

Paper 3: Contextualizing effect sizes by examining population, study, and outcome characteristics: Extended investigation using multilevel models and mixtures

Hendricks Brown, University of South Florida

This presentation examines the variation in impact with effect sizes by describing traditional meta-analyses and what they intend to discover; examining impact with multi-level models and mixtures; and outlining the inherent limits of meta-analyses and reporting the findings.

Traditional Meta-Analyses

There are several traditional questions that can be addressed by effect size modeling in meta-analysis: (a) is there an overall protective or harmful effect across all studies; (b) does effect vary by population characteristics, intervention, or study design characteristics; and (c) how much variation is explained by these factors?

There are several requirements for determining the overall mean. First, the outcome measure must be polarized with “good” and “bad” ends. Researchers often try to develop measures on a continuum, but do not always succeed. Second, outcome measures need to be placed on the same scale in order to compare them. Effect sizes are calculated for continuous outcomes, but log odds ratios are often used for dichotomous variable. Finally, the analytical method and results need to be summarized.

Predictors of Effect Size from Trials

One way to analyze data with traditional analyses is using regression modeling at the level of a trial. This equation is expressed as Program Effect Size = Population Characteristics + Intervention Characteristics + Study Design Characteristics + Error. Each of these components is described below.

Population characteristics are typically at the level of the study, not within the study. A meta-analysis that illustrates this is “Bereavement Intervention Meta-analysis Ignores Differential Impact by Gender” (Kato & Mann, 1999). Two out of 11 studies reported results separately by gender or adjusted for gender. In another analysis, the impact was found to be lower when more distressed children were excluded and the impact was lower for those where on average parent death occurred longer ago” (Currier, Holland, & Neimeyer, 2007).

Intervention characteristics determine if the intervention varies in their effects. The following examples are traditional effect size modeling in meta-analyses where typically each study ends with one outcome, scaled as an effect size. Durlak and Wells (1997) reviewed 177 mental health prevention programs and reported a single effect size. In 2000, Tobler and Roona reviewed 207 drug prevention program trials and divided the outcomes into follow-up intervals. They also averaged all effect sizes within follow-up intervals, whereas most analyses only use the first follow-up period. They concluded that interactive school drug prevention programs were more effective than didactic ones. Brown, et al., (2000) reviewed 214 psycho-social-education prevention programs for children age 0 - 6. The authors identified 1, 451 outcomes or an average of 7 outcomes per trial.

Trial design characteristics can impact the effect size. Wilson, et al., (2003) concluded that randomized trials for aggressive behavior had systematically greater impact (0.34) than nonrandomized (0.13).

Finally, error, or variability, comes from two sources: (a) standard error (SE) for each effect size calculated and included in the analyses and (b) heterogeneity of effect size's beyond these SE's which are assumed to be normally distributed and estimated from a mixed effects model.

Multi-level Modeling

Multi-level or mixed models can explore the effect of multiple outcomes within a trial to determine if the impact is general or long-lasting. In addition, researchers can investigate heterogeneity with mixtures modeling to examine if further variations are detectable.

A Random Slope Model to incorporate known standard errors is expressed as:

$$ES_{\text{Trial}} = \text{FIXED Trial EFFECTS} + \beta_{\text{Trial}} SE_{\text{Trial}} + \epsilon_{\text{Trial}}$$

A Random Slope Model to incorporate multiple outcomes is expressed as:

$$ES_{\text{Trial, Outcome}} = \text{FIXED Trial EFFECTS} + \text{FIXED Outcome EFFECTS} + \beta_{\text{Trial}} SE_{\text{Trial}} + \epsilon_{\text{Trial}} + \epsilon_{\text{Outcome}}$$

Formal modeling of effect sizes in multi-level modeling can be completed with Mplus 4.2 Models. The software calculates two-level analyses (trial, outcome); random slopes (internal standard error); regression of effect size on trial and outcome characteristics; and adjusts for correlated outcomes within a trial.

Summarizing Studies with Effect Sizes

It can be difficult to summarize studies with effect sizes. Putting every measure on a common scale may hide important effects within trials. An example is a meta-analysis of bereavement interventions for children conduct by Curran et al., (2007). The meta-analysis revealed very low effect sizes for child bereavement interventions. However, the effects were stronger for high-risk children suggesting low-risk children were not as likely to be helped because of their "resiliency trajectory."

In conclusion, it is possible to examine contextual factors with traditional meta-analyses. There can be difficulties, however, since statistical power is often low, and there may be publication bias. Many meta-analyses end up summarizing all outcomes of a single trial using one effect size. This collapsing of outcomes to make a single effect size for a trial can hide important structure. Multilevel modeling can provide a useful means of analyzing multiple outcome level effect sizes to understand how outcome characteristics, type of outcome, length of follow-up, and interactions affect intervention impact. Such modeling requires some adjustment for correlation between multiple outcomes and measurement times within a trial. Finally, meta-analyses, although the best tool available, may miss the most important message when used for

policy implications. In particular, key predictor variables within a study that show important interactions with intervention condition may not be replicated across studies.