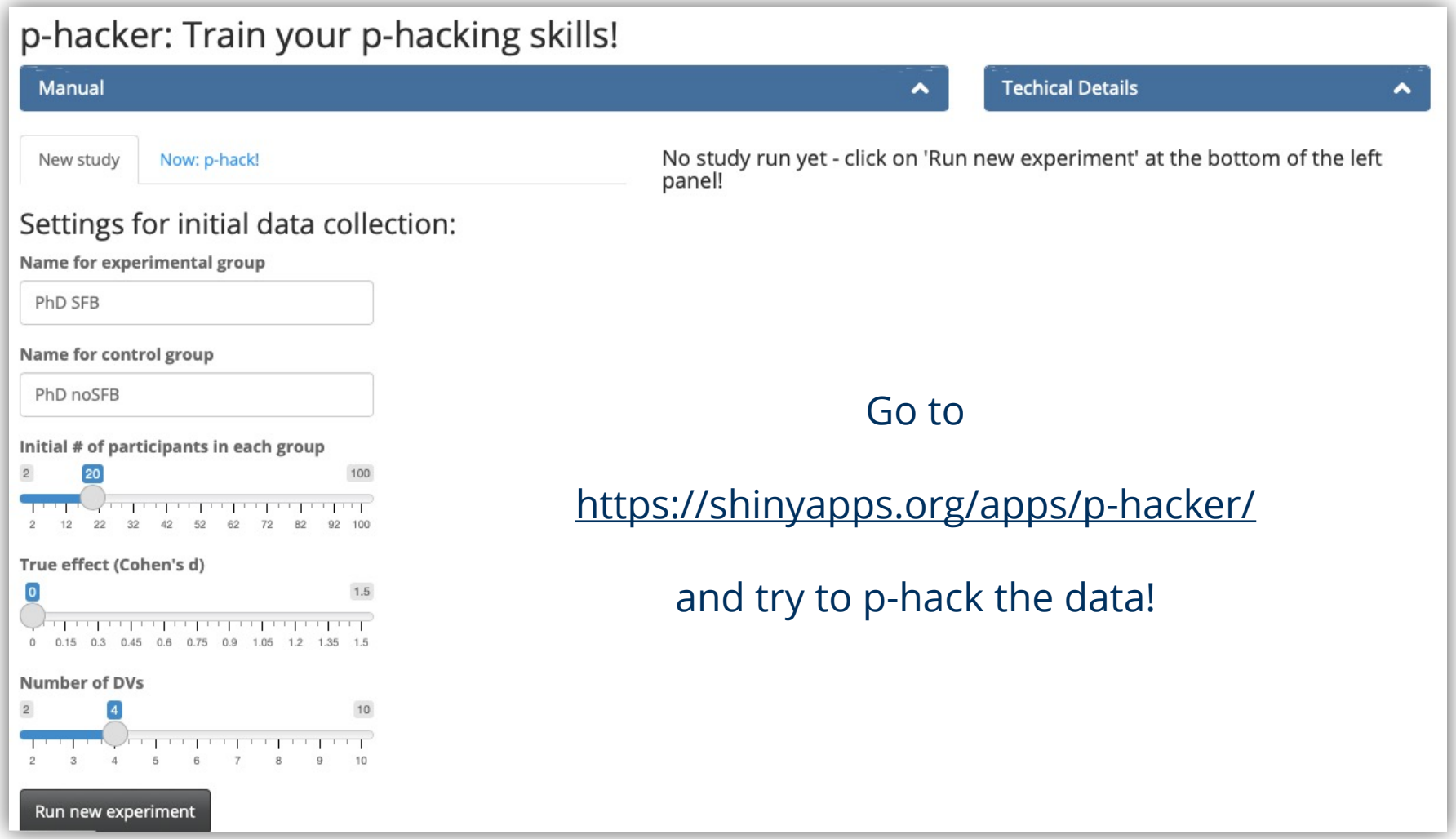


Prereg:
8%
success rate...

Blog Post | Kaplan & Irvin (2015). *PLOS ONE, 10*(8), e0132382–12.

Exercise

Exercise



Go to

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

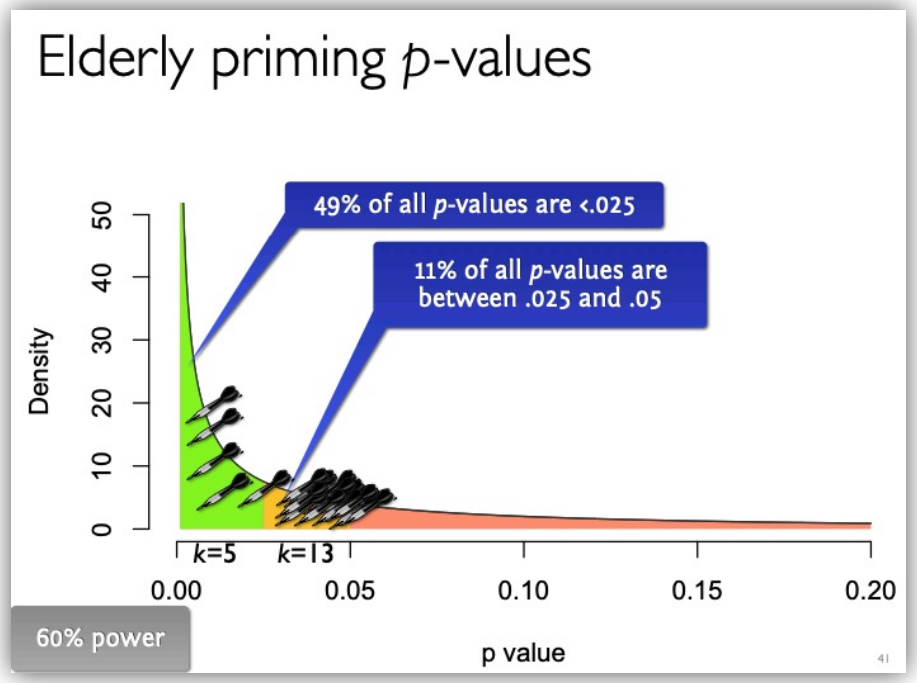
and try to p-hack the data!

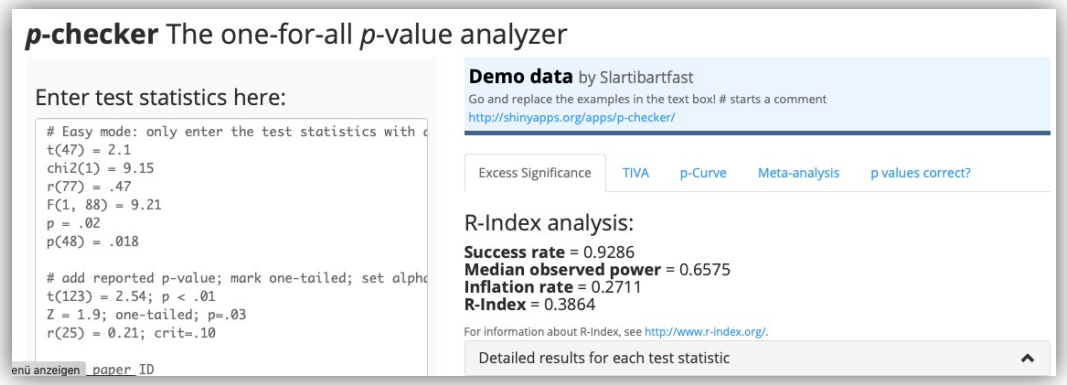
***P*-hacking**
What we did *not* consider

Tools to detect *p*-hacking

- *P*-curve
- Replication Index (R-Index)
- Test of insufficient variance (TIVA)
- Begg and Mazumdar test
- Egger's regression test
- Precision effect test (PET)
- Precision effect estimate with standard error (PEESE)
- …

→ Practice with *P*-checker!





https://shinyapps.org/apps/p-checker/

Solutions

Solutions
How to prevent *p*-hacking

Be honest / Don't do intentional *p*-hacking

- 21 word solution (Simmons et al., 2012)

> We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study.

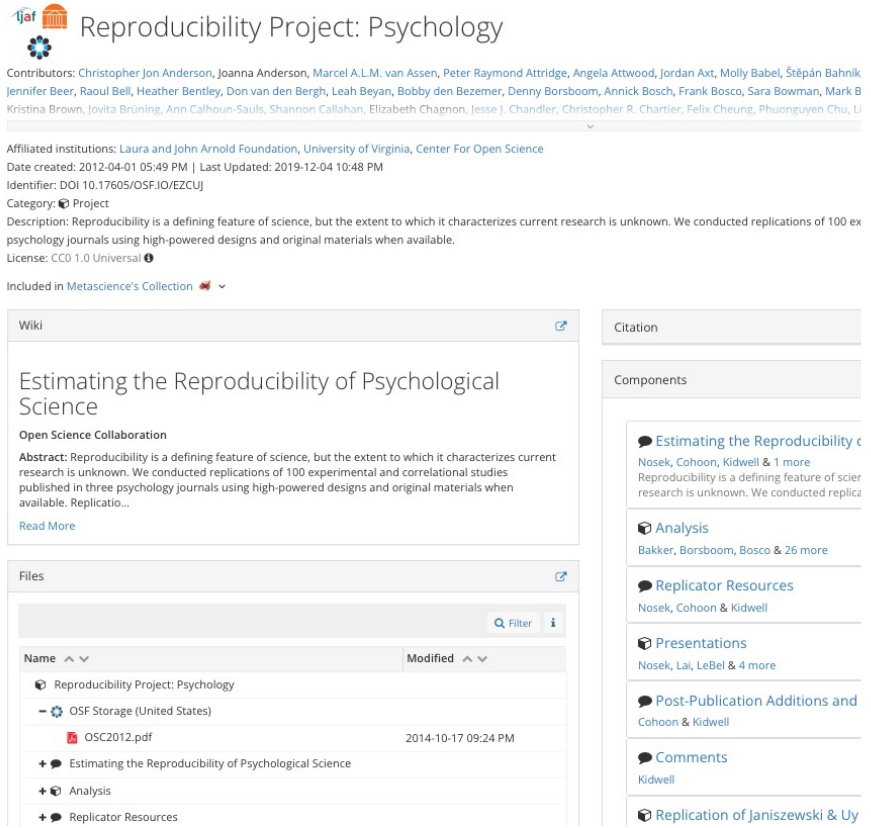
- Standard Reviewer Statement for the Disclosure of Sample, Conditions, Measures, and Exclusions

> "I request that the authors add a statement to the paper confirming whether, for all experiments, they have reported all measures, conditions, data exclusions, and how they determined their sample sizes. The authors should, of course, add any additional text to ensure the statement is accurate. This is the standard reviewer disclosure request endorsed by the Center for Open Science [see http://osf.io/hadz3]. I include it in every review."

Solutions
How to prevent *p*-hacking

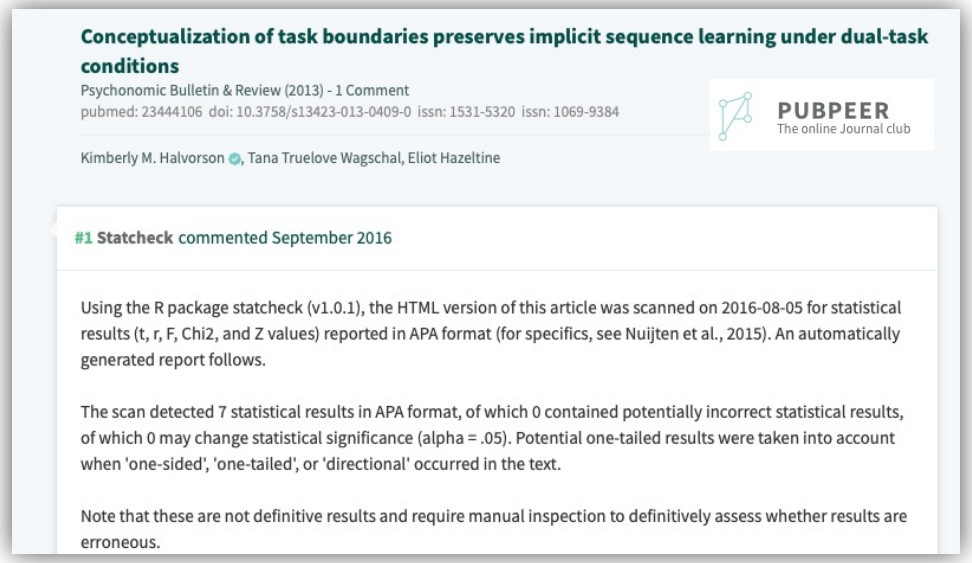
Don’t do unintentional p-hacking

- Provide open material for an existing project
  - E.g. on OSF (https://osf.io), publish open data, open material, reproducible analysis code, preprints, postprints, supplemental material
  - Get a persistent URL and even a doi
- Identify questionable research practices
  - E.g., *p*-checker, replication index (https://replicationindex.com)
- Preregister your studies

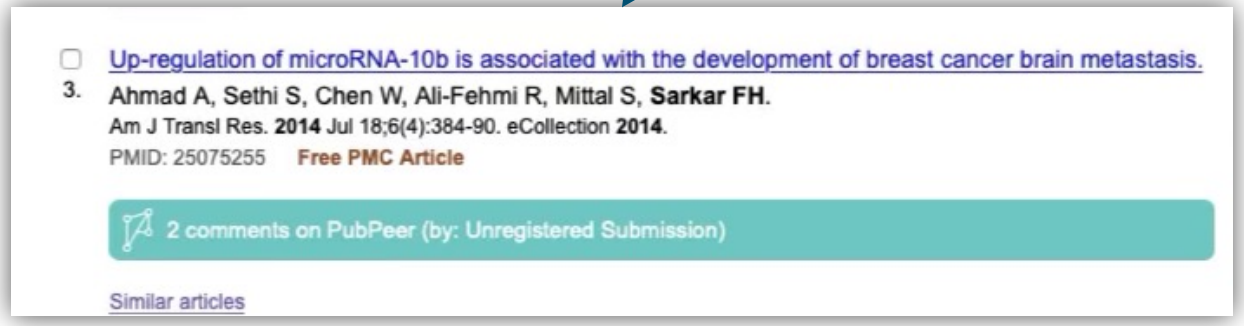


Solutions
How to prevent *p*-hacking

Do post-publication peer review



PubPeer browser plugin: Automatic alert for PubPeer comments at Google, journal websites etc.



Who to ask?

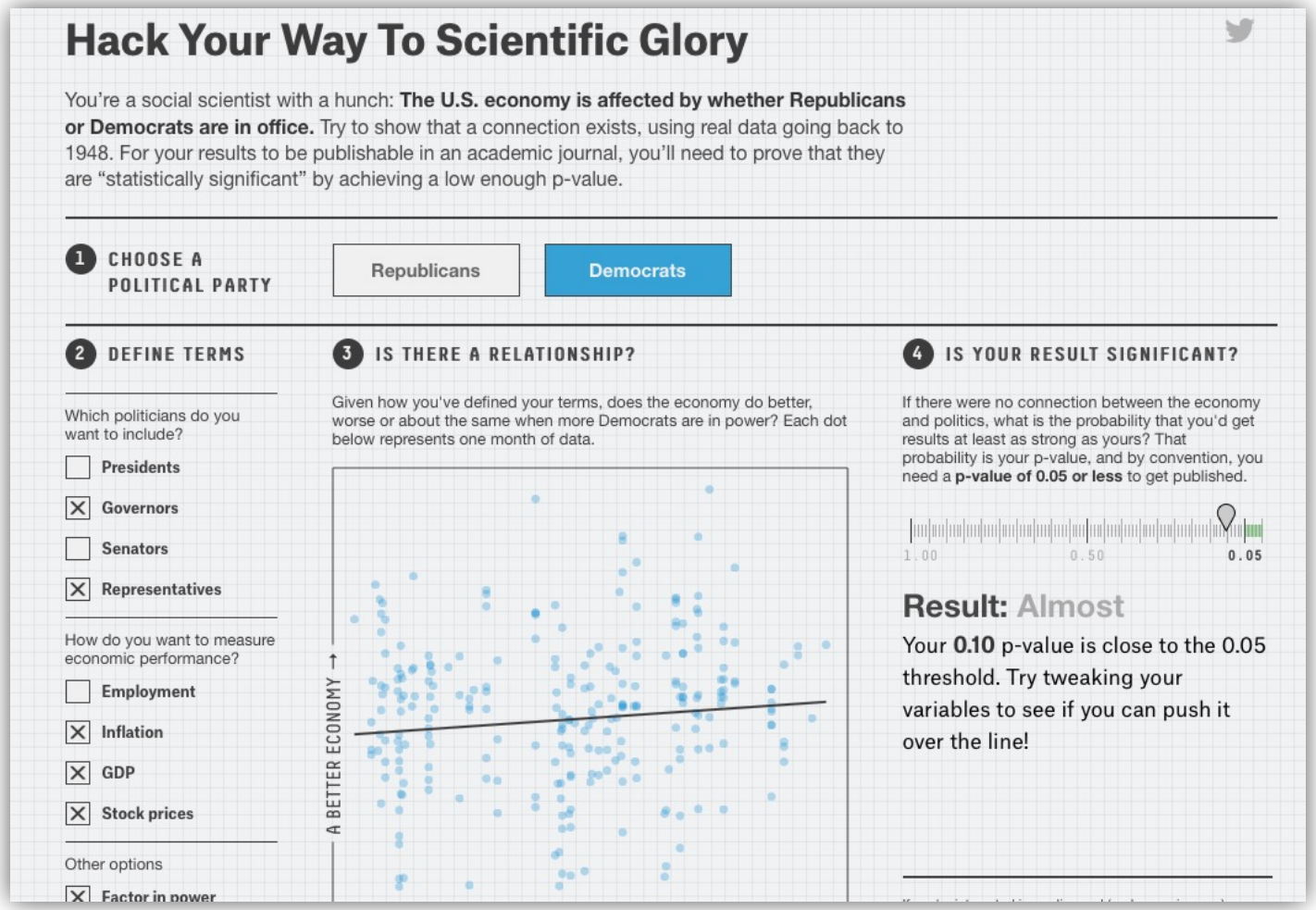
- Your local open science initiative
  https://tu-dresden.de/mn/psychologie/die-fakultaet/open-science

OPEN SCIENCE INITIATIVE
OSIP Open Science Initiative der Fakultät Psychologie der TUD

Interview with a researcher

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

Exercise



https://projects.fivethirtyeight.com/p-hacking/

Thank you!