The Role of Cherry Picking in the Replication Crisis
Xinran Liu and Samantha F. Anderson, 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
- Original Study False Positive Rate

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