Published in final edited form as:

Child Dev Perspect. 2008 November 28; 2(3): 198–305. doi:10.1111/j.1750-8606.2008.00065.x.

Examining How Context Changes Intervention Impact: The Use of Effect Sizes in Multilevel Mixture Meta-Analysis

C. Hendricks Brown¹, Wei Wang¹, and Irwin Sandler²

¹ University of South Florida

² Arizona State University

Abstract

In describing the impact of an intervention, a single effect size, odds ratio, or other summary measure is often employed. This single measure is useful in calibrating the effect of one intervention against others, but it is less meaningful when the intervention displays variation in impact. A single intervention trial can show differential effects when subgroups respond differentially, when impact varies by environmental context, or when there is varying impact with different outcome measures or across follow-up time. This article presents a multilevel mixture modeling approach for meta-analyses that summarizes these sources of impact variation across trials and measured outcomes.

Keywords

meta-analysis; effect sizes; multilevel modeling; mixture modeling

Effect sizes (ESs) and related scale-invariant measures, such as odds ratios, risk differences, and hazard ratios, are often used to calibrate the strength of an intervention's impact in a single experimental trial. By combining ESs from an identical or similar interventions tested across different trials conducted as separate studies, it is possible to obtain a single measure of impact representing overall effect across different trial conditions (Hedges, 1982; Hedges & Olkin, 1985; Lipsey & Wilson, 2001). Different interventions that are tested in different trials can be ranked based on the magnitude of their ESs. This approach allows policy makers to select between competing strategies even if these interventions have never been compared head to head in a trial. One example of the use of this approach has been the comparison of prevention versus service programs to reduce delinquency. Such comparison reveals that, in terms of reduction in criminality, prevention benefits exceed those of incarceration (Greenwood, Model, Rydell, & Chiesa, 1996). Finally, the ESs from a broad array of related trials can be treated as a dependent variable in a regression model to identify common factors in interventions predicting higher impact. This analytical model led Tobler et al. (2000) to conclude that interactive curricula led to lower drug use than did lecturing.

Traditional meta-analyses of intervention impact use a single mean to assess the impact of a single intervention across replicated trials (Petrosino, Boruch, Rounding, McDonald, & Chalmers, 2000; The Cochrane Collaboration, 2007). Many of the well-known meta-analyses—such as those of 177 mental health prevention programs for youth (Durlak & Wells, 1997), 207 school-based drug prevention programs (Tobler et al., 2000), 221 delinquency prevention

programs (Wilson et al., 2003a, Wilson, Lipsey, & Soydan, 2003b), and 69 depression prevention programs (Horowitz & Garber, 2006; Jané-Llopis, Hosman, Jenkins, & Anderson, 2003)—all chose a single ES for each trial. However, in all these meta-analyses, there were typically 4–10 distinct, relevant outcome analyses for each trial corresponding to multiple outcome variables, multiple follow-up times, and analyses involving distinct subgroups (Brown, Berndt, Brinales, Zong, & Bhagwat, 2000). In cases like these, the reduction to one index of program effect is often insufficient, given that impact within a trial may vary over time, across different outcome measures, and across subgroups and environmental contexts (Brown & Faraone, 2004; Brown & Liao, 1999; Brown et al., 2008; Faraone, Brown, Glatt, & Tsuang, 2002).

In this article, we discuss multilevel mixture models that examine the influence of contextual factors on intervention impact both within and across trials. We illustrate how multilevel mixture modeling in meta-analyses helps identify and interpret contextual factors that influence impact within or across trials.

Multilevel approaches to meta-analysis are not new; indeed, the random effects approach to address site heterogeneity is equivalent to a multilevel approach (Hedges & Olkin, 1985; Raudenbush & Bryk, 1985). Various methods have been proposed to deal with special cases. Rosenthal and Rubin (1986) handled more than one outcome variable by forming a single composite index and correcting for a common correlation. Raudenbush (1994) presented the fundamental multilevel model in meta-analysis with extensions by Goldstein, Yang, Omar, Turner, and Thompson (2000) and Hox (2002). To account for heterogeneity in effects beyond the inherent variation in the effects within each trial, all these models involve continuous random effects that are assumed to have normal distributions. In this article, we also include discrete mixtures that allow for different clustering of ESs on top of these continuous mixtures of random effects. These so-called hybrid models with both discrete and continuous mixtures have been applied in the multilevel framework, especially to examine different growth trajectories (Asparouhov & Muthén, 2008; Muthén, 2004; Muthén & Asparouhov, 2008).

Building on this work, we apply a multilevel mixture modeling approach to examine context at both individual and trial level using data from group-based randomized trials. The method handles standard errors that reflect group-randomized trials, and a new mixture modeling approach identifies variation in impact not captured by measured covariates. We provide representative computer code for carrying out these analyses in Mplus (Muthén & Muthén, 2008). For nonmethodologists, the technical formulae are kept at a minimum, and the meaning of these formulae can be obtained from the corresponding text.

A RATIONALE FOR INCLUDING WITHIN- AND BETWEEN-EXPERIMENT CONTEXTUAL FACTORS

The examination of contextual factors within and across studies is especially important in the prevention field, where a fundamental question is whether an intervention varies in its benefit or harm across different levels of baseline risk (Brown et al., 2008). Such information can identify subgroups harmed by an intervention, and it may lead to different interventions targeting different risk groups. Meta-analyses can be helpful in assessing promising strategies, but they are often limited to examining impact of contextual factors at the level of the study, with little attention to variation within trials. In one of the few examples of a meta-analysis that compares impact at both the individual and study (ecological) level, Berlin, Santanna, Schmid, Szczech, and Feldman (2002) found a strong significant interaction between individual-level risk and intervention impact across the studies. However, their meta-analysis using the percent of subjects at high risk as an ecological level covariate found no difference

in impact. Thus, meta-analyses involving only ecological measures can draw erroneous conclusions of effect at the individual level.

When the intervention is administered in a group context, one or more ecological levels may have important roles in affecting outcomes (Schwartz, 1994). Wilson et al. (2003a) reported that intervention impact on aggressive behavior is dramatically stronger in those trials where youth were at higher risk. Specifically, for interventions aimed solely at youth with highest individual-level risk just short of a diagnosis, the average ES was strong ($ES = 0.41$). For interventions that were coded as selective based on individual risk factors, which included samples with moderately elevated personal risk, such as being male, the average ES was 0.26. For those selective interventions based on group-level risk alone, such as living in a high-crime area, the average ES was 0.13. Finally, for so-called universal interventions aimed at the general population, such as a school-based program, the average ES was 0.09, lowest among all the types of interventions. Thus, higher average effects occur in trials with more at-risk populations. This might suggest that preventive interventions aimed at the general population have little effectiveness in addressing aggressive behavior. But this conclusion would be wrong: Universal preventive interventions can be quite effective with high-risk youth. Using methods to examine variation in risk within universal early-intervention trials (Brown et al., 2008), our colleagues (Kellam et al., 2008) and others (Eddy, Reid, & Fetrow, 2000) have shown dramatic improvements along the life course from universal classroom-based interventions for those children who are aggressive, with much less impact on the majority of nonaggressive children. Only by examining this variation within a study can one compare the impact of a universal intervention on aggressive youth. Again, meta-analyses need to examine both ecological and individual-level risk.

Differences in intervention impact may also occur across time. A study of group-based intervention for children who had recently experienced the death of a parent, found a pattern of significant program interaction—specifically, positive effects for girls but not for boys (Sandler et al., 2003). However, these effects were found at the 11-month follow-up assessment but were not present at the immediate posttest. Thus, estimating a single ES for this study that pooled across gender and across time of assessment would have missed this program effect, as it would in other trials as well (DeGarmo, Patterson, & Forgatch, 2004; Wolchik, Sandler, Weiss, & Winslow, 2007).

Impact may also differ across outcome measures in a single study. For example, in early analyses of the Multimodal Treatment Study of Children with ADHD (MTA), there were few differences between the effect of medical management alone and the effect of medical management plus behavioral treatment. However, over time and across several externalizing, internalizing, and school performance measures, the combined medical and behavioral treatment showed benefit over medical management alone (aMTA Cooperative Group, 1999a; MTA Cooperative Group, 1999b).

MULTILEVEL MODELS OF EFFECT SIZES TO INCORPORATE WITHIN- AND BETWEEN-STUDY VARIATION IN IMPACT

In this section, we describe an analytic method for examining both within- and between-study variation in ESs. Multilevel regression analyses is used to account for heterogeneity due to known covariates at both levels, and mixture modeling is used to identify the extent of variation not explained by available covariates.

For our example, we examine the impact of school-based drug prevention programs in rural areas (Brown, Guo, Singer, Downes, & Brinales, 2007), using expanded summaries of outcomes on drug use obtained from references in Tobler's broader meta-analysis (Tobler et

al., 2000). Beginning with a population of trials that satisfied our inclusion/exclusion criteria of rural school-based drug prevention programs ($n = 24$), we compiled all 435 documented analyses of intervention impact on drug-related behavioral outcomes. These analyses included main effects analyses and subgroup analyses for distinct outcome measures and at each time point. Here, we concentrate on continuous measure ESs ($n = 93$), each coded so that a positive value indicates benefit.

Three quarters of the ESs were positive, indicating an overall beneficial intervention impact. More important was the variation in ESs, as shown in a smoothed histogram in Figure 1. In the central mode, the ESs were close to zero, but two other modes existed; one with ESs around 0.5 and a smaller cluster of negative impacts of the intervention around -0.5 . Standard multilevel modeling identified several outcome analysis-level factors that explained some of this variation (Brown, Guo, et al., 2007). Here, we incorporate discrete mixtures to reflect the trimodality in Figure 1.

Multilevel meta-analyses require the inclusion of standard errors for each outcome analysis. Few of the articles reported standard errors taking into account intraclass correlations for group-based trials (Murray, 1998; Raudenbush, 1997), so we chose the larger of two standard errors, one calculated by assuming independence and one that incorporated a small intraclass correlation appropriate for drug prevention studies (Brown, Guo, et al., 2007).

Each of these within-trial outcome analyses requires coding for a main effect or subgroup analysis, follow-up time, and type of outcome measure. All these within-trial covariates are measured at the level of the outcome analysis (first level). They include outcome analyses involving individual-level covariates such as gender.

The second level in the study is at the trial level, and covariates at this trial level were also coded. These covariates include characteristics of the intervention, its delivery, quality of the trial design, and characteristics of the overall population averaged over schools, such as average age and overall rate of drug use at baseline.

The two-level data for each outcome analysis were then analyzed in mixed-model regressions taking into account fixed-effect predictors for Level 1 and Level 2 covariates, random effects incorporating the measured standard error for each outcome, and random effects examining unexplained variation at both levels. Symbolically, the multilevel model—without discrete mixtures—when combined across first and second levels is

$$\begin{aligned} ES_{Outcome,Trial} = & \alpha X_{Outcome,Trial} + \beta X_{Trial} \\ & + a_{Outcome,Trial} SE_{Outcome,Trial} \\ & + b_{Trial} Z_{Trial} + \varepsilon_{Outcome,Trial} + \varepsilon_{Trial} \end{aligned} \quad (1)$$

For the most general model, each ES for each outcome in each trial is predicted by six terms. These involve three types of factors at the outcome level within a trial and three types of factors at the trial level. Listed first are the measured covariates contributing to fixed effects at the level of each outcome analysis, $X_{Outcome, Trial}$, as well as covariates at the level of the trial, X_{Trial} . The vectors of regression coefficients α and β examine the strength of these fixed factors, and they, along with their appropriate standard errors, can be used to assess the strength of these outcome level and trial level covariates. The second line of Equation 1 includes two distinct types of random slopes; they differ from the fixed effects in the first line because the coefficients themselves are permitted to vary across (Level 1) outcomes. With the means of these random slopes fixed to zero, inferences are based on the magnitude of the variances for these random slopes. The two random slope terms incorporate measured outcome level

standard errors for each analysis and random slopes related to trial level covariates. The first of these random slopes involves the outcome analysis standard error (SE), multiplied by a random slope $\alpha_{\text{Outcome Trial}}$ that has a standard normal distribution with mean 0 and variance 1. These restrictions ensure that internal standard errors are correctly handled, and they account for the “known variance” type of random errors that are incorporated in mixed-effect meta-analyses. The second term on this line involves random slopes at the level of the trial. This could reflect heterogeneity in overall impact for, say, trials conducted in more rural settings.

On the last line are two residual variances. The first of these is an outcome level residual variance that is not explained by the standard errors or the fixed outcome level effects. The second term involves $\varepsilon_{\text{Trial}}$ and allows for extra variability to occur at the trial level.

Examination of the role of contextual factors is based on the terms remaining in a best fitting multilevel model. In the above equation, if a particular covariate is used as a fixed effect but is not included as a random slope (or the random slope has negligible variance) then the coefficient of the fixed covariate fully reflects the magnitude of this contextual influence. A significant positive coefficient implies that intervention impact increases in proportion to this factor. If both fixed and random factors are included for a single covariate, with a significant fixed-effect coefficient and a variance in the random slope that differs from zero, then this factor contributes overall to intervention impact, but there is still unexplained heterogeneity associated with this covariate. Another unobserved variable may interact with the covariate to explain such variation. All the random effects in this equation other than that for standard error reflect sources of unexplained variation in impact.

We now present a new analytic method using discrete mixtures in meta-analyses. Discrete mixture modeling allows for heterogeneity that comes from an unmeasured categorical variable. For purposes of illustration, assume that the observed data consist of multiple clusters of data but that there are no covariates available to distinguish these clusters. For example, one may have a mixed sex data set for young people’s aggression with a single variable reflecting each youth’s aggressive behavior. Because male aggressive behavior is generally higher than female aggressive behavior, we would anticipate that the data may be bimodal when youth’s gender is unknown. A simple mixture analysis on this single variable could be performed by assuming that the representative outcome is a mixture of two normal distributions having different means for males and females. Here, the mixing proportion would be the proportion of males in the sample, and the density of the outcome would be a weighted average of these normal densities for males and females. Statistical inference on mixtures comes from estimating the proportion in each mixture class and the means and variance parameters for each class. These discrete mixture models can produce multimodality in the ES distribution in more complex models, even when covariates and random slopes (i.e., continuous mixtures) are included in these models. One advantage of these methods is that they allow detection of heterogeneity in the effects across outcomes (Level 1 mixtures) or trials (Level 2 mixtures) that are not accounted for by measured covariates.

In model terms, discrete mixtures can be added to Equation 1 by allowing some or all of the fixed coefficients to depend on the class to which the mixture belongs. Thus, a simple shift in the overall mean for two classes in a mixture model can be represented by having two equations like that of Equation 1, one for each class, where all coefficients are the same except for the class-specific intercepts. Alternatively, if some coefficients are permitted to vary across mixture classes, then this points to heterogeneity in these effects.

Returning to the evaluation of drug prevention programs in rural settings, we can illustrate this process by comparing the results of a single-class (nonmixture) model with those of a two-class mixture model involving differences in impact across classes in both the effects of length

of follow-up (Time: 1 if > 6 months follow-up, 0 if < 6 months) and among the most rural versus less rural communities where these trials took place (Rural: 1 if mostly rural, 0 if partly rural, based on authors' descriptions or percentages). These and similar models can be fit using marginal maximum likelihood analyses in Mplus (Muthén & Muthén, 2008), and decisions between competing models can be based on the Bayesian Information Criterion (BIC; Schwarz, 1978), as well as on graphical (Wang, Brown, & Bandeen-Roche, 2005) and other model-fit procedures (Carlin, Wolfe, Brown, & Gelman, 2001). In our example, we began by fitting all ESs in a standard two-level random effects model—without a discrete mixture—and obtained nonsignificant effects for both the first-level follow-up time (defined as Time above) and the second-level rurality measure. Furthermore, the overall average ES was very close to zero, and there was a very large unexplained residual variability in the Level 1 ESs with a standard deviation of 0.23 (95% confidence interval [CI] 0.19–0.26).

We then examined a two-class mixture model that used the same Level 1 and Level 2 covariates. In this second model, we found that the longer follow-up time and mostly rural sites led to significantly higher impact for the vast majority—91%—of the outcomes of these trials (95% CI = 0.022–0.130 for Time and 0; 0.31–0.331 for Rural). For the remaining 9% of all outcomes in the second class, the worst outcomes occur in the mostly rural communities (95% CI = −0.84 to −0.488). We concluded that this two-class mixture improved the fit over the one-class model, based on very large differences in BIC (−77.706) and likelihood ratio tests ($\chi^2 = 91.304$ on 3 df, $p < .0001$). The estimates for the two-class model are given below:

Class 1 (91% of sample):

$$\begin{aligned} \text{Class 1 (91\% of sample):} \\ ES_{\text{Outcome,Trial}} = & 0.076\text{Time}_{\text{Outcome,Trial}} + 0.18\text{Rural}_{\text{Trial}} \\ & + 1SE_{\text{Outcome,Trial}} + 0.10\epsilon_{\text{Outcome,Trial}} + 0.09\epsilon_{\text{Trial}} \end{aligned} \quad (2)$$

Class 2 (9% of sample):

$$\begin{aligned} \text{Class 2 (9\% of sample):} \\ ES_{\text{Outcome,Trial}} = & -0.664\text{Rural}_{\text{Trial}} + 1SE_{\text{Outcome,Trial}} \\ & + 0.10\epsilon_{\text{Outcome,Trial}} + 0.09\epsilon_{\text{Trial}} \end{aligned} \quad (3)$$

All coefficients for fixed effects are significantly different from zero, as indicated by the CIs reported above. The coefficients given for the random error terms are standard deviations of the unexplained random variation at Levels 1 and 2, respectively. The Level 1 residual standard deviation in this mixture model is still significantly greater than zero (95% CI = 0.09–0.12) but only half the size in the nonmixture model.

The role of contextual effects can be seen clearly in this two-class mixture model. For the first class consisting of 91% of the outcomes, the predicted impact is nearly zero unless the outcome is measured beyond 6 months or the sample has a majority of rural schools. When both of these Level 1 and Level 2 factors are present, the predicted ES is the sum of these coefficients, 0.257. For class 2, which consists of only 9% of the outcomes, the intervention effect is nearly zero unless the schools were mostly rural, in which case the impact is strongly negative (−0.664). There is still unexplained variance at the outcome level but not at the trial level. Together, these effects closely replicate the three-modal distribution in Figure 1. The annotated code in the Appendix illustrates how to fit this model.

DISCUSSION

The procedures we have described allow assessment of known and unknown sources of variability in intervention impact. The overall average ES in this analysis is slightly positive, telling us that there is slight overall benefit of school-based drug prevention programs. But an average ES does not provide any explanation of the variation in impact, as do multilevel meta-analyses that include all relevant analyses from each trial. This information may be critical to making informed policy statements.

In our example, a histogram identified heterogeneity in effects across subjects, outcomes, or studies. Although these could potentially be explained by measured covariates at Levels 1 and 2, our mixture analysis recognized the existence of some negative outcomes, but no specific factors in our data predicted these poor outcomes. We note that the small proportion in this class may indicate spurious findings, but the large magnitude of these negative ESs suggests vigilance in searching for causes of potentially harmful outcomes in school-based prevention of substance abuse.

The methodology presented here is a generic tool to examine variation in impact. However, theory often suggests specific treatment interactions that are important to examine and that have implications for program delivery. One critical covariate to examine is how the intervention varies by baseline risk at Level 1 (Brown et al., 2008). Many universal (Brown & Liao, 1999) and selective (Dawson-McClure, Sandler, Wolchik, & Millsap, 2004) preventive interventions show their largest impact on those at highest risk at baseline, but this need not be the case, especially for programs that mix elements of primary and secondary prevention. A drug prevention program that gives direct messages to “not use drugs” might be of potential benefit to nonusers but have adverse effects on those who had begun using drugs before the intervention started. If a universal program was directed more toward those who already have engaged in drug use, the program might adversely affect low-risk subjects. This has been suggested by research on anorexia prevention programs, which indicates that a message to not engage in bulimia for weight loss may lead some to initiate purging (Mann et al., 1997). So far there seems to be no parallel risk in suicide prevention programs, there being no evidence that simply talking about suicide increases adolescents’ suicidal thoughts (Brown, Wyman, Brinales, & Gibbons, 2007; Gould et al., 2005).

We believe that it is a fundamental scientific and public policy flaw to report only main effects in analyses of universal interventions without examining how impact varies in relation to the underlying baseline risk factors that are targeted by the intervention. Without reporting these analyses, there is no way to determine if there is beneficial impact on those at most or least risk.

Many meta-analyses ignore time of impact. Programs that have a positive effect at posttest but diminish over time are of much less value than those that last across developmental periods (e.g., Hawkins, Kosterman, Catalano, Hill, & Abbott, 2005; Kellam et al., 2008; Olds et al., 1997; Wolchik et al., 2002). Meta-analyses that derived a single estimate from pooling these effects across time would seriously misrepresent the findings.

We provide several cautions about interpreting these models. First, the collection of all reported outcomes of a set of trials, as in our example of rural, school-based drug prevention, can no longer be considered fully representative. Reported impact analyses may be extensive in one trial that has lots of positive results and more limited in another that shows little impact (Cooper, 1998). Thus, we view the full collection of ESs that is available as an observational study, even though the trials have experimental research designs. Conclusions drawn from these multilevel analyses provide less rigorous evidence than would a single hypothesized analysis based on a single well-conducted randomized field trial (Rubin, 1992). In addition, the process of selecting

best fitting multilevel models and then testing coefficients can introduce spurious findings just as it can in stepwise procedures for multiple regression. Many mixture models turn out to be highly sensitive to certain model restrictions, and one should always determine how stable the estimates are to such restrictions; here our results are quite stable. Determining the number of mixture classes is challenging (Nylund, Asparouhov, & Muthén, 2007), and the numerical task of fitting mixture models requires a large number of random starts and adjustment of the convergence criteria to assure that the optimal solution has been obtained (Muthén & Muthén, 2008). Finally, there may be more than one well-fitting statistical model, and they may lead to quite different interpretations (Wang et al., 2005).

Acknowledgments

We thank our colleagues in the Prevention Science and Methodology Group for their suggestions on this article and especially Dr. Bengt Muthén for his development of multilevel hybrid mixture models that have heavily influenced this article. This work has been supported by National Institute on Drug Abuse (NIDA) under Grant R01 MH040859-14S1, by National Institute of Mental Health (NIMH) and NIDA under Grant R01 MH40859, and by NIMH under Grants P30 MH068685 and R01 MH049155. An earlier version of this article was presented at the Application of Effect Sizes in Research on Children and Families: Understanding Impacts on Academic, Emotional, and Behavioral Outcomes Roundtable, sponsored by the Administration for Children and Families and the National Institutes of Health.

References

- Asparouhov, T.; Muthén, BO. Multilevel mixture models. In: Hancock, GR.; Samuelsen, KM., editors. *Advances in latent variable mixture models*. Charlotte, NC: Information Age Publishing; 2008. p. 27-51.
- Berlin JA, Santanna J, Schmid CH, Szczech LA, Feldman HI. Individual patient-versus group-level data meta-regressions for the investigation of treatment effect modifiers: Ecological bias rears its ugly head. *Statistics in Medicine* 2002;21:371–387. [PubMed: 11813224]
- Brown CH, Berndt D, Brinales JM, Zong X, Bhagwat D. Evaluating the evidence of effectiveness for preventive interventions: Using a registry system to influence policy through science. *Addictive Behaviors* 2000;25:955–964. [PubMed: 11125782]
- Brown, CH.; Faraone, SV. Prevention of schizophrenia and psychotic behavior: Definitions and methodologic issues. In: Stone, WS.; Faraone, SV.; Tsuang, MT., editors. *Early clinical intervention and prevention in schizophrenia*. Totowa, NJ: Humana Press; 2004. p. 255-284.
- Brown CH, Guo J, Singer LT, Downes K, Brinales JM. Examining the effects of school-based drug prevention programs on drug use in rural settings: Methodology and initial findings. *Journal of Rural Health* 2007;23(Suppl):29–36. [PubMed: 18237322]
- Brown CH, Liao J. Principles for designing randomized preventive trials in mental health: An emerging developmental epidemiologic paradigm. *American Journal of Community Psychology* 1999;27:677–714.
- Brown CH, Wang W, Kellam SG, Muthén BO, Petras H, Toyinbo P, et al. Methods for testing theory and evaluating impact in randomized field trials: Intent-to-treat analyses for integrating the perspectives of person, place, and time. *Drug and Alcohol Dependence* 2008;95S:S74–S104. [PubMed: 18215473]
- Brown CH, Wyman PA, Brinales JM, Gibbons RD. The role of randomized trials in testing interventions for the prevention of youth suicide. *International Review of Psychiatry* 2007;19:617–631. [PubMed: 18092240]
- Carlin JB, Wolfe R, Brown CH, Gelman A. A case study on the choice, interpretation, and checking of multilevel models for longitudinal, binary outcomes. *Biostatistics* 2001;2:397–416. [PubMed: 12933632]
- Cooper, HM. *Synthesizing research: A guide for literature reviews*. 3. Thousand Oaks, CA: Sage; 1998.
- Dawson-McClure SR, Sandler IN, Wolchik SA, Millsap RE. Risk as a moderator of the effects of prevention programs for children from divorced families: A six-year longitudinal study. *Journal of Abnormal Child Psychology* 2004;32:175–190. [PubMed: 15164859]

- DeGarmo DS, Patterson GR, Forgatch MS. How do outcomes in a specified parent training intervention maintain or wane over time? *Prevention Science* 2004;5:73–89. [PubMed: 15134313]
- Durlak JA, Wells AM. Primary prevention mental health programs for children and adolescents: A meta-analytic review. *American Journal of Community Psychology* 1997;25:115–152. [PubMed: 9226860]
- Eddy JM, Reid JB, Fetrow RA. An elementary school-based prevention program targeting modifiable antecedents of youth delinquency and violence: Linking the Interests of Families and Teachers (LIFT). *Journal of Emotional & Behavioral Disorders* 2000;8:165–176.
- Faraone SV, Brown CH, Glatt SJ, Tsuang MT. Preventing schizophrenia and psychotic behaviour: Definitions and methodological issues. *Canadian Journal of Psychiatry* 2002;47:527–537.
- Goldstein H, Yang M, Omar R, Turner R, Thompson S. Meta-analysis using multilevel models with an application to the study of class size effects. *Journal of the Royal Statistical Society Series C-Applied Statistics* 2000;49:399–412.
- Gould MS, Marrocco FA, Kleinman M, Thomas JG, Mostkoff K, Cote J, et al. Evaluating iatrogenic risk of youth suicide screening programs: A randomized controlled trial. *Journal of the American Medical Association* 2005;293:1635–1643. [PubMed: 15811983]
- Greenwood P, Model K, Rydell CP, Chiesa J. The economic benefits of diverting children from crime. *Challenge* 1996;39:42–44.
- Hawkins JD, Kosterman R, Catalano RF, Hill KG, Abbott RD. Promoting positive adult functioning through social development intervention in childhood: Long-term effects from the Seattle Social Development Project. *Archives of Pediatrics & Adolescent Medicine* 2005;159:25–31. [PubMed: 15630054]
- Hedges LV. Estimation of effect size from a series of independent experiments. *Psychological Bulletin* 1982;92:490–499.
- Hedges, LV.; Olkin, I. *Statistical methods for meta-analysis*. Orlando, FL: Academic Press; 1985.
- Horowitz JL, Garber J. The prevention of depressive symptoms in children and adolescents: A meta-analytic review. *Journal of Consulting & Clinical Psychology* 2006;74:401–415. [PubMed: 16822098]
- Hox, JJ. The multilevel approach to meta-analysis. In: Hox, JJ., editor. *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Erlbaum; 2002. p. 139–156.
- Jané-Llopis E, Hosman C, Jenkins R, Anderson P. Predictors of efficacy in depression prevention programmes. Meta-analysis. *British Journal of Psychiatry* 2003;183:384–397. [PubMed: 14594912]
- Kellam SG, Brown CH, Poduska J, Ialongo N, Wang W, Toyinbo P, et al. Effects of a universal classroom behavior management program in first and second grades on young adult behavioral, psychiatric, and social outcomes. *Drug and Alcohol Dependence* 2008;95S:S5–S28. [PubMed: 18343607]
- Lipsey, M.; Wilson, D. *Practical meta-analysis*. Thousand Oaks, CA: Sage; 2001.
- Mann T, Nolen-Hoeksema S, Huang K, Burgard D, Wright A, Hanson K. Are two interventions worse than none? Joint primary and secondary prevention of eating disorders in college females. *Health Psychology* 1997;16:215–225. [PubMed: 9152699]
- MTA Cooperative Group. A 14-month randomized clinical trial of treatment strategies for attention-deficit/hyperactivity disorder. *Archives of General Psychiatry* 1999a;56:1073–1086.
- MTA Cooperative Group. Moderators and mediators of treatment response for children with attention-deficit/hyperactivity disorder. *Archives of General Psychiatry* 1999b;56:1088–1096.
- Murray, DM. *Design and analysis of group-randomized trials*. New York: Oxford University Press; 1998.
- Muthén, B. Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In: Kaplan, D., editor. *The Sage handbook of quantitative methodology for the social sciences*. Thousand Oaks, CA: Sage; 2004. p. 345–368.
- Muthén, BO.; Asparouhov, T. Growth mixture modeling: Analysis with non-Gaussian random effects. In: Fitzmaurice, G.; Davidian, M.; Verbeke, G.; Molenberghs, G., editors. *Longitudinal data analysis: A handbook of modern statistical methods*. Boca Raton, FL: Chapman & Hall/CRC Press; 2008. p. 143–165.
- Muthén, LK.; Muthén, BO. *Mplus user's guide*. 5. Los Angeles: Muthén & Muthén; 2008.

- Nylund KL, Asparouhov T, Muthen BO. Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling* 2007;14:535–569.
- Olds DL, Eckenrode J, Henderson CR Jr, Kitzman H, Powers J, Cole R, et al. Long-term effects of home visitation on maternal life course and child abuse and neglect. Fifteen-year follow-up of a randomized trial. *Journal of the American Medical Association* 1997;278:637–643. [PubMed: 9272895]
- Petrosino A, Boruch RF, Rounding C, McDonald S, Chalmers I. The Campbell Collaboration Social, Psychological, Educational and Criminological Trials Register (C2-SPECTR) to Facilitate the Preparation and Maintenance of Systematic Reviews of Social and Educational Interventions. *Evaluation and Research in Education* 2000;14:206–219.
- Raudenbush, SW. Random effects model. In: Cooper, H.; Hedges, L.V., editors. *The handbook of research synthesis*. New York: Russell Sage Foundation; 1994. p. 301–321.
- Raudenbush SW. Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods* 1997;2:173–185.
- Raudenbush SW, Bryk AS. Empirical Bayes metaanalysis. *Journal of Educational Statistics* 1985;10:75–98.
- Rosenthal R, Rubin DB. Meta-analytic procedures for combining studies with multiple effect sizes. *Psychological Bulletin* 1986;99:400–406.
- Rubin DB. Meta-analysis: Literature synthesis or effect size surface estimation? *Journal of Educational Statistics* 1992;17:363–374.
- Sandler IN, Ayers TS, Wolchik SA, Tein JY, Kwok OM, Lin K, et al. Family Bereavement Program: Efficacy of a theory-based preventive intervention for parentally-bereaved children and adolescents. *Journal of Consulting and Clinical Psychology* 2003;71:587–600. [PubMed: 12795581]
- Schwarz G. Estimating the dimension of a model. *Annals of Statistics* 1978;6:461–464.
- Schwartz S. The fallacy of the ecological fallacy: The potential misuse of a concept and the consequences. *American Journal of Public Health* 1994;84:819–824. [PubMed: 8179055]
- The Cochrane Collaboration. *The Cochrane Manual*. 2007. [updated May 23, 2007]. Retrieved July 12, 2007, from <http://www.cochrane.org/admin/manual.htm>
- Tobler NS, Roona MR, Ochshorn P, Marshall DG, Streke AV, Stackpole KM. School-based adolescent drug prevention programs: 1998 meta-analysis. *Journal of Primary Prevention* 2000;20:275–336.
- Wang CP, Brown CH, Bandeen-Roche K. Residual diagnostics for growth mixture models: Examining the impact of a preventive intervention on multiple trajectories of aggressive behavior. *Journal of the American Statistical Association* 2005;100:1054–1076.
- Wilson SJ, Lipsey MW, Derzon JH. The effects of school-based intervention programs on aggressive behavior: A meta-analysis. *Journal of Consulting and Clinical Psychology* 2003a;71:136–149. [PubMed: 12602434]
- Wilson SJ, Lipsey MW, Soydan H. Are mainstream programs for juvenile delinquency less effective with minority youth than majority youth? A Meta-analysis of outcomes research. *Research on Social Work Practice* 2003b;13:3–26.
- Wolchik SA, Sandler IN, Millsap RE, Plummer BA, Greene SM, Anderson ER, et al. Six-year follow-up of preventive interventions for children of divorce: A randomized controlled trial. *Journal of the American Medical Association* 2002;288:1874–1881. [PubMed: 12377086]
- Wolchik, S.; Sandler, IN.; Weiss, L.; Winslow, E. *New Beginnings: An empirically-based intervention program for divorced mothers to promote resilience in their children*. In: Briesmeister, JM.; Schaefer, CE., editors. *Handbook of parent training: Helping parents prevent and solve problem behaviors*. 3. New York: Wiley; 2007. p. 25–66.

APPENDIX

Mplus Code for Multilevel Mixture Models

The following annotated Mplus code (Muthén & Muthén, 2008) illustrates the use of two-level modeling and mixture modeling in the same model. The two-level model uses WITHIN and BETWEEN keywords to identify the levels for the respective variables, and CLUSTER is used

to signify that trials have multiple outcomes. The keyword CLASSES identifies the name of the discrete mixture variable. In the ANALYSIS statement, the TWOLEVEL and MIXTURE keywords correspond to a multilevel mixture model, with RANDOM indicating that a continuous random slope model is used. A number of parameters in other ANALYSIS statements are included to make certain a full global search is performed. The specific model is determined in the MODEL statement. Here the keywords %WITHIN% and %BETWEEN% deal with the two distinct levels, whereas %OVERALL%, %Cls#1%, and %Cls#2% identify parts of the mixture models that pertain to both classes or each one separately. Regression models are use the ON keyword, for example, ES ON Time regresses the effect sizes on the within level variable Time. Random slopes are phrased as b | ES ON SE; this introduces a continuous random variable that is multiplied by the respective standard error for each separate outcome. The restrictions [b @ 0]; and b @ 1; force this random slope to have mean zero and variance 1, therefore compensating for the within variance for each analysis. Finally, parenthetical expressions such as [ES] (equate1); and similar terms allow one to equate model parameters across classes.

VARIABLE:

WITHIN= ! Define covariates at level of outcome

Time

SE

Cls; ! Cls is indicates the class in a mixture model

BETWEEN= ! Define covariates at level of trial

Rural;

CLUSTER= ! Identify variable representing higher level

TrialNo;

CLASSES= ! Identify name of class variable for mixture

Cls (2); ! Cls has two unobserved categories

ANALYSIS:

TYPE= ! Define type of analysis to perform

TWOLEVEL ! Within and between covariates allowed

RANDOM ! Include random slopes in the model

MIXTURE; ! Include discrete mixtures

! Technical specification of fitting method

ALGORITHM=INTEGRATION;

ESTIMATOR=ML;

STARTS=500 5; ! Large number of searches

LRTSTARTS=2 1 120 10;
 MITERATIONS=4000;
 STITERATIONS=20;
 MODEL: ! Terms included in the model
 %WITHIN% ! Terms at the outcome level
 %OVERALL% ! Common model for both classes
 b | ES ON SE; ! Predict ES with random slope an
 ES ON Time; ! Regress ES on outcome level
 %Cls#1% ! Additional terms for first class
 ES ON Time; ! Regress ES on outcome level
 %Cls#2% ! Additional terms for first class
 ES ON Time; ! Regress ES on outcome level vars
 %BETWEEN% ! Between Level Predictors
 %OVERALL% ! Same terms for both classes
 [b @ 0]; ! Fix mean of random slopes to 0
 b @ 1; ! Fix variance of random slopes to
 [ES];
 ES;
 ES ON Rural; ! Regress ES on rural
 %Cls#1% ! Additional terms for first class
 ES ON Rural; ! Class 1 regression of ES on Rural
 [ES] (equate1); ! Equate variances in the 2 classes
 ES (equate2);
 %Cls#2% ! Additional terms for second class
 ES ON Rural; ! Class 2 regression of ES on Rural
 [ES] (equate1);
 ES (equate2);

Figure 1.
Density of effect sizes for the rural subset of Tobler School Drug Prevention Programs.