Introduction to Meta-Analysis

- a bit of history
- definitions, strengths & weaknesses
- what studies to include ???
- "choosing" vs. "coding & comparing" studies
- what information to code along with each effect size ???

What got all this started?

The two events that seem to have defined & stimulated meta-analysis in Psychology

- In 1952, Hans J. Eysenck reviewed the available literature and concluded that there were no favorable effects of psychotherapy – guess how that went over…
  - 20 additional years of empirical research "failed to resolve the debate"
- In 1978, Gene V. Glass statistically aggregated 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

The statistical ideas behind meta-analysis predate Glass's work…

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"
- Source of the idea of aggregating 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 to decide "whether or not" there is an effect (though we really don't "believe" in the H0:")
- Significance testing is not well suited to this task
  - highly dependent on sample size
  - Most errors are Type II errors (e.g., Butcher's 59%)
  - question of comparability of studies of "same study"
- 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 is meta-analysis applicable?

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

Kinds of Research Amenable to Meta-Analysis

- Central Tendency Research
  - prevalence rates & averages
- Between Group Contrasts
  - Experimental designs
  - Non-experimental & Natural Groups designs
- Within-Groups Contrasts
  - Experimental designs
  - Non-experimental & Pre-Post designs
- Studies of Statistical Association Between Variables
  - measurement research (e.g., reliability & validitty)
  - individual differences research

The "Parts" of a meta-analysis

- Each study / analysis is a "case" in the meta analysis
  - simple studies will have single analysis giving a single ES
  - more complex studies may yield several ESs
- Effect Size (ES) is the "dependent variable" in the meta analysis
  - is comparable across studies
  - represents the magnitude & direction of the effect of interest
  - is independent of sample size
- Other "important" attributes of the study / analysis producing the effect size are the "independent variables" in the meta analysis – these have to be coded into the database

What are the strengths of meta-analysis ?

- A disciplined and quantitative approach to combining and comparing empirical research findings
- Is a non-hierarchical approach – doesn't favor earlier or later studies as a "starting place" to which we compare other studies
- Protects against over-interpreting differences across studies
- Can handle a large numbers of studies (this would overwhelm traditional approaches to review)
- Allows us to evaluate what attributes of a study are related to smaller vs. larger effect sizes
- Allows us to better balance concerns about "maximum effect size" and "maximum representativeness" when designing studies
- Allows us to plan smarter, more sensitive, and more useful studies!

What are the weaknesses* of meta-analysis ?

- Requires a huge amount of effort
- "Apples and oranges"; comparability of studies is often in the "eye of the beholder" (Wilson)
- Most meta-analyses include "blemished" studies
- Various forms of subjectivity…
  - What studies to include in the meta analyses
  - What study attributes to code
  - Coding of those attributes
- Often can't obtain study results or can't summarize as effect sizes
- Analysis of between study differences is fundamentally correlational

* None of these should impress you!

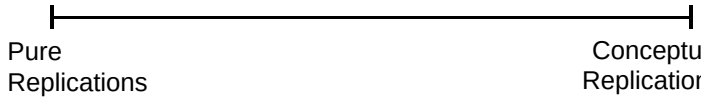
Which Studies to Include?

A bit of an aside…

- The main meta analytic question ***used*** to be…

  "What is the size of the effect under study?"
- Leading to the question → "What studies should we include?"
- The answer ***used*** to be "all ***comparable*** studies"
- You might imagine that answer led to much argument…
- Are studies using… … comparable?
  - …different operationalizations / measures of the DV…
  - …experimental and non-experimental designs…
  - …different populations (or subpopulations)…
  - …different → tasks … stimuli … equipment… settings …

The Replication Continuum



You have to be able to argue that the collection of studies chosen for meta-analysis 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.

(Lipsey & Wilson, 1993)

Which Studies to Include?

- The main meta analytic question is now more commonly…

  "What things influence the size of the effect under study?"
- Leading to the answer → Every study of "the effect"
- Leading to the question → "What attributes should we include?"
- The answer is → "all important attributes"
- Lots of coding, from careful methodological evaluation of each study!!! This is often the hardest part of the meta analysis!!!!

Said differently…

Meta analyses were primarily used in the past to "combine effect sizes from comparable studies," usually to ask if the effect was "non-zero" (e.g., Glass & Smith).

Meta analyses are primarily used currently to "examine relationships between how a study is conducted and the effect sizes obtained from that study."

So (finally) … Which Studies to Include?

- You must have explicit criteria for what studies you include
- Those criteria must conform to the "standards of your people"
- "Published studies" won't hack it !
  - Because studies retaining Null are less likely to be published, including only published studies biases effect size estimates away from 0.
- Potential sources for identification of documents
  - computerized bibliographic databases
  - authors working in the research domain
  - conference programs
  - dissertations
  - review articles
  - hand searching relevant journals
  - government reports, bibliographies, clearinghouses

Which Information to Include about each study?

Coding your database, so that you know "all the important stuff" about each study has 4 purposes. The coding will help you ,,,

- identify groups of studies that are "replications"
- compare studies to understand what design elements are related to the size of the effect found
- Adjust/correct individual effect sizes to give more useful values
- It will give you a better understanding of the research literature than you can possibly imagine!!!
  - Folks who survive the meta-analysis process often say that this was the most valuable result of their study
  - You will see details, similarities, differences, genus & mistakes in a literature that you thought you knew!!!!

What attributes of the study to code ???

Everything that may be influencing the results & effect size !!!

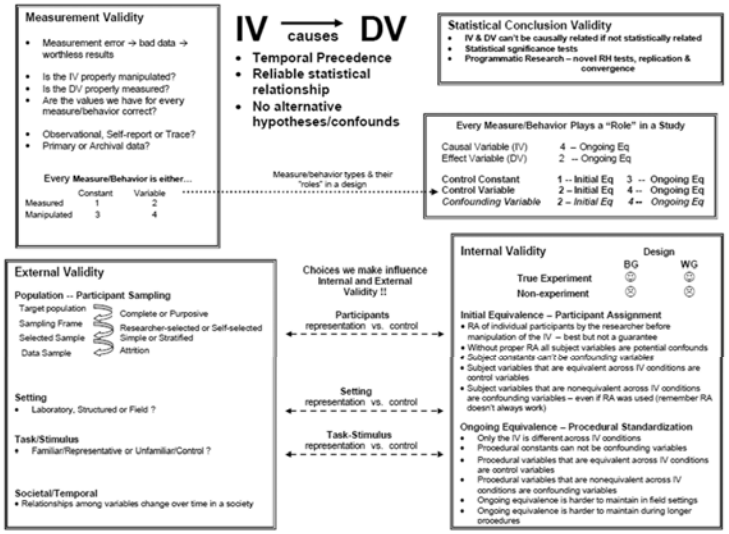
On the next several pages are summaries of materials we've used in previous classes to characterize and evaluate research designs – any of the attributes listed could add value to your meta-analysis.

Put differently…
Everything that can influence study results & statistical conclusion validity can also influence the effect size found!!!

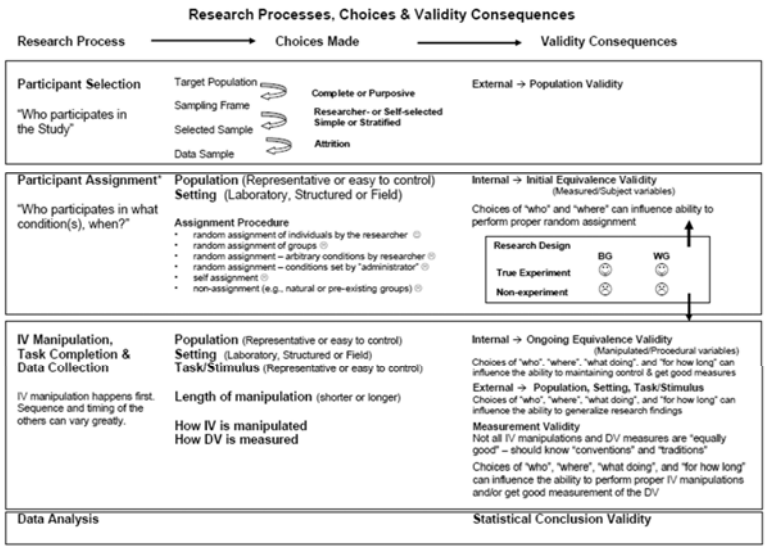
- All-the-Words page – organizing the design/validity jargon
- Validity Net – organizing the jargon around article critiquing
- Researcher Choices – organizing jargon around designing studies
- Relationships among the types of research validity
- Variance sources in research designs & procedures

Suggested Data to Code Along with the Effect Size

Suggested Data to Code Along with the Effect Size



Suggested Data to Code Along with the Effect Size



Suggested Data to Code Along with the Effect Size

External Validity

Do the who, where, what & when of our study represent what we intended want to study?

Measurement Validity

Do the measures/data of our study represent the characteristics & behaviors we intended to study?

Internal Validity

Are there confounds or 3rd variables that interfere with the characteristic & behavior relationships we intend to study?

Statistical Conclusion Validity

Do our results represent the relationships between characteristics and behaviors that we intended to study?

- did we get non-representative results "by chance" ?
- did we get non-representative results because of external, measurement or internal validity flaws in our study?

Suggested Data to Code Along with the Effect Size

$SS_{Total} = SS_{IV} + SS_{subcon} + SS_{proccon} + SS_{Indif} + SS_{wcsubinf} + SS_{wcprocinf}$

Sources of variability…

SS_{IV} → IV ☺

SS_{subcon} → subject variable confounds (initial eq problems)

$SS_{proccon}$ → procedural variable confounds (ongoing eq pbms)

SS_{indif} → population individual differences ☺

$SS_{wcsubinf}$ → within-condition subject variable influences

$SS_{wcprocinf}$ → within-condition procedural variable influences

Suggested Data to Code Along with the Effect Size

1. A label or ID so you can backtrack to the exact analysis from the exact study – you will be backtracking!!!
2. Sample size for each group *
3. Sample attributes (e.g., mean age) #
4. DV construct & specific operationalization / measure #
5. Point in time (after/during TX) when DV was measured #
6. Reliability & validity of DV measure *
7. Standard deviation of DV measure *
8. Type of statistical test used *#
9. Between group or within-group comparison / design #
10. True, quasi-, or non-experimental design #
11. Details about IV manipulation or measurement #
12. External validity elements (pop, setting, task/stimulus) #
13. "Quality" of the study #
    - better yet → data about attributes used to eval quality!!!

* Used to "adjust" effect sizes

\# Used to compare studies