Reviews

Summary of reviewers notes and ratings on criteria.

Effect Size, Heterogeneity, and Power in Meta-Analyses of Direct Replications: A Comprehensive Simulation Study

Unit / Sub Division D - Measurement and Research Methodology / Division D - Section 2:
Unit: Quantitative Methods and Statistical Theory

Review Worksheet **(3/3)**

Review #2437723

Criteria	Rate
Objectives or purposes	3 / 5
Perspective(s) or theoretical framework	3 / 5
Methods, techniques, or modes of inquiry	3 / 5
Data sources, evidence, objects or materials	4 / 5
Results and/or substantiated conclusions or warrants for arguments/point of view	3 / 5
Scientific or scholarly significance of the study or work	3 / 5
Relationship to meeting theme	3 / 5

Comments to the Author/Submitter

The study attempts to fit several theoretical distributions to effect size and effect size variance. It is not clear what is accomplished by fitting theoretical distributions. It is likely that the distributions are across research study topics. Moreover, the empirical distribution could be used with Efron's bootstrap to obtain a bootstrap distribution of effect size and effect size variance. The bootstrap distribution would give confidence interval for true effect size and effect size variance.
Effect size and effect size variance are discussed as being correlated, so a simulation should draw from a bivariate distribution instead of two marginal distributions.
In the Preliminary Results section discussed Monte Carlo simulation results as estimates. They are not estimates. They are just simulation results.

Review #2437719

Criteria	Rate
Objectives or purposes	3 / 5
Perspective(s) or theoretical framework	3 / 5
Methods, techniques, or modes of inquiry	2 / 5
Data sources, evidence, objects or materials	3 / 5
Results and/or substantiated conclusions or warrants for arguments/point of view	3 / 5
Scientific or scholarly significance of the study or work	3 / 5
Relationship to meeting theme	2 / 5

Comments to the Author/Submitter

The proposal aims to show how the heterogeneity of effect sizes influences meta-analysis in direct replications. Direct replication is an interesting field in experimental science. However, heterogeneity could happen in any meta-analysis. The heterogeneity can be tested and if the results show the heterogeneity, appropriate approaches can be applied to handle the potential bias. The current proposal focuses on a different population, however it doesn’t improve the methodology of meta-analysis. The topic doesn't fit well in the section.

Review #2437721

Criteria	Rate
Objectives or purposes	4 / 5
Perspective(s) or theoretical framework	4 / 5
Methods, techniques, or modes of inquiry	3 / 5
Data sources, evidence, objects or materials	5 / 5
Results and/or substantiated conclusions or warrants for arguments/point of view	4 / 5
Scientific or scholarly significance of the study or work	3 / 5
Relationship to meeting theme	2 / 5

Comments to the Author/Submitter

This proposal considers an interesting and important problem concerning variation in effect sizes across direct replications. The proposed work will evaluate multiple facets of effect size heterogeneity (relationship with effect size, impact on meta-analytic

estimates, and implications for statistical power), which could provide insight into replication failure rates and help inform the design of direct replication studies. While the authors' simulation-based approach appears generally appropriate, the proposal would be strengthened by offering more detailed information on the methodology, specifically the simulation parameters and the statistical tests that will be used.

Specific comments/questions:

The term “estimated true effects” is somewhat confusing in this context. I believe what the authors are referring to with this term is the estimates of the true effect size. It would be helpful to clarify this terminology

The authors state that the effect size estimates they will use as a basis for their analysis are “accurate estimations of their respective true effect sizes". What is the basis for this claim?

The simulation design is interesting, but it relies on the strong assumption that the distributions for both parameters are: 1) correctly specified, and 2) can be accurately estimated from the observed data. To gauge the plausibility of these assumptions requires validating the fitted distributions on out-of-sample direct replication data. Without doing so, it will be hard to assess the credibility and external validity of the study’s findings.

It is unclear to me why the Monte Carlo simulation step is needed to estimate the averages for the two parameters of interest. Rather than repeatedly sampling from the fitted (closed-form) distributions, why not just calculate these values analytically (e.g., using the standard estimator for the mean of distribution)?

To expand the scope of this work, the authors might consider expanding their simulation study to incorporate additional factors such as publication bias or predictors of heterogeneity

Review Worksheet **(1/1)**

Review #2437730

Criteria	Rate
Objectives or purposes	4 / 5
Perspective(s) or theoretical framework	3 / 5
Methods, techniques, or modes of inquiry	3 / 5

Data sources, evidence, objects or materials	5 / 5
Results and/or substantiated conclusions or warrants for arguments/point of view	4 / 5
Scientific or scholarly significance of the study or work	4 / 5

Comments to the Author/Submitter

This is an interesting point of view that many researchers may ignore.
This paper proposed a new method to demonstrate his arguement and compared the result with a conventional analysis.