Practical Meta-Analysis

David B. Wilson
American Evaluation Association
Orlando, Florida, October 3, 1999

The Great Debate

- 1952: Hans J. Eysenck concluded that there were no favorable effects of psychotherapy, starting a raging debate
- 20 years of evaluation research and hundreds of studies failed to resolve the debate
- 1978: To proved Eysenck wrong, Gene V. Glass statistically aggregate the findings of 375 psychotherapy outcome studies
- Glass (and colleague Smith) concluded that psychotherapy did indeed work
- Glass called his method "meta-analysis"

The Emergence of Meta-Analysis

- **Ideas behind meta-analysis predate Glass' work by several decades**
  - **R. A. Fisher (1944)**
    - **"When a number of quite independent tests of significance have been made, it sometimes happens that although few or none can be claimed individually as significant, yet the aggregate gives an impression that the probabilities are on the whole lower than would often have been obtained by chance" (p. 99).**
    - **Source of the idea of cumulating probability values**
  - **W. G. Cochran (1953)**
    - **Discusses a method of averaging means across independent studies**
    - **Laid-out much of the statistical foundation that modern meta-analysis is built upon (e.g., inverse variance weighting and homogeneity testing)**

The Logic of Meta-Analysis

- Traditional methods of review focus on statistical significance testing
- Significance testing is not well suited to this task
  - highly dependent on sample size
  - null finding does not carry to same "weight" as a significant finding
- Meta-analysis changes the focus to the direction and magnitude of the effects across studies
  - Isn't this what we are interested in anyway?
  - Direction and magnitude represented by the effect size

When Can You Do Meta-Analysis?

- Meta-analysis is applicable to collections of research that
  - are empirical, rather than theoretical
  - produce quantitative results, rather than qualitative findings
  - examine the same constructs and relationships
  - have findings that can be configured in a comparable statistical form (e.g., as effect sizes, correlation coefficients, odds-ratios, etc.)
  - are "comparable" given the question at hand

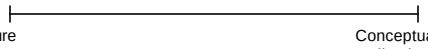
Forms of Research Findings Suitable to Meta-Analysis

- **Central Tendency Research**
  - **prevalence rates**
- **Pre-Post Contrasts**
  - **growth rates**
- **Group Contrasts**
  - **experimentally created groups**
    - **comparison of outcomes between treatment and comparison groups**
  - **naturally occurring groups**
    - **comparison of spatial abilities between boys and girls**
- **Association Between Variables**
  - **measurement research**
    - **validity generalization**
  - **individual differences research**
    - **correlation between personality constructs**

Effect Size: The Key to Meta-Analysis

- **The effect size makes meta-analysis possible**
  - **it is the "dependent variable"**
  - **it standardizes findings across studies such that they can be directly compared**
- **Any standardized index can be an "effect size" (e.g., standardized mean difference, correlation coefficient, odds-ratio) as long as it meets the following**
  - **is comparable across studies (generally requires standardization)**
  - **represents the magnitude and direction of the relationship of interest**
  - **is independent of sample size**
- **Different meta-analyses may use different effect size indices**

The Replication Continuum



Pure Replications — Conceptual Replications

You must be able to argue that the collection of studies you are meta-analyzing examine the same relationship. This may be at a broad level of abstraction, such as the relationship between criminal justice interventions and recidivism or between school-based prevention programs and problem behavior. Alternatively it may be at a narrow level of abstraction and represent pure replications.

The closer to pure replications your collection of studies, the easier it is to argue comparability.

Which Studies to Include?

- **It is critical to have an explicit inclusion and exclusion criteria (see handout)**
  - **the broader the research domain, the more detailed they tend to become**
  - **developed iteratively as you interact with the literature**
- **To include or exclude low quality studies**
  - **the findings of all studies are potentially in error (methodological quality is a continuum, not a dichotomy)**
  - **being too restrictive may restrict ability to generalize**
  - **being too inclusive may weaken the confidence that can be placed in the findings**
  - **must strike a balance that is appropriate to your research question**

Searching Far and Wide

- **The "we only included published studies because they have been peer-reviewed" argument**
- **Significant findings are more likely to be published than nonsignificant findings**
- **Critical to try to identify and retrieve all studies that meet your eligibility criteria**
- **Potential sources for identification of documents**
  - **computerized bibliographic databases**
  - **authors working in the research domain**
  - **conference programs**
  - **dissertations**
  - **review articles**
  - **hand searching relevant journal**
  - **government reports, bibliographies, clearinghouses**

Strengths of Meta-Analysis

- Imposes a discipline on the process of summing up research findings
- Represents findings in a more differentiated and sophisticated manner than conventional reviews
- Capable of finding relationships across studies that are obscured in other approaches
- Protects against over-interpreting differences across studies
- Can handle a large numbers of studies (this would overwhelm traditional approaches to review)

Weaknesses of Meta-Analysis

- **Requires a good deal of effort**
- **Mechanical aspects don't lend themselves to capturing more qualitative distinctions between studies**
- **"Apples and oranges"; comparability of studies is often in the "eye of the beholder"**
- **Most meta-analyses include "blemished" studies**
- **Selection bias posses continual threat**
  - **negative and null finding studies that you were unable to find**
  - **outcomes for which there were negative or null findings that were not reported**
- **Analysis of between study differences is fundamentally correlational**