

## **Abstract Title Page**

**Title:** Consequences of Outcome Reporting Bias in Education Research

**Authors and Affiliations:**

Williams, R.T., University of Memphis

Polanin, J.R., Vanderbilt University

## Abstract Body

### **Background / Context:**

Publication bias is a term that typically refers to the well-known tendency for studies lacking statistically significant results to be less likely to be published in peer-reviewed journals. This happens because authors are less likely to submit, while editors and reviewers are less likely to accept for publication, papers that lack statistically significant results for their primary outcomes (see Dickersin, 2005 for a review).

Outcome reporting bias (ORB) is a phenomenon associated with publication bias (Rothstein, Sutton, & Borenstein, 2005). Conceptually, ORB occurs when an author censors (i.e., does not report) an outcome analysis because the results are not statistically significant. For instance, a researcher may be interested in the effects of a new educational curriculum on a variety of outcomes (e.g., math, reading, and science). During the preliminary stages of the analysis, the researcher's findings indicate that the program has significant effects on two of the three outcomes. Under the assumption of ORB, therefore, the published version of the analysis would include only the statistically significant findings, excluding the non-significant findings.

Unfortunately, most of the extant literature on this phenomenon derived from medicine. Research protocols, developed prior to conducting the study, provide operational details on the study's methods and analysis plan. Chan, Hróbjartsson, Haahr, Gøtzsche, and Altman (2004) collected the protocols of randomized trials reviewed by two scientific-ethical committees (similar to institutional review boards) in Denmark. The researchers compared the outcomes reported in the protocols with the outcomes reported in published reports and found evidence of ORB. 71% of statistically significant outcomes were reported versus 56% of non-significant findings, resulting in an odds ratio of 2.4 (i.e., the odds of an outcome being reported were 2.4 times greater for statistically significant outcomes than the odds for non-statistically significant outcomes). Chan and Altman (2005) found similar results in comparing the protocols for trials funded by the Canadian Institutes of Health Research with their published reports.

Outcome reporting bias represents more than a problem for the understanding of intervention effects, however, as it has the potential to undermine the validity of meta-analyses. If primary researchers do not provide a full and accurate reporting of the study's methods and results, then inferences from the review will likely be biased (Orwin & Cordray, 1985). For example, in a review of 90 meta-analyses published in the journal *Psychological Bulletin*, Ferguson and Brannick (2012) found that most review indicated evidence of bias due to publication status. Of the 90 reviews, 19 directly reported publication status as a moderator of effect size magnitude, with most of the 19 showing significantly smaller effect sizes for unpublished studies. Moreover, a recent simulation study conducted by Francis (2012) illustrated the biased mean effect size estimates possible when failing to account for unpublished studies. The simulation measured extant meta-analyses for the presence of *too much* positive replication, positing that a meta-analysis of only positive intervention effects is itself biased. Sutton (2005) expounded on multiple methodological studies of publication bias, concluding that publication bias has the potential to significantly impact meta-analyses, and should be accounted for (i.e., tested for the presence of) within the review.

The educational sciences require similar investigations. In a recent study, Pigott, Valentine, Polanin, Williams, and Canada (in press) examined dissertations and published research for evidence of ORB. The authors searched and screened dissertations implementing an educational interventions conducted from 2001-2005 across 96 institutions. The screening procedure yielded 621 such studies, of which, 79 had published matches. The authors of the studies conducted 1,599 analyses. The means odds ratio of publishing statistically significant effects (i.e., the odds of a statistically significant outcome in a dissertation appearing in the published version), was 2.41 (95% C.I. = 1.79, 3.25). Said differently, the probability of publishing a statistically significant effect was 71%, whereas the probability of publishing a non-significant effect was only 29%.

Given the implications ORB has on the validity of meta-analyses, and the impact meta-analyses have on policymakers and practitioners, it is important to recognize and mitigate the problem. Moreover, a number of calls have been made to examine publication bias more extensively in the social sciences, especially education (Rothstein, Sutton, & Borenstein, 2005). A thorough and thoughtful examination of the impact of ORB must be considered, and this project intends to extend the extent literature.

**Purpose / Objective / Research Question / Focus of Study:**

The purpose of this study is to investigate the consequences of ORB in education research and how to improve estimation in meta-analysis when ORB is present.

This project is guided by two research questions:

1. What are the consequences of ORB in meta-analyses of treatment effect estimates in education?
2. How does a beta density weight function perform in mitigating the adverse effects of ORB in education research?

The focus of the study is on a simulation project designed to assess the properties of the beta-density function to mitigate the problem of ORB. We intend to provide recommendations for reviewers conducting meta-analyses in education.

**Significance / Novelty of study:**

Previous methodological work has been conducted to guide the hypothesized impact, but to date no empirical studies have investigated the effects of ORB on meta-analytic results in education research. This work examines the effects of ORB and evaluates the performance of the beta density weight function to mitigate those effects.

**Statistical, Measurement, or Econometric Model:**

Citkowicz and Vevea (under review) propose using the beta density weight function to account for the selection process when statistically significant results are favored over nonsignificant results. The beta density is defined as

$$f(p|a, b) = \frac{1}{B(a, b)} p^{a-1} (1 - p)^{b-1}, \quad (1)$$

where  $a$  and  $b$  are the shape parameters,  $p$  is the p-value of the effect size estimate, and  $B(a, b)$ , is the beta function.

Citkowicz and Vevea identify three important reasons for using the beta density instead of alternative selection models. First, only  $a$  and  $b$  are estimated, making it possible to detect and correct for selection bias with a relatively small number of effects. Second, when the parameters  $a$  and  $b$  are fixed at 1.0, the density becomes a uniform distribution, representing the absence of selection bias. This characteristic makes it efficient to compare the results of an adjusted and unadjusted model. Third, the beta density can detect selection in either direction. Though the purpose of this study is to detect selection in a specific direction (i.e. over selection of statistically significant effects), the density is able to detect and adjust for selection in either or both directions.

**Usefulness / Applicability of Method:**

This study illustrates the potential problems for meta-analytic results in the presence of known ORB and it investigates the utility of a possible solution in the beta density weight function.

**Data Collection and Analysis:**

Data for this study were simulated based on parameter estimates from Pigott et al. (in press). We first estimated the effects of know ORB under general or typical study conditions in education research. Pigott et al. showed that the average education intervention in their study had statistical power of about .46. As such we fixed treatment and control group sample sizes to estimate a standardized mean difference of .20. For this part of the simulation study, we generated a population of 1000 standardized mean difference effect size estimates from a normal distribution with mean effect size  $d$ , equal to .20, and variance,  $\sigma^2$ , equal to the variance of the sampling distribution of the mean difference. From that population of effects we randomly sampled  $m$  effects, 40 in this scenario<sup>1</sup>, with replacement, specifying the probability of selection,  $ps$ , for statistically significant effects at .71, consistent with Pigott et al. Between-study heterogeneity was fixed at zero for this part of the analysis (i.e. a fixed effect analysis). A weighted mean effect estimate was generated using inverse variance weights. This process was repeated for 1000 replicates (i.e. 1000 meta-analyses).

As the second part of this work we investigated the performance of the beta density weight function using the same conditions described above.

---

<sup>1</sup> This value was based on a review of the Campbell Collaboration education library of systematic reviews. Across 15 reviews, the mean number of included studies was 36.8.

### **Findings / Results:**

The distribution of these fixed effect unadjusted mean estimates for the typical education meta-analytic scenario, and their 95% confidence intervals, are presented in Figure 1. Across all of the replicates in this simulation, the mean estimate was approximately .24, or about 20% above the expected value. Furthermore, parameter recover was low. These estimates recovered the parameter value only 31% of the time, far below the nominal 95%.

Figure 1 Here

The distribution of fixed effect adjusted mean estimates is presented in Figure 2. The same parameter values used in the first simulation were repeated here, except that the beta density weight function was applied in hopes of correcting for the unbalanced selection process favoring statistically significant estimates.

Figure 2 Here

The results of this simulation were very positive. The mean estimate across all 1000 replicates was about .27. And while this number is still upwardly biased, inference under this model is greatly improved. That is, parameter recovery was very close to the nominal 95%, at about 96.5%.

While the results mentioned here are limited to one set of parameter values, we are currently investigating the performance of the density function under a number of parameter combinations, and under both fixed and random effects models, relevant to education research. The parameter values under investigation are presented in Table 1. We will present the comprehensive results of these simulations as part of the complete paper.

Table 1 Here

### **Conclusions:**

This study investigated the consequences of ORB in education research. It showed that under typical circumstances (i.e. those found from Pigott et al.) the estimated ORB in education research will tend to positively bias meta-analytic results. To combat this problem, we used the beta density weight function as an efficient approach to adjusting for selection bias in meta-analysis (Citkowicz & Vevea, in press). The results indicated that estimates were still upwardly biased on average but parameter coverage of the mean effect was restored to nearly the nominal value. The beta density has great potential as an efficient adjustment to the problem of ORB in education research.

Meta-analysis is a uniquely important part of replication research and it is uniquely dependent on the quality of replicate results. This project fits well with the overall theme of the 2014 SREE conference, *The Role of Replication*, by highlighting the potential pitfalls of biased outcome reporting. In order for meta-analysts to provide insightful information to policymakers and practitioners about what works, reviews in education must be much more equipped to address the consequences of ORB.

## Appendices

### Appendix A. References

- Chan, A., & Altman, D. G. (2005). Identifying outcome reporting bias in randomised trials on PubMed: Review of publications and survey of authors. *BMJ (Clinical Research Ed.)*, 330(7494), 753.
- Chan A, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. (2004). Empirical evidence for selective reporting of outcomes in randomized trials: Comparison of protocols to published articles. *JAMA*, 291, 2457-2465. doi:10.1001/jama.291.20.2457.
- Citkowicz, M., & Vevea, J. L. (under revision). *A parsimonious weight function for modeling publication bias*.
- Dickersin, K. (2005). Publication bias: Recognizing the problem, understanding its origins and scope, and preventing harm. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 11-33). Chichester, UK: Wiley.
- Ferguson, C.J., & Brannick, M.T. (2012). Publication bias in psychological science: Prevalence, methods for identifying and controlling, and implications for the use of meta-analysis. *Psychological Methods*, 17, 120-128.
- Francis, G. (2012). Publication bias and the failure of replication in experimental psychology. *Psychonomic Bulletin & Review*, 19, 975-991.
- Orwin, R. G., & Cordray, D. S. (1985). Effects of deficient reporting on meta-analysis: A conceptual framework and reanalysis. *Psychological Bulletin*, 97, 134-147. doi:10.1037//0033-2909.97.1.134
- Pigott, T.D., Valentine, J.D., Polanin, J.R., & Williams, R.T. (in press). Outcome reporting bias in education doctoral dissertations. *Educational Researcher*.

## Appendix B. Tables and Figures

Not included in page count.

Figure 1. Distribution of unadjusted fixed effect mean difference effect estimates ( $d = .20$ ,  $m = 40$ ,  $ps = .71$ ,  $power = .46$ )

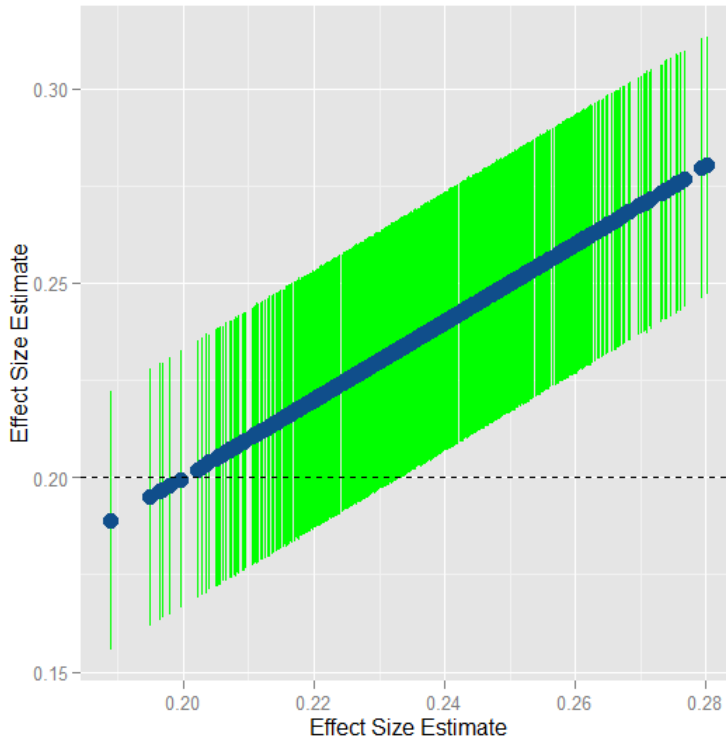


Table 1. Parameter values for simulation study.

Parameter	Value1	Value 2	Value 3
<i>power</i>	.20	.46	.80
<i>m</i>	10	40	60
$I^2$	0	.33	.50
<i>ps</i>	.50	.71	.90

*Note.* *power* is the statistical power of the primary study effects; *m* is the number of effect sampled,  $I^2$  is the proportion of the total heterogeneity that is among samples in a meta-analysis; and *ps* is the probability of selection for statistically significant effects.

Figure 2. Distribution of adjusted fixed effect mean difference effect estimates ( $d = .20$ ,  $m = 40$ ,  $ps = .71$ ,  $power = .46$ )

