

BCS 206: Research Lab Course

Replication crisis

Levels of description of “science”

Goal

Understanding

Process/algorithm

Scientific method

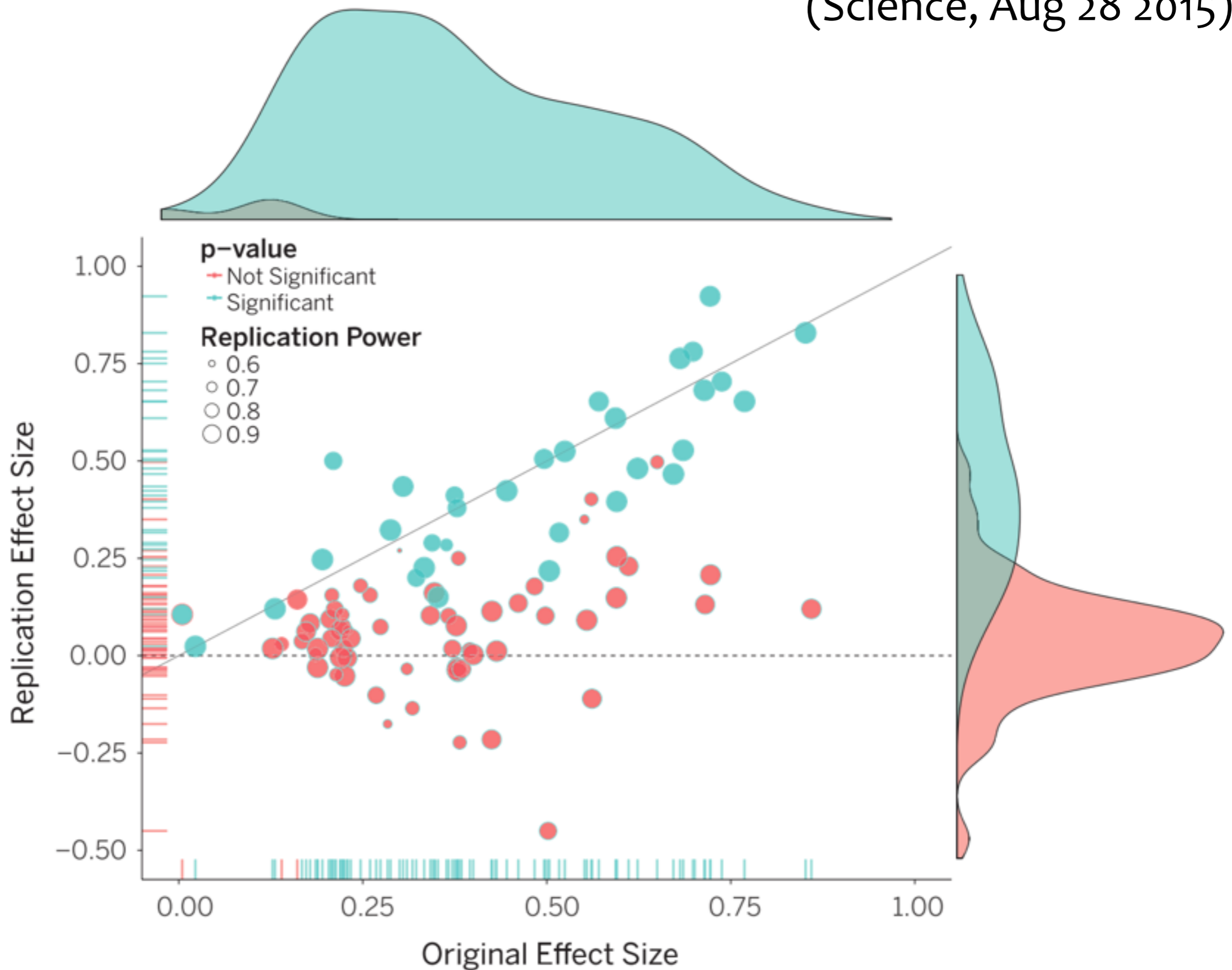
- Data characterization
- Hypotheses generation
- Predictions
- Experimentation/
observation

Implementation

Scientific
community

Estimating the reproducibility of psychological science

(Science, Aug 28 2015)



Crucial elements

- Effect size and statistical power of study
- True positive and false positive
- Pre-study odds of hypothesis being true
- Bias in estimate of hypothesis being true
 - “p-value hacking” to “get a paper”
 - reverse less common: e.g. conflicts of interests, oversight
- Multiple tests (e.g. by different groups)

Ioannidis 2005

The more likely wrong...

1. ... the smaller the studies
2. ... the smaller the effect size
3. ... higher the number of tested hypotheses
4. ... the greater the design+analysis flexibility in a field
5. ... the stronger the conflicts of interest in a field
6. ... the “hotter” a field

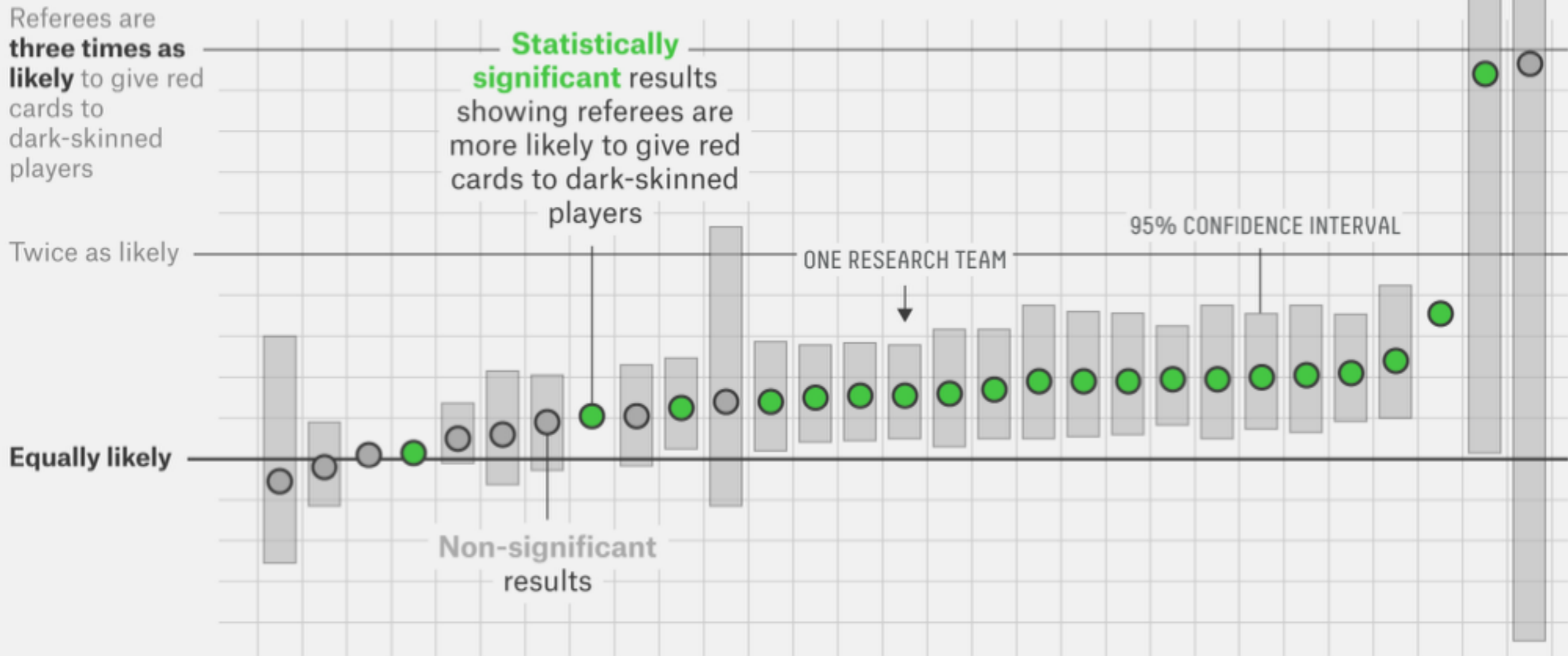
p -hacking demo

- multiple comparisons (“fishing expedition”) without correcting for it
- selective reporting of data/suppression of counter-evidence

p-value ambiguity with the best of intentions

Same Data, Different Conclusions

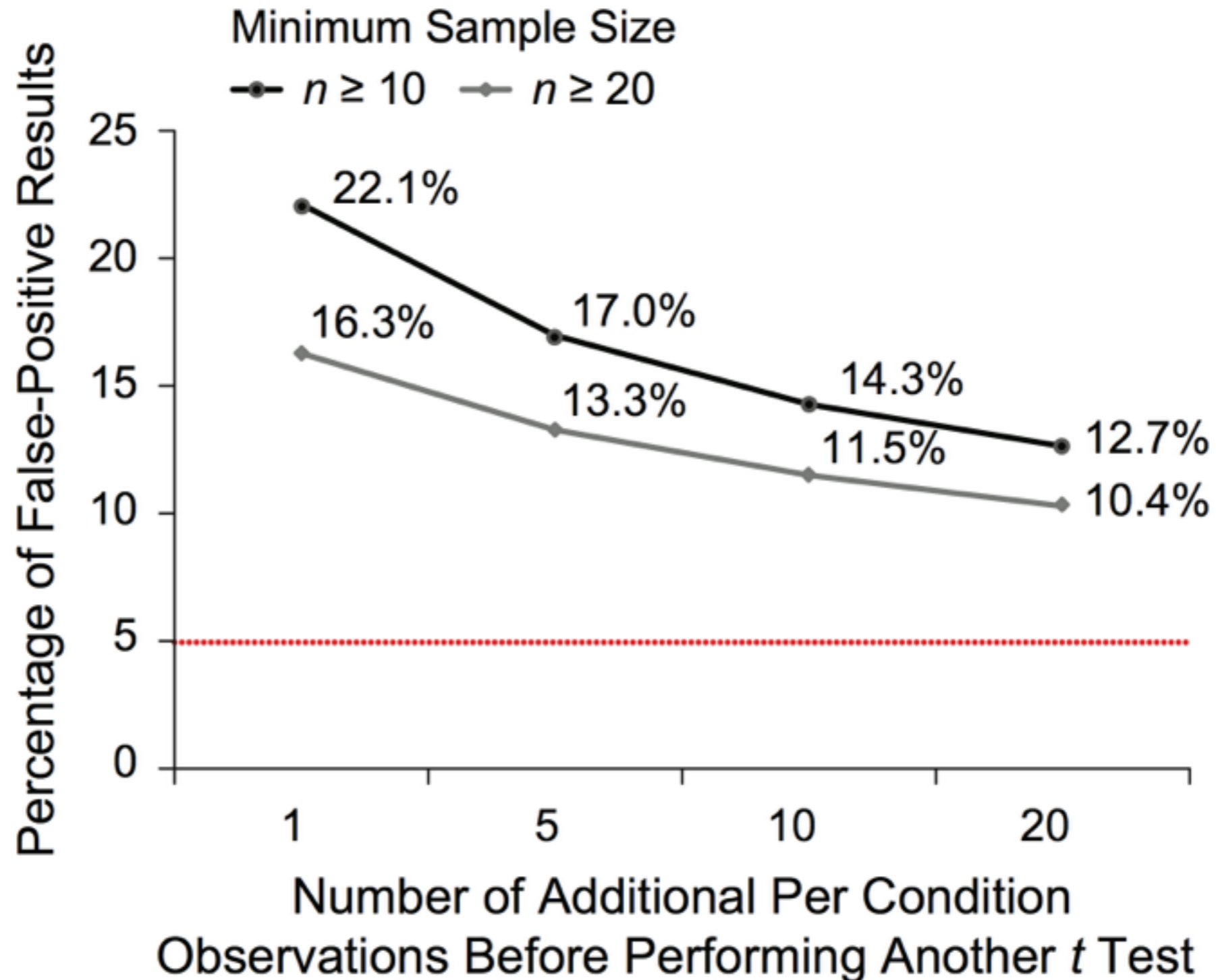
Twenty-nine research teams were given the same set of soccer data and asked to determine if referees are more likely to give red cards to dark-skinned players. Each team used a different statistical method, and each found a different relationship between skin color and red cards.



“Questionable” research practices

(Simmons, Nelson, Simonsohn 2011)

- continuing data selection until significance



9 circles of scientific hell

