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

Melissa A. Rodgers & James E. Pustejovsky

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.
- Two common methodological problems in meta-analysis:
  - Outcome Reporting Bias (ORB)
  - Systematically biases pooled effect estimates and threatens validity of results.
  - Most methods to detect ORB assume univariate effect size estimates.
  - Dependent Effect Sizes
  - Primary studies often contribute multiple, statistically dependent effect sizes.
  - Multiple outcomes, treatment group comparisons, and longitudinal designs.
  - Many methods to handle dependency: ad hoc solutions and multivariate models.
- 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.

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 analysis.

3 Parameter Selection Model

- 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.
- Increased heterogeneity and a smaller study sample (k = 20) decreases the rejection rate to the nominal level (α = 0.05) when dependence is aggregated.

Trim & Fill

- 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 modeled with robust variance estimation.

Regression Test Variants

- 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 selectivity 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.

Power

- 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 (α = 1).
- There is no difference in power between the regression tests when dependence is aggregated or modeled.
- The 3PSM has substantially higher power than the other detection tests, but is miscalibrated in the absence of outcome reporting bias (α = 0).

Discussion

- 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.

Table 1: Simulation Parameters

<table>
<thead>
<tr>
<th>Experimental Factors</th>
<th>Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>True underlying effect size (μ)</td>
<td>0.0, 0.2, 0.5, 0.8</td>
</tr>
<tr>
<td>Between-study heterogeneity (ϕ²)</td>
<td>0.1, 0.2, 0.4</td>
</tr>
<tr>
<td>Number of studies (k)</td>
<td>20, 50, 80</td>
</tr>
<tr>
<td>Correlation between outcomes (ρ)</td>
<td>0.4, 0.8</td>
</tr>
<tr>
<td>Average (μ̄)</td>
<td>0.0, 0.2, 0.4, 0.6, 0.8, 0.9, 1.0</td>
</tr>
<tr>
<td>Standard Deviation (σ̄)</td>
<td>0.003, 0.05</td>
</tr>
<tr>
<td>Outcome Reporting Bias Censoring (r)</td>
<td>0.0, 0.2, 0.4, 0.6, 0.8, 0.9, 1.0</td>
</tr>
</tbody>
</table>

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

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

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

Figure 4: Power of all methods to detect outcome reporting when dependent effects are aggregated or modeled.
References


