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

Eval Café
Evaluation Center at WMU
January, 2012

Robert McCowen, IDPE
robert.h.mccowen@wmich.edu

Overview

- Background and Context
- Applications of Meta-Analysis
- Step-By-Step Process
- Conclusion

Background and Context

- Meta-analysis is:
  - A statistical technique for integrating the conclusions of multiple studies.
  - A straightforward but labor-intensive exercise.
  - A powerful but narrow quantitative tool.

Background and Context

- Meta-analysis lets you:
  - Estimate the central tendency of study outcomes.
  - Test the pattern of outcome variations.
  - Estimate the overall effects and relationships of variables.
  - Predict results of future studies.

Background and Context

- Meta-analysis in evaluation:
  - Summarizes what is already known (and is complementary to a thorough literature review).
  - Yields relatively objective evidence, and is largely value-neutral.*

**Your mileage may vary.*

Applications of Meta-Analysis

- A handful of published meta-analyses:
  - 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.

Applications of Meta-Analysis

- Meta-analysis is appropriate for research/evaluation topics that:
  - Produces quantitative, empirical results.
  - Examines a single construct, or a closely related set of constructs.
  - Produces an effect size measure (mean difference, correlation coefficient, odds ratio, etc.)

Applications of Meta-Analysis

- Effect size is the master key.
  - Using effect size allows direct comparison between disparate studies.
  - Effect size is the "dependent variable" in meta-analysis.

Applications of Meta-Analysis

- Any standardized index can be an effect size…
  - Standardized mean difference *d* and Hedges' *g*
  - Correlation *r* and Fisher's *z*
  - Odds ratio
  - Risk ratio
- …as long as it meets the following criteria:
  - Must be comparable across studies
  - Must represent both magnitude and direction of the relationship of interest
  - Must be independent of sample size

Meta-Analysis in One Simple Step

(And At Least Four Complicated Ones)

1) Formulate Research Question
2) Comprehensive Literature Search
3) Code Primary Studies
4) Summary Effect Size and Bias
5) Interpretation

Step 1: Formulate the Research Question

- Simple Skeletons:
  - What is the effect of [intervention] on [construct]?
  - How effective is [intervention] in [changing] [construct]?
- This is where you remember that meta-analysis is *not* a methodology for original evaluation or research.

Example 1:

"How effective was the neighborhood watch movement in reducing crime?"

—Bennett, Holloway, & Farrington (2008)

Step 1: Formulate Research Question

- Meta-analysis can examine multiple outcomes, as well as moderators and mediators—but the more complex the question, the more complex the coding.

Example 2:

"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)

Step 2: Comprehensive Literature Search

- Develop criteria for inclusion/exclusion of studies
  - Include or exclude low quality studies?
    - Do you include them?
    - How do you tell they're low quality?
  - Possible criteria:
    - Language
    - Sample size
    - Type of publication
    - Study design
    - Data collection method

Step 2: Comprehensive Literature Search

- Finding candidate studies:
  - Cast a wide net
  - Multiple databases
    - Record your search terms, and keep them consistent.
  - Follow citations
  - Ask authors
- Make a serious effort to find *everything*.

Step 3: Code Primary Studies

- Develop a coding protocol
  - Should include everything you need, and no more!
    - If it takes only 5 minutes to code for an item, and you have 20 studies, each item adds *at least* 5*20*2=200 minutes to coding time.
  - Plan for at least two coders per study, so you can report reliability
- At a minimum, code for:
  - Final effect size
  - All information required to calculate effect size (e.g. mean, variance, sample size)

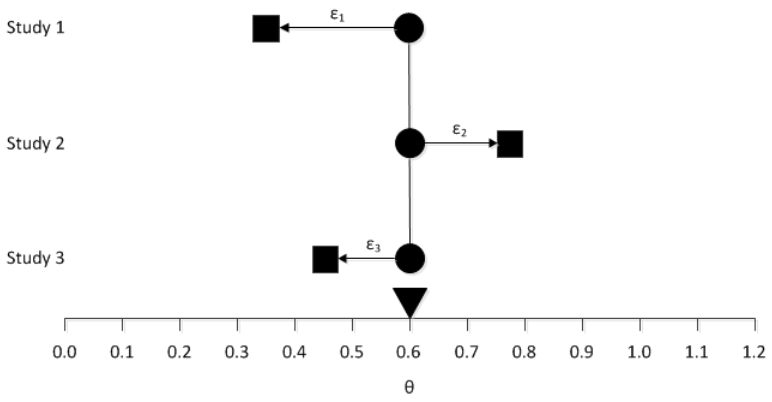
Step 3: Code Primary Studies

- Also consider coding:
  - Methodological information (sampling, assignment, group equivalence, pretest, attrition…)
  - Study quality/validity/confidence in results
  - Publication type (e.g. journal article, thesis, dissertation, book chapter)
- Also include anything unique to your field that might influence results or should be part of the analysis
  - Education -> grade level(s) of participants
  - Development -> country or region(s)
  - Public Health -> mean and variance of participant age

Step 4: Summary Effect Size and Bias

- In order to calculate a summary effect size (the goal of the entire exercise), decide between fixed-effects and random-effects models.
  - Some fields report both.
- Fixed-effects:
  - All studies share a single true effect size
  - Variation between studies is due *solely* to error
- Random-effects:
  - Each study samples from an underlying distribution of possible true effects
  - Variation between studies is due to *both* error and actual differences

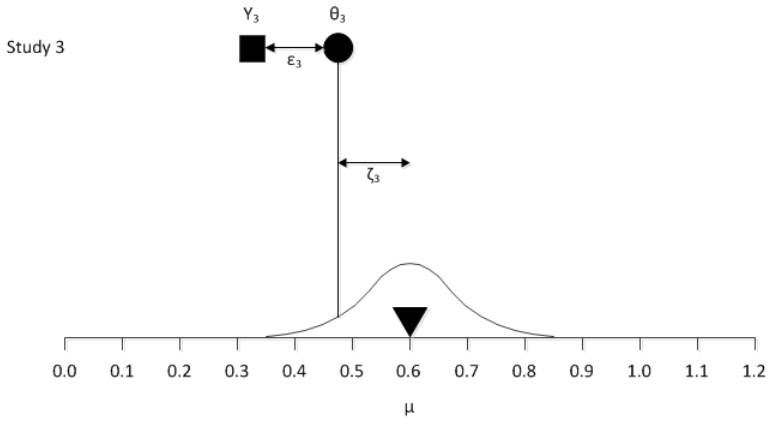
Step 4: Summary Effect Size and Bias



$$Y_i = \theta + \varepsilon_i$$

Fixed-effect model: True effects and sampling error

Step 4: Summary Effect Size and Bias



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

Random-effects model: True effect and observed effect

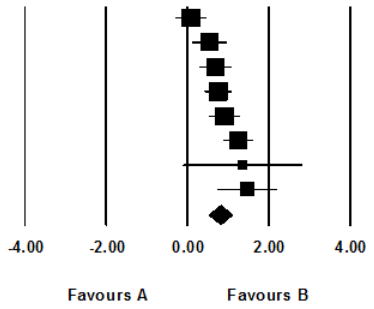
Step 4: Summary Effect Size and Bias

- Publication bias is likely the largest threat to validity for meta-analysis.
- Quantitative analysis can indicate if bias is likely, as well as indicating the direction and approximate magnitude of bias
- The models used to assess publication bias assume:
  - Studies with large N are more likely to be published than smaller studies.
  - Studies that find significant, large effects are more likely to be published than those that do not.

Step 5: Interpretation

Meta Analysis

Study name	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.

Conclusion: Strengths

- Meta-analysis is a *structured, quantitative* review of existing literature.
  - It can find patterns among studies that are obscured by even systematic qualitative review.
- Meta-analysis can handle an arbitrarily large number of studies.
  - Traditional review techniques are limited by the researcher or evaluator's ability to spot and recall patterns and links.
- Relative to an individual researcher, meta-analysis is objective and value-free.
  - It is not subject to familiar forms of cognitive bias.

Conclusion: Weaknesses

- The “apples and oranges” criticism.
  - Not a problem for measurement, but may be counter-intuitive for some members of the intended audience.
- Meta-analysis is exclusively quantitative.
  - Qualitative differences between studies are obscured.
  - Valid qualitative studies cannot be included.
- Publication bias is an *unavoidable* threat to validity.
  - “Grey literature”, desk-drawer problem, etc.
- Most meta-analyses include low-quality studies.
  - Hard to say how this affects the final conclusion, but “garbage in, garbage out” applies.

Further Reading

- Borenstein, M., Hedges, L.V., Higgins, J.P.T., & Rothstein, H.R. (2010). *Introduction to Meta-Analysis*. London: Wiley.
- Cooper, H., Hedges, L.V., & Valentine, J.C. (Eds.). (2009). *The Handbook of Research Synthesis and Meta-Analysis*. New York, NY: Russell Sage Foundation.
- Schwandt, T.A. (2000). Meta-analysis and everyday life: The good, the bad, and the ugly. *American Journal of Evaluation* 21(2), 213-219.

Questions and Discussion