Meta-analysis and systematic review of the effectiveness of school-based programs to reduce multiple violent and antisocial behavioral outcomes.

Chapter · January 2012

DOI: 10.13140/2.1.2616.6402

2 authors:

Aaron Alford
National Network of Public Health Institutes
42 PUBLICATIONS **199** CITATIONS

SEE PROFILE

James H Derzon
RTI International
55 PUBLICATIONS **2,288** CITATIONS

SEE PROFILE

44

Meta-Analysis and Systematic Review of the Effectiveness of School-Based Programs to Reduce Multiple Violent and Antisocial Behavioral Outcomes

Aaron A. Alford and James Derzon

THE BATTELLE CENTERS FOR PUBLIC HEALTH RESEARCH AND EVALUATION, ARLINGTON, VA

Abstract

Early models of violence prevention targeted a single outcome by intervening in a single suspected causal pathway. This limited approach has increasingly fallen out of favor, as evidence has accumulated that violent and antisocial outcomes are driven by a highly inter-correlated cloud of potential risk factors. Policy makers and consumers are often interested in whether or not a program actually prevents or reduces the problem behaviors that it is designed to address. The current study combined the methods of a traditional meta-analysis and traditional systematic review to examine the evidence of effectiveness of school-based programs in simultaneously reducing both violent and antisocial behavioral outcomes. Overall, none of the programs report evidence of being effective for all outcomes, and only one successfully impacted more than one distal outcome. When considering all forms of evidence, no program showed uniformly positive evidence across all outcomes and domains considered. Implications for science and practice are discussed.

Increasingly, prevention science is moving towards integrated models of prevention. Early models of violence prevention targeted a single outcome by intervening in a single suspected causal pathway (Domitrovich, Bradshaw, Greenberg, Embry, & Ialongo, 2009). This limited approach has increasingly fallen out of favor, as evidence has accumulated that violent and antisocial outcomes are driven by a highly inter-correlated risk factors (Wei, Loeber, & White, 2004). Prevailing evidence suggests that there is a complex interplay of predisposition, training and skills, and environmental stimuli that predispose youth to acting or responding inappropriately (Silberg, Rutter, D'Onofrio, & Eaves, 2003). Theoretically, each individual receiving an intervention has a unique cluster, or at least uniquely weighted cluster, of risk factors. That is, within individual,

each risk factor likely has a unique duration, intensity and interaction with individual lability. By targeting a cluster of risk factors or common factors that lead to multiple related outcomes, it may be possible to reduce the net burden of risk for each of those outcomes. Most school-based programs target clusters of risk factors simultaneously with the intent of reducing all the distal outcomes related to those risk factors (Hansen, Dusenbury, Bishop, & Derzon, 2007).

What Counts as Evidence of Effectiveness?

When discussing outcomes, it is essential to maintain conceptual clarity. The reigning theoretical framework, linking risk and protective factors with distal outcomes, suggests that effectively targeting proximal outcomes (risk and protective factors) should lead to changes in common problem behaviors (Hawkins, Catalano, & Miller, 1992; Jessor & Jessor, 1977). When documenting program effectiveness, it is tempting to claim that there is evidence of effectiveness for multiple outcomes based on positive changes in the risk and protective factors known to contribute to the distal behavioral outcome the program seeks to prevent. Many programs implicitly or explicitly target multiple proximal outcomes such as attitudes, opinions, knowledge and behaviors that are believed to mediate distal outcomes such as alcohol use, tobacco use, drug use, and aggression. However, change in proximal outcomes at the aggregate level shows limited correspondence with change in the distal outcome of interest (Alford & Derzon, 2011; Najaka, 2000; Najaka, Gottfredson, & Wilson, 2001). Support for the claim that trial-wide change in proximal risk and protective factors translates reliably into change in distal behavioral outcomes remains elusive, even if support for these relations is well documented at the individual level (Derzon, 2007; Lipsey & Derzon, 1998). Thus, while it is clear that proximal outcomes (i.e., risk and protective factors) mediate distal outcomes at the individual level, it is less clear that group-wide evidence of change in proximal outcomes reliably translates into change in distal outcomes.

For the purposes of exploration, development and understanding, reliance on proximal outcomes is more than warranted (MacKinnon, Fairchild, & Fritz, 2007; Pearl, 2000). Knowing why a program works, or does not work, can lead to a finer understanding of prevention programs as well as the mechanisms of violent and aggressive behavior. Recent work in prevention science has also indicated that a basic understanding of the mechanisms of a program is central to effective content delivery (e.g., Derzon, Springer, Sale, & Brounstein, 2005).

Despite the interest and theoretical importance of linking risk and protective factors with outcomes, policy makers and consumers are often interested in whether or not a program actually prevents or lessens the problem behaviors that the program is designed to reduce. Program adopters require an understandable and common metric with which they can compare the many prevention programs publically available. In the United States, the focus has primarily been distal behavioral outcomes (McBride, 2003). For these reasons, we focus this investigation solely on the effectiveness of programs that measure the effect on distal behavioral outcomes representative of violent and other antisocial behaviors.

Meta-Analysis and Systematic Reviews

Meta-analyses of school-based programs tend to either aggregate study findings to produce a summary estimate of intervention impact (e.g., Derzon, 2006; Wilson, Gottfredson, & Najaka, 2001; Wilson, Lipsey, & Derzon, 2003) or to report summary results at the level of strategy or mechanism of change for one or more discrete outcomes (see, e.g., Ennett et al., 2003; Hansen et al., 2009). Inherent in each of these approaches is the pooling of evidence across programs and studies to make evidence-based claims of the average effectiveness of the interventions that contribute data to each outcome. In meta-analysis, the observation that different programs and

different studies may contribute evidence to different outcomes is typically not explicitly examined. Meta-analysis was developed to account for the instability of evidence inherent in much primary research in the social sciences. Although pooling evidence increases power to detect effects and estimate homogeneity, pooling evidence from different studies obscures whether the same implementation is effective for multiple outcomes or if different trials contribute to different results. Thus, while meta-analysis can tell us if programs are, on average, effective across a range of outcomes, it does not typically provide evidence that the same program is effective across multiple outcomes.

While meta-analysis provides a unique set of methodologies for combining the information from many studies or evaluations to arrive at a single estimate of the magnitude and direction of effectiveness, it relies on unconditioned estimates (main effects) of an intervention's effectiveness. As it is a quantitative method, it can also provide a test of significance for an aggregate estimate, and can be used to test if the included estimates vary more than would be expected by sampling error. The estimates of effectiveness from individual trials are combined using standardized estimates of impact (effect sizes). However, the estimates of effect are derived from original research reports, and are dependent on the reporting of the original authors.

A meta-analyst must rely on the original author to provide clear reporting and is limited by the available information and the analyses reported by the original author. Transformations are available to convert many common statistics into standardized effect sizes, but many results from newer, more sophisticated methods, such as growth mixture models, cannot be transformed. Because of this limitation, meta-analysis is largely restricted to combining evidence in the form of main effects. A main effect is the effect of the intervention on the outcome of interest, regardless of potential covariates. Controlling for a covariate removes the covariates' variance from the main effect (Keef & Roberts, 2004), but the resulting finding is no longer representative of the main effect, and may not be comparable to effect sizes from other studies. Therefore, including covariate-adjusted results is not typical in meta-analysis (Morris & DeShon, 2002; Peterson & Brown, 2005).

As a result, studies that use these methods are lost to a meta-analysis under most circumstances (Morris & DeShon, 2002). When aggregating at the program level, this creates the potential for bias. Most programs are evaluated by a single researcher or a limited number of researchers. Researchers tend to favor a particular analysis frame or may conduct sophisticated analyses to answer specific research questions. If these analyses produce findings that cannot be transformed into a common metric, results from these sophisticated studies are systematically lost to meta-analysis.

A systematic review does not have the same limitations. Systematic reviews can capture results from all forms of statistical tests, as they do not rely on a standardized quantitative estimate. That same feature is also the primary limitation when used for decision making purposes. Systematic reviews do not provide a strong option for combining statistical tests when combining estimates of effectiveness across multiple tests or studies. Traditionally, the option for summarizing evidence in this framework has been vote counting, a procedure that relies on counting the number of statistically significant findings across studies, but which has been found lacking (Borenstein, Hedges, Higgins, & Rothstein, 2009; Cooper & Hedges, 1994).

When used alone, each method has specific weaknesses that are difficult to address without the addition of the second method. Meta-analysis allows for the estimation of a single quantitative estimate of effectiveness for each outcome, but is limited to only a subset of the evidence available to estimate effectiveness. Systematic reviews are more inclusive, but cannot provide a common quantitative estimate of program impact.

To determine the effectiveness of school-based programs on multiple outcomes, the present study uses a novel combination of traditional meta-analysis and traditional systematic review

to maximize the information available to estimate if programs are effective across multiple outcomes.

Methods

The current study combined the methods of a traditional meta-analysis and traditional systematic review to examine the evidence of effectiveness of school-based programs in reducing both violent and antisocial behavioral outcomes.

Program Eligibility

The present study breaks out evidence from a larger study documenting the effectiveness of school-based programs in reducing substance use and antisocial behavior outcomes (Alford, Derzon, Hagan, & Crosse et al., 2011) which sought to identify all programs that are available to U.S. schools either through open source documents or through the marketplace and have been identified or marketed as a promising or effective program. From a practical perspective, we assessed whether publically available programs that schools could identify from lists of effective programs were effective across a range of problem behaviors of interest to the U.S. Department of Education. Thus, we used a hybrid two-stage search technique to identify programs and evaluations of those programs. First, 491 programs were identified through an examination of a cumulative listing of 12 existing lists of effective programs intended to reduce problem behavior (Mihalic, 2008, additional information is available from the first author [A. A.]), by contacting developers, and through marketing materials. In many cases, the developers have listed evaluation reports in their marketing materials. Second, over 6,000 reports, documents, and articles for these 491 programs were identified by searching each of the program names in the following sources: Applied Social Sciences Index and Abstracts (ASSIA), Dissertation Abstracts, ERIC, GoogleScholar, MEDLINE, National Criminal Justice Reference Service (NCJRS), National Technical Information Service (NTIS), Periodical Abstracts (PerAbs), PsycINFO, Social Science Abstracts (SocialSciAbs), Sociological Abstracts, What Works Clearinghouse (WWC), Wilson Select Plus, and WorldCat. When an abstract appeared to contain relevant evaluation information, the corresponding document was retrieved. Through these procedures over 6,000 reports were identified and retrieved. References cited in the retrieved reports were scanned for additional evaluation documents.

Each retrieved report was subjected to an eligibility screen by two reviewers to ascertain whether or not the evaluation met basic inclusion and exclusion criteria. Differences between reviewers were reconciled by discussion. The following inclusion criteria were used during this pre-screening process:

1. The report included findings from a program evaluation.
2. The program evaluation included measurements of a behavioral outcome (ATOD or violence outcome).
3. The program evaluated was school-based or included school staff.
4. Studies were also retained if they met modest methodological quality standards.
    a. The evaluation used an experimental or quasi-experimental study design, including using a non-treatment or standard treatment comparison group.
    b. Differential attrition was no greater than 20% difference between the treatment and control groups.
    c. The planned intervention was similar to the model program.

d. There were no historical effects uniquely impacting either the treatment or comparison groups.

Reports that passed the above criteria were then screened by coders to confirm the program tested and identify the study sample involved. Groups of reports that presented findings for the same study sample were clustered under a common study identifier. Reports that provided estimates from more than one independent trial were assigned individual identifiers for each independent trial. Once all evidence was sorted into their respective trials and categorized by program, two additional eligibility criteria were applied.

5. The program was evaluated using two or more independent samples (this could include two independent research trials or a single trial that included multiple sites).
6. The program evaluated was available in a manualized form during 2009.

For the current study, we differentiate between three levels of evidence: program-level, report-level, and study-level evidence. A report is any document that details the findings from a program evaluation. Study-level evidence is all of the reports that contain evaluations of a program applied to an independent and unique sample. Program-level evidence is all of the evidence from independent evaluation studies of a given program.

These distinctions are important for several reasons. Report-level evidence is potentially misleading due to the tendency of researchers to publish multiple documents using evidence from the same study sample. At the report-level, estimates of the true density of evidence are systematically inflated. Using report-level evidence can also bias the results of meta-analyses by inflating the weights of programs with a higher ratio of reports to independent studies. These biases can be avoided by clustering data according to the study sample they were derived from, rather than relying on report-level data. To examine the evidence of effectiveness across several outcomes, therefore, data were clustered first by independent study sample and then by program.

Selecting Outcome Estimates

Similar to researchers reporting evidence from a single study sample in multiple reports, the effectiveness of the intervention using multiple measures representing the same outcome construct is often reported. In these cases, we retained only one measure per study. The retained measure was chosen based on representativeness and sensitivity to change. For this study, representativeness was defined as the capacity to capture the full range of outcomes in each class of outcomes. For example, a measure capturing past month use of alcohol is more representative of alcohol use than a measure capturing past month use of beer. Sensitivity was defined as the capacity to capture change without becoming unstable. As an illustration, lifetime use can only capture onset and is not sensitive to cessation, and weekly use is too sensitive to temporally local change. Past month use is both stable and sensitive to change, and was believed preferable to the lifetime or weekly use.

Many researchers likewise report program effectiveness estimates at posttest and at multiple follow-up periods. When this occurred, we selected the posttest estimate (as it is the most commonly reported estimate), or the follow-up estimate closest in time to the posttest estimate when a posttest value was not reported for the distal outcome.

The data for the systematic review and meta-analysis were combined according to the following hierarchy:

1. If a program provided one or more effect sizes for an outcome, the aggregate effect size or effect size was abstracted and the findings abstracted for the systematic review was dropped.
2. If a program only provided data to the systematic review, the results from the systematic review were carried forward and contributed to the findings table developed for this evidence summary.

Following these principles and procedures, evidence for eight categories of outcome was coded: alcohol use, tobacco use, marijuana use, substance use (all other substances), aggressive and disruptive behavior, antisocial behavior, delinquent behavior, and physical aggression. Both of the coding frameworks discussed below included all eight outcomes.

Data Coding

Coding the data was a two-step process. Once reports were clustered by sample, data from each report were abstracted into a framework capturing the assigned sample size by condition, attrition, outcome type, notes on the design of the study, the point estimate provided by the report, and the confidence intervals for the point estimate, whether or not the estimate was covariate adjusted, the covariates used in the adjustment procedure, and the presence or absence of the information necessary to create a standardized measure of effect.

Point estimates that could not be standardized were captured in the coding framework for the systematic review. The test statistic, confidence interval and direction were recorded along with interpretation of the statistic within the context of the study. If only one study contributed a test of effectiveness, the unstandardized measure was used to determine of effectiveness for the program and could be positive and significant, null, or negative and significant.

If the study reported evidence sufficient for calculating an effect size, then the effect size and information relating to the individual study from which it was derived was coded into a second framework. This framework provided the data for the meta-analysis. Main effects findings were transformed into an effect size, specifically Hedges *d*. Hedges *d* has a normal distribution centered on zero and standard deviation of one and is calculated as: Hedges $d = \Sigma_m = \frac{\overline{X}_{G1} - \overline{X}_{G2}}{s_p}$ where $\overline{X}_{G1}$ is the mean for Group 1 at posttest, $\overline{X}_{G2}$ is the mean for Group 2 at posttest and s_p is the pooled standard deviation of the two groups.

Hedges (1981) has demonstrated that standardized estimates based on small samples can be upwardly biased. The correction for this is: Hedges Unbiased Estimate $= \Sigma'_m = \left[1 - \frac{3}{4N-9}\right]\Sigma_m$ where N is the total sample size $(n_{G1} + n_{G2})$, Σ_m is the biased standardized mean difference shown in the formula above, n_{G1} is the number of subjects in Group 1, and n_{G2} is the number of subjects in Group 2.

Calculating the inverse variance is accomplished using the following equation and the terms described above: Inverse variance weight $= w_m = \frac{1}{\Sigma^2} = \frac{2n_{G1}n_{G2}(n_{G1} + n_{G2})}{2(n_{G1} + n_{G2})2 + n_{G1}n_{G2}(\Sigma'_m)^2}$

Fixed-effect modeling acknowledges the greater precision of effectiveness estimates from larger samples from the same study population (e.g., to reduce the impact of non-response bias on a study-sample's grand mean) and was used when the same study sample produced multiple estimates of a construct (i.e., the measures were determined to be equally sensitive). Because the assumption of a common treatment effect cannot be assumed across different study samples, random-effects modeling was used to combine estimates from different study samples (DerSimonian & Laird, 1986). Random effects modeling reduces the role of weights in obtaining mean estimates and assumes there are both measured and unmeasured influences affecting the finding obtained in the primary study. The program Comprehensive Meta-analysis, Version 2.0 was used to obtain all mean estimates.

Choosing the Effect Size to Represent the Program.

Within each program, if more than one study contributed a test of effectiveness for a single outcome, the following decision rules were applied:

1. Two or more positive and significant tests of effectiveness for a single outcome: The overall effect for the outcome was determined to be positive and significant.
2. Two or more negative and significant tests of effectiveness for a single outcome: The overall effect for the outcome was determined to be negative and significant.
3. Two or more null tests of effectiveness for a single outcome: The overall effect for the outcome was determined to be null.
4. Two or more estimates of effect that disagree: The overall evidence of effectiveness was determined to be contradictory.

In some cases, researchers choose to only report p-values for tests of effectiveness. In these cases, the estimate of effectiveness was not included in the present study. It is not uncommon for meta-analysts to transform reported data or make limited inferences for the sake of transformation. However, it is considered poor practice to create effect sizes from data that is largely inferred, unless there is an extenuating circumstance. Furthermore, the practice of reporting only p-values has largely fallen out of favor in most fields of research, as the practice is consider antithetical to the ethic of transparency. In keeping with these practices, we did not include estimates of effectiveness based solely on p-values.

Results

Four hundred ninety-one programs were identified during the canvass of the literature. Of these, 334 were excluded because they were designed to impact outcomes other than ATOD use, aggression, or violence. Forty-two were excluded because no implementation materials were publicly available. Of the remaining 115 programs, 24 were excluded for having only a single evaluation study, 41were excluded for not meeting methodological criteria, and another four were excluded for having evidence from fewer than two independent samples. The remaining 46 programs had an adequate research base to support an investigation of the evidence for program effectiveness for substance use or antisocial behavior outcomes; of these 24 programs included at least one violence or aggression outcome (see Figure 44.1).

The reports and studies evaluating these programs were examined for evidence of effectiveness on violence or aggression behavioral outcomes. Of the 24 programs that had an adequate research base for assessing the effectiveness of the program in reducing antisocial behavior outcomes, 13 provided an estimate of effectiveness for the meta-analysis, and 18 programs provided an estimate of effectiveness for the research synthesis. A standard meta-analytic treatment of the 18 programs providing main effect data would conclude these programs were generally effective and effective in improving three ot the four outcomes examined (see Figure 44.2). Using our current approache, two programs were found to be effective for at least one outcome by the meta-analysis, and a total of eight programs were found to have evidence of effectiveness when considering the estimates provided by both the meta-analysis (standardized estimates of effect) and the systematic review (non-standardized estimates of effect). The systematic review found that 6 programs had null, negative or mixed effects on at least one outcome. The meta-analysis found that 12 programs had null or negative effects on at least one outcome. Four programs were found to have consistently positive effects across more than one study. The bibliography of reports used to establish the findings included in this review is available from the authors.

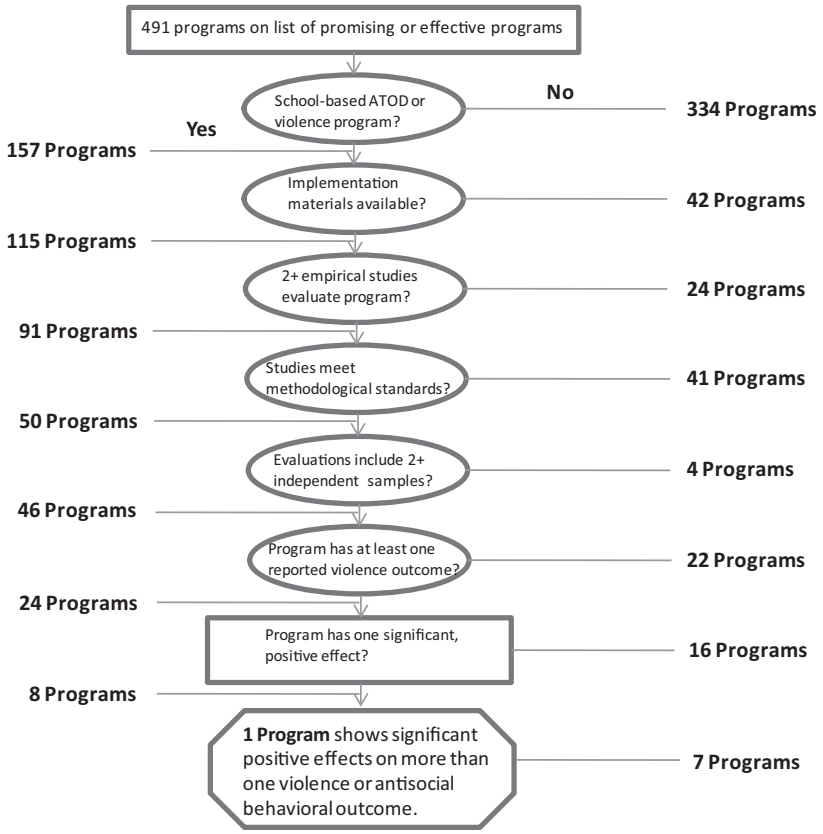

Figure 44.1 Flow of evidence through phases of the review

Table 44.1 displays all summary effectiveness estimates for each of the studies and programs reporting behavioral outcome measures. The shaded rows are the summary estimates for each program while the unshaded rows provide evidence from each independent study reporting evidence of the effectiveness of the program. Numbers indicate that a standardized (meta-analytic) estimate of the program's effect on the outcome could be calculated. Symbols are used to represent the findings coded using systematic review methods. A positive sign represents a positive

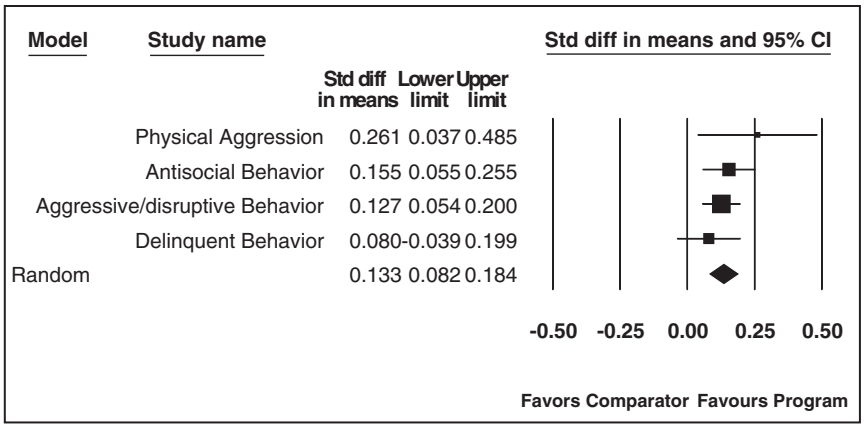

Figure 44.3 Grand means across programs by outcome

and significant finding while negative signs indicate a negative and significant finding. The "θ" symbol is used to indicate null (statistically nonsignificant) findings. At the program-level (the shaded rows) the question mark summarizes inconsistent findings across studies that could not be meta-analytically summarized. Empty cells indicate that no evidence for the outcome was reported. Bolded values indicate that the result is statistically significant.

Eight programs reported data for multiple outcomes (see Table 44.1).[1] Of these, one exhibited a positive and significant effect on more than one outcome of interest. Of the six effect sizes contributed by this study, only one was not significant and positive.

Discussion

Forty-six of the 491 programs identified using lists of effective programs reported evidence of effectiveness on the distal outcomes summarized in this investigation. Of these, 24 provided evidence of effectiveness on a violence or antisocial outcome.

Findings from the current study reveal that publically available programs can impact multiple violenct and antisocial behaviors. However, none of the programs report evidence of being effective for all outcomes, and only one successfully impacted more than one distal outcome. When considering all forms of evidence, no program showed uniformly positive evidence across all outcomes and domains considered. Unfortunately, these findings show there is limited evidence from manualized programs that have been carefully evaluated in more than one trial to support the claim that programs are effective for multiple outcomes. While we could not test whether targeting more intermediate outcomes increased the effect of the intervention on a given outcome, the evidence summarized here suggests that targeting multiple intermediate outcomes is not sufficient to reliably improve behavioral outcomes across indicators or domains. It has been posited that reducing the net burden of risk should improve youth outcomes across a spectrum of problem behaviors (e.g., Jessor & Jessor, 1977), but it seems also likely that not all outcome behaviors are equally sensitive to reductions in risk factors.

The evidence presented here is comprehensive in its coverage of school-based programs, and our inclusion of evidence from systematic reviews makes it more inclusive than traditional meta-analysis. By combining systematic review with meta-analysis, it was possible to identify evidence from a larger number of programs than would be possible using meta-analysis alone. By combining meta-analytic results with those of a systematic review, legitimate findings could be retained even when they could not be included in aggregate effect sizes.

Unfortunately, unlike the main effects summarized by the meta-analysis, there is no accepted means of weighting and combining effectiveness evidence captured by the systematic review. For this study, a program could only receive a rating of positive and significant only if all studies reporting an outcome reported a positive and significant finding for the outcome. Thus, while this approach protects against Type 1 error (rejecting the null hypothesis when it is, in fact, true), the summary results provided by the systematic review may be overly conservative as the program would receive a positive and significant rating in our systematic review framework only if all trials reported positive findings for the outcome.

Future Directions and Implications for Practice

Multi-target school-based prevention programs are an exciting but relatively unexplored development in the field of prevention science. Such programs have potential for reducing the net burden of risk within individuals, but this potential effect on multiple distal outcomes has not been well documented to date. Frequently, tests of change in risk factors have been carried out through global means-driven tests of change for individual risk factors. Trials linking changes in

Table 44.1 Estimates of effectiveness for all programs, by study and outcome

Effectiveness of Programs Across Studies									
		ATOD use outcomes				*Antisocial behavior outcomes*			
Name	*Study*	*Alcohol Use*	*Tobacco Use*	*Marijuana Use*	*Substance use*	*Aggressive/ Disruptive*	*Antisocial Behavior*	*Delinquent Behavior*	*Physical Aggression*
Program 1						0.18			
	Study 1					0.18			
	Study 2					–			
	Study 3					θ			
Program 2		0.311		0.319	0.411	0.301		0.194	–0.122
	Study 1	0.311		0.319	0.411			0.194	–0.122
	Study 2					0.55			
	Study 3					0.063			
	Study 4					θ			
Program 3					θ	θ			
	Study 1				θ	θ			
Program 4						**+**			
	Study 1					+			
	Study 2					+			
Program 5						**+**			
	Study 1					+			
	Study 2					+			
Program 6		**+**	θ	θ		θ		θ	θ
	Study 1	+	θ	θ		θ		θ	θ
Program 7					0.21	–0.044		0.193	
	Study 1				0.294	0.006			
	Study 2				0.18			0.193	
	Study 3					–0.229			
Program 8						0.158			
	Study 1					0.158			
	Study 2					θ			
	Study 3					θ			
Program 9		–0.031	**+**	**+**	–0.147			–0.112	
	Study 1	–0.031	+	+	–0.147			–0.112	
Program 10						0.109			
	Study 1					–0.049			
	Study 2					θ			
	Study 3					**0.27**			
Program 11								θ	
	Study 1							θ	
Program 12						**+**			
	Study 1					+			

Effectiveness of Programs Across Studies

		ATOD use outcomes				*Antisocial behavior outcomes*			
Name	*Study*	*Alcohol Use*	*Tobacco Use*	*Marijuana Use*	*Substance use*	*Aggressive/ Disruptive*	*Antisocial Behavior*	*Delinquent Behavior*	*Physical Aggression*
Program 13						0.133			
	Study 1					0.133			
Program 14						**0.579**			**0.93**
	Study 1					0.064			
	Study 2					**0.483**			
	Study 3					**1.286**			**1.178**
	Study 4					**0.567**			**0.68**
Program 15						0.135			
	Study 1					0.135			
	Study 2					θ			
	Study 3					θ			
Program 16		**0.216**	**+**	**+**	0.116	**+**			
	Study 1	**0.216**	+	+	0.116	+			
Program 17		–0.059	**0.336**	θ	**0.142**	**+**			
	Study 1		**0.343**						
	Study 2				**0.127**				
	Study 3		**0.766**						
	Study 4	+	θ						
	Study 5	–0.059							
	Study 6		**0.162**						
	Study 7		**0.365**						
	Study 8		**0.496**						
	Study 9		θ	+	+	+			
	Study 10		θ	θ	+	+			
	Study 11	+	+	θ	**0.158**				
	Study 12		θ						
	Study 13		**0.166**						
	Study 14		θ						
	Study 15	θ	**+**	+					
	Study 16	θ	θ	θ					
	Study 17	+	**+**	+					
Program 18						0.126			
	Study 1					**0.24**			
	Study 2					–0.032			
	Study 3					θ			
Program 19						θ			
	Study 1					θ			

(*continued*)

Table 44.1 Continued

Effectiveness of Programs Across Studies									
		ATOD use outcomes				*Antisocial behavior outcomes*			
Name	*Study*	*Alcohol Use*	*Tobacco Use*	*Marijuana Use*	*Substance use*	*Aggressive/ Disruptive*	*Antisocial Behavior*	*Delinquent Behavior*	*Physical Aggression*
Program 20		0.255	**0.488**	0.238			0.026	0.24	
	Study 1	0.255	**0.488**	0.238				0.24	
	Study 2						0.026		
Program 21		+	+					+	
	Study 1	+	+					+	
Program 22					0.098		0.075		0.07
	Study 1				θ				
	Study 2				θ				θ
	Study 3				0.098		0.075		0.07
Program 23						?	**0.831**		
	Study 1					θ			
	Study 2					θ			
	Study 3								
	Study 4				–	–			
	Study 6						**1.082**		
	Study 7						0.568		
	Study 9					θ			
Program 24		0.036	-0.121	0.073	**0.155**				θ
	Study 1	0.036	-0.121	0.073	**0.155**				θ
	Study 2		θ	θ					
	Study 3	θ	θ	θ	+				
	Study 4	θ	θ	θ	θ				

Key
Programs are not identified at the request of the funder
Bold indicates the result is significant
Shaded numeric findings are random effects modeled grand means if multiple studies contributed data
+ indicates a significant positive finding from evidence that could not be standardized
– indicates a significant negative finding from evidence that could not be standardized
θ indicates a null finding from evidence that could not be standardized
? indicates mixed findings across multiple non-standardized results

proximal outcomes to multiple distal outcomes using individual-level data are needed to ascertain whether or not the observed changes occur in different pools of individuals or if individuals are experiencing significant declines in multiple behaviors. Moreover, study trials need to provide more comprehensive results of their findings. It is currently the exception, and not the rule, that authors provide evidence of effectiveness for multiple outcomes. Providing this evidence would allow for a finer comparison of the relative effectiveness of delivering multi-target programs or multiple single-target programs to at-risk populations.

By focusing primarily on distal outcomes, the fuller picture available to prevention research may be clouded. Due to the political climate, outcomes research in the US and Canada has largely focused on mechanism of change and intermediate outcomes (McBride, 2003). The present study documents that despite the plethora of programs being advertized as effective, the

Table 44.2 Implications for practice

There is relatively thin evidence of the effectiveness of school-based programs in reducing ATOD and antisocial behaviors from well-conducted studies of "off-the-shelf" programs with multiple trials.
Although each of these programs have been tested in multiple trials, there are relatively few replications of particular outcomes.
There is relatively little support for programs being effective for multiple antisocial behavior and violent outcomes.
Compared to other reviews and required data studies these trials report relatively few negative findings.
The warrant for choosing evidence-based programs to reduce youth antisocial behavior and violent behaviors remains elusive.

actual evidence of their effectiveness across multiple trials and outcomes is thinner than might be imagined given the current support for evidence-based prevention programming. Implications for practice considering the results of meta-analysis and systematic review of program effectiveness are delineated in Table 44.2. Across the spectrum of programs examined, many have not been evaluated by two or more independent trials, many do not report evidence of their effectiveness on distal outcomes and, among those that do report effectiveness evidence, many do so using methods that limit the utility of their findings for comparative effectiveness research. From this perspective, researchers are encouraged to include main effects estimates in their reports for all distal outcomes their interventions are designed to impact. Providing these data will permit answering definitively what programs and under what conditions programs are effective in improving multiple outcomes for youth.

Notes

This project has been funded at least or in part with Federal funds from the U.S. Department of Education under contract number ED-04-CO-0059/0027. The content of this publication does not necessarily reflect the views or policies of the U.S. Department of Education nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

Bibliography of reports used to code findings are available from the authors and online.

1. Program names are omitted at the request of the funder of this research.

References

Alford, A. A., & Derzon, J H. (2011). *A meta-analytic evaluation of the effectiveness of targeting mediating factors to reduce violent and aggressive outcomes.* Unpublished manuscript.

Alford, A. A., Derzon, J., Hagen, C. A., & Crosse, S. (2011). Do publically available school-based intervention programs effectively impact more than one outcome? Unpublished manuscript.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis.* West Sussex, UK: Wiley. doi: 10.1002/9780470743386

Cooper, H., & Hedges, L.V. (1994.). *The handbook of research synthesis.* New York, NY: Russell Sage Foundation.

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials, 7,* 177–188. doi: 10.1016/0197-2456(86)90046-2

Derzon, J. H. (2006). How effective are school-based violence prevention programs in preventing and reducing violence and other antisocial behaviors? A meta-analysis. In S. R. Jimerson & M. J. Furlong (Eds.), *Handbook of school violence and school safety: From research to practice* (pp. 429–442). Mahwah, NJ: Erlbaum.

Derzon, J. H. (2007). Using correlational evidence to select youth for preventive interventions. *Journal of Primary Prevention, 28,* 421–447. doi: 10.1007/s10935-007-0107-7

Derzon, J. H. Springer, F., Sale, L., & Brounstein, P. (2005). Estimating intervention effectiveness: Synthetic projection of field evaluation results. *Journal of Primary Prevention, 26,* 321–343. doi: 10.1007/s10935-005-5391-5

Domitrovich, C. E., Bradshaw, C. P., Greenberg, M. T., Embry, D., & Ialongo, S. I. (2009).

Integrated models of school-based prevention: Logic and theory. *Psychology in the Schools*, *47*(1), 71–88. doi: 10.1080/19415531003616862

Ennett, S. T., Ringwalt, C. L., Thorne, J., Rohrbach, L. A., Vincus, A., Simons-Rudolph, A., & Jones, S. (2003). A comparison of current practice in school-based substance use prevention programs with meta-analysis findings. *Prevention Science, 4,* 1–14. doi: 10.1023/A:1021777109369

Hansen, W. B., Derzon, J., Dusenbury, L, Bishop, D., Campbell, K., & Alford, A. (2009). Operating characteristics of prevention programs: Connections to drug etiology. In L. M. Scheier (Ed.), *Handbook of drug use etiology* (pp. 597–616). Washington, DC: American Psychological Association.

Hansen, W. B., Dusenbury, L., Bishop, D., & Derzon, J. H. (2007). Substance abuse prevention program content: Systematizing the classification of what programs target for change. *Health Education Research, 22,* 351–360. doi: 10.1093/her/cyl091

Hawkins, J. D., Catalano, R. F., & Miller, J. Y. (1992). Risk and protective factors for alcohol and other drug problems in adolescence and early adulthood: Implications for substance abuse prevention. *Psychological Bulletin, 112,* 64–105. doi: 10.1037/0033-2909.112.1.64

L. V. Hedges (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*(2), 107–128.

Jessor, R., & Jessor, S. L. (1977). *Problem behavior and psychosocial development: A longitudinal study of youth*. New York, NY: Academic Press.

Keef, S. P., & Roberts, L. A. (2004). The meta-analysis of partial effect sizes. *British Journal of Mathematical and Statistical Psychology*, *57*, 97–129.

Lipsey, M. W and Derzon, J. H. (1998) Predictors of serious delinquency in adolescence and early adulthood: A synthesis of longitudinal research. In R. Loeber & D. P. Farrington (Eds.), *Serious and violent offenders: Risk factors and successful interventions* (pp. 86–105). Thousand Oaks, CA: Sage.

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology, 58,* 593–614. doi: 10.1146/annurev.psych.58.110405.085542

McBride, N. (2003). A systematic review of school drug education. *Health Education Research, 18*(6), 729–742. doi: 10.1093/her/cyf050

Mihalic, S. (2008). Matrix of programs as identified by various federal and private agencies. Retrieved August 2, 2010, from http://www.colorado.edu/cspv/blueprints/matrix.html

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups design. *Psychological Methods*, *7*(1), 105–125. doi: 10.1037/1082-989X.7.1.105

Najaka, S. B. (2000). *A meta-analytic inquiry into the relationship between risk-factors and problem behavior.* Dissertation, University of Maryland, Maryland Department of Criminology and Criminal Justice, College Park.

Najaka, S. S., Gottfredson, D. C., & Wilson, D. B. (2001). A meta-analytic inquiry into the relationship between selected risk factors and problem behavior. *Prevention Science, 2,* 257–271. doi: 10.1023/A:1013610115351

Pearl, J. (2000). *Causality: Models, reasoning and inference.* New York, NY: Cambridge University Press. doi: 10.1017/S0266466603004109

Peterson, R. A., & Brown, S. (2005). On the use of beta coefficients in meta-analysis. *Journal of Applied Psychology*, *90*(1), 175–181. doi: 10.1037/0021-9010.90.1.175

Silberg, J. Rutter, M., D'Onofrio, & B., Eaves, L. (2003). Genetic and environmental risk factors in adolescent substance use. *Journal of Child Psychology and Psychiatry*, *44*(5), 664–676. doi: 10.1111/1469-7610.00153

Wei, E. H., Loeber, R., & White, H. R. (2004). Teasing apart the development associations between alcohol and marijuana use and violence. *Journal of Contemporary Criminal Justice*, *20*(2), 166–183. doi: 10.1177/1043986204263777

Wilson, D. B., Gottfredson, D. C., & Najaka, S. S. (2001). School-based prevention of problem behaviors: A meta-analysis. *Journal of Quantitative Criminology, 17,* 247–272. doi: 10.1023/A:1011050217296

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. doi: 10.1037/0022-006X.71.1.136