Evaluation Review
Volume 33 Number 4
August 2009 396-414

10.1177/0193841X09334711
http://erx.sagepub.com
hosted at
http://online.sagepub.com

The Social Construction of "Evidence-Based" Drug Prevention Programs

A Reanalysis of Data From the Drug Abuse Resistance Education (DARE) Program

Dennis M. Gorman
J. Charles Huber, Jr.
Texas A&M Health Science Center School of Rural Public Health

This study explores the possibility that any drug prevention program might be considered "evidence-based" given the use of data analysis procedures that optimize the chance of producing statistically significant results by reanalyzing data from a Drug Abuse Resistance Education (DARE) program evaluation. The analysis produced a number of statistically significant differences between the DARE and control conditions on alcohol and marijuana use measures. Many of these differences occurred at cutoff points on the assessment scales for which post hoc meaningful labels were created. Our results are compared to those from evaluations of programs that appear on evidence-based drug prevention lists.

Keywords: *evidence-based practice; drug prevention; Drug Abuse Resistance Education (DARE); data analysis practices*

Authors' Note: This work was supported by grant number R01 NS 49611-01 from the National Institute of Alcohol Abuse and Alcoholism. We thank Dr. Richard R. Clayton for sharing his DARE data set with us. We also thank Dr. Mark J. Schlesinger for suggesting the idea of the social construction of solutions. The opinions expressed in this article are entirely those of the authors. Please address correspondence to Dennis M. Gorman, Department of Epidemiology & Biostatistics, School of Rural Public Health, TAMU 1266, College Station, TX 77843; e-mail: gorman@srph.tamhsc.edu.

Introduction

As in many areas of public policy, the idea of evidence-based practice has been enthusiastically embraced by the field of drug prevention during the last decade (Hallfors, Pankratz, and Hartman 2007; Weiss et al. 2008). A key component of this move toward evidence-based practice has been the production of "best practice" lists of approved drug prevention programs (e.g., Schinke, Brounstein, and Gardner 2002; US Department of Education Safe, Disciplined, and Drug-Free Schools Expert Panel 2002; National Institute on Drug Abuse 2003; National Registry of Evidence-based Programs and Practices 2008). It is claimed by those who develop these lists that there exists a body of scientific evaluation research that demonstrates the efficacy of the programs they recommend. Prevention practitioners, they argue, should start to use these programs and abandon those interventions for which no such scientific evidence exists. The Drug Abuse Resistance Education (DARE) program is probably the most well-known drug prevention intervention said to be unsupported by empirical evidence (Sherman et al. 1998; US General Accounting Office 2003; Weiss et al. 2008).

Against the almost overwhelming support for the development of best practice lists in the drug prevention field, there has emerged a small critical literature focused on both the methods and criteria used to select interventions (e.g., Gorman 2002; Petrosino 2003) and the quality of the evaluation research associated with those programs that appear most often on the lists. With regard to the latter, it has been argued that evaluations of many of the most well-known and advocated-for drug prevention programs use data analysis and presentation practices that serve to verify the hypothesis that the program "works" rather than to critically test this hypothesis (Gorman 2003, 2005a). These practices include multiple subgroup analysis, post hoc sample refinement, use of one-tailed significance tests, and changes in the way that outcome variables are constructed across publications from the same evaluation (Brown 2001; Gorman 2002; 2005b; Gandhi et al. 2007; Gorman, Conde, and Huber 2007; Midford 2008). The use of such practices raises concern that it is the data analysis techniques used to generate "evidence," rather than the actual interventions, which distinguishes "evidence-based" programs from their "unproven" counterparts.

The potential for such a problem to occur in the identification of best practices would be limited were stringent criteria used to evaluate a program's effects on behavioral outcomes. For the most part, the criteria used to establish the evidence base of programs selected by best practice lists has focused on three methodological issues: the quality of the study design used

to evaluate the program (e.g., whether random allocation to study conditions was used), the types of measures used to assess outcomes (e.g., whether they are reliable and valid), and the extent to which the study was successfully implemented (e.g., final sample size and attrition) (Gorman 2002; Petrosino 2003). Absent from these criteria is any assessment of data analysis practices. Moreover, those criteria that pertain to program effects on behavioral outcomes are easily met because they generally require demonstration of just a single statistically significant effect. The US Department of Education model program criteria, for example, required just "one evaluation that has demonstrated *an effect* on substance use, violent behavior, or other conduct problems one year or longer beyond baseline" (Safe, Disciplined, and Drug-Free Schools Expert Panel 1999, 5, emphasis added). The University of Colorado's *Blueprints*, which includes drug prevention programs among its model violence prevention programs, also only "requires *a sustained effect* at least 1 year beyond treatment, with no subsequent evidence that this effect is lost" (Mihalic et al. 2001, 3, emphasis added). Finally, the procedure used by the National Registry of Evidence-based Programs and Practices (NREPP) for designating a program as "evidence-based" simply requires that "The intervention demonstrates *one or more positive outcomes* ($p \leq .05$) in mental health and/or substance use behavior among individuals, communities, or populations" (Department of Health and Human Services 2007, 30815, emphasis added). None of these lists addresses how these single effects are produced in the data analysis process (e.g., whether the isolated effects reported are offset by many more null results or whether they are the outcome of data dredging).

Evaluators' use of data analysis techniques that maximize the potential to produce positive program effects coupled with the use of a very minimal standard by review systems to assign evidence-based status to prevention programs creates the possibility that almost any intervention could be rated as "evidence-based." In the most detailed analysis of this issue conducted so far, Gandhi et al. (2007) reviewed the evidence pertaining to five programs that were most often recommended in seven widely used best practice lists (including *Blueprints*, the US Department of Education list, and the NREPP list). They found that for four of the programs (CASASTART, Project ALERT, Project Northland, and the Midwestern Prevention Project), there was very limited evidence demonstrating their effectiveness in reducing substance use but that each program met the inclusion criteria of the lists on which it appeared because these allowed for isolated statistically significant effects from just one or two studies. Gandhi and colleagues were left wondering

whether these programs were in fact any more effective in preventing substance use and abuse than the DARE program.

Gandhi et al. (2007) also speculated as to whether the evaluations of DARE might have produced isolated statistically significant effects (thereby enabling the program to gain best practice status) if the analysis and presentation of data used in these had been conducted in a manner similar to the evaluations they reviewed (see also Gorman 2003). In the current study, we explored this possibility by reanalyzing data from a large-scale evaluation study of the DARE program, the original analyses of which had produced null findings for substance use outcome variables. Although the original evaluation used conservative data analysis procedures suitable for genuine hypothesis testing, we used some simple analysis techniques that increased the probability of finding a statistically significant difference between participants who received the DARE program and those in the control group. These practices are, however, ones that have been used in evaluations of programs that appear on a number of evidence-based drug prevention lists.

Methods

The Original DARE Evaluation Study Design, Measures, and Findings

The data used in the reanalysis were from the DARE evaluation conducted by Clayton and colleagues at the University of Kentucky (Clayton et al. 1991; Clayton, Catterello, and Johnson 1996; Lynam et al. 1999). The original study began in the 1987–1988 academic year and involved 2,071 6th grade students from 31 schools (Clayton, Catterello, and Johnson 1996). Twenty-three of the schools (1,550 students) were randomly assigned to receive the DARE program and eight schools (521 students) were assigned to the control condition. Participants were followed up immediately after the program (4 months after the baseline assessment) and each subsequent year until 10th grade (Clayton, Catterello, and Johnson 1996). The attrition rate rose from 7% at the immediate posttest to 44.8% at the 10th grade follow-up (Clayton, Catterello, and Johnson 1996). In addition, a final follow-up was conducted 10 years after the program was delivered, at which point 1,002 (48.4%) of the original 2,071 participants were reassessed (Lynam et al. 1999).

The main behavioral outcomes reported in the follow-up studies were lifetime, past year, and past month use of cigarettes, alcohol, and

marijuana. Each was assessed on a 7-point scale. In the case of alcohol and marijuana use, the scale items were 1 (0 times), 2 (1–2 times), 3 (3–5 times), 4 (6–9 times), 5 (10–19 times), 6 (20–39 times), and 7 (40 or more times) for each of the measures (i.e., lifetime, past year, and past month). For cigarette use, the items were 1 (none), 2 (<1 cigarette), 3 (1–5 cigarettes), 4 (6–10 cigarettes), 5 (11–20 cigarettes), 6 (21–30 cigarettes), and 7 (31 or more cigarettes) for each of the three measures.

The results of the original analyses conducted by Clayton and colleagues were consistent in showing that the DARE program had no effect on participants' use of cigarettes, alcohol, and marijuana. Clayton et al. (1991), for example, found that the program had no effect on self-reported drug use at the immediate posttest assessment. Summarizing the results of the 5-year follow-up, Clayton, Catterello, and Johnson (1996) observed that these were largely consonant with the findings of prior evaluations of the short-term effects of the program in showing no statistically significant effects with respect to cigarette, alcohol, or marijuana use. Finally, Lynam and colleagues concluded that the data from their 10-year follow-up added to the "accumulating literature on DARE's lack of efficacy in preventing or reducing substance use" (Lynam et al. 1999, 593). These results have been widely accepted within the drug prevention field and are among the reasons why the DARE program is considered an ineffective prevention strategy (Sherman et al. 1998; US General Accounting Office 2003; Weiss et al. 2008).

The Reanalysis of the DARE Data Set

We reanalyzed the University of Kentucky DARE data set using a number of procedures that have been used in recent evaluations of programs that appear on evidence-based lists of drug prevention programs. First, we used one-tailed tests in assessing the statistical significance of the differences between study conditions at follow-up. One-tailed tests are routinely used in evaluations of the Life Skills Training program as well as in evaluations of other "evidence-based" drug prevention programs (e.g., Botvin et al. 1995; Goldberg et al. 2000; Griffin et al. 2003; Spoth et al. 2006; Sun et al. 2006).[1] Second, we conducted two basic forms of subgroup analysis, one by gender and one by baseline drug use status (specifically restricting the analysis to cigarette, alcohol, and marijuana nonusers). The differential effects of drug prevention programs by gender are frequently reported in the evaluation literature, and a number of research reports restrict analyses

to baseline nonusers or separate out this group from baseline drug users (e.g., Perry et al. 1996; Botvin et al. 1999; Ellickson et al. 2003).

Although the original study involved six follow-up points, only five were included in our reanalysis. We excluded the 4-month posttest as the best practice lists typically specify that behavioral effects must be present a minimum of 1 year following the implementation of an intervention program.

As for the nine drug use outcome measures, we created six dichotomous variables for each by dividing the 7-point scale used to assess these at every possible cut point (i.e., 1 vs. 2–7, 1–2 vs. 3–7, 1–3 vs. 4–7, 1–4 vs. 5–7, 1–5 vs. 6–7, and 1–6 vs. 7). Continuous quantity and frequency measures of cigarette, alcohol, and marijuana use have been dichotomized and interpreted in terms of variables such as *ever smoked, past month drinking*, and *recent substance use* in a number of previous evaluations of drug prevention programs (e.g., Botvin et al. 1995, 1999; Eisen, Zellman, and Murray 2003; Sussman et al. 2003; Spoth et al. 2005). In the case of the current research, we simply ran every combination possible using the 7-point measurement scales, and we were not initially concerned with whether the dichotomous variables created could be meaningfully interpreted. However, once we identified statistically significant differences between the DARE and control group, we set about identifying those dichotomous variables that could be meaningfully interpreted in terms of categories such as *monthly use* or *weekly use* (see Results).

In the analysis, the proportion of respondents that reported the higher level of dichotomous cigarette, alcohol, or marijuana use was calculated for both the DARE and control groups. A logistic regression model was then used to compute odds ratios for the dichotomous substance use variable and treatment status adjusting for race ($0 = White$, $1 = non\text{-}White$). Odds ratios less than one indicate a beneficial effect of the DARE program. Wald chi-squared tests were used to assess the statistical significance of the adjusted odds ratios. This process was repeated for all possible cut points of the 7-point scales and for all five follow-up assessments.

Results

Analysis of all Dichotomous Variables

Table 1 shows the results of the analysis for cigarette, alcohol, and marijuana use for both the total sample and for the subgroup of individuals who were nonusers of each drug at baseline. Because we used six cutoff points

Table 1
Number of Statistically Significant Differences[a] Between DARE Condition and Control Condition Using Every Possible Cutoff Point for Each Outcome Variable[b]: All Participants and Baseline Nonusers[c]

	Baseline Users and Nonusers			Baseline Nonusers Only		
Drug Use Variable	All	Males	Females	All	Males	Females
Cigarettes						
Lifetime	0	0	0	0	0	0
Past Year	0	0	0	0	0	0
Past Month	0	0	0	0	0	0
Alcohol						
Lifetime	1	2	1	0	0	0
Past Year	2	3	0	2	5	0
Past Month	3	2	1	2	2	1
Marijuana						
Lifetime	1	5	1	3	6	1
Past Year	3	4	0	4	4	0
Past Month	5	4	1	5	4	1

a. One-tailed tests of significance (with $\alpha < .05$) were used in all the analyses.
b. For each cell of the table, 30 comparisons were made between the DARE condition and the control condition (6 cutoff points for each measure $\times$ 5 follow-up points). Of course, under normal procedures, when making so many post hoc comparisons between study conditions, some adjustment to *p* values should be made. When the Bonferroni adjustment was applied to the analysis presented in Table 1, none of the statistically significant differences presented in the table remained. However, such adjustments are seldom reported in the school-based drug prevention evaluation literature. In addition, the single statistically significant effect required by the evidence-based practice lists for designation of "model" or "exemplary" status does not have to be the product of a data analysis procedure that entailed adjustments for multiple post hoc comparisons.
c. The baseline nonuser analysis was restricted to those who reported no drug use at baseline.

for each variable and had five follow-up assessments, each cell of the table represents the number of statistically significant differences between the DARE group and control group that resulted from making a total of 30 comparisons. It can be seen that none of the 270 comparisons made for the total sample for cigarette use were statistically significant. Nor did any of the 270 comparisons made between DARE baseline nonusers and control nonusers produce a statistically significant result.

For the three sets of alcohol variables, there were a total of 15 statistically significant differences between the two study conditions, 4 for

lifetime use, 5 for past year use, and 6 for past month use. Among the baseline nonusers, there were 12 statistically significant differences between the DARE and control participants, 7 for the past year measures and 5 for the past month measures. In the gender subgroup analysis, there were more differences found for males than for females in both the total sample and baseline nonuser sample (7 vs. 2 and 7 vs. 1, respectively).

In the case of the three sets of marijuana variables, there were 24 statistically significant differences between the two study conditions found in the total sample analyses and 28 in the baseline nonuser subgroup analyses. As with the alcohol data, the gender subgroup analysis showed that there were more differences between the DARE and control participants for males than for females in both the total sample (13 vs. 2) and baseline nonuser sample (14 vs. 2).

Differences on "Meaningful" Dichotomous Variables

As noted in the Methods section, we created the six dichotomous variables from each of the original 7-point scales by simply dividing the scale at every possible cut point. Thus, the differences reported in Table 1 for the alcohol use measures and marijuana use measures are the result of this data dredging process and take no account of whether the dividing point on the scale creates a dichotomous variable that is in any way meaningful. It would, for example, be difficult to argue that it makes any sense dichotomizing the past month marijuana use measure at the point that separates those who report use five or fewer times from those that report use six or more times.

However, one obvious point at which to dichotomize all of the continuous scales is at the point that separates *no use* of the drug from *any use*. Indeed, in the case of lifetime alcohol and marijuana use, this is really the only meaningful point at which to dichotomize the 7-point scale. Because the maximum score on the lifetime scale was "40 or more," there were no other cut points that appeared to separate out meaningful categories of lifetime alcohol or marijuana users. The distinction between those who never used the drug and those reporting *any use* of it was also meaningful for the past month and past year measures of both alcohol and marijuana. In addition, for both the past year and past month measures, there were two additional dichotomized variables used in our analysis that could be interpreted in a meaningful way. In the case of the past year measure, the cut point "10–19 times" a year could reasonably be interpreted as *monthly use* of either alcohol or marijuana, and the cut point "40 or more times" could

Table 2
Statistically Significant Differences[a] Between DARE Condition and Control Condition on Dichotomous Measures of Ever Used Alcohol, Monthly Use of Alcohol, Weekly Use of Alcohol, and Daily Use of Alcohol: All Participants and Baseline Nonusers

				Proportion		Adjusted	
Sample	Follow-Up	Measure	Variable[b]	DARE	Control	Odds Ratio	*p* Value
All Participants							
Total	7th grade	Past year use	Weekly use	0.007	0.024	0.28	.008
Total	7th grade	Past month use	Weekly use	0.017	0.035	0.46	.021
Males	7th grade	Past year use	Weekly use	0.008	0.047	0.16	.002
Males	7th grade	Past month use	Weekly use	0.020	0.058	0.33	.007
Females	9th grade	Lifetime use	Any use	0.607	0.687	0.70	.039
Baseline Nonusers							
Total	7th grade	Past year use	Weekly use	0.004	0.018	0.21	.016
Males	7th grade	Past year use	Weekly use	0.005	0.035	0.15	.011
Males	7th grade	Past month use	Weekly use	0.014	0.044	0.30	.017

a. Using one-tailed tests of statistical significance.
b. The lifetime use *any use* variable was created by dichotomizing the 7-point scale at the 2nd point on the scale (i.e., those reporting use "1–2 times" or more were classified as "users" and those reporting "none" as "nonusers"). The past year *weekly use* variable was created by dichotomizing the 7-point scale at the final point on the scale (i.e., those reporting use "40 or more" times were classified as "weekly users"). The past month *weekly use* variable was created by dichotomizing the 7-point scale at the 4th point on the scale (i.e., those reporting use "6–9 times" or more were classified as "weekly users"). The 3rd point on the scale was not used as this states "3–5 times," which could refer to use less than once a week.

reasonably be interpreted as *weekly use* (this was the final point on the scale and so there was no option that specified a greater number of occasions). In the case of the past month measures, the cut point "6–9 times" could be reasonably interpreted as *weekly use* and the cut point "40 or more times" as *daily use*. Thus, there was a total of seven dichotomous variables that could be interpreted in terms of meaningful categories of both alcohol and marijuana use (lifetime—*any use*; past year—*any use*; past month—*any use*; past year—*monthly use*; past year—*weekly use*; past month—*weekly use*; and past month—*daily use*). Each of these variables could be assessed at each of the five follow-up points, making a total of 35 comparison between the DARE and control groups.

Table 2 presents the results for the seven alcohol use variables. All but one of the statistically significant differences between the DARE and

control groups occurred at the 7th grade follow-up and pertained to the *weekly use* variables derived from the past month and past year measures. These differences were observed for the total sample and for males in both the analysis that involved all participants (i.e., baseline users and nonusers) and the analysis confined to baseline nonusers. In the all-participant analysis, the male DARE participants were 67% less likely than controls to report *weekly use* of alcohol at 7th grade using the variable derived from the past month measure and 84% less likely using the variable derived from the past year measure. Comparable figures were found in the baseline nonuser analysis. The only difference found for females occurred in the analysis based on both baseline users and nonusers. This was for the *any use* variable (derived from the lifetime measure) at the 9th grade follow-up.

A total of 24 statistically significant differences between the DARE and control groups was found on the seven "meaningful" dichotomous marijuana variables, 12 in the analyses conducted with all participants (i.e., baseline users and nonusers) and 12 in the analyses conducted with just baseline nonusers. Table 3 presents the results for all participants. Differences at 7th grade were found for the *monthly use* variable derived from the past year measure (for both the total sample and males) and for the *any use* and *weekly use* variables derived from the past month measure (again, for both the total sample and males). In the total sample, the DARE group was 61% less likely than the controls to report *monthly use* (using the past year measure), 56% less likely to report *any use* (using the past month measure), and 71% less likely to report *weekly use* (using the past month measure). The comparable reductions among males were 71%, 70%, and 83%, respectively. The only long-term effect in the total sample was found at the 10-year follow-up on the *any use* variable derived from the past month measure (with DARE participants being 27% less likely to report use). Among males, there were also differences between the two groups on the *any use* variable derived from the past year measure at 7th grade and on *any use* variable derived from the lifetime measure at both 7th and 8th grade. The remaining two differences observed were for females at the 10-year follow-up on the *any use* variables derived from the lifetime and past month measures.

Table 4 shows that there was also a statistically significant difference between the DARE and control females at the 10-year follow-up on these same variables when the analysis was restricted to baseline nonusers. All but one of the other differences observed for the nonusers occurred in the total sample or male subsample at the 7th grade follow-up. Among males, there was a lower proportion reporting *any use* of the drug on all

Table 3
Statistically Significant Differences[a] Between DARE Condition and Control Condition on Dichotomous Measures of Ever Used Marijuana, Monthly Use of Marijuana, Weekly Use of Marijuana, and Daily Use of Marijuana: All Participants

				Proportion		Adjusted	
Sample	Follow-Up	Measure	Variable[b]	DARE	Control	Odds Ratio	*p* Value
Total	7th grade	Past year use	Monthly use	0.012	0.030	0.39	.015
Total	7th grade	Past month use	Any use	0.019	0.041	0.44	.011
Total	7th grade	Past month use	Weekly use	0.007	0.024	0.29	.008
Total	10 years	Past month use	Any use	0.250	0.312	0.73	.036
Males	7th grade	Lifetime use	Any use	0.093	0.145	0.61	.031
Males	7th grade	Past year use	Monthly use	0.014	0.046	0.29	.009
Males	7th grade	Past year use	Any use	0.056	0.098	0.54	.028
Males	7th grade	Past month use	Any use	0.022	0.070	0.30	.002
Males	7th grade	Past month use	Weekly use	0.006	0.035	0.17	.006
Males	8th grade	Lifetime use	Any use	0.128	0.188	0.63	.031
Females	10 years	Lifetime use	Any use	0.573	0.669	0.66	.027
Females	10 years	Past month use	Any use	0.200	0.290	0.60	.014

a. Using one-tailed tests of statistical significance.
b. The lifetime use, past year and past month *any use* variables were each created by dichotomizing the 7-point scale at the 2nd point on the scale (i.e., those reporting use "1–2 times" or more were classified as "users" and those reporting "none" as "nonusers"). The past year *monthly use* variable was created by dichotomizing the 7-point scale at the 5th point on the scale (i.e., those reporting use "10–19 times" or more were classified as "monthly users"). The past month *weekly use* variable was created by dichotomizing the 7-point scale at the 4th point on the scale (i.e., those reporting use "6–9 times" or more were classified as "weekly users").

three of the derived variables (i.e., lifetime, past year, and past month). The DARE participants were 47% less likely than the controls to report *any use* on the variable derived from the lifetime measure and 59% and 76% less likely on the *any use* variables derived from the past year and past month measures. There were also significant differences on the *monthly use* variable derived from the past year measure and the *weekly use* variable derived from the past month measure. Differences on these derived variables were also found in the total sample analysis, as were differences on the *any use* variables derived from the past year and past month measures. The DARE participants were 79% less likely than the controls to report *monthly use* on the variable derived from the past year measure and 86% less likely to report *weekly use* on the variable derived from the past month measure.

Table 4
Statistically Significant Differences[a] Between DARE Condition and Control Condition on Dichotomous Measures of Ever Used Marijuana, Monthly Use of Marijuana, Weekly Use of Marijuana, and Daily Use of Marijuana: Baseline Nonusers

				Proportion		Adjusted	
Sample	Follow-Up	Measure	Variable[b]	DARE	Control	Odds Ratio	*p* Value
Total	7th grade	Past year use	Any use	0.040	0.066	0.58	.025
Total	7th grade	Past year use	Monthly use	0.005	0.024	0.21	.003
Total	7th grade	Past month use	Any use	0.013	0.036	0.35	.005
Total	7th grade	Past month use	Weekly use	0.003	0.021	0.14	.002
Total	10 years	Past month use	Any use	0.247	0.307	0.74	.044
Males	7th grade	Lifetime use	Any use	0.070	0.124	0.53	.017
Males	7th grade	Past year use	Any use	0.042	0.095	0.41	.005
Males	7th grade	Past year use	Monthly use	0.008	0.042	0.19	.005
Males	7th grade	Past month use	Any use	0.017	0.066	0.24	.001
Males	7th grade	Past month use	Weekly Use	0.004	0.036	0.11	.004
Females	10 years	Lifetime use	Any use	0.567	0.664	0.66	.030
Females	10 years	Past month use	Any use	0.196	0.285	0.61	.019

a. Using one-tailed tests of statistical significance.
b. The lifetime use, past year, and past month *any use* variables were each created by dichotomizing the 7-point scale at the 2nd point on the scale (i.e., those reporting use "1–2 times" or more were classified as "users" and those reporting "none" as "nonusers"). The past year *monthly use* variable was created by dichotomizing the 7-point scale at the 5th point on the scale (i.e., those reporting use "10–19 times" or more were classified as "monthly users"). The past month *weekly use* variable was created by dichotomizing the 7-point scale at the 4th point on the scale (i.e., those reporting use "6–9 times" or more were classified as "weekly users").

Finally, there was also a significant difference at the 10-year follow-up between the two groups on the lifetime *any use* variable.

Discussion

This study involved a reanalysis of data from an evaluation of the DARE program conducted by Clayton and colleagues (Clayton et al. 1991; Clayton, Catterello, and Johnson 1996; Lynam et al. 1999) that is widely acknowledged to have shown the program to be ineffective in reducing substance use (Sherman et al. 1998; US General Accounting Office 2003; Weiss et al. 2008). We were, however, able to produce a number of

statistically significant differences at the 7th grade follow-up between the DARE condition and the control condition on measures of alcohol use and marijuana use using some very basic "data dredging" techniques. Moreover, many of these differences were found at cutoff points on the assessment instruments for which a meaningful label, such as *monthly use* or *weekly use*, could be fit in a post hoc manner. These differences could be framed and presented, if one so wished, to make the case that the DARE program is effective in reducing alcohol use and marijuana use and is especially effective with male nonusers of these drugs. Much could be made of the percentage reductions in alcohol and marijuana use found at the 1-year follow-up, a number of which were in the order of 70–80%. In addition, the failure to find many significant effects for females could be easily explained away (e.g., boys are more likely than girls to relate to the police officers who taught the DARE program), as could the failure to find any effects for cigarette smoking (e.g., underage drinking and marijuana use are law enforcement issues, whereas cigarette smoking is more of a health issue). The statistically significant effects we found for alcohol and marijuana use at 7th grade would also enable the DARE program to be considered an "evidence-based" prevention practice on lists such as those of the US Department of Education Safe, Disciplined, and Drug-Free Schools Expert Panel (2002) and Department of Health and Human Services (2007) because these require just one statistically significant difference between an intervention and control group and do not prohibit the type of data analysis practices that we used.

Lest it be thought that we are exaggerating the potential for isolated program effects to elevate an intervention to evidence-based status, it is instructive to compare our results to those of an evaluation conducted by Eisen, Zellman, and Murray (2003) of the Lions-Quest curriculum that is described on the NREPP list of evidence-based programs (National Registry of Evidence-based Programs and Practices 2006). In this study of 7,426 6th graders, data were collected pertaining to nine separate measures of drug use (assessed using 5- and 7-point scales that were dichotomized in the data analysis). The results showed just two statistically significant differences when data from the total sample were analyzed. These were for marijuana use, specifically the variables *recent marijuana use* and *lifetime marijuana use*. There was also one statistically significant difference when data from baseline binge drinkers were analyzed (for *recent binge drinking*). On the basis of these results, it was concluded that the "curriculum delivered in the seventh grade can help reduce the prevalence of lifetime and monthly marijuana use through the end of the eighth grade" and that the "program was especially

helpful in reducing the prevalence of binge drinking among students who had initiated regular binge drinking by the end of the sixth grade" (Eisen, Zellman, and Murray 2003, 895). The fact that no effects were found for the other six measures of substance use (including cigarette use) was not mentioned at all in the discussion section of the article. Moreover, in naming Lions-Quest an evidence-based program, the NREPP Web site focuses exclusively on these three statistically significant effects; as with the original research report, the fact that the majority of analyses conducted showed no statistically differences between the groups is entirely ignored (National Registry of Evidence-based Programs and Practices 2007)

Nor is this an isolated case. As noted above, the most detailed analysis of the interface between evaluation research and the development of best practice lists is the study conducted by Gandhi et al. (2007) that reviewed the evaluation research pertaining to five widely advocated "evidence-based" programs. In the case of the four programs supported by very scant evidence, they found that there were just one or two evaluation studies available for each and that these reported, at best, just a few positive effects on substance use outcome variables. The original Project ALERT evaluation, for example, produced no effects on cigarette, alcohol, or marijuana use at the 6-, 12-, and 18-month follow-ups when data from the total sample were analyzed and only isolated statistically significant effects when multiple subgroup analyses (based on participants' baseline levels of drug use) were performed (Ellickson and Bell 1990; Weiss et al. 2008). And later follow-ups showed no effect at all on drug use (Ellickson, Bell, and McGuigan 1993). Yet, Project ALERT appeared on four of the seven best practice lists reviewed by Gandhi et al. (2007), most notably the original NREPP list (Schinke, Brounstein, and Gardner 2002). Indeed, the program continues to appear on the revised NREPP evidence-based list of substance use prevention programs despite the fact that a subsequent evaluation by Ellickson and colleagues produced findings comparable to those of the original evaluation (i.e., some immediate effects at 8th grade that dissipated by 9th and 10th grade) (Ellickson et al. 2003; Longshore et al. 2007), and an independent evaluation found absolutely no effects on drug use (St. Pierre et al. 2005). Once again, examination of the NREPP Web site shows that it is the handful of statistically significant effects from each of the individual evaluation studies that are reported (National Registry of Evidence-based Programs and Practices 2006). A focus on these isolated positive findings, along with a total disregard of the plethora of null results, serves to create the impression that ALERT is an effective "evidence-based" prevention program.

Sociologists have argued that the definition of a behavior or an activity as a "social problem" is determined to some extent on the "claims-making" activities of those who find the behavior or activity unacceptable (Spector and Kitsuse 1987). Thus, to varying degrees all social problems are socially constructed, and the content and character of these social constructs can change over time. The analysis presented herein, along with the critical literature that is emerging in the area of evidence-based drug prevention (e.g., Brown 2001; Gorman 2005b; Gandhi et al. 2007; Midford 2008), suggests that social solutions, such as social problems, are also socially constructed, with program evaluators and the developers of evidence-based lists performing the role of claims makers. Evaluators have considerable leeway in the type and number of analyses that they conduct, which of these they choose to report out, and how they choose to interpret the results they obtain. A research culture based on hypothesis verification, rather than falsification, permits evaluators to conduct multiple analyses of their data sets, to selectively report out their findings, and to emphasize positive results over null or negative results. Consequently, more and more preventive interventions are socially constructed as "effective." Our reanalysis indicates that with a suitably sympathetic set of analyses, the DARE program might also be considered an "effective" prevention practice.

For their part, the evidence-based list developers are quite happy with the isolated statistically significant effects that program evaluators present them with and pay no attention as to how these were arrived at. This in turn means that more and more interventions will be socially constructed as "evidence-based." Here too, our reanalysis showed that the DARE program could quite easily meet the criteria of many of the evidence-based lists that have been developed in the United States over the past decade, most notably that of the NREPP. Of course, consideration of the preponderance of evidence from the University of Kentucky DARE evaluation (especially Clayton and colleagues' original analysis) shows that the program is not an effective drug prevention program. However, our reanalysis suggests that what differentiates DARE from many of the programs on evidence-based lists might not be the actual intervention but rather the manner in which data analysis is conducted, reported, and interpreted.

Notes

1. There is some debate in the literature as to whether it is appropriate to use one-tailed or two-tailed tests in evaluations of drug prevention programs. Our use of one-tailed tests, and the implication that this is inappropriate, is based on two issues. First, the rationale of our study is

that some data analysis strategies are chosen by evaluation researchers in the drug prevention field to maximize the chances of showing a positive effect of the program. Insofar as a one-tailed *p* value of .05 is equivalent to a two-tailed *p* value of .10, then the use of one-tailed tests clearly serves this purpose. Second, standard clinical trials textbooks advise against the use of one-tailed tests if there is even the possibility that the test statistic could provide evidence of iatrogenic effects (e.g., Piantadosi 1997; Friedman, Furberg, and DeMets 1998; Moyé 2000). In the case of the evaluation of drug prevention programs, there are examples in the literature of school-based programs (including some of those that appear on prominent evidence-based practice lists) that have been found to increase drug use at follow-up (e.g., Botvin et al. 1990; Ellickson and Bell 1990; St. Pierre et al. 2005; Hallfors et al. 2006).

References

Botvin, G. J., E. Baker, L. Dusenbury, E. M. Botvin, and T. Diaz. 1995. Long-term follow-up of a randomized drug abuse prevention trial in a white middle-class population. *Journal of the American Medical Association* 273:1106-12.

Botvin, G. J., E. Baker, A. D. Filazzola, and E. M. Botvin. 1990. A cognitive-behavioral approach to substance abuse prevention: One-year follow-up. *Addictive Behaviors* 15:47-63.

Botvin, G. J., K. W. Griffin, T. Diaz, N. Miller, and M. Ifill-Williams. 1999. Smoking initiation and escalation in early adolescent girls: One-year follow-up of a school-based prevention intervention for minority youth. *Journal of the American Medical Women's Association* 54:139-43.

Brown, J. H. 2001. Youth, drugs and resilience education. *Journal of Drug Education* 31: 83-122.

Clayton, R. R., A. Catterello, L. E. Day, and K. P. Walden. 1991. Persuasive communication and drug prevention: An evaluation of the DARE program. In *Persuasive communication and drug abuse prevention*, ed., L. Donohew, H. E. Sypher, and W. J. Bukoski, 295-313. Hillsdale, NJ: Lawrence Erlbaum.

Clayton, R. R., A. Catterello, and B. M. Johnstone. 1996. The effectiveness of Drug Abuse Resistance Education (Project DARE): 5-year follow-up results. *Preventive Medicine* 25:307-18.

Department of Health and Human Services. 2007. Notice regarding Substance Abuse and Mental Health Services Administration's National Registry of Evidence-Based Programs and Practices (NREPP): Priorities for NREPP reviews. *Federal Register* 72:30814-6. http://www.nrepp.samhsa.gov/pdfs/FRN060407.pdf (accessed August 8, 2008).

Eisen, M., G. L. Zellman, and D. M. Murray. 2003. Evaluating the Lions-Quest "Skills for Adolescence" drug education program: Second-year behavior outcomes. *Addictive Behaviors* 28:883-97.

Ellickson, P. L., and R. M. Bell. 1990. Drug prevention in junior high: A multi-site longitudinal test. *Science* 247:1299-305.

Ellickson, P. L., R. M. Bell, and K. McGuigan. 1993. Preventing adolescent drug use: Long-term results of a junior high program. *American Journal of Public Health* 83:856-61.

Ellickson, P. L., D. F. McCaffrey, B. Ghosh-Dastidar, and D. L. Longshore. 2003. New inroads in preventing adolescent drug use: Results from a large-scale trial of Project ALERT in middle schools. *American Journal of Public Health* 93:1830-6.

Friedman, L. M., C. D. Furberg, and D. L. DeMets. 1998. *Fundamentals of clinical trials*. 3rd ed. New York: Springer.

Gandhi, A. G., E. Murphy-Graham, A. Petrosino, S. S. Chrismer, and C. H. Weiss. 2007. The devil is in the details: Examining the evidence for "proven" school-based drug abuse prevention program. *Evaluation Review* 31:43-74.

Goldberg, L., D. P. MacKinnon, D. L. Elliot, E. L. Moe, G. Clarke, and J. -W. Cheong. 2000. The Adolescent Training and Learning to Avoid Steroids Program: Preventing drug use and promoting health behaviors. *Archives of Pediatric and Adolescent Medicine* 154:332-8.

Gorman, D. M. 2002. Defining and operationalizing "research-based" prevention: A critique (with case studies) of the US Department of Education's Safe, Disciplined and Drug-Free Schools Exemplary Programs. *Evaluation & Program Planning* 25:295-302.

———. 2003. The best of practices, the worst of practices: The making of science-based primary prevention programs. *Psychiatric Services* 54:1087-9.

———. 2005a. Drug and violence prevention: Rediscovering the critical rational dimension of evaluation research. *Journal of Experimental Criminology* 1:1-23.

———. 2005b. Does measurement dependence explain the effects of the Life Skills Training program on smoking outcomes? *Preventive Medicine* 40:479-87.

Gorman, D. M., E. Conde, and J. C. Huber Jr. 2007. The creation of evidence in "evidenced-based" drug prevention: A critique of the Strengthening Families Program Plus Life Skills Training evaluation. *Drug and Alcohol Review* 26:585-93.

Griffin, K. W., G. J. Botvin, T. R. Nichols, and M. M. Doyle. 2003. Effectiveness of a universal drug abuse prevention approach for youth at high risk for substance use initiation. *Preventive Medicine* 36:1-7.

Hallfors, D., H. Cho, V. Sanchez, S. Khatapoush, H. M. Kim, and D. Bauer. 2006. Efficacy vs effectiveness trial results of an indicated "model" substance abuse program: Implications for public health. *American Journal of Public Health* 96:2254-9.

Hallfors, D. D., M. Pankratz, and S. Hartman. 2007. Does federal policy support the use of scientific evidence in school-based prevention programs? *Prevention Science* 8:75-81.

Longshore, D., D. F. McCaffrey, P. A. St. Clair, and P. L. Ellickson. 2007. School-based drug prevention among high-risk adolescents: Effects of Project ALERT. *Health Education and Behavior* 34:351-68.

Lynam, D. R., R. Milich, R. Zimmerman, S. P. Novak, T. K. Logan, C. Martin, C. Leukefeld, and R. Clayton. 1999. Project DARE: No effects at 10-year follow-up. *Journal of Consulting and Clinical Psychology* 67:590-3.

Midford, R. 2008. Is this the path to effective prevention? *Addiction* 103:1169-70.

Mihalic, S., K. Irwin, D. Elliott, A. Fagan, and D. Hansen. 2001. *Blueprints for violence prevention*. Washington, DC: US Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention.

Moyé, L. A. 2000. *Statistical reasoning in medicine: The intuitive p-value primer*. New York: Springer.

National Institute on Drug Abuse. 2003. *Preventing drug use among children and adolescents: A research-based guide for parents, educators, and community leaders.* 2nd ed. Bethesda, MD: US Department of Health and Human Services.

National Registry of Evidence-based Programs and Practices. 2006. *Project ALERT*. http://www.nrepp.samhsa.gov/programfulldetails.asp?PROGRAM_ID=109#studies (accessed December 10, 2008).

National Registry of Evidence-based Programs and Practices. 2007. *Lions Quest skills for adolescence*. http://www.nrepp.samhsa.gov/programfulldetails.asp?PROGRAM_ID=99#studies (accessed December 10, 2008).

National Registry of Evidence-based Programs and Practices. 2008. *SAMHSA's National registry of evidence-based programs and practices*. http://www.nrepp.samhsa.gov/index.asp (accessed December 10, 2008).

Perry, C. L., C. L. Williams, S. Veblen-Mortenson, T. L. Toomey, K. A. Komro, P. S. Anstine, P. G. McGovern, J. R. Finnegan, J. L. Forster, A. C. Wagenaar, and M. Wolfson. 1996. Project Northland: Outcomes of a communitywide alcohol use prevention program during early adolescence. *American Journal of Public Health* 86:956-65.

Petrosino, A. 2003. Standards for evidence and evidence for standards: The case of school-based drug prevention. *Annals of the American Academy of Political and Social Science* 587:180-207.

Piantadosi, S. 1997. *Clinical trials: A methodologic perspective.* New York: Wiley.

Safe, Disciplined, and Drug-Free Schools Expert Panel. 1999. *Guidelines for submitting safe, disciplined, and drug-free schools programs for designation as promising or exemplary*. Washington, DC: US Department of Education.

Schinke, S., P. Brounstein, and S. Gardner. 2002. *Science-based Prevention programs and principles, 2002*. Rockville, MD: Center for Substance Abuse Prevention, Substance Abuse and Mental Health Services Administration.

Sherman, L. W., D. C. Gottfredson, D. L. MacKenzie, J. Eck, P. Reuter, and S. D. Bushway. 1998. *Preventing crime: What works, what doesn't, what's promising* (National Institute of Justice Research in Brief). Washington, DC: US Department of Justice.

Spector, M., and J. Kitsuse. 1987. *Constructing social problems*. New York: Aldine de Gruyter.

Spoth, R. L., S. Clair, C. Shin, and C. Redmond. 2006. Long-term effects of universal preventive interventions on methamphetamine use among adolescents. *Archives of Pediatric and Adolescence Medicine* 160:876-82.

Spoth, R. L., K. Randall, C. Shin, and C. Redmond. 2005. Randomized study of combined universal family and school preventive interventions: Patterns of long-term effects on initiation, regular use, and weekly drunkenness. *Psychology of Addictive Behaviors* 19:372-81.

St. Pierre, T. L., D. W. Osgood, C. C. Mincemoyer, D. L. Kaltreider, and T. J. Kauh. 2005. Results of an independent evaluation of Project ALERT delivered in schools by cooperative extension. *Prevention Science* 6:305-17.

Sun, W., S. Skara, P. Sun, C. W. Dent, and S. Sussman. 2006. Project towards no drug abuse: Long-term substance use outcomes evaluation. *Preventive Medicine* 42:188-92.

Sussman, S., P. Sun, W. J. McCuller, and C. W. Dent. 2003. Project towards no drug abuse: Two-year outcomes of a trial that compares health educator delivery to self-instruction. *Preventive Medicine* 37:155-62.

US Department of Education Safe, Disciplined, and Drug-Free Schools Expert Panel. 2002. *Exemplary and promising safe, disciplined, and drug-free schools programs 2001*. Jessup, MD: US Department of Education.

US General Accounting Office. 2003. *Youth illicit drug use prevention: DARE long-term evaluations and federal efforts to identify effective programs*. GAO-03-172R. Washington, DC: US General Accounting Office. http://www.gao.gov/new.items/d03172r.pdf (accessed November 25, 2008).

Weiss, C., E. Murphy-Graham, A. Petrosino, and A. G. Gandhi. 2008. The fairy godmother and her warts: Making the dream of evidence-based policy come true. *American Journal of Evaluation* 29:29-47.

Dennis M. Gorman, PhD, is a professor of social epidemiology in the Department of Epidemiology & Biostatistics at the Texas A&M Health Science Center School of Rural Public Health. His research interests include ecological studies of the relationship between alcohol availability and violence, evaluation of public policies to reduce alcohol- and drug-related problems, and the role of science in the evaluation of primary prevention programs.

J. Charles Huber, Jr., PhD, is an assistant professor of biostatistics at the Texas A&M Health Science Centers School of Rural Public Health. His research interests include statistical consulting, statistical genetics, longitudinal and multilevel modeling, categorical data analysis, and simulation methods.