Correspondence

Sources of bias and need for caution in interpreting the results of Spoth *et al.*'s (2017) PROSPER study

A recent paper published in Psychological Medicine reported results from a 7.5-year follow-up of subjects at age 19 years who took part in a community-based intervention known as PROSPER (Spoth *et al.* 2017). The study examined the effects on substance misuse and associated risk behaviors of participation in a family-based program delivered in 6th grade (Strengthening Families Program 10–14; SFP 10–14) combined with one of three 7th grade school-based programs (All Stars, Life Skills Training or Project ALERT). Although much of the discussion in the paper focused on PROSPER as a delivery system, claims were also made concerning the 'positive intervention effects' of the programs tested (Spoth *et al.* 2017, p. 8). The findings of the study were said to 'indicate the potential benefits of developmentally well-timed interventions in early adolescence for substance-related outcomes during emerging adulthood' (Spoth *et al.* 2017, p. 10). Such claims of intervention effectiveness are problematic for at least three reasons: the potential for selection bias in the sample used in the study; the low rate of participation in the SFP 10–14; and the possibility that the positive results reported were the result of flexible data analysis and selective reporting.

The data analysis used in the PROSPER study includes all subjects in the intervention condition who provided data at follow-up, irrespective of whether or not they participated in the SFP 10–14 in 6th grade or the school-based program in 7th grade. However, unlike earlier follow-up assessments, the 7.5-year follow-up did not attempt to contact the entire sample, but randomly selected from among subjects who completed the 6th grade baseline assessment and were still enrolled in the same school district in 9th grade. Spoth *et al.* (2017) present data from 1985 subjects selected from this group (1004 in the intervention group and 981 in the control group). They fail, however, to report the number of eligible subjects contacted in order to generate this sample. Since 21% of subjects did not complete the 9th grade follow-up assessment and 28% did not complete the 10th grade assesssment (Spoth *et al.* 2011, Figure 1), it is unlikely there were

no refusals when a random sample was drawn from these same subjects at the 7.5-year follow-up. In order to assess the potential for selection bias, one needs to know the number of subjects who refused to be re-assessed in each study condition. For the intervention condition, one also needs to know if there was differential refusal according to participation of subjects in the intervention programs.

It is unclear from Spoth *et al.* (2017) whether any of the individuals who remained in the intervention condition at age 19 actually received the family-based program in 6th grade. A total of 5515 intervention group subjects were pretested in 6th grade and eligible for participation in the SFP 10–14 component of the intervention (Spoth *et al.* 2017). Of these, just 1064 (19%) attended at least one of the seven SFP 10–14 sessions (Spoth *et al.* 2011). This very low participation creates the possibility that few if any of the 1004 intervention group subjects assessed at 7.5 years actually participated in the program. At one extreme, it is possible that the random sample of 1004 subjects assessed at age 19 was drawn exclusively from the 1064 subjects in the intervention group who took part in the SFP 10–14 in 6th grade. At the other extreme, it is possible these 1004 subjects all came from the 4451 who did not participate in the program. If the sample is truly random (and assuming no differential attrition related to SFP 10–14 participation), one would expect about 191 (19%) of the 1004 subjects assessed at 7.5 years to have participated in the program in 6th grade. One needs to know the exact number in order to assess the validity of Spoth et al's (2017) claim that the SFP 10–14 is an effective substance use prevention intervention.

It should be noted that while Spoth *et al.* (2017) include a CONSORT-like flow diagram of subjects (detailing those eligible to take part in the study at its inception, those randomized to study conditions and those successfully followed-up), this does not strictly adhere to the CONSORT requirements that guide reporting of trials. Specifically, the CONSORT diagram should include 'the numbers of participants who were randomly assigned, *received intended treatment*, and were analysed for the primary outcome' flowing through the trial at each assessment point (Schulz *et al.* 2010, emphasis added).

The final reason that caution should be used in claiming that the PROSPER interventions are effective is that the results reported could simply be the result of running multiple analyses and selective reporting. Spoth and colleagues claim that their results clearly show 'the pattern of statistically significant PROSPER

Table 1. *Statistical significance of outcomes reported at the 6.5-year and 7.5-year follow-ups in the PROSPER evaluation*[a]

Variable	6.5 years follow-up (Spoth *et al.* 2013)	7.5 years follow-up (Spoth *et al.* 2017)
Only reported in Spoth *et al.* (2017)		
Lifetime marijuana use		✓
Lifetime cocaine use		✓
Lifetime ecstasy use		✓
Lifetime methamphetamine use		✓
Lifetime LSD use		✓
Lifetime narcotics use (non-prescription)		✓
Lifetime amphetamine use (non-prescription)		✗
Lifetime illicit substance use index (0–5)[b]		✓
Lifetime prescription drug misuse index		✓
Past year narcotics		✗
Frequency cigarette use		✓
Frequency drinking		✗
Frequency narcotics (non-prescription)		✗
Drug-related problems		✓
Alcohol-related problems		✗
Only reported in Spoth *et al.* (2013)		
Illicit substance initiation index (0–5)[c]	✓	
Past year driving after drinking	✗	
Past year inhalants	*	
Past year methamphetamine	✓	
Reported in both Spoth *et al.* (2013, 2017)		
Past month cigarette use	*	✗
Past month drunkenness	✗	✗
Past year marijuana use	*	✗
Frequency drunkenness	✗	✗
Frequency driving after drinking	✗	✗
Frequency marijuana use	✓	✓

[a] ✓ = Statistically significant at $p < 0.05$ using a two-tailed test of statistical significance; ✗ = Not statistically significant at $p < 0.05$ using a two-tailed test of statistical significance; * = Statistically significant at $p < 0.05$ using a one-tailed test of statistical significance (Spoth *et al.* 2013 used 1-taled tests).

[b] Ever used: (a) methamphetamines, (b) ecstasy, (c) LSD (or other hallucinogens), (d) cocaine, and (e) GHB or Rohypnol.

[c] Ever used: (a) methamphetamines, (b) ecstasy, (c) marijuana, (d) drugs prescribed by a doctor for someone else, and (e) Vicodin, Percocet or Oxycontin not prescribed by a doctor.

intervention effects on primary substance use outcomes, previously found through the 12th grade, was still evident at age 19' (Spoth *et al.* 2017, p. 8). However, as shown in Table 1, there was very little consistency in the variables reported at the 6.5-year (12th grade) follow-up and the 7.5-year (aged 19) follow-up, let alone any consistent pattern of statistically significant results. A total of 21 comparisons between PROSPER and control participants on substance use outcomes are reported in Spoth *et al.* (2017), 11 of which showed statically significant differences between the study conditions using two-tailed tests of significance [note: the 6.5-year analysis reported in Spoth *et al.* (2013) used one-tailed tests of significance]. Eight of the 11 statistically significant differences pertain to lifetime use of illicit and non-prescribed substances. None of these lifetime measures

were included in the 6.5-year follow-up (Spoth *et al.* 2013). Only one of the three other variables (frequency of marijuana use) for which a statistically significant difference was found at the 7.5-year follow-up was included in the 6.5-year analysis. There are five other variables common to each analysis, and none of the differences in scores on these between the PROSPER and control groups were statistically significant using two-tailed tests. Thus the 'pattern' of statistically significant effects Spoth *et al.* (2017) claim exist is simply not present.

Table 1 also raises concerns that the results reported in the PROSPER study may be the product of outcome manipulation and analytic flexibility. Such practices are common in substance use prevention research (Holder, 2010; Gorman, 2015) and their use increases the likelihood that a study will produce false discoveries

(Ioannidis, 2005; Simmons *et al.* 2011). Their use is also evident in an earlier evaluation of the SFP 10–14 program (Gorman *et al.* 2007). In the PROSPER study, there is no obvious reason why lifetime use of substances, frequency of cigarette use and drug-related problems should be outcomes at the 7.5-year follow-up, but not the 6.5-year follow-up. Nor is there any obvious reason to include past year inhalant use and methamphetamine use as outcomes at the 6.5-year follow-up but not the 7.5-year follow-up. 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 the 6.5-year and 7.5-year follow-ups. Each index is comprised of five substances, but only two substances are common to both indexes (methamphetamines and ecstasy). Thus, what appears at first sight to be a consistent effect across the two follow-ups, is not, and no reason is given for the change in the composition of the index.

Spoth *et al.* (2017) advise readers of their paper to adjust the significance levels they present based on the number of statistical tests reported (although they refrain from conducting any such adjusted analysis themselves). The data presented in Table 1 suggests that the 21 substance use tests presented in the paper may not have been the only ones conducted in the analyses. A published document that prespecified the exact analyses to be conducted and the precise outcomes to be included in the 7.5-year follow-up would dispel any such concerns. In the absence of this, one has no way of knowing if the results presented by Spoth *et al.* (2017) are false discoveries that arise through analytic flexibility, outcome manipulation and multiple comparisons. It is worth noting that two recent evaluations of the SFP 10–14 that did prespecify their outcomes and analyses plans in a clinical trials registry and published protocol both produced null results (Baldus *et al.* 2016; Foxcroft *et al.* 2017).

Financial support

None.

Declaration of interest

None.

References

Baldus C, Thomsen M, Sack PM, Bröning M, Arnaud N, Daubmann A, Thomasius R (2016). Evaluation of a German version of the strengthening families programme 10–14: a randomized controlled trial. *European Journal of Public Health* **26**, 953–959.

Foxcroft DR, Callen H, Davies EL, Okulicz-Kozaryn K (2017). Effectiveness of the strengthening families programme 10–14 in Poland: cluster randomized controlled trial. *European Journal of Public Health* **27**, 494–500.

Gorman DM (2015). 'Everything works': the need to address confirmation bias in evaluations of drug misuse prevention interventions for adolescents. *Addiction* **110**, 1539–1540.

Gorman DM, Conde E, Huber JC 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**, 585–593.

Holder HD (2010). Prevention programs in the 21st century: what we do not discuss in public. *Addiction* **105**, 578–581.

Ioannidis JPA (2005). Why most published research findings are false. *PLoS Medicine* **2**, e124.

Schulz KF, Altman DG, Moher D, CONSORT Group (2010). CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMC Medicine* **18**, 18. doi: 10.1186/1741-7015-8-18.

Simmons JP, Nelson LD, Simonsohn U (2011). False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science* **22**, 1359–1366.

Spoth R, Redmond C, Clair S, Shin C, Greenberg M, Feinberg M (2011). Preventing substance misuse through community-university partnerships: randomized controlled trial outcomes $4^{1/2}$ years past baseline. *American Journal of Preventive Medicine* **40**, 440–447.

Spoth R, Redmond C, Shin C, Greenberg M, Feinberg M, Schainker L (2013). PROSPER community-university partnership system effects on substance misuse through $6^{1/2}$ years past baseline from a cluster randomized controlled intervention trial. *Preventive Medicine* **56**, 190–196.

Spoth R, Redmond C, Shin C, Greenberg MT, Feinberg ME, Trudeau L (2017). PROSPER delivery of universal preventive interventions with young adolescents: long-term effects on emerging adult substance misuse and associated risk behaviors. *Psychological Medicine*. doi: 10.1017/S0033291717000691.

D. M. GORMAN
Department of Epidemiology and Biostatistics, Texas A&M University, TAMU 1266, College Station, TX 77843-1266, USA

Address for correspondence: D. M. Gorman, Department of Epidemiology and Biostatistics, School of Public Health, Texas A&M University, TAMU 1266, College Station, TX 77843-1266, USA.
(Email: gorman@sph.tamhsc.edu)