Meta-Analysis De-Mystified: A Step-by-Step Workshop

A Presentation at the Annual Conference of
the American Evaluation Association
in Anaheim, California, 2011

Pedro Mateu, Western Michigan University
Kristin A. Hobson, Western Michigan University
Robert H. McCowen, Western Michigan University

Overview

- Background and Context
- Applicability in the Real World
  - When can you conduct a meta-analysis?
  - What is the key in the process?
- Step by Step Process
- Conclusions
- Questions

Background and Context

- **Definition**

  Meta-analysis seeks to integrate conclusions across multiple studies and disciplines by applying statistical analyses to groups of studies.

- **Why it is important?**

  Through meta-analysis, evaluators can estimate not only the central tendency of study outcomes, test the pattern of outcome variations, and estimate the overall effects and relationships of variables, but also predict results of future evaluations.

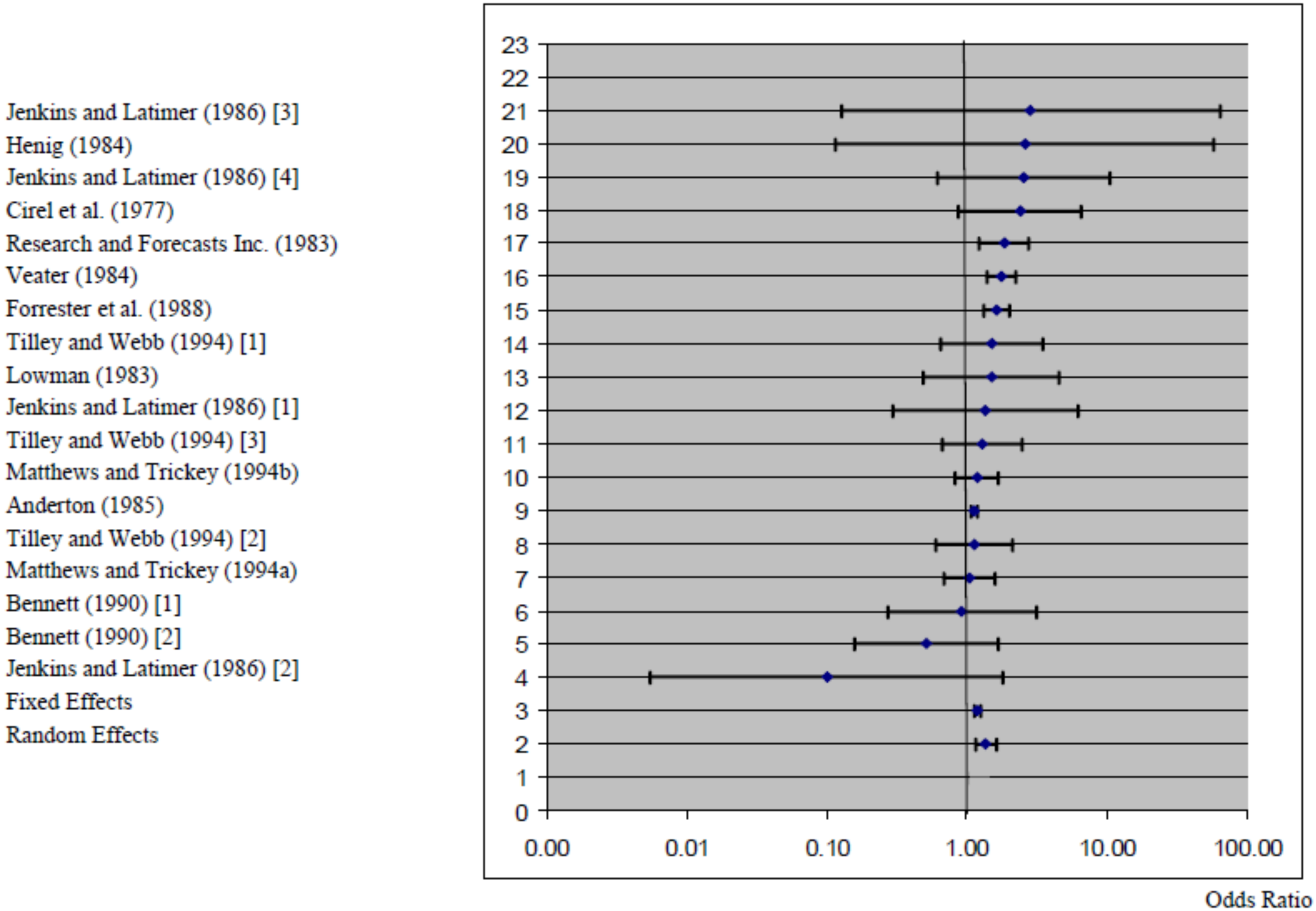
Background and Context

- **Value**

  The application of this tool is valuable because meta-analysis yields objective, defensible, and largely value-neutral evidence, which policy- and decision-makers could reference when forming and revising policies and programs.

- **Caveats**
  - Narrative reviews are not meta-analyses.
  - Meta-evaluation is different from meta-analysis.

A Visual Summary of Meta-Analysis



Source: Bennett, T., Holloway, K., & Farrington, D. (2008). The Effectiveness of Neighborhood Watch. *Campbell Systematic Reviews 18*. Doi: 10.4073/csr.2008.18

Applicability in the Real World

- Which meta-analyses have been published?
  - Anderson, C. A., et al. (2010). Violent video game effects on aggression, empathy, and prosocial behavior in eastern and western countries: A meta-analytic review. Psychological Bulletin, 136(2), 151-173.
  - DuBois, D. L., Holloway, B. E., Valentine, J. C., & Cooper, H. (2002). Effectiveness of mentoring programs for youth: A meta-analytic review. American Journal of Community Psychology, 30(2), 157-197.
  - Peterson, J. L., & Shibley Hyde, J. (2010). A meta-analytic review of research on gender differences in sexuality, 1993–2007. Psychological Bulletin, 136(1), 21-38.
  - Wilson, S. J., & Lipsey, M. W. (2000).Wilderness challenge programs for delinquent youth: a metaanalysis of outcome evaluations. *Evaluation and Program Planning*, *23*(1), 1-12.

When can you conduct a meta-analysis?

- Meta-analysis is applicable to collections of research that
  - Are empirical, rather than theoretical
  - Produce quantitative results, rather than qualitative findings
  - Examine the same constructs and relationships
  - Have findings that can be configured in a comparable statistical form (e.g., as effect sizes, correlation coefficients, odds-ratios, proportions)
  - Are "comparable" given the question at hand

What is the key in the process?

- The effect size makes meta-analysis possible
  - It is the "dependent variable"
  - It standardizes findings across studies such that they can be directly compared
- Any standardized index can be an "effect size" as long as it meets the following criteria
  - Is comparable across studies
  - Represents both magnitude and direction of the relationship of interest
  - Is independent of sample size

Step by Step Process

1. Formulate the research question
2. Set inclusion and exclusion criteria
3. Identify studies
4. Develop a coding protocol
5. Code studies
6. Select effect size and precision
7. Convert effect sizes
8. Determine effect size model
9. Compute heterogeneity statistics
10. Determine publication bias
11. Interpret results

Step 1: Formulate the Research Question

1. Formulating the question to be answered
   - The statement of the topic will guide study selection, coding of information, and data analysis.
   - Set of examples

Example 1:

"How effective are challenge programs in reducing the subsequent antisocial behavior of juveniles with behavior problems? What are the characteristics of the least and most successful programs? Do these programs have favorable effects on other outcomes such as relations with peers, locus-of-control, and self-esteem?"

—Lipsey & Wilson (2001)

Example 2:

How effective was the neighborhood watch movement in reducing crime?

—Bennett, Holloway, & Farrington (2008)

Example 3:

The purpose of this meta-analysis is to review the research on feedback in higher education mathematics and science classrooms to determine to if there is an overall statistically significant gain in student achievement when feedback is employed.

—Bentz, Engelman, & McCowen (2011)

Step 2: Set Inclusion and Exclusion Criteria

2. Criteria for inclusion and exclusion of studies in the review
   - Method Quality Dilemma: include or exclude low quality studies?
   - Inclusion / Exclusion criteria
     - Language
     - Sample size
     - Type of publication
     - Study design
     - Data collection

Step 3: Identify Studies

3. Search strategy for identification of studies
   - Checking databases where studies can be searched.
   - Searching one or two databases is generally inadequate

Methods: Data Sources

- **Searched databases:**
  - Cochrane Central Register of Controlled Trials (CENTRAL)
  - Biological and Agricultural Index (since 1983)
  - Education Resources Information Center (ERIC) (since 1966)
  - Dissertations and Theses (since 1860)
  - Dissertation Abstracts, ASSIAbi
  - Google Scholar, MEDLINE (since 1965)
  - ProQuest (since 1971)
  - PsycINFO (since 1887)
  - PubMed (since 1949)
  - Science Direct, Scopus (since 1995)
  - Web of Science (Science Citation Index Expanded (since 1900)
  - Social Sciences Citation Index (since 1956)
  - JSTOR
  - International Bibliography of the Social Sciences (IBSS)

- Additional primary studies can be searched for by the reviewing reference lists of retrieved studies.

- **Don’t forget the search terms.**

Step 4: Develop a Coding Protocol

4. Development of Coding Protocol
   - Description of studies meeting the eligibility criteria
   - Coding forms and manual
   - File structure: flat vs hierarchical
   - Type of information to code
   - Report identification
   - Methodology: Type of intervention
   - Effect Size
     - Statistical methods used to calculate effect size
     - Statistic used for reporting final effect sizes
   - Confidence ratings

Step 5: Code Studies

- Minimum of two coders per study
- Measure inter-rater reliability
  - Coefficient of Agreement: proportion of observations on which raters agree.
  - Cohen's Kappa: extent to which agreement exceeds that which would be expected by chance.

Step 5 Continued: Coefficient of Agreement

		Coder 2		
		Characteristic Present (j_1)	Characteristic Not Present (j_2)	Row Total
Coder 1	Characteristic Present (i_1)	c_{11}	c_{21}	$n = c_{11} + c_{21}$
	Characteristic Not Present (i_2)	c_{12}	c_{22}	$n = +c_{12} + c_{22}$
	Column Total	$n = c_{11} + c_{12}$	$n = c_{21} + c_{22}$	$N = (c_{11} + c_{12} + c_{21} + c_{22})$

$$p_o = \sum_{i=1}^{c} \sum_{j=1}^{c} p_{ij} = \frac{c_{11} + c_{22}}{N}$$

where *c* denotes the total number of cells, *i* denotes the ith row, and *j* denotes the jth column.

Step 5 Continued: Cohen's Kappa

- Cohen's kappa k is calculated as

$$\kappa = \frac{p_0 - p_e}{1 - p_e}$$

- Here, p_o is observed agreement, and p_e is expected agreement.

Step 5 Continued: Cohen's Kappa

		Coder 2		Marginal Row Probabilities
		Characteristic Present (j_1)	Characteristic Not Present (j_2)	$p_{i.}$
Coder 1	Characteristic Present (i_1)	c_{11}	c_{21}	$p_{1.} = \frac{(c_{11} + c_{21})}{N}$
	Characteristic Not Present (i_2)	c_{12}	c_{22}	$p_{2.} = \frac{(c_{12} + c_{22})}{N}$
Marginal Column Probabilities	$p_{.j}$	$p_{.1} = \frac{(c_{11} + c_{12})}{N}$	$p_{.2} = \frac{(c_{21} + c_{22})}{N}$	$N = (c_{11} + c_{12} + c_{21} + c_{22})$

$$p_{\mathrm{e}} = \sum_{i=1}^{c}\sum_{j=1}^{c} p_{i.}p_{.j} = p_{.1}p_{1.} + p_{.2}p_{2.}$$

Step 6: Select Effect Size and Precision

- Select a comparable effect size measure for each evaluation/study, including its variance
  - Raw Mean Difference, *D*, from independent and dependent groups
  - Standardized Mean Difference, *d* and *g*, from independent and dependent groups

Step 6 Continued: Effect Sizes and Precision

- Effect sizes based on binary data: For risk ratios (RR), odds ratios (OR), and risk differences (RD) data are typically represented in 2×2 tables with cells A, B, C, and D.

	Events	Non-Events	N
Treated	A	B	n_1
Control	C	D	n_2

- For the purpose of meta-analysis, computations of RR and OR are conducted using a log scale. For RD all computations are performed on the raw units.

Step 6 Continued: Effect Sizes and Precision

- The estimate of the correlation population parameter ρ is r where

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

- For meta-analyses, r is converted to Fisher's z.

$$z = 0.5 \times \ln\left(\frac{1+r}{1-r}\right)$$

Step 7: Convert Effect Sizes

- Comprehensive Meta-Analysis 2.0 (software suggested) will automate this process and many effect sizes calculators are also useful.

http://www.campbellcollaboration.org/resources/effect_size_input.php

- To convert from the logs odd ratio to *d*

$$d = \text{Log Odds Ratio} \times \frac{\sqrt{3}}{\pi}$$

- To convert from *r* to *d*

$$d = \frac{2r}{\sqrt{1-r^2}}$$

Step 7 Continued:
Converting Among Effect Sizes

- To convert from d to r

$$r = \frac{d}{\sqrt{d} + a}$$

- Where a is a correction factor when $n_1 \neq n_2$

$$a = \frac{(n_1 + n_2)^2}{n_1 n_2}$$

Step 7 Continued: Confidence Intervals

- Assuming that an effect size is normally distributed

$$LL_Y = \bar{Y} - 1.96 \times SE_Y$$

- And

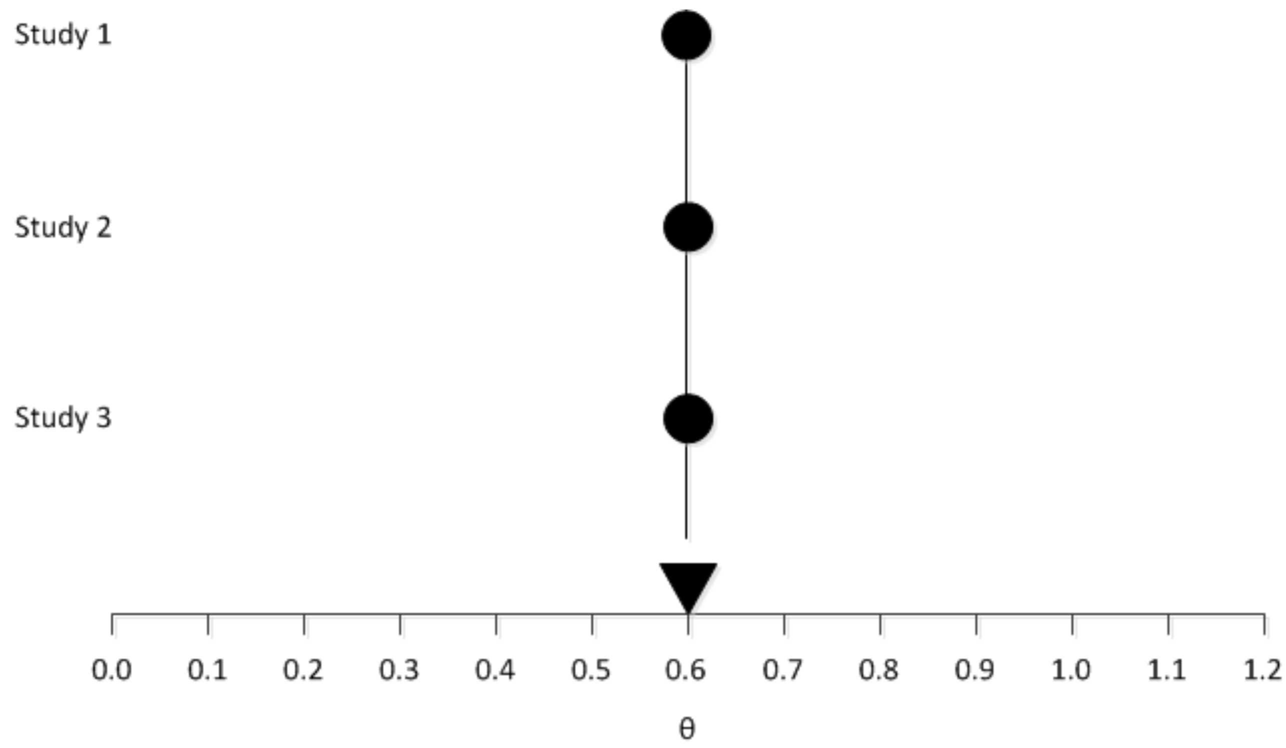
$$UL_Y = \bar{Y} + 1.96 \times SE_Y$$

- 1.96 is the Z-value corresponding to confidence limits of 95% (with error of 2.5% at either end of the distribution)

Step 8 Determine Effect Size Model: Fixed-Effect Models

- Under the fixed-effect model it is assumed that all studies share a common (true) effect size
- All factors that could influence the effect size are the same across all studies (the true effect is the same, thus *fixed*)
- The true (unknown) effect size is θ

Step 8 Continued: Fixed-Effect Models

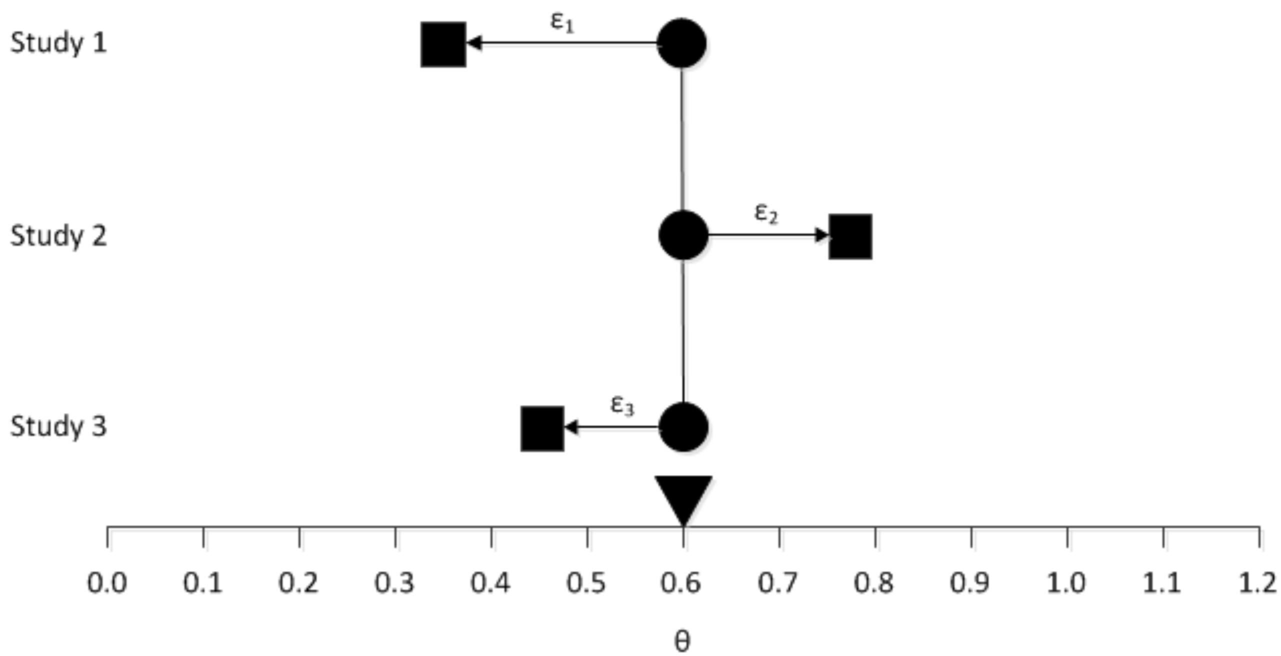


Fixed-effect model: True effect

Step 8 Continued: Fixed-Effect Models

- Given that all studies share the same true effect, the observed effect size varies from study to study only because of random (sampling) error.
- Although error is random, the sampling distribution of the errors can be estimated.

Step 8 Continued: Fixed-Effect Models



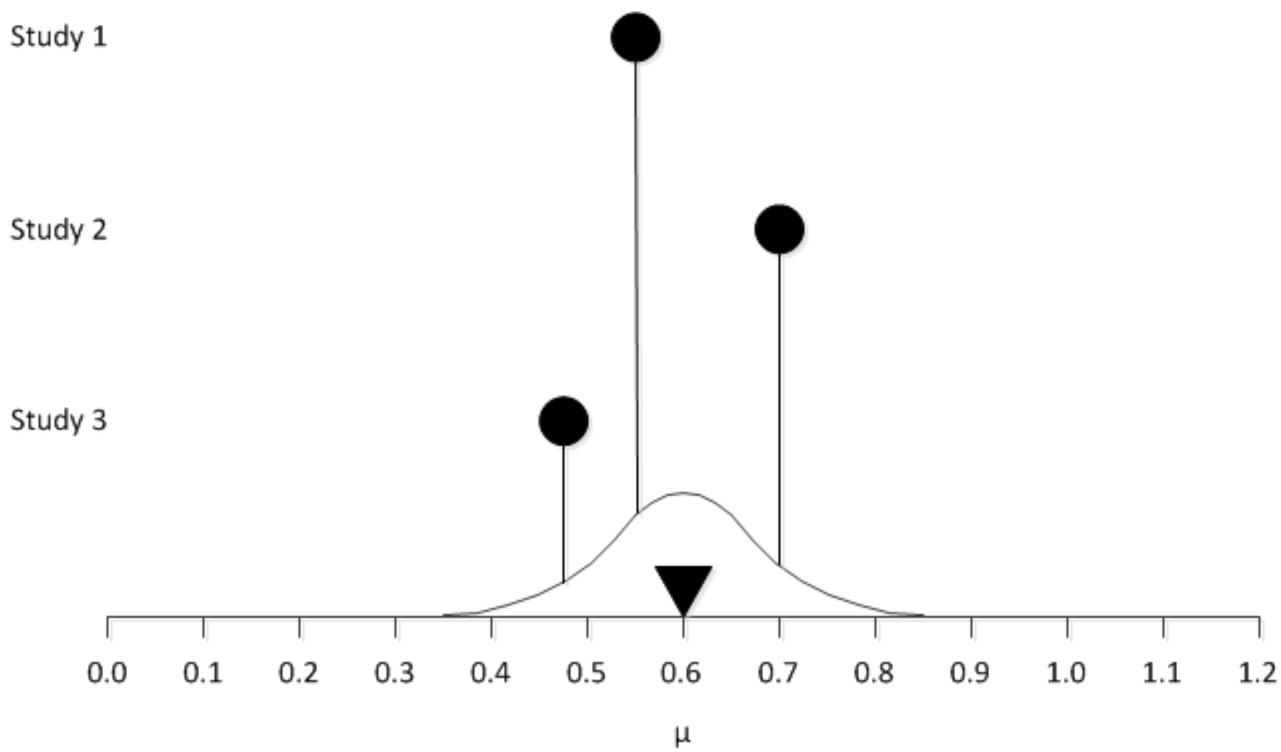
$$Y_i = \theta + \varepsilon_i$$

Fixed-effect model: True effects and sampling error

Step 8 Continued:
Random-Effects Models

- Does not assume that the true effect is identical across studies
- Because study characteristics vary (e.g., participant characteristics, treatment intensity, outcome measurement) there may be different effect sizes underlying different studies

Step 8 Continued: Random-Effects Models

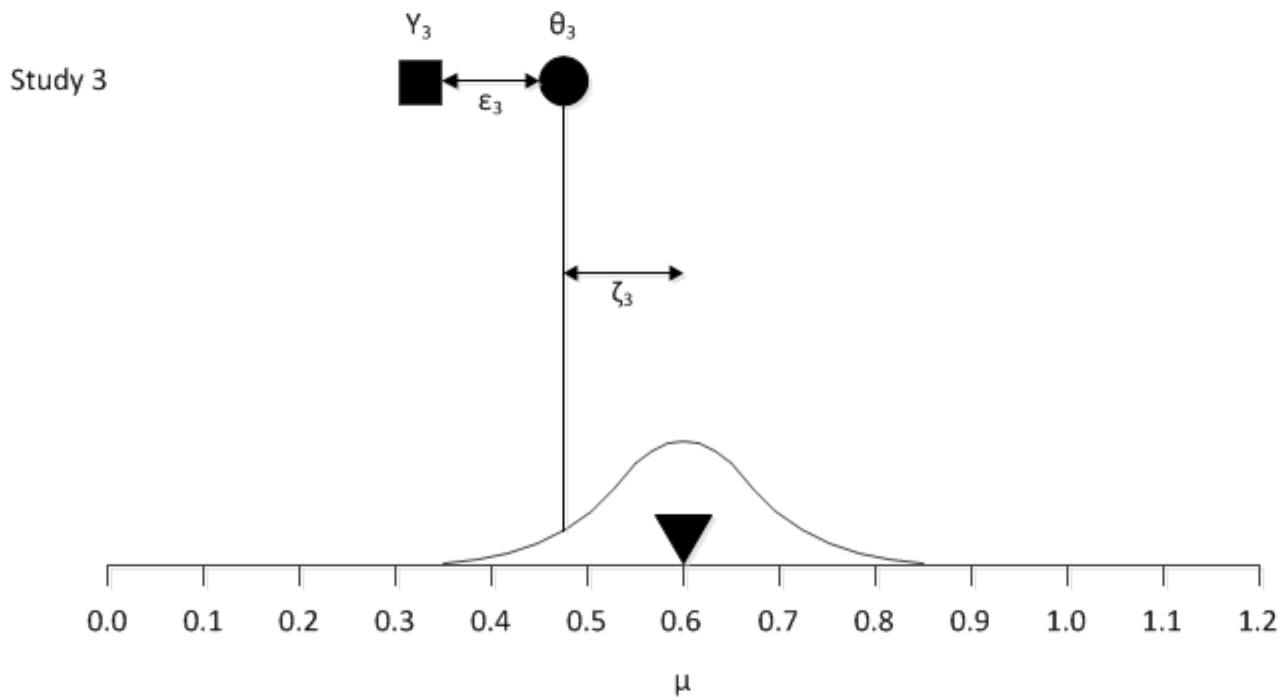


Random-effects model: True effects

Step 8 Continued:
Random-Effects Models

- If the true effect size for a study is θ_i then the observed effect will be less than or greater than θ_i due to sampling error
- The distance between the summary mean and the observed effect consists of true variation in effect sizes (ζ_i) and sampling error (ε_i)

Step 8 Continued: Random-Effects Models



$$Y_i = \mu + \zeta_i + \varepsilon_i$$

Random-effects model: True effect and observed effect

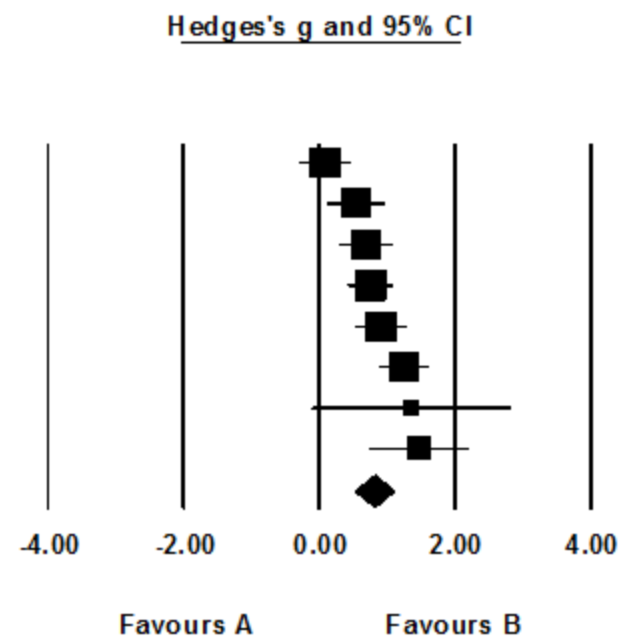
Step 8 Continued: Random-Effects Models

- The distance from μ to each θ_i depends on the standard deviation of the true effects across studies, which is represented as τ and τ^2 for its variance
- Each study's variance is a function of both the within-study variance and τ^2 and is the sum of these two values

Steps 6, 7 and 8: Let’s see an example

Meta Analysis

Study name	Statistics for each study						
	Hedges's g	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Freestone (C3)	0.091	0.197	0.039	-0.294	0.476	0.463	0.643
Butler	0.540	0.221	0.049	0.106	0.974	2.439	0.015
Krause (Co-Op)	0.690	0.202	0.041	0.293	1.087	3.408	0.001
Freestone (C1)	0.751	0.177	0.031	0.404	1.098	4.240	0.000
Freestone (C4)	0.911	0.200	0.040	0.520	1.302	4.565	0.000
Freestone (C2)	1.254	0.194	0.038	0.874	1.634	6.464	0.000
Lee	1.355	0.754	0.569	-0.123	2.833	1.796	0.072
Krause (Ind)	1.470	0.374	0.140	0.736	2.203	3.928	0.000
	0.799	0.154	0.024	0.498	1.101	5.194	0.000



Source: Bentz, Engelman, & McCowen (2011). Formative Assessment and Feedback in Higher Education Mathematics and Science Classrooms: A Meta-Analysis. Manuscript.

Step 9: Compute Heterogeneity Statistics

- Although we are usually concerned with the dispersion in true effect sizes, observed dispersion includes both true variance and random error
- The mechanism used to isolate true variance is to compare the observed dispersion with the amount expected if all studies shared a common effect size
  - The excess is assumed to reflect real differences among studies
  - This portion of the variance is used to create indices of heterogeneity

Step 9 Continued: An Example

		Effect Size and 95% CI			Test of Null		Heterogeneity				τ^2			
Model	N	M	LL	UL	Z	p	Q	$df(Q)$	p	I^2	τ^2	SE_{τ^2}	V_{τ^2}	τ
Fixed	12	0.94	0.85	1.02	-1.44	0.15	29.72	11	0.00	62.99	0.05	0.05	0.00	0.23
Random	12	0.87	0.72	1.06	-1.42	0.16								

Source: Hobson, K. A., Mateu, P., & Fields, J. K. (2011, April). A meta-analysis of studies of the effects of measles, mumps, and rubella vaccines on the development of autism among children. Poster presented at the 5th Annual Western Michigan University Research and Creative Activities Poster Day, Western Michigan University, Kalamazoo, MI.

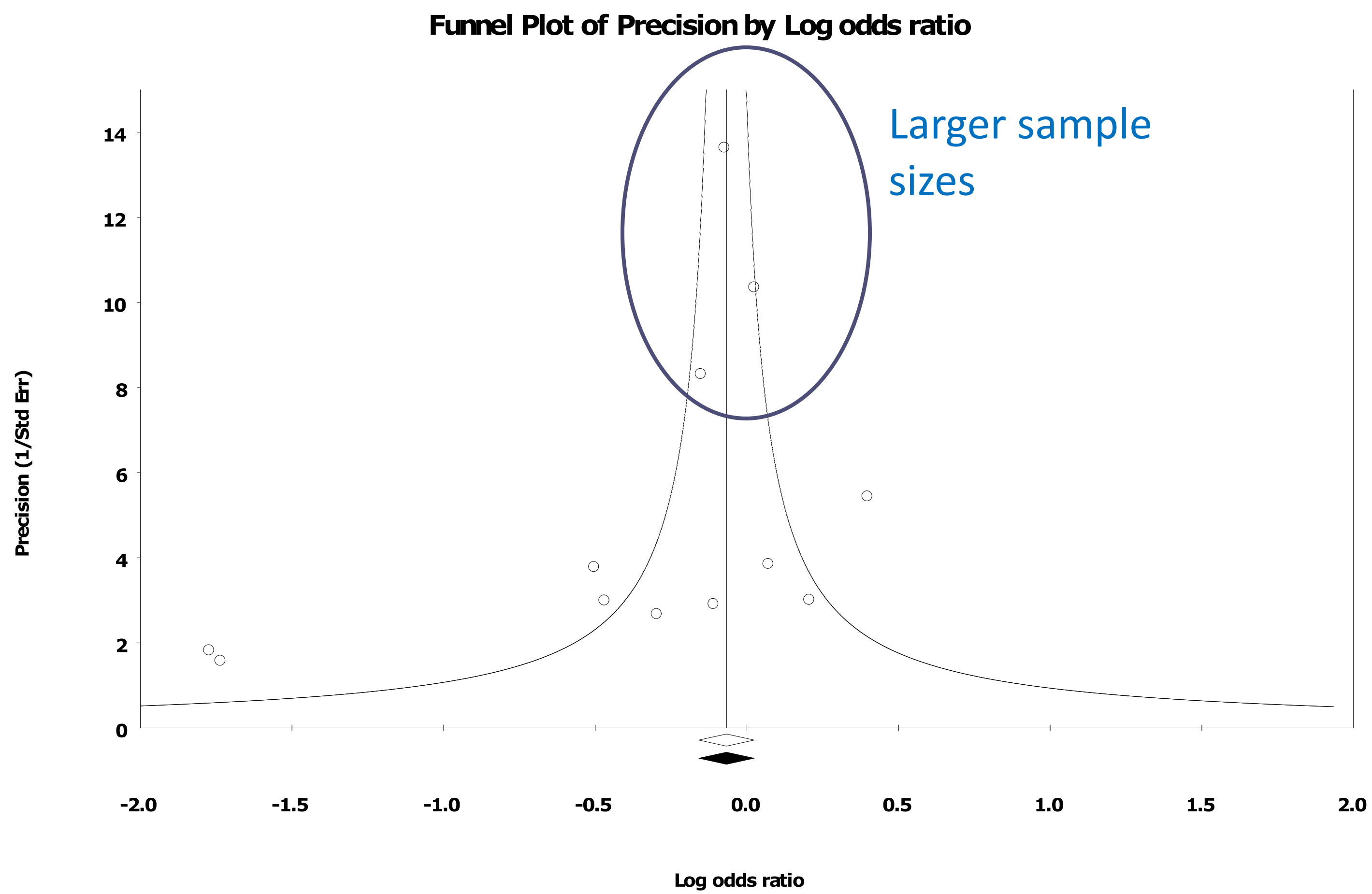
Step 10: Determine Publication Bias

- Publication bias is concerned with biases that arise from missing studies in a meta-analysis.
- Publication bias methods are used to determine if bias is likely, the impact of bias, and to make adjustments.
- The models used to assess publication bias assume
    - Large studies are likely to be published regardless of statistical significance

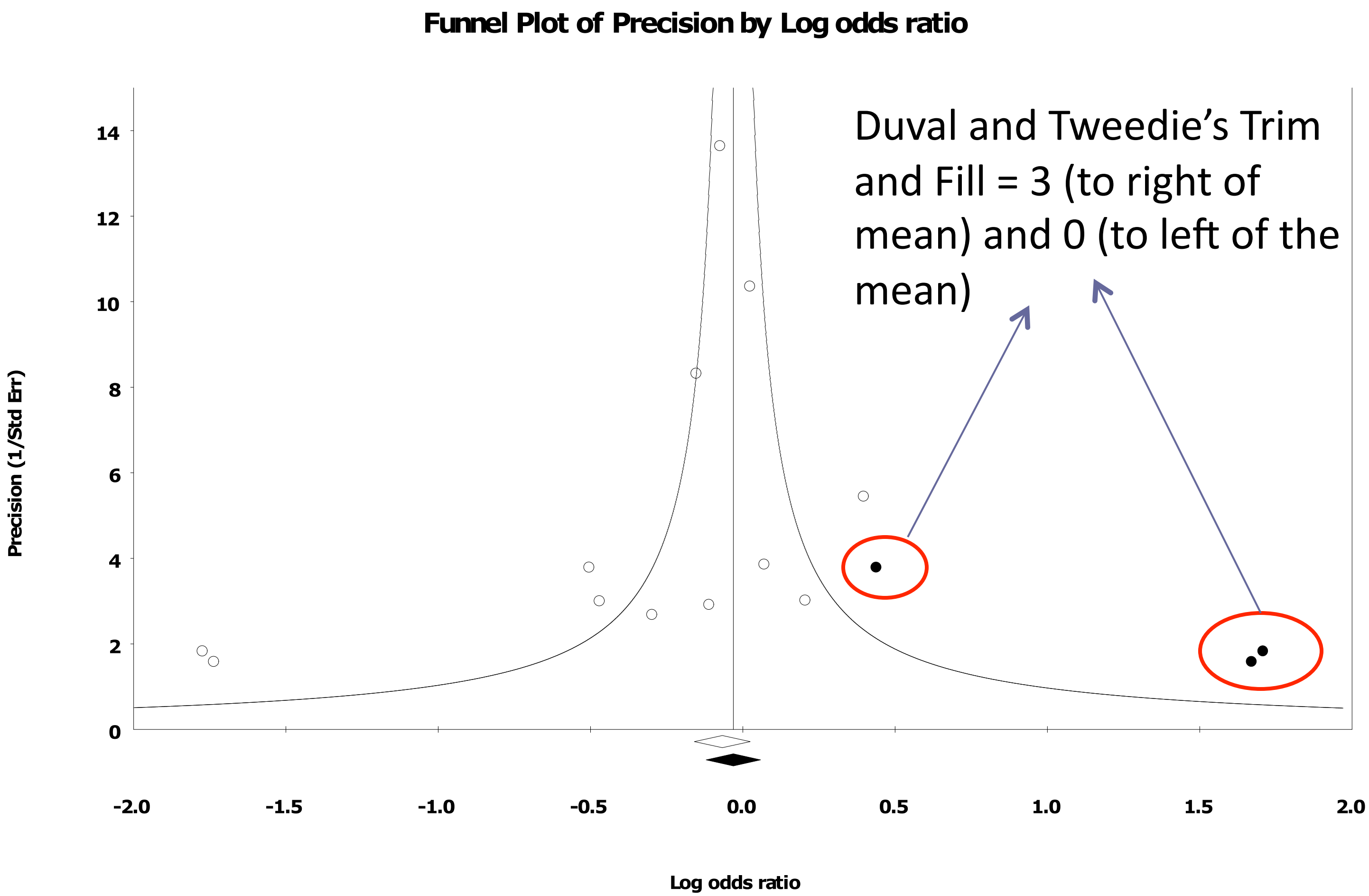
Step 10 Continued: Determine Publication Bias

- Evidence of Bias
  - Funnel plots
- Tests for publication bias
  - Duval and Tweedie's Trim-and-Fill procedure results in X imputed studies to the right of mean or to the left of the mean, based on an adjusted effect size.
  - Rank correlation (Kendall's τ) between the standardized effect size and its variances.
  - Egger's linear regression method.
  - Rosenthal's Fail-Safe *N* indicates the number of additional studies that should be located and included in the analysis to yield a statistically non-significant overall effect.

Step 10 Continued:
Funnel Plot (To Left of Mean)



Step 10 Continued: Funnel Plot (To Right of Mean)



Step 11: Interpret Results

- Example: Let's recall slide 5 →

Conclusion:

Strengths of Meta-Analysis (1)

- Imposes discipline on the process of summarizing research findings.
- Represents findings in a more differentiated and sophisticated manner than conventional reviews.
- Capable of finding relationships across studies that are obscured by other approaches.

Conclusion:

Strengths of Meta-Analysis (2)

- Protects against over-interpreting differences across studies.
- Can handle large numbers of studies (this would overwhelm traditional approaches to research review).

Conclusion:
Weaknesses of Meta-Analysis (1)

- Requires a good deal of effort.
- Mechanical aspects don’t lend themselves to capturing more qualitative distinctions between studies.
- “Apples and oranges” criticism.
- Most meta-analyses include “blemished” studies to one degree or another (e.g., a randomized design with attrition).

Conclusion:

Weaknesses of Meta-Analysis (2)

- Selection bias poses a continual threat.
  - Negative and null finding studies that you were unable to find.
  - Outcomes for which there were negative or null findings that were not reported.
- Analysis of between study differences is fundamentally correlational.

Questions?

Contact Information

Western Michigan University
Interdisciplinary Ph.D. in Evaluation

Pedro Mateu
pedro.f.mateu@wmich.edu

Kristin A. Hobson
kristin.a.hobson@wmich.edu

Robert H. McCowen
robert.mccowen@wmich.edu