

Invited Letter Rejoinder

Psychological Medicine, 48 (2018).

doi:10.1017/S0033291717002951

First published online 26 October 2017

Sources of bias in Gorman critique of bias: again a need for more reasonable, valid conclusions with true dialogue

In the first paragraph of his commentary ‘Sources of bias and need for caution in interpreting the results of Spoth *et al.*’s (2017) PROSPER Study,’ Dr. Gorman references our ‘claims of intervention effectiveness’ that he asserts are problematic for three reasons. Briefly stated, the reasons Gorman proffers are: (1) sample selection bias; (2) a low rate of participation in (strengthening families program: for parents and youth) SFP 10-14; and (3) flexible data analyses and selective reporting. Our response begins with his misrepresentation of the intervention effectiveness claims made in our 2017 article. We then proceed to address each of Gorman’s three reasons for attributions of bias in our study.

The most important issue raised by this critique concerns the need for constructive criticism and true dialogue. We believe that constructive criticism requires: (1) careful attention to methodological detail in both critiqued outcome reports and related publications in multi-study programs of research; (2) sound reasoning about purported biases with critical evaluation of related assumptions; (3) consideration of the requirements of large-scale longitudinal studies; and (4) thorough literature reviews with balanced conclusions. Notably, critics’ representation of the literature should reference prior responses to similar critiques of the research in question. For example, responses by the authors, co-investigators, and other prevention scientists to prior critiques by Gorman provide repeated rejoinders to criticisms similar to those in the critique herein (e.g. Ellickson & Bell, 1994; Ennett *et al.* 1995; Hawkins & Catalano, 2003; Botvin & Griffin, 2005; Hawkins & Catalano, 2005; Graham, 2008; Spoth *et al.* 2008, 2009; Sussman *et al.* 2014; Midford *et al.* 2015; Rulison *et al.* 2016). While we acknowledge the many benefits of constructive criticism from fellow researchers in improving the quality of our scientific work, the following illustrates that many of the attributes of constructive criticism are lacking in this critique. Most importantly, we believe true dialogue requires timely opportunities to respond to constructive criticism (see below).

In the present critique, several statements clearly misrepresent the primary research question addressed by the PROSPER trial – the question guiding its experimental design. The tested intervention is a multicomponent delivery system for evidence-based programs in school and community settings. Although analyses intending to disentangle intervention component effects can be performed, PROSPER’s experimental design, as such, most directly supports tests of the overall, multicomponent PROSPER intervention outcomes. The PROSPER intervention components include various community-building activities organized by a community team, plus implementation of school-based and family-focused programs selected from a menu. Thus, the statement in the critique’s opening paragraph, ‘...claims were also made concerning positive intervention effects *of the programs tested*’ (our emphasis) is misleading since the study is not designed to test the effects of individual programs. This misleading characterization of the PROSPER study is of relevance to Gorman’s stated concerns about low rates of participation in SFP 10-14 and the implications of stated rates in the young adult sample. Further, Gorman’s quote about our claims of demonstrated benefits of ‘developmentally well-timed interventions’ is followed by a description of the effects of PROSPER *as a whole*, including reference to how ‘PROSPER community teams delivered’ the developmentally well-timed interventions.

Briefly stated, Gorman mischaracterizes the PROSPER study as a test of SFP 10-14. In his participation rate and sample bias criticisms, he ignores the fundamental issue that the PROSPER study is an examination of a delivery system for a multicomponent sequence of interventions, focusing attention on participation rates in one intervention component (SFP-10-14) exclusively. This overlooks the fact that the vast majority of intervention condition youth also received one of three school-based interventions, in the context of building prevention knowledge, collaboration and social capital in participating communities. The extent to which SFP 10-14 participants are represented in the age 19 sample is relevant only as far as it reflects the representativeness of the age 19 sample vis-à-vis the baseline 6th grade sample. In fact, the percentage of young adults in the age 19 follow up assessment that had attended SFP 10-14 was 21.7%, close to the overall rate of 17% participation in the full sample. The subsample that was studied at age 19 was neither made up of ‘few if any’ who attended SFP 10-14, nor made up ‘exclusively’ by those who attended SFP 10-14, as Gorman indicates could be the case.

In this vein, Gorman fails to note the explicitly-stated rationale for the sample selection strategy in the 2017 article. Following the entire sample was neither financially viable nor necessary to address research aims. In addition, Gorman references 'differential refusal' concerns but also chooses not to acknowledge our examination of differential participation by condition at age 19, an examination that addressed key socio-demographic and all outcome variables in the article. The 2017 article references analyses showing no significant condition-specific participation by baseline measure effects, conducted to examine the possibility of differential attrition. Finally in this connection, Gorman states that the CONSORT diagram does not 'strictly adhere to CONSORT requirements,' focusing on missing information about those who 'received the intended treatment.' He misleads the reader by suggesting that the intended treatment was only SFP 10-14. As noted above, the intended treatment extended well beyond SFP 10-14 and the vast majority received that treatment. Moreover, the referenced type of CONSORT diagram has been reviewed and accepted by multiple journals.

Gorman's third reason for claiming bias concerns results that could be '...the product of outcome manipulation and analytic flexibility.' These claims are of the greatest concern to us, particularly his sweeping assertion, repeatedly stated without reference to published counterpoints, that 'such practices are common in substance use prevention research' and in particular reference to assessments of the SFP 10-14 program (see Spoth et al. 2009 and Rulison et al. 2016 for examples of earlier responses to his unfounded critiques).

First, although we agree with Gorman's cautions about the need for consistency in selection and analysis of measures when conducting follow-up assessments, his summary of outcomes reported in our 6.5 and 7.5 year follow ups in Table 1 is simply incorrect. For example, at the top of the table and labeled 'Only Reported in Spoth et al. (2017),' three of the first five variables are incorrectly represented. Specifically, lifetime/new use of marijuana, ecstasy, and methamphetamines were reported in the 6.5 year follow up article cited, placed in an online supplement (as clearly noted in the article text in the measures section), as a result of space constraints and journal guidance. The other two measures in the first section of his Table 1 were only reported at the 7.5 year time point because the base rates were too low at the high school stage to allow meaningful analysis.

Second, Gorman's critique reflects a failure to understand the need to test intervention outcomes that are most relevant to participants' developmental stage in a trial that cuts across three separate developmental

stages. Gorman fails to consider age-related patterns of substance use, including low base rates, ceiling effects on lifetime measures like alcohol use, as well as the irrelevance of other measures outside of a specific developmental stage. The issues with low base rates and related low cell frequencies (e.g. base rates of cocaine use were too low to meet analytic requirements during the high school stage) were highlighted in a response to an earlier Gorman critique (Spoth et al. 2009), as they had been by Ellickson & Bell (1994) previously.

The development phase-related measurement of prescription drug misuse provides another case in point concerning Gorman's charge of analytic flexibility. Young adulthood is a developmental stage during which rates of prescription drug misuse are at their highest levels (Center for Behavioral Health Statistics & Quality, 2016). At the data collection point for the 2017 paper (age 19, post high school), the expanded measurement package – since we were no longer limited to collecting data in one class period – allowed for a developmentally appropriate expansion of the measurement items concerning this type of misuse. This, in turn, allowed the analysis of a separate index focused on the increasingly important issue of prescription drug misuse, incorporating one of the lifetime items that previously was used in an earlier, broader index of illicit use employed at the 6.5 year high school time point (11th and 12th grades). There also was an expanded set of marijuana use items, allowing for reporting on three separate measures (lifetime, current, frequency) in the 2017 paper.

The purposefully expanded set of measures and reporting of outcomes relate to the following statement by Gorman. 'But perhaps the *best indicator* of the presence of analytic flexibility in the PROSPER study is the *Illicit Substance Use Index that appears in both* (our emphases) the 6.5 year and the 7.5 year follow ups.' The measure is not in both reports. In retrospect, we understand that similar measure names across the two studies ('*Lifetime* illicit substance use index' v. 'Illicit substance use index' in the 2013 and 2017 reports, respectively) could reasonably raise a question in a comparative analysis. It also is reasonable, however, to expect an inquiry, before a charge of scientific bias is made based on this 'best indicator.' Elimination of the 'Lifetime' label reflects developmental phase-related improvements in measurement afforded by an expanded set of measures. In addition, we could have readily addressed any concern that a significant finding on the more focused 'Illicit substance use index' was primarily due to measurement changes.

Third, more generally, Gorman states there is 'very little consistency' in reported variables between the 6.5 and the 7.5 year outcome studies, '...let alone any

consistent pattern of statistically significant results.' As suits his argument, Gorman chooses to focus on differences in the individual measures employed, rather than on the broader interpretation that both studies find significant effects across a range of substances; this type of selective focus has been noted previously (e.g. Hawkins & Catalano, 2005).

Fourth, Gorman notes the failure to pre-specify planned analyses. We agree that this is an increasingly common practice that benefits the science. In fact, our research group has done so in our PROSPER replication RCT recently underway. The criticism overlooks common standards when the original proposal was written almost two decades ago; also, the original 5-year grant did not carry expectations for follow-up assessments into young adulthood. Further, Gorman implicitly suggests support for his claims by citing null findings from two recent SFP 10-14 evaluations, again ignoring the fact that the PROSPER study is not an evaluation of SFP 10-14, as well as overlooking other details relevant to those studies. For example, he ignores differences in designs, samples, methods, country and cultural contexts, native languages, duration of follow-ups, and the number of years elapsed since the initial rural US-based program evaluations – all important considerations for a balanced discussion concerning the generalizability of SFP 10-14 findings and its current efficacy with rural US populations and elsewhere.

In conclusion, we are in wholehearted agreement with the need for constructive criticism and open dialogue. Productive scientific discourse requires, however, careful attention to the detail of a study's research questions, design, methods, and measures, along with the complexities of long-term longitudinal studies and multi-study programs of research, critical evaluation of assumptions, and a balanced review of relevant literature in its historical context. This response to Gorman documents inadequacies in his critique concerning each of these points. It adds to our own and our colleagues' prior responses to similar critiques by Gorman. This said, our greatest concern is Gorman's repeated calls for critical dialogue while demonstrating a willingness to publish critiques – and more worrisome, suggestions of scientific impropriety – that were not first brought to the attention of those critiqued or did not afford an opportunity for true dialogue. For example, on two separate occasions, we learned about full-article critiques with specific, wide-ranging claims of bias only after they were published. Further, we are troubled by Gorman's frequent statements of researcher bias such as 'outcome manipulation and analytic flexibility' supported by a selective use and omission of relevant information, misleading tabular summaries, and superficial or completely-missing consideration of alternative

explanations of divergent findings across time points or studies. We have also been disappointed that journals publish such critiques without a more thorough vetting of their scientific merit. We very much appreciate the opportunity provided by *Psychological Medicine* to respond in this particular case. We call for truly open exchanges about the significant challenges in optimizing preventive intervention science accompanied by well-reasoned strategies for addressing them.

Acknowledgements

Work on this paper was supported by research grant DA13709 from the National Institute on Drug Abuse and co-funding from the National Institute on Alcohol Abuse and Alcoholism.

Declaration of Interest

None.

Ethical Standards

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

References

- Botvin GJ, Griffin KW (2005). Prevention science, drug abuse prevention, and Life Skills Training: comments on the state of the science. *Journal of Experimental Criminology* **1**, 63–78 (Response to Gorman).
- Center for Behavioral Health Statistics and Quality (2016). *2015 National Survey on Drug Use and Health: Detailed Tables*. Substance Abuse and Mental Health Services Administration: Rockville, MD.
- Ellickson PL, Bell RM (1994). Ellickson and Bell respond. *American Journal of Public Health* **84**, 500–501.
- Ennett ST, Tobler NS, Ringwalt CL, Flewelling RL (1995). Ennett and colleagues respond. *American Journal of Public Health* **85**, 873–874.
- Graham K (2008) Balancing type I and type II error: response to Gorman 2008. *Addiction* **103**, 1753–1755.
- Hawkins JD, Catalano RF (2003). Assessing outcomes of prevention interventions. *Psychiatric Services* **54**, 1548.
- Hawkins JD, Catalano RF (2005). Doing prevention science: a response to Dennis M. Gorman and a brief history of the quasi-experimental study nested within the Seattle Social Development Project. *Journal of Experimental Criminology* **1**, 79–86.
- Midford R, Foxcroft DR, Cahill H, Ramsden R, Lester L (2015). Rejoinder to Dennis Gorman's critique of: 'Preventing alcohol harm: early results from a cluster randomized, controlled trial in Victoria, Australia of

- comprehensive harm minimisation school drug education'. *International Journal of Drug Policy* **26**, 721–722.
- Rulison K, Feinberg M, Gest SD, Osgood W** (2016). Response to Gorman *et al.*, 'Comment on Rulison *et al.* (2015) diffusion of intervention effects'. *Developmental Psychology* **58**, 693. [http://www.jahonline.org/article/S1054-139X\(16\)30005-2/pdf](http://www.jahonline.org/article/S1054-139X(16)30005-2/pdf).
- Spoth R, Trudeau L, Redmond C, Shin C** (2008). Finding a path to more reasonable conclusions about prevention – response to Midford (basis for Gorman commentary). *Addiction* **103**, 1169–1173. Letter to editor. <http://dx.doi.org/10.1111/j.1360-0443.2008.02270.x>.
- Spoth R, Trudeau L, Redmond C, Shin C** (2009). Further clear examples of the need for more reasonable conclusions and critiques about prevention. *Addiction* **104**, 154–155. (Response to Gorman.) <http://dx.doi.org/10.1111/j.1360-0443.2008.02460.x>.
- Spoth R, Redmond C, Shin C, Greenberg M, Feinberg M, Trudeau L** (2017). PROSPER delivery of universal preventive interventions with young adolescents: Long-term effects on emerging adult substance misuse and associated risk behaviours. *Psychological Medicine* **47**, 2246–2259.
- Sussman S, Valente TW, Rohrbach LA, Dent CW, Sun P** (2014). Commentary – project towards no drug abuse: an evidence-based drug abuse prevention program. *Journal of Primary Prevention* **35**, 233–237.

R. SPOTH^{1,*}, C. REDMOND¹, C. SHIN¹,
M. GREENBERG², M. FEINBERG² AND L. TRUDEAU¹
¹Partnerships in Prevention Science Institute, Iowa State University, Ames, Iowa, USA
²Prevention Research Center, The Pennsylvania State University, University Park, Pennsylvania, USA

*Address for correspondence: R. Spoth, Ph.D., Partnerships in Prevention Science Institute, Iowa State University, 2625 N. Loop Drive, Suite 2400, Ames, IA 50010, USA.
(Email: rlspoth@iastate.edu)