# White Hat P-Hacking

Bayesian Parables to Understand the Reproducibility Crisis

Omkar Venkatesh

# P-hacking?

- [https://projects.fivethirtyeight.com/p-hacking/](https://projects.fivethirtyeight.com/p-hacking/)
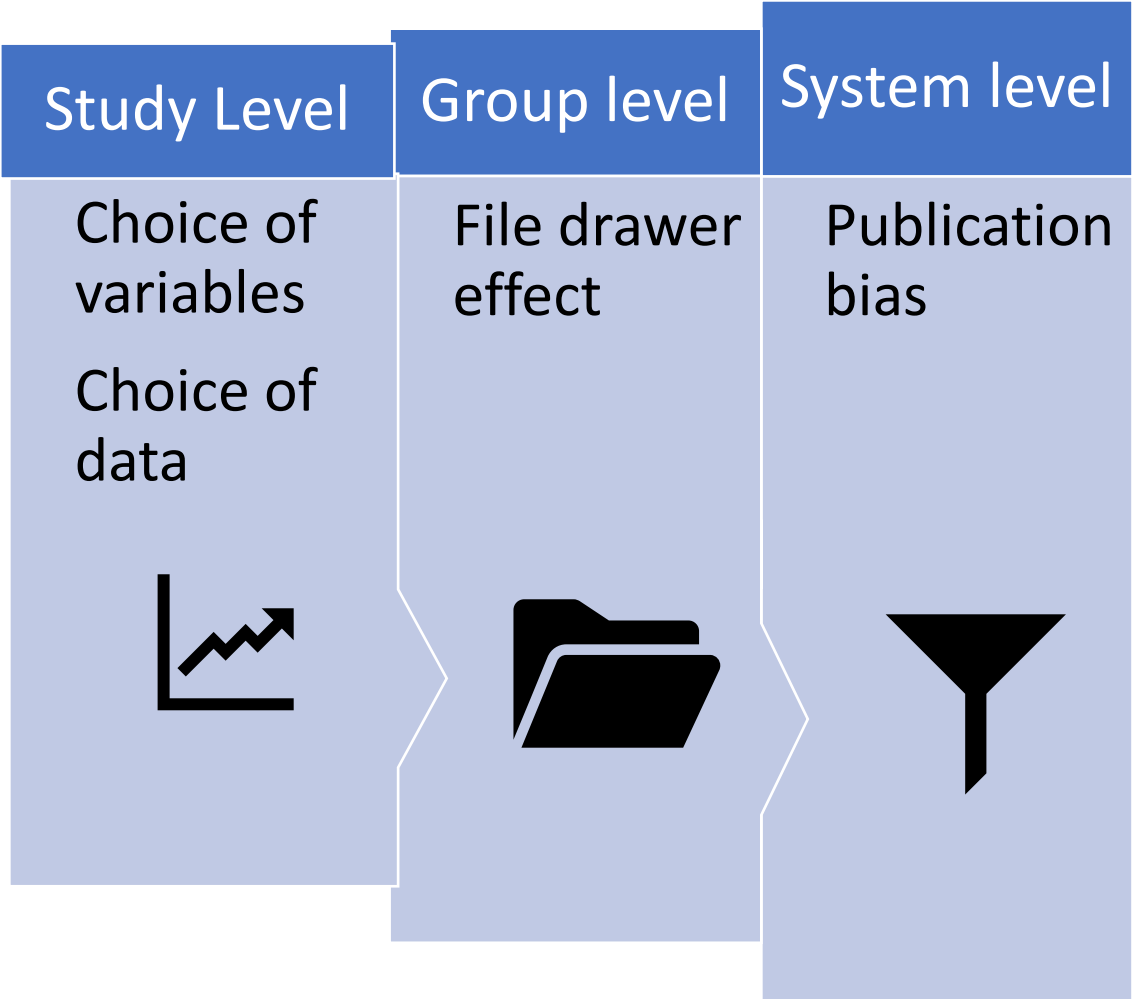
# The P-Hacker

Non-reproducible results are almost never due to nefarious intent. Or even ignorance. But adopting this perspective might help us understand the concepts in a new way.
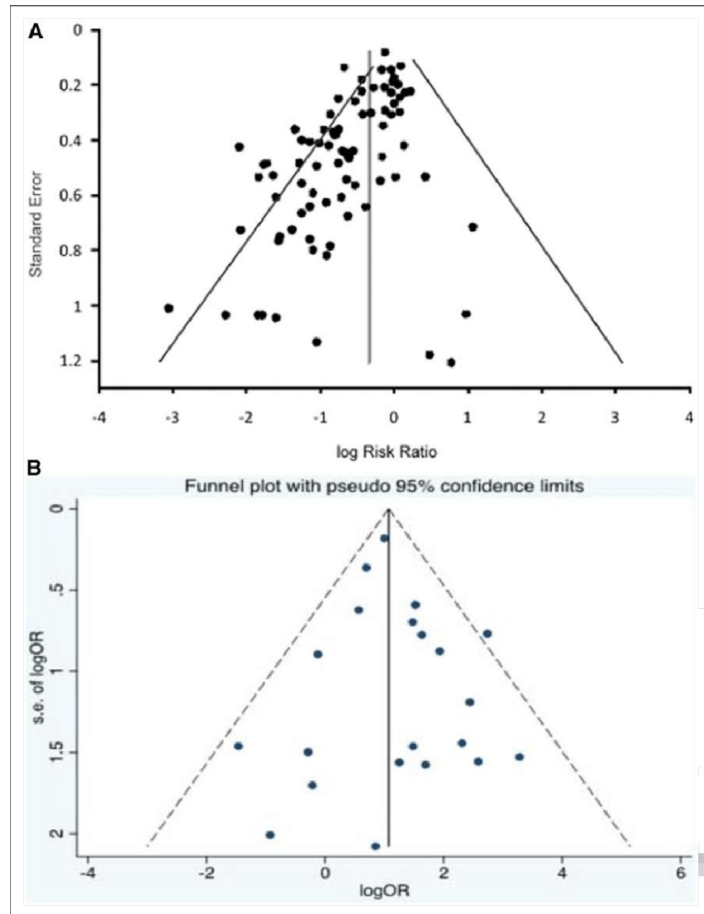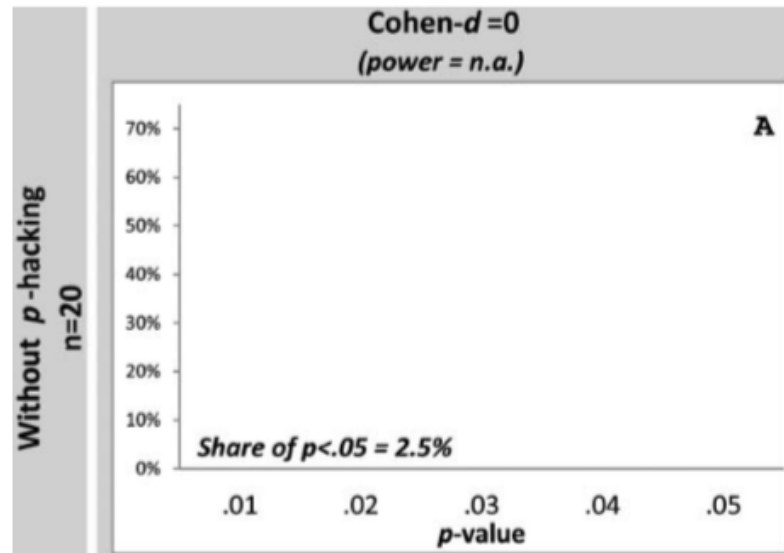
# The Multiple Layers of P-hacking

| Study Level | Group level | System level |
|---|---|---|
| Choice of variables | File drawer effect | Publication bias |
| Choice of data | | |

# How can this stuff be caught?

# What if we get caught? ~~Enter~~ Funnel Plots Exit



But r(n, d) ≠ 0

# P-Curves?



**Cohen-*d* =0**
*(power = n.a.)*

A

*Share of p<.05 = 2.5%*

Without *p*-hacking
n=20

*p*-value
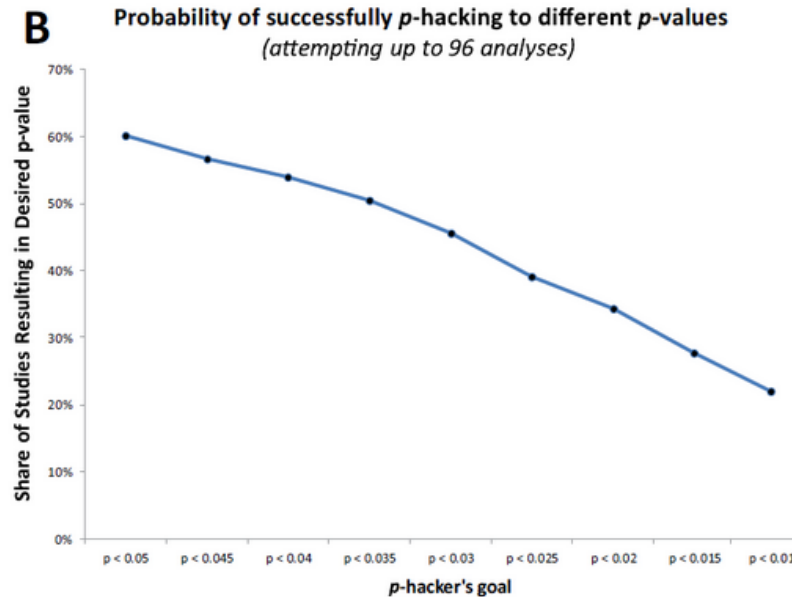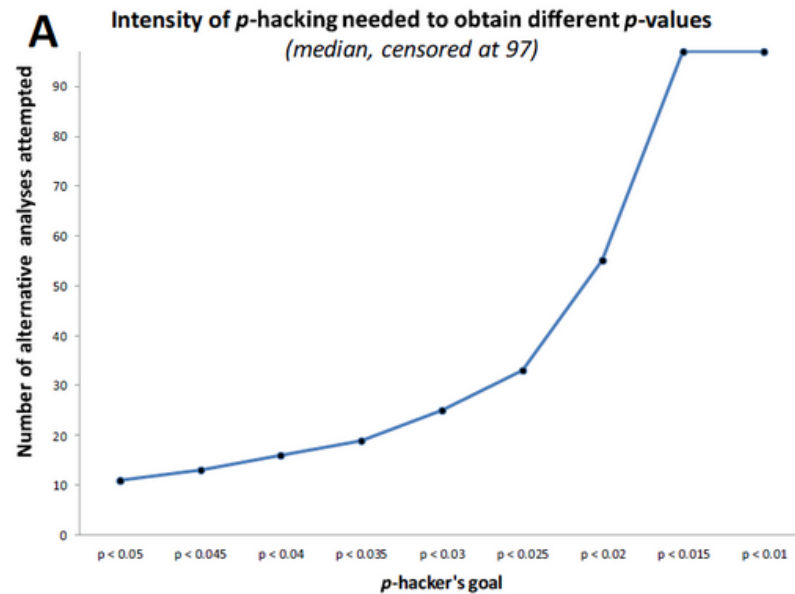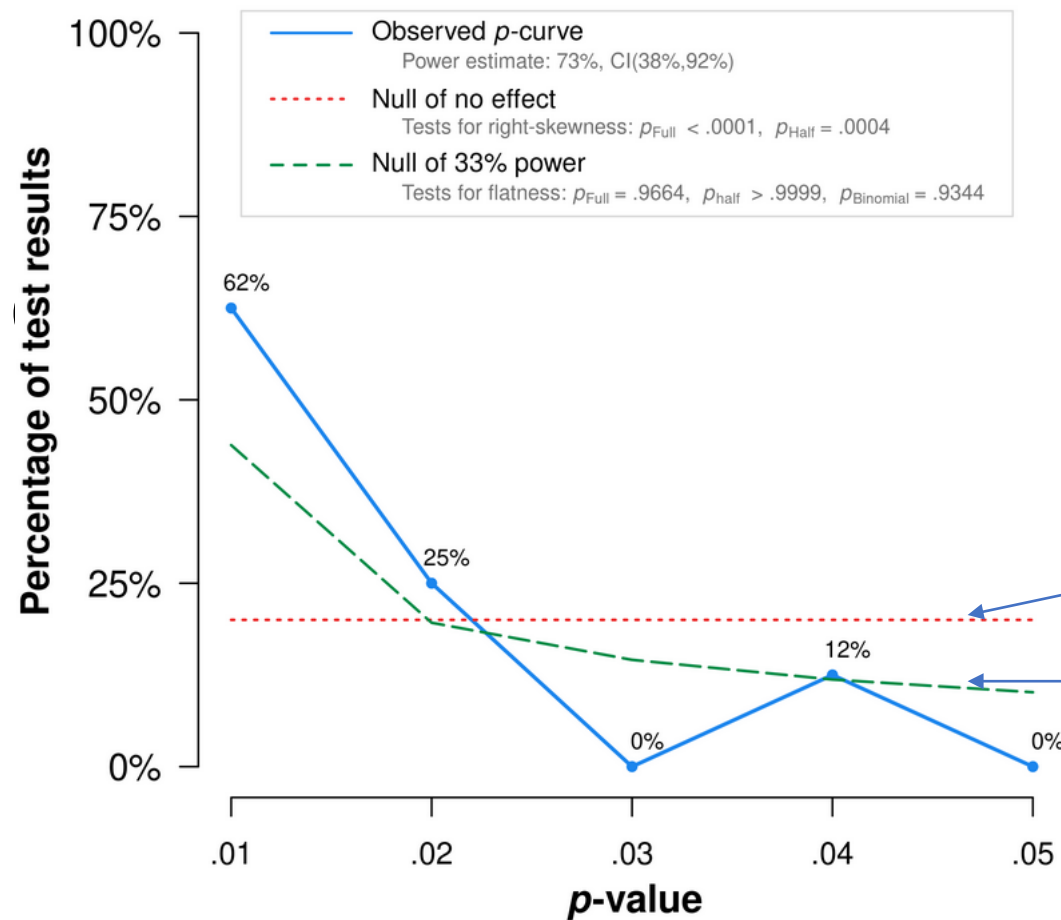
# Wait, what if we tried harder?

- The key to p-curves: limited ambition
  - You don't p-hack to .0001, you're happy with .05

When do I support for null?

Answer: This can happen when power is high and effect size is large
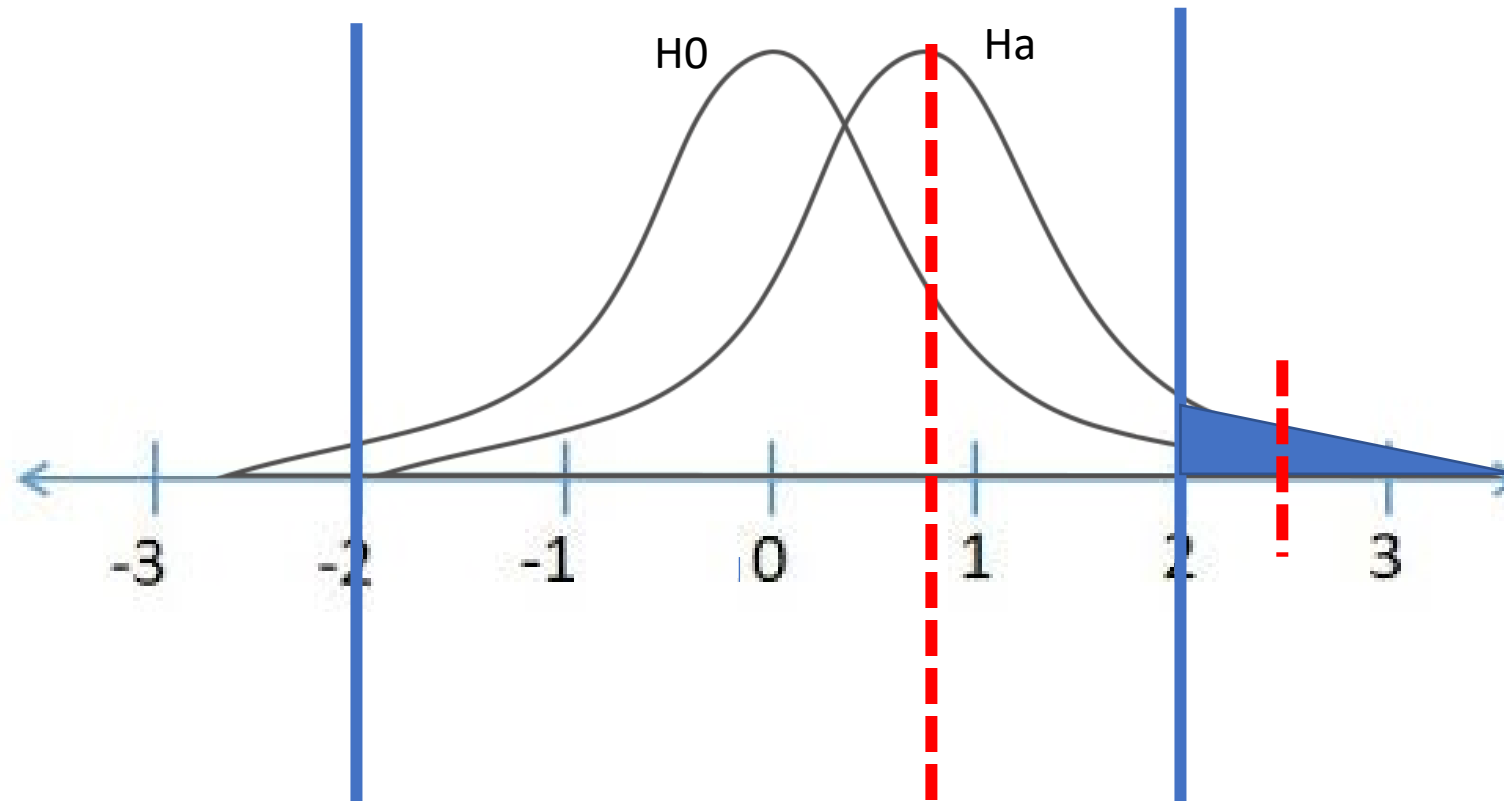
P(Data|Null)

P(Data|Alternative)

A deeper dive into power…

# What about low power?

# Doing a meta-analysis, how can I jack up effect size?

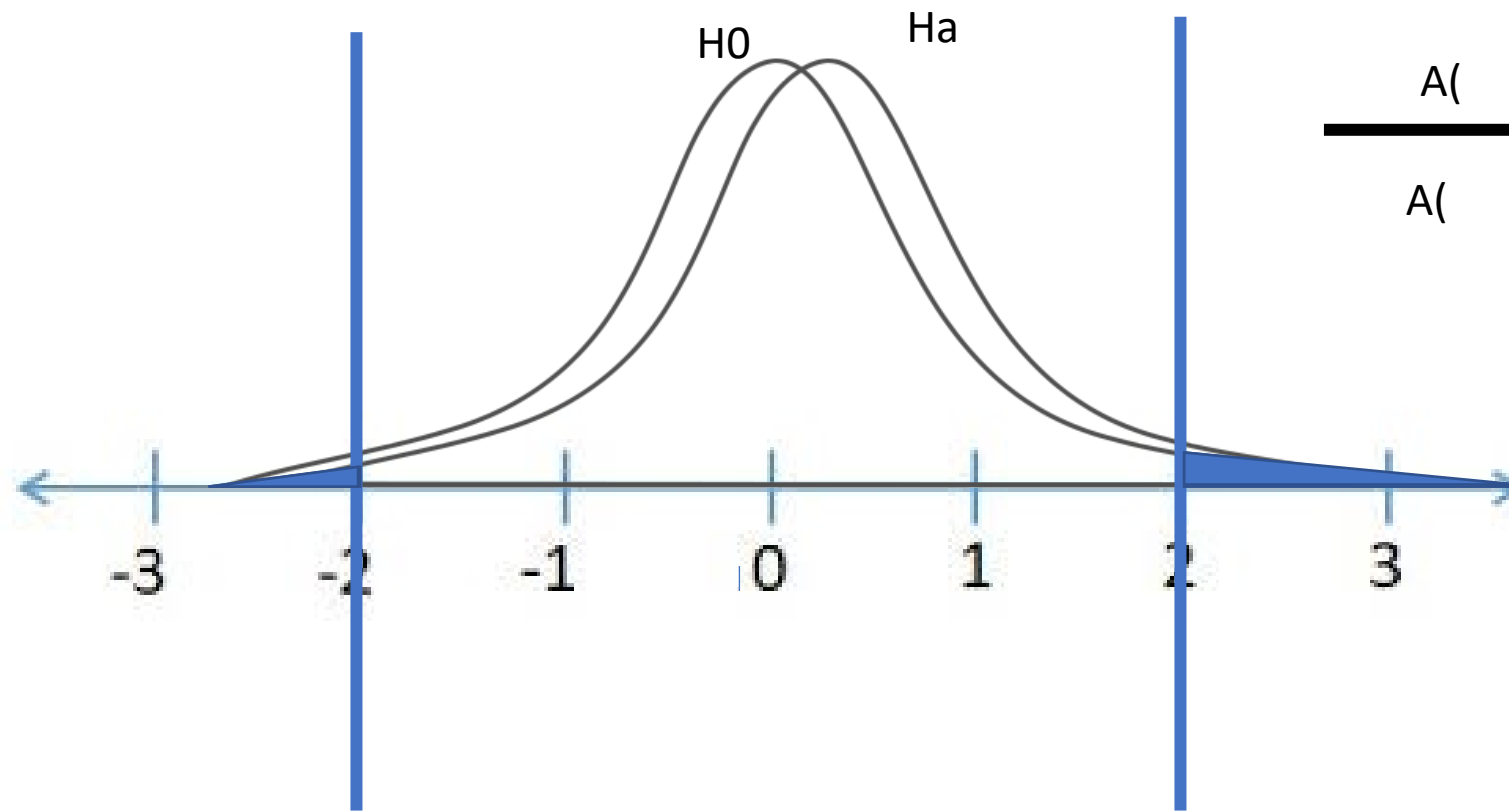- Use low power studies and one-sided tests



H0    Ha

True effect: <1
Average significant result: >2

# Can I use low power to produce results that "contradict" the scholarly consensus?

- Yep

# What if editors catch on and make me do a post-hoc power analysis?

- Use the effect size from your study, since it's likely an overestimate, as seen before

# Summary

- P-Hacking is often unavoidable

- Biased distributions of p-values can demonstrate p-hacking

- p<.05 doesn't simply mean evidence for null

- Low power reduces the informative value of significant p values
  - Can also lead to magnitude errors in meta analysis
  - Can also lead to "sign" errors

# What's this all about, anyway?

- Human need for certainty & binary labels
- The sneakiness of multiple comparisons
  - "Garden of forking paths"
- Proposed solutions
  - Bayes factors
  - Lower the significance threshold
  - Preregistration
  - Cross-validation (hold-out)

# Contact

- Twitter @OmkarGV

- This presentation is heavily indebted to the works/ideas of Uri Simonsohn, Andrew Gelman, Daniel Lakens and others. No infringement is intended.