Psychology Replication Crisis

Emily Robinson
Metis Investigation 7/21

The Reproducibility Project in Psychology

- Researchers at Center for Open Science coordinated replication teams who conducted **replications of 100 studies** published in psychology journals

- Replication effects (Mr = .197, SD = .257) **half the magnitude** of original effects (Mr = .403, SD = .188)

- **Only 36%** of replications **had significant results** (compared to **95%** of the original studies)

KATIE M. PALMER SCIENCE 03.03.16 2:00 PM

PSYCHOLOGY IS IN CRISIS OVER WHETHER IT'S IN CRISIS

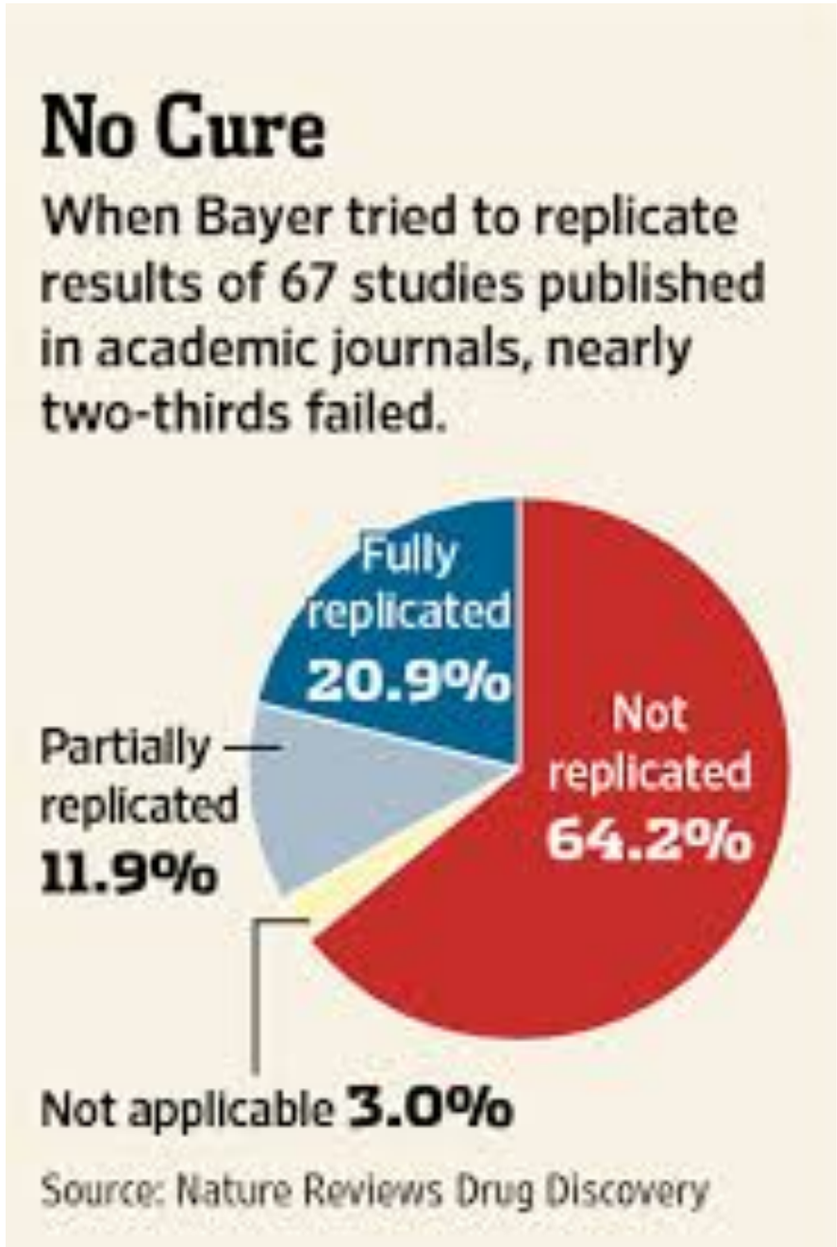
- Some in psychology community disagreed on:
  - The methodology and analysis of the reproducibility project
  - Commentary in Science on statistical errors

Replication = Bullying?

Medical studies also typically fail to replicate

- Initiative by Bayer Healthcare to replicate 67 pre-clinical studies led to reproducibility rate of 20-30% (Prinz et al., 2011)

- Researchers at Amgen were only able to replicate 6 of 53 influential cancer biology studies (Begley & Ellis, 2012)



Why is this happening?

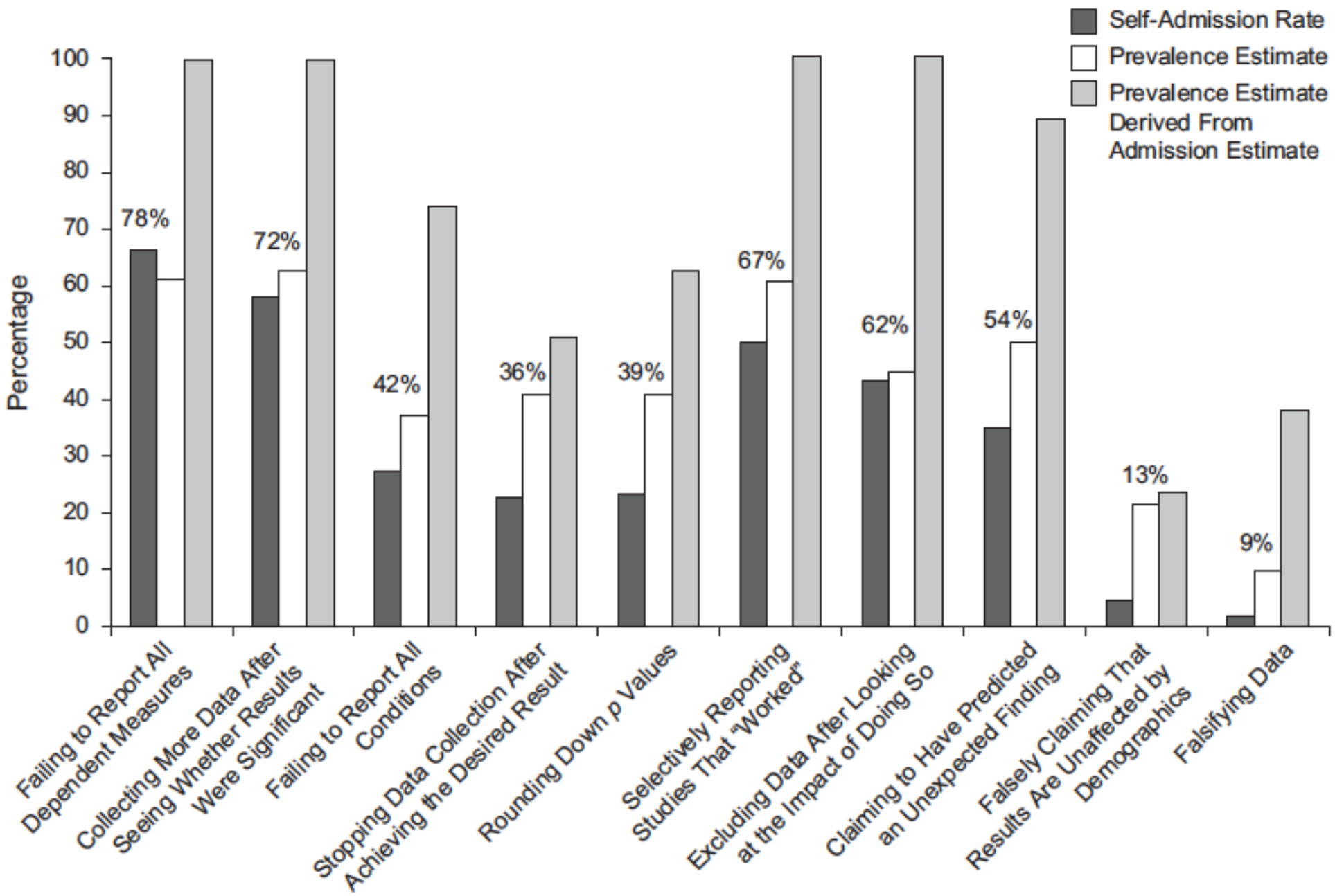
- **Incentive problem**: the publication practice rewards reporting of "optimized data," or statistically significant results (http://bit.ly/29BJrGF)

- **Deliberate fraud:** pretty rare

- **P-hacking:** data mining to uncover statistically significant patterns without an a priori hypothesis

  - "it's **easy to find a p < .05** comparison even if nothing is going on… and **[come] up with good stories** (plausible even to themselves, as well as to their colleagues and peer reviewers) **to back up any statistically-significant comparisons** they happen to come up with."

Questionable Research Practices are Common in Psychology



Garden of Forking Paths (Gelman & Loken, 2013

- **One-to-many** mapping from **scientific** to **statistical hypotheses**

- "[Scientists] start with a somewhat-formed idea in their mind of what comparison to perform, and they **refine** that idea in **light of the data**"

Example

- 2013 study in top psychology journal:
  - Finding: **women at peak fertility 3x more likely to wear red or pink shirts**
- **Implicit choices:**
  - Exclusion criteria: e.g. age of women, sufficient precision in cycle stage
  - Which colors: red and pink, just pink, or just red?
  - What is “peak fertility”?

Thank You!