Preregistration

Assessing Publication Bias in Meta-Analyses: A Simulation-Based Estimation Approach Based on the Joint Distribution of Effect Size and Sample Size

Jan Luca Schnatz

08 March 2024

Table of contents

Note: This preregistration was written using the secondary data analysis preregistration template by Van den Akker et al. (2021).

1 Study Information

1.1 Working Title

Assessing Publication Bias in Meta-Analyses: A Simulation-Based Estimation Approach Based on the Joint Distribution of Effect Size and Sample Size

1.2 Authors of Preregistration

Jan Luca Schnatz 𝕏

1.3 Research Questions

(1) How does publication bias influence the joint probability distribution of effect size and sample size?

(2) How can the magnitude of publication bias in meta-analyses be estimated and effect sizes under publication bias be corrected from the joint distribution of sample size and effect size?

1.4 Hypotheses

1.4.1 Hypothesis I

The estimated publication bias severity $\hat{\omega}_{\mathrm{PBS}}$ is positively associated with the Fisher z-transformed Spearman correlation coefficients of the unsigned effect sizes and sample sizes in each meta-analysis. In other words, when the proposed method estimates high publication bias (i.e., low probabilities for $\hat{\omega}_{\mathrm{PBS}}$) we expect the correlation coefficients for each meta-analysis to be more negative and conversely. In statistical terms, this expectation implies that the regression coefficient β_1 will be greater than zero.

$$\mathcal{H}_0 : \beta_1 \leq 0$$
$$\mathcal{H}_1 : \beta_1 > 0$$

1.4.2 Hypothesis II

In instances where the true effect size is exactly zero ($\delta = 0$) and substantial publication bias is evident, the distribution of effect sizes and sample sizes displays a symmetric hollowing around zero (Light & Pillemer, 1984). This symmetry results in the difference $\Delta_{\widehat{\mu}_d,\,\hat{\delta}}$ between the estimated $\widehat{\mu}_1$ and the observed mean effect size $\hat{\delta}$ remaining invariant irrespective of the magnitude of publication bias. However, when the true effect size exceeds zero ($\delta > 0$), publication bias leads to an overestimation of the true effect (i.e., $\hat{\delta} > \delta$), and conversely, underestimation (i.e., $\hat{\delta} < \delta$) when $\delta < 0$. If the mean parameter $\widehat{\mu}_d$ from the Gaussian effect distribution is a better estimator of the true effect size δ under publication bias than the average effect size $\hat{\delta}$, then the relationship between the difference $\Delta_{\widehat{\mu}_d,\,\hat{\delta}}$ of those parameters and the publication bias severity $\widehat{\omega}_{\mathrm{PBS}}$ should follow a curvilinear inverse U-shaped pattern. In other words, when the observed difference $\Delta_{\widehat{\mu}_d,\,\hat{\delta}}$ is approaching zero, publication bias severity is expected to decrease (indicated by larger values for $\widehat{\omega}_{\mathrm{PBS}}$). Conversely, when the difference $\Delta_{\widehat{\mu}_d,\,\hat{\delta}}$ diverges from zero in both negative and positive directions, publication bias severity is expected to increase (i.e. lower values for $\widehat{\omega}_{\mathrm{PBS}}$). In statistical terms, it is anticipated that the quadratic term β_2 of the regression model is smaller than zero.

$$\begin{aligned} \mathcal{H}_0 : \quad & \beta_2 \geq 0 \\ \mathcal{H}_1 : \quad & \beta_2 < 0 \end{aligned}$$

1.4.3 Hypothesis III

The mean $\widehat{\mu}_d$ of the Gaussian effect size distribution is equivalent to the average effect size $\hat{\delta}$ across multisite replication studies within the equivalence bounds Δ_L and Δ_U.

$$\Delta_{\widehat{\mu}_d,\,\hat{\delta}} = \widehat{\mu}_d - \hat{\delta}$$

$$\mathcal{H}_{01} : \Delta_{\widehat{\mu}_d,\,\hat{\delta}} \leq \Delta_L \quad \cap \quad \mathcal{H}_{02} : \Delta_{\widehat{\mu}_d,\,\hat{\delta}} \geq \Delta_U$$
$$\mathcal{H}_1 : \Delta_L > \Delta_{\widehat{\mu}_d,\,\hat{\delta}} > \Delta_U$$

1.4.4 Hypothesis IV

In the context of large-scale replication research, particularly within scholarly publishing frameworks like registered reports, where study protocols undergo peer review and approval prior to data collection, publication bias should be inherently absent. Consequently, it is expected that the estimated publication bias severity $\widehat{\omega}_{\mathrm{PBS}}$ for multisite replication studies is greater (i.e., lower

publication bias) in comparison to normal meta-analysis. In statistical terms, when the regressor is a binary indicator of the type of research synthesis (i.e., multisite replications as the reference group and normal meta-analysis) and the outcome is the estimated publication bias severity $\hat{\omega}_{PBS}$, the regression coefficient β_1 is expected to be greater than zero.

$$\begin{aligned} \mathcal{H}_0: & \quad \beta_1 \leq 0 \\ \mathcal{H}_1: & \quad \beta_1 > 0 \end{aligned}$$

2 Data Description

2.1 Dataset Description

To answer the second research question, we require empirical data from an extensive dataset covering a wide range of meta-analyses across ideally multiple psychological subfields as well as multisite replication studies where information regarding both sample size and effect sizes has been systematically collected. Given these criteria, the dataset used for this secondary data analysis is sourced from the primary study conducted by Linden & Hönekopp (2021). Within this study, the authors investigated effect size heterogeneity in meta-analyses across cognitive, organizational, and social psychology as well as large-scale multisite replication studies. The dataset includes 150 meta-analyses and 57 multisite replication projects, encompassing a diverse array of primary studies within each subdiscipline.

2.2 Public Availability

The dataset is openly available through the Open Science Framework (OSF). This includes supplementary information on the meta-analyses and close replications as well as raw data files. There are no barriers to accessing the data, and it is publicly available for further analysis.

2.3 Data Access

The dataset can be accessed directly through the provided DOI link: https://doi.org/10.17605/osf.io/yr3xd. The raw data is located in the *supplementary info - raw data for MA and close reps.xlsx* file and the supplementary data regarding the reference for each meta-analysis / multiple replication is located in the *Supplementary information and tables - reviewed May 2020.pdf* file.

2.4 Access Date

The dataset was downloaded on 03.10.2023 and first accessed on 29.11.2023 by Jan Luca Schnatz.

2.5 Data Collection Method

The data collection method described aligns with the procedures outlined in Linden & Hönekopp (2021). For each specified subfield (cognitive, organizational and social psychology), data from 50 meta-analyses were gathered, driven primarily by feasibility considerations rather than power considerations.

The inclusion criteria for each meta-analysis were as follows:

1. The meta-analysis / close replication had to address a substantive psychological effect.
2. Effect sizes must be reported as standardized mean differences (Cohen´s d or Hedges´ g) or as correlations (Pearson´s r or Fisher´s z).
3. Effect sizes and sample-size information of the original primary studies must be provided.
4. The article must be available in the English language.

The data collection procedure involved two steps: initially identifying a sufficient number of eligible meta-analyses and subsequently constructing a subset of 50 meta-analyses for each subfield. To locate eligible studies, the authors systematically searched the PsycINFO database using the query "meta-analy" (journals only) in the abstract field, restricting the search to PsycINFO classifications "3000 Social Psychology," "3600 Industrial and Organizational Psychology," "2340 Cognitive Processes," "2343 Learning and Memory," and "2346 Attention." Because this initial search did not yield enough eligible meta-analyses, the authors conducted a systematic search in the Web of Science database (articles only) using the query "meta-analy," limited to the categories "Psychology Social," "Psychology Applied," and "Psychology." In the second step, all eligible meta-analyses were examined for inclusion criteria in random order until the authors reached the target of 50 meta-analyses per subfield.

2.6 Data Codebook

The references for each meta-analysis and replication study are available within the OSF repository, specifically in the *Supplementary Information and Tables - reviewed May 2020.pdf* file. The raw data is organized in Excel files, with each meta-analysis or close replication located in a separate sheet. Each sheet consists of two columns: *N* representing sample size and *d_obs* representing the observed effect sizes measured as Cohen's d. Each row within the sheet corresponds to an individual primary study.

3 Variables

3.1 Manipulated Variables

Not applicable

3.2 Measured Variables

Measured variables include the sample size of primary studies and the observed effect size (Cohen's d) in both meta-analyses and close replication studies. In instances where a meta-analysis or replication study used a different effect size measure than Cohen's d, the authors converted the effect size measure accordingly.

3.3 Units of Analysis

The units of analysis include all 150 meta-analyses from cognitive, organizational, and social psychology, as well as all 57 multisite replication projects, collectively reflecting 9232 effect sizes and sample size (7227 primary studies of meta-analyses and 2005 primary replication studies).

3.4 Missing Data

No missing data is reported in the dataset, all required information is available for the specified variables.

3.5 Outliers

No outliers will be removed from the dataset.

3.6 Sampling Weights

No sampling weights are applied in this dataset.

4 Knowledge of Data

4.1 Previous Work with Dataset

The data was not used previously.

4.2 Prior Knowledge

The conceptualization of the study was partly influenced by a recent preprint authored by Linden et al. (2023), wherein the same data of Linden & Hönekopp (2021) was used as the present study. In this study Linden et al. (2023) investigated the relationship between sample size and effect size in meta-analysis and close replications by means of the correlation between sample size and effect size (using Spearman and Pearson correlations). The author of the present study checked the shared analysis scripts of Linden et al. (2023) for computational reproducibility, thus engaging in the dataset prior to the preregistration. However, the present simulation-based estimation method for publication bias and sample size planning (see Statistical Analyses), is a novel approach not undertaken before this preregistration using the data.

5 Statistical Analyses

5.1 Simulation-based Publication Bias Estimation and Effect Size Correction (SPEEC)

5.1.1 Simulation Framework

The simulation-based framework for estimating the magnitude of publication bias in a single meta-analysis involves the following key steps:

1. Simulation of theoretical joint population distributions for sample size and effect size from assumed marginal distributions (see section Marginal and Joint Distributions)

2. Application of publication bias (see section Selection Bias Definition and Selection Bias Application)

3. Comparison of the distance between the theoretical simulated joint distribution and the empirical joint distribution of the meta-analysis (see section Comparison of Empirical and Simulated Joint Distribution)

4. Algorithmic optimization of the bias and distribution parameters to minimize the distance between theoretical and empirical joint distribution (see section Algorithmic Optimization of Parameters)

The simulation framework to estimate the magnitude of publication bias and sample size planning is implemented as an open-source R package called speec by the author of the preregistration. The package can be found at the following URL: https://github.com/jlschnatz/speec

5.1.2 Marginal and Joint Distributions

To simulate random samples from the joint distribution of sample size and effect size, it is necessary to specify their assumed marginal distributions. We simulate k observations (i.e., single studies) for both marginal distributions where i is the index of a single study.

The marginal distribution for sample size ν is inherently a discrete distribution. When modelling count data, researchers commonly utilize the Poisson distribution and negative-binomial distribution. In the context of psychology, empirical sample size distributions often exhibit considerable variance and skewness (Cafri et al., 2010; Marszalek et al., 2011; Sassenberg & Ditrich, 2019; Shen et al., 2011; Szucs & Ioannidis, 2017). This variability cannot be captured independently of the mean in a Poisson distribution. Therefore, the negative binomial distribution can be considered the more appropriate choice as it is able to account for overdispersion by introducing a second distribution parameter. The alternative parametrization with the dispersion parameter ϕ_n and mean parameter μ_n will be used to model the total sample size:

$$n_1, n_2, ..., n_k \quad \text{where} \quad N \stackrel{\text{i.i.d.}}{\sim} \mathcal{NB}(\phi_n, \mu_n) \quad \text{for } i = 1, ..., k \tag{1}$$

Regarding the assumed marginal distribution for effect size δ, we assume a normal distribution with parameters the mean μ_d and variance σ_d^2 that stems from a two-sample t-test design. To account for the fact that as the sample size increases, the precision in estimating δ also increases, we calculate the standard error from the means $\bar{x}_i$ from which the effect sizes arise. Next we calculate a normalization factor γ_i so that $\overline{\gamma} = 1$:

$$\begin{aligned} \sigma_{\bar{x}_i}^2 &= \sigma_d^2 / n_i \\ \gamma_i &= \sigma_{\bar{x}_i}^2 / \overline{\sigma^2}_{\bar{x}} \end{aligned} \tag{2}$$

With this normalization factor the total variance of the individual variances $\text{Var}(\gamma \cdot \sigma_d^2) = \sigma_d^2$. The resulting study-specific effect sizes d_k are defined as

$$d_1, d_2, ..., d_k \quad \text{where} \quad D \sim \mathcal{N}(\mu_d, \gamma_i \cdot \sigma_d^2) \quad \text{for } i = 1, ..., k \tag{3}$$

5.1.3 Selection Bias Definition

5.1.3.1 Publication Bias

Selection bias, in the form of publication bias, arises when the likelihood of a study being published is conditional on its statistical significance (Dickersin, 1990). To simulate publication bias selection, a two-tailed p-value is computed for each individual study i given the total sample size n_i and effect size d_i assuming the effect sizes originate from a two-sample t-test design. If the total sample size is even, the individual sample sizes of the two groups n_{1i} and n_{2i} are simply

chosen to be $n_i/2$. If the total sample size is odd, the individual sample sizes of the two groups are $\lceil n_i/2 \rceil$ and $\lfloor n_i/2 \rfloor$ respectively.

The calculation of each p-value p_i is derived from its empirical t-value t_i

$$t_i = |\frac{d_i}{\sqrt{\frac{1}{n_{1i}} + \frac{1}{n_{2i}}}}| \tag{4}$$

$$p_i = 2 \cdot P(t_i \mid df_i) \tag{5}$$

where $P(t_i \mid df_i)$ is the cumulative central t-distribution with degress of freedom $df_i = n_{1i} + n_{2i} - 2$. Given each p-value p_i, publication bias is introduced by assigning each study i a weight

$$\omega_{\text{PBS}_i}(p_i) = \begin{cases} \omega_{\text{PBS}} & \text{for} \quad p_i \geq \alpha \\ 1 & \text{otherwise} \end{cases} \tag{6}$$

given $\omega_{\text{PBS}} \in \mathbb{R} : 0 \leq \omega_{\text{PBS}} \leq 1$. This weight denotes the probability of a study being selected conditional on the p-value and the type I error rate α. If $p_i \geq \alpha$, a publication bias weight ω_{PBS} is assigned, else the probability of a study being selected is 1, thus indicating no publication bias. The type I error rate is fixed at the common threshold of $\alpha = .05$.

5.1.4 Selection Bias Application

Following the computation of bias weights for publication bias, the selection bias applied to the random samples (d_i, n_i) can be defined as

$$P(S_i = 1) = \omega_{\text{PBS}_i} \tag{7}$$

where S_i is a binary function that indicates if a study was selected.

$$S_i = \begin{cases} 0 & \text{not selected} \\ 1 & \text{selected} \end{cases} \tag{8}$$

Consequently, we can define the selected effect sizes and sample sizes after the application of selection bias as

$$(d_i', n_i') = (d_i, n_i \mid S_i = 1) \quad \text{for} \quad i = 1, \dots, i = k \tag{9}$$

and the effect sizes and sample sizes that were censored within the selection process as

$$(d_i'', n_i'') = (d_i, n_i \mid S_i = 0) \quad \text{for} \quad i = 1, \dots, i = k \tag{10}$$

5.1.5 Comparison of Empirical and Simulated Joint Distribution

To estimate the bias parameter and distributional parameters, we employ a distance measure to quantify how different the simulated theoretical joint distribution is from the empirical joint distribution. This distance measure serves as a loss function for parameter optimization. The first step to compute the loss function involves the estimation of the joint probability density of the empirical and theoretical distributions via two-dimensional kernel density estimation (KDE) with a bivariate normal kernel that is evaluated on a square grid. This step is implemented using the bkde2D() function from the R package *KernSmooth* (Wand, 2023) using the bandwidth method by Sheather & Jones (1991) and the grid size g is chosen to be $g = 2^7 + 1$ in each direction. The limits of the KDE estimation, defining the bounds of the rectangle covered by the grid, are determined based on the empirical joint distribution of sample size and effect size. More specifically, the upper and lower quantiles (q_{l_n}, q_{l_d}, q_{u_n} and q_{u_d}) of the inner 99th percentile of the cumulative distributions are computed using the ML estimates for the distribution parameters of the normal distribution and the negative binomial distribution. The bounds are then defined as the absolute minimum/maximum of the calculated quantiles and the empirical random samples of sample size and effect size ($\min \mathbf{d} \wedge q_{l_d}$, $\max \mathbf{d} \wedge q_{u_d}$) and ($\min \mathbf{n} \wedge q_{l_n}$, $\max \mathbf{n} \wedge q_{u_n}$). This is done to ensure that both the empirical and theoretical joint distribution are evaluated on the same bounds. From the estimated joint density the divergence between the empirical and theoretical joint distributions is computed using the Kullback-Leibler divergence (Kullback & Leibler, 1951). The KL divergence for a discrete set of possible outcomes as returned from the 2-dimensional KDE is defined as

$$D_{\mathsf{KL}}(P \parallel Q) = \sum_{x \in X} p(x) \ln \left(\frac{p(x)}{q(x)} \right) \tag{11}$$

Given the estimated joint density of the empirical distribution $\widehat{f_e}$ and the theoretical distribution $\widehat{f_t}$, the KL divergence is computed using the formula

$$D_{\mathsf{KL}}(\widehat{f_e} \parallel \widehat{f_t}) = \sum_{u=1}^{g} \sum_{v=1}^{g} \widehat{f_e}(u, v) \ln \left(\frac{\widehat{f_e}(u, v)}{\widehat{f_t}(u, v)} \right) \tag{12}$$

5.1.6 Algorithmic Optimization of Parameters

To determine the magnitude of publication bias we will use an optimization approach to minimize the KL divergence, which will serve as the loss function. Fixed parameters in the optimization will include the type error $\alpha = .05$. Parameters that are optimized are the four distribution parameters of the joint distribution of sample size and effect size $[\mu_d, \sigma_d^2, \phi_n, \mu_n]$ and the bias parameter for publication bias ω_{PBS}. The specific details of the optimization approach are currently under consideration.

5.2 Inference Testing of Hypotheses

5.2.1 Hypothesis I

We will analyze the first hypothesis utilizing a beta-regression model. Specifically, we will use the *betareg* R package (Cribari-Neto & Zeileis, 2010) to conduct the analysis with a logit link function to model the mean μ of the beta-distributed outcome (for all other hypotheses that rely on beta-regression, the same package will be used).

$$g(\mu_i) = \beta_0 + \beta_1 x_i$$
$$g(\mu) = \log(\frac{\mu}{1 - \mu})$$

The outcome variable will be the estimated publication bias weight $\hat{\omega}_{\text{PBS}}$ and the predictor variable the Fisher z-transformed correlation coefficients z_r of the relationship between effect size and sample size from each of the 150 meta-analyses. We will test β_1 using a one-sided test and expect β_1 to be positive (see Hypothesis I).

5.2.2 Hypothesis II

To test hypothesis III, we will use a beta-regression model with a logit link. The outcome will be the estimator $\hat{\omega}_{\text{PBS}}$ of the meta-analyses and the regressor will be the difference $\Delta_{\hat{\mu}_d, \hat{\delta}}$ between the estimated mean $\hat{\mu}_d$ of the normal effect size distribution and the average effect size $\hat{\delta}$. We will test the quadratic term β_2 using a one-sided test and expect that $\beta_2 < 0$ (see Hypothesis II).

$$g(\mu_i) = \beta_0 + \beta_2 x_i^2$$

5.2.3 Hypothesis III

To test whether μ_d is equivalent to the average meta-analytical and multisite replication study effect size estimate $\hat{\delta}$ within the specified equivalence bounds, we will conduct an equivalence test using the TOST (two one-sided tests) procedure for two dependent means (Lakens, 2017). We will perform the analysis with the *TOSTER* R package (Caldwell, 2022). The SESOI will define the equivalence bounds and will be set based on a sensitivity power analysis (see Smallest Effect Size of Interest).

5.2.4 Hypothesis IV

To test hypothesis IV, we will again use a beta-regression model with a logit link. The outcome will be $\hat{\omega}_{\text{PBS}}$ of the meta-analyses and multisite replications and the regressor will be a binary indicator variable which specifies the type of research synthesis (normal meta-analysis MA or multisite replication MR). The reference level will be the normal multisite replication studies. We will test the regressor β_1 using a one-sided test and expect that $\beta_1 < 0$.

$$D_i \begin{cases} 0 & \text{for MA} \\ 1 & \text{for MR} \end{cases}$$
$$g(\mu_i) = \beta_0 + D_i\beta_1$$

5.3 Smallest Effect Size of Interest

For all four hypotheses, we will establish the smallest effect size of interest (SESOI) based on effect sizes that we can reliably detect, considering the constraints imposed by the sample size resources available for this secondary data analysis (Lakens, 2014). More specifically, we conducted three simulation-based ($\mathcal{H}_1$, $\mathcal{H}_2$, $\mathcal{H}_4$) and one analytical ($\mathcal{H}_3$) sensitivity power analysis to determine which effect sizes we have at least 80% power to detect, taking into account the constraints of the sample size and a fixed significance level $\alpha = .05$ (details see section Power Analysis). The resulting SESOI for each hypothesis is presented in Table 1. The SESOI for the equivalence hypothesis will define the equivalence bounds for the TOST procedure ($\Delta_L = -0.17$ and $\Delta_U = 0.17$).

Table 1

SESOI Specification for each Hypothesis

Hypothesis	SESOI	Units
$\mathcal{H}_1$	1.28	*OR*
$\mathcal{H}_2$	0.59	*OR*
$\mathcal{H}_3$	0.17	raw
$\mathcal{H}_4$	1.28	*OR*
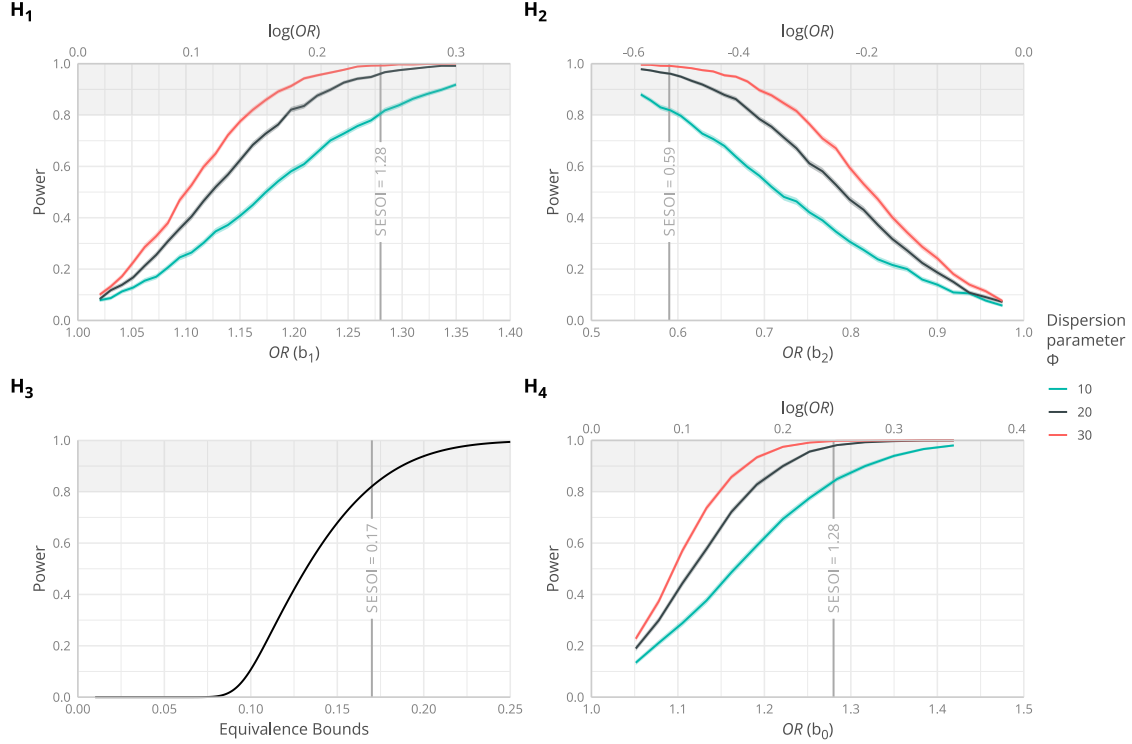
5.4 Power Analysis

The simulation-based sensitivity power analysis targeted a statistical power $1-\beta$ of 0.8 with a fixed type I error rate of $\alpha = .05$. Samples sizes varied across hypotheses: $n = 150$ for hypotheses 1 (only meta-analysis) and $n = 207$ for hypotheses 2 and 4 (both meta-analysis and multisite replication studies) and $n = 57$ (only multisite replication studies) for hypothesis 3. Predictor variables´ distribution assumptions were specified as follows: Hypothesis 1 assumed a normal distribution ($\mu = -0.1; \sigma = 0.5$) for the Fisher z-transformed sample size effect size correlation coefficients, with the linear regression coefficient as the parameter of interest. Hypothesis 3 assumed equal means $\Delta = 0$ and a standard deviation $\sigma_{\text{diff}} = \sqrt{0.3^2 + 0.3^2}$ for the difference scores of $\hat{\mu}_d$ and $\hat{\delta}$, with the same assumptions for hypothesis 2, incorporating a quadratic regression coefficient as the parameter of interest. Beta-regression on $\hat{\omega}_{\text{PBS}}$ in hypotheses 1,

2, and 4 involved simulations for different dispersion parameter $\phi = \{10, 20, 30\}$, as lower dispersion parameters result in reduced statistical power. We chose a conservative approach to define the SESOI for the parameters of interest ensuring a minimum power of 80% for the smallest simulated dispersion parameter $\phi = 10$.

5.5 Inference Criteria

For all hypotheses, we will use frequentist inference criteria. We will conclude that each of the null hypotheses is rejected if the p-value is smaller than .05 ($\alpha = .05$).

5.6 Violation of Statistical Assumptions

One important assumption of the beta regression models is that $0 < \hat{\omega}_{\mathrm{PBS}} < 1$. It is important to note that the optimization approach permits $\hat{\omega}_{\mathrm{PBS}}$ values to vary within the range of $0 \leq \hat{\omega}_{\mathrm{PBS}} \leq 1$ in principle. To handle this problem, we will utilize a transformation proposed by Smithson & Verkuilen (2006). The transformed variable is expressed as $y' = \frac{y \cdot (n-1) + 0.5}{n}$, where n denotes the sample size. This transformation subtly adjusts the bounds to slightly narrow the range between 0 and 1.

5.7 Sensitivity Analysis

Before estimating the parameters of interest using the optimisation approach, we will test different hyperparameter constellations with simulated meta-analytical data to identify a configuration for which the estimated parameters are stable and accurate across optimisations.

5.8 Exploratory Analysis

Regarding hypotheses II, we will explore, whether the data is equidispersed (dispersion remaining invariant across different values of $\Delta_{\hat{\mu}_d, \hat{\delta}}$) or whether there is further dependence of the dispersion parameter ϕ on the regressor $\Delta_{\hat{\mu}_d, \hat{\delta}}$ to account for the heteroskedasticity. We will perform a likelihood ratio test to compare both models.

6 Integrity Statement

The author of this preregistration stated that they filled out this preregistration to the best of their knowledge and that no other preregistration exists pertaining to the same hypotheses and dataset.

7 References

Cafri, G., Kromrey, J. D., & Brannick, M. T. (2010). A meta-meta-analysis: Empirical review of statistical power, type i error rates, effect sizes, and model selection of meta-analyses published in psychology. *Multivariate Behavioral Research*, *45*(2), 239–270.

Caldwell, A. R. (2022). *Exploring equivalence testing with the updated TOSTER r package*.

Cribari-Neto, F., & Zeileis, A. (2010). Beta regression in R. *Journal of Statistical Software*, *34*(2), 1–24. https://doi.org/10.18637/jss.v034.i02

Dickersin, K. (1990). The existence of publication bias and risk factors for its occurrence. *Jama*, *263*(10), 1385–1389.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, *22*(1), 79–86.

Lakens, D. (2014). Performing high-powered studies efficiently with sequential analyses. *European Journal of Social Psychology*, *44*(7), 701–710.

Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, *8*(4), 355–362.

Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research*. Harvard University Press.

Linden, A. H., & Hönekopp, J. (2021). Heterogeneity of research results: A new perspective from which to assess and promote progress in psychological science. *Perspectives on Psychological Science*, *16*(2), 358–376. https://doi.org/10.1177/1745691620964193

Linden, A. H., Pollet, T. V., & Hönekopp, J. (2023). Publication bias in psychology: A closer look at the correlation between sample size and effect size. *PsyArXiv*. https://doi.org/10.31234/osf.io/s4znd

Marszalek, J. M., Barber, C., Kohlhart, J., & Cooper, B. H. (2011). Sample size in psychological research over the past 30 years. *Perceptual and Motor Skills*, *112*(2), 331–348.

Sassenberg, K., & Ditrich, L. (2019). Research in social psychology changed between 2011 and 2016: Larger sample sizes, more self-report measures, and more online studies. *Advances in Methods and Practices in Psychological Science*, *2*(2), 107–114.

Sheather, S. J., & Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society: Series B (Methodological)*, *53*(3), 683–690.

Shen, W., Kiger, T. B., Davies, S. E., Rasch, R. L., Simon, K. M., & Ones, D. S. (2011). Samples in applied psychology: Over a decade of research in review. *Journal of Applied Psychology*, *96*(5), 1055.

Smithson, M., & Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods*, *11*(1), 54. https://doi.org/https://doi.org/10.1037/1082-989X.11.1.54

Szucs, D., & Ioannidis, J. P. (2017). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLoS Biology*, *15*(3), e2000797.

Van den Akker, O., Weston, S., Campbell, L., Chopik, B., Damian, R., Davis-Kean, P., Hall, A., Kosie, J., Kruse, E., Olsen, J., et al. (2021). Preregistration of secondary data analysis: A

template and tutorial. *Meta-Psychology*, *5*, 2625. https://doi.org/10.15626/MP.2020.2625

Wand, M. (2023). *KernSmooth: Functions for kernel smoothing supporting wand & jones (1995)*. https://CRAN.R-project.org/package=KernSmooth