Evaluating Meta-Analytic Methods to Detect Outcome Reporting Bias in the Presence of Dependent Effect Sizes

Melissa A. Rodgers & James E. Pustejovsky

THE UNIVERSITY OF TEXAS AT AUSTIN

TEXAS Education
The University of Texas at Austin
College of Education

Methodological Challenges for Meta-Analysis

- Meta-analysis is a set of statistical tools for synthesizing results from multiple, primary studies on a common research topic (Glass, 1976).
- Two common methodological problems in meta-analysis:
  - Outcome Reporting Bias (ORB)
    - Selective reporting and publication based on statistical significance of results (Rothstein et al., 2006).
    - Systematically biases pooled effect estimates and threaten validity of results (Sutton, 2009).
    - Most methods to detect ORB assume univariate effect size estimates (Sutton, 2009).
  - Dependent Effect Sizes
    - Primary studies often contribute multiple, statistically dependent effect sizes.
    - Multiple outcomes, treatment group comparisons, and longitudinal designs (Gleser & Olkin, 2009).
    - Many methods to handle dependency: ad hoc solutions and multivariate models (Becker, 2000).
- Little available research on how to assess the presence of selective outcome reporting when synthesis includes dependent effect sizes.
  - Few methodological and applied studies have incorporated both (e.g. Bediou, 2018, Hwang et al., 2018, Kirkham, 2013, Stevens et al., 2018).
- Need to identify, evaluate, and disseminate methods that simultaneously address both of these challenges.

Simulation Study - Method

- Simulated two-group designs with standardized mean difference effect sizes, based on a two-level model.
- Each study included multiple correlated outcomes, creating dependent effects.
- Analysis conducted using R packages (metaphor::trimfill() and clubSandwich), and custom written R code for 3PSM.
- A one-sided p-value of 0.025 is used to introduce outcome reporting censoring and for one-sided detection tests for outcome reporting bias.

Table 1
Simulation Parameters

Experimental Factors	Levels
True underlying effect size (μ)	0.0, 0.2, 0.5, 0.8
Between-study heterogeneity (τ^2)	0.1, 0.2, 0.4
Number of studies (k)	20, 50, 80
Correlation between outcomes (ρ_i)	
Average (μ_ρ)	0.4, 0.8
Standard Deviation (υ_ρ)	.0001, .05
Outcome Reporting Bias Censoring (π)	0.0, 0.2, 0.4, 0.6, 0.8, 0.9, 1.0

- Dependence Methods & ORB Detection tests:
  - Ignore or aggregate (simple average) and application of univariate detection tests:
    - Trim & Fill (Duval & Tweedie, 2000)
    - 3 Parameter Selection Model (Hedges & Vevea, 2005)
    - Regression Test variants (Egger et al., 1997, Pustejovksy & Rodgers, 2018)
  - Multivariate meta-regression using robust variance estimation to account for dependence, combined with ORB Regression Test Variants.
    - Referred to as Egg Sandwich.
- Performance Criteria:
  - Type I error rates in the absence of outcome reporting bias (π = 0)
  - Power to detect outcome reporting bias when selection introduced at varying levels of censoring (π > 0).

3 Parameter Selection Model

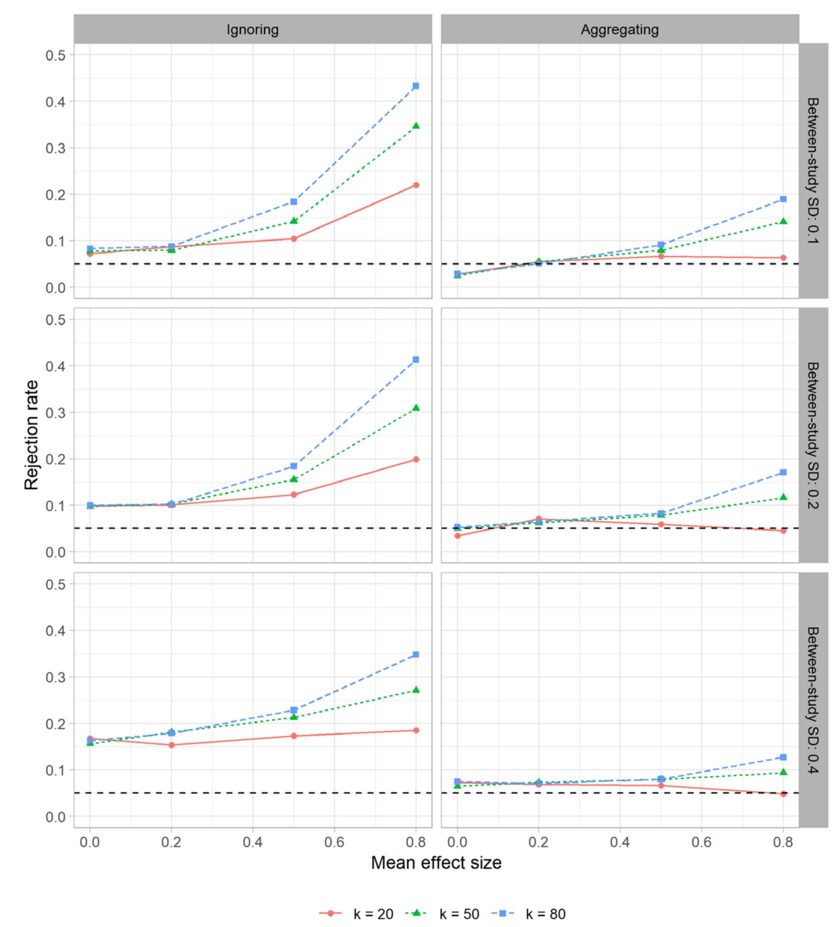

*Figure 1: Type-I error rates for 3PSM test when dependence is ignored, or aggregated for samples of **k = 80** studies.*

Trim & Fill

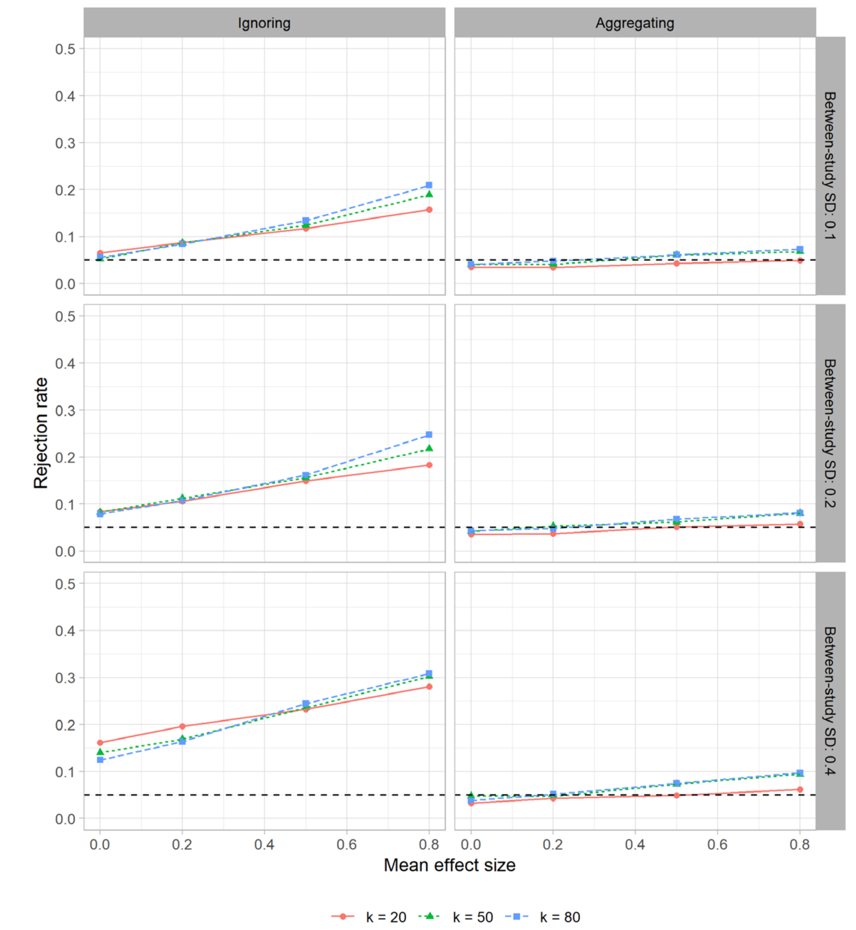

*Figure 2: Type-I error rates for Trim & Fill test when dependence is ignored, or aggregated for samples of **k = 80** studies.*

- Ignoring dependence inflates Type-I error rates for the 3PSM and Trim & Fill methods, especially as true effect size and the study sample (k) increases.
- Aggregation also inflates Type-I error rates for these methods if the true effect size exceeds $\mu = 0.2$.
- Increased heterogeneity and a smaller study sample (k = 20) decreases the rejection rate to the nominal level (α = 0.05) when dependence is aggregated.

Power

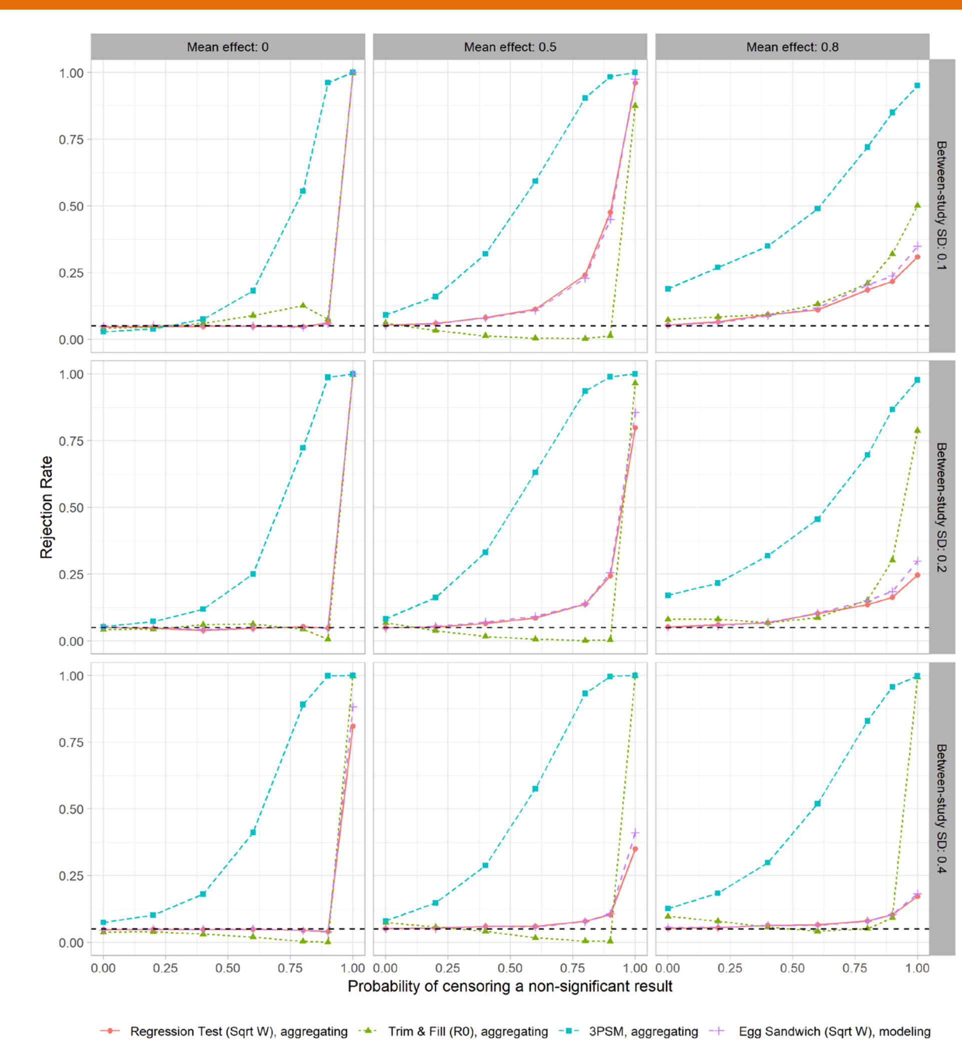

*Figure 4: Power of all methods to detect selective publication when dependent effects are aggregated or modeled for samples of **k = 80** studies.*

- Across degrees of selective publication, Regression Tests rates have limited power, particularly when the true effect size is (μ = 0 or 0.8); adequate power is only obtained with a moderate true effect size (0.5), low heterogeneity, and strong selective publication censoring (π = 1).
- There is no difference in power between the regression tests when dependence is aggregated or modelled.
- The 3PSM has substantially higher power than the other detection tests, but is miscalibrated in the absence of outcome reporting bias (π = 0).

Regression Test Variants

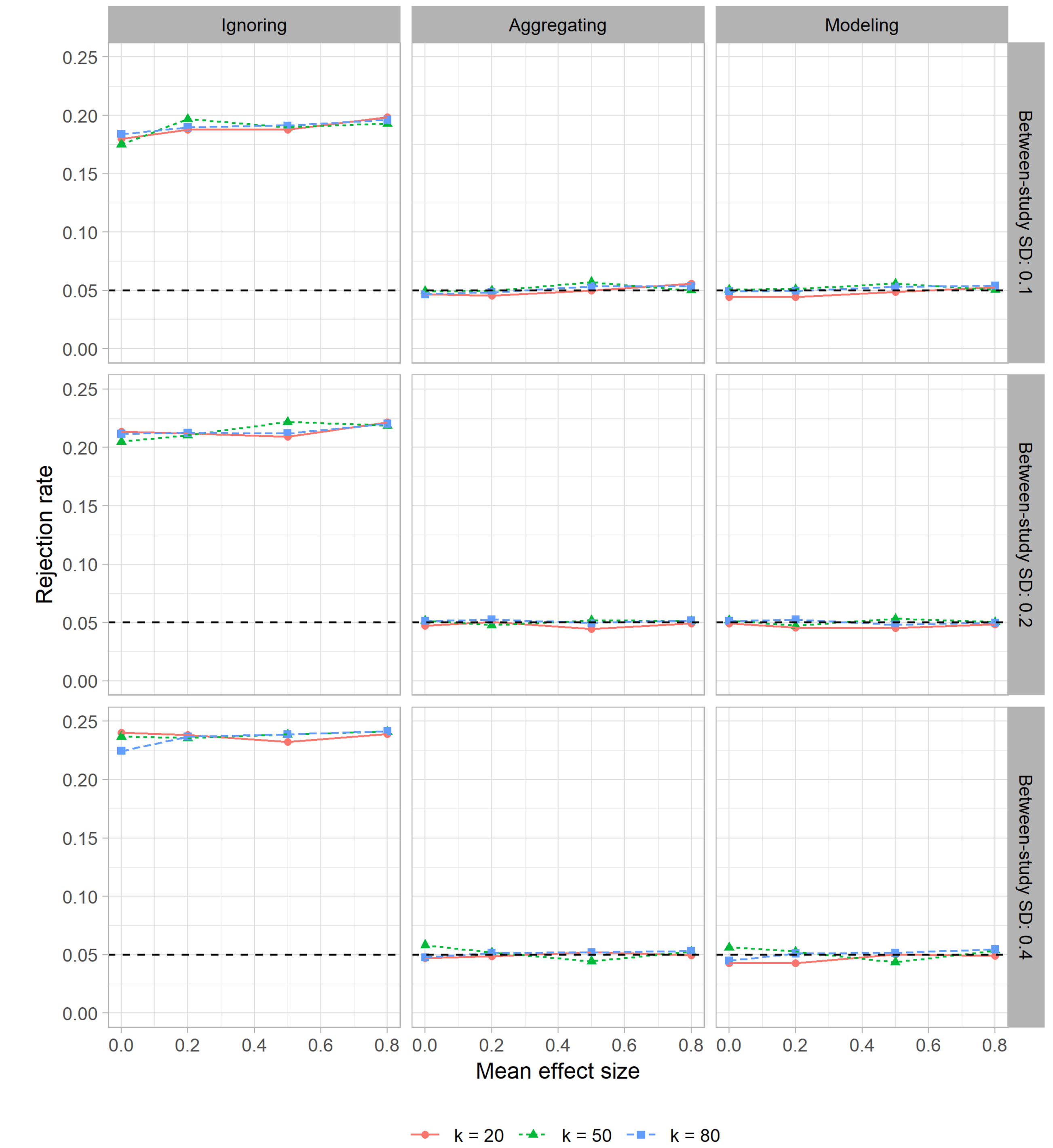

*Figure 3: Type-I error rates for Regression Test variants when dependent effects are ignored, aggregated or modeled for samples of **k = 80** studies.*

- Regression test variants have inflated Type I error rates when dependency is ignored.
- For all levels of heterogeneity and study sample sizes examined, the nominal alpha level is maintained when dependent effects are aggregated or modelled with robust variance estimation.

Discussion

- Results provide guidance to applied researchers who wish to apply valid and powerful methods to detect selective outcome reporting when synthesizing dependent effects.
  - Do not ignore dependence; doing so inflates Type-I error rates for all univariate detection methods evaluated in this study.
  - Regression test variants based on aggregating or modeling dependent effect sizes with robust variance estimation results in proper Type-I error.
  - Regression tests that maintain Type-I error rates have little to no power to detect selection bias, except under strong censoring
  - Power is lower when between study heterogeneity is high.
- Future research should consider developing multivariate methods to test for selective outcome reporting; specifically, refining the 3PSM test to handle dependency.

- Limitations
  - Evaluation of performance with single effect size index (standardized mean difference).
  - Simple, two group between subject design, with correlated multiple outcomes.
  - Limited number of methods available to handle dependence and detect publication bias.

References

Becker, B. J. (2000). Multivariate meta-analysis. In Handbook of applied multivariate statistics and mathematical modeling (pp. 499–525). Elsevier.

Bediou, B., Adams, D. M., Mayer, R. E., Tipton, E., Green, C. S., & Bavelier, D. (2018). Meta-analysis of action video game impact on perceptual, attentional, and cognitive skills. Psychological Bulletin, 144 (1), 77–110. doi:10.1037/bul0000130

Borenstein, M., Cooper, H., Hedges, L., & Valentine, J. (2009). Effect sizes for continuous data. The handbook of research synthesis and meta-analysis, 2, 221-235.

Duval, S., & Tweedie, R. (2000). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association, 95* (449), 89–98.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ, 315* (7109), 629–634.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. Educational Researcher, 5 (10), 3–8.

Gleser, L. J., & Olkin, I. (2009). Stochastically dependent effect sizes. In The handbook of research synthesis and meta-analysis (pp. 357–376). New York: Russell Sage Foundation.

Hedges, L., & Vevea, J. (2005). Selection method approaches. In *Publication bias in meta-analysis: Prevention, assessment, and adjustments* (pp. 145–174). Chichester, England: John Wiley & Sons.

Hwang, H., & DeSantis, S. M. (2018). Multivariate network meta-analysis to mitigate the effects of outcome reporting bias. Statistics in Medicine.

Kirkham, J. J., Riley, R. D., & Williamson, P. R. (2012). A multivariate meta-analysis approach for reducing the impact of outcome reporting bias in systematic reviews. Statistics in Medicine, 31 (20), 2179–2195. doi:10.1002/sim.5356.

Pustejovsky, J. E., & Rodgers, M. (2018). Testing for funnel plot asymmetry of standardized mean differences. *Research Synthesis Methods,* 1-15.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2006). Publication bias in metaanalysis: Prevention, assessment and adjustments. John Wiley & Sons.

Stevens, E. A., Rodgers, M. A., & Powell, S. R. (2017). Mathematics interventions for upper elementary and secondary students: A meta-analysis of research. Remedial and Special Education, 0741932517731887.

Sutton, A. (2009). Publication bias. In The handbook of research synthesis and meta-analysis (pp. 435–445). Russell Sage Foundation.