DRK-12 RESEARCH METHODS WEBINAR SERIES | SEPTEMBER 2020

AN INTRODUCTION TO META-ANALYTIC TECHNIQUES

Jeffrey Valentine, PhD| Emily Tanner-Smith, PhD

MAKING RESEARCH RELEVANT

AIR®

Welcome!

Take a moment to introduce yourself in the chat box.

Please tell us your name, organization, and affiliation with the DRK-12 program (e.g., principal investigator [PI], project team member, evaluator, or aspiring PI).

DRK-12 Research Methods Webinar Series

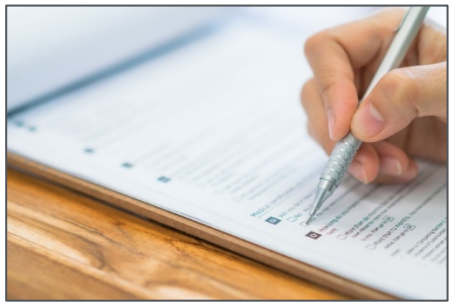
Melissa Rasberry, EdD
Principal Technical Assistance Consultant

Learning Outcomes

Following this session, participants will be able to

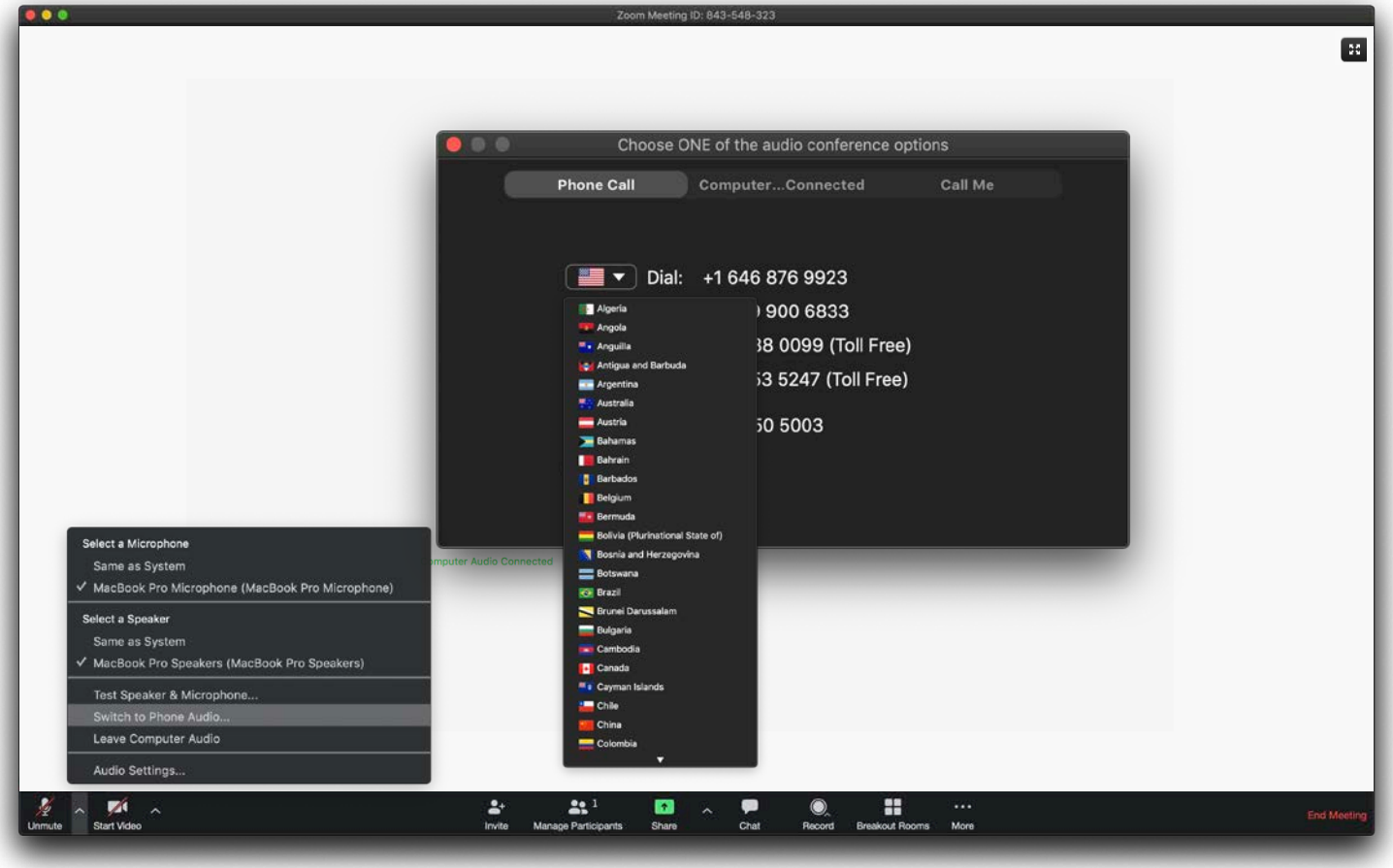
- Understand meta-analysis terminology
- Identify the importance and benefits of meta-analysis
- Understand key considerations when synthesizing evidence using meta-analysis
- Consider ways meta-analytic techniques could further new learning in STEM education

Today's Webinar

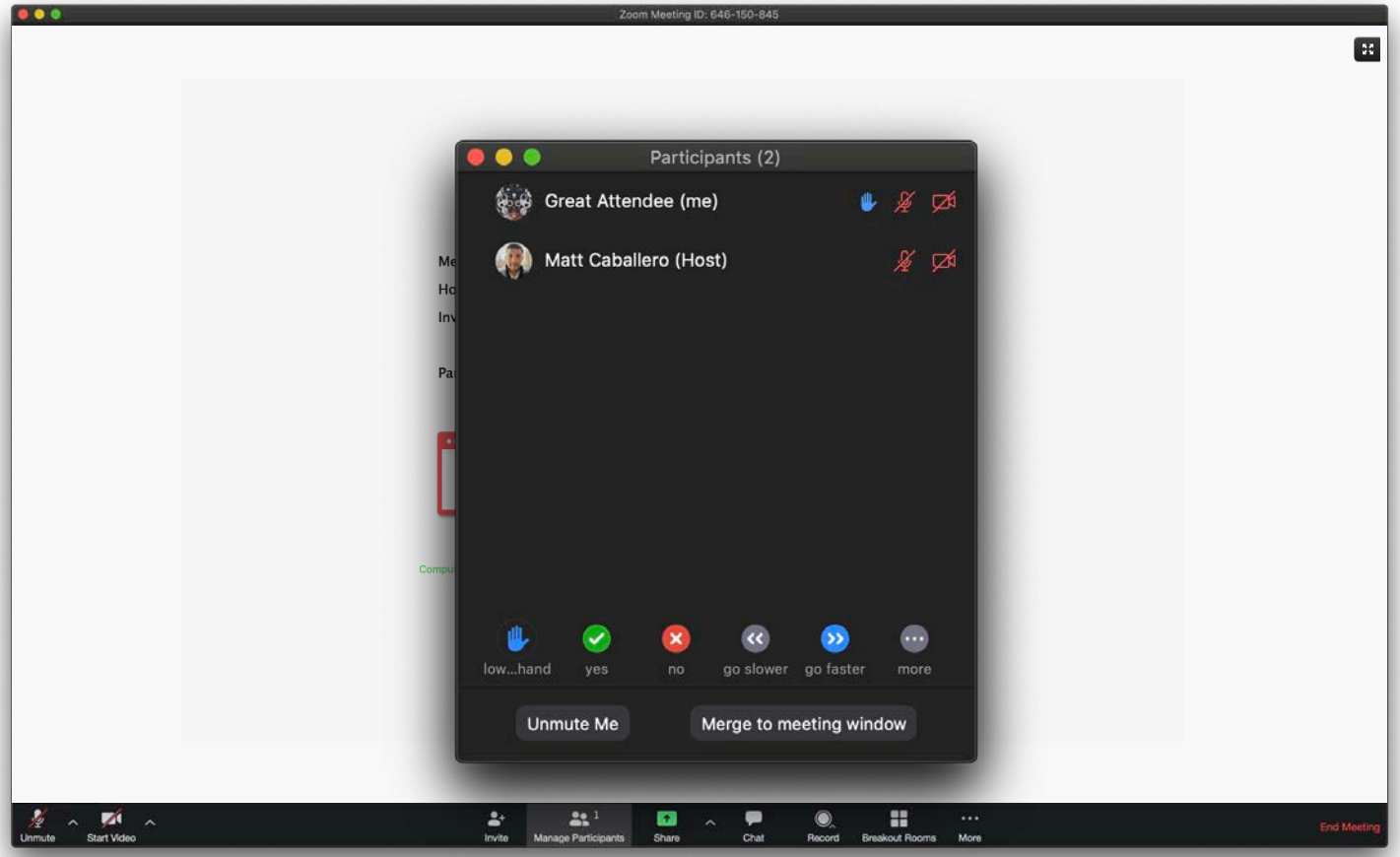


http://cadrek12.org/

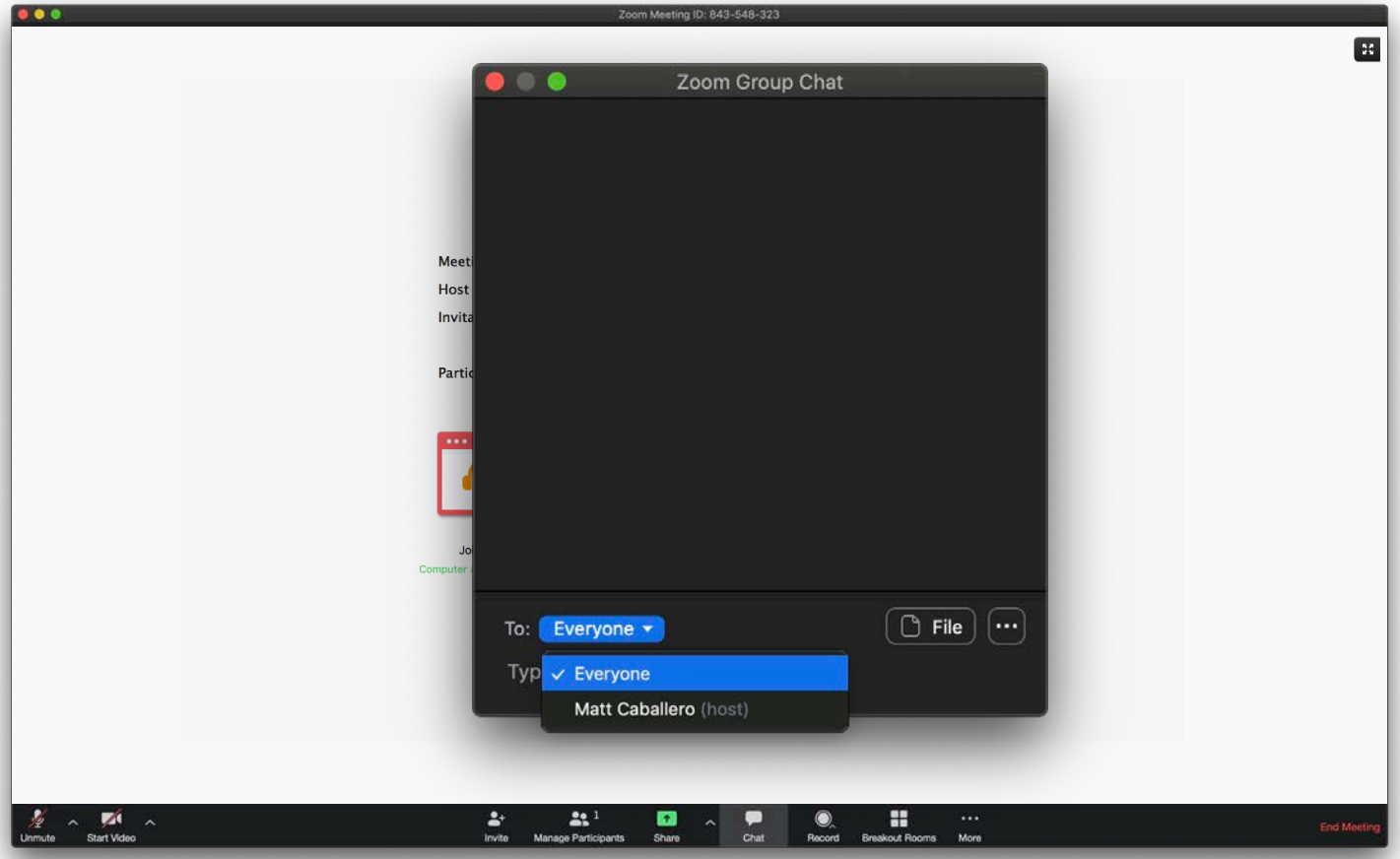
How to Use Zoom



How to Use Zoom

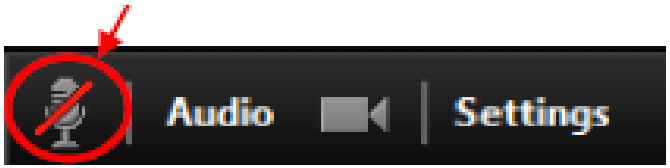


How to Use Zoom



How to Use Zoom

Mute your mic. This helps to minimize audio feedback. Mute your audio by clicking on the microphone icon located in the lower left-hand corner of the menu bar.



Audio is muted

Use chat. Connect with participants via private chat or with a comment to everyone.

Ask questions. If you have a technical question, leave your message in the chat.

Meet the Presenters

Emily Tanner-Smith, PhD

University of Oregon
etanners@uoregon.edu

Jeffrey Valentine, PhD

University of Louisville
jeff.valentine@louisville.edu

Overview of Today's Webinar

1. Meta-analysis methods for quantitative synthesis

2. Methods for assessing publication and small study bias

Meta-Analysis Methods for Quantitative Synthesis

Emily Tanner-Smith, PhD

Definition of Meta-Analysis

Systematic review

Summary of the research literature that uses **explicit**, **reproducible** methods to **identify, extract** information from, and **analyze** relevant studies.

Meta-analysis

A meta-analysis involves **statistically combining** the results of studies.

What Types of Questions Can Be Addressed in a Meta-Analysis?

Intervention effectiveness

What are the effects of *x* intervention on *y* outcome for *z* population?

Group differences

How does group *z1* differ from group *z2* on some characteristic *y*?

Associations

How does *x1* relate to *x2* in population *z*? (direction and strength of correlation)

Change over time

How does some behavior/attitude change from time 1 to time 2?

Diagnostic test accuracy

Which test (A vs. B) has better sensitivity/ specificity in diagnosing or predicting *y*?

Prevalence

What is the prevalence of some condition in population *z*?

Examples From STEM Education Meta-Analyses

How does computer-based scaffolding for STEM education affect students' cognitive outcomes? Do outcomes vary by learner characteristics (grade level, baseline achievement)?[1]

How effective is computer-supported collaborative learning in STEM education? Does it vary according to implementation context (mode of collaboration, type of technology, pedagogical approach)?[2]

How do the effects of STEM professional development programs on student outcomes vary by study methodology (design, outcome measurement, statistical adjustments)?[3]

[1] Belland et al. (2017), doi:10.3102/0034654316670999; [2] Jeong et al. (2019), doi:10.1016/j.edurev.2019/100284; [3] Lynch et al. (2019), doi:10.3102/0162373719849044)

Effect Sizes: The Building Blocks of Meta-Analysis

Effect sizes are the unit of analysis in a meta-analysis.

Effect sizes represent the magnitude and direction of a quantity of interest, independent of sample size.

Because different primary studies use different scales to measure the same construct, standardized effect sizes are often used to ensure results are on a common scale.

Effect sizes (and their standard errors) are typically collected during the data extraction/coding phase of a systematic review.

Commonly Used Effect Size Indexes

Measures of Central Tendency, Event Counts

- Mean
- Proportion
- Incidence rate

Associations Between Variables

- Pearson correlation coefficient
- Phi coefficient
- Point-biserial correlation
- Biserial correlation

Change Over Time

- Mean change score (unstandardized or standardized)

Group Contrasts

- Mean difference (unstandardized or standardized)
- Ratio of means
- Odds ratio
- Risk ratio, risk difference
- Incidence rate ratio
- Incidence rate difference

The Standardized Mean Difference Effect Size (*d*)

Compares the means of two groups, standardizes the difference.

Hedges' adjustment (*g*) used to correct for small sample bias

$$d = \frac{\bar{x}_{G1} - \bar{x}_{G2}}{\sqrt{\frac{s_{G1}^2(n_{G1} - 1) + s_{G2}^2(n_{G2} - 1)}{n_{G1} + n_{G2} - 2}}}$$

$$g = d * \left[1 - \frac{3}{4df\ - 1}\right]$$

$$SE_g = \sqrt{\frac{n_{G1} + n_{G2}}{n_{G1} * n_{G2}} + \frac{d^2}{2(n_{G1} + n_{G2})} * \left[1 - \frac{3}{4df\ - 1}\right]}$$

Common Steps in a Meta-Analysis

Descriptive Statistics	Central Tendency and Dispersion	Predicting Variance	Assessing Robustness
Summarizing and describing included studies, including quality/risk of bias	Examining the distribution of effect sizes • Mean effect size (central tendency) • Assessing heterogeneity (dispersion)	Explaining heterogeneity in effects across values of moderators (meta-regression)	Exploring robustness of findings • Publication/ small study bias analyses • Sensitivity analyses

The Role of Weights in Meta-Analysis

Meta-analysis techniques use weighting to account for differences in precision across studies.

Larger sample sizes are assumed to provide more precise parameter estimates, so they carry more weight

Typically use inverse variance weights
(smaller variances -> more precise estimates -> more weight)

Mean Effect Size Estimation

Weighted mean effect size computation is straightforward:

$$\bar{\theta}_w = \frac{\sum_{i=1}^{k}(w_i * y_i)}{\sum_{i=1}^{k}(y_i)}$$

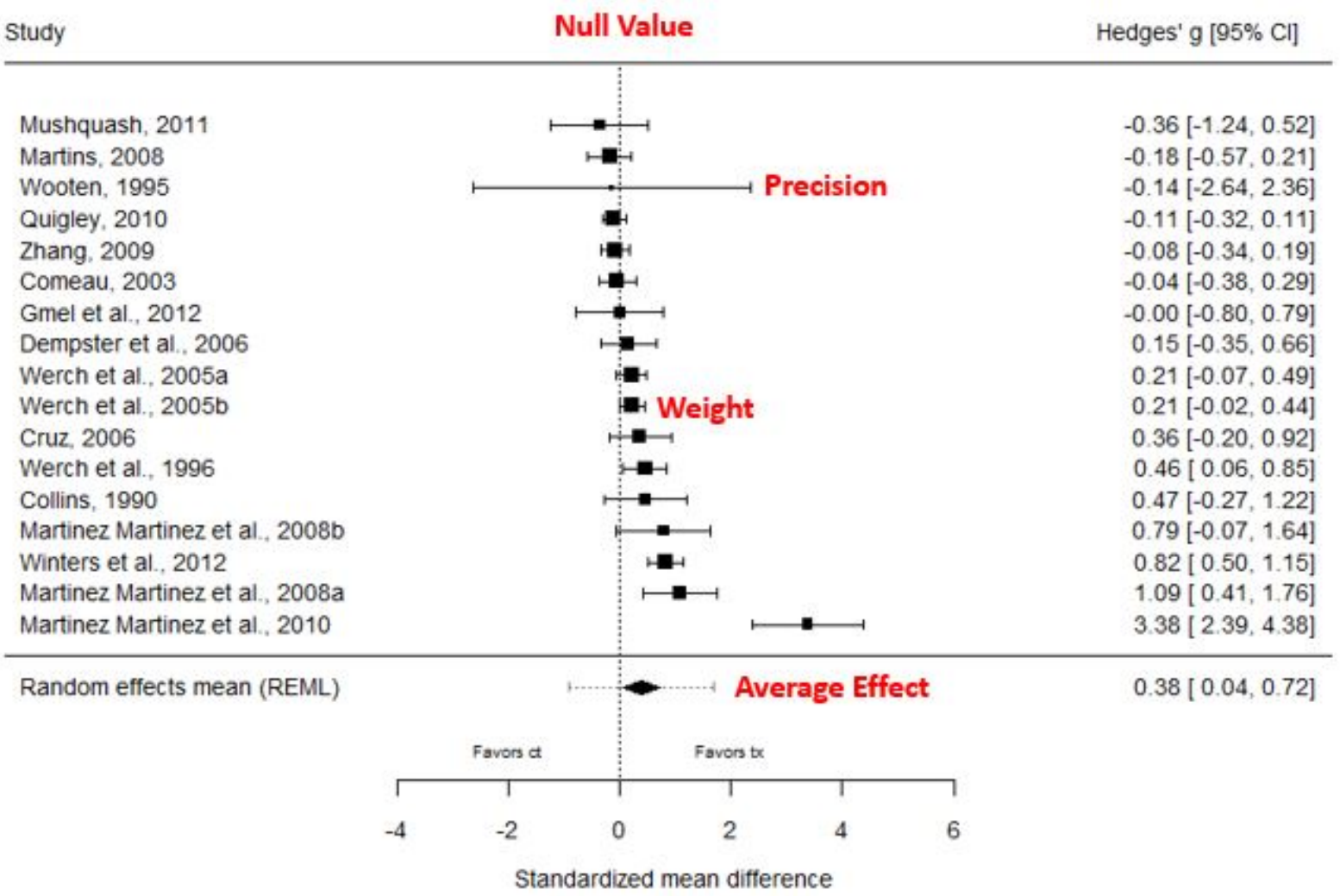
$$se_{\bar{\theta}_w} = \sqrt{\frac{1}{\sum_{i=1}^{k}(w_i)}}$$

$$95\%\ \text{CI} = \bar{\theta}_w \pm z_{crit} * se_{\bar{\theta}_w}$$

$$Z = \frac{\bar{\theta}_w}{se_{\bar{\theta}_w}}$$

where $\bar{\theta}_w$ is the weighted mean effect size that pools observed effect size estimates y_i from *k* included studies using inverse variance weights w_i.

Forest Plot Visualization of a Simple Meta-Analysis



Source: Data from Hennessy & Tanner-Smith (2015), doi: 10.1007/s11121-014-0512-0

Fixed versus Random Effects Models (and Weights)

Fixed Effect

Assumes a common effect size in the population (θ)

Goal is to estimate common effect size θ

Assumes observed variation is solely due to sampling error

Inverse variance weight

$$w_i = \frac{1}{v_i}$$

Random Effects

Assumes a distribution of multiple effect sizes in the population (θs)

Goal is to estimate mean of distribution of θs: μ

Assumes observed variation is due to sampling error and variance of distribution of θs

Inverse variance weight

$$w_i = \frac{1}{v_i + \tau^2}$$

Assessing Heterogeneity

Heterogeneity refers to variation in the true population effect sizes (θs).

- Observed variability in effect size estimates includes both true heterogeneity and random error.

Assessing Heterogeneity

Heterogeneity refers to variation in the true population effect sizes.

- Observed variability in effect size estimates includes both true heterogeneity and random error.

There are numerous graphical and statistical tools available for assessing heterogeneity.

- Is there graphical evidence of variability? (e.g., forest plot, Galbraith plot)
- Is there statistical evidence of variability? (Q)
- What is the variance/standard deviation of the distribution of the true effects? (τ^2, τ)
- What proportion of observed variation is attributable to true heterogeneity? (I^2)
- What is the dispersion of true effects around μ? (prediction interval)

Explaining Heterogeneity

In primary research, we use t-tests, ANOVA, and regression to compare means for two or more groups; similar logic is used in meta-analysis, but now the focus is on now studies rather than participants.

Meta-regression is the most flexible tool. It permits inclusion of multiple categorical and/or continuous predictors:

$$\theta_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + u_i + \varepsilon_i$$
$$u_i \sim N(0, \tau^2)$$
$$\varepsilon_i \sim N(0, v_i)$$

where true effects in the population θ_i are modeled as a function of *p* study-level predictors and τ^2 reflects the amount of residual heterogeneity in true effects not accounted for by the predictors in the model.

Statistical Dependencies in Meta-Analyses

Traditional meta-analysis models assume effect size estimates are statistically independent, but dependencies are common.

Reductionist approaches can be used so that one effect size per study is included in any given meta-analytic model.

Integrative approaches can be used to so that all (dependent) effect sizes in a study can be included in the meta-analytic model.

Recommended Resources

Borenstein, M. (2019). *Common mistakes in meta-analysis and how to avoid them*. Englewood, NJ: Biostat, Incorporated.

Centre for Reviews and Dissemination. (2009). *Systematic reviews*: *CRD's guidance for undertaking reviews in healthcare*. University of York: Centre for Reviews and Dissemination. Retrieved from https://www.york.ac.uk/crd/guidance/

Moeyaert, M., Ugille, M., Natasha Beretvas, S., Ferron, J., Bunuan, R., & Van den Noortgate, W. (2017). Methods for dealing with multiple outcomes in meta-analysis: A comparison between averaging effect sizes, robust variance estimation and multilevel meta-analysis. *International Journal of Social Research Methodology*, *20*(6), 559-572.

Schwarzer, G., Carpenter, J. R., & Rücker, G. (2015). *Meta-analysis with R* (Vol. 4784). Springer.

Wilson, D. B. (n.d.). Practical meta-analysis effect size calculator. Retrieved from https://campbellcollaboration.org/escalc/html/EffectSizeCalculator-Home.php

Q&A

Publication Bias

Jeffrey Valentine, PhD

Reporting Biases

Publication bias occurs when the decision about whether to publish a study depends on the study's results.

Outcome reporting bias occurs when the decision about whether to include an outcome in a study depends on the outcome's results.

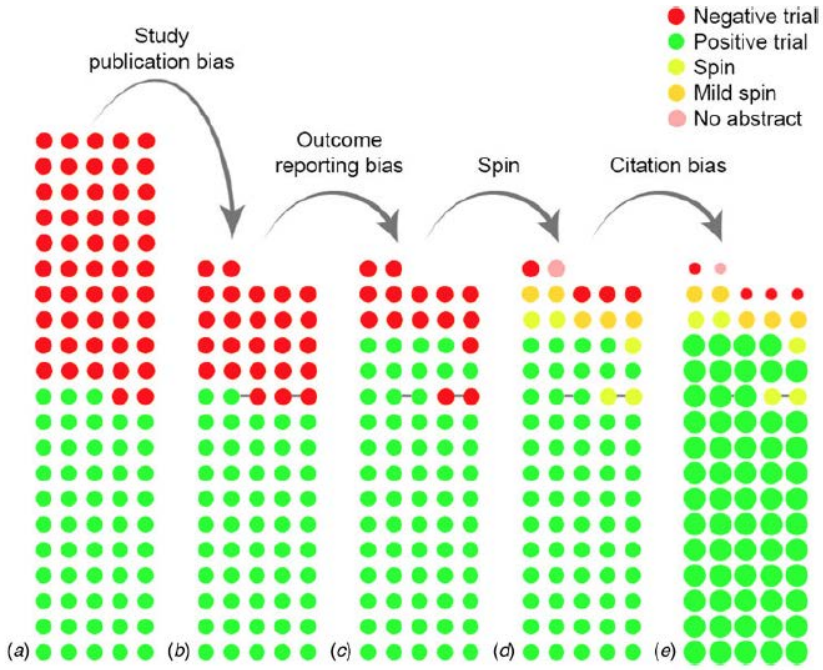


Figure from de Vries, Y. A., Roest, A. M., de Jonge, P., Cuijpers, P., Munafò, M. R., & Bastiaansen, J. A. (2018). The cumulative effect of reporting and citation biases on the apparent efficacy of treatments: The case of depression. *Psychological Medicine 48*, 2453–2455. https:// doi.org/10.1017/S0033291718001873

Reporting Bias

Studies with statistically significant results are

about 2x

more likely to be published.

Researchers have investigated what happens to studies that were initiated. Editor and peer reviewer preferences appear to have a role, but “we” seem to be the primary culprits.

Relationship Between Sample Size and Effect Size

All publication bias methods interpret the relationship between study size and effect size as evidence of publication bias.

Publication bias is sometimes referred to as "small study effects."

In theory, sample size (variance) and effect size (mean difference) should be independent.

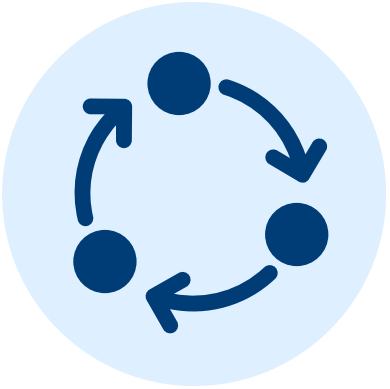
In reality, this is often not so.

A negative relationship between study size and effect size (e.g., smaller studies have larger effects) might be evidence of publication bias.

Methods for Detecting and Addressing Bias

A vigorous literature search is the best defense

Graphical approaches to detecting publication bias



Statistical approaches to detecting and possibly assessing publication bias

The Importance of the Literature Search

Bottom line message:

There are no “very good” methods for addressing publication bias. Prevention is our best defense.

Graphical and Statistical Approaches to Bias

The general approach to assessing publication bias involves answering three questions:

Graphical and Statistical Approaches to Bias

The general approach to assessing publication bias involves answering three questions:

- Is there any evidence of publication bias?

Graphical and Statistical Approaches to Bias

The general approach to assessing publication bias involves answering three questions:

- Is there any evidence of publication bias?
- It is possible that the entire effect is an artifact of publication bias?

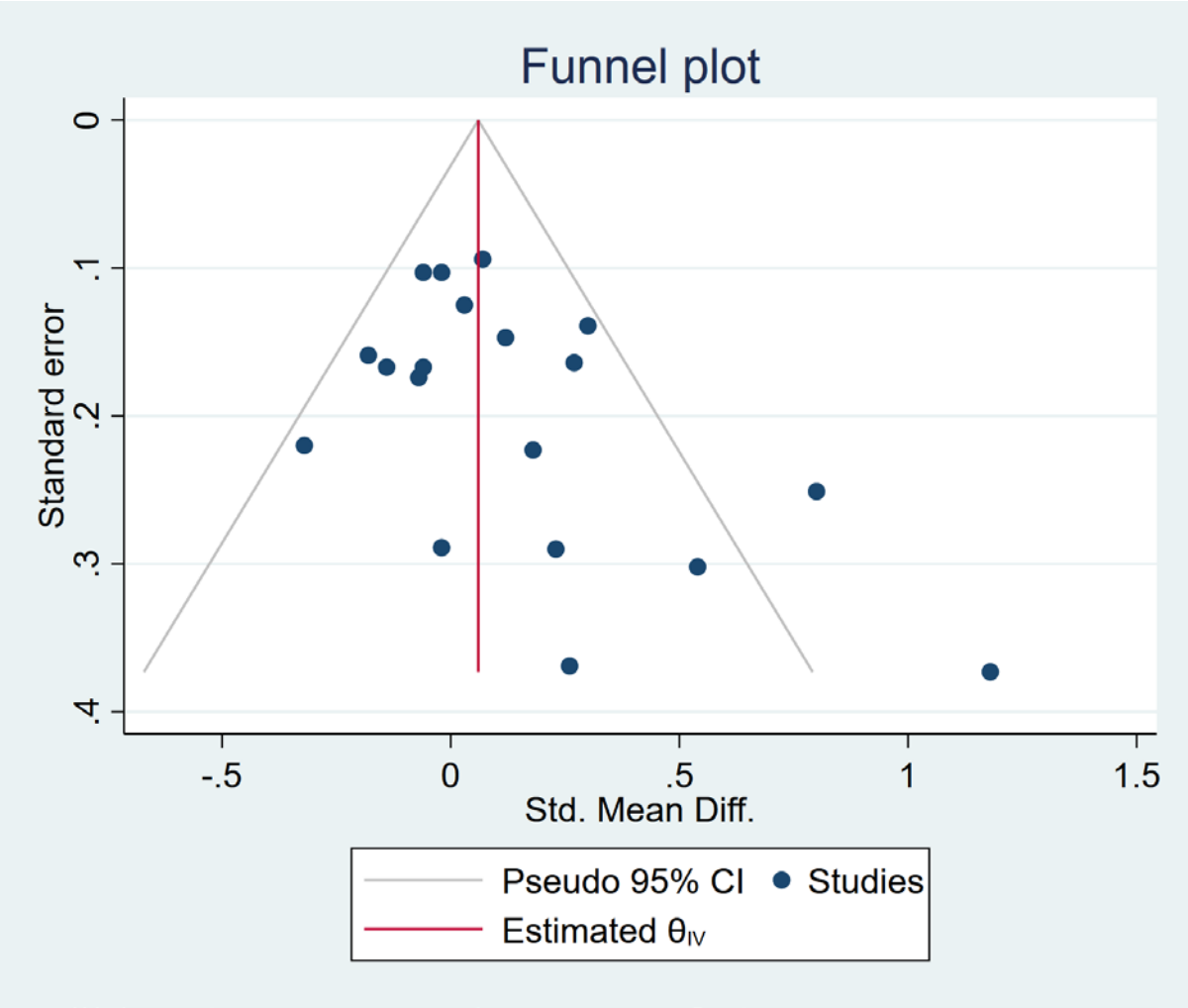
Graphical and Statistical Approaches to Bias

The general approach to assessing publication bias involves answering three questions:

- Is there any evidence of publication bias?
- It is possible that the entire effect is an artifact of publication bias?
- How much of an impact might publication bias have?

Graphical Approaches to Assessing Publication Bias

The most popular approach is the **funnel plot.** It is an informal method for detecting publication bias.



Funnel Plots

Funnel plots are examined for "gaps," which suggest missing studies.

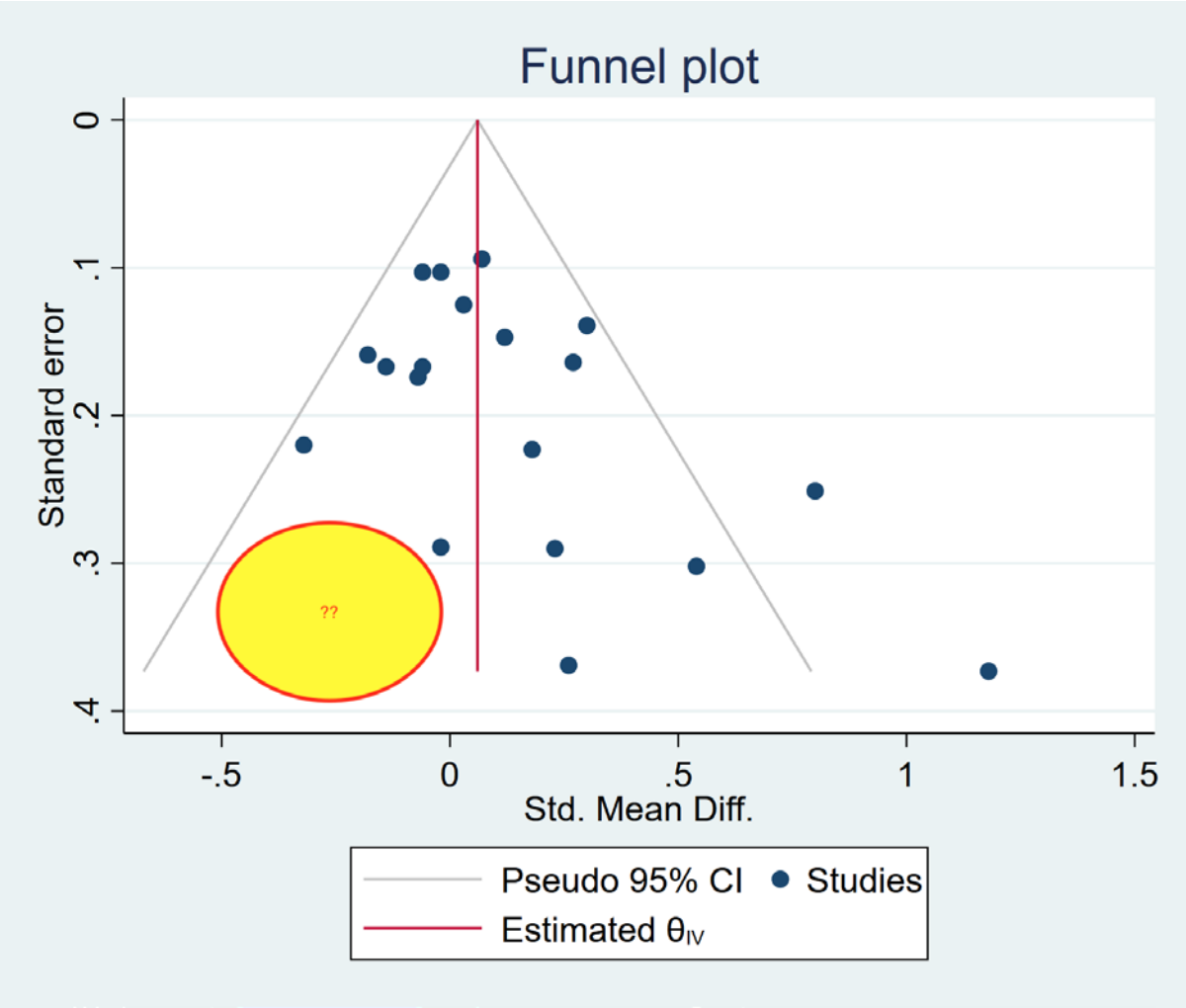


Figure source: https://www.stata.com/new-in-stata/meta-analysis/i/meta_funnel.png

Funnel Plots

Funnel plots are **deceptively simple**, often **highly ambiguous,** and let researchers see **what they want to see**.

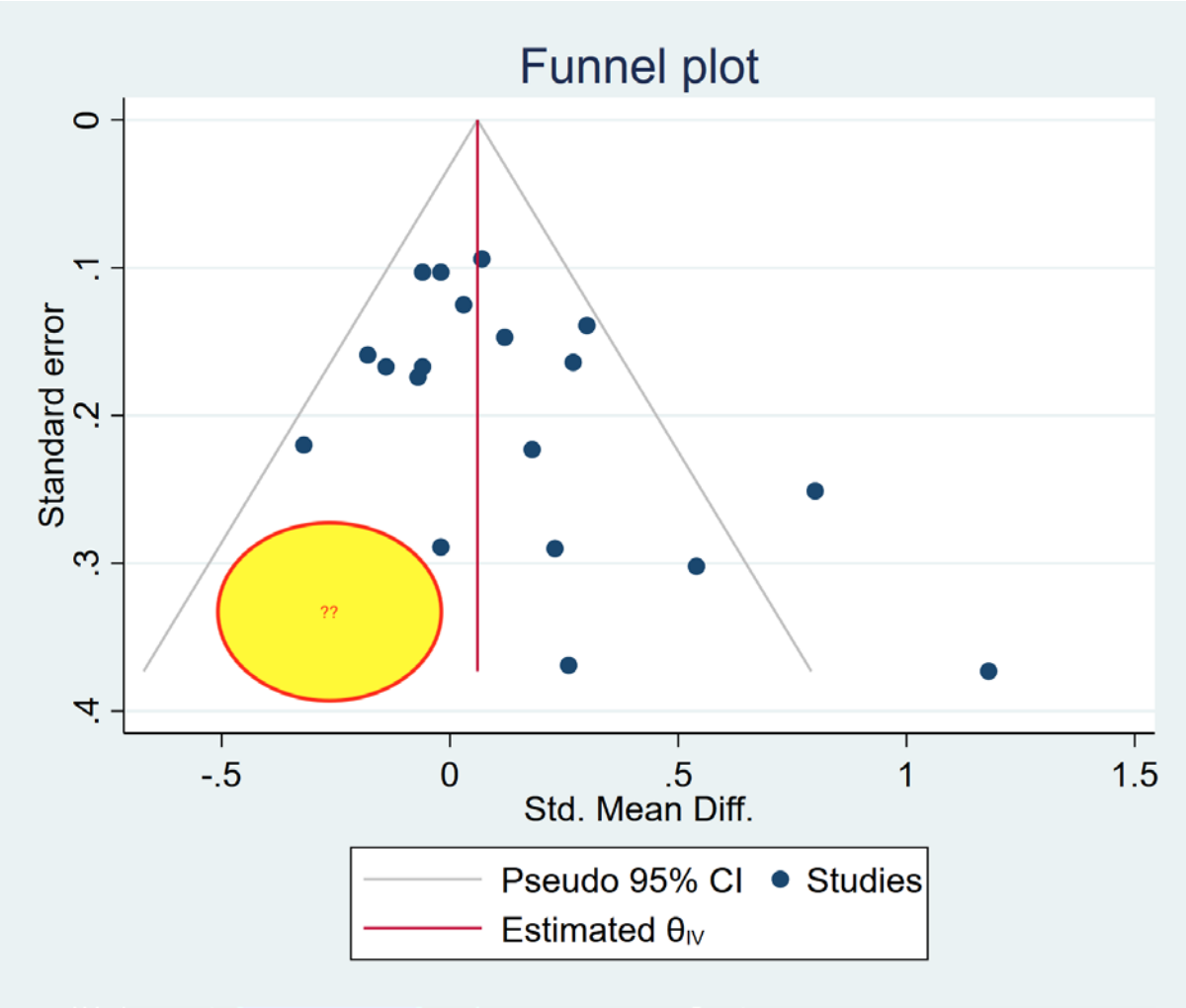


Figure source: https://www.stata.com/new-in-stata/meta-analysis/i/meta_funnel.png

Funnel Plots

Funnel plots only address **one of the three** questions we want to ask.

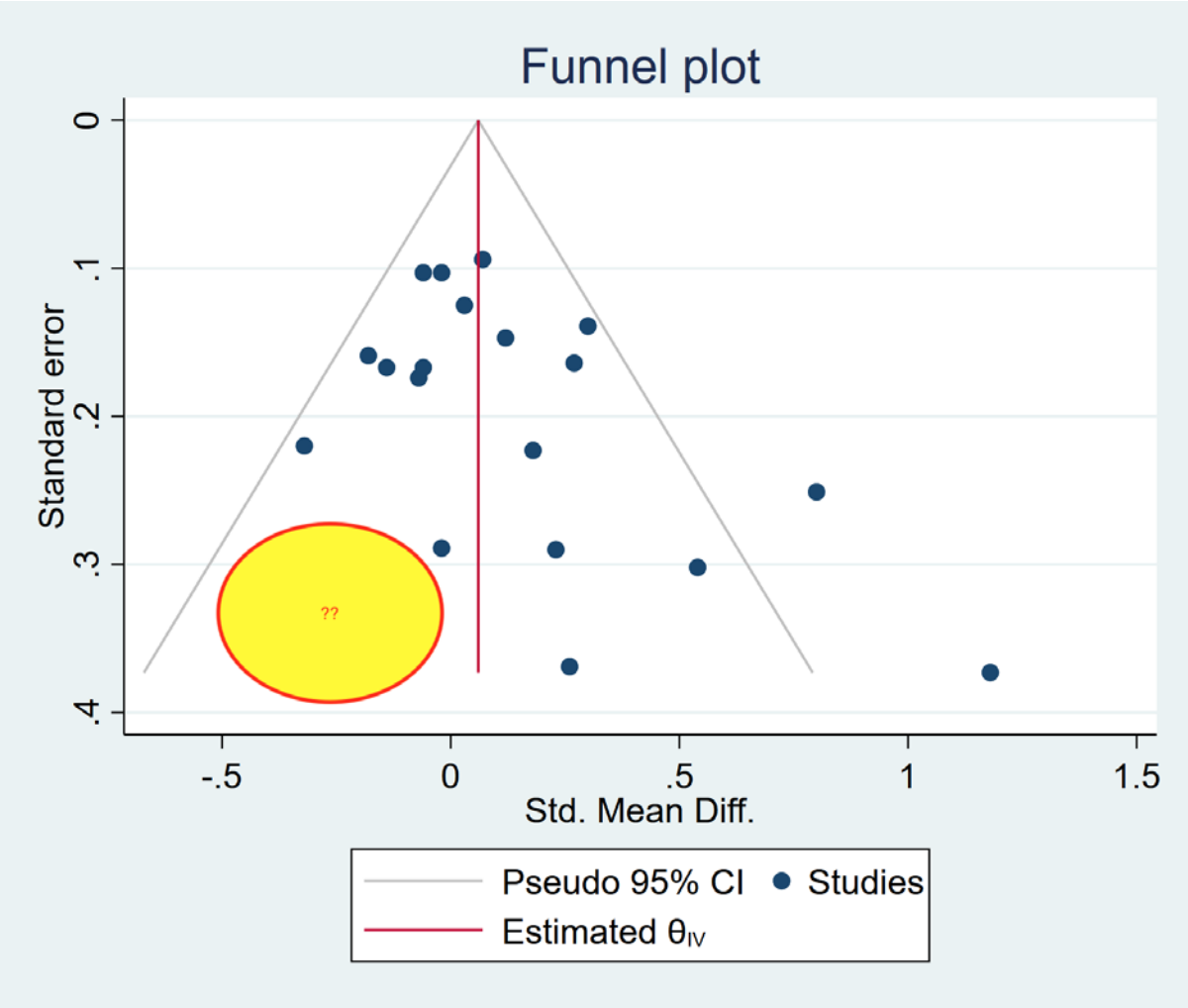


Figure source: https://www.stata.com/new-in-stata/meta-analysis/i/meta_funnel.png

Funnel Plots

Funnel plots only address **one of the three** questions we want to ask.

Therefore, it is best to **use funnel plots as an adjunct to statistical approaches.**

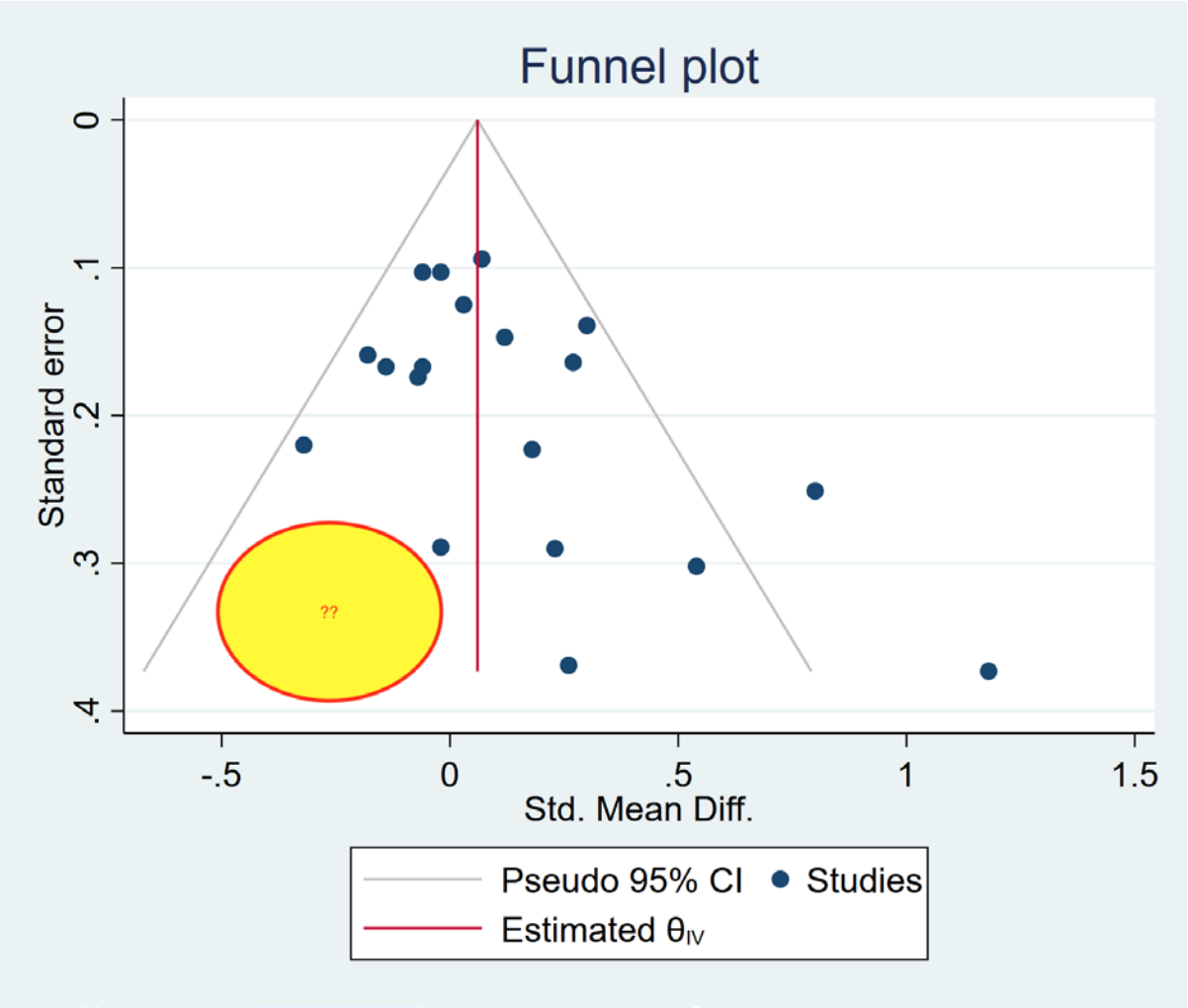


Figure source: https://www.stata.com/new-in-stata/meta-analysis/i/meta_funnel.png

Statistical Approaches to Assessing Publication Bias

None are "very good."

Popular

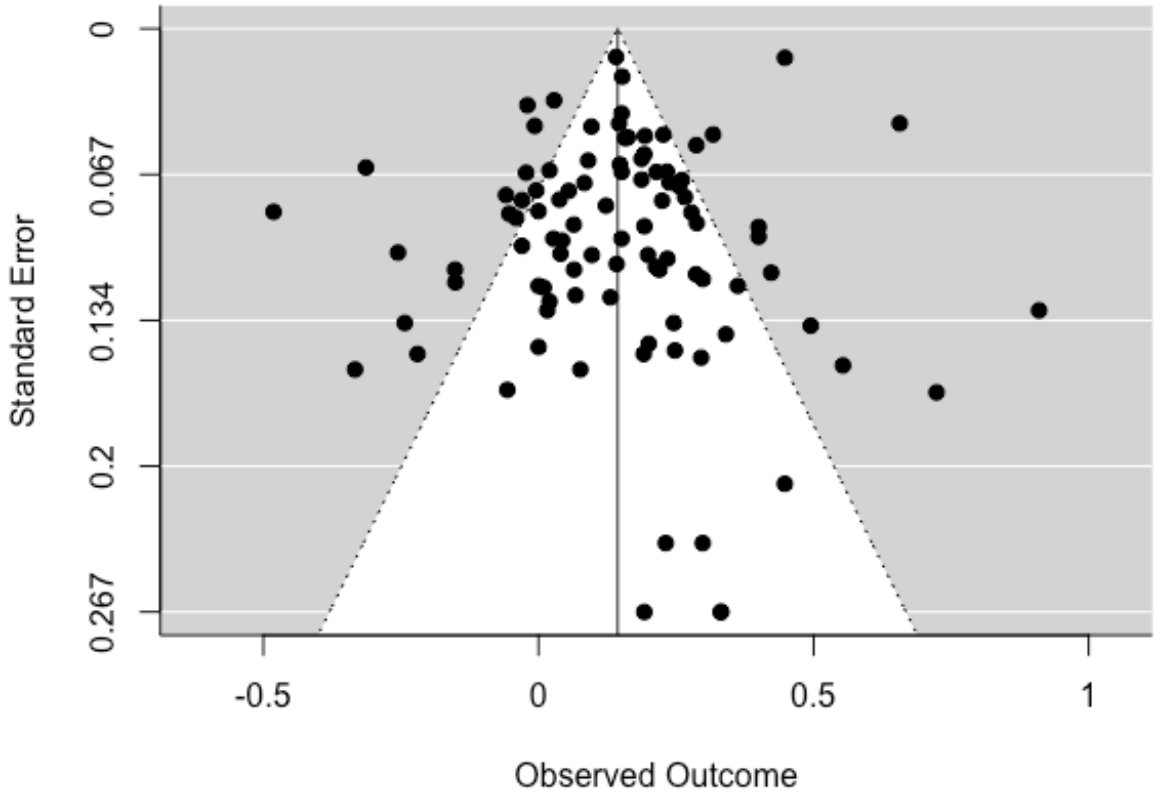
The most popular statistical approach is Rosenthal's fail-safe N (not recommended).

Alternatives

Better statistical approaches include trim and fill, Egger's regression test, and the rank test.

Trim and Fill Uses an Algorithm to Detect Funnel Plot Asymmetry

If a funnel plot is asymmetric, the algorithm identifies and trims the most influential study causing asymmetry.

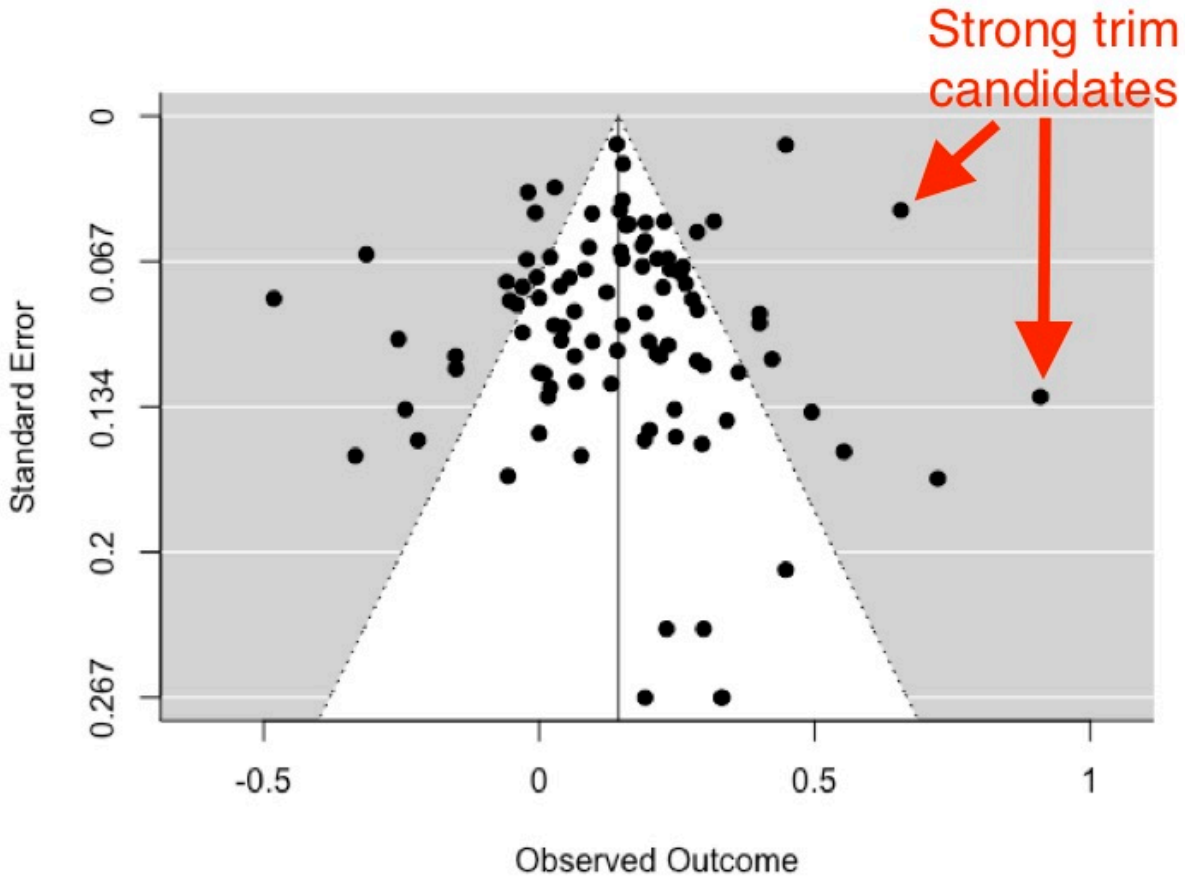


Data source: Fong, C. J., Davis, C. W., Kim, Y., Kim, Y. W., Marriott, L., & Kim, S. (2017). Psychosocial factors and community college student success: A meta-analytic investigation. *Review of Educational Research, 87*(2), 388-424.

Trim and Fill Uses an Algorithm to Detect Funnel Plot Asymmetry

This process (symmetry test, trim) repeats until the algorithm concludes that the plot is symmetric.

That's the "trim" part.

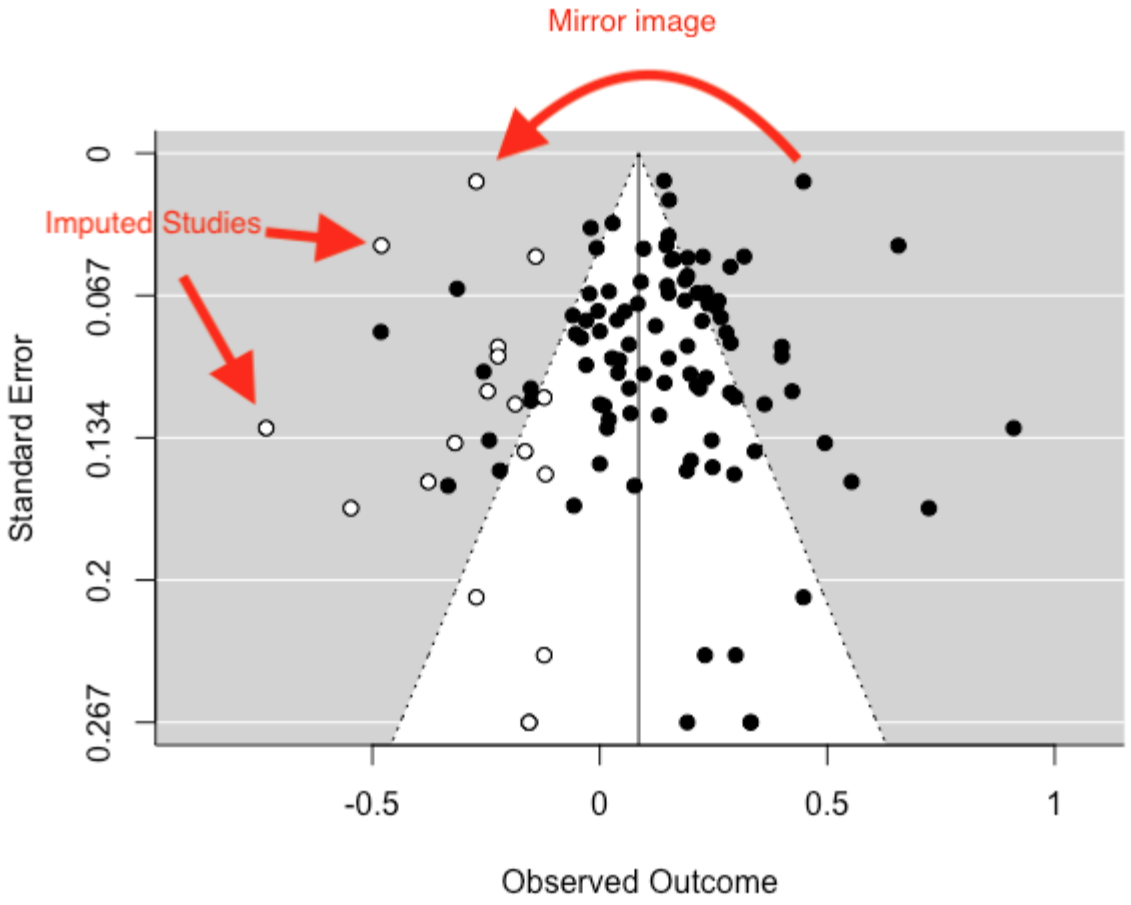


Data source: Fong, C. J., Davis, C. W., Kim, Y., Kim, Y. W., Marriott, L., & Kim, S. (2017). Psychosocial factors and community college student success: A meta-analytic investigation. *Review of Educational Research, 87*(2), 388-424.

Trim and Fill Uses an Algorithm to Detect Funnel Plot Asymmetry

The algorithm **adds** the trimmed studies back and **imputes missing studies** that mirror the trimmed studies.

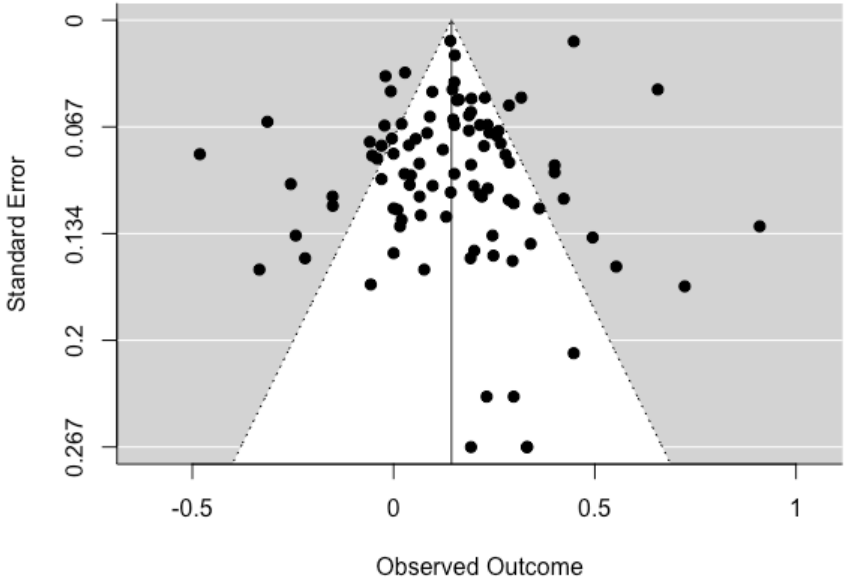
This is the "fill" part.

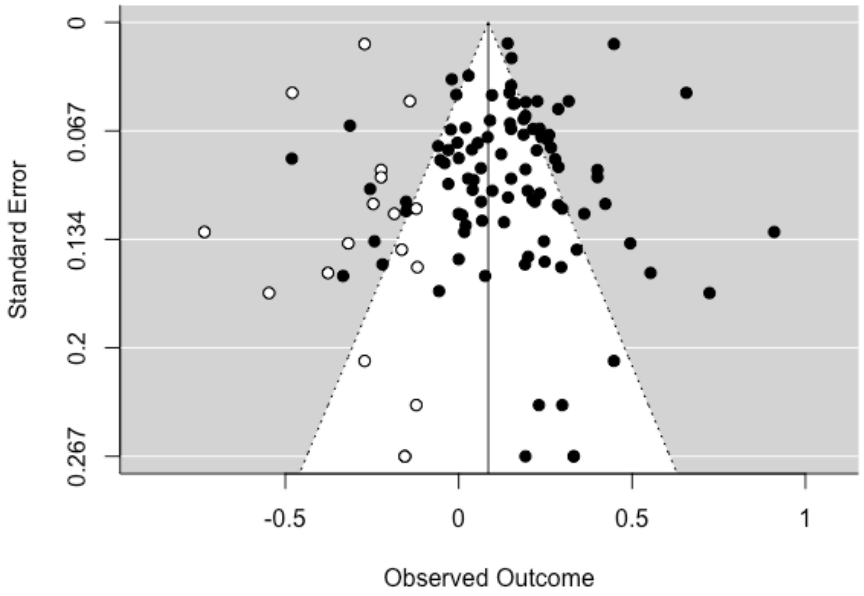


Data source: Fong, C. J., Davis, C. W., Kim, Y., Kim, Y. W., Marriott, L., & Kim, S. (2017). Psychosocial factors and community college student success: A meta-analytic investigation. *Review of Educational Research, 87*(2), 388-424.

Trim and Fill Output Addresses All Three Questions

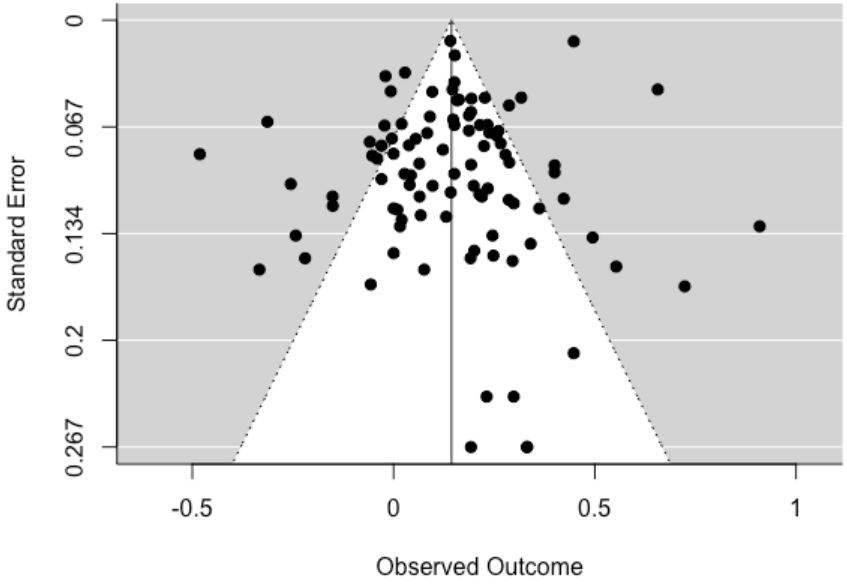
Question about publication bias	How trim and fill addresses the question
Is there any evidence of publication bias?	If no studies are imputed, there is no evidence of publication bias (per the algorithm).
Is it possible that the entire effect is an artifact of publication bias?	The algorithm re-computes the meta-analytic mean and statistical significance based on imputed studies.
How much of an impact might publication bias have?	The output allows us to compare the importance of the unadjusted and the adjusted effect sizes.

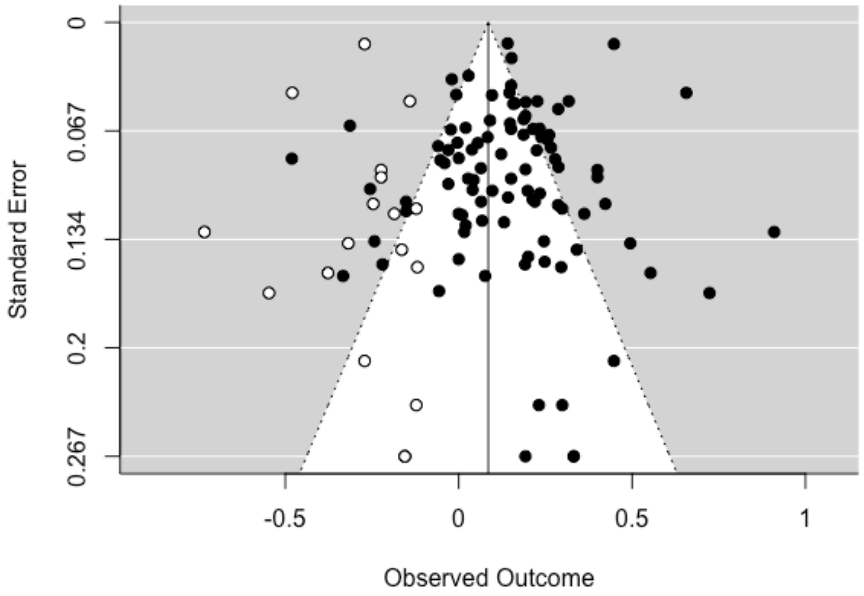




Q: Is publication bias present?

Data source: Fong, C. J., Davis, C. W., Kim, Y., Kim, Y. W., Marriott, L., & Kim, S. (2017). Psychosocial factors and community college student success: A meta-analytic investigation. *Review of Educational Research, 87*(2), 388-424.





Q: Is publication bias present?

A: It appears so – the algorithm imputed 18 "missing" studies. There were 95 studies originally.

Data source: Fong, C. J., Davis, C. W., Kim, Y., Kim, Y. W., Marriott, L., & Kim, S. (2017). Psychosocial factors and community college student success: A meta-analytic investigation. *Review of Educational Research, 87*(2), 388-424.

Q: Could the entire effect be attributable to publication bias?

	Mean correlation	SE	p
Original analysis	+.12	.029	< .001
Analysis with original + 18 imputed "missing" studies	+.09	.027	< .001

Data source: Fong, C. J., Davis, C. W., Kim, Y., Kim, Y. W., Marriott, L., & Kim, S. (2017). Psychosocial factors and community college student success: A meta-analytic investigation. *Review of Educational Research, 87*(2), 388-424.

Q: Could the entire effect be attributable to publication bias?

	Mean correlation	SE	p
Original analysis	+.12	.029	< .001
Analysis with original + 18 imputed "missing" studies	+.09	.027	< .001

A: It appears not – the re-estimated effect size is still statistically significant and not statistically significantly different from the original estimate.

Data source: Fong, C. J., Davis, C. W., Kim, Y., Kim, Y. W., Marriott, L., & Kim, S. (2017). Psychosocial factors and community college student success: A meta-analytic investigation. *Review of Educational Research, 87*(2), 388-424.

Q: How much of an impact might publication bias have had?

	Mean correlation	SE	p
Original analysis	+.12	.029	< .001
Analysis with original + 18 imputed "missing" studies	+.09	.027	< .001

Data source: Fong, C. J., Davis, C. W., Kim, Y., Kim, Y. W., Marriott, L., & Kim, S. (2017). Psychosocial factors and community college student success: A meta-analytic investigation. *Review of Educational Research, 87*(2), 388-424.

Q: How much of an impact might publication bias have had?

	Mean correlation	SE	p
Original analysis	+.12	.029	< .001
Analysis with original + 18 imputed "missing" studies	+.09	.027	< .001

A: Some – the mean correlation re-estimated with the imputed studies is somewhat smaller in magnitude than the original estimate, even though it is not statistically significantly different from the original estimate.

Is the correlation with the imputed studies is still large enough to be meaningful? My sense is probably "yes."

Data source: Fong, C. J., Davis, C. W., Kim, Y., Kim, Y. W., Marriott, L., & Kim, S. (2017). Psychosocial factors and community college student success: A meta-analytic investigation. *Review of Educational Research, 87*(2), 388-424.

The Importance of the Literature Search

Bottom line message:

There are no “very good” methods for addressing publication bias. Prevention is our best defense.

Why?

Rosenthal's fail-safe N essentially always yields the "right" answer (publication bias is not a problem).

Even better methods:

- Require a large number of studies (25+)
- Are confused by heterogeneity
- Assume that a relationship between study size and effect size is due to publication bias and not something else

Relationship Between Sample Size and Effect Size

All publication bias methods interpret the relationship between study size and effect size as evidence of publication bias.

However, large and small studies might have different effect sizes for other reasons, including:

- If **implementation quality** is related to effect size, and larger studies are harder to implement with fidelity, effect sizes will be smaller in larger studies.
- **Larger studies** might have **more funding**, which might plausibly translate into better study design.

Both graphical and statistical approaches will interpret these situations as evidence of publication bias, but they are not.

Publication Bias Tests As Sensitivity Analyses

- A good suggestion is to triangulate across the different publication bias tests.
- Think of them as sensitivity analyses – they do not provide a definitive answer one way or the other.
- Remember that narrative reviews are almost always based solely on published studies and cannot do anything to address publication bias.
- Even though meta-analysts are in an imperfect situation, that situation is still much better than that of most narrative reviews.

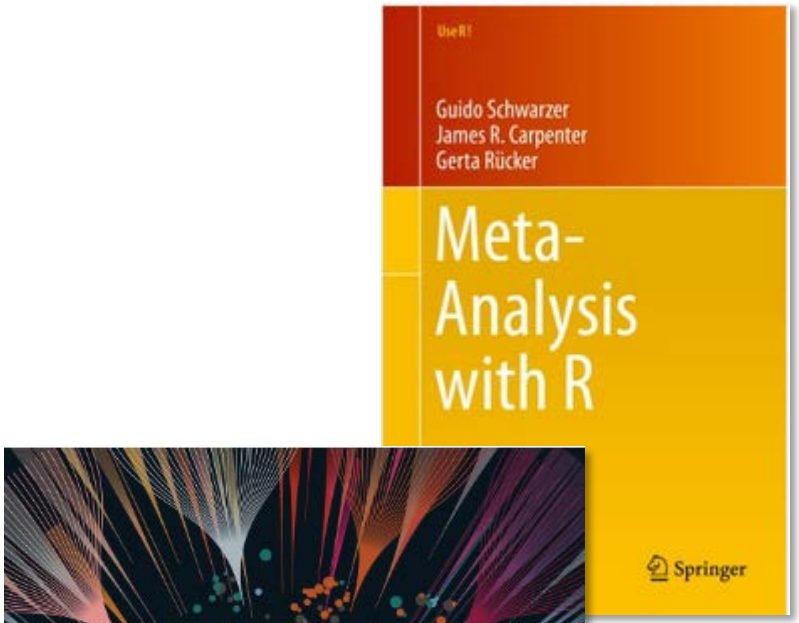
Summary: What Should We Do About Publication Bias?

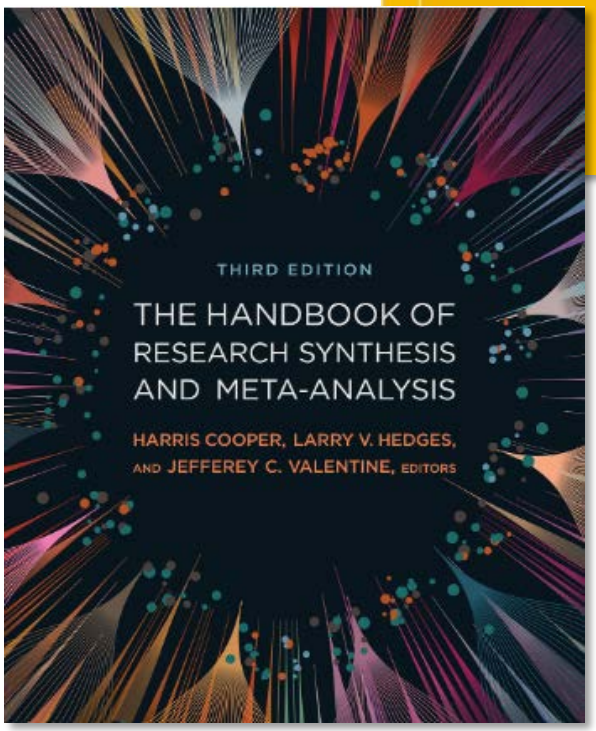
As a producer of a **systematic review:**	As a consumer of a **systematic review:**
• Start with a robust literature search. • Produce a funnel plot. • Use publication status in a moderator analysis if you can (multivariate model preferred). • Run multiple publication bias tests (at least two) if you can.	• Look for a robust literature search. • Examine the ratio of published to unpublished studies. • Examine funnel plots, if present. • Examine any statistical tests conducted for publication bias. • Generally, be skeptical of claims made by study authors about publication bias.

Recommended Resources

Vevea, J. L., Coburn, K., & Sutton, A. (2019). Publication bias. In H. Cooper, L. V. Hedges, and J. C. Valentine (Eds), *The handbook of research synthesis and meta-analysis* (3rd ed.), pp. 383-432. New York: Russell Sage Foundation.

Schwarzer, G., Carpenter, J. R., & Rücker, G. (2015). *Meta-analysis with R* (Vol. 4784). Springer.





Q&A

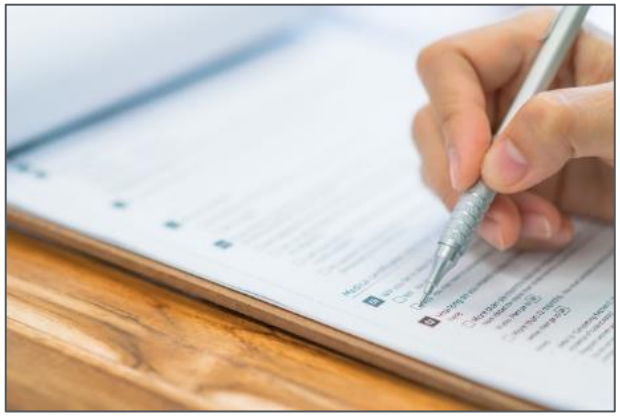
Looking Forward

Melissa Rasberry, EdD

Handling Statistical Dependencies and Other Advanced Topics

Emily Tanner-Smith, PhD

Looking Forward



Please fill out a feedback survey following the webinar.

Recording will be available soon on the CADRE website.

http://cadrek12.org/

Handling Statistical Dependencies and Other Advanced Topics

Emily Tanner-Smith, PhD

Statistical Dependencies in Meta-Analyses

Traditional meta-analysis approaches assume effect size estimates are statistically independent, but dependencies are common in the STEM education literature:

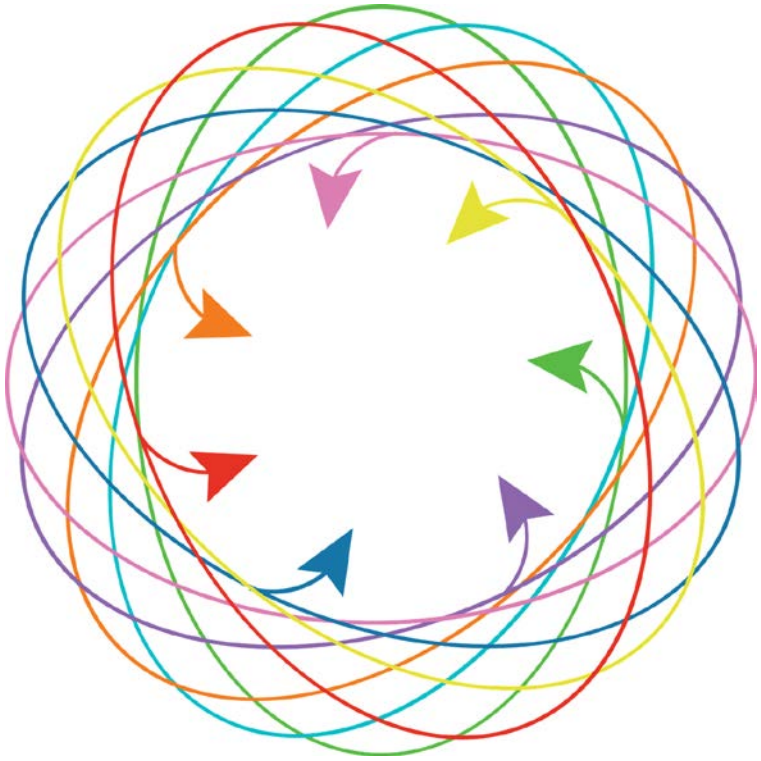
- Multiple participant subgroups within a study
- Multiple intervention or comparison groups within a study
- Multiple outcome measures within a study
- Multiple time-points within a study
- Multiple analyses within a study

Reductionist Approaches to Handling Statistical Dependencies

Reductionist approaches include one effect per study in any given meta-analysis model.

- Random selection
- Selection using decision rule(s)
- Averaging/synthesizing effect sizes within a study

Integrative Approaches to Handling Statistical Dependencies

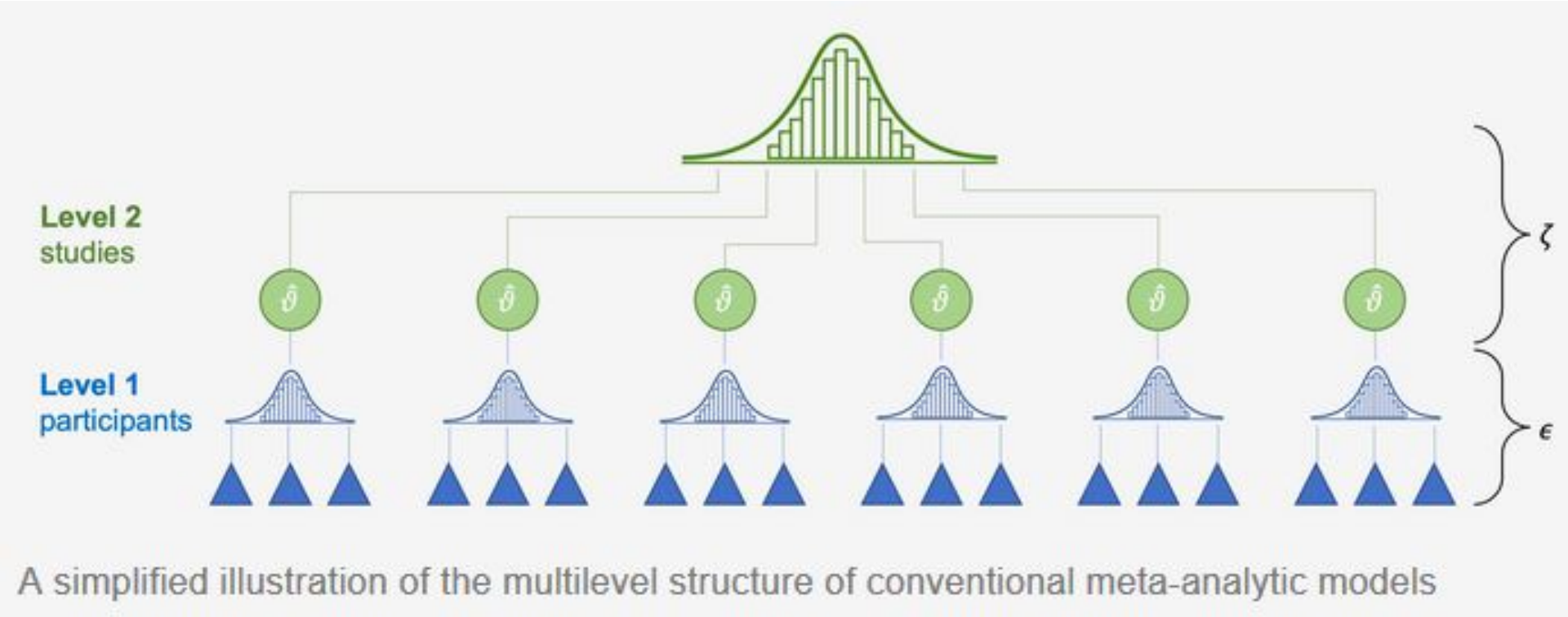


Integrative approaches include multiple effects per study in any given meta-analysis model:

- (3-level) multilevel meta-analysis
- Full multivariate meta-analysis
- Robust variance estimation

Multilevel Meta-Analysis

Standard meta-analysis models presuppose a (2-level) multilevel data structure.

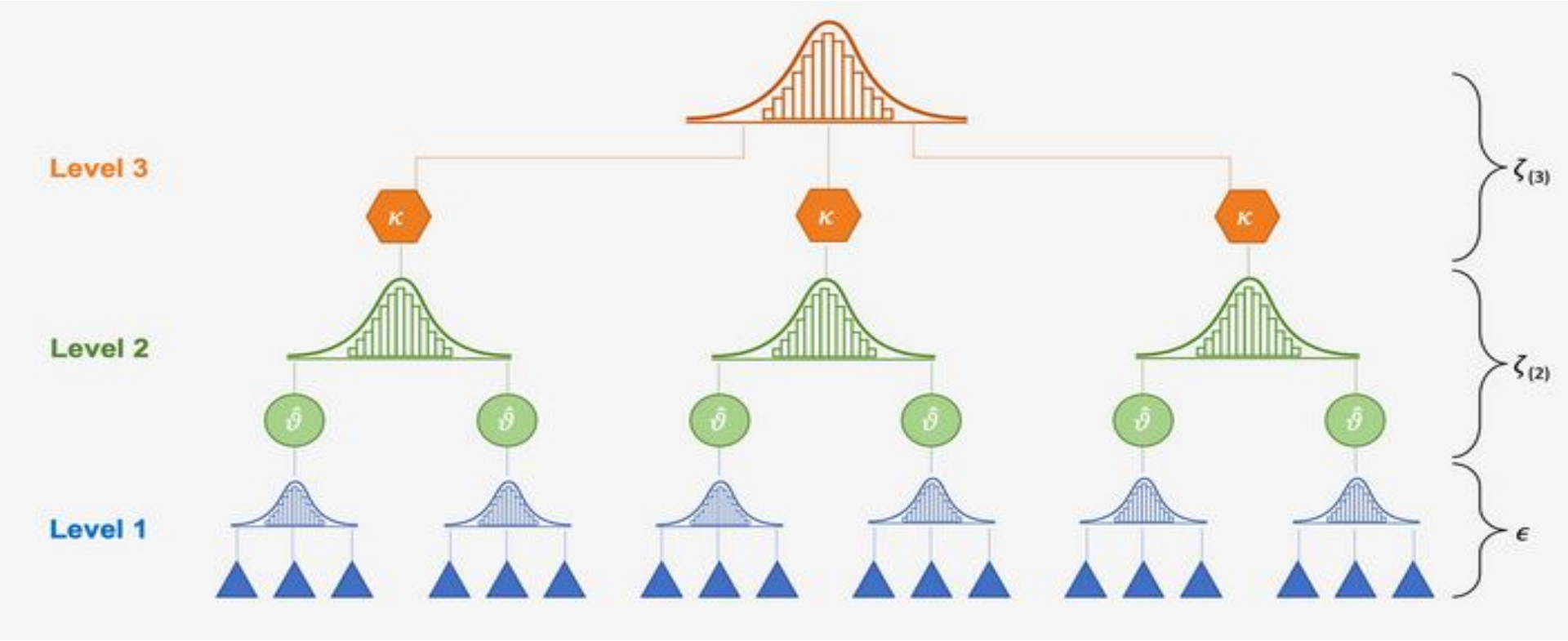


A simplified illustration of the multilevel structure of conventional meta-analytic models

Figure Source: Harrer, M., Cuijpers, P., Furukawa, T., & Ebert, D. D. (2019). "Multilevel meta-analysis. In: *Doing meta-analysis in R: A hands-on guide*. Retrieved from https://bookdown.org/MathiasHarrer/Doing_Meta_Analysis_in_R/mlma.html

Multilevel Meta-Analysis

(3-level) Multilevel meta-analysis can be used to synthesize effect sizes that are



Can be extended to handle multiple effect sizes from the same study[1-2]

Figure Source: Harrer, M. Cuijpers, P., Furukawa, T., & Ebert, D. D. (2019). "Multilevel" meta-analysis. In: *Doing meta-analysis in R: A hands-on guide*. Retrieved from https://bookdown.org/MathiasHarrer/Doing_Meta_Analysis_in_R/mlma.html

[1] Van den Noortgate et al. (2013), doi:10.3758/s13428-012-0261-6; [2] Van den Noortgate et al. (2015), doi:10.3758/x13428-014-0527-2

Full Multivariate Meta-Analysis

Multivariate meta-analysis is another method for jointly synthesizing evidence on multiple correlated effects, for example:

- Algebra scores, Calculus scores
- Math GRE scores, Science GRE scores

Can provide reduced bias and improved precision, particularly when:

- Within-study correlations between outcomes are known
- Outcomes are missing at random across studies

Robust Variance Estimation

Extends the standard meta-regression model to use heteroskedastic-robust clustered standard errors that account for dependencies.

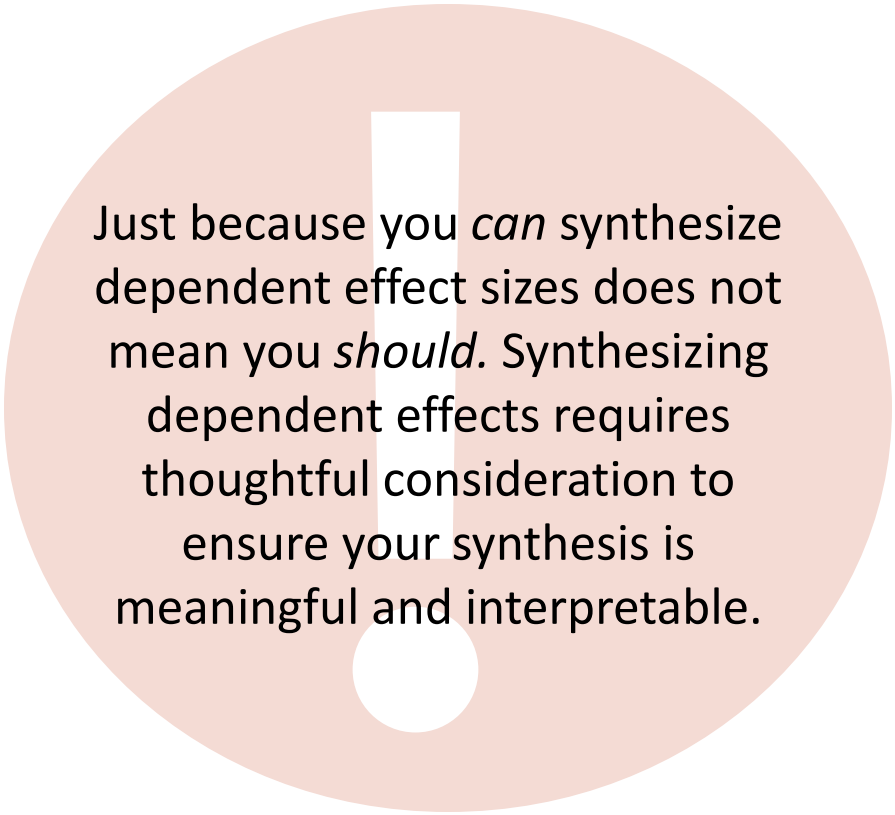
One of the most flexible integrative approaches available:

- Can simultaneously handle multiple types of dependencies
- Does not require accurate estimate of covariance structure between dependent effects
- Can be used with any form of weights, although inverse-variance weights are the most efficient

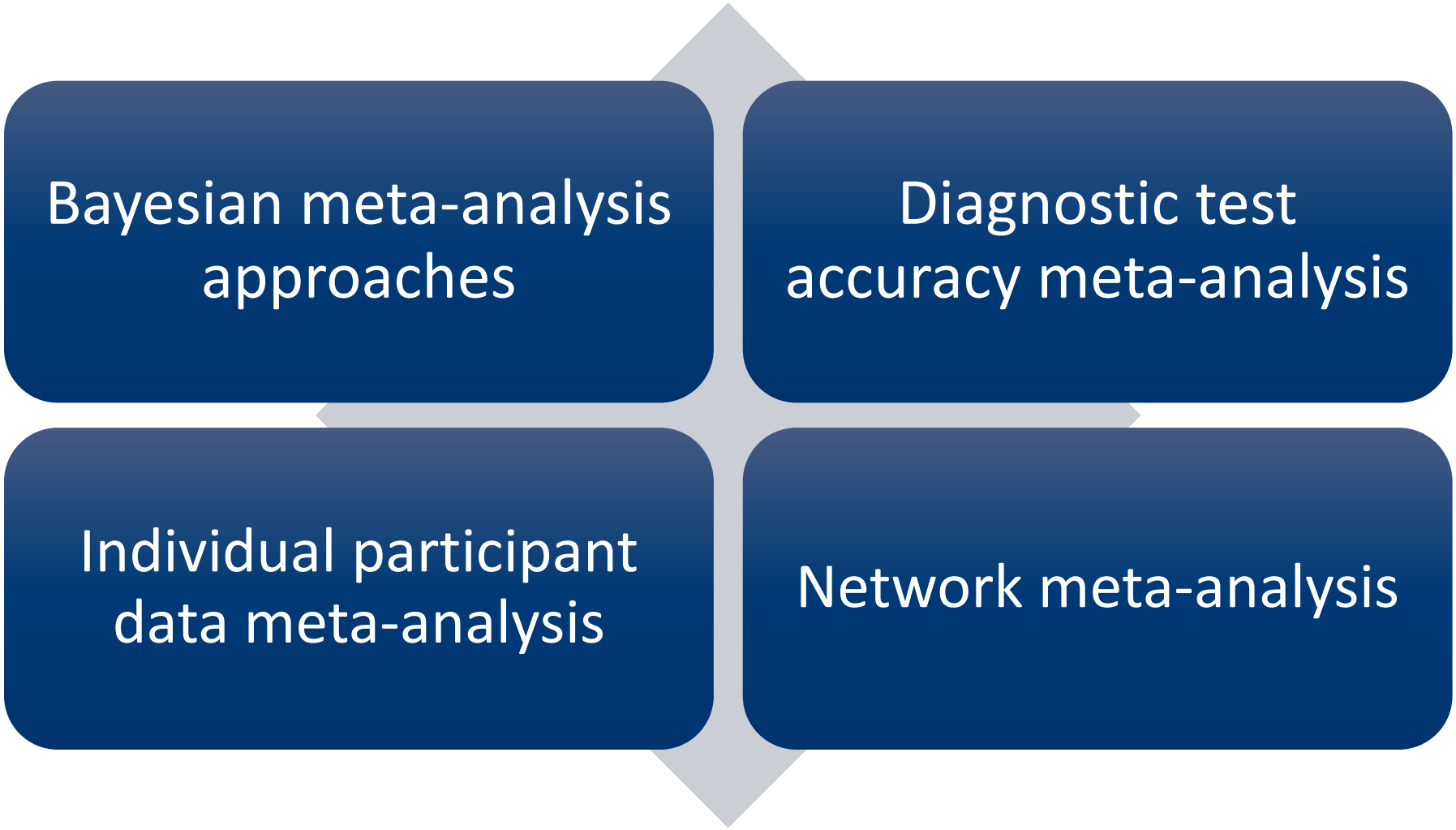
Robust Variance Estimation

Limitations of the method:

- May not perform well with a small number of studies (apply small sample adjustments)
- Inefficient estimation in some scenarios
- Does not provide meaningful estimates of heterogeneity (τ^2 is a nuisance parameter)



Other Advanced Meta-Analysis Methods



Examples From STEM Education Meta-Analyses

How do digital games affect K-16 student learning outcomes, and how do those effects vary depending on game mechanics and game design features?[1]

How does computer-based scaffolding affect cognitive learning when used by students working in different sized groups in problem-centered STEM instruction?[2]

Does reading comprehension vary based on reading media type (paper-based vs. screen-based reading)?[3]

[1] Clark et al. (2016), doi:10.3102/0034654315582065; [2] Kim et al. (2020), doi:10.1007/s10648-019-09502-3; [3] Kong et al. (2018) doi:10.1016/j.compedu.2018.05.005