Meta-Analysis: A Gentle Introduction to Research Synthesis

Jeff Kromrey

Lunch and Learn
27 October 2014
Discussion Outline

- Overview
- Types of research questions
- Literature search and retrieval
- Coding and dependability
- Effect sizes
- Describing results
- Testing hypotheses
- Threats to validity
- Reporting meta-analyses
- References worth pursuing
Overview

- Summarization of empirical studies using quantitative methods
  - Results
    - Estimated weighted mean effect size
    - Confidence interval around mean effect size (or test null hypothesis about mean effect size)
    - Homogeneity of effect sizes
    - Tests of moderators
Overview: Why Meta-Analyze?

- Strength in numbers
  - Several ‘non-significant’ differences may be significant when combined
- Strength in diversity
  - Generalizability across variety of participants, settings, instruments
  - Identification of moderating variables
- Good way to look at the forest rather than the trees
  - What do we think we know about a phenomenon?
  - How well do we know it?
  - What remains to be investigated?
- It’s fun!
Overview:

Stages of Meta-Analysis

- Formulate problem
- Draw sample / collect observations
- Measure observations
- Analyze data
- Interpret data
- Disseminate
Types of Research Questions: Treatments

- Does the treatment (in general) appear effective?
  - How effective?

- Does treatment effectiveness vary by
  - Participant characteristics?
  - Treatment characteristics?
  - Research method characteristics?

- Does the treatment appear ineffective in some conditions?
Types of Research Questions: Relationships

- What is the relationship (in general)?
  - Direction?
  - Strength?
- Does direction or strength of relationship vary by
  - Participant characteristics?
  - Treatment characteristics?
  - Research method characteristics?
- Is the relationship *not evident* in some conditions?
Decisions to make before searching the literature

- Inclusion/Exclusion criteria for sources
  - Types of publication
    - Language and country of publication
    - Dissemination: Journal, presentation, unpublished
  - Study characteristics
    - Participant characteristics
    - Information reported
    - Timeframe
    - Type of design
    - Measures
Literature Search and Retrieval

- Decisions to make before searching the literature
  - Search strategies
    - Keywords
    - Databases
      - ERIC, PsychInfo, GoogleScholar, Web of Science
    - Other
      - Key researchers
      - Listservs
      - Websites
      - Reference sections of articles
Coding of Studies

- **Record**
  - Study inclusion/exclusion characteristics
  - Effect size(s)
    - Multiple measures?
    - Subsamples?
    - Different times?
  - Other relevant variables
    - Research design (sampling, controls, treatment, duration)
    - Participant attributes (age, sex, race/ethnicity, inclusion/exclusion)
    - Settings (geography, classrooms, laboratory)
    - Dissemination characteristics (journal, conference, dissertation, year, Dr. B)
Coding of Studies (Cont’d)

- Written codebook and coding forms
- The Goldilocks principle: not too coarse, not too fine.
- Training and calibration of coders
  - Beware of drift
- Estimating reliability of coders
Study Coding Form

Meta-Analysis Coding Part I:  *Increased levels of stress will reduce the likelihood of XXX treatment success.*

STUDY TITLE:
I. Qualifying the study: *Answer the following questions as either “yes” or “no”.*

Does the study involve women participating in an XXX treatment program?

Does the study focus on the relationship between stress and XXX treatment outcomes?

Was the study conducted between January 1995 and December 2013?

Does the study employ a prospective design?

Does the study report outcome measures of stress or anxiety as well as XXX treatment outcomes?

If the answer to each of the above questions is yes, the study qualifies for inclusion in the meta-analysis.
II. Coding the study:

A. Publication Characteristics
1. Title of the study:______________________________

2. Year of Publication:____________________________

3. Authors:_____________________________________

B. Ecological Characteristics
1. Age of Female Participants: Mean:_____ Range:___

2. Country:_____________________

3. Race:
   White    N:_______  %:_______
   Black    N:_______  %:_______
   Hispanic N:_______  %:_______
   Asian / Pacific Islander N:_______  %:_______
   American Indian N:_______  %:_______
   Other     N:_______  %:_______
Duration of Psychoeducational Intervention (Please choose)
  a. Daily for duration of XXX treatment
  b. 1 – 3 sessions during XXX treatment
  c. 6 weeks during XXX treatment
  d. 8 weeks during XXX treatment
  e. 10 weeks during XXX treatment
  f. Other: ________________________________

Length of Psychoeducational Intervention (Please choose)
  a. 1 hour
  b. 1.5 hours
  c. 2 hours
  d. Other: ________________________________

Frequency of Psychoeducational Intervention (Please choose)
  a. Daily
  b. Weekly
  c. Bi-Weekly
  d. Other: ________________________________
Effect Size

- How false is the null hypothesis?
  - How effective is the treatment?
  - How strong is the relationship?
- Independent of sample size (more or less)
- Useful in primary studies and in meta-analysis
  - Links to power
  - Descriptive statistic (big enough to care?)
Effect Size (Cont’d)

- Jacob Cohen
  - *Statistical Power Analysis for the Behavioral Sciences*
  - Anytime a statistical hypothesis is tested, an effect size is lurking in there somewhere
  - Small, medium, large effects
  - *Medium effect size is big enough to be seen by the naked eye of the careful but naïve observer*
Effect Size:
Standardized Mean Difference

- Population effect size
  \[ \Delta = \frac{\mu_1 - \mu_2}{\sigma} \]

- Sample effect size
  \[ d = \frac{\bar{X}_1 - \bar{X}_2}{\hat{\sigma}_{pooled}} \]

- Small=.20, Medium=.50, Large=.80
Effect Size:
Chi-square Tests

- Population effect size
  \[ \phi = \sqrt{\sum_j \frac{(\pi_{aj} - \pi_{oj})^2}{\pi_{oj}}} \]

- Sample effect size
  \[ w = \sqrt{\frac{\chi^2}{N}} \]

- Small=.10, Medium=.30, Large=.50
Effect Size:
ANOVA and Regression

- **ANOVA**
  \[ \hat{f} = \sqrt{\frac{(k-1)F}{N}} \]

- **Regression (test of R^2)**
  \[ \hat{f}^2 = \frac{R^2}{1-R^2} = \frac{signal}{noise} \]

- **Regression (test of R^2 change)**
  \[ \hat{f}^2 = \frac{\Delta R^2}{1-R_L^2} = \frac{\Delta signal}{remaining\ noise} \]
Effect Size:
Correlation

- Pearson Product Moment Correlation is an effect size
- Commonly transformed to $z$ for aggregation and analyses

\[ z_r = .5 \log_e \left[ \frac{1 + r_{xy}}{1 - r_{xy}} \right] \]

- Small=.10, Medium=.30, Large=.50
Effect Size: Computing from Reported Statistics

Article Information:

\[ \bar{X}_1 = 12.58, \sigma_1 = 3.22, n_1 = 20 \quad \bar{X}_2 = 10.37, \sigma_2 = 2.92, n_2 = 24 \]

Knowing

\[ \hat{\sigma}_i = \sqrt{\frac{SS_i}{n_i - 1}} \quad \text{and} \quad SS_i = (n_i - 1) \hat{\sigma}_i^2 \]

\[ SS_1 = (19)(3.22^2) = 196.999, \quad SS_2 = (23)(2.92^2) = 196.107 \]

\[ \hat{\sigma}_{pooled} = \sqrt{\frac{SS_1 + SS_2}{n_1 + n_2 - 2}} = \sqrt{\frac{196.999 + 196.107}{20 + 24 - 2}} = 3.06 \]

\[ \text{and} \quad d = \frac{12.58 - 10.37}{3.06} = 0.72 \]
Effect Size: Computing from Reported Statistics

Article Information: $t(54) = 4.52, p < .05$

$$d = \frac{2t}{\sqrt{df}}$$

$$d = \frac{2(4.52)}{\sqrt{54}} = 1.23$$
Effect Size: Caveats

- Sensitivity to violations of assumptions
  - $VIN$
- Robust effect sizes
  - Improved estimation
  - Difficulty/impossible to compute without original data
- Multiple effect sizes from same sample or correlated samples
  - Special analyses
Describing Results: Graphical Displays

Funnel Plot Display for Example Data (k=10)
Describing Results: Graphical Displays

Forest Plot

- study3
- study2
- study1
- study4
- study5
- study6
- study8
- study7
- study9
- study10
- Total

Effect Size

-0.6 -0.4 -0.2 0.0 0.2 0.4 0.6 0.8 1.0
Describing Results: Graphical Displays

Stem and Leaf Plot

<table>
<thead>
<tr>
<th>Stem</th>
<th>Leaf</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>44</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>333</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0</td>
<td>111</td>
<td></td>
</tr>
<tr>
<td>-0</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>-0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Describing Results: Graphical Displays

Box Plot for Effect Sizes

Effect Size

Pub
No Moderator
Describing Results: Graphical Displays

Box Plot for Effect Sizes
All observed effect sizes from a single population
Testing Hypotheses

Observed effect sizes from two populations

\[ \delta \]

\[ d_1 \] \rightarrow \text{Females}

\[ d_2 \]

\[ d_3 \]

\[ d_4 \] \rightarrow \text{Males}

\[ d_5 \]

\[ d_6 \]
Testing Hypotheses:
Fixed Effects vs. Random Effects

![Graph 1: Population Effect Size vs. Frequency](image1)

![Graph 2: Population Effect Size vs. Frequency](image2)
Testing Hypotheses: Fixed Effects vs. Random Effects

- **Fixed Effects**
  - Assumes one population effect size
  - Effect size variance = sampling error (subjects)
    - Weights represent study variance due to sampling error associated with the subjects (sample size)
Random Effects

- Assumes population effect size is a normal distribution of values (i.e. not one effect size)
- Effect size variance = sampling error (subjects) + random effects (study)
- Weights represent study variance due to sampling error associated with the subjects (sample size) and sampling of studies (random effects variance component)
Fixed Effects vs. Random Effects: Which model to use?

Aspects to consider:

- Statistics – decision based on the outcome of the homogeneity of effect sizes statistic (conditionally random-effects)
- Desired Inferences – decision based on the inferences that the researcher would like to make
  - Conditional Inferences (fixed effect model): Researcher can only generalize to the studies included in the meta-analysis
  - Unconditional Inferences (random effect model): Researcher can generalize beyond the studies included in the meta-analysis
- Number of studies – when the number of studies is small fixed effects may be more appropriate
Testing Hypotheses: Estimation of Weights

- Fixed effects weight
  \[ v_i = \frac{1}{\sigma_i^2} \]

- Random effects weight
  \[ w_i = \frac{1}{\sigma_i^2 + \tau^2} \]

For standardized mean difference:
\[ \sigma_i^2 = \frac{n_1 + n_2}{n_1 n_2} + \frac{d_i^2}{2(n_1 + n_2)} \]
Testing Hypotheses:
Weighted Mean Effect Size

- Fixed effects
  \[ \bar{\delta} = \frac{\sum v_i d_i}{\sum v_i} \]
  \[ \text{var}(\bar{\delta}) = \frac{1}{\sum v_i} \]

- Random effects
  \[ \bar{\delta} = \frac{\sum w_i d_i}{\sum w_i} \]
  \[ \text{var}(\bar{\delta}) = \frac{1}{\sum w_i} \]
Also called Random Effects Variance Component (REVC), symbolized with $\tau^2$.

Used to calculate random effects weights.

Three methods to calculate:
- Observed variance
- Q based
- Maximum likelihood
Testing Hypotheses: Estimates of Effect Size Variance

- Observed variance: $\tilde{\sigma}_\theta^2 = s^2 - \frac{\sum \sigma_i^2}{k}$

- Q based: $\tilde{\sigma}_\theta^2 = \frac{Q - (k - 1)}{c}$ where $c = \sum v_i - \frac{\sum v_i^2}{\sum v_i}$

- Maximum likelihood:

$$l(y_i; \Delta, \tau^2) = (2\pi)^{-\frac{k}{2}} \prod_{i=1}^{k} \left( \frac{1}{\sigma_i^2 + \tau^2} \right)^{\frac{1}{2}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^{k} \frac{(y_i - \Delta)^2}{\sigma_i^2 + \tau^2} \right\}$$
Testing Hypotheses:
Significance Testing and Confidence Intervals (CI)

- Significance testing: 
  \[ Z = \frac{\bar{\delta} - \delta_0}{\sqrt{\text{var}(\bar{\delta})}} \]

- Confidence interval (95% CI):
  \[ \bar{\delta} \pm 1.96 \sqrt{\text{var}(\bar{\delta})} \]
Testing Hypotheses: Mean and Individual Effect Size Differences

- Focused test of between group differences

\[ Q_{BET} = \sum \left( \frac{1}{\text{var}(\bar{\delta}_j)} (\bar{\delta}_j - \bar{\delta}_\cdot) \right) \]

- General test of homogeneity of effect sizes

\[ Q = \sum \left( \frac{1}{\text{var}(d_i)} (d_i - \bar{\delta}_\cdot) \right) \]
Testing Hypotheses:
Meta-Analytic Regression Model

\[ d_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + e \]

- Generalization of the Q test
  - Continuous or categorical moderators
  - \( X_i \) are potential moderating variables

- Test \( \beta_j \) for moderating effect
Threats to Validity

- **Sources**
  - Primary studies – unreliability, restriction of range, violations of assumptions, missing effect sizes (publication bias), incompatible constructs, and poor quality
  - Meta-analysis processes – incomplete data collection (publication bias), inaccurate data collection, poor methodology, and inadequate power
Threats to Validity

- Apples and Oranges
- Dependent Effect Sizes
- File Drawer/Publication Bias
- Methodological Rigor
- Power
Threats to Validity

● Apples and Oranges
  ● Are the studies being analyzed similar regarding:
    ● Constructs examined
    ● Measures
    ● Participants (sampled from same population?)
    ● Analyses

● Dependent Effect Sizes
  ● Participants cannot contribute to the mean effect size more than once without special treatment (see Hedges et al., 2010; Owens et al., 2011)
Threats to Validity: Publication Bias

- **Publication Bias** = Studies unavailable to the meta-analyst due to lack of publication acceptance or submission (termed “file drawer problem” by Rosenthal, 1979)

- Pattern in the literature

<table>
<thead>
<tr>
<th></th>
<th>Effect Size</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Small</td>
<td>Large</td>
</tr>
<tr>
<td>Variance</td>
<td>Small (N=large)</td>
<td>Published (Stat Sig)</td>
<td>Published (Stat Sig)</td>
</tr>
<tr>
<td></td>
<td>Large (N=small)</td>
<td>Not Published (not Stat Sig)</td>
<td>Published (Stat Sig)</td>
</tr>
</tbody>
</table>
Threats to Validity: Publication Bias

- Publication Bias Detection Methods
  - Visual interpretation
    - Funnel plot display
  - Statistical methods
    - Begg Rank Correlation (variance or sample size)
    - Egger Regression
    - Funnel Plot Regression
    - Trim and Fill
Threats to Validity

- Methodological Rigor of Primary Studies
  - Set criteria for inclusion
  - Include various levels of rigor; then code and use in meta-analytic analyses (moderators or quality weights)

- Power
  - Enough studies collected to support the validity of hypothesis tests?
Reporting Meta-Analyses: Pertinent Information to Include

- Details regarding the search criteria and retrieval
- Coding process including rater reliability
- Describe effect sizes graphically
- Analyses
  - Mean effect size (significance test and / CI)
  - Fixed vs. Random Effects model
  - Homogeneity of effect sizes
  - Tests for moderators
- How threats to validity were addressed
For Further Reading & Thinking

For Further Reading & Thinking

For Further Reading & Thinking


For Further Reading & Thinking


For Further Reading & Thinking

Thank You

Now, let’s just talk...