# The Role of Cherry Picking in the Replication Crisis

Xinran Liu and Samantha F. Anderson, Department of Psychology

Department of Psychology

## Background

In response to a "**crisis of confidence**", where low replication success rates and high-profile replication failures have threatened to undermine trust in social science research, researchers have suggested that **questionable research practices** (QRPs) are a contributor. The influence of QRPs on aspects central to a replicable, cohesive social science literature are underexplored. The common practice of "**cherry picking**" occurs when a researcher collects data on multiple versions of a variable (e.g., dependent variable; DV) but reports results only for the version that offers the strongest possible support for their hypothesis.

## Method

- We simulated original study research literatures where
  1. Researchers engage in cherry picking
  2. All use 2-group, between-subjects design
  3. "Statistically significant" studies published
- Based on these original studies, researchers conduct replication studies

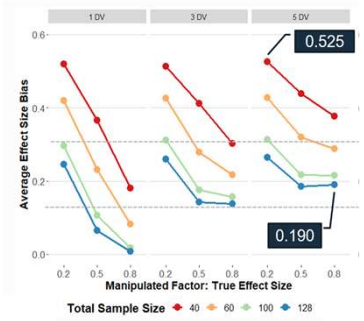
- **Manipulated (Varied) Factors:**
  - True Cohen's *d* effect size: *0, 0.2, 0.5, 0.8*
  - Cherry picking (# of DVs tested): 1, 3, 5
  - Correlation among DVs: $r = .3$, $r = .7$
  - Sample size: *N = 40, 60, 100, 128*

- **Evaluation Criteria**
  - Original study **false positive rate**
  - Original study effect size **bias**
    - 1. In published studies
    - 2. If all studies were published
  - Replication study statistical **power**

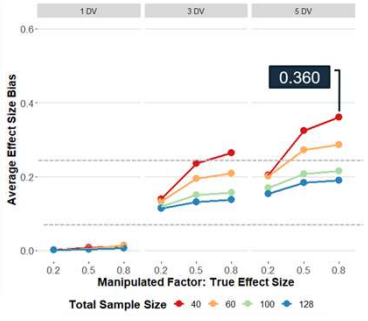
- **Definitions**
  - **False positive rate**: proportion of studies incorrectly rejecting true null hypothesis
  - **Bias**: difference between average reported effect size and true effect size
  - **Power**: proportion of studies correctly rejecting false null hypothesis

## Results

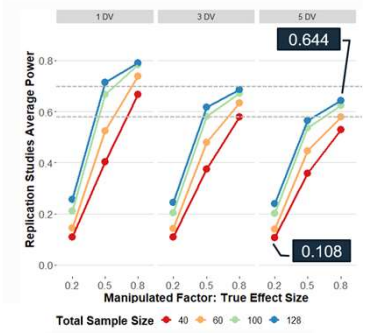
### Effect Size Bias for Published Studies




### Effect Size Bias for All Studies




### Replication Study Average Power




### Original Study False Positive Rate

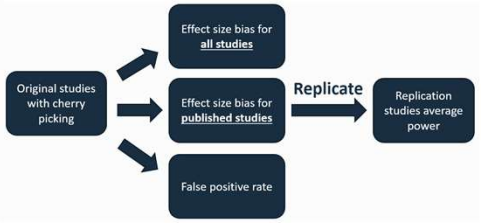
| 1 DV | | | | |
|---|---|---|---|---|
| | 40 | 60 | 100 | 128 |
| 0.2 | 0.089 | 0.121 | 0.165 | 0.204 |
| 0.5 | 0.335 | 0.467 | 0.700 | 0.803 |
| 0.8 | 0.689 | 0.863 | 0.979 | 0.995 |
| **3 DVs** | | | | |
| 0.2 | 0.234 | 0.284 | 0.369 | 0.442 |
| 0.5 | 0.626 | 0.784 | 0.930 | 0.970 |
| 0.8 | 0.934 | 0.988 | 0.999 | 1.000 |
| **5 DVs** | | | | |
| 0.2 | 0.336 | 0.402 | 0.503 | 0.562 |
| 0.5 | 0.759 | 0.886 | 0.976 | 0.993 |
| 0.8 | 0.975 | 0.998 | 1.000 | 1.000 |

*all figures assume $r = 0.3$.

## Research Objectives

We investigate consequences of cherry picking: on "original" studies and "replication" studies:

- Does cherry picking lead to
  - biased original study effect sizes?
  - original study false positives?
  - lower replication study power?




## Conclusions

- Cherry picking → up to 21% false positives
- Cherry picking → effect size bias
  - Stronger effect when:
    - More severe cherry picking (more DVs)
    - Smaller sample size
    - Smaller true effect size
- Cherry picking → lower replication power
  - Researchers rely on published effect sizes to plan replication sample sizes
  - When original study effect size is biased, this hinders replication study power
- In absence of cherry picking, effect sizes still biased due to "publication bias"
  - But if we published every study, cherry picking would have even *larger* effect
- **Cherry picking can distort research literatures and hinder replication. We emphasize rigorous research practices and reporting in the social sciences**

## References

Anderson, S. F. (2022). ***Psychological Methods.*** Anderson, S. F. (2020) ***Psychological Methods.*** Anderson, S. F. (2020). ***Psychological Methods.*** Anderson, S. F. & Maxwell, S. E. (2017). ***Multivariate Behavioral Research.*** Dallow, N., & Fina, P. (2011). ***Pharmaceutical Statistics.*** Gelman, A., & Loken, E. (2014). ***American Scientist.*** John, L. K., Loewenstein, G., & Prelec, D. (2012). ***Psychological Science.*** Judd, C. M. et al. (2012). ***Journal of Personality and Social Psychology.*** Klein, R. A. et al. (2014). ***Social Psychology.*** Lehrer, J. (2010). ***The New Yorker.*** Maxwell, S. E. et al. (2015). ***American Psychologist.*** Monin, B., & Oppenheimer, D. M. (2014). ***Social Psychology.*** Open Science Collaboration (2015). ***Science.*** Schooler, J. (2011). ***Nature.*** Shrout, P. E., & Rodgers, J. L. (2018). ***Annual Review of Psychology.*** Simmons, J. P. et al. (2011). ***Psychological Science.*** Simonsohn, U. et al. (2014). ***Journal of Experimental Psychology: General.*** Wells, G. L., & Windschitl, P. D. (1999). ***Personality & Social Psychology Bulletin.*** Wicherts, J. M. et al. (2016). ***Frontiers in Psychology.***