A Hands on Introduction to Systematic Reviews and Meta-Analysis

Part 2: Meta-Analysis (Summarizing the Data)

Effect Size

- An *effect size* quantifies the magnitude of the relationship among variables
  - For example, let's say we are comparing 10 boys and 10 girls on reading speed
    - Boys: M = 36, SD = 7
    - Girls: M = 29, SD = 6
- Unstandardized Effect Size
  - $M_{diff} = M_{Boys} - M_{girls} = 36 - 29 = 7$
  - This is interpretable if the units are interpretable (e.g., if reading speed was measured in seconds or minutes)

Effect Size

- Standardized Effect Size
  - E.g., Cohen's *d*
    - Quantifies differences in means in 'standard deviation' units
  - $d = \frac{M_{Boys} - M_{Girls}}{\sqrt{\frac{(n_{Boys}-1)SD^2_{Boys} + (n_{Girls}-1)SD^2_{Girls}}{n_{Boys}+n_{Girls}-2}}} = \frac{36-29}{\sqrt{\frac{(10-1)7^2+(10-1)6^2}{10+10-2}}} = 1.07$
  - Thus, boys and girls differ by a little more than one SD
  - This is interpretable regardless of the units of measurement, and is comparable across studies which use different scales, measures, etc.
    - Which will obviously be useful in meta-analysis

Effect Size

- Example 2: Correlation between income and depression
- $r = .24$
  - Increasing income by one standard deviation is, on average, associated with a .24 increase in depression
- Since correlation values are inherently standardized (range from −1 to 1), we would almost always adopt a standardized metric to explore the correlation among variables

Confidence Intervals for Effect Sizes

- Confidence Interval (CI)
  - A range of values over which we expect the true (population) parameter to fall
  - E.g., 95% CI
    - If we sampled repeatedly from the population and calculated a CI for each effect size from each sample, 95% of the CIs would contain the population parameter
  - Importance of Confidence Intervals
    - CIs provide information regarding measurement precision
      - E.g. 1: $d = .28$; 95% CI = {.27, .29}
      - E.g. 2: $d = .28$; 95% CI = {.15, .41}
    - The second CI measures the effect of interest much less precisely, even though the effect size is the same
  - In general, studies with larger sample sizes have narrower CIs

Effect Size in Meta-Analysis

- Effect size is the Outcome/Dependent Variable
  - This will require the computation of effect sizes or transforming from one effect size to another
  - Standardized effect sizes are almost always used in meta-analysis
    - A standardized index must be comparable across studies, represent the magnitude and direction of the relationship of interest, and be independent of sample size
  - It is also possible to use unstandardized effect sizes, but this requires that the exact same scales/variables are used in each study (and that no transformations, modifications, etc. were made to any variables)

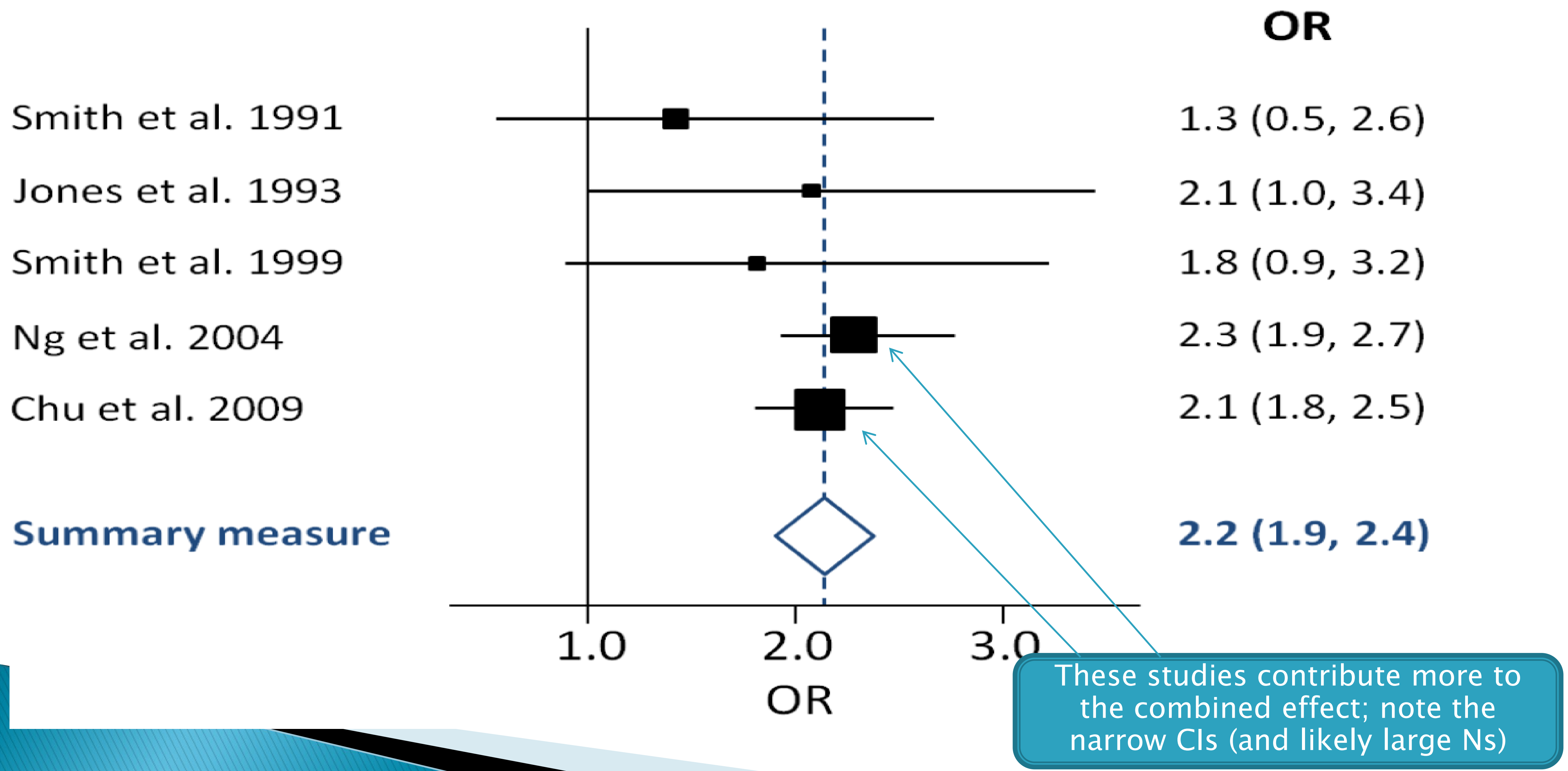
Effect Size in Meta-Analysis

- Note that in some testing situations it might be tricky to obtain a proper effect size estimate
  - Take, for example, a study that looks at the difference between boys and girls in vocabulary development over 6 months from 18 months to 24 months using a repeated measures analysis
- Typical methods that convert *t/z* statistics to *d* will be incorrect for repeated measures studies, and corrections need to be applied in order to minimize bias

Forest Plot

- A visual representation of effect sizes (and confidence intervals for the effect sizes) from multiple studies included in a meta-analysis
  - Recall: all effects must be measured in the same metric (e.g., *d*, correlation)
- The area of the effect size icons (usually squares) on the plot indicates the “weight” of the study to the combined effect
  - E.g., larger N studies have a higher weight
- The plot also shows the *combined* effect size, and confidence interval for the effect size, across studies

Forest Plot Example – Odds Ratios



Statistical Models

- There are two popular models available for conducting a meta-analysis
    - In other words, two models available for arriving at a “combined” measure of effect size

- Fixed Effects Model
    - Assumes that all the studies investigated the same population, and therefore estimate the same population effect size
        - Highly questionable

- Random Effects Model
    - Allows for the possibility that the studies investigated somewhat different populations, and therefore estimate different population effect sizes
        - Another way to say this is that we expect some “true” variability in effect sizes

Fixed Effects vs Random Effects

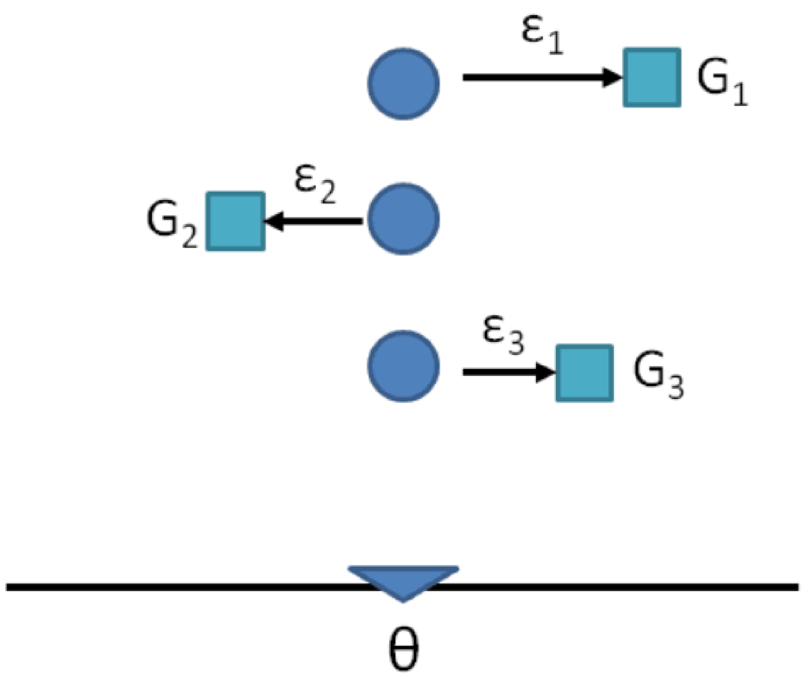
- It is difficult to imagine a setting in which multiple studies conducted in different locations, with different samples, and with potentially different measures are all studying the same population (and thus are after a single population effect size)

- The random effects model is more realistic and provides a basis for understanding the heterogeneity of effect sizes
  - Further, the models give the same answer if there is only a single population, so it is hard to find a reason for a researcher to prefer a fixed effects model

Fixed Effects vs Random Effects

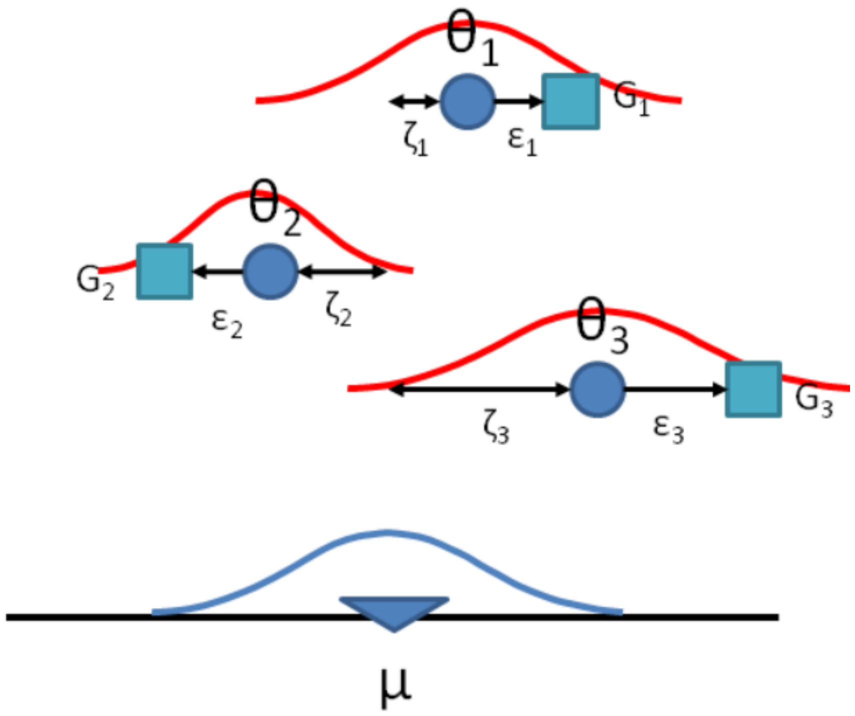
Fixed Effects Model

$$G_i = \theta + \varepsilon_i$$



Random Effects Model

$$G_i = \mu + \zeta_i + \varepsilon_i$$



Fixed Effects Meta-Analysis

- For a set of *S* effect size measures (γ)
  - $\hat{\Upsilon}_F = \frac{\sum_{i=1}^{S} w_i \hat{\Upsilon}_i}{\sum_{i=i}^{S} w_i}$
  - $w_i = \frac{1}{s^2(\hat{\Upsilon}_i)}$
  - $s^2(\hat{\Upsilon}_F) = \frac{1}{\sum_{i=i}^{S} w_i}$

This info is used to generate a mean effect size and a CI around the mean effect size

Fixed Effects Meta-Analysis Example

- Study 1: $M_1 = 12$, $M_2 = 14$, $SD_1 = 3$, $SD_2 = 3$, $n_1 = 22$, $n_2 = 32$
- Study 2: $M_1 = 14$, $M_2 = 16$, $SD_1 = 2$, $SD_2 = 2$, $n_1 = 25$, $n_2 = 52$
- Study 3: $M_1 = 11$, $M_2 = 13$, $SD_1 = 4$, $SD_2 = 4$, $n_1 = 142$, $n_2 = 128$

- Cohen's *d* Values
  - $d = \frac{M_1 - M_2}{\sqrt{\frac{(n_1-1)SD_1^2+(n_2-1)SD_S^2}{n_1+n_2-2}}}$
  - $d_1 = \frac{M_1 - M_2}{\sqrt{\frac{(n_1-1)SD_1^2+(n_2-1)SD_S^2}{n_1+n_2-2}}} = \frac{12-14}{\sqrt{\frac{(22-1)3^2+(32-1)3^2}{22+32-2}}} = -.67$
  - $d_2 = \frac{M_1 - M_2}{\sqrt{\frac{(n_1-1)SD_1^2+(n_2-1)SD_S^2}{n_1+n_2-2}}} = \frac{14-16}{\sqrt{\frac{(25-1)2^2+(52-1)2^2}{25+52-2}}} = -1.00$
  - $d_3 = \frac{M_1 - M_2}{\sqrt{\frac{(n_1-1)SD_1^2+(n_2-1)SD_S^2}{n_1+n_2-2}}} = \frac{11-13}{\sqrt{\frac{(142-1)4^2+(128-1)4^2}{142+128-2}}} = -.50$

Fixed Effects Meta-Analysis Example

- Study 1: $M_1 = 12$, $M_2 = 14$, $SD_1 = 3$, $SD_2 = 3$, $n_1 = 22$, $n_2 = 32$
- Study 2: $M_1 = 14$, $M_2 = 16$, $SD_1 = 2$, $SD_2 = 2$, $n_1 = 25$, $n_2 = 52$
- Study 3: $M_1 = 11$, $M_2 = 13$, $SD_1 = 4$, $SD_2 = 4$, $n_1 = 142$, $n_2 = 128$

- Variances of the *d* values
  - $s^2(d) = \frac{n_1+n_2}{n_1 n_2} + \frac{d^2}{2(n_1+n_2-2)}$
  - $s^2(d_1) = \frac{n_1+n_2}{n_1 n_2} + \frac{d^2}{2(n_1+n_2-2)} = \frac{22+32}{(22)(32)} + \frac{-.67^2}{2(22+32-2)} = .085$
  - $s^2(d_2) = \frac{n_1+n_2}{n_1 n_2} + \frac{d^2}{2(n_1+n_2-2)} = \frac{25+52}{(25)(52)} + \frac{-1.00^2}{2(25+52-2)} = .073$
  - $s^2(d_3) = \frac{n_1+n_2}{n_1 n_2} + \frac{d^2}{2(n_1+n_2-2)} = \frac{142+128}{(142)(128)} + \frac{-.50^2}{2(142+128-2)} = .016$

Notice that the study with the smallest variance for its associated effect size has the largest N

Fixed Effects Meta-Analysis Example

- Study 1: $M_1 = 12$, $M_2 = 14$, $SD_1 = 3$, $SD_2 = 3$, $n_1 = 22$, $n_2 = 32$
- Study 2: $M_1 = 14$, $M_2 = 16$, $SD_1 = 2$, $SD_2 = 2$, $n_1 = 25$, $n_2 = 52$
- Study 3: $M_1 = 11$, $M_2 = 13$, $SD_1 = 4$, $SD_2 = 4$, $n_1 = 142$, $n_2 = 128$

- Weights
  - $w = \frac{1}{s^2(d)}$
  - $w_1 = \frac{1}{s^2(d)} = \frac{1}{.085} = 11.73$
  - $w_2 = \frac{1}{s^2(d)} = \frac{1}{.073} = 13.78$
  - $w_3 = \frac{1}{s^2(d)} = \frac{1}{.016} = 63.34$

Notice that the study with the largest weight for its associated effect size has the smallest variance/largest N

Fixed Effects Meta-Analysis Example

- $\hat{\Upsilon}_F = \frac{\sum_{i=1}^{S} w_i \hat{\Upsilon}_i}{\sum_{i=i}^{S} w_i} = \frac{(11.73)(-.67)+(13.78)(-1.00)+(63.34)(-.5)}{11.73+13.78+63.34} = -.60$

- $s^2(\hat{\Upsilon}_F) = \frac{1}{\sum_{i=i}^{S} w_i} = \frac{1}{11.73+13.78+63.34} = .011$

- $\text{SE}(\hat{\Upsilon}_F) = \sqrt{s^2(\hat{\Upsilon}_F)} = \sqrt{.011} = .10$

- $95\%\text{CI}(\hat{\Upsilon}_F) = \hat{\Upsilon}_F \pm (1.96)\text{SE}(\hat{\Upsilon}_F) =$
  $\{(-.60 - 1.96 * .10), (-.60 + 1.96 * .10)\} = \{-.80, -.40\}$

Fixed Effects Meta-Analysis Example

- Note: You are not going to be doing any of these "hand calculations" yourself
  - All of the calculations will be done using software

- The example was simply to provide you with an idea of how the process of meta-analysis is carried out "behind-the-scenes"

Random Effects Meta-Analysis

- For a set of S effect size measures (γ)
  - $\hat{\Upsilon}_R = \frac{\sum_{i=1}^{S} w_i^* \hat{\Upsilon}_i}{\sum_{i=i}^{S} w_i^*}$
  - $w_i^* = \frac{1}{s^2(\hat{\Upsilon}_i)+\tau^2}$
  - $\tau^2 = \frac{Q-(S-1)}{\sum_{i=i}^{S} w_i - \frac{\sum_{i=i}^{S} w_i^2}{\sum_{i=i}^{S} w_i}}$ for Q > S-1
  - $Q = \sum_{i=i}^{S} w_i\,(\hat{\Upsilon}_i - \hat{\Upsilon}_F)^2$
  - $s^2(\hat{\Upsilon}_R) = \frac{1}{\sum_{i=i}^{S} w_i^*}$

Weights are more similar across studies given the addition of the constant τ^2

Heterogeneity of Effect Sizes

- A simple goodness-of-fit test can be used to test for excessive heterogeneity
  - Q $\sim \chi^2_{df=S-1}$
    - We computed Q on the previous slide
    - We reject the null that there is no population heterogeneity if Q $\geq \chi^2_{\alpha,df=S-1}$
- The problem with this approach is that the test has low-power when S is small

Heterogeneity of Effect Sizes

- A better approach to quantifying heterogeneity is to use an effect size measure

$$I^2 = \frac{Q - S + 1}{Q}$$

- Tells us what proportion of the observed variance in effect sizes is due to true differences in effect sizes, rather than sampling error
- I^2 ranges from 0 to 1, with larger values indicating more heterogeneity

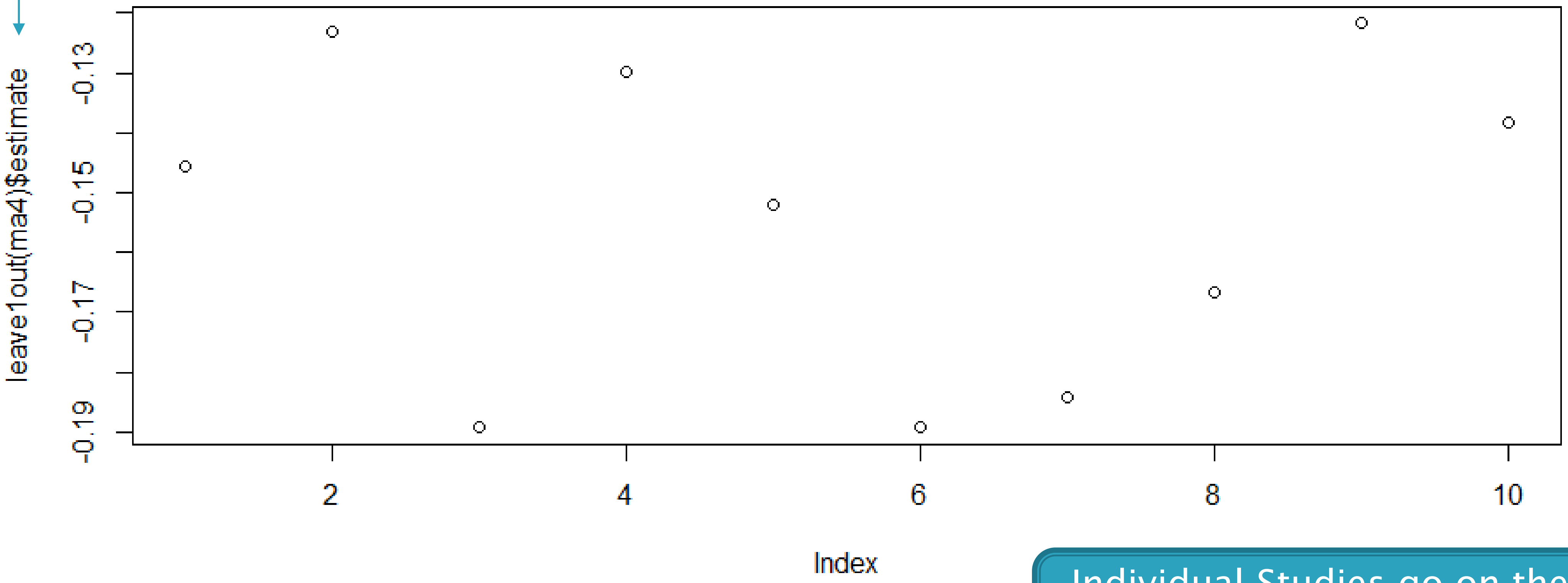
Outliers (Influential Cases)

- There are different ways of assessing the effect of outliers, but the main issue relates to what effect each study has on the combined effect size

- The easiest way to observe the effect of outliers is through “leave-one-out” analyses
  - Cook’s distance
    - A measure of the influence of individual cases on the combined effect
      - Popular cutoff is 4/(Number of Studies)
  - Plot the combined effect, as a function of which study is left out

Plot of "Leave-One-Out" Analyses

Combined Effect Size

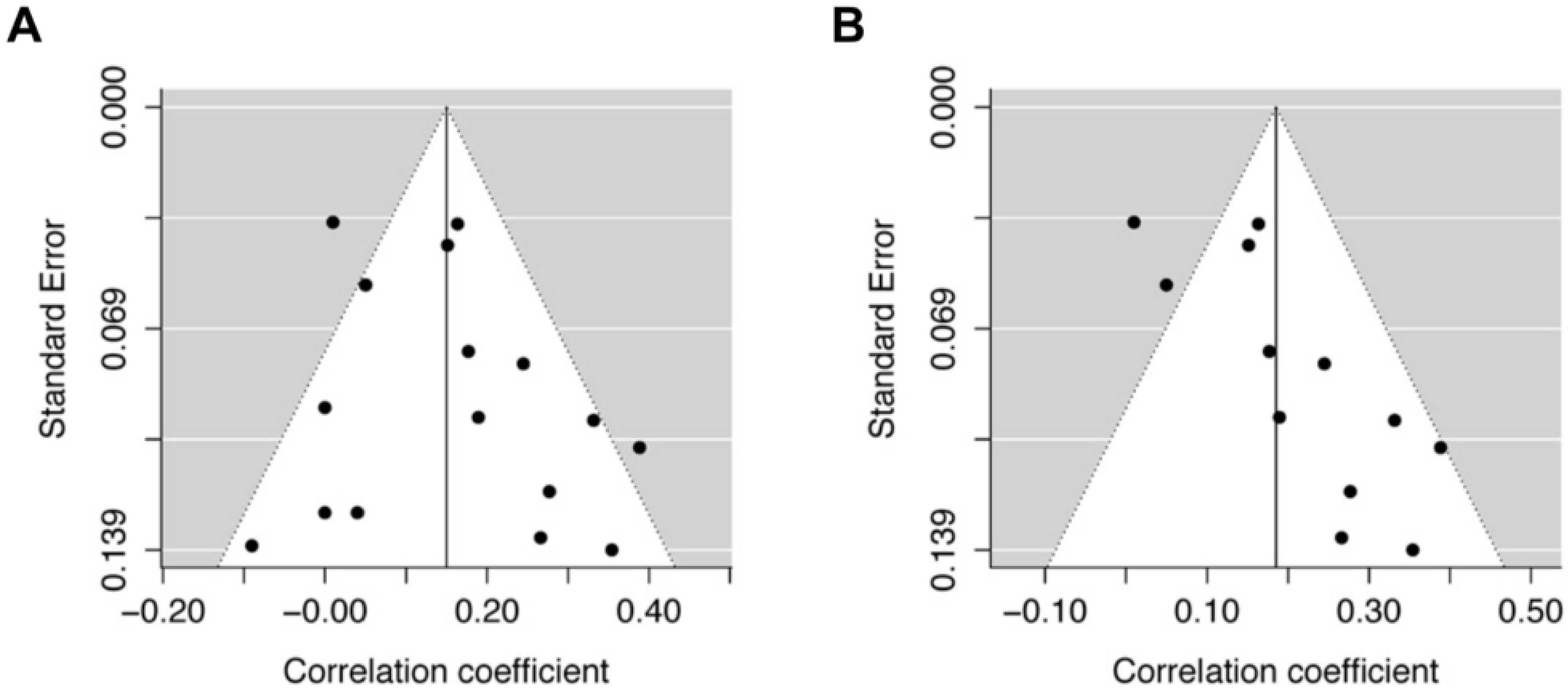


Individual Studies go on the X-axis, here studies 1-10

Publication Bias

- When there is publication bias (e.g., studies with statistically significant effects getting published), studies with small sample sizes tend to have large effects
  - A large effect size is needed for an effect based on a small N to be statistically significant
- If we plot effect sizes against sample size/standard error, publication bias would show up in terms of “asymmetry”
  - Small N studies would all tend to have large effects
- This plot is called a *funnel plot*

Funnel Plot: Which Shows Evidence of Publication Bias?



Conclusion: Why Meta-Analysis?

- Focuses on effect sizes, not statistical significance
- Combines multiple studies for a more precise estimate of the effect size
- Provides a rationale for small-N research
  - I.e., the results will be combined with other studies for a more precise estimate of the effect size