

Methodological Challenges Examining Subgroup Differences: Examples from Universal School-based Youth Violence Prevention Trials

Albert D. Farrell · David B. Henry · Amie Bettencourt

Published online: 28 January 2011
© Society for Prevention Research 2011

Abstract This article reviews the literature on school-based universal violence prevention programs to illustrate key methodological challenges for investigating subgroup differences in prevention effects. The variety of potential moderating factors examined within this literature is discussed within the context of a social-ecological model. Our review of this literature identified the following methodological issues: the need for a clear a priori theoretical basis for selecting potential moderators, inflated Type I error rates that result from large numbers of comparisons, the absence of explicit tests of moderation, interpretive issues arising from a restricted range on moderator variables, the failure to report effect size estimates, the presence of potential confounding factors, and the importance of examining factors that might operate at multiple ecological levels. These points are illustrated using examples of studies, primarily within youth violence prevention research, that have

identified factors within the individual, school, and community that moderate the outcomes of preventive interventions. We conclude with general recommendations for future work. These include the benefits of using the social-ecological model to provide a basis for moving from exploratory to more theory-driven confirmatory models of subgroup differences, the potential merits of qualitative research designed to identify factors that may influence the effectiveness of intervention efforts for specific subgroups of individuals, and the provision of effect size estimates and confidence intervals for effect sizes in prevention reports.

Keywords Prevention · Moderation · Subgroup analyses · Violence prevention

Preventive interventions are designed to disrupt the development of a disorder by reducing exposure to risk factors and strengthening promotive and protective factors (Coie et al. 1993). To be effective, such programs must address the factors most relevant to their target populations (Davis et al. 2003). Because most disorders are multiply determined, individuals within a target population may vary in their exposure to risk factors, and their levels of promotive and protective factors (Coie et al. 1993). This is particularly true for universal prevention programs aimed at a broad population. Because prevention programs are planned and implemented within social contexts such as schools, classrooms, or neighborhoods, factors that vary among these contexts such as group norms, cultural beliefs, family characteristics, neighborhood collective efficacy, and other setting factors also are likely to influence their effectiveness. These differences may result in considerable variability in outcomes across individuals targeted by the same prevention strategy. Identification of subgroups in the population and associated factors related to differences in

This research was supported by cooperative agreements U49/CE000730 (VCU) and U49/CE000732 (UIC) from the Centers for Disease Control and Prevention (CDC). The research and interpretations reported are the sole responsibility of the authors and are not necessarily endorsed by CDC or represent the views, opinions, or policies of the CDC or their staff.

A. D. Farrell (✉)
Department of Psychology, Virginia Commonwealth University,
P.O. Box 842018, Richmond, VA 23284-2018, USA
e-mail: afarrell@vcu.edu

D. B. Henry
School of Public Health, University of Illinois at Chicago,
Chicago, IL, USA
e-mail: dhenry@uic.edu

A. Bettencourt
Bloomberg School of Public Health, Johns Hopkins University,
Baltimore, MD, USA
e-mail: bettencouaf@gmail.com

outcomes can improve the effectiveness of prevention efforts by targeting them at the individuals who are most likely to benefit (Yale et al. 2003), and by guiding the development of interventions to meet the needs of individuals who do not benefit from existing interventions.

The importance of examining individual differences in intervention outcomes has long been recognized. In his classic 1966 article on psychotherapy outcome research, Kiesler challenged psychotherapy researchers to conduct factorial studies to identify the most effective strategies for producing change in specific subgroups of patients. Paul (1967) similarly argued that “The question to which all outcome research should ultimately be directed is the following: *What* treatment, by *whom*, is most effective for *this* individual with *that* specific problem, and under *which* set of circumstances?” (p. 111). Although the basic problem has long been identified, addressing it has presented significant challenges.

Variability in intervention effects is particularly likely for preventive interventions because they are typically directed at populations rather than at individuals, as in traditional psychotherapy. This mode of delivery is likely to lead to heterogeneous subgroups receiving the same intervention. The importance of examining the consistency of intervention effects across subgroups was incorporated into the Standards of Evidence for identifying effective prevention programs adopted by the Society for Prevention Research (Flay et al. 2005). The *Standards* suggests that subgroup analyses be conducted on heterogeneous samples with respect to variables such as age, gender, ethnicity/race, and risk levels. Such analyses were considered central to determining the extent to which the effects of a specific intervention may be generalizable.

This article reviews studies evaluating universal school-based violence prevention programs as a context for discussing key challenges to subgroup analysis and to form the basis of recommendations for addressing them. Consistent with the focus of this special issue, this article will address methodological issues evident within this work rather than on more substantive conclusions regarding specific factors that moderate intervention effects.

Studies of Subgroup Differences in Universal School-based Violence Prevention Effects

Youth violence is a complex phenomenon determined by multiple factors operating at different levels of the social ecology. Because no single theory or developmental pathway adequately accounts for youth violence (Flannery et al. 2007; Lipsey and Derzon 1998), specific prevention programs are unlikely to achieve consistent effects across different individuals or social settings. Prevention efforts

themselves vary in their targeted populations and processes, including programs designed to address individual-level factors and programs that focus on social structures such as peers, schools, and families (Farrell and Vulin-Reynolds 2007; US Department of Health and Human Services [USDHHS] 2001), through universal interventions administered to entire populations (e.g., school-wide curricula), selective interventions targeted at high risk youths, and indicated programs for youths displaying elevated levels of aggression (Farrell and Vulin-Reynolds 2007).

For this article, we identified studies examining subgroup differences in effects of youth violence prevention programs using several sources. We began by searching PsychInfo using the terms *violence*, *aggression*, or *bullying* paired with *prevention*. We added studies cited in 20 youth violence prevention literature reviews and additional studies identified by members of the research team. This process identified a total of 300 studies evaluating the effects of youth violence prevention programs, of which 130 examined subgroup effects. These studies focused on individual and setting factors that influenced intervention impact. We did not include studies that examined differences related to quality or fidelity of implementation. We also restricted our review to universal programs implemented alone or in combination with selective interventions because they are typically directed at fairly heterogeneous populations and are thus particularly likely to be susceptible to subgroup differences. This eliminated 36 studies that focused exclusively on selective or indicated interventions. Most universal violence prevention interventions are implemented in school settings (Farrell et al. 2001a). We therefore restricted our review to this setting to avoid additional complexities that might be unique to other settings. This eliminated one study conducted in a community setting, and three that evaluated preschool interventions. Because we were interested in effects on the general population of intervention participants, we eliminated six studies that restricted their focus to effects of combined universal and selective interventions on youth meeting criteria for the selective intervention. We eliminated five studies that restricted their analyses of moderating effects to other outcomes (i.e., social competence, condom use, beliefs) because we were interested in factors that moderated intervention effects on problem behaviors. Finally, we eliminated 15 studies that examined factors influencing changes within an intervention group without any comparison group. This process led to the identification of 68 studies¹ that examined subgroup differences in evaluations of universal school-based violence prevention programs. Although this is likely not a comprehensive list of all studies that have been conducted

¹ A table listing the studies included in the review may be obtained from the first author.

in this area, it provides a reasonable sample of those that have appeared in the literature for the purpose of highlighting methodological issues.

We used Bronfenbrenner's (1979) social-ecological model as a framework to organize our discussion. This model recognizes that an individual's behavior is influenced by his or her personal characteristics and social environment. This model is a key component of the Center for Disease Control and Prevention's framework for prevention (Dahlberg and Krug 2002) which considers the influence of individual factors (e.g., biological and personal history), interpersonal relationships (e.g., peers and family), community influences (e.g., schools, neighborhoods), and society (e.g., societal norms, social policies). Dahlberg and Krug argued for the development of prevention efforts that address multiple levels within this model. This suggests the need to understand how patterns of risk and protective factors within individuals and in their social structures might create subgroups for whom the outcomes of prevention efforts differ.

Individual-level Factors as Moderators of Prevention Effects

The wide variation in individual-level factors associated with youth violence makes it highly unlikely that any single prevention program implemented on a school-wide basis will produce uniform effects. Researchers have examined differences in the effects of universal school-based violence prevention programs across subgroups that differ on a variety of individual-level factors believed to moderate intervention effects. These include demographic characteristics (e.g., gender, ethnicity, age), initial levels of aggression and other problem behaviors, risk factors, and degree of participation in the intervention.

Demographic Characteristics Of the 68 studies we identified, 54 examined differences across one or more demographic variables. Most (50 studies) examined gender differences, followed by age or grade differences (16), race or ethnic differences (15), poverty status (2), and English proficiency (1). These studies often did not provide a clear rationale for subgroup analyses. Some cited gender or developmental differences in patterns of aggression as a reason for examining subgroup differences in effects (e.g., Griffin et al. 2007). In other cases, researchers used such analyses to support claims of consistent effects across demographic subgroups. For example, based on finding few subgroup differences in intervention effects, the Conduct Problems Prevention Research Group (CPPRG 1999b) concluded: "Evidence of differential intervention effects across child gender, race, site, and cohort was minimal" (p. 648), and Aber et al. (2003) interpreting similar findings in

their study, concluded: "Interactions with gender or family socioeconomic status (school lunch eligibility) were negligible and not above a rate expected by chance. Significant interaction effects for race/ethnicity were few and weak and lacked a discernible pattern." (p. 341).

Overall, our review did not find a sufficiently consistent pattern of support for moderation across specific demographic subgroups to warrant drawing even general conclusions. Even in those studies where differences across variables such as gender were found, these effects were rarely consistent across all outcomes or waves of data (e.g., Farrell and Meyer 1997). Moreover, the wide variety in types of interventions, research designs, study populations, and outcomes examined in these studies further precludes drawing any general conclusions.

Individual Differences at Baseline A total of 20 of the 68 studies examined the extent to which intervention effects varied as a function of scores on pretest measures of aggression or related indicators. This number does not include several studies (e.g., CPPRG 1999a; Metropolitan Area Child Study Research Group [MACS] 2002) that conducted analyses of intervention effects on youth with high levels of aggression but did not compare these results to a low aggression subgroup. The rationale for examining pretest aggression as a moderator was often based on predicted differences in the responses expected for individuals at different levels of aggression. Stoolmiller et al. (2000), for example, hypothesized that moderately aggressive children would be most likely to respond to a universal school-based violence prevention program because those at low levels of aggression have little room for improvement and the intervention may not be intense enough to produce change in children at high levels of aggression. In some cases researchers examined the moderating effects of pretest aggression as a continuous variable (e.g., Stoolmiller et al. 2000). In three cases these analyses were conducted as part of a more general strategy in which pretest levels of each outcome measure were included as a moderator of intervention effects on that outcome (e.g., Reid et al. 1999). Others coded baseline aggression categorically using cutoffs based on the distribution of scores within their sample (e.g., Foshee et al. 2005; Tolan et al. 2004). Perhaps the most sophisticated approach was taken by Segawa and colleagues (2005) who used growth mixture modeling to examine differences in intervention effects for distinct classes of individuals that differed in their growth trajectories of change in aggression.

Regardless of the method used, the majority of studies (i.e., 12 of 17 unique intervention trials) found greater benefit for individuals at higher levels of initial aggression on one or more outcomes (e.g., Farrell et al. 2001b; Reid et al. 1999). However, as with studies of demographic

variables, a consistent pattern of moderation was not generally found across all outcomes, waves of data, or grade levels. In contrast, Foshee et al. (2005) found stronger effects for a dating violence prevention program among adolescents who reported lower baseline levels of severe violence perpetration, but similar effects were not found among those reporting high baseline levels of perpetration. Four studies (e.g., Smokowski et al. 2004) found no difference in intervention effects across levels of aggression. One methodological issue not addressed in many of these studies concerns the potential role of other variables correlated with baseline aggression. For example, although less consistent support has been found for moderating effects of gender and ethnicity, it would seem appropriate to control for Gender \times Intervention and Ethnicity \times Intervention effects to rule out the possibility that these factors are responsible for any observed moderation. We found one study that included these interaction terms (i.e., Smokowski et al. 2004), but the model controlled for the moderating effects of pretest aggression before, rather than after accounting for gender and ethnic differences in intervention effects.

We identified two outcome studies that examined the extent to which the effects of prevention programs varied as a function of composite measures of individual-level risk factors. Aber and colleagues (1998) conducted a quasi-experimental study in which they compared results for schools that differed in their degrees of implementing a universal violence prevention program. Within their study they evaluated the extent to which a composite measure of risk based on students' level of depression and scores on achievement tests moderated the effects of different degrees of implementation. There were no significant Risk Composite \times Implementation Profile interactions on the four measured outcomes—interpersonal negotiation strategies, attributional bias, aggressive fantasies, and conduct problems. Two recent reports (Multisite Violence Prevention Project [MVPP] 2008, 2009) evaluating the relative and combined efficacy of a school-based universal intervention and a selective family intervention investigated the extent to which intervention effects were moderated by a risk index constructed from ten individual-level variables representing social-cognitive variables, peer influences, and parental influences. Risk factors were drawn from an initial set of 13 variables based on their ability to predict changes in aggression after controlling for gender, ethnicity, family structure, and site. Because boys reported a higher number of risk factors than girls, the analyses controlled for Gender \times Intervention interactions. Analyses revealed a linear relation between the number of risk factors and effects of the universal intervention on one of three measures of aggression, and on both overt and relational victimization (MVPP 2009). Across all three outcomes, students with

high levels of pretest risk at intervention schools had lower posttest scores on aggression than their counterparts at control schools. In contrast, those with low levels of pretest risk at intervention schools had higher posttest scores on aggression than their counterparts at control schools. A similar pattern of risk moderation was found in the effects of the universal intervention on social-cognitive processes targeted by the intervention (MVPP 2008).

Family Factors Although much of the literature evaluating family factors as moderators of intervention effects has focused on selective family interventions (e.g., Dishion et al. 2002), it is reasonable to assume that family factors might also influence how individuals respond to other intervention modalities. Parental factors such as monitoring and involvement, parental support for fighting, and parental support for nonviolence have been found to exert direct effects on adolescents' aggression, and to serve a protective function by moderating the effects of peer and school risk factors (Farrell et al. 2011). Given the salience of parental influences, it seems quite plausible that the extent to which parental influences support or oppose the goals of school-based interventions would impact their effectiveness.

We identified three studies that evaluated family variables as a moderator of the effects of universal school-based violence prevention interventions. The SAFEChildren project evaluated the effects of a year of academic tutoring coupled with a 22-session group-based family intervention designed to enhance parenting practices and family functioning and to improve parents' relationships with their children's schools (Tolan et al. 2004). This approach was implemented as a universal intervention in high-risk neighborhoods. Family risk was based on measures assessing parenting practices and family relationships. Separate analyses of the high risk-families revealed reduced child aggression, increased child concentration, and increased parental monitoring among those randomly assigned to receive the intervention. A second study by Reid et al. (1999) evaluated the extent to which the impact of a school-based prevention program for conduct problems was moderated by mothers' level of aversive verbal behavior. Mothers with high initial levels of aversive verbal behavior changed the most, but, as in the SafeChildren study, explicit tests of moderation by family variables were not conducted.

Spoth et al. (1998) examined the extent to which the impact of a universal school-based family intervention on proximal outcomes (i.e., parenting behaviors and response to peer pressure) was moderated by family risk factors using data from two outcome studies conducted with rural populations. In one study families in nine schools were randomly assigned to an intervention or control group. The second study randomly assigned 33 schools to intervention and control conditions. Based on theory and prior research

the investigators hypothesized that outcomes for higher-risk families would be equal to or stronger than those for lower-risk families. Families in their study were divided into five risk groups based on a risk index constructed from family variables including demographic characteristics (e.g., family structure, family income and parents' education), and measures of mothers', fathers', and adolescents' emotional adjustment. Significant overall main effects were found for parenting behaviors, but not for response to peer pressure. These effects were not moderated by family risk.

Participation or Dosage It is reasonable to assume that individuals who do not fully participate in an intervention are less likely to experience its benefits. Participation rates or dosage may vary at both the classroom or school level, and be related to a variety of individual characteristics. Unlike individual-level characteristics such as demographic variables and levels of risk factors that may be equally distributed across treatment conditions through random assignment, individuals are typically not randomly assigned to dosage. This makes it challenging to identify suitable comparison groups that match those at high and low levels of participation on potential confounding variables. Recent applications of statistical models (Jo 2002) have provided approaches that address this potential bias by identifying subgroups of individuals within the control group that resemble those in the intervention group that differed in their level of participation. This approach has the further advantage of providing valuable information about the characteristics of individuals who are least likely to participate in the intervention. Stuart et al. (2008), investigating the impact of the Family-School Partnership Intervention conducted with first grade students, found effect sizes nearly twice as large when the sample was restricted to a comparison of those in the intervention and control conditions classified as participants. These methods could be of particular value to researchers examining subgroup differences in prevention programs. In particular, they provide an opportunity to determine the extent to which degree of participation in the intervention may serve as the underlying mechanism through which a variety of previously identified individual-level variables influence intervention impact.

Moderators at Ecological Levels Beyond the Individual

In contrast to research examining individual-level moderators of universal school-based violence prevention interventions, few studies have examined moderators above the individual level of analysis. A variety of school-level factors have the potential to influence the impact of prevention programs implemented in school settings. The school environment may counteract the impact of preven-

tion programs designed to change norms by creating informal social norms according to which aggression is associated with higher social status (Fagan and Wilkinson 1998) and by providing exposure to deviant peers (Dishion et al. 1994). Risk factors such as normative beliefs, and peer and family influences may influence youth not only at the individual level, but also may cluster to create a strong influence at the school level (Henry et al. 2010). Moreover, school policies and staff may reinforce or discourage behaviors that are the focus of intervention efforts. Finally, school factors may affect intervention impact through fidelity of program implementation. Studies have found that organizational factors affect implementation (Gottfredson et al. 1993; Gregory et al. 2007). Program developers, in particular, have emphasized the importance of factors such as school readiness and staff commitment (Meyer et al. 2000b) that may influence quality of implementation and ultimately their impact.

Classroom and School-level Variables that Moderate Prevention Effects Patterns of risk and protective factors that represent the shared experiences of students within the same classrooms or attending the same schools may also influence the impact of prevention efforts. For example, Dishion et al. (1999) described the negative impact that can occur when interventions are conducted in small groups of deviant peers. Although such effects may be most relevant to selective interventions, researchers have also examined the extent to which peer factors at the classroom and school levels influence the impact of universal interventions. We identified three studies that examined the extent to which classroom norms moderate the impact of school-based violence prevention programs.

Kellam et al. (1998) hypothesized that students in first grade classrooms with higher levels of aggression would benefit most from the Good Behavior Game, a universal, classroom-based intervention. Tests of this hypothesis within the context of a randomized trial involving 18 schools and 40 classrooms found support for this pattern for boys, but not for girls. This highlights the possibility that individual and contextual factors might jointly moderate the effects of universal interventions.

Aber et al. (1998) found partial support for the hypothesis that effects of a universal prevention curriculum would be weaker in classrooms where there were strong norms supporting the use of aggression. In particular, they found the clearest intervention effects on hostile attribution biases and aggressive fantasies in classrooms in which the prevailing belief was that aggression was wrong. Similar differences in effects were not found on aggression or interpersonal negotiation strategies.

Support for school characteristics as a moderator of intervention effects was also found by Henry and Schoeny

(2007), who tested the extent to which different aspects of school norms supporting aggression or nonviolent alternatives to aggression moderated the effects of the universal school-based interventions implemented within MACS and MVPP. In the MVPP study, class norms at sixth grade entry affected subsequent levels of aggression in control schools, but exerted no effect at schools assigned to receive the universal intervention, suggesting that defining subgroups by norms may be helpful in understanding why some studies find strong effects for these interventions and others do not.

We found only one other study that examined a specific school-level variable as a moderator of intervention effects. Kam et al. (2003) provided a clear argument regarding the importance of principal leadership as a key factor influencing the impact of a school-based intervention. They compared the impact of the PATHS curriculum in three intervention and three control schools. They reported significant interactions between level of implementation, degree of principal support and intervention effects such that higher levels of implementation produced significant effects on several outcomes, but only when level of principal support was high. This study was, however, limited in that only three schools were included and the moderating impact of principal support was based on comparing changes in outcomes for students at the school with the highest versus the lowest level of support.

Differences in Effects Across School Settings The majority of studies examining school variables as moderators of the effects of universal school-based interventions have involved simple comparisons of the intervention effects across participating schools or sites, at times focusing on site differences in poverty, to explore the consistency of effects. For example, Fraser and colleagues (2005) conducted a social-emotional skills intervention with third graders in two schools that differed in SES levels and ethnic composition. They compared cohorts that received different interventions within the same school to evaluate the impact of the curriculum, but found few significant school differences in intervention effects, leading them to conclude that the effects of the intervention were robust.

Leadbeater et al. (2003) examined moderation of intervention effects by school levels of poverty. The intervention was designed to reduce victimization and promote social competence. In a quasi-experimental design, 11 schools that had successfully implemented the intervention were compared to 5 other schools. The intervention was associated with lower physical victimization in schools with average or high levels of poverty. Similar effects were not found in schools at low levels of poverty.

The CPPRG (1999b) evaluated the effects of a combined universal and selective intervention on first graders in a

large study involving approximately 12 schools per site in four sites that differed in location, ethnic composition, and income levels. Analyses did not identify any significant Site \times Intervention interactions leading the authors to conclude that there were: “no major differences in effects of intervention as a function of rural versus urban school location, percentage of children below the poverty level, or ethnic composition of the classrooms.” (p. 655). In a more recent report examining impacts of 3 years of intervention on longer-term outcomes, the CPPRG (2010) conducted a more direct test of the moderating impact of school disadvantage as measured by the percentage of students who qualified for free or reduced lunch at each school. They hypothesized that stronger effects would occur in schools low on school disadvantage based on negative influences associated with disadvantaged schools. These hypotheses were partially supported by their findings of stronger intervention effects on teacher ratings of student problem behaviors in less disadvantaged schools. Similar effects were not found on peer sociometric ratings of aggression and hyperactivity.

Neighborhood Factors as Moderators of Prevention Effects

Risk at the community level also has consequences for how preventive interventions are implemented and for whom they have effects. A variety of community level factors that place youth at risk for violence have been identified including poverty, community disorganization, and high rates of crime and drug use (Hawkins et al. 1998). The extent to which community factors moderate intervention effects is rarely examined within an individual study because of the resources such an undertaking would require. Community factors are more typically addressed by attempting to match interventions to community needs. Yale et al. (2003), for example, argued for use of a developmental epidemiological approach focusing on neighborhood rather than individual level risk to identify neighborhoods to target for prevention efforts. Because many of the characteristics of students within a given school reflect the neighborhoods they serve, studies in the preceding section that examined variables such as school disadvantage also reflect the moderating impact of neighborhood factors. In this section, we describe two studies that more directly examined the moderating influence of neighborhood factors.

Aber et al. (1998) hypothesized that poor and dangerous neighborhoods would place children at greater risk for violence and consequently make it more difficult for a school-based intervention to have a positive effect. They examined the impact of neighborhood homicide and poverty rates within the context of their quasi-experimental

evaluation of the impact of four levels of program implementation at 26 elementary schools in New York City. Their analyses revealed mixed support for their hypotheses. The positive impact of the intervention on children's social cognitions was evident for children in neighborhoods characterized by low to medium rates of both homicide and poverty, but positive effects were not found for those in neighborhoods with high homicide and poverty rates. Similar moderating effects were not found for interpersonal behaviors.

The MACS (2002) study evaluated three increasingly intensive intervention strategies in schools located in poor, high crime neighborhoods in a large city, and schools in impoverished areas of a smaller city. Because this study examined effects on a high-risk sample it was not included in the original review, but is discussed here because of the limited number of studies that have examined the moderating effects of neighborhood variables. The interventions were a universal curriculum-based classroom intervention (Level A), the Level A intervention plus a small group social skills training intervention for youth at elevated risk for aggression (Level B), and the Level B intervention plus a group-based intervention for families of higher risk youth (Level C). The Level C intervention was found to be effective in reducing child aggression, but only when offered in second and third grades in schools in the less impoverished communities. Such moderation of effects does not appear to have been attributable to school-level differences in resources. This is apparent because the intervention found to be effective among younger children in lower-risk neighborhoods was a multi-context intervention that did not take place primarily in the school. It appears, rather, that moderation of the effects of the MACS intervention was a function of community level risk associated with levels of poverty and crime.

Challenges and Recommendations

Our review identified a variety of limitations and challenges faced by researchers examining subgroup differences in the effects of universal school-based violence prevention programs. In this section, we discuss these and provide some recommendations for improving work in this area. Specific challenges involve issues related to the role of theory, research design, and statistical analyses.

Theory and Measurement

Although there were exceptions, relatively few studies of subgroup differences provided an explicit theoretical rationale for why intervention effects would be expected to differ across the factors they examined. For example, the finding that

subgroup differences had been encountered in previous studies was the most common justification for examining differences across subgroups defined by gender, and race/ethnicity. These studies often examined multiple potential moderating variables in separate analyses of multiple outcomes and in some cases across multiple waves of data. In other cases researchers examined subgroup differences in an effort to show that intervention effects were consistent across different subgroups. As Smith and Sechrest (1991) argued in their discussion of psychotherapy outcome research, important moderators of treatment effects are most likely to be discovered through “deliberate tests of theoretically driven a priori hypotheses” (p. 242) rather than post hoc analyses of a myriad of potential factors.

Most examinations of subgroup differences in the effects of universal school-based violence prevention programs have focused on individual-level demographic factors and baseline levels of aggression. More could be gained by examining group differences based on intervention logic models that explicate the specific patterns of risk and protective factors an intervention is designed to address, the mechanisms used to address these factors, and individual and contextual factors believed to influence intervention effects. These logic models can provide specific hypotheses regarding those individuals most likely to benefit from a given intervention approach, the broader contextual factors that may moderate outcomes, and the mechanisms responsible for variability in outcomes across individuals, families, schools, and communities.

For example, interventions such as Responding in Peaceful and Positive Ways (RIPP; Meyer et al. 2000a) and Second Step (Frey et al. 2000) focus on teaching specific social-cognitive skills such as problem solving, emotion regulation, and empathy. The logic model implicit in such interventions is that these skills are related to the development and maintenance of aggressive behavior, that participants targeted by the intervention have deficiencies in these skills, that their participation in the intervention will increase their levels of these skills, and that these changes will result in reductions in their subsequent use of aggression. Each of these assumptions represents a testable hypothesis. More specifically, data could be collected to establish the extent to which individuals within a specific target population show deficiencies in the skills targeted by the intervention. Measures of these skills could then be included in outcome batteries evaluating intervention effects to test the intervention's action theory by establishing whether the intervention increases levels of these skills, and its conceptual theory by determining the extent to which changes in these skills are related to reductions in aggressive behavior (see MacKinnon 2008).

Logic models also provide a framework for understanding the multiple mechanisms that might moderate intervention

effects. In contrast to indicated and selective interventions, universal interventions do not typically involve screening individuals for participation in the intervention. It is therefore likely that participants in interventions such as RIPP and Second Step will vary in the extent to which they have deficiencies in the skills targeted by these interventions. Their logic models would therefore predict that these interventions would be more effective for participants with deficiencies in these skills. Subgroup differences might also emerge in the extent to which specific instructional techniques produce their desired effects on the skills they target. For example, cultural factors might influence the degree to which participants find an intervention's focus and instructional techniques relevant (Wright and Zimmerman 2006). Contextual factors might also moderate the extent to which mastery of these skills is sufficient to produce the anticipated effects on aggression. For example, findings that interventions are less effective with students in classrooms where there are strong norms supporting aggression (e.g., Aber et al. 1998) suggest the need to direct more intensive efforts at measuring classroom norms and developing and implementing interventions designed to alter them so that they support the use of these skills. This suggests the need to construct logic models that not only specify the underlying mechanisms of change, but that also consider potential moderators that may influence each link within the models. Greater reliance on theory would move the field beyond simply identifying factors that moderate intervention effects toward an understanding of the underlying mechanisms responsible for these differences. For example, finding gender differences in intervention effects is less useful to developers of interventions than an understanding of the factors associated with gender that might be responsible for this differential effect (e.g., perceived peer norms, perceived consequences of nonviolent responses, differences in social orientation). Such an approach would inform the development of effective interventions and establish the limits of their generalizability.

Further work also is needed to generate hypotheses related to the underlying processes responsible for moderated effects. For example, Farrell and colleagues (2008, 2010) conducted a series of qualitative studies designed to identify factors at the individual, peer, school, family, and neighborhood levels that influenced the extent to which individuals would use effective nonviolent responses to conflict. This research identified important individual and contextual factors that would discourage individuals from responding nonviolently. These included peer factors such as concerns about status, family factors such as parents who support the use of aggression, and school factors such as teachers who would not respond to a student's request for help. The presence of these factors would presumably reduce the potential effectiveness of an intervention that focused on teaching individual-level skills. This could be

directly tested by determining if these factors moderate the impact of a specific intervention.

The number and variety of contextual responses obtained in the Farrell et al. (2010) study also suggest that researchers need to pay greater attention to the measurement of potential moderators. The social-ecological model provides a useful framework for suggesting potential contextual moderators of prevention effects. It has been known for many years that the influence of individual factors may depend on their prevalence in a group (Asch 1955). For example, Dodge et al. (2006) summarized the large literature supporting the hypothesis that group interventions tend to lead members to influence each other, with high-risk members benefiting and relatively low-risk members worsening. Grabosky (1996) also noted such unintended negative consequences in a variety of interventions to reduce crime and delinquency. Although the findings reviewed by Dodge et al. (2006) were based on selective and targeted intervention programs, there is evidence that such effects may also occur in universal interventions (e.g., MVPP 2008, 2009). This suggests the need to consider key individual-level characteristics such as normative beliefs and behavior not only at the individual-level, but also at the classroom or school level. This will require methods of measuring and quantifying characteristics of social settings that are suitable for higher levels of analysis (Tseng and Seidman 2007). Constructing measures of school or classroom characteristics requires first a definition of the construct that is appropriate for an organizational level of analysis (Shinn and Rapkin 2000). Constructs such as "beliefs" have different meanings when describing an individual and a group or organization.

Developing measurement approaches appropriate to the school or classroom level of analysis is also necessary. Observational approaches, such as those developed by Pianta and colleagues (2004), provide measures of classroom climate that are independent of individual reports and are excellent candidates for variables that define subgroups for which intervention effects may vary. The use of aggregated individual scores for organization-level measurement is common (Rousseau 1985) and some approaches to creating aggregated measures take variability in individual reports into account (e.g., Raudenbush and Sampson 1999), allowing the consensus within a school or classroom setting to be modeled along with the mean levels (e.g., Henry et al. 2010). Groups also provide incentives to their members that promote unanimity of beliefs and consistency of behavior, and they differ in the range of behaviors or beliefs that will be tolerated among their members (Jackson 1966). These additional characteristics can refine and enhance the measurement of organizational characteristics. For example, Henry and Chan (2010) found that adding the degree of consensus on norms for

nonviolence and the range of acceptable nonviolent behaviors to mean approval of nonviolence predicted variance in aggression and associated attitudes to a greater extent than did mean approval alone.

Research Design

Our review also identified a variety of issues related to the design of studies examining subgroup differences. Analyses of subgroup differences are conducted within the context of outcome studies designed to examine intervention effects. As such, they need to meet the same basic requirements of sound intervention studies in terms of their overall design, measurement issues, intervention fidelity, etc. (Farrell et al. 2001a; Tolan and Brown 1998). Our initial review of studies examining subgroup differences identified numerous studies with serious flaws. For example, we excluded 15 studies that examined differences in intervention effects without any comparison group.

Restricted range of variables defining subgroups is also a serious design barrier to understanding subgroup differences. Many of the studies we reviewed examined subgroup differences within the context of existing data sets. Because the parent studies were not typically designed with the intention of examining subgroup differences, they were often less than optimal in terms of the degree to which they sampled the distribution of the variable(s) defining the subgroups of interest. For example, finding no significant differences across individuals in a sample with a narrow range of family income levels is not sufficient to indicate that intervention effects are robust across individuals differing in family income levels. A restricted range will reduce the possibility of finding moderation and will also limit the extent to which the findings generalize to samples outside the observed range.

Consideration of school and community level variables as either factors directly moderating intervention effects, or as contextual factors that influence the role of individual-level moderators is particularly challenging. Subgroup studies often examined differences across schools or sites to determine if effects were consistent across a variety of factors such as urban versus rural location, ethnic composition, and income level (Aber et al. 1998; CPPRG 2010). For example, differences in intervention effects across the two communities examined in the MACS (2002) were attributed to differences in community level risk associated with levels of poverty and crime. However, the comparison of results across only two communities makes it difficult to rule out other potential differences that may have influenced intervention effects.

Few studies include a sufficiently large or diverse sample of schools or communities to provide the depth and scope needed to examine the moderating effects of

school and community characteristics on outcome processes. Considering the scale that would be required, it is unlikely that any single study will have such a scope. Influences at this level might be better addressed through approaches such as meta-analysis that compare the effects of interventions implemented in schools and communities that differ on important characteristics (e.g., Wilson et al. 2003). Such an approach will require a research base in which interventions are implemented in settings that differ on important risk characteristics. It will also require more careful assessment and consistent reporting of both individual-level characteristics and the characteristics of the school and community settings in which they are implemented. Archiving of research data consistent with recent changes in federal regulations, and the opportunity to provide extended tables of means, grouped by variables that define subgroups, should facilitate building a suitable research base for meta-analytic investigation of subgroup differences in effects. However, as Lipsey (2003) noted, there are serious complicating factors involved in investigating moderators in meta-analysis.

The social-ecological model that guided the development of many violence prevention programs also has important implications for the design of studies to examine subgroup differences. For example, examining differences across ethnic groups at the individual level also requires consideration of ethnic composition at the school level. The extent to which intervention effects differ for African American, Latino, and White students may vary depending on which group is in the majority at participating schools (Hanish and Guerra 2000). Context also plays a subtle, but important role in how subgroups are constructed. For example, forming groups considered high or low on aggression is often based on cutoffs defined by the distribution of scores within a participating sample rather than in absolute terms (Huesmann et al. 1996). This means that whether an individual is classified into a “high aggressive” group will depend not only on their level of aggression, but also on the overall level of aggression within their group.

Statistical Analysis

The examination of subgroup differences in intervention effects also presents challenges for the statistical methods used by researchers. Many of these issues are addressed elsewhere in this special issue. As with research design, studies in this area need to address the same basic issues as outcome studies. These include the selection of appropriate statistical models that take into account features of the research design such as the clustering of individuals within schools or schools within communities (MacKinnon and Lockwood 2003). Others are more specific to analyses of

subgroup effects. For example, several studies were found that examined subgroup effects based on whether significant main effects were found in each subgroup. Finding that an effect significantly differs from zero in one group and not in another does not establish that the two effects differ from each other. This is further complicated by differences in the power to detect differences across subgroups that differ in their sample sizes. We also found studies that determined that a specific factor moderated intervention effects, but did not report effect size estimates for subgroups or establish whether any subgroup effects were significantly different from zero. A significant interaction between pretest aggression and intervention conditions should be followed by estimating effect sizes across different levels of pretest aggression to determine both the direction and magnitude of effects.

Small sample sizes that result from dividing a sample into subgroups can compromise the statistical power of tests of effects conducted within subgroups. Evidence that subgroup effect sizes differ could also be obtained by calculating confidence intervals for effect sizes, which would facilitate comparisons across subgroups and studies of different sizes. Indeed, understanding subgroup differences in prevention effects would be greatly enhanced if inclusion of confidence intervals for effect sizes were common practice. At present, however, confidence intervals for effect sizes are seldom reported and the noncentral probability distributions and iterative methods required for their calculation may be unfamiliar to many researchers (see Steiger 2004; Thompson 2002).

As was previously noted, many studies of subgroup differences, particularly across demographic variables, are exploratory and include separate analyses across multiple outcomes and waves. This results in highly inflated Type I error rates (i.e., probability of rejecting the null hypothesis of no subgroup effects when no such effects actually exist). Few of the studies we reviewed acknowledged this problem. A notable exception is Aber et al. (2003), whose interpretation of findings and conclusions explicitly took into account the number of significant effects relative to the number of comparisons they conducted. The published literature may represent a biased tip of the iceberg on this problem because studies that find significant intervention effects are more likely to be published than those that do not. This “file drawer” problem (Rosenthal 1979) can lead to a biased pattern of findings within the literature. This is not to negate the potential of exploratory studies to inform prevention practices. The presence of subgroup differences, even when exploratory, may provide useful information for improving the effects of an intervention. It does suggest the need to be more circumspect in their interpretation. In contrast, other studies have attempted to show consistency of intervention effects across subgroups. These researchers

are essentially attempting to affirm the null hypothesis of no differences. This raises concerns regarding Type II error (i.e., not rejecting the null hypothesis when true subgroup differences exist). Our previous recommendation regarding reporting of effect sizes and confidence intervals for effect sizes, coupled with a venue for reporting studies whose effects did not reach statistical significance would help address this challenge, as would care taken to conduct studies with sufficient power to detect subgroup effects.

A fourth issue is the frequent practice of conducting separate analyses of potential moderating variables. This approach does not take into account the likelihood that moderators will be correlated with each other, producing confounded results. This is a problem in individual studies and in meta-analyses, where moderators may be correlated with study selection criteria (Lipsey 2003). For example, a number of studies examined gender differences and influences of pretest aggression as separate moderators of intervention effects. Higher levels of physical aggression are typically found for males across most measures of aggression and age groups. Several studies have also found that gender moderates intervention effects. It would therefore be appropriate to include not only a gender main effect, but also the Gender \times Treatment interaction terms in any test of Pretest Aggression \times Treatment interactions to determine if the moderating effects of pretest aggression are simply an artifact of gender. A good example of such an approach is provided by the MVPP (2008, 2009) study, which controlled for gender as a moderator within the context of an examination of Risk \times Condition interactions. This makes it possible to conclude that the moderating effects of level of risk are not simply an artifact of gender differences.

A final issue is the relative absence of studies that test for the presence of subgroups whose defining characteristics are unknown. If such subgroups are present in the data, it is likely that assumptions underlying multivariate analysis, such as the assumption of linearity of regression and the assumption of homoscedasticity, have been violated. Two studies in our review made use of such methods. The Segawa et al. (2005) study used growth mixture modeling to detect different substance use trajectories. The Aber et al. (1998) study used cluster analysis to detect subgroups defined by patterns of participation in the intervention. We also found a study of a dating violence intervention (Jaycox et al. 2006) that used latent class analysis to resolve discrepant information about dating status, thus defining eligibility for analysis.

Methods that have become widely available to researchers in the past decade, such as growth mixture modeling and latent class analysis, along with improvements to older methods, such as model-based clustering (Fraley and Raftery 1998) make detecting the presence of sub-

populations much more straightforward than was previously the case. Models with differing numbers and configurations of subgroups can be compared with the Bayesian information criterion (BIC; Kass and Raftery 1995) and other tests (e.g., Lo et al. 2001). More relevant to the issue of subgroups in tests of preventive interventions are methods such as the complier average causal effect (CACE; Stuart et al. 2008; Yau and Little 2001) that make it possible to estimate which control group members would have participated in the intervention had they been given the opportunity. We hope that, with the emergence of such methods, examining data for the presence of previously unidentified subgroups will become regular practice in the analysis of prevention trials.

Summary and Conclusions

The literature on youth violence prevention provides support for the notion that intervention effects may differ across subgroups, particularly for universal interventions that focus on a broad population. A variety of factors within the individual and within the broader social environment may impact intervention effects. The importance of these effects is underscored by examples from the literature indicating that interventions may not only produce stronger effects for some individuals, but may sometimes produce adverse effects for others (e.g., MACS 2002; MVPP 2008, 2009; Stoolmiller et al. 2000). Although the scope of this brief review was limited to prevention efforts aimed at reducing youth violence, it is likely that similar effects may be found for prevention efforts directed at other disorders, particularly those that attempt to produce behavioral change.

A major focus of this article was on the use of the social-ecological model as a framework for examining subgroup differences. This model differentiates factors likely to influence intervention effects that operate at the individual level versus contextual factors within an individual's social environment. This model represents a useful heuristic for organizing these factors, but it also has important implications for how factors operating at different levels might be investigated and addressed. Research designs for evaluating individual-level factors require the assessment of potential characteristics of individuals and their environment that can facilitate or impede the action of preventive interventions. Beyond the individual level, the studies reviewed suggested two types of mechanisms of moderation. One mechanism involved the barriers or facilitating factors for implementation. For example, interpersonal relationships among the school staff may impact delivery of an intervention and the organization of the school may moderate longer-term dissemination of an intervention. Other contextual factors

may moderate intervention effects because they affect uptake of intervention content. At the school level, there was evidence for moderation through an element of the organizational culture of the classroom or school, namely norms. The relative absence of studies testing moderation systematically in different communities limits the extent to which we can speculate on underlying processes, but it is possible that community disadvantage and its attendant stress on individuals makes implementation of intervention content more difficult than would be the case in less disadvantaged communities.

The studies reviewed thus far are sufficient to encourage further research aimed at identifying moderators of preventive interventions at multiple levels of analysis, along with the processes underlying such moderation. Taken together, these findings raise the possibility of improving the effectiveness of universal interventions through clarifying the role that pre-existing social setting norms play in their effects. As with the analyses of individual risk moderation presented earlier, there is the suggestion in these analyses of potential negative effects as well as positive effects for subgroups of schools that differ in their pre-existing levels of risk. If universal interventions may have negative effects in schools with strong pre-existing norms against aggression, might it not make sense to measure existing norms and use the information gained to decide whether or not to employ a universal intervention in a particular school?

References

- Aber, J. L., Jones, S. M., Brown, J. L., Chaundry, N., & Samples, F. (1998). Resolving conflict creatively: Evaluating the developmental effects of a school-based violence prevention program in neighborhood and classroom context. *Development and Psychopathology*, *10*, 187–214.
- Aber, J. L., Brown, J. L., & Jones, S. M. (2003). Developmental trajectories toward violence in middle childhood: Course, demographic differences, and response to school-based intervention. *Developmental Psychology*, *3*, 324–348.
- Asch, S. E. (1955). Opinions and social pressure. *Scientific American*, *193*, 31–35.
- Bronfenbrenner, U. (1979). Contexts of child rearing: Problems and prospects. *American Psychologist*, *34*, 844–850.
- Coie, J. D., Watt, N. F., West, S. G., Hawkins, J. D., Asamow, J. R., Markman, H. J., et al. (1993). The science of prevention: A conceptual framework and some directions for a national research program. *American Psychologist*, *48*, 1013–1022.
- Conduct Problems Prevention Research Group. (1999a). Initial impact of the Fast Track prevention trial for conduct problems: I. The high-risk sample. *Journal of Consulting and Clinical Psychology*, *67*, 631–647.
- Conduct Problems Prevention Research Group. (1999b). Initial impact of the Fast Track prevention trial for conduct problems II: Classroom effects. *Journal of Consulting and Clinical Psychology*, *67*, 648–657.
- Conduct Problems Prevention Research Group. (2010). The effects of a multiyear universal social—emotional learning program: The

- role of student and school characteristics. *Journal of Consulting and Clinical Psychology*, 78, 156–168.
- Dahlberg, L. L., & Krug, E. G. (2002). Violence—a global public health problem. In E. Krug, L. L. Dahlberg, J. A. Mercy, A. B. Zwi, & R. Lozano (Eds.), *World report on violence and health* (pp. 1–56). Geneva, Switzerland: World Health Organization.
- Davis, C. H., MacKinnon, D. P., Schultz, A., & Sandler, I. (2003). Cumulative risk and population attributable fraction in prevention. *Journal of Clinical Child and Adolescent Psychology*, 32, 228–235.
- Dishion, T. J., Patterson, G. R., & Griesler, P. C. (1994). Peer adaptation in the development of antisocial behavior: A confluence model. In L. R. Huesmann (Ed.), *Aggressive behavior: Current perspectives* (pp. 61–95). New York: Plenum Press.
- Dishion, T. J., McCord, J., & Poulin, F. (1999). When interventions harm: Peer groups and problem behavior. *American Psychologist*, 54, 755–764.
- Dishion, T. J., Kavanagh, K., Schneiger, A., Nelson, S., & Kaufman, N. K. (2002). Preventing early adolescent substance use: A family-centered strategy for the public middle school. *Prevention Science*, 3, 191–201.
- Dodge, K. A., Lansford, J. E., & Dishion, T. J. (2006). The problem of deviant peer influences in intervention programs. In K. A. Dodge, T. J. Dishion, & J. E. Lansford (Eds.), *Deviant peer influences in programs for youth: Problems and solutions* (pp. 3–13). New York: Guilford.
- Fagan, J., & Wilkinson, D. L. (1998). Social contexts and functions of adolescent violence. In D. S. Elliott, B. A. Hamburg, & K. R. Williams (Eds.), *Violence in American schools: A new perspective* (pp. 55–93). New York: Cambridge University Press.
- Farrell, A. D., & Meyer, A. L. (1997). The effectiveness of a school-based curriculum for reducing violence among urban sixth-grade students. *American Journal of Public Health*, 87, 979–984.
- Farrell, A. D., & Vulin-Reynolds, M. (2007). Violent behavior and the science of prevention. In D. J. Flannery, A. T. Vazsonyi, & I. D. Waldman (Eds.), *The Cambridge handbook of violent behavior and aggression* (pp. 767–786). New York: Cambridge University Press.
- Farrell, A. D., Meyer, A. L., Kung, E. M., & Sullivan, T. N. (2001a). Development and evaluation of school-based violence prevention programs. *Journal of Clinical Child Psychology*, 30, 207–220.
- Farrell, A. D., Meyer, A. L., & White, K. S. (2001b). Evaluation of Responding in Peaceful and Positive Ways (RIPP): A school-based prevention program for reducing violence among urban adolescents. *Journal of Clinical Child Psychology*, 30, 451–463.
- Farrell, A. D., Erwin, E. H., Bettencourt, A., Mays, S., Vulin-Reynolds, M., Sullivan, T. N., et al. (2008). Individual factors influencing effective nonviolent behavior and fighting in peer situations: A qualitative study with urban African American adolescents. *Journal of Clinical Child and Adolescent Psychology*, 37, 397–411.
- Farrell, A. D., Mays, S., Bettencourt, A., Erwin, E. H., Vulin-Reynolds, M., & Allison, K. W. (2010). Environmental influences on fighting versus nonviolent behavior in peer situations: A qualitative study with urban African American adolescents. *American Journal of Community Psychology*, 46, 19–35.
- Farrell, A. D., Henry, D., Mays, S., & Schoeny, M. (2011). Parents as moderators of the impact of peers and school environment. *Child Development*, 82, 146–161.
- Flannery, D. J., Vazsonyi, A. T., & Waldman, I. D. (Eds.). (2007). *The Cambridge handbook of violent behavior and aggression*. New York: Cambridge University Press.
- Floy, B. R., Biglan, A., Boruch, R. F., Castro, F. G., Gottfredson, D., Kellam, S., et al. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science*, 6, 151–175.
- Foshee, V. A., Bauman, K. E., Ennett, S. T., Suchindran, C., Benefield, T., & Linder, G. F. (2005). Assessing the effects of the dating violence prevention program "Safe Dates" using random coefficient regression modeling. *Prevention Science*, 6, 245–258.
- Fraley, C., & Raferty, A. E. (1998). How many clusters? Which clustering method? Answers via model-based cluster analysis. *The Computer Journal*, 41, 478–488.
- Fraser, M. W., Galinsky, M. J., Smokowski, P. R., Day, S. H., Terzian, M. A., Rose, R. A., et al. (2005). Social information-processing skills training to promote social competence and prevent aggressive behavior in the third grade. *Journal of Consulting and Clinical Psychology*, 73, 1045–1055.
- Frey, K. S., Hirschstein, M. K., & Guzzo, B. A. (2000). Second Step: Preventing aggression by promoting social competence. *Journal of Emotional and Behavioral Disorders*, 8, 102–112.
- Gottfredson, D. C., Gottfredson, G. D., & Hybl, L. G. (1993). Managing adolescent behavior: A multiyear, multischool study. *American Educational Research Journal*, 30, 179–215.
- Grabosky, P. N. (1996). Unintended consequences of crime prevention. In R. Homel (Ed.), *Crime prevention studies, vol 5: The politics and practice of situational crime prevention* (pp. 25–57). Monsey, NY: Criminal Justice Press.
- Gregory, A., Henry, D., Schoeny, M. & The Metropolitan Area Child Study Research Group. (2007). School climate and implementation of a preventive intervention. *American Journal of Community Psychology*, 40, 250–260.
- Griffin, J. P., Jr., Chen, D., Eubanks, A., Brantley, K. M., & Willis, L. A. (2007). A randomized violence prevention trial with comparison: Responses by gender. *Journal of School Violence*, 6, 65–81.
- Hanish, L. D., & Guerra, N. G. (2000). The roles of ethnicity and school context in predicting children's victimization by peers. *American Journal of Community Psychology. Special Issue: Minority Issues in Prevention*, 28, 201–223.
- Hawkins, J. D., Herrenkohl, T., Farrington, D. P., Brewer, D., Catalano, R. F., Harachi, T. W., et al. (1998). A review of predictors of youth violence. In R. Loeber & D. P. Farrington (Eds.), *Serious & violent juvenile offenders: Risk factors and successful interventions* (pp. 106–146). Thousand Oaks, CA: Sage Publications.
- Henry, D., & Chan, W. Y. (2010). Cross-sectional and longitudinal effects of sixth grade setting-level norms for nonviolent problem solving on aggression and associated attitudes. *Journal of Community Psychology*, 38, 1007–1022.
- Henry, D. B., & Schoeny, M. E. (2007, June). *Moderation of universal intervention effects by norms and ethnicity (S-11372: "Violence and poverty: Testing theoretical models for African American youth")*. Paper presented at the Biennial Meeting of the Society for Community Research and Action, Pasadena, CA.
- Henry, D. B., Farrell, A. D., Schoeny, M. E., Tolan, P. H., & Dymnicki, A. B. (2010). *Influence of school-level variables on aggression and associated attitudes of middle school students*. Manuscript submitted for publication.
- Huesmann, L. R., Maxwell, C., Eron, L., Dahlberg, L. L., Guerra, N. G., Tolan, P. H., et al. (1996). Evaluating a cognitive/ecological program for the prevention of aggression in urban children. *Preventive Medicine*, 12, 120–128.
- Jackson, J. (1966). A conceptual and measurement model for norms and roles. *Pacific Sociological Review*, 9, 35–47.
- Jaycox, L. H., McCaffrey, D., Eiseman, B., Aronoff, J., Shelley, G. A., & Collins, R. L. (2006). Impact of a school-based dating violence prevention program among Latino teens: Randomized controlled effectiveness trial. *Journal of Adolescent Health*, 39, 694–704.
- Jo, B. (2002). Estimation of intervention effects with noncompliance: Alternative model specifications. *Journal of Educational and Behavioral Statistics*, 27, 385–409.
- Kam, C., Greenberg, M. T., & Walls, C. T. (2003). Examining the role of implementation quality in school-based prevention using the PATHS curriculum. *Prevention Science*, 4, 55–63.

- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*, 773–795.
- Kellam, S. G., Ling, X., Merisca, R., Brown, C. H., & Ialongo, N. (1998). The effect of the level of aggression in the first grade classroom on the course and malleability of aggressive behavior into middle school. *Development and Psychopathology*, *10*, 165–185.
- Kiesler, D. J. (1966). Some myths of psychotherapy research and the search for a paradigm. *Psychological Bulletin*, *65*, 110–136.
- Leadbeater, B., Hoglund, W., & Woods, T. (2003). Changing contents? The effects of a primary prevention program on classroom levels of peer relational and physical victimization. *Journal of Community Psychology*, *31*, 397–418.
- Lipsey, M. W. (2003). Those confounded moderators in meta-analysis: Good, bad, and ugly. *Annals of the American Academy of Political and Social Science*, *587*, 69–81.
- Lipsey, M. W., & Derzon, J. H. (1998). Predictors of violent or serious delinquency in adolescence and early adulthood: A synthesis of longitudinal research. In R. Loeber & D. P. Farrington (Eds.), *Serious & violent juvenile offenders: Risk factors and successful interventions* (pp. 86–105). Thousand Oaks, CA: Sage Publications, Inc.
- Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, *88*, 767–778.
- MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. New York: Taylor & Francis Group/Lawrence Erlbaum Associates.
- MacKinnon, D. P., & Lockwood, C. M. (2003). Advances in statistical methods for substance abuse prevention research. *Prevention Science*, *4*, 155–171.
- Metropolitan Area Child Study Research Group (MACS). (2002). A cognitive-ecological approach to preventing aggression in urban settings: Initial outcomes for high risk children. *Journal of Consulting and Clinical Psychology*, *70*, 179–194.
- Meyer, A. L., Farrell, A. D., Northup, W., Kung, E., & Plybon, L. (2000a). *Promoting nonviolence in early adolescence: Responding in peaceful and positive ways*. New York: Kluwer.
- Meyer, A., Nicholson, R., Danish, S., Fries, E., & Polk, V. (2000b). A model to measure program integrity of peer-led health promotion programs in rural middle schools: Assessing the implementation of the sixth grade goals for health program. *Journal of Educational and Psychological Consultation*, *11*, 223–252.
- Multisite Violence Prevention Project. (2008). Impact of a universal school-based violence prevention program on social-cognitive outcomes. *Prevention Science*, *9*, 231–244.
- Multisite Violence Prevention Project. (2009). The ecological effects of universal and selective violence prevention programs for middle school students: A randomized trial. *Journal of Consulting and Clinical Psychology*, *77*, 526–542.
- Paul, G. L. (1967). Strategy of outcome research in psychotherapy. *Journal of Consulting Psychology*, *31*, 109–118.
- Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2004). *Classroom Assessment Scoring System [CLASS]*. Unpublished measure. University of Virginia.
- Raudenbush, S. W., & Sampson, R. J. (1999). Econometrics: Toward a science of assessing ecological settings, with application to the systematic social observation of neighborhoods. *Sociological Methodology*, *29*, 1–41.
- Reid, J. B., Eddy, J. M., Fetrow, R. A., & Stoolmiller, M. (1999). Description and immediate impacts of a preventive intervention for conduct problems. *American Journal of Community Psychology*, *27*, 483–517.
- Rosenthal, R. (1979). The 'file drawer problem' and tolerance for null results. *Psychological Bulletin*, *86*, 638–641.
- Rousseau, D. M. (1985). Issues of level in organizational research: Multi-level and cross-level perspectives. *Research in Organizational Behavior*, *7*, 1–37.
- Segawa, E., Ngwe, J. E., Li, Y., & Flay, B. R. (2005). Evaluation of the effects of the Aban Aya Youth Project in reducing violence among African American adolescent males using latent class growth mixture modeling techniques. *Evaluation Review*, *29*, 128–148.
- Shinn, M., & Rapkin, B. D. (2000). Cross-level research without cross-ups in community psychology. In J. Rappaport & E. Seidman (Eds.), *Handbook of community psychology* (pp. 669–695). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Smith, B., & Sechrest, L. (1991). Treatment of aptitude \times treatment interactions. *Journal of Consulting and Clinical Psychology*, *59*, 233–244.
- Smokowski, P. R., Fraser, M. W., Day, S. H., Galinsky, M. J., & Bacallao, M. L. (2004). School-based skills training to prevent aggressive behavior and peer rejection in childhood: Evaluating the making choices program. *Journal of Primary Prevention*, *25*, 233–251.
- Spoth, R., Redmond, C., Shin, C., Lepper, H., Haggerty, K., & Wall, M. (1998). Risk moderation of parent and child outcomes in a preventive intervention: A test and replication. *American Journal of Orthopsychiatry*, *68*, 565–579.
- Steiger, J. H. (2004). Beyond the F Test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychological Methods*, *9*, 164–182.
- Stoolmiller, M., Eddy, J. M., & Reid, J. B. (2000). Detecting and describing preventive intervention effects in a universal school-based randomized trial targeting delinquent and violent behavior. *Journal of Consulting and Clinical Psychology*, *68*, 296–306.
- Stuart, E. A., Perry, D. F., Le, H.-N., & Ialongo, N. S. (2008). Estimating intervention effects of prevention programs: Accounting for noncompliance. *Prevention Science*, *9*, 288–298.
- Thompson, B. (2002). What future quantitative social science research could look like: Confidence intervals for effect sizes. *Educational Researcher*, *31*, 24–31.
- Tolan, P. H., & Brown, C. H. (1998). Evaluation research on violence interventions: Issues and strategies for design. In P. K. Trickett & C. J. Schellenbach (Eds.), *Violence against children in the family and the community* (pp. 439–464). Washington, DC: American Psychological Association.
- Tolan, P. H., Gorman-Smith, D., & Henry, D. (2004). Supporting families in a high-risk setting: proximal effects of the SAFE-Children preventive intervention. *Journal of Consulting and Clinical Psychology*, *72*, 855–869.
- Tseng, V., & Seidman, E. (2007). A systems framework for understanding social settings. *American Journal of Community Psychology*. Special Issue: *Systems Change*, *39*, 217–228.
- US Department of Health and Human Services (USDHHS) Public Health Service (PHS); Office of the Surgeon General. (2001). *Youth violence: A report of the Surgeon General*. Rockville MD: US Department of Health and Human Services.
- Wilson, S. J., Lipsey, M. W., & Derzon, J. H. (2003). The effects of school-based intervention programs on aggressive behavior: A meta-analysis. *Journal of Consulting and Clinical Psychology*, *71*, 136–149.
- Wright, J. C., & Zimmerman, M. A. (2006). Culturally sensitive interventions to prevent youth violence. In N. G. Guerra & E. P. Smith (Eds.), *Preventing youth violence in a multicultural society* (pp. 221–269). Washington, DC: American Psychological Association.
- Yale, M. E., Scott, K. G., Gross, M., & Gonzalez, A. (2003). Using developmental epidemiology to choose the target population for an intervention program in a high-risk neighborhood. *Journal of Clinical Child and Adolescent Psychology*, *32*, 236–242.
- Yau, L. H. Y., & Little, R. J. (2001). Inference for the complier-average causal effect from longitudinal data subject to noncompliance and missing data, with application to a job training assessment for the unemployed. *Journal of the American Statistical Association*, *96*, 1232–1244.