Selective outcome reporting in meta-analysis of dependent effect size estimates

James E. Pustejovsky

University of Wisconsin - Madison

Stanford Quantitative Sciences Unit

February 8, 2022

Outline

- Selective reporting in meta-analysis
- Dependent effect sizes
- A generalized excess significance test

Selective reporting in meta-analysis

Dependent effect sizes

A generalized excess significance test

Selective reporting of primary study findings

- **Selective reporting** occurs if "affirmative" ("positive") findings are more likely to be reported and available for inclusion in meta-analysis.
    - Bias in the publication process (journal/editor/reviewer incentives).
    - Strategic decisions by authors.
- Strong concerns about selective reporting across social, behavioral, and health sciences.
    - Registries of medical trials (Chan et al., 2004; Turner et al., 2008) and social science survey experiments (Franco et al., 2014).
    - Surveys of social science researchers (John, Loewenstein, & Prelec, 2012; Fiedler & Schwarz, 2016).
    - Systematic reviews of dissertations (Pigott et al., 2013; O'Boyle, Banks, & Gonzalez-Mule, 2016; Cairo et al., 2020)
- For a given meta-analysis, we expect strength of selection to depend on
    - Rigor of the systematic review search process.
    - Whether effect sizes are from focal or ancillary analysis.

Implications of selective reporting for meta-analysis

- Selective reporting **distorts the evidence base** available for systematic review/meta-analysis.
    - Inflates average effect size estimates from meta-analyses.
    - Biases estimates of heterogeneity (Augusteijn et al., 2019).

- When conducting a meta-analysis, we need to investigate:
    - Whether selective reporting is of concern (***detecting*** selective reporting)
    - Extent of biases arising from selective reporting (***correcting*** for selective reporting)

Tools for investigating selective reporting

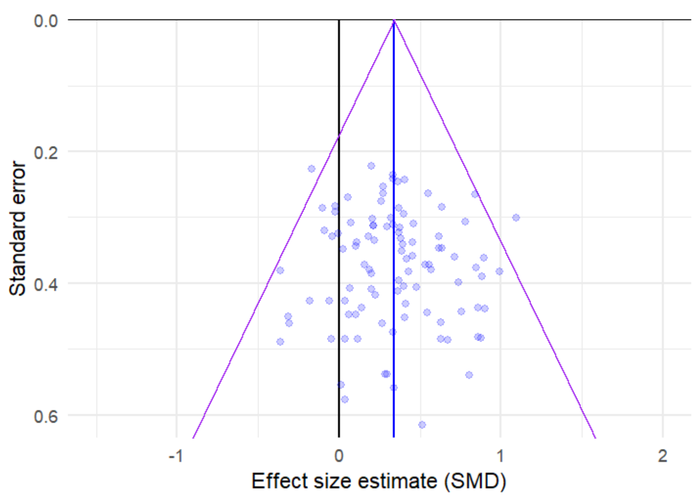
- Graphical diagnostics
    - Funnel plots
    - Contour-enhanced funnel plots
    - Power-enhanced funnel plots (sunset plots)

- Tests/adjustments for funnel plot asymmetry
    - Trim-and-fill
    - Egger's regression
    - PET/PEESE
    - Kinked meta-regression
- Selection models
    - Weight-function models
    - Copas models
    - Sensitivity analysis
- p-value diagnostics
    - Test of Excess Significance
    - p-curve
    - p-uniform / p-uniform*

Funnel plots

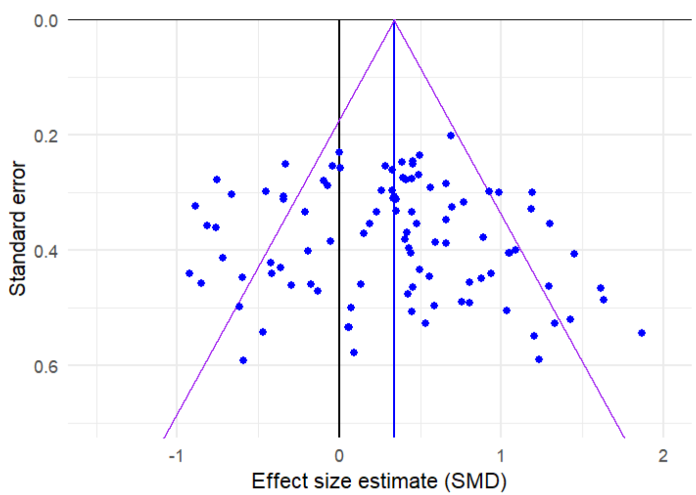
- A funnel-plot is a scatter plot of effect size estimates versus a measure of study precision (e.g., standard error).

Constant effect



- Effect size estimates will mostly fall within the funnel of $\hat{\mu} \pm 1.96SE$

Random effects



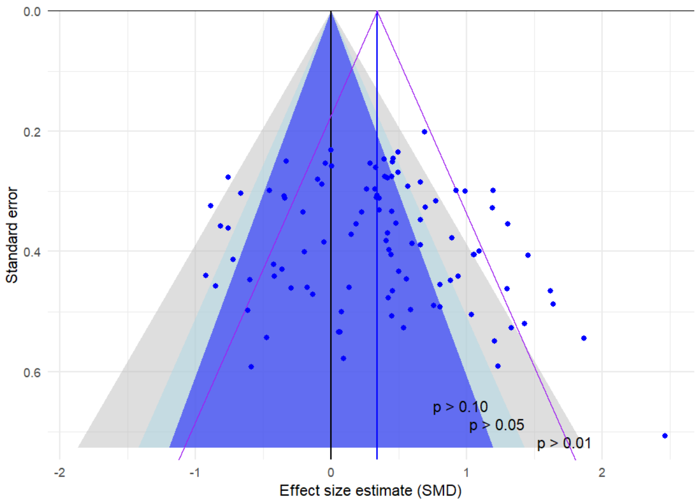
- Estimates outside the funnel indicate heterogeneity

Contour-enhanced funnel plots

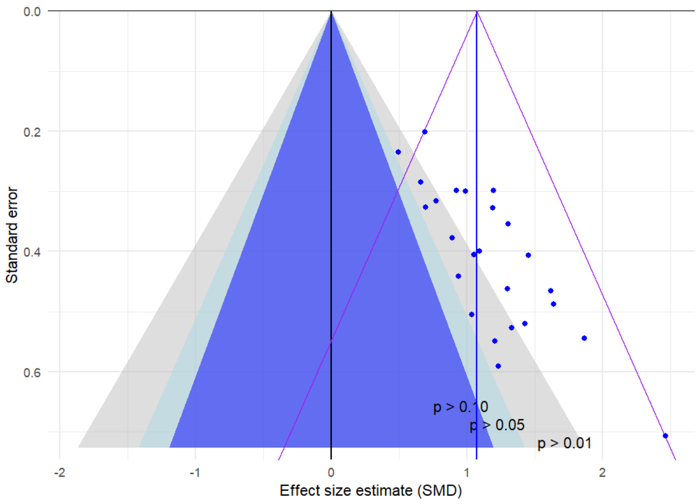
- Contour-enhanced funnel plots add shading to indicate regions where effect size estimates are statistically significant.

Selective reporting creates asymmetry

Non-selected data

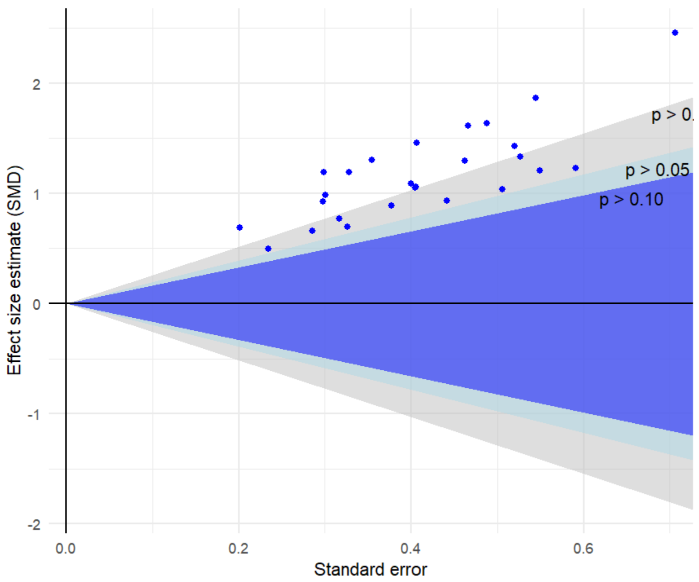


Affirmative effects only



Asymmetry tests/adjustments

- Egger's regression / PET / PEESE, rank correlation test
- Infer selective reporting from the presence of asymmetry.



- But asymmetry can have other causes!

Selection models

- Big literature
  - Iyengar & Greenhouse (1988)
  - Hedges & Vevea (1995)
  - Copas & Shi (2001)
- Infer selective reporting based on the ***shape of the effect size distribution***.
- Can accomodate moderators.
- But existing methods assume 1 effect size estimate per study.
  - Does not accomodate dependent effects.

Selective reporting in meta-analysis

Dependent effect sizes

A generalized excess significance test

Dependent effect size estimates

Multiple outcomes measured on a common set of participants

Treatment	O P Q
Control	O P Q

d_{O1}
d_{P1}
d_{Q1}

Outcomes measured at multiple follow-up times

Treatment	O_1	O_2	O_3
Control	O_1	O_2	O_3

d_{12}
d_{22}
d_{32}

Multiple treatment conditions compared to a common control

Treatment T	O
Treatment U	O
Control	O

d_{T3}
d_{U3}

Multiple correlations from a common sample

	a	b	c	d
a				
b				
c				
d				

r_{ab}
r_{ac}
r_{ad}

Dependent effect sizes are prevalent

- Tanner-Smith & Lipsey (2015). Brief alcohol interventions for adolescents and young adults: A systematic review and meta-analysis.
    - 185 studies, 1446 effect size estimates
    - 1-108 effect size estimates per study (median = 6, IQR = 3-12)
    - Multiple outcome measures, multiple follow-up times, multiple treatment conditions, multiple comparison groups
- Lehtonen et al. (2018). Is bilingualism associated with enhanced executive functioning in adults?
    - 152 studies, 891 effect size estimates
    - 1-40 effect size estimates per study (median = 4)
- Bediou et al. (2018). Meta-Analysis of Action Video Game Impact on Perceptual, Attentional, and Cognitive Skills.
    - 70 cross-sectional studies, 88 samples, 194 effect size estimates
    - 1-28 effect size estimates per study (median = 2)

Limited tools for investigating selective reporting with dependent effect sizes

- Ad hoc modifications to the data
    - Aggregate effect sizes to remove dependence
    - Conduct analysis within sub-groups
- Robust Egger's regression test (Rodgers and Pustejovsky, 2020):

$$T_{ij} = \beta_0 + \beta_1 (SE)_{ij} + \epsilon_{ij}$$

    - Meta-regression of effect size on a measure of precision (such as standard error).
    - Use robust variance estimation (clustering by sample) to account for effect size dependency.
    - Limited power except when there is very strong selective reporting.
    - Asymmetric funnel plots are suggestive but ambiguous.

Selective reporting in meta-analysis

Dependent effect sizes

A generalized excess significance test

An exploratory test of excess significance (TES)

- Ioannidis and Trikalinos (2007) proposed an intuitive diagnostic for selective reporting based on **statistical significance at level** α.
    - k: Total number of effect sizes (assuming one ES per sample)
    - O: observed number of statistically significant effect sizes
    - P_j: Estimated power of study j, assuming a common effect model or random effects model.
    - $E = \sum_{j=1}^{k} P_j$: expected number of statistically significant effect sizes
- A binomial approximation for O in the absence of selective reporting:

$$O \dot{\sim} Binom(k, E/k) \qquad \text{or} \qquad \frac{O - E}{\sqrt{E(k-E)/k}} \dot{\sim} N(0,1)$$

- Excess of statistically significant effect sizes indicates selective reporting.

Problems with TES

- Binomial approximation isn't correct (because P_j are usually heterogeneous).
- Does not account for uncertainty in power estimates.
- Requires independent effect sizes.
- Many different, somewhat arbitrary ways of estimating power.
  - Creates analytic flexibility in how TES is applied.

Goal: Generalize TES

- Account for uncertainty in power estimates
- Allow for dependent effect sizes
- Allow for systematic predictors / covariates
- Proper null distribution

A meta-regression model

$$\mathbf{T}_j = \mathbf{X}_j\boldsymbol{\beta} + \mathbf{u}_j + \mathbf{e}_j$$

- $\mathbf{T}_j$: set of effect size estimates for sample j
- $\mathbf{X}_j$: covariate matrix for sample j
- $\boldsymbol{\beta}$: Meta-regression coefficients
- $\boldsymbol{\theta}$: parameters describing random effects $\mathbf{u}_j$.
- $\mathbf{W}_j$: Weighting matrix for estimating meta-regression

Estimation

- $\boldsymbol{\theta}$ estimated by full/restricted maximum likelihood estimation or method of moments.
- $\boldsymbol{\beta}$ estimated by weighted least squares.

TES as estimating equations

- Meta-regression estimating equations:

$$\mathbf{S}_{\boldsymbol{\beta}} = \sum_{j=1}^{k} \mathbf{X}_j' \mathbf{W}_j \left(\mathbf{T}_j - \mathbf{X}_j \boldsymbol{\beta} \right)$$

$$\mathbf{S}_{\boldsymbol{\theta}} = \frac{\partial l_R(\boldsymbol{\beta}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

- An additional estimating equation:

$$S_\pi = \sum_{j=1}^{k} \left[O_j - E_j(\boldsymbol{\beta}, \boldsymbol{\theta}) \right]$$

where

  - O_j: number of statistically significant effect sizes from study j
  - E_j: expected number of statistically significant effect sizes, given the model parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$

- In the absence of publication bias, $\mathbb{E}\left(S_\pi\right) = 0$.

Generalized excess significance test

- A cluster-robust score test statistic (Rotnizky & Jewell, 1990):

$$Z^{GEST} = \frac{\hat{S}_\pi}{\sqrt{V^{CR}}}$$

where V^{CR} is a cluster-robust estimate of $\mathbb{V}\text{ar}(S_\pi)$, accounting for estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$.

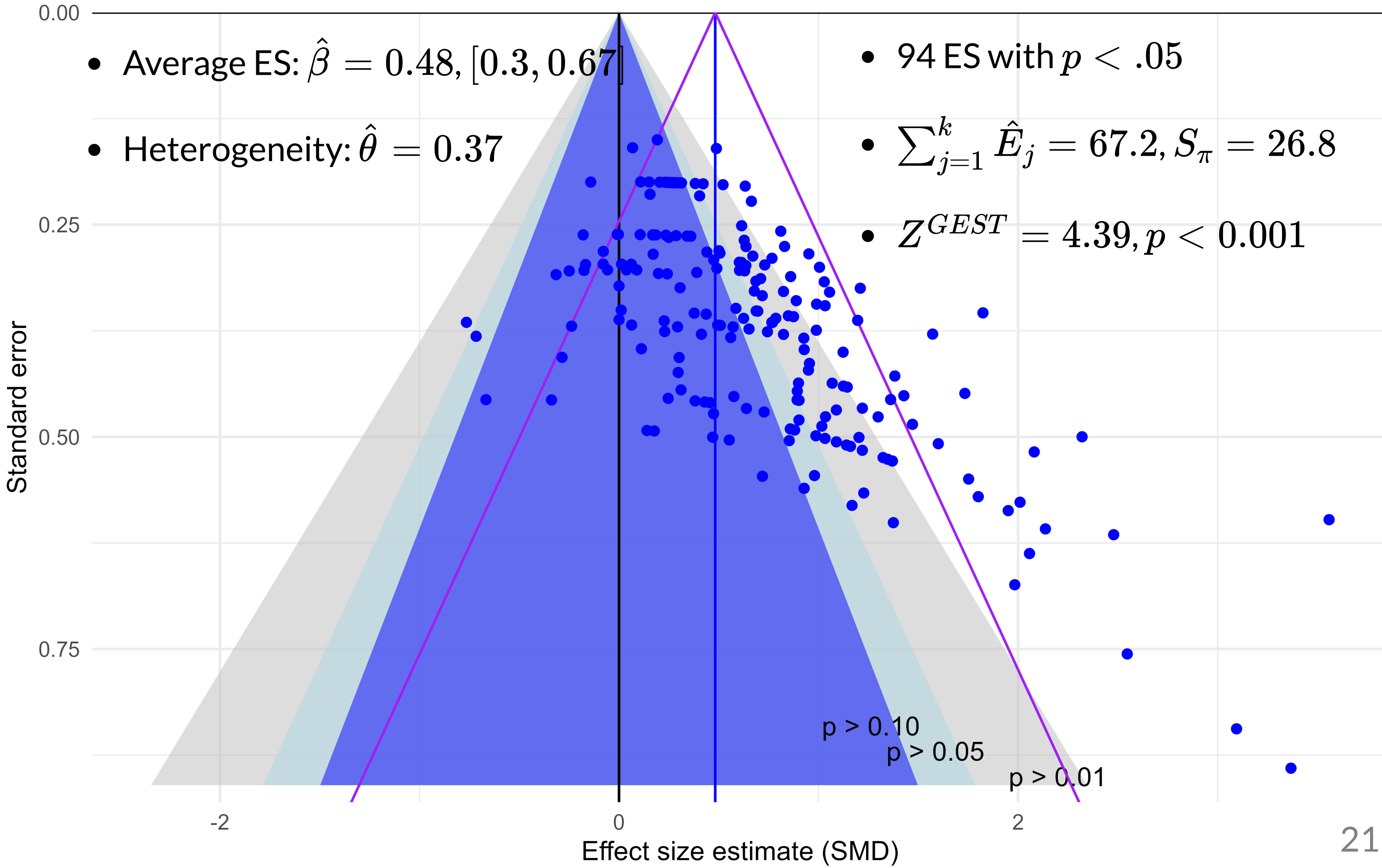
- Large-sample approximation (for large-enough k):

$$Z^{GEST} \dot{\sim} N(0,1)$$

in the absence of selective reporting.

- Selective reporting indicated if $Z^{GEST} > \Phi^{-1}(1-\alpha)$.

Bediou et al. (2018). Meta-Analysis of Action Video Game Impact on Perceptual, Attentional, and Cognitive Skills.

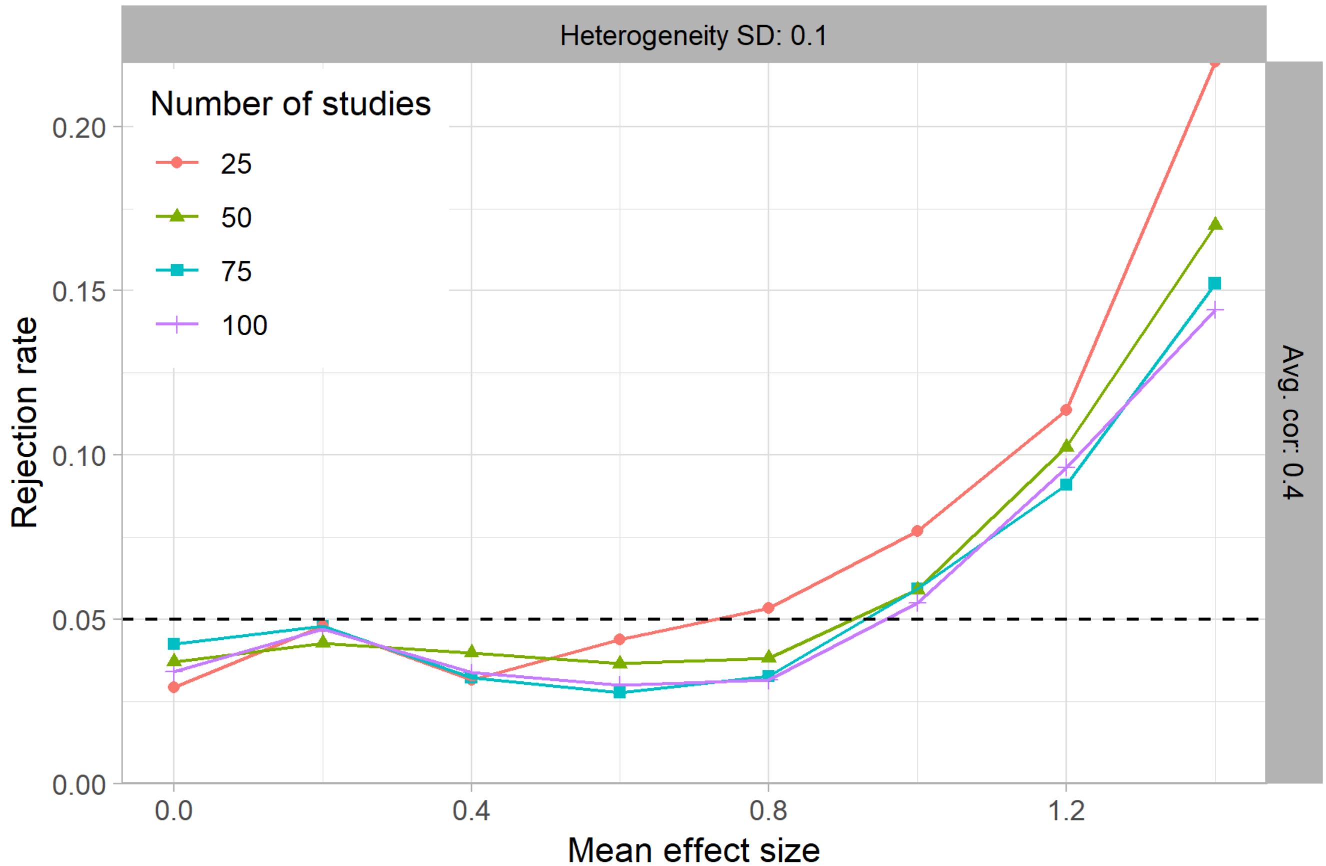


- Average ES: $\hat{\beta} = 0.48, [0.3, 0.67]$
- Heterogeneity: $\hat{\theta} = 0.37$

- 94 ES with $p < .05$
- $\sum_{j=1}^{k} \hat{E}_j = 67.2, S_\pi = 26.8$
- $Z^{GEST} = 4.39, p < 0.001$

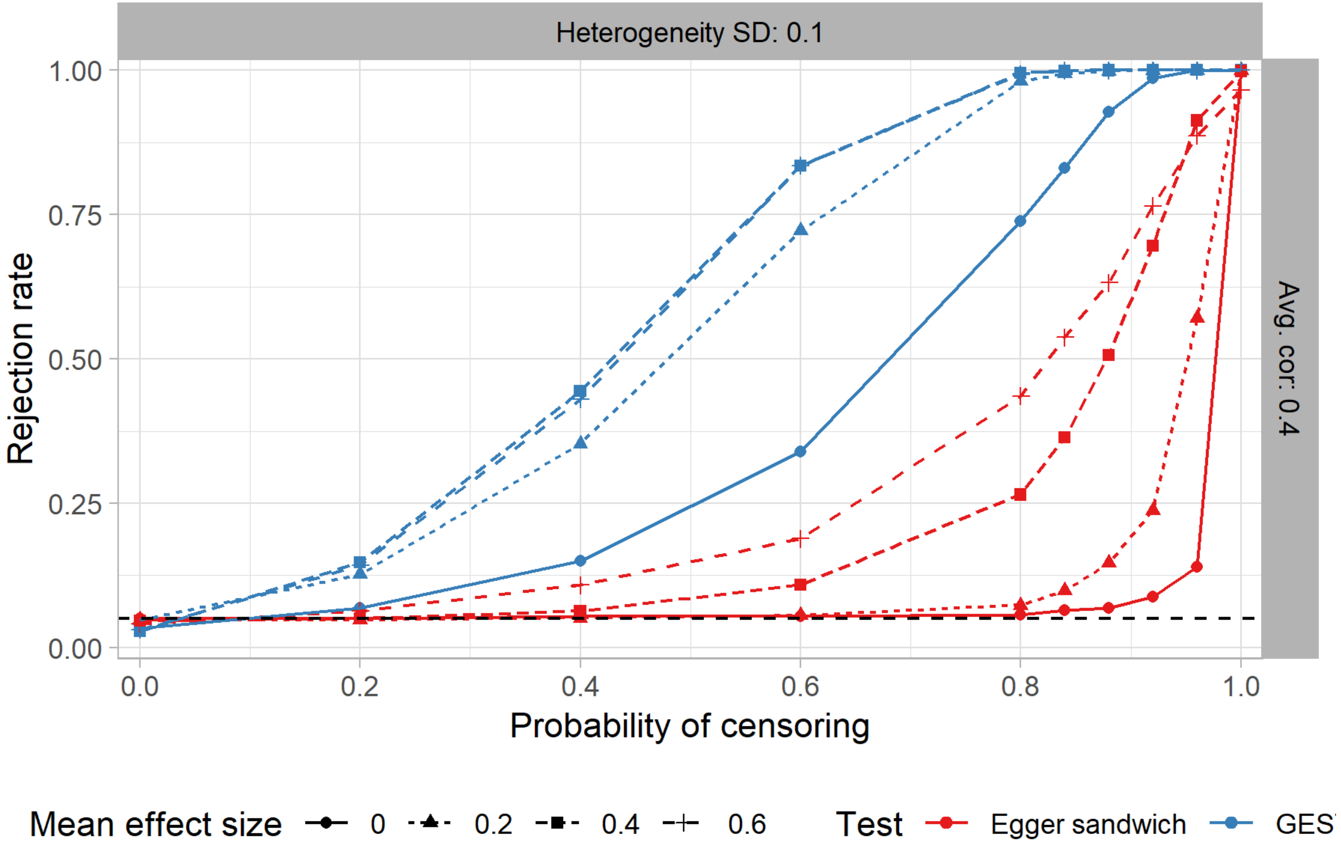
Simulations: Type I error rates

(Correlated standardized mean differences)



Simulations: Power comparison

$k = 50$

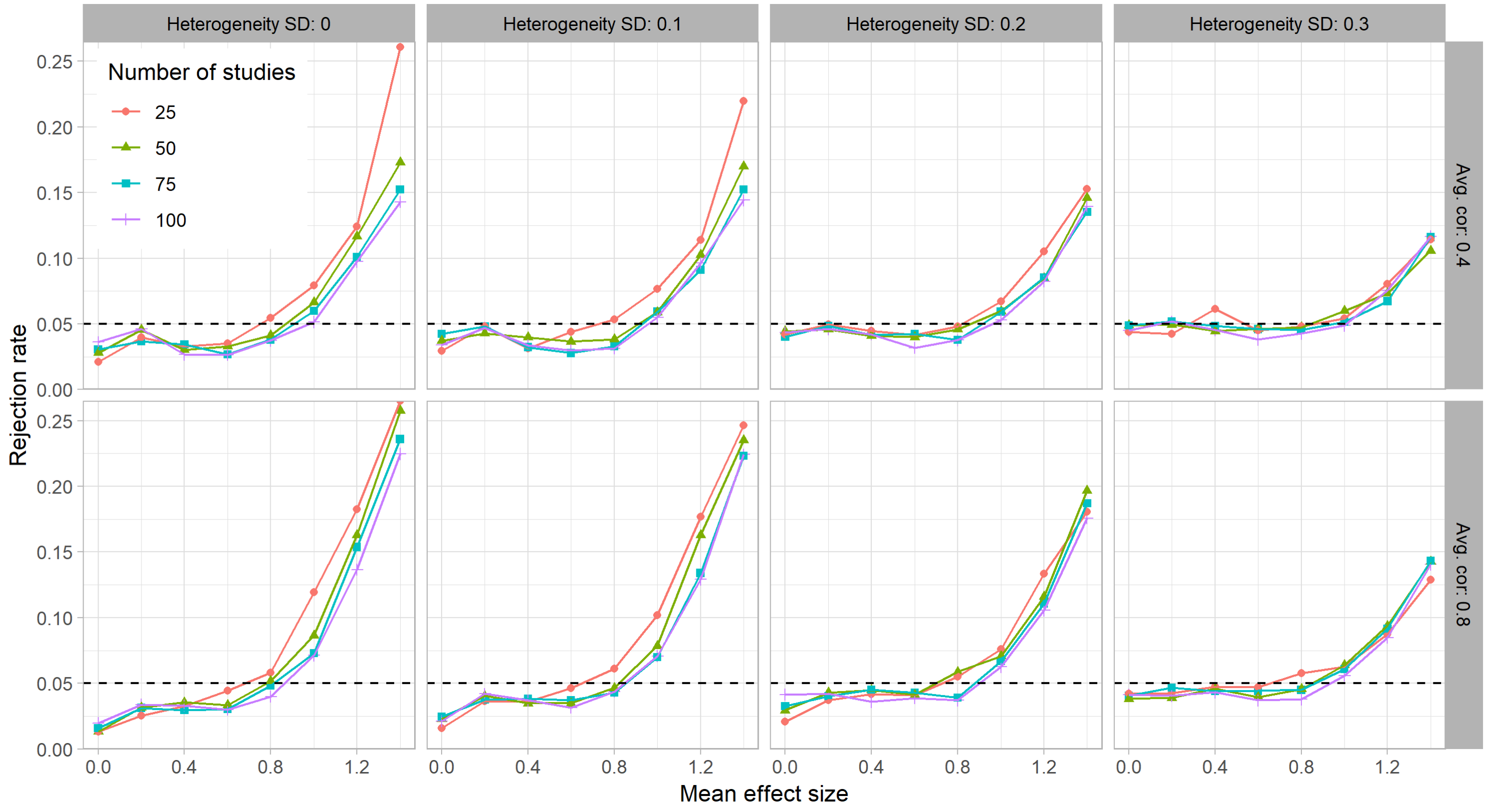


Discussion

- GEST requires consistent estimation of mean and variance of the effect size distribution *in the absence of selection*.
  - Can accommodate meta-regression models.
  - Can use weighting schemes that are not inverse-variance.
- Type I error rates are inflated when average effects are large and homogeneous.
  - Small sample refinements still under investigation (cluster wild bootstrap?).
- GEST estimates expected power *marginally* for each effect size.
  - Does not consider the joint pattern of statistical significance.
- Outstanding need for models that
  - capture both selective outcome reporting and study-level selection.
  - accommodate pre-registered studies, known to be fully reported.
  - *estimate* strength of selection rather than using an assumption.

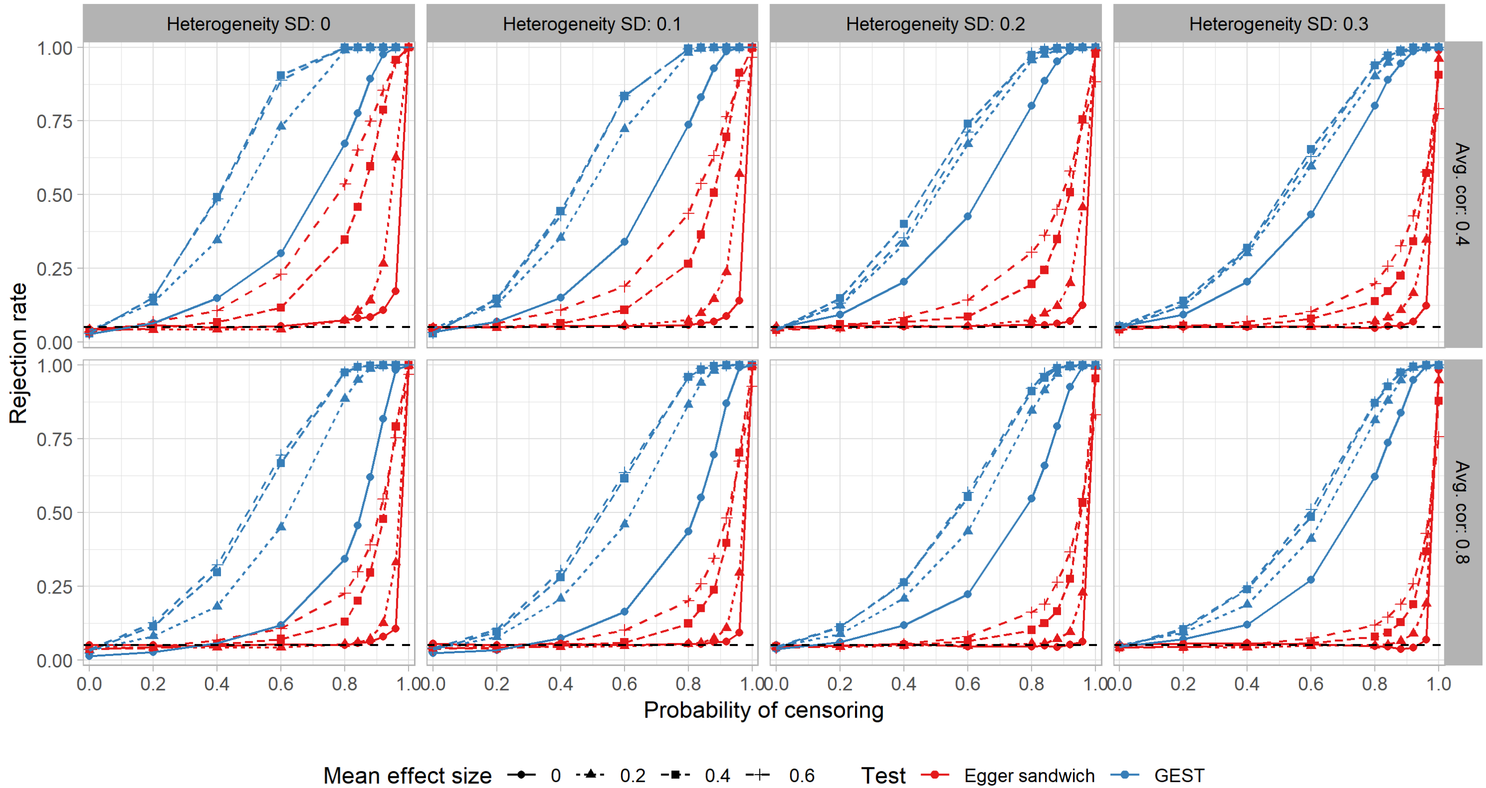
Simulations: Type I error rates

(Correlated standardized mean differences)



Simulations: Power comparison

$k = 50$



Simulations: Power comparison

$k = 100$