Efficacy vs Effectiveness Trial Results of an Indicated "Model" Substance Abuse Program: Implications for Public Health

Denise Hallfors, PhD, Hyunsan Cho, PhD, Victoria Sanchez, DrPH, Shereen Khatapoush, PhD, Hyung Min Kim, MA, and Daniel Bauer, PhD

Objectives. The US Department of Education requires schools to choose substance abuse and violence prevention programs that meet standards of effectiveness. The Substance Abuse and Mental Health Services Agency certifies "model" programs that meet this standard. We compared findings from a large, multisite effectiveness trial of 1 model program to its efficacy trial findings, upon which the certification was based.

Methods. 1370 high-risk youths were randomized to experimental or control groups across 9 high schools in 2 large urban school districts. We used intent-to-treat and on-treatment approaches to examine baseline equivalence, attrition, and group differences in outcomes at the end of the program and at a 6-month follow-up.

Results. Positive efficacy trial findings were not replicated in the effectiveness trial. All main effects were either null or worse for the experimental than for the control group.

Conclusions. These findings suggest that small efficacy trials conducted by developers provide insufficient evidence of effectiveness. Federal agencies and public health scientists must work together to raise the standards of evidence and ensure that data from new trials are incorporated into ongoing assessments of program effects. (*Am J Public Health.* 2006;96:2254–2259. doi:10.2105/AJPH.2005.067462)

Substance abuse prevention research has made considerable strides during the past 40 years. Prevention investigators now implement rigorous epidemiologically based designs and theoretically informed interventions to test whether decreasing hypothesized risk factors and increasing protective factors will result in lower adolescent substance use.[1] Prevention science has attempted to follow a logical progression in program testing, including hypothesis development, pilot studies, efficacy trials, effectiveness trials, and dissemination studies. This process was advocated in a seminal article by Flay[2] describing the phases of research in the development of health promotion programs. Flay's article helped to establish prevention research and enhance acceptability among other disciplines, but some have argued that it also had an unanticipated negative effect—the proliferation of many small-scale efficacy studies of unknown generalizability, which were only rarely followed by effectiveness trials.[3]

Efficacy trials are a test of whether an intervention does more good than harm when delivered under optimal conditions.[2] Program implementers are usually highly trained professionals, and the conditions are closely monitored to ensure that targeted participants receive a maximal dose of the intervention. Effectiveness trials, on the other hand, test whether an intervention does more good than harm under *real-world* conditions. For example, in school-based effectiveness trials, regular school personnel are recruited as implementers. Implementation variation is expected, and measuring the amount and quality of intervention delivered is an essential component in the analysis of impact.[1] Flay, however, differentiated between a *treatment effectiveness trial,* in which implementation fidelity is maintained as much as possible, and an *implementation effectiveness trial,* in which implementation is allowed to vary naturally or by planned comparison.[2]

We conducted a treatment effectiveness trial. The trial evaluated Reconnecting Youth, an "indicated" drug abuse prevention program. Unlike universal programs that target all youths, indicated prevention programs target individuals who have already started exhibiting problem behaviors.[4,5] Reconnecting Youth consists of a 1-semester class, taken for academic credit, with the objective of improving academic achievement (i.e., grade point average [GPA] and attendance), preventing or reducing illegal drug use, and improving mood management (depression, anger, and anxiety). Developers reported a number of significant positive findings from the Reconnecting Youth efficacy trial.[6] Compared with the control group, the experimental group showed decreased drug control problems and consequences; decreased hard drug use; increased GPA; increased self-esteem; increased school bonding; and decreased deviant peer bonding. No negative effects (e.g., better gains for the comparison than the experimental group) were reported. Findings were strongest immediately after the intervention and tended to deteriorate at 6 months.

On the basis of the efficacy trial data, Reconnecting Youth was designated as a "model program" by the Substance Abuse and Mental Health Services Administration, US Department of Health and Human Services (SAMHSA),[7] and as a "research-based drug abuse prevention program" by the National Institute on Drug Abuse.[8] Such designations are critical to the successful diffusion of prevention intervention programs, because federal education policy requires school districts that receive Safe and Drug-Free School funds to select programs on the basis of research evidence of program effectiveness.[9]

Our study is rare; few prevention programs have been tested beyond the small efficacy trials conducted by program developers and

implemented under ideal conditions. When positive effects from efficacy trials are reported, programs are deemed "model" and marketed as such, usually without independent replication and further testing and development. Yet very little is known about how representative or robust the results are from such studies.[3] In fact, evidence from published efficacy trials indicates that effects are not robust. Tobler et al.[10] found a large drop (from 0.35 to 0.08) in the effect size of school-based interactive programs as the number of students in the study increased. Their explanation was that the fidelity of implementation decreased when programs went to scale. Lipsey,[11] in a meta-analysis of juvenile delinquency interventions, found that researcher implementation under optimal conditions made the largest single contribution to the R^2 change in effect size, adding 0.11.

When the effectiveness of Reconnecting Youth was tested, 2 important differences may have affected comparisons with the original efficacy trial. First, the original Reconnecting Youth trial was not a true randomized control trial. Although students meeting high-risk criteria were randomly selected for participation and randomly assigned to experimental and control conditions, assignment was accomplished before invitation, resulting in a differential refusal rate. Refusals were higher among the experimental group than among the control group, and nonequivalence in outcome variables was found at baseline, with the experimental group showing lower GPA, higher truancy, and higher drug use.[6] The key advantage of the randomized control design is that random allocation of participants increases the likelihood that unknown factors are equally distributed among the experimental and control groups, reducing the potential for bias. Meta-analyses,[10,11] however, have found that random assignment alone has a negligible impact on effect sizes when compared with studies with well-matched comparison groups.

Second, the efficacy trial analyses had a total sample size of 259 youths from 4 high schools in 1 school district. Our study included 1370 youths from 9 high schools in 2 demographically different school districts. The larger sample size and replication by independent evaluators implies the likelihood of smaller effect sizes, approximating what might be experienced by end users who attempt to implement with fidelity.

An important consideration in evaluating evidence from a randomized control trial is the analysis. The intent-to-treat approach is believed to provide the most reliable and unbiased evidence about effectiveness.[12,13] This approach compares all randomly assigned experimental subjects to control subjects, regardless of whether they actually met the eligibility requirements, actually received the treatment, experienced crossover to some other treatment, or in some other way deviated from the research protocol. Thus, the analysis is based on the treatment that they were intended to receive. Generally, it is assumed that the intent-to-treat approach is the most internally valid measure of the true treatment effect because it preserves the baseline equivalence in the study and minimizes the likelihood of type 1 error.[14] It is also more externally valid, because the level of adherence in the trial is similar to that which can be expected in the community.[15]

Although the intent-to-treat approach provides unbiased comparisons, in order to avoid dilution of the treatment effect, studies often include analyses that take into account the amount of treatment respondents actually received. This is known as the actual treatment approach, on-treatment approach, or explanatory approach. Explanatory analyses may yield relevant information and useful clarifications in the evaluation of treatment but also have the potential to introduce serious bias. Bias arises from the evidence that participants who adhere to treatment tend to do better than those who do not adhere. Excluding nonadherent participants from the analysis leaves those who may be more likely to have a better outcome and destroys the unbiased comparison afforded by randomization.[16]

We report both types of analyses and compared findings to those in the Reconnecting Youth efficacy trial. This large, independent replication of a model indicated prevention program represents an important contribution to knowledge about intervention effects for high-risk adolescents.

METHODS

Participants

Participants in the study were 9th- to 11th-grade students recruited from 9 high schools in 2 large urban school districts in the United States. One school district was located in a large city in the Southwest (site A) and the other in a large metropolis on the Pacific coast (site B). Four schools in site A and 5 schools in site B participated in this study. About 49% of the participants were male. In site A, 87% of the participants were Hispanic, 9% were Black, 4% were White, and 4% were American Indian or other race. The vast majority (90%) qualified for the federal free or reduced-price lunch program. In site B, 40% were Asian/Pacific Islander, 21% were Hispanic, 15% were Black, 10% were White, and 12% were American Indian or other race. Sixty-one percent of site B students qualified for free or reduced-price lunches. Study participants were randomly selected from each school's pool of students identified as being at high risk for school dropout. Criteria for high-risk status included being in the top 25% for truancy and bottom 50% for GPA, or being referred for participation by a school teacher or counselor. Students assented to participate in the study with signed parental consent. Participants who completed the study's questionnaire at the study's baseline were randomly assigned to 1 of 2 conditions: the experimental group, consisting of high-risk students who were assigned to take the Reconnecting Youth class, or the control group, consisting of high-risk students not assigned to take the class. A total of 1370 high-risk students participated in the Reconnecting Youth study for 2002–2004, with 695 students assigned to the experimental group and 675 students assigned to the control group across the 2 sites.

Intervention

Reconnecting Youth classes were conducted at each participating school. Classes were offered during regular school hours for elective credit as a half-year course. The curriculum included 55 core lessons and 24 booster lessons focusing on 4 main themes: self-esteem, decisionmaking, personal control, and interpersonal communication. Students

monitored their school attendance, academic achievement, drug use, and moods, and set personal goals in each of these areas. Students practiced skills related to the 4 main themes and their goals and also practiced giving and receiving positive social support. The targeted student/teacher ratio for the Reconnecting Youth class was 10 to 12 students per teacher. During fall 2002 through fall 2004, a total of 41 Reconnecting Youth classes were conducted: 21 in 5 schools in site A and 20 classes in 4 schools in site B. Teachers were trained by developers, according to protocol. Additional supervision was provided by local coordinators as well as 2 site research coordinators, who also monitored implementation.

Measures

The High School Questionnaire (HSQ) is an instrument developed by Eggert et al.[6,17] to evaluate Reconnecting Youth. The HSQ was administered at 3 time points: at student invitation (time 1), at the end of the next semester (time 2), and at the end of the following semester (time 3). HSQ multi-item scales have demonstrated acceptable reliability and validity in both the efficacy trial[6] and the present effectiveness trial.[18] For our study, the questionnaire was adapted to an audio computer-assisted interview format. Students who left school or moved away from their school districts were sent a paper version of the survey at follow-up.

A total of 25 outcome and mediator variables were examined. Academic performance variables of GPA and truancy were obtained from school records; all other variables were self-reported, including (1) substance use outcomes including past 30-day alcohol use, marijuana use, and cigarette smoking; and composite indexes of other drug use, adverse drug consequences, drug control problems, and drug use progression; (2) emotional states including hopelessness, stress, anxiety, anger, self-esteem, depression, and perceived acceptability of suicide; (3) behavior indicators including delinquent behaviors, prosocial weekend activities, and partying; and (4) posited mediators such as school connectedness, conventional peer bonding, high-risk peer bonding, personal coping strategies, personal control, and perceived family support. In addition, demographic variables assessing age, gender, and grade (9th to 11th) were also included.

Statistical Analyses

Preliminary analyses were conducted to examine attrition rates by group and site at each time point. We also examined whether dropouts at the follow-up time point differed significantly between groups (experimental vs control group; on-treatment vs control group), on either outcome measures or demographic characteristics from study adherers.

Substantive data analyses were conducted in 3 phases: (1) group equivalence; (2) group differences at time 2; and (3) group differences at time 3. Analyses were conducted with both the intent-to-treat approach and the on-treatment approach. Group equivalence was evaluated using time 1 demographic and outcome variables. Significance tests were conducted using t tests for the continuous variables and χ^2 tests for the categorical variables. Relations were considered significant at $P<.05$. Outcome changes before and after the intervention were then compared between all experimental and control students (intent-to-treat approach), and then just the students who attended 50% or more of the Reconnecting Youth class and control students (on-treatment approach). Analysis of covariance (ANCOVA) was used (with SAS PROC GLM [SAS Institute, Cary, NC]) to evaluate program effect with control for the outcome value at time 1, as well as for gender and grade. These models estimated the effects of intervention (experimental vs control group; on-treatment vs control group), and interaction between intervention and school. Immediate postprogram effects were assessed by comparing time 2 data across groups, and 6-month follow-up effects by comparing time 3 data across groups.

An important consideration for these analyses was that students were nested within schools, producing a probable violation of the assumption that the residuals of the ANCOVA model were independent (e.g., a common school "culture" would lead the residuals of students from the same school to be correlated). One approach to addressing this lack of independence would be to use a multilevel model. Multilevel models assume that the upper-level units are sampled from a population and use this sample to estimate the characteristics of that population. In this case, however, schools were not sampled randomly from a population but were chosen explicitly, making this a fixed factor in the study design. Moreover, the number of upper-level units (schools) was too few (9) to make multilevel modeling feasible.[19] As such, we controlled for differences across schools by including "school" as a fixed factor in the ANCOVA models (a strategy sometimes referred to as the fixed-effects approach to analyzing multilevel data). By conditioning on school membership, the residuals of the ANCOVA models could reasonably be assumed to be independent. Because schools were nested within sites, contrast codes for the school factor were used to test differences in program effects across the 2 sites.

RESULTS

Attrition

A total of 1220 students completed the HSQ at time 2 and 1178 completed at time 3, yielding an overall attrition rate of 10.9% and 14.0%, respectively (Table 1). Attrition rates by group (experimental and control) were not significant ($P>.05$) at time 2 or time 3. Likewise, no differences in attrition by site were found. For the on-treatment approach, attrition was lower in the on-treatment group versus control at time 2 ($\chi^2=6.91$, $P=.01$), and at time 3 ($\chi^2=3.99$, $P=.05$); site differences were not significant at either time point.

Compared with those who remained in the study, dropouts were more likely to be male and riskier by indicators such as lower GPA and school attendance; higher depression; and higher cigarette and marijuana use. However, no significant differences were found when we compared dropouts from the experimental and control groups at either time point.

Reconnecting Youth Class Attendance

Across both sites, 47% of students assigned to the experimental group attended at least 50% of Reconnecting Youth classes. The main reasons for experimental protocol noncompliance were a move from the school, class schedule conflicts, or counselor refusal to place the student in an elective class because the student was behind in core credits. Another 6% (n=43) enrolled in the class but attendance was less than 50%.

TABLE 1—Attrition Rate Comparison, by Group: United States, 2002–2004

	Experimental Group					Control Group					On-Treatment Group				
	Time 1, No.	Time 2, No.	Time 3, No.	T1–T2, %	T1–T3, %	Time 1, No.	Time 2, No.	Time 3, No.	T1–T2, %	T1–T3, %	Time 1, No.	Time 2, No.	Time 3, No.	T1–T2, %	T1–T3, %
Site A	349	311	307	10.89	12.03	335	305	298	8.96	11.04	158	154	146	2.53	7.59
Site B	346	297	290	14.16	16.18	340	307	283	9.71	16.76	169	158	150	6.51	11.24
Total	695	608	597	12.52	14.10	675	612	581	9.33	13.93	327	312	296	4.59	9.48

TABLE 2—Intent-to-Treat Analyses: Significant Results of Analysis of Covariance Immediately After Intervention and at 6-Month Follow-Up

Outcomes	Experimental Group, Adjusted Mean (SE)	Control Group, Adjusted Mean (SE)	Group F Value (*P*)	School × Group F Value (*P*)
Immediately after intervention (time 2) (n = 1220)[a]				
Progression of drug use	1.98 (0.08)	2.02 (0.08)	. . .	2.28 (.02)
6-month follow-up (time 3) (n = 1178)[b]				
Conventional peer bonding	1.78 (0.03)	1.87 (0.03)	5.24 (.02)	. . .
High-risk peer bonding	1.76 (0.04)	1.62 (0.04)	8.09 (.01)	. . .
Prosocial weekend activities	3.87 (0.10)	4.23 (0.10)	6.89 (.01)	. . .

Note. Predictors = group and school × group; covariates = baseline score, grade, and gender.
[a]n = 608 in the experimental group and n = 612 in the control group.
[b]n = 597 in the experimental group and n = 581 in the control group.

Baseline Equivalence

We examined the baseline equivalence between the experimental and the control group and between the on-treatment and the control group on demographic and outcome variables. Attendance (84% vs 86%, *P*=.01) was lower and smoking (0.67 vs 0.49, *P*=.02) was higher in the experimental group than in the control group. Students in the on-treatment group had lower personal control than those in the control group (1.91 vs 2.01, *P*=.02) and were younger (14.99 vs 15.21, *P*=.00).

Reconnecting Youth Program Effects

Table 2 reports the results from the intent-to-treat analyses. The table presents the adjusted time 2 means for the experimental and control groups, with control for time 1 values and other covariates. Only significant effects are reported. Immediately after the intervention, no main program effects were found, and only 1 variable (progression of drug use) showed a school-by-group interaction effect. That is, small positive effects were found in 4 schools, whereas small negative effects were found in 5 schools. At the 6-month follow-up, 3 negative main program effects were found (conventional peer bonding, high-risk peer bonding, and prosocial weekend activities).

Table 3 reports significant adjusted mean differences from the on-treatment analyses. Immediately after the intervention, 1 negative main program effect was found (anger). Two school-by-group interaction effects were found (perceived family support and smoking). For perceived family support, 3 schools showed positive program effects and 6 schools showed negative effects. For smoking, 5 schools showed positive effects, and 4 schools showed negative effects. At 6-month follow-up, negative main program effects were found on high-risk peer bonding and prosocial weekend activities. One school-by-group interaction effect was found (school connectedness). Four schools showed positive effects and 5 schools showed negative effects.

Using contrast codes for the school effect, we tested whether the program effects differed by site. With the intent-to-treat approach, we found 2 significant site-by-group interaction effects, on progression of drug use (F=4.65, *P*=.03) and smoking (F=5.29, *P*=.02) immediately after the intervention;

TABLE 3—On-Treatment Analysis: Significant Results of Analysis of Covariance Immediately After Intervention and 6-Month Follow-Up

Outcomes	On-Treatment Group, Adjusted Mean (SE)	Control Group, Adjusted Mean (SE)	Group F Value (*P*)	School × Group F Value (*P*)
Immediately after intervention (time 2) (n = 924)[a]				
Number	312	612	. . .	. . .
Anger	1.39 (0.05)	1.22 (0.03)	7.87 (.01)	. . .
Perceived family support	1.58 (0.05)	1.61 (0.03)	. . .	2.33 (.02)
Smoking	0.57 (0.08)	0.58 (0.05)	. . .	2.92 (.01)
6-month follow-up (time 3) (n = 877)[b]				
Number	296	581	. . .	. . .
High-risk peer bonding	1.82 (0.06)	1.62 (0.04)	8.84 (.01)	. . .
Prosocial weekend activities	3.84 (0.15)	4.23 (0.10)	4.52 (.03)	. . .
School connectedness	13.46 (0.22)	13.41(0.14)	. . .	2.02 (.04)

Note. Predictors = group and school × group; covariates = baseline score, grade, and gender.
[a]n = 312 in the on-treatment group and n = 612 in the control group.
[b]n = 296 in the on-treatment group and n = 581 in the control group.

we found no significant interaction effects at 6-month follow-up. With the on-treatment approach, we found 1 significant site-by-group interaction on direct coping strategies ($F=4.71$, $P=.03$) immediately after the intervention, and on hard drug use ($F=3.90$, $P=.05$) 6 months after the intervention. In all cases, site A had more schools with positive program effects than site B.

DISCUSSION

Findings from the Reconnecting Youth treatment effectiveness trial failed to replicate positive findings from the efficacy trial. Immediately after the intervention, intent-to-treat analyses showed no significant main effects, but 1 school-by-group interaction effect. Moreover, Reconnecting Youth failed to meet the requirement to do more good than harm. By the second follow-up period, the experimental group had worse outcomes than the control group on conventional peer bonding, high-risk peer bonding, and prosocial weekend activities.

The effectiveness trial also uncovered a number of difficulties with implementing a semester-long course for high-school youths at risk for dropout. First, school guidance counselors hesitated—and sometimes refused—to enroll such youths in an elective course because of their need to fulfill core credit courses or obtain academic remediation. High-risk youths tend to be very mobile, changing schools, dropping out for periods, or attending alternative schools. Thus, despite strenuous efforts by research and school staff, less than half the students assigned to the experimental group were actually exposed to a substantial portion of the intervention.

Given the low compliance with the intervention, on-treatment analyses were conducted to examine if there were any additional positive effects among students who actually attended the class. Instead, immediate posttreatment analyses showed worse anger outcomes for intervention students than for control students. There were no other main effects, but significant differences by school in perceived family support and smoking. At the 6-month follow-up, the experimental group had worse outcomes than the control group on high-risk peer bonding and prosocial weekend activities. In no case was the experimental group better off than the control group, and school differences appeared only for school connectedness.

There are 2 important implications of these findings. First, study findings suggest that efficacy trials conducted in 1 location by developers do not provide adequate evidence for widespread dissemination or designation as "model" programs. Because our study was a treatment effectiveness trial, the program was implemented with greater fidelity than can be expected under normal conditions,[18] and yet positive outcomes could not be replicated. According to the SAMHSA Web site, "Effective programs are well-implemented, well-evaluated programs, meaning they have been reviewed by the National Registry of Effective Programs (NREP) according to rigorous standards of research."[20] Model programs meet the further requirement of having high-quality materials and training available. Over 100 programs meet the criteria for effective programs, and about 60 of these are considered model programs. Reconnecting Youth was listed as a model program.

Second, negative outcomes may be related to iatrogenic effects of grouping high-risk youths.[21] It is particularly troubling to see increases in high-risk peer bonding and decreases in conventional peer bonding and prosocial weekend activities (such as doing homework, club or church activities, and family activities), because these are the posited mediators of prosocial behavioral change.

There are a number of limitations to this report. First, we used audio computer-assisted interviews instead of the original paper and pencil HSQ, and the survey was shortened in consultation with the developers. How this affected the comparison with efficacy trial results is unknown, but there are several advantages to the audio computer-assisted interview format. For example, the survey is read to respondents through headphones (a help to challenged readers), and there is less missing data.[18,22] Second, because we tested the program in 2 large, diverse, urban settings, findings may not be generalizeable to other types of settings. Given the school and site differences, it is possible that specific outcomes may be related to region or population.

It is important to note that the Reconnecting Youth program was developed more than 10 years ago. New federal No Child Left Behind legislation has placed increasing pressures on schools to focus on core academic competencies and also requires schools to choose effective programs for Safe and Drug-Free Schools (Title IV) funding. Schools are thus left with increasingly difficult strategic decisions, particularly in high schools. For example, should at-risk students attend a public health prevention program that addresses their substance use? Or should they take a remedial math or science class? Federal agencies, such as SAMHSA, provide a real service by examining the evidence of effectiveness and certifying prevention programs to inform decisionmakers. But our findings indicate that more needs to be done to provide accurate and useful information. To their credit, SAMHSA has recently made efforts toward such needed change.[23]

In conclusion, our findings suggest that small efficacy trials are not sufficient evidence for certifying that a program is ready for dissemination. Although public health advocates argue that schools should address health problems as a necessary prerequisite to learning, credible evidence of effectiveness (without harm) are needed. Federal agencies and public health scientists must work together to raise the standards of evidence and ensure that data from new trials are incorporated into ongoing assessments of program effects. ■

About the Authors

Denise Hallfors, Hyunsan Cho, and Hyung Min Kim are with Pacific Institute for Research and Evaluation, Chapel Hill, NC. Victoria Sanchez is with the School of Nursing and Daniel Bauer is with the Psychology Department, University of North Carolina, Chapel Hill. Shereen Khatapoush is with the Daniel Bryant Youth and Family Treatment Center, Council on Alcoholism and Drug Abuse, Santa Barbara, Calif.

Requests for reprints should be sent to Denise Hallfors, PhD, 1516 E Franklin St, Suite 200, Chapel Hill, NC 27514 (e-mail: hallfors@pire.org).

This article was accepted March 20, 2005.

Contributors

D. Hallfors is the principal investigator of the study, originated the ideas for this article, and supervised all aspects of its development and execution. H. Cho is the methodologist for the project and supervised the design and the analyses. V. Sanchez and S. Khatapoush coordinated data acquisition efforts. H.M. Kim conducted the

analyses. D. Bauer consulted on analyses for the revised article.

Acknowledgments

This research was supported by the National Institute on Drug Abuse (grant R01DA13666).

We acknowledge the hard work of our field teams and the cooperation of administrators and teachers in the 2 school districts.

Human Participant Protection

The study was approved by the institutional review board of Pacific Institute for Research and Evaluation and is reviewed annually. It was last approved on July 20, 2005.

References

1. Kellam SG, Langevin DJ. A framework for understanding "evidence" in prevention research and programs. *Prev Sci.* 2003;4:137–153.

2. Flay BR. Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Prev Med.* 1986;15:451–474.

3. Glasgow RE, Klesges LM, Dzewaltowski DA, Bull SS, Estabrooks P. The future of health behavior change research: what is needed to improve translation of research into health promotion practice? *Ann Behav Med.* 2004;27:3–12.

4. Institute of Medicine. *Reducing Risks for Mental Disorders.* Washington, DC: National Academy Press; 1994.

5. *Preventing Drug Use Among Children and Adolescents: A Research Based Guide.* Washington, DC: National Institute on Drug Abuse, National Institutes of Health; 1997.

6. Eggert LL, Thompson EA, Herting JR, Nicholas LJ, Dicker BG. Preventing adolescent drug abuse and high-school dropout through an intensive school-based social network development program. *Am J Health Promot.* 1994;8:202–215.

7. Schinke S, Brounstein P, Gardner S. *Science-Based Prevention Programs and Principles, 2002.* Rockville, Md: Center for Substance Abuse and Prevention, Substance Abuse and Mental Health Services Administration; 2002. DHHS publication SMA 03-3764.

8. Robertson EB, David SL, Rao SA. *Preventing Drug Use Among Children and Adolescents: A Research-Based Guide for Parents, Educators, and Community Leaders.* 2nd ed. Washington, DC: National Institutes of Health, National Institute on Drug Abuse; 2003. NIH publication 04-4212(A).

9. No Child Left Behind Act of 2001, Pub L No. 107-110, 115 Stat 1425 (2002). Available at: http://www.ed.gov/legislation/ESEA02/. Accessed April 20, 2006.

10. Tobler NS, Roona MR, Ochshorn P, Marshall DG, Streke AV, Stackpole KM. School-based adolescent drug prevention programs: 1998 meta-analysis. *J Primary Prev.* 2000;20:275–336.

11. Lipsey MW. Juvenile delinquency treatment: a meta-analytic inquiry into the variability of effects. In: Cook TD, Cooper H, Cordray DS, et al., eds. *Meta-Analysis for Explanation: A Casebook.* 1st ed. New York, NY: Russell Sage Foundation; 1992:83–127.

12. Boruch RF. *Randomized Experiments for Planning and Evaluation: A Practical Guide.* Thousand Oaks, Calif: Sage Publications; 1997.

13. Rossi PH, Freeman HE, Lipsey MW. *Evaluation: A Systematic Approach.* 6th ed. Thousand Oaks, Calif: Sage Publications; 1999.

14. Fergusson D, Aaron SD, Guyatt G, Herbert P. Post-randomization exclusions: the intent to treat principle and excluding patients from analysis. *BMJ.* 2002; 325:652–654.

15. Montori VM, Guyatt GH. Intention-to-treat principle. 2001;165:1339–1341.

16. Altman DG. Clinical trials. In: *Practical Statistics for Medical Research.* London, England: Chapman & Hall; 1991:440–476.

17. Eggert LL, Thompson EA, Herting JR, Nicholas LJ. Prevention research program: reconnecting at-risk youth. *Issues Ment Health Nurs.* 1994;15:107–135.

18. Cho H, Hallfors D, Sanchez V. Evaluation of a high-school peer group intervention for at-risk youth. *J Abnorm Child Psychol.* 2005;33:363–374.

19. Snijder TB, Bosker R. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling.* Thousand Oaks, Calif: Sage Publications; 1999.

20. SAMHSA Model Programs. Effective Substance Abuse and Mental Health Programs for Every Community. Available at: http://modelprograms.samhsa.gov. Accessed April 17, 2005.

21. Dishion TJ, McCord J, Poulin F. When interventions harm: peer groups and problem behavior. *Am Psychol.* 1999;54:755–764.

22. Hallfors D, Watson K, Khatapoush S, Kadushin K, Saxe L. A comparison of paper vs. computer assisted self interview for school alcohol, tobacco, and other drug surveys. *Eval Program Plann.* 2000;23:149–155.

23. Advancing evidence-based practice through improved decision support tools: reconceptualizing NREPP, *71 Federal Register* 13132–13155 (2006).