

Generalizability: The Role of Meta-Analysis

Mark W. Lipsey
Vanderbilt University

Workshop on Strengthening Benefit-Cost Methodology for
the Evaluation of Early Childhood Interventions

National Research Council, March 4-5, 2009
Washington DC



The concept of “generalization”

The same intervention will produce the same (or nearly the same) effect despite variation on some specified dimensions of generalizability, e.g., characteristics of recipients or providers, settings, location, time, etc.

Knowing what effect is produced, however, requires appropriate research; “effects” are thus estimates from research findings.



Ideal (?): Generalizability evidence from each effectiveness study

A study with both strong internal validity and strong external validity would require:

- ✧ A relatively large research sample randomly drawn from the relevant population (external validity)
- ✧ Random assignment to intervention and control conditions with low attrition, etc. (internal validity)

Rare combination; typically strive for internal validity with a not-too-unrepresentative convenience sample




Next best (?): Meta-analysis of multiple studies with internal validity and diverse convenience samples

- n Effects are represented as effect size statistics to provide a common metric across studies
- n Analysis focuses on the central tendency and variation of the distribution of effect sizes
- n Key question is whether effects vary and, if so, how it is related to study characteristics
- n Meta-analysis can thus be characterized as the empirical study of the generalizability of intervention effects




Lessons from Meta-Analysis



Definitional problems about what constitutes the “same” intervention

- n Few interventions have a manual, script, or recipe that unambiguously defines them
- n Developers often modify the recipe from study to study– same or different?
- n Providers often adapt and change the recipe in practice– challenge to generalization from research to practice.



Sufficient research to support generalization exists only for intervention “types”

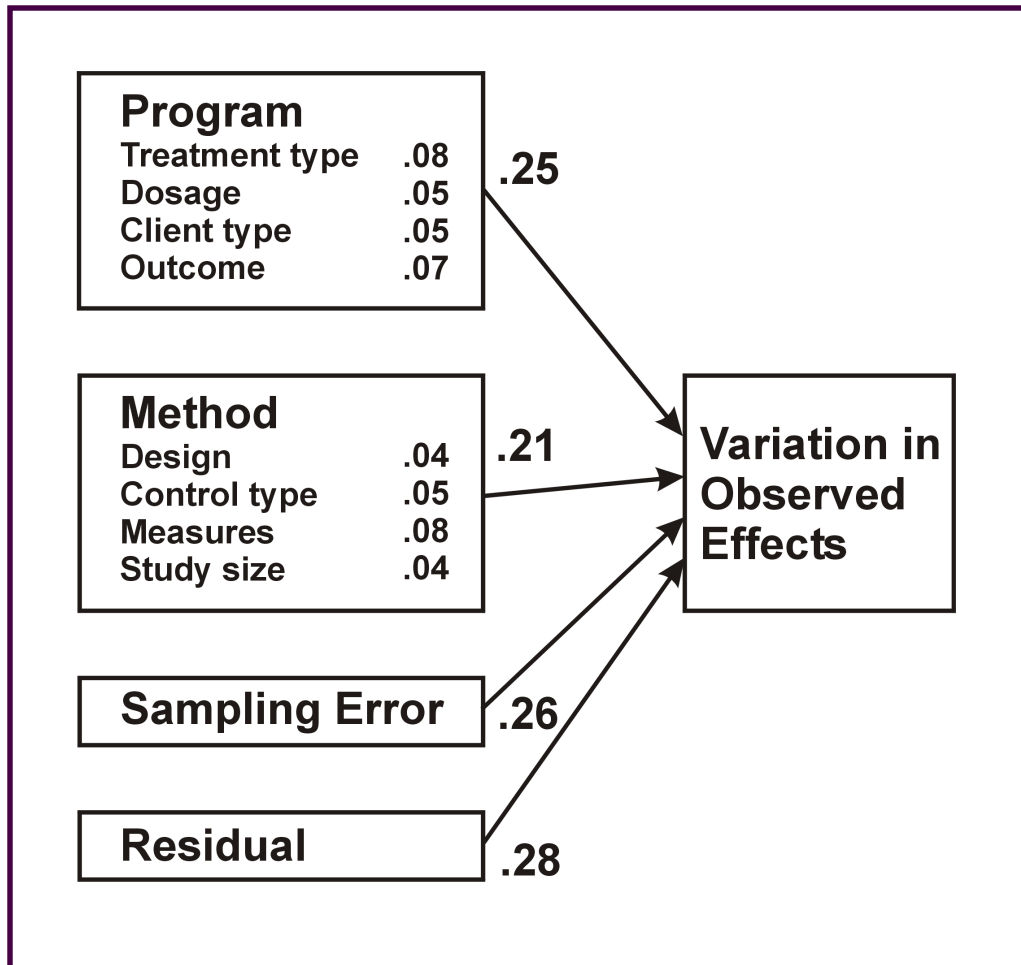
- n Very few defined (e.g. manualized) interventions have enough research to adequately explore generalizability
- n Meta-analyses and reviews typically focus on an intervention type defined generically
- n No established intervention typologies are available to guide classification into types
- n Classification is a judgment call very much in the eye of the beholder



Variability abounds

- n Effect size distributions can be statistically tested for “excess” variation (Q-test, I^2)— greater between-study variance than expected from subject level sampling error
- n Effects for a given type of intervention almost always show great variability— typically many times more than expected from subject sampling
- n Such variability is inconsistent with any presumption of broad generalizability of effects

Many sources of variance in observed effects of social interventions



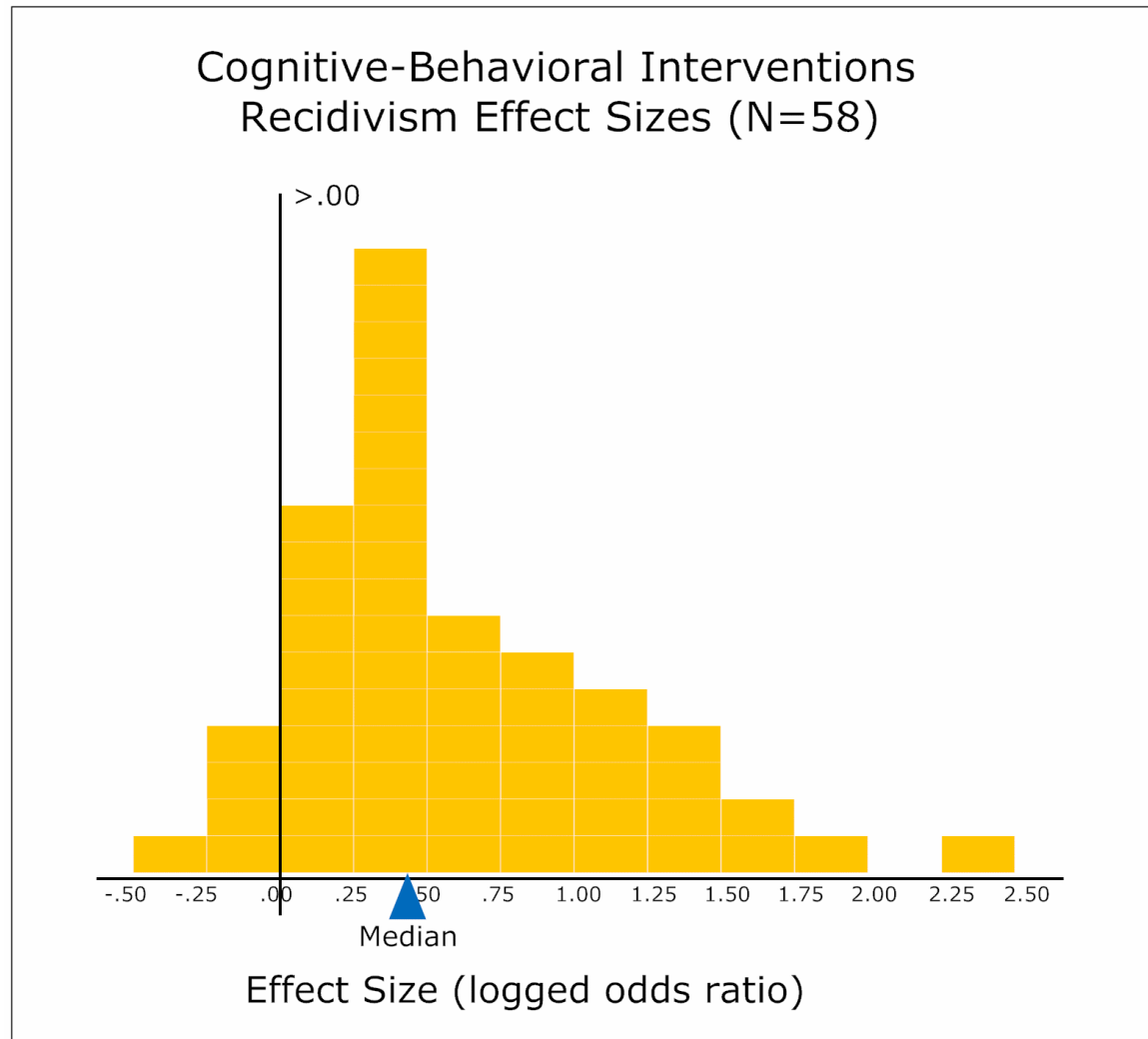
Estimates based on 300 meta-analyses of intervention studies



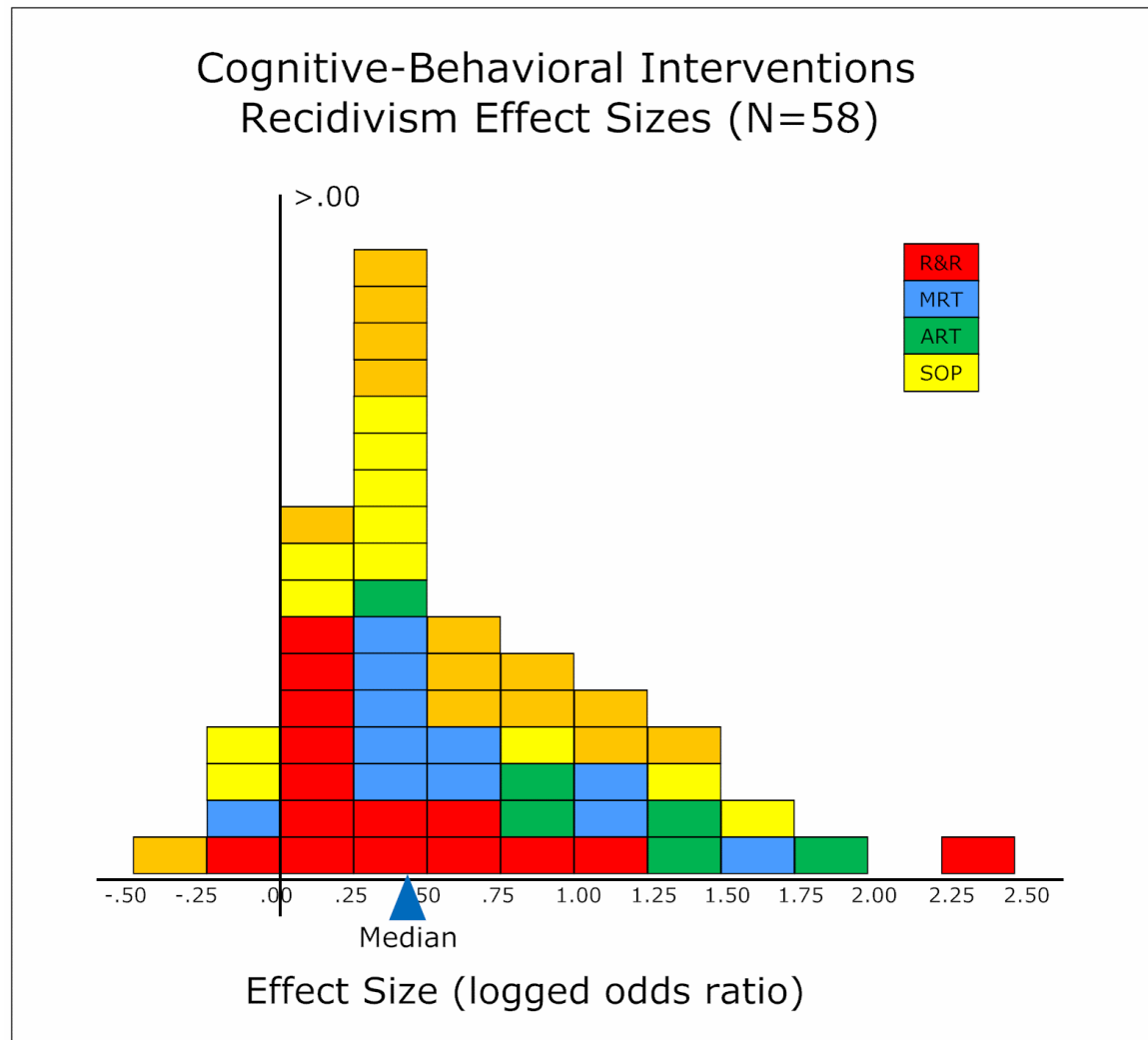
Large amount of variation in effects associated with method differences

- n Obscures view of “real” effects— observed effects partly stem from researcher input
- n Influence of outcome measurement and control group choice is especially notable (design differences better known)
- n Publication bias can also be a problem
- n Biggest problem: Method differences are confounded with the substantive variables most relevant for generalization, e.g., subject characteristics, setting, research context, etc.

Variation in effects for a given type of intervention (CBT for offenders)



Variation in effects for specific interventions within a type






Typical sources of variation associated with a given intervention

n Implementation

- ✕ Fidelity
- ✕ Quantity/dose
- ✕ Quality

n Developer involvement

- ✕ R&D programs vs. routine practice
 - n approximately 2x effect size difference



Moderators usually associated with generalizability may also matter

- n Subject characteristics

- ✕ Gender, ethnicity, age, SES, etc.
- ✕ Risk level may be important

- n Organizational, geographic setting

- n Provider personnel– type & training

Note: Typically unexplained variation in effects after all moderators accounted for



Response surface perspective (Rubin)

- n Response surface defined by the multiple dimensions of interest along which effects vary, though missing data in many cells
- n The best current meta-analysis approach for exploring generalizability is construction of multivariate models that allow ‘prediction’ of expected outcomes for given scenario of method, intervention, subject, setting, etc. characteristics
- n Requires a relatively large number of diverse studies; limited by inadequate reporting




In short ...


- n Establishing or exploring generalizability adequately through intervention research with representative probability samples is not a practical option
- n An alternative is to examine the relationship between cross-study variability in effects and moderator variables that represent dimensions of potential generalizability via meta-analysis



But, assessing the generalizability of the effects found in research for a given intervention is a challenging task

- ✧ Unresolved definitional issues about what constitutes the ‘same’ intervention from study to study
- ✧ Insufficient number and diversity of studies to allow exploration of generalizability for the most specifically defined ‘protocol’ interventions

- 
- ✧ Even with the ‘same’ intervention, natural variation in implementation produces varying effects, further complicating the matter of defining the intervention at issue
 - ✧ Social interventions typically show great variability of effects across studies, much of which is related to characteristics of the research rather than characteristics of the intervention or its recipients
 - ✧ It is difficult to disentangle the ‘real’ intervention effects from the confounded method variables

- 
- ✧ Addressing generalization across recipient characteristics, settings, etc. requires estimating the variation associated with those moderators net of that associated with method and implementation variation.
 - ✧ Meta-analysis attempts this estimation with multivariate models, but this requires many diverse studies that provide adequate information on the relevant study features.

And, so, there is much work yet to be done before we will know much about the generalizability of intervention effects