

Commentary—Project Towards No Drug Abuse: An Evidence-Based Drug Abuse Prevention Program

Steve Sussman · Thomas W. Valente ·
Louise A. Rohrbach · Clyde W. Dent ·
Ping Sun

Published online: 1 May 2014
© Springer Science+Business Media New York 2014

Abstract Project Towards No Drug Abuse (Project TND) is a research-based program that has been examined over seven group-randomized controlled trials. In all trials, an effect has been found on hard drug use. An effect has been found on alcohol use in four trials, and on cigarettes and marijuana in two trials. (Arguably, an effect is found on marijuana in three trials.) Program effects on violence-related behavior were established in earlier trials, though such effects were not assessed in later trials. Certainly, as in most scientific studies, there are limitations in the interpretation of the effects obtained, particularly regarding cigarette and marijuana use and violence-related behavior, and more empirical work is needed. For Project TND, however, numerous trial replications

have been completed, and the effects within each trial could not have been obtained by chance alone more than 10 % of the time (two-tailed). These results suggest that Project TND is indeed evidence-based. To be evidence-based means that evidence has accumulated to suggest that the program is likely to work, at least under conditions that are comparable to those in which it has been tested.

Keywords Project TND · Analysis · Response

Project Towards No Drug Abuse (Project TND) is an evidence-based program. In its simplest meaning, to be an evidence-based program means to rely on data outside of one's subjective opinion that, for example, a program is "fun" or "effective," or that "it must have worked because everyone showed up." The quality and quantity of evidence is important. It is desirable to be able to provide some type of comparison to the program. This comparison group may be an untreated

Commentary on Is *Project Towards No Drug Abuse* (Project TND) an Evidence-Based Drug and Violence Prevention Program? A Review and Reappraisal of the Evaluation Studies.

This commentary refers to the article available at doi:[10.1007/s10935-014-0348-1](https://doi.org/10.1007/s10935-014-0348-1).

S. Sussman (✉) · T. W. Valente · L. A. Rohrbach ·
P. Sun

Institute for Health Promotion and Disease Prevention
Research, Keck School of Medicine, University of
Southern California, Soto Street Building, 2001 North
Soto Street, Room 302A, Los Angeles, CA 90033, USA
e-mail: ssusma@usc.edu

T. W. Valente
e-mail: tvalente@usc.edu

L. A. Rohrbach
e-mail: rohrbac@usc.edu

P. Sun
e-mail: sping@usc.edu

C. W. Dent
Oregon Health Authority, Salem, OR, USA
e-mail: Clyde.DENT@dhsosha.state.or.us

population similar to the treated population that is ascertained from survey data, or from preselected comparison groups. More confounders are controlled for when a trial involves multiple preselected units that are randomly assigned to conditions. Project TND has had seven clustered-randomized controlled trials (individuals nested within schools or classes that were randomly assigned to conditions), which tested different variations of the program. As we have reported in our 2012 paper (Sussman, Sun, Rohrbach, & Spruijt-Metz, 2012), a recent book chapter (Sussman, 2014), and on our web site (<http://tnd.usc.edu>), the project showed effects on hard drug use in all trials, on alcohol use in four of the trials (particularly at higher levels of use), and on cigarette smoking and marijuana use in two of the trials. (Arguably, an effect is found on marijuana in three trials.) We did not understand the conditions that influenced variability in effects on cigarette smoking and marijuana use. The alcohol results were promising, but here too we lacked a complete understanding of why we achieved effects in some trials but not others. The effects on these three drugs are based on accepted standards of evidence (Flay et al., 2005), but the results are inconsistent across the trials. We achieved effects on violence in the earlier studies, but then did not measure this construct in some of the later studies to manage subject recruitment. Human Subjects concerns had increased over time regarding the assessment of violence. These concerns negated subject selection without written parental consent if we had continued to request collection of these items, which would have dramatically reduced subject recruitment. Hard drug use effects were obtained across all trials and this appears to be the major strength of the program. We doubt any drug abuse prevention program has been submitted to so many trials and replicated a program effect on a drug outcome in all of them. This appears to be a consistent effect.

It is of great scientific and practical importance to independently examine the published findings for evidence-based interventions. Gorman (2014) intends to provide a critical evaluation of the results and assessment of the methodology and data analysis techniques employed in the Project TND evaluations. After examining the details, Gorman lists four major methodological issues that he contends undermine the validity of this body of work and the extent to which the later studies on TND can be seen as replicating the

findings of the earlier studies. These issues include: (1) the use of one-tailed tests of statistical significance, (2) a lack of consistency across the studies in how hard drug use was measured, (3) issues pertaining to the manner in which the data were entered into the analysis in those studies that used the frequency measure of hard drug use, and (4) concerns about subject recruitment and very high attrition that occurred, particularly in the first four studies. We provide comments that address these four issues.

The Use of One-Tailed Tests of Statistical Significance

There is still lack of agreement about whether one- or two-tailed tests should be employed in evaluating outcomes of public health intervention programs. Based on a review of 85 published evaluations of school-based drug use prevention curricula, Ringwalt and others found that 20 % of studies reported one-tailed tests (Ringwalt, Paschall, Gorman, Derzon, & Kinlaw, 2011). The theoretical assumptions behind the use of one- versus two-tailed tests are exclusively different. According to some statistical theorists, using a one-tailed test to examine positive program effects dismisses the examination of potential harmful program effects that might be seen on the other tail. By looking at both tails in a one-tailed test, some theorists argue, one is actually relaxing the standard for the p value required for acceptance of an effect.

However, other statisticians have argued that directional hypotheses require directional tests (e.g., Neyman & Pearson, 1967), to avoid misallocation of the error probabilities between the null and alternative hypotheses. Also, the possibility for a program to achieve positive effects may not be the same as the possibility to achieve null or negative effects. Thus, to consider these uneven possibilities and maintain an overall type-1 error of 5 %, some researchers have proposed ‘more scientific’ ways of doing significance tests (e.g., uneven tails, or two tailed but relaxing the p value for possible significance to .10; the latter option is suggested by Ringwalt et al. 2011). The ‘more scientific’ ways of doing significance tests may make even more sense when a series of trials have been conducted, and in which no iatrogenic effect has been detected.

In our research on Project TND, we took the position that a directional theoretical hypothesis

(participants exposed to the program would exhibit less drug use than those not exposed versus the null position of no beneficial effect) permitted us to engage in a one-tailed test. We did, however, as is ethically required, check the other tail for potential iatrogenic effects in all trials. There were none, other than the findings obtained by Valente et al. (2007), which were published. We generally provided (or have gladly provided on request) enough data in our presentations for one to derive exact p values for both two-tailed and one-tailed tests, and effect sizes, so that the readers could decide on the acceptability of the evidence provided. Still, we acknowledge that perhaps we should have explicitly reported exact p values for both two-tailed and one-tailed tests, and effect sizes, so that the reader could have obtained the most complete description of our results, as perhaps so should have Gorman (2014). (We report some results in these papers which Gorman (2014) does not mention; see “Notes” section.)

Lack of Consistency on How Hard Drug Use Outcome Was Measured

The TND program has been tested and refined with a series of trials that were administered during a span of more than 10 years. The samples were different from trial to trial, and some of the trials were evaluated by different research teams (i.e., TND research team membership has varied across trials). The positive and convergent findings based on related, but somewhat independent assessments and evaluations, could be recognized as a boost to the validity of the TND program. Variations did not constitute dramatically different measures, the stem questions asked were identical across studies, and rather minor changes were made to individual response options by collapsing rarely endorsed adjacent categories.

Manner in Which Data Were Entered in Analysis: Studies that Used Frequency Measures of Hard Drug Use

In the later trials, TND data analysts have started to employ more comprehensive statistical tests. In the latest trials, substance use indicators were evaluated as both binary and count data. The earlier trials were

not analyzed with the same types of models; analysts were limited by the tools commonly available when the data were analyzed. On the other hand, the replications of the earlier findings may indicate the robustness of the earlier, more traditional statistical models. It is unfortunate that the modern tools were not readily available in the past. However, we do not view the application of new statistical techniques in the most recent studies as a major methodological issue that undermines the validity of the findings from the earlier studies. Statistical and methodological tools continue to evolve, and we believe that the use of the best methods and statistics available at the time has been one of the strengths of the TND evaluations.

Under this criticism, Gorman also takes issue with the variety of transformations we used in the response scales. We note that all transformations were clearly specified in our Methods sections and applied uniformly across comparison groups. The transformations used were monotonic and the resulting transformed scales are highly correlated with the original; they were performed to aid in interpretation or improve the error distribution in the analysis models. The minor variation in the weight given to high frequency users, while theoretically possible, would not in practice result in different conclusions in the study outcomes, given the sample sizes used in these studies.

Recruitment Limitations and Very High Attrition

As indicated in the Methods sections of TND papers, the majority of failure to obtain subject participation (recruitment) was due to chronic absenteeism. That is, at continuation high schools, many more youths are on the enrollment rosters than ever attended even one class. This “selection bias” has been discussed by TND researchers in their outcome papers, where they state that study results only apply to youths that attend school. Youth *en masse* did not “refuse” participation in the TND program, as Gorman (2014) suggested.

We agree that it is important to control for attrition in drug use prevention program research. Attrition control has been especially important in research on TND, since the primary target group for the program is high-risk youth. (A propensity score for attrition was adjusted for in the analysis of the later trials to attempt

to grapple with attrition at follow-up.) In all of these trials, we tested for and failed to find differential attrition by condition on key variables such as baseline drug use. Thus, the high attrition is a matter of external validity, not internal validity.

Technically, the unit of assignment, treatment and analysis in our studies is the school (or classroom, in two of the trials). Students are representatives of these larger units. Under that view, we lost no units to follow-up—all schools (or classrooms) were represented in the follow-up assessments. The issue then becomes one whether the followed students adequately represented their schools (or classrooms).

Among those surveyed at baseline, attrition was as high as 30–40 %. Just because attrition was that high does not itself dictate that the results of the evaluations are invalid. The level of bias attrition may introduce to the validity of the study is related to whether the missing data is completely at random (MCAR), conditionally at random (MAR), or not at random (MNAR). Although the TND investigators demonstrated comparability in selected study variables between retained samples and whole samples (or the lost-to-follow-up samples), and comparability in the attrition rate across program conditions, it is still possible that the missing data are MAR or MNAR. All missing data situations are likely partly MAR and partly MNAR; thus, if properly adjusted, the attrition-induced bias is related to both the attrition rate and how highly correlated the cause of missingness is with the outcome. In simulations work, Collins, Schafer, and Karri (2001) have demonstrated that at 25 % attrition, there is little bias even if the correlation between the outcome and cause of missingness is .90. They further demonstrated that even at the high attrition rate of 50 %, the bias is still limited if the correlation between the outcome and cause of missingness is .40 (Collins, Schafer, & Karri, 2001; Graham, 2009). In our view, while attrition could bias the estimation, one should not automatically dismiss the validity of estimations from studies in which there is 30–40 % attrition. To support Gorman's claim that attrition invalidates the results of the TND trials, more details on a theory and mathematical proofs (such as how the attrition could reduce the reported program effect) are needed. We stress, again, that with comparable attrition across the study conditions, which we found in all of the trials, non-random attrition may limit external validity, but does not threaten internal validity.

Project TND has been tested extensively in continuation high schools. The cost of attempting to conduct interventions among continuation high school youth is that the research is challenging; samples often are hard to obtain and follow. The benefits, however, are substantial given that this population is at high risk for a lifetime of substance abuse, delinquency, no or under-employment, and other poor outcomes. Consequently, program effects have been obtained among a population that has a strong and demonstrated need for prevention programs.

Central Issues When Considering Evidence-Based Programs

We agree with Gorman (2014) that what constitutes an “evidence-based” program (EBP) remains a salient issue in the prevention field. Although the criteria for identifying evidence-based programs are derived from agreed upon scientific standards, such as the rigor of evidence (e.g., appropriateness of methodology), the methods used to collect and analyze the data, the magnitude and consistency of the effects, and the generalizability of the findings (e.g., Flay et al. 2005), certainly the application of these standards has varied across registries and lists. Importantly, to be evidence-based does not mean a program is “model” or “exemplary.” We are not sure that such a prototypical program exists in the drug abuse prevention or cessation fields. We agree that such labeling of programs might thwart critical research.

The determination that Project TND was “good” evidence-based programming was not made by the TND researchers themselves. Certainly, if even inconsistent (but replicated at least once) effects are found across different drugs using randomized controlled trial designs, this would seem a rational basis to consider a substance abuse prevention program evidence-based. The “bar” for gauging whether a program is “evidence-based” could, of course, be raised high enough such that no drug abuse prevention program would qualify. If the public mandates using some type of program with at-risk or older youth, a program such as Project TND would appear, based on available programs, a reasonable choice.

In summary, we agree with the general view that program development groups, including Project TND, should conduct additional critical evaluations of the

results and assessment of the methodology and data analysis techniques employed in their substance use prevention and control programs. We also agree that researchers should be able to report non-effects without perceived pressure to report positive effects. However, it appears that this particular critical evaluation is trying to support a predefined claim with vague or misleading evidence. We believe that the author did not offer enough details to support his bold conclusion that there is little in the seven evaluations to support the view that Project TND is an effective drug use or violence prevention program. On the other hand, we do agree that a much more comprehensive meta-analysis that summarizes across the outcomes papers or aggregates data from all trials would be useful to assess the program effects of TND.

Notes

We wish to point out a few additional results, not reported by Gorman (2014). In the Rohrbach et al. (2010) study, the program condition showed an effect on marijuana use relative to the control condition ($p < .1$, two tailed). We suggest that this effect be considered potentially meaningful. Then an effect on marijuana use would have been considered ‘found’ in three of the seven trials. In Sussman et al. (2012), one frequency measure of hard drug use revealed an effect as a function of any programming with a $p < .026$, one tailed, which Gorman indicated would not be significant if it was two tailed. However, as was reported in the paper, another measure of hard drug use frequency, a hard drug use index which was composed of the log of the average number of times of different types of hard drug use in the last 30 days, also demonstrated a program effect comparing any programming to control ($p < .023$, one tailed; Sussman et al., 2012). We would comfortably conclude that the TND program generated a positive effect on frequency of hard drug use (two-tailed p values were .052 on the first measure and .046 on the second measure). A composite substance use index (across all measured substances) was examined in two of the studies (Sussman, Sun, Rohrbach, & Spruijt-Metz, 2012; Valente et al. 2007). In these studies a program condition effect (any

programming and the TND networked condition, respectively) was achieved at $p < .01$ (one tailed and two tailed, respectively).

Acknowledgments This paper was supported by a grant from the National Institute on Drug Abuse (DA020138).

References

- Collins, L. M., Schafer, J. L., & Karri, C. M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, *6*, 330–351.
- Flay, B. R., Biglan, A., Boruch, R. F., Castro, F. G., Gottfredson, D., Kellam, S., et al. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science*, *6*, 151–175.
- Gorman, D. (2014). Is Project Towards No Drug Abuse (Project TND) an evidence-based drug and violence prevention program? A review and reappraisal of the evaluation studies. *The Journal of Primary Prevention*. doi:10.1007/s10935-014-0348-1.
- Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*, 549–576.
- Neyman, J., & Pearson, E. S. (1967) [1933]. *The testing of statistical hypotheses in relation to probabilities a priori*. Joint statistical papers. New York, N.Y.: Cambridge University Press (pp. 186–202).
- Ringwalt, C., Paschall, M. J., Gorman, D., Derzon, J., & Kinlaw, A. (2011). The use of one-versus two-tailed tests to evaluate prevention programs. *Evaluation & the Health Professions*, *34*, 135–150.
- Rohrbach, L. A., Sun, P., & Sussman, S. (2010). One-year follow-up evaluation of the Project Towards No Drug Abuse (TND) dissemination trial. *Preventive Medicine*, *51*, 313–319.
- Sussman, S. (2014). Why might Project TND be impacting on drug use? A speculation considered across seven research trials. In L. M. Scheier (Ed.), *Handbook of drug prevention*. Washington, DC: American Psychological Association (in press).
- Sussman, S., Sun, P., Rohrbach, L., & Spruijt-Metz, D. (2012). One-year outcomes of a drug abuse prevention program for older teens and emerging adults: Evaluating a motivational interviewing booster component. *Health Psychology*, *31*, 476–485.
- Valente, T. W., Ritt-Olson, A., Stacy, A., Unger, J. B., Okamoto, J., & Sussman, S. (2007). Peer acceleration: Effects of a social network tailored substance abuse prevention program among high risk adolescents. *Addiction*, *102*, 1804–1815.