An Introduction to Meta-analysis

Douglas G. Bonett

10/24/2018
Overview

• What is a “meta-analysis”?
• Some common measures of effect size
• Three kinds of meta-analysis statistical models
• Combining results from multiple studies
• Comparing results from multiple studies
• Assessing and adjusting for publication bias
• Converting from one effect-size measure to another
What is a “Meta-analysis”

A meta-analysis uses summary information (e.g., sample correlations, sample means, sample proportions) from \( m \geq 2 \) studies. In each study, a sample of participants is obtained from some study population and the sample data are used to estimate a population “effect size” parameter (e.g., mean difference, standardized mean difference, odds ratio, correlation, slope coefficients, etc.).

Statistical methods are then used to compare or combine the \( m \) effect size estimates.
Benefits of a Meta-analysis

A comparison of effect size estimates can assess interesting differences across the \( m \) study populations. If a statistical analysis indicates that the population effect sizes are meaningfully different across the study populations, certain characteristics of the study populations (length of treatment, age of participants, etc.) could be due to scientifically interesting moderator variables. The discovery of a moderator variable could clarify current theory and suggest new avenues of inquiry.

The \( m \) effect size estimates also can be combined to obtain a single, more precise, estimate of an effect size that generalizes to all \( m \) study populations.
Some Common Effect-size Measures

*Pearson correlation* \((\rho_{xy})\) – describes the strength of a linear relation between two quantitative variables \((x\ and\ y)\)

*Point-biserial correlation* \((\rho_{pb})\) – describes the strength of a relation between a quantitative variable and a dichotomous variable

*Standardized mean difference* \((\delta)\) – describes the mean difference between treatment conditions (or subpopulations) in standard deviation units

*Odds ratio* \((\omega)\) – describes the strength of relation between two dichotomous variables
Meta-analysis Statistical Models

Let $\theta_k$ present some measure of effect size (correlation, standardized mean difference, odds ratio, etc.) for study $k$ ($k = 1$ to $m$). The estimator of $\theta_k$ and its estimated variance will be denoted as $\hat{\theta}_k$ and $\text{var}(\hat{\theta}_k)$, respectively. $\hat{\theta}_k$ is a random variable and a statistical model may be used to represent its large-sample expected value and random deviation from expectation.

Three basic types of statistical models – the constant coefficient model, the varying coefficient model, and the random coefficient model – may be used to represent the estimators from multi-study designs.
The constant coefficient model may be expressed as

\[ \hat{\theta}_k = \theta + \varepsilon_k \]

where \( \theta \) is the large-sample expected value of each \( \hat{\theta}_k \) and \( \varepsilon_k \) is a random parameter prediction error. The constant coefficient model assumes that every \( \hat{\theta}_k \) (\( k = 1 \) to \( m \)) is an estimator of the same quantity (\( \theta \)). This assumption is referred to as the effect-size homogeneity assumption. The constant coefficient model is a type of fixed-effect model because \( \theta \) is an unknown constant.
The varying coefficient model may be expressed as

$$\hat{\theta}_k = \theta + \nu_k + \varepsilon_k$$

where $\theta = m^{-1} \sum_{k=1}^{m} \theta_k$, and $\nu_k$ represents effect-size heterogeneity due to all unknown or unspecified differences in the characteristics of the $m$ study populations. The varying coefficient model is also a fixed-effects model because $\theta$ and each $\nu_k$ are unknown constants.
The random coefficient model may be expressed as

\[ \hat{\theta}_k = \theta^* + \nu_k + \epsilon_k \]

where \( \nu_k \) is assumed to be a normally distributed random variable with mean 0 and standard deviation \( \tau \). To justify the claim that \( \nu_k \) is a random variable, two-stage cluster sampling can be assumed. One way to conceptualize two-stage cluster sampling is to randomly select \( m \) study populations from a superpopulation of \( M \) study populations and then take a random sample of size \( n_k \) from each of the \( m \) randomly selected study populations.
Meta-analysis Statistical Models (continued)

All three models can be used to combine results from $m$ studies to estimate a single effect size parameter.

constant coefficient (CC) model: $\theta$

varying coefficient (VC) model: $\theta = \sum_{k=1}^{m} \frac{\theta_k}{m}$

random coefficient (RC) model: $\theta^* = \sum_{k=1}^{M} \frac{\theta_k}{M}$
Limitations of Each Model

In the CC model, the estimate of $\theta$ is biased and inconsistent unless $\theta_1 = \theta_2 = \ldots = \theta_m$.

In the CC and VC models, the single effect size describes the $m$ study populations while $\theta^*$ in the RC model describes the average of all $M$ effect sizes in the superpopulation.

In the RC model, the estimate of $\theta^*$ is biased and inconsistent if the variances of $\hat{\theta}_k$ are correlated with $\hat{\theta}_k$ (which is a common situation).
The RC model is difficult to justify unless the \( m \) studies are assumed to be a random sample from some definable superpopulation.

A confidence interval for the standard deviation of the superpopulation effect sizes (\( \tau \)) in the RC model is hypersensitive to minor violations of the superpopulation normality assumption.

The confidence interval for \( \sum_{k=1}^{m} \theta_k / m \) (VC model) is wider than the confidence interval for \( \theta \) (CC model).

**Recommendation:** The VC model should be used in most cases unless the assumptions of the CC or RC models can be satisfied.
General Confidence Interval for $\theta = \sum_{k=1}^{m} \theta_k / m$

An estimate of $\theta = \sum_{k=1}^{m} \theta_k / m$ is

$$\hat{\theta} = \sum_{k=1}^{m} \hat{\theta}_k / m$$

and the estimated variance of $\hat{\theta}$ is

$$\text{var}(\hat{\theta}) = \sum_{k=1}^{m} \text{var}(\hat{\theta}_k) / m^2$$

and an approximate $100(1 - \alpha)\%$ confidence interval for $\theta$ is

$$\hat{\theta} \pm z_{\alpha/2} \sqrt{\text{var}(\hat{\theta})}.$$
Example: Correlations

The Pearson correlations between Beck Depression Inventory (BDI) scores and drug use was reported in $m = 4$ different studies. Hypothetical data are given below.

<table>
<thead>
<tr>
<th>Study</th>
<th>$\hat{\rho}_{xy}$</th>
<th>$n$</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.40</td>
<td>55</td>
<td>[.15, .60]</td>
</tr>
<tr>
<td>2</td>
<td>.65</td>
<td>90</td>
<td>[.51, .76]</td>
</tr>
<tr>
<td>3</td>
<td>.60</td>
<td>65</td>
<td>[.42, .74]</td>
</tr>
<tr>
<td>4</td>
<td>.45</td>
<td>35</td>
<td>[.14, .68]</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>.53</td>
<td>[.41, .62]</td>
</tr>
</tbody>
</table>

Note that the CI for the average correlation is narrower than any of the single-study CIs. Furthermore, the CI for the average correlation describes all four study populations.
Example: Odds Ratios

Three studies each used a two-group nonexperimental design to compare adults diagnosed with traumatic brain injury (TBI) with a matched control group that did not have TBA. In each study, the participants were asked if they had experienced suicidal thoughts (ST) in the past 30 days (yes or no). Hypothetical data are given below.

<table>
<thead>
<tr>
<th>Study</th>
<th>TBI ST</th>
<th>TBI No ST</th>
<th>No TBI ST</th>
<th>No TBI No ST</th>
<th>(\hat{\omega})</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>41</td>
<td>59</td>
<td>22</td>
<td>78</td>
<td>2.43</td>
<td>[1.32, 4.50]</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>4</td>
<td>8</td>
<td>15</td>
<td>3.85</td>
<td>[0.96, 15.6]</td>
</tr>
<tr>
<td>3</td>
<td>13</td>
<td>5</td>
<td>22</td>
<td>30</td>
<td>3.33</td>
<td>[1.07, 10.3]</td>
</tr>
</tbody>
</table>

Average (geometric) 3.14 [1.67, 5.93]
Example: Standardized Mean Differences

Three studies used a two-group design to compare the self-confidence of employees who reported having an abusive parent with employees who did not report having an abusive parent. Hypothetical results are given below.

<table>
<thead>
<tr>
<th>Study</th>
<th>( \delta )</th>
<th>SE</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.64</td>
<td>0.263</td>
<td>[0.12, 1.16]</td>
</tr>
<tr>
<td>2</td>
<td>0.57</td>
<td>0.306</td>
<td>[-0.03, 1.17]</td>
</tr>
<tr>
<td>3</td>
<td>0.43</td>
<td>0.329</td>
<td>[-0.21, 1.07]</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>0.55</td>
<td>0.174</td>
</tr>
</tbody>
</table>

We can be 95% confident that the population mean self-confidence score averaged over the three study populations is 0.21 to 0.89 standard deviations greater for employees who had an abusive parent than employees who did not have an abusive parent.
Assessing Moderator Effects

A variable that influences the value of an effect-size parameter is called a *moderator variable*. Moderator variables can be assessed using a linear statistical model. The $m$ estimators, $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_m$ may be represented by the following VC linear model

$$\hat{\theta} = X\beta + Z\xi + \epsilon$$

where $X$ is a design matrix that codes known quantitative or qualitative characteristics of the $m$ study populations, $Z$ is a matrix of unknown or unspecified predictor variables.
Estimate of $\beta$

An ordinary least squares estimator of $\beta$ is

$$\hat{\beta} = (X'X)^{-1}X'\hat{\theta}$$

with an estimated covariance matrix

$$\text{cov}(\hat{\beta}) = (X'X)^{-1}X'\hat{V}X(X'X)^{-1}$$

where $\hat{V}$ is a diagonal matrix with $\text{var}(\hat{\theta}_k)$ in the $k^{th}$ diagonal element.
Confidence Interval for $\beta$

The estimated variance of $\hat{\beta}_t$ is the $t^{th}$ diagonal element of $\text{cov}(\hat{\beta})$ which will be denoted as $\text{var}(\hat{\beta}_t)$.

An approximate $100(1 - \alpha)\%$ confidence interval for $\beta_t$ is

$$
\hat{\beta}_t \pm z_{\alpha/2} \sqrt{\text{var}(\hat{\beta}_t)} .
$$
Five studies examined the relation between job performance and amount of college education (bachelor vs masters). Average employee tenure in each study varied from 2.1 years to 15.8 years across the five studies. The researcher believes that the relation between job performance and the amount of college education is moderated by tenure. Hypothetical results are given below.

<table>
<thead>
<tr>
<th>Study</th>
<th>Ave Tenure</th>
<th>$\hat{\delta}$</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.1</td>
<td>0.68</td>
<td>0.142</td>
</tr>
<tr>
<td>2</td>
<td>4.7</td>
<td>0.51</td>
<td>0.093</td>
</tr>
<tr>
<td>3</td>
<td>9.2</td>
<td>0.34</td>
<td>0.159</td>
</tr>
<tr>
<td>4</td>
<td>10.7</td>
<td>0.20</td>
<td>0.078</td>
</tr>
<tr>
<td>5</td>
<td>15.8</td>
<td>0.15</td>
<td>0.201</td>
</tr>
</tbody>
</table>

$\hat{\beta}_1 = .123$  95% CI = [0.011, 0.018]

We are 95% confident that each additional year is associated with a 0.011 to 0.018 reduction in $\delta$. 

---

Example – Linear Model
Publication Bias

Effect size estimates from studies with small sample sizes are more likely to be the result of large anomalous sampling errors that have exaggerated the effect size and small $p$-values, and a meta-analysis that includes these studies will produce biased results. This type of bias is often referred to as publication bias.

To implement the proposed approach, fit a linear model with an additional predictor variable $(x_t)$ that is equal to $\sqrt{n_k^*} - \min \sqrt{n_k^*}$ where $n_k^* = 1/n_k$ for 1-group designs and $n_k^* = 1/n_{1k} + 1/n_{2k}$ for 2-group designs in study $k$. 
Example: Publication Bias

Six different studies examined the correlation between alcohol use and age among 21 to 35 year old adults. The six studies sampled from similar study populations and there were no obvious moderator variables to include in the analysis. The sample correlations and sample sizes are given below.

<table>
<thead>
<tr>
<th>Study</th>
<th>n</th>
<th>$\hat{\rho}_{xy}$</th>
<th>$\sqrt{1/n_k} - \min(\sqrt{1/n_k})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>52</td>
<td>-.591</td>
<td>0.077</td>
</tr>
<tr>
<td>2</td>
<td>98</td>
<td>-.383</td>
<td>0.039</td>
</tr>
<tr>
<td>3</td>
<td>76</td>
<td>-.399</td>
<td>0.053</td>
</tr>
<tr>
<td>4</td>
<td>159</td>
<td>-.347</td>
<td>0.017</td>
</tr>
<tr>
<td>5</td>
<td>258</td>
<td>-.308</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>60</td>
<td>-.477</td>
<td>0.067</td>
</tr>
</tbody>
</table>
The 95% confidence interval for $\beta_1$ is [-0.005, 0.003]. The confidence interval includes 0 and is sufficiently narrow to conclude that $\beta_1$ is close to 0. The predicted correlation at the largest sample size is -0.41 with a 95% confidence interval of [-0.49, -0.33].

In comparison, the estimated average correlation is -0.42 with a 95% confidence interval of [-0.48, -0.35]. The results for the predicted correlation at the largest sample size are essentially the same as the results for the average correlation which suggests that the publication bias is not serious.
Linear Contrasts

A linear contrast can be expressed as $c_1 \theta_1 + c_2 \theta_2 + \cdots + c_m \theta_m$ where $c_k$ is called a contrast coefficient. A linear contrast can be expressed more concisely using summation notation as $\sum_{k=1}^{m} c_k \theta_k$.

For instance, suppose two studies used military samples and three studies used civilian samples. To assess the possible moderating effect of military vs. civilian, we could estimate

$$(\theta_1 + \theta_2)/2 - (\theta_3 + \theta_4 + \theta_5)/3$$

Linear contrasts are especially useful when comparing group means or proportions from two or more studies.
Confidence Interval for $\sum_{k=1}^{m} c_k \theta_k$

An estimate of $\sum_{k=1}^{m} c_k \theta_k$ is

$$\sum_{k=1}^{m} c_k \hat{\theta}_k$$

and the estimated variance of $\sum_{k=1}^{m} c_k \hat{\theta}_k$ is

$$\sum_{k=1}^{m} c_k^2 \text{var}(\hat{\theta}_k)$$

and an approximate $100(1 - \alpha)\%$ confidence interval for $\sum_{k=1}^{m} c_k \theta_k$ is

$$\sum_{k=1}^{m} c_k \hat{\theta}_k \pm Z_{\alpha/2} \sqrt{\sum_{k=1}^{m} c_k^2 \text{var}(\hat{\theta}_k)}.$$
Suppose three studies examined the effectiveness of a resilience training program. Study 1 measured resilience at pretreatment and 1-month post-treatment. Study 2 measured resilience at pretreatment and 2-month post-treatment. Study 3 measured resilience at 1-month, 2-months and 6-month post-treatment. The population means that are estimated in each study are summarized below.

<table>
<thead>
<tr>
<th>Study</th>
<th>Pretest</th>
<th>1-month</th>
<th>2-month</th>
<th>6-month</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\mu_1$</td>
<td>$\mu_2$</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>2</td>
<td>$\mu_3$</td>
<td>$\mu_4$</td>
<td>$\mu_5$</td>
<td>---</td>
</tr>
<tr>
<td>3</td>
<td>---</td>
<td>$\mu_6$</td>
<td>$\mu_7$</td>
<td>$\mu_8$</td>
</tr>
</tbody>
</table>

Some interesting linear contrasts:

\[
\frac{\mu_1 + \mu_3}{2} - \frac{\mu_2 + \mu_4 + \mu_6}{3}
\]

\[
\frac{\mu_2 + \mu_4 + \mu_6}{3} - \frac{\mu_5 + \mu_7}{2}
\]

\[
\frac{\mu_5 + \mu_7}{2} - \mu_8
\]

\[
(\mu_1 - \mu_2) - (\mu_3 - \mu_4)
\]

\[
(\mu_4 - \mu_5) - (\mu_6 - \mu_7)
\]
Effect-size Conversions

Meta-analysis results for one type of effect size can be converted into another effect size.

\[ \rho_{xy} \rightarrow \omega \rightarrow \delta \rightarrow \rho_{pb} \]

black arrow = exact conversion
blue arrow = approximate conversion
Some Effect-size Conversion Formulas

\( \rho_{pb} = \frac{\delta}{\sqrt{\delta^2 + \frac{1}{pq}}} \)  \( \delta = \frac{\rho_{pb} \sqrt{1/pq}}{\sqrt{1 - \rho_{pb}^2}} \)

\( \delta \approx \frac{\ln(\omega)}{1.7} \)

\( \omega \approx \exp[1.7(\delta)] \)

\( \rho_{xy} \approx \frac{\omega^{3/4} - 1}{\omega^{3/4} + 1} \)

\( \omega \approx \left[\frac{1 + \rho_{xy}}{1 - \rho_{xy}}\right]^{4/3} \)

\( \rho_{pb} \approx \frac{\ln(\omega)}{\sqrt{\ln(\omega)^2 + \frac{2.89}{pq}}} \)

\( \omega \approx \exp\left[\frac{\rho_{pb} \sqrt{\frac{2.89}{pq}}}{\sqrt{1 - \rho_{pb}^2}}\right] \)

where \( p = n_1 / (n_1 + n_2) \) and \( q = 1 - p \)

\( \rho_{xy} \approx (\rho_{pb} / h) \sqrt{pq} \)

\( \rho_{pb} \approx \rho_{xy}(h) / \sqrt{pq} \)

where \( h = \exp(-z^2 / 2) / 2.51 \) and \( z \) is the point on the standard normal curve that is exceeded with probability \( p \).
Some Resources

These slides are posted on the CSASS web site:

https://csass.ucsc.edu/seminars/index.html

Meta-analysis R functions and a meta-analysis eBook can be retrieved from:

https://people.ucsc.edu/~dgbonett/meta.html
Some References

Suggested reference texts


References for varying coefficient methods


Questions or comments?