Addiction

FOR DEBATE

doi:10.1111/j.1360-0443.2009.02752.x

# Prevention programs in the 21st century: what we do not discuss in public

**Harold Holder**

Prevention Research Center, Pacific Institute for Research and Evaluation, El Cerrito, CA, USA

## ABSTRACT

Prevention research concerning alcohol, tobacco and other drugs faces a number of challenges as the scientific foundation is strengthened for the future. Seven issues which the prevention research field should address are discussed: lack of transparency in analyses of prevention program outcomes, lack of disclosure of copyright and potential for profit/income during publication, *post-hoc* outcome variable selection and reporting only outcomes which show positive and statistical significance at any follow-up point, tendency to evaluate statistical significance only rather than practical significance as well, problem of selection bias in terms of selecting subjects and limited generalizability, the need for confirmation of outcomes in which only self-report data are used and selection of appropriate statistical distributions in conducting significance testing. In order to establish a solid scientific base for alcohol, tobacco and drug prevention, this paper calls for discussions, disclosures and debates about the above issues (and others) as essential. In summary, the best approach is always transparency.

**Keywords** Prevention research, selection bias, self-report, statistical analyses.

*Correspondence to:* Harold Holder, Prevention Research Center, Pacific Institute for Research and Evaluation, 1609 Roger Court, El Cerrito, CA 94530, USA. E-mail: holder@prev.org

Submitted 12 December 2008; initial review completed 23 February 2009; final version accepted 24 July 2009

Prevention research concerning alcohol, tobacco and other drugs faces a number of challenges as the scientific foundation is strengthened for the future. In many ways, there have been two basic traditions for prevention: (i) programs developed typically by either researchers or program practitioners; and (ii) public policy, which is designed usually to reduce problems and implemented by government. Of course, there are hybrids, but in general this bifurcation serves us. For public policy, research typically undertakes evaluations of the effects of policies in reducing substance abuse problems at the population level, often utilizes existing archival data of substance-involved problems, e.g. drink driving crashes, and has to address critical design and analysis challenges [1]; see Babor *et al.* [2] for a summary of alcohol policy research outcomes.

Prevention program research as part of addiction studies is less mature and has not engaged actively the challenges faced by clinical research, with which it shares many features. One of the reasons for advances in clinical research is the frequent replication of behavioral treatment and pharmacotherapy modalities by a variety of independent researchers where no researcher has a potential for personal gain. In addition, in the wake of large-scale clinical trials involving several independent researchers, there has been agreement on the most appropriate set of outcome variables for evaluating treatment outcomes as well as intermediate causal variables. Such collaborative trials have not occurred in prevention program research. Researcher-designed prevention programs have a relatively brief history, and many issues are often not discussed publically and noted in scientific papers submitted for peer review. Here I identify seven of these issues in order to stimulate more public discourse.

First, the greatest number of researcher-designed prevention programs, especially in the United States, concern school, youth and family education. Usually these are developed and tested by researchers or prevention advocates. Many researcher-designed programs utilize funds provided by the national government. These research funds are highly competitive and research proposals are subjected to intense independent scientific review. With few exceptions, funded educational programs are subsequently designed, implemented and

evaluated by the principal investigator and his/her staff. No other independent investigators are involved and, to my knowledge, there are few published cases of successful replications of these programs by independent researchers not involved initially in the original program design and testing. On the surface, this is not necessarily a problem if there is full transparency and other researchers are invited to re-analyze data sets on which the published scientific papers are based. In fact, the Society for Prevention Research (SPR) [3] calls for appropriate transparency in its Conflict of Interest and Disclosure Statement, namely:

> After research results are published, SPR members should not unreasonably withhold the data on which their conclusions are based from other competent professionals who seek to verify their substantive claims through reanalysis, and who intend to use such data only for that purpose.

This is rarely conducted in practice, as data sets are often viewed as proprietary. Unfortunately, SPR even allows for this exception by noting that '. . . study data should be shared as long as the confidentiality of the participants can be adequately protected . . . and unless legal rights concerning proprietary data preclude their release' (Society for Prevention Research) [3]. There are very few re-analyses of results from researcher-designed programs by independent experts.

Secondly, some educational programs have been copyrighted by the researchers who designed them and subsequently made available on a fee basis to schools and communities. Publication and distribution rights are also sometimes sold to for-profit firms from which the original researchers may receive additional funds based upon actual sales or consultant agreements [4]. Again, selling scientific work for future profit is not in and of itself a problem or bad science. However, as the designers are the ones who created, implemented and evaluated the programs, the field of substance abuse prevention faced with potential conflicts of interest has recently required more public disclosure when a researcher is also selling his/her educational program for profit. Unfortunately, this is not always the case in publishing a paper in a scientific journal. For example, even *Addiction*'s submission system requires only a disclosure of sources of research funding and 'connections of any of the authors with the tobacco, alcohol, pharmaceutical and gaming industries (or bodies such as social aspect organisations that receive funding from them) irrespective of whether it relates to the current research', not whether the researcher can obtain any income from his/her prevention program if sold to others.

Thirdly, a troublesome issue is *post-hoc* outcome variable selection and thus reporting only outcomes which show positive and statistically significant results at any follow-up point. There have been questions raised about published studies which appear to show only positive results and which may not report negative or non-significant results, i.e. not reporting *all* results for *all* dependent variables. For discussion, for example, see Gorman *et al.* [5]. This issue has also been addressed by SPR in its Standards of Evidence by requiring that 'Results must be reported for every measured outcome, regardless of whether they are positive, non-significant or negative' and 'efficacy can be claimed only for constructs with a consistent pattern of statistically significant positive effects. That is, when multiple indicators are used, most or all must be in the positive direction and at least one must be statistically significant' [3]. Thus, scientific papers should disclose researcher-purposeful selection of outcome variable results, especially if the same person designed, implemented and evaluated the program. Analyzing a large number of outcome variables increases the possibility of 'multiple comparison bias' or Type I error, i.e. finding positive results by chance. This problem is confounded further if only selected significant positive findings are reported. If a scientific paper submitted for journal review does not disclose specifically that there has been author selection of the outcome variables reported, then readers may assume understandably that all outcome variables are being presented.

Fourthly, an important consideration in evaluating prevention program outcomes is statistical versus practical significance. Traditionally, statistical significance has been set at $P = 0.05$ or lower. The higher the value set for statistical significance the wider the confidence interval and the greater the possibility of 'significance'. However, what is to be made of practical significance between a variable which is statistically significant at 0.08 (thus only 3 percentage points higher than 0.05 for significance)? On the other hand, a statistically significant finding may actually represent little practical significance; see Gandhi *et al.* [6]. For example, let us say that a prevention program finds a reduction in self-reported drug use to be 4 percentage points lower comparing pre- to post-test results, which is statistically significant at 0.05, due possibly to the large number of subjects involved. Can we consider a 4-percentage-point reduction at a single point in time to be of practical significance? More important (and significant from a public health perspective) is whether there is convincing collaborative evidence of effect across several independent but linked variables; for example, is there consistent evidence of effect across most patterns of use? On the other hand, if a variety of outcome measures are collected and some are significant at one time-period and others non-significant and vice versa at another time-point, then I would argue that there is limited confirmatory

evidence of effect. An example of inconsistency might occur if, at the conclusion of a prevention program, drug use had a statistically non-significant decline 'over the life-time' but the 30-day decline was statistically significant, and later, at a 6-month follow-up, life-time use declined statistically significantly and the past 30-day use decline was non-statistically significant. In this hypothetical example, I would be unimpressed whether the paper used a 0.05 or a 0.08 cut-off value (or even 0.0001).

Fifthly, as with clinical studies, some prevention programs are subjected to selection bias, i.e. where the 'pool' of subjects includes those willing to participate even if available subjects are assigned randomly to experimental and control conditions. In clinical research, there is recognition that clients enrolled in treatment are not necessarily representative of all dependent individuals in the community. This most certainly occurs in prevention research, when participants who are most highly motivated are typically those who agreed to participate originally and thus remain in the program to completion. Outcomes for those who refused enrollment or who drop out without completion, i.e. non-completers, can include most often those with non-positive outcomes. Sometimes dropouts are excluded from the outcome analyses (whether or not reported). Even if reported, we often do not know if poor outcomes are the result of low personal motivation reflecting poor candidacy for improvement using this prevention program or whether the lack of exposure to the full program is the causal factor. Selection bias is especially problematic where candidates volunteer to participate, for example, in a parental educational program where volunteer parents are assigned randomly to condition, and neither control nor the experimental groups are necessarily representative of all parents in a community. Any effects resulting from such a parental educational program can only be generalized to the parents involved, not to *all* parents in a community (as some claim or infer).

Sixthly, also as with clinical trials, most outcome variables in prevention programs are based upon self-report. Unfortunately, in controlled conditions where self-reports of substance use are compared to biological measures, there is substantial reporting error (see [7–9]). Of course, if self-report bias is consistent or random, then confidence in the validity of observed differences (experimental to control in a prevention program) is increased. However, prevention program researchers have rarely undertaken tests of self-report measurement validity or reliability in order to actually document the potential direction of bias. In clinical trials there is a considerable effort to obtain verification of self-reported drinking via independently interviewing spouses or significant others as well as the collection of blood markers. For me, potential error may arise from normative effects on self-reported substance use by children and youth participating in some type of school or family educational program in which abstinence is clearly the desired adult norm. In my estimation, this potential normative error has not been ruled out adequately via measurement validation, i.e. children and youth may recognize quickly that abstinence is 'desired' by important adults, especially by the youngest ages in which the greatest effects have been reported in the scientific journals.

Seventhly, an issue often raised regarding prevention program research concerns statistical analyses; for example, the use of single- versus two-tailed *t*-tests to determine statistical significance. Two-tailed *t*-tests are used typically by researchers as the more conservative approach to testing statistical significance, i.e. the confidence interval surrounding the point estimates of potential effects are much narrower in a two-tailed test and the chance of error reduced. Prevention program researchers who utilize a single-tailed *t*-test should provide a strong rationale for this analysis decision, as a single-tailed *t*-test has more opportunity to find significant differences, control-to-experimental groups. In general, one chooses a one-tailed *P*-value when both the following are true: (a) one knows which group will be expected to have the larger mean before data collection; and thus (b) when the opposite group resulted with the larger mean then any difference would be considered to be by chance and 'not statistically significant'. In other words, in utilizing a one-tailed *t*-test, a prevention program evaluation infers a priori knowledge of whether the control or experimental group is expected to always have the higher mean value. Without a strong rationale about the distribution of the dependent variable, conventional conservative statistical analyses suggest using a two-tailed *t*-test, which makes finding statistical significance more difficult. Unfortunately, there are a number of examples where this convention is violated.

In summary, in order to establish a solid scientific base for alcohol, tobacco and drug prevention, discussions, disclosures and debates about the above issues (and others) are essential. The best approach is always transparency. I recognize that reporting program failure, i.e. lack of significant or consistent effects, has important implications for future research funding (or even personal income). Competition for research funds in all countries puts tremendous pressure on researchers to find positive findings, and in the case of researcher-designed prevention programs the pressure is intense to demonstrate a significant effect. Even if the researcher is not selling his/her program for profit, reporting positive findings increases the possibility of future research, so in many cases the stakes are rather high. However, reporting successes as well as a failure to yield signifi-

 

cant outcomes or effects is essential to establish a trustable and strongly scientific foundation for prevention research.

At least five conditions should be met in the future for scientific papers reporting the effects of a prevention program, as follows.

**1** Publication of non-significant results and providing an adequate justification for journals to accept such papers.

**2** Presentation of all outcome results using an appropriate theoretical rationale developed *before* the study began and avoid *post-hoc* selection of outcome results. If there is researcher selection of specific outcomes, this should be disclosed with an appropriate rationale for so doing.

**3** Disclosure of any potential for personal profit or income from the prevention program for which positive outcomes are being reported by the program designers themselves.

**4** Public discussion of methodological and statistical effects which match or do not match the basic standards of science and whether the results can be of practical significance, i.e. avoid use of statistical significance as an end in itself and thus encourage readers to determine for themselves how practically significant are final results.

**5** Recognition in scientific papers about prevention programs of potential selection biases (how the original pool of candidates were drawn and inherent limitations on generalizability from these subjects) and self-report bias (especially if there is no independent verification of self-report results). If the researchers have evidence to rule out either bias, then this should be also reported.

Prevention research will be the better for it.

**Declaration of interest**

None.

**References**

1. Biglan A., Avery D., Wagenaar A. C. The value of interrupted time–series experiments for community intervention research. *Prev Sci* 2000; **1**: 31–49.
2. Babor T., Caetano R., Casswell S., Edwards G., Giesbrecht N., Graham K. *et al. Alcohol: No Ordinary Commodity: Research and Public Policy*. New York: Oxford University Press; 2003 (revised edition, in press, expected 2010).
3. Society for Prevention Research. *Standards of Evidence: Criteria for Efficacy, Effectiveness and Dissemination*. Falls Church, VA: Society for Prevention Research; 2004 available at: http://www.prevention research.org (accessed 9 April 2009).
4. Gorman D. M., Conde E. Conflict of interest in the evaluation and dissemination of 'model' school-based drug and violence prevention programs. *Eval Program Plann* 2007; **30**: 422–9.
5. Gorman D. M., Conde E., Huber J. C. Jr. The creation of evidence in 'evidence-based' drug prevention: a critique of the Strengthening Families Program Plus Life Skills Training evaluation. *Drug Alcohol Rev* 2007; **26**: 585–93.
6. Gandhi A. G., Murphy-Graham E., Petrosino A., Chrismer S. S., Weiss C. H. The devil is in the details: examining the evidence for 'proven' school-based drug abuse prevention. *Eval Rev* 2007; **31**: 43–74.
7. Johnson M., Voas R. A., Miller B. A., Holder H. D. Predicting drug use at electronic music dance events: self reports and biological measurement. *Eval Rev* 2009; **33**: 211–25.
8. Buchan B., Dennis M. L., Tims F. M., Diamond G. S. Cannabis use: consistency and validity of self-report, on-site urine testing and laboratory testing. *Addiction* 2002; **97**: S98–108.
9. Cook R. F., Bernstein A. D., Arlington, T. L., Andrews C. M., Marshall G. A. Methods for assessing drug use prevalence in the workplace: a comparison of self-report, urinalysis, and hair analysis. *Subst Use Misuse* 1995; **30**: 403–26.