The “science” of drug and alcohol prevention: the case of the randomized trial of the Life Skills Training program

Dennis M. Gorman *

Department of Epidemiology and Biostatistics, School of Rural Public Health, The Texas A&M University System Health Science Center, 3000 Briarcrest Drive, Suite 310, Wells Fargo Building, Bryan, TX 77802, USA

Abstract

Changes requiring greater accountability among federal agencies in the United States, along with specific criticisms of prevention activities funded by agencies such as the Department of Education, have led to an increased emphasis on what are called “science-based” or “research-based” interventions in recent years. Federal agencies such as the Department of Education and National Institute on Drug Abuse (NIDA) have produced documents describing such interventions and advocating their widespread use and dissemination. The most widely advocated of these prevention interventions is the Life Skills Training (LST) program, the effectiveness of which, its supporters argue, has been demonstrated using rigorous research methods. The research study that has attracted most attention is the randomized trial conducted with white middle-class adolescents in New York State, as this purports to demonstrate that the LST program can reduce alcohol and illicit drug use 6 years after initial implementation. In contrast to the advocates for the LST program, I argue that this longitudinal trial does not meet the rigorous methodological standards claimed on its behalf. Indeed, it violates one of the fundamental principles of a randomized trial by restricting key analyses to selective subsamples of the experimental group. I estimate that about 7.5% of those who initially received the LST intervention in the trial were included in the most recent set of analyses reported. This falls considerably short of the proportion of intervention group participants required at follow-up in a methodologically sound controlled trial. Thus, contrary to what its advocates claim, the study tells us little about the long-term effectiveness of the LST program in reducing alcohol and illicit drug use among adolescents.

© 2002 Elsevier Science B.V. All rights reserved.

Keywords: School-based drug prevention; Evaluation; Drug prevention; Life Skills Training

Science and drug prevention in the United States

Since the escalation in spending on federal drug control policies and programs in the mid-1980s in the United States, prevention has been a mainstay of the country’s efforts to reduce illicit drug use among adolescents. The federal drug prevention budget increased almost three-fold in real terms between 1986 and 1988, and almost tripled again over the next 3 years (Gorman, 1997). Much of this increase was accounted for by the growth in the Department of Education’s budget. This increased from $4.7 to 663.7 million (in constant dollars) between 1981 and 1992 (Gorman, 1998, table 2). Following some decline in the mid-1990s, the Department’s budget was close to $700 million dollars for Fiscal Year 2000 (Office of National Drug Control Policy, 2000, page 44).

The decline in the Department of Education’s drug prevention budget that occurred in the mid-1990s was in part a result of criticism directed at the appropriateness and efficacy of the activities that were funded under its Safe and Drug-Free Schools and Communities Program (US General Accounting Office, 1997). Moreover, this criticism occurred within the broader context of demands for smaller, more effective and less costly government that found expression in the Government Performance and Results Act of 1993. By the close of the 1990s, federal agencies, including those responsible
for overseeing the country’s drug control policies, were required to focus their accountability procedures on performance measures of effectiveness (PME) rather than simply providing descriptive accounts of how their programs functioned (US General Accounting Office, 1996).

The PME System developed by the Office of National Drug Control Policy (ONDCP) first appeared in 1998 and was elaborated upon and slightly amended the following year (Office of National Drug Control Policy, 1998, 1999). The PME System comprises five goals, the first of which is to: “Educate and enable America’s youth to reject illegal drugs as well as alcohol and tobacco” (Office of National Drug Control Policy, 1999, page 55). Each of the five ONDCP goals is broken down into a set of measurable objectives. Three of the nine objectives included under Goal 1 refer to the use of research and science to guide and inform prevention efforts (Objectives 4, 8 and 9; Office of National Drug Control Policy, 1999, pages 86, 95 and 97). For example, Goal 1, Objective 4 is to “Provide students in grades K-12 with alcohol, tobacco, and drug prevention programs and policies that are research based” (Office of National Drug Control Policy, 1999, page 86).

The Department of Education and Department of Health and Human Services are designated as the “reporting agencies” for objectives 4, 8 and 9 of Goal 1 of ONDCP’s PME System. As a result, these agencies and their subsidiaries have produced a plethora of documents describing “model” or “exemplary” programs over the past few years. These are programs that are recommended for widespread use and dissemination on the basis of scientific research apparently demonstrating their efficacy. For example, the Department of Education recently released its list of exemplary prevention programs (Safe, Disciplined, and Drug-Free Schools Expert Panel, 2001). The Guidelines describing the criteria for selection as an exemplary program state that it “must have at least one evaluation that has demonstrated an effect on substance use, violent behavior, or other conduct problems 1 year or longer beyond baseline” (Safe, Disciplined, and Drug-Free Schools Expert Panel, 1999, page 5). Similarly, in its 1997 Research-Based Guide, the National Institute on Drug Abuse (NIDA) describes ten programs that were “developed as part of a research protocol” and that have been “studied scientifically” (National Institute on Drug Abuse, 1997, page 19).

The Life Skills Training program—the exemplary ‘science-based’ program

The most prominent drug prevention program in terms of its consistency in being named on these lists of research-based interventions is the Life Skills Training (LST) program developed by Gilbert Botvin. In addition to being named on the Department of Education and NIDA lists of science-based programs, the LST website informs visitors that the program has also received awards and recognition from the American Medical Association, the American Psychological Association, the Center for Substance Abuse Prevention, the National Centers for Disease Control, the New York State Governor, and the private drug prevention agency Drug Strategies, Inc. (Life Skills Training, 2001a). It has also been featured in newspapers and magazines such as the New York Times, the Miami Herald, and Time (Life Skills Training, 2001b; Van Biema, 1996). And, in a somewhat ironic twist, the LST program is also the prevention intervention of choice of Philip Morris USA (America’s leading cigarette manufacturer), which also proclaims its commitment to research-based programs (Philip Morris USA, 2001).

Many prevention researchers in the United States have also advocated the use of the LST program on the basis of the evaluation research that has been conducted by Botvin and colleagues (e.g. Dusenbury, Falco & Lake, 1997; Rohrbach, D’Onofrio, Backer & Montgomery, 1996). In a recent review article, Botvin himself cites the program as an example of the science-based approach to prevention. The hallmark of this approach he argues “is careful testing using rigorous research methods” (Botvin, 2000, page 888). Thus, if there is an exemplary “science-based” alcohol and drug prevention program it is, by all accounts, the LST program.

The study most often cited in these documents heralding the efficacy of the LST program is the randomized trial conducted with white middle-class adolescents in New York State. Results from this trial have been published in the Journal of Consulting and Clinical Psychology (Botvin, Baker, Dusenbury, Tortu & Botvin, 1990), the Journal of the American Medical Association (JAMA; Botvin, Baker, Dusenbury, Botvin & Diaz, 1995), and most recently Addictive Behaviors (Botvin et al., 2000). The latter two publications are especially noteworthy as they claim to have demonstrated that the LST program can reduce alcohol, marijuana, and other illicit drug use 6 years after the intervention was administered.

---

1 The wording of the objectives changed slightly from the 1998 to 1999 ONDCP reports, with science and research becoming more prominent in the latter. For example, Objective 4 referred to “programs and policies that have been evaluated and tested and are based on sound practices and procedures” in the 1998 document, while the 1999 report referred to “programs and policies that are research based”. Similarly, Objective 8 in the 1998 report referred only to “a set of principles upon which prevention programming can be based”, while the same objective in the 1999 document specified “research-based principles” (Office of National Drug Control Policy, 1998, pages 60 and 65, and Office of National Drug Control Policy, 1999, pages 86 and 95).
The quality of science in the Life Skills Training evaluation

In contrast to Botvin and the various federal agencies and other prevention researchers that advocate the use of the LST, I argue that the 6-year randomized control trial of the LST does not meet the rigorous methodological standards claimed. Specifically, it violates one of the fundamental principles of a randomized trial by restricting key analyses to selective sub-samples of the experimental group. Consequently, the findings presented in the recent Addictive Behaviors paper and in the earlier paper published in JAMA do not support the claims made on behalf of the LST program concerning its general effectiveness in reducing alcohol, marijuana and other illicit drug use among adolescents. (Here I do not discuss the effects of the LST program on cigarette use.)

The nature of the claims concerning alcohol and illicit drug use made on behalf of the LST program, on the basis of the data from the JAMA study, are evident in Botvin’s recent review article. Here he states that: “At the follow-up conducted at the end of the 12th grade, the prevalence of cigarette smoking, alcohol use, and marijuana use for students in the prevention condition schools was as much as 44% lower than for controls. The strongest effects were found in schools where the program was delivered with the highest integrity” (Botvin, 2000, page 893, emphasis added). What this statement obscures, however, is that data collected from all of those subjects that were followed up at 6 years (called the “full sample” by the authors of the paper), show that there were statistically significant differences between the LST and control groups on only one of the six measures pertaining to marijuana and alcohol use. The prevalence rates for having “three or more drinks per occasion”, for example, were 0.57 and 0.55 for the two LST groups (one of whom comprised teachers trained in person, and one teachers trained by videotape) and 0.59 for the control group. Those for weekly marijuana use were 0.06 for the two LST groups and 0.09 for the controls. The only statistically significant difference reported with the full sample in relation to either alcohol or marijuana use was from a measure of “drunkenness”. Program effects on measures of alcohol consumption and use of marijuana emerged only when a new set of analyses was conducted with a sub-sample of LST participants whose teachers had successfully delivered 60% of the program. Thus, the key to finding positive results for use of any drug other than tobacco lies in restricting data analysis to this “high fidelity” sub-group. Unfortunately, this selective data analysis fundamentally undermines the randomization process considered so integral to the methodological rigor of the study.

The purpose of using randomization in an evaluation trial is to control for extraneous factors that can influence the intended outcome, and hence obscure or be confounded with intervention effects. One of the most important extraneous factors that can influence the outcome of a program evaluation is the process of uncontrolled selection that leads some individuals to participate in an intervention and others to opt out or be excluded (Rossi, Freeman & Lipsey, 1999; Treasury Board of Canada, 1998). It is not difficult to think of numerous extraneous factors (e.g. parental attitudes and opinions, neighborhood context, presence of deviant peer subgroups) that might affect a teacher’s motivation, willingness and/or ability to deliver the LST program, and that might also affect students’ use of alcohol and drugs. Randomization deals with such uncontrolled selection by using chance procedures to allocate units to experimental and control conditions—those who are willing and able to take part in the intervention are just as likely to end up in the control group as they are the intervention group. Other than chance variation, the intervention is the only difference between the experimental and control groups following randomization (at least this is the intention). However, it is important to note that randomization is only intended to bring about comparability of the entire randomized groups (in this case two LST groups and one control group). If it is successful in inducing such group comparability this does not ensure comparability of any subsets of these groups (Gail, 1985).

In the recent Addictive Behaviors paper Botvin and colleagues state that one of the main strengths of their longitudinal evaluation of the LST program is its use of random assignment (Botvin et al., 2000). Unfortunately, having addressed the possibility of bias from uncontrolled selection at the outset of the trial through random allocation of schools to study conditions, it is introduced into the study at the point at which the analysis shifts to the high fidelity sub-sample. Differential dropout from the intervention group is one of the primary threats to the internal validity of randomized trials, as it can introduce the type of bias that randomization was intended to deal with (Treasury Board of Canada, 1998). Once the size of the experimental group is substantially diminished, either through the actions of the participants or the research team, one can no longer be certain that it is comparable to the control group. Moreover, those intervention group subjects who remain “in” at follow-up are likely to differ from those who are “out” at follow-up.

What proportion of intervention group subjects need to remain in a trial at follow-up in order for the study to remain methodologically sound? In a 1993 report, the National Research Council Committee on Drug Abuse Prevention Research concluded that an evaluation with a gross follow-up rate under 75% was of “dubious
validity to assess effects even on relatively common behaviors” (Gerstein & Green, 1993, page 7). In discussing this issue in relation to the evaluation of substance abuse treatment programs, McLellan et al. (1997), drawing on recommendations of the Food and Drug Administration (FDA), propose that 70% of subjects need to be followed up in order to ensure that the data obtained are representative of the entire treated sample. Anything lower, they suggest, “should be regarded critically” (McLellan et al., 1997, page 15).

The LST trial discussed in this paper has a long follow-up period of more than 6 years, and so in judging its methodological rigor it might be unreasonable to expect follow-up rates quite as high as 70–75%. However, data from previous longitudinal follow-up evaluations of school-based substance abuse prevention programs suggest that it does provide a rough benchmark against which such a judgement can be made. For example, Clayton, Cattarello and Johnstone (1996) followed up 55% of participants in a DARE evaluation over a 5-year period, while a recent school-based smoking prevention trial that followed subjects over more than 10 years reported attrition rates of just 7% (see Petersen, Kealy, Mann, Marek & Sarason, 2000).

Botvin and colleagues were, in fact, fairly successful in following-LST program participants and control group subjects of the original 5954 who took part in the trial at baseline, 60.4% were re-assessed at 6 years (this is the “full-sample” in the JAMA paper). However, this success was undermined by the deliberate exclusion of intervention group subjects that occurred when the analysis shifted to the high fidelity sub-sample. As the left-hand side of Fig. 1 shows, about one third of the subjects in the intervention group in the LST trial who were followed up at 6 years were jettisoned from the study when the analysis shifted to the high fidelity subgroup. No similar refinement of the control group was undertaken (see Gorman, 1998 for more details). Thus, the LST high fidelity group represents about 40% of the original intervention sample—far lower than the 70–75% standard suggested above.

The JAMA paper states that those excluded from the high fidelity sample and those included were “virtually identical” in terms of demographic characteristics such as race and parental education. Since, however (as the title of the JAMA paper makes evident), there is very little demographic variability in the sample, it is unlikely that the two groups would differ in terms of demographic characteristics such as race or parental education. Moreover, no data are presented which allow one to rule out the possibility that the high fidelity sub-sample is self-selected, either as a result of individual-level characteristics (e.g. teacher motivation) or characteristics of some larger group of which the individuals are a part (e.g. the administrative and organizational capacity of the school that the children attend). As such, the differences in alcohol and marijuana use that were observed at follow-up with the high fidelity sub-sample might simply be the result of such extraneous confounding factors, and nothing to do with the LST program.

The results presented in the Addictive Behaviors paper, and the claims made on behalf of these, raise

---

* Estimated, as the exact size of the LST group at baseline is not reported in Botvin et al. (1990). At the 6-year follow-up, 68% of the 3,597 subjects for whom data were available were in the LST group. Since there was no report of differential attrition, this figure was used to estimate the size of the baseline intervention group—i.e., 68% of the baseline sample (n=5,554) = 4,049.

**Fig. 1.** Estimated Number and Percent of Baseline Experimental Group Subjects used in the analyses of the 6.5-Year follow-up data from the randomized trial of the LST program.
even more questions than those of the JAMA paper. With regard to the claims made, the results are said to demonstrate that the LST program “can produce prevention effects that last beyond the end of high school” and to “provide additional support for the long-term effectiveness of [this] broad-spectrum, cognitive-behavioral, universal prevention approach” (Botvin et al., 2000, page 773). These claims are tempered only by two sentences in the penultimate paragraph that point to the “small sample size” and its possible impact on the generalizability of the study findings.

Contrary to Botvin and colleagues, I would argue that the small sample size, rather than simply cautioning one in interpreting the findings presented in the Addictive Behaviors paper, should lead to their rejection. As the right-hand side of Fig. 1 shows, the LST sub-group included in the Addictive Behaviors analysis was comprised of just 302 individuals. This is about 7.5% of those who initially took part in the intervention at baseline. Even if one considers only those subjects available to follow-up at 6 years (i.e. only those included in the JAMA “full sample”), just 12% make it into the sub-group included in the Addictive Behaviors analysis. These follow-up figures are obviously far lower than those suggested by the National Research Council Committee (Gerstein & Green, 1993), McLellan et al. (1997) and the FDA for obtaining representative data from a program evaluation.

As to how this sub-sample of LST participants was selected, the Addictive Behaviors paper states that it comprised students from the larger study that were asked to complete additional survey items pertaining to illicit drug use, and that these data were collected 6 years and 6 months after the baseline assessment. However, no information is given as to how this small sub-group was constituted. Was it created through random sampling of the baseline sample reported in the Journal of Clinical and Consulting Psychology or a random sampling of the JAMA full sample? Alternatively, was it a further refinement of the JAMA high fidelity sub-group? What was the response rate on the mail survey used to collect data from this group? Did the research team select these individuals (and if so, on what basis), or did they select themselves?

In order to make claims concerning the generalizability of evaluation findings one must be able to demonstrate that the sample of target units employed in the analysis is an unbiased sample of the broader target audience (Rossi et al., 1999). (Recall that, as noted above, randomization of units to study conditions does not ensure comparability of any subsets of these conditions.) From the data presented in the Addictive Behaviors paper, there is no way of knowing whether the results can even be generalized to the more than two thousand other LST participants who were followed-up but omitted from the analysis, let alone beyond this to any broader target population of youth. Indeed, in terms of one of the few demographic variables that Botvin et al. (2000) do report, the Addictive Behaviors sub-sample clearly differs from the larger baseline sample from which it was drawn-just 40% of the former were male compared with 52% of the latter.

Conclusions

To conclude, the evidence from the 6-year randomized trial of the LST program suggesting that it is effective in reducing alcohol, marijuana and other illicit drug use only emerges in analyses that are conducted with small sub-samples of program participants. Refining the data analysis samples in this manner violates one of the fundamental principles of a randomized trial, and therefore cannot be considered an application of “rigorous research methods”. Moreover, I would argue that requiring that the judgment of the effectiveness of the LST program, or indeed any intention program, be based on analyses of data from all of those who were reassessed at follow-up is by no means an unreasonably stringent methodological requirement².

References


² For example, an “intent-to-treat analysis” would require that all those who dropped out of the intervention remain part of the trial throughout its course (including data analysis). This is a far more rigorous standard to apply to a randomized trial than the one that I have suggested here. It should be noted that there are compelling methodological reasons for conducting an intent-to-treat analysis (see Begg, 2000; Gail, 1985). Moreover, such an analysis is feasible in the field of substance abuse prevention (see Petersen et al., 2000).


