

DRUG & ALCOHOL FINDINGS *Research analysis*

This entry is our analysis of a study considered particularly relevant to improving outcomes from drug or alcohol interventions in the UK. The original study was not published by Findings; click [Title](#) to order a copy. Free reprints may be available from the authors – click [prepared e-mail](#). [Links](#) to other documents. [Hover over](#) for notes. [Click to](#) highlight passage referred to. [Unfold extra text](#)  The Summary conveys the findings and views expressed in the study. Below is a commentary from Drug and Alcohol Findings.

Send email for updates

[SEND](#) [About updates](#)

► [Title and link for copying](#) ► [Comment/query to editor](#) ► [Tweet](#)

► [PROSPER community-university partnership delivery system effects on substance misuse through 6½ years past baseline from a cluster randomized controlled intervention trial.](#)



DOWNLOAD PDF
for saving to
your computer

Spoth R., Redmond C., Shin C. et al.
Preventive Medicine: 2013, 56, p. 190–196.

Unable to obtain a copy by clicking title? Try asking the author for a reprint by adapting this [prepared e-mail](#) or by writing to Dr Spoth at rspoth@iastate.edu. You could also try this [alternative](#) source.

Evaluated drug prevention programmes for adolescents are typically implemented by research teams, raising questions over real-world applicability and sustainability, but an important US trial is said to have robustly demonstrated the public health potential of a system in which the communities themselves take primary responsibility.

SUMMARY The PROSPER (*PRO*moting School-community-university Partnerships to Enhance Resilience) model for disseminating substance use prevention is not a set programme but a way of engaging and supporting communities to (hopefully sustainably) implement the programmes they choose. It uses the leverage afforded by the existing infrastructure in US states of [cooperative extension systems](#) [offering non-formal educational programmes to help people use research-based knowledge to improve their lives] run by [land grant universities](#) [founded to extend higher education to broad segments of the US population].

In the [PROSPER model](#) local teams of 8–12 community stakeholders (from schools and human service agencies plus parent and youth representatives) associated with schools in the area are co-led by local cooperative extension staff. These teams are linked to state-level university researchers and cooperative extension faculties via prevention coordinators connected with the land grant university. Coordinators also provide ongoing, proactive technical assistance to community teams to optimise team functioning and the delivery of prevention programmes. Though technical assistance and expert support is provided, the local teams [have primary responsibility](#) for implementing the programmes they choose, promising a more sustainable and lower cost alternative to projects led by outside experts or by staff dedicated to the project.

To join the study, communities in the US states of Iowa and Pennsylvania [had to](#) have a school district enrolment of 1300 to 5200, of whom at least 15% were eligible for free or reduced-cost school lunches, netting communities with an appreciable percentage of poorer families. Also considered essential were qualified extension and school system personnel to serve as team co-leaders. The recruitment process included visits to school district administrators to describe the project. All the districts [had just one](#) high school.

Of 68 otherwise eligible districts, 20 did not meet staffing requirements; just five refused to join the study. Willing and suitable communities joined until the required 28 were recruited, 14 in each state. They consisted of rural towns and small cities with populations of about 7000 to 46,000. After matching on location and school district size, districts were randomly allocated to carry on as usual (the [control](#) districts) or to implement the PROSPER model, 14 in each arm of the study. During the first year of the trial two of the PROSPER school districts [dropped out](#) and were replaced. Control districts [did not necessarily](#) neglect



Key points From summary and commentary

In small rural US communities, the PROSPER trial tested a method for implementing family- and school-based substance use prevention programmes for adolescents in which the communities themselves to take primary responsibility.

Researchers on this major trial saw its findings as indicative of robust preventive impacts, showing the PROSPER model had the potential for improving public health.

Main questions over the findings are whether they would be replicated outside the type of communities recruited to the study, and whether even in these communities they really were a “robust” demonstration of the effectiveness of the PROSPER model.

prevention; virtually all engaged in some prevention efforts, and six of the 14 implemented interventions of the type available through PROSPER.

Aided by the other tiers of the system, the 14 local PROSPER teams selected what central experts saw as evidence-based prevention programmes intended to be universally applicable rather than just for high-risk families. The selection was to include one of three family-focused programmes. Of these, all 14 teams chose the version of the [Strengthening Families Program](#) developed by a project headed by the lead researcher in the study, and delivered it to pupils in the sixth grade, typically aged 11–12. For the programme parents and children were invited to attend seven weekly two-hour sessions typically in the evening; 17% of eligible families attended at least one session; of these families, 90% attended at least four.

The following year school-based substance use education programmes were chosen and delivered to the now 7th-grade pupils. [Life Skills Training](#) and [Project Alert](#) were each selected by four teams; the [All Stars](#) curriculum was selected by the other six. Each was delivered during class periods, generally by one of the schools' own teachers, and reached around 9 in 10 of the intended pupils. This pattern of interventions was implemented for two successive school years beginning in 2002.

In 2002 and 2003 about 90% of the intended sixth-grade pupil sample provided some baseline data via confidential written questionnaires before the interventions started. Two-thirds lived with both biological parents, 85% were described as Caucasian, and about a third received free or subsidised school lunches. They were followed up by researchers annually until the 12th grade, which typically consists of 18-year-olds ending their school education. At each follow-up point, on average 86% of eligible pupils completed the questionnaires. By the final high-school follow-up the sample had reduced from the 10,849 who completed pre-intervention baseline questionnaires to about 7784, representing 72% of pupils who completed baseline measures and 65% of all the pupils in the relevant school years.

Main findings

Presented findings focused on the 11th grade and the final 12th grade follow-ups, between which a further tenth of the sample could no longer be re-assessed. A 'lifetime illicit substance use index' scored from 0 to 5 captured the range of substances ever used by the pupils. It asked if they had ever used methamphetamine, ecstasy, cannabis (by smoking), drugs or medications prescribed for someone else, or certain prescription-only painkillers when these had not been prescribed by a doctor.

By 12th grade this index averaged 1.43 in PROSPER school districts and 1.68 in control districts, a statistically significant difference indicating that on average slightly fewer different types of drugs had been tried by pupils in PROSPER districts. Of the individual drugs, in PROSPER communities methamphetamine, ecstasy, cannabis and inhalants ['glue-sniffing'] were all significantly less likely to have been tried among pupils who had not used them before the interventions ('uptake'). Uptake of drinking and of drinking to the point of drunkenness were virtually identical in PROSPER and non-PROSPER districts. Results were similar in the 11th grade.

In respect of current/recent use of individual substances, at 12th grade pupils in PROSPER districts were significantly less likely to have smoked cigarettes in the past month, or over the past year to have used cannabis, methamphetamine or inhalants. Except for no significant impact on inhalant use, results were similar in the 11th grade. Across the whole sample, the largest relative reduction in the proportion of pupils was in respect of past-year methamphetamine use at the 12th grade, a reduction from about 4% to about 3%, representing 31% fewer users. At neither grade were there any statistically significant differences relating to drinking, assessed in terms of past-month drunkenness and past-year drink-driving.

The study also assessed **how often** pupils said they had recently got drunk, drove after drinking, or smoked cannabis. At the 12th grade, only in respect of cannabis did pupils in PROSPER districts use significantly less often, a reduction from 2.05 to 1.83 in a frequency index scored from 0 to 7. Results were similar at the 11th grade, except for a lower frequency of having been drunk (index scores 2.46 v. 2.66 in non-PROSPER districts) and a larger – but still not statistically significant – difference in driving after drinking.

Growth in substance use across the six-and-a-half years of the follow-up tended to be less steep among pupils in PROSPER school districts. These results were statistically significant for growth in the current/recent use of all assessed substances and frequency of use measures, except in respect of inhalants and having been drunk (though growth in the *frequency* of drunkenness was significantly retarded).

Often the association of PROSPER with retarded development of substance use was stronger among pupils at higher risk of substance use at the start of the study, defined as those who (typically aged about 11–12) had already tried one or more of alcohol, cigarettes, or cannabis. Greater effects on high-risk pupils were most apparent in the 11th-grade analysis. By the 12th grade it was statistically significant for the lifetime illicit substance use index, and over the past year for cannabis use and

frequency of use; measures related to smoking, drinking or methamphetamine showed no significant effect. This pattern was the same for differences in the growth of substance use over the entire six-and-a-half years of the follow-up.

The authors' conclusions

Overall, the effects of interventions delivered via the PROSPER delivery system on long-term adolescent substance use outcomes were robust, both in terms of growth across the middle- and high-school years, and in results at the final two follow-ups. Positive long-term effects were observed for lifetime illicit use and for current use and frequencies of use, for all types of substances. Relative reduction rates (indicating how much smaller the proportion engaging in that behaviour was in PROSPER districts) ranged from 3.3% to 31.4%. Notably, intervention effects on growth in substance use from 6th to 12th grades were significant for all outcomes, except past-month drunkenness and past-year inhalant use, indicating that the differences between PROSPER and control districts increased over time.

Effects on outcomes related to alcohol were relatively weak compared to those for other substances, perhaps because of relatively high starting levels of use and its greater acceptability. Intervention effects among the higher-risk subsample were comparable to or, in most cases, stronger than those among the lower-risk subsample, countering the contention that only lower-risk populations benefit from universally applied interventions.

Other findings from the study suggested how the interventions chosen by PROSPER districts might have worked. In middle school these reduced the influence on other students of adolescents who used substances, by reducing the centrality of substance-using youth in their friendship networks. Also [there were](#) positive effects on key factors thought to underly substance misuse in later adolescence, including social skills, parenting, and family environment, and the children's attitudes and beliefs about substance use. Finally, early initiation of substance use [was reduced](#) by the interventions; early initiation is one of the most powerful predictors of later substance misuse, underscoring the critical importance of the developmental timing of interventions; PROSPER teams delivered interventions when pupils were beginning to experiment with substance use, but before they had transitioned to more frequent or serious use.

The relative reduction rates have potential public health significance. For example, if the rate for the frequency of cannabis use was to hold when scaling up the PROSPER partnership model, for every 100 12th-graders in non-intervention school districts using cannabis more often than yearly, there would about 86 in PROSPER districts. However, it has not yet been established whether results from study communities would generalise to populations which differ in characteristics such as ethnicity and geographic location.

FINDINGS COMMENTARY PROSPER's evaluation design provided the framework for a stringent test of the PROSPER delivery model and its menu of interventions – not just in one school, but in 28 school districts across two US states, and not just in the immediate wake of the interventions, but up to six years later, and in respect of a sub-sample, up to age 19. It was an impressive culmination to a consistent programme of work dating back decades with important implications for mounting sustainable prevention programmes. Its importance warrants searching analysis and its methods offer an opportunity to address issues common to much substance use prevention research. The commentary is correspondingly extensive but readers who just want the main points can confine themselves to the [Commentary in brief](#). Our conclusion is that the trial was not a robust demonstration of the effectiveness of the PROSPER system or its menu of interventions, one contested by the authors, whose [comments](#) are publicly available.

A major question was whether the interventions could be delivered through an implementation system [which handed much of the work](#) to representatives of the communities themselves – relatively 'real world' conditions compared to entirely researcher-controlled trials, and a system which might more sustainably embed preventive interventions in these types of communities. A second question was whether the interventions they selected – from a menu set by the central team on the basis that they had been shown to be effective – remained effective in these circumstances. If they did, it would represent a breakthrough to more sustainable, community-based substance use prevention.

On the face of it, both questions – whether the interventions could be implemented this way, and whether they would be effective – were answered in the affirmative, leading the researchers to talk of the "potential public health value" of the PROSPER model, suggesting it could underpin widespread dissemination of effective prevention. Relative to control communities, on nearly every measure published to date, children recruited to the trial in

PROSPER communities were less likely to develop the substance use patterns the programme was aiming to prevent. On the balance of probabilities, it is likely that there were preventive impacts. However, criteria closer to 'beyond reasonable doubt' are commonly applied in scientific evaluations.

Over the reality of the programme's public health potential, there are two main concerns:

- 1 First, and acknowledged by the researchers, whether the results would be replicated outside the type of communities recruited to the study.
- 2 Secondly, whether even in these communities, these results meet conventional scientific criteria for being confident they represent real effects; whether they really were a "robust" demonstration of the effectiveness of the PROSPER model.

These concerns are unpacked under the headings: [Would the system work elsewhere?](#) and [Were the interventions really shown to be effective?](#) You can find the bottom line under the heading, [Focus on the final in-school follow-up](#), where the findings of the trial are revisited taking into account methodological concerns. During this process the PROSPER trial is used to explore common features which undermine confidence in prevention programme evaluations – though the trial and its reporting are no more flawed than several others, and in significant respects, more robust than most.

Also examined are the findings that in some respects pupils most likely to use substances benefited most from PROSPER (['High risk' pupils benefit most](#)), attempts to identify the most active ingredients in the PROSPER mix ([Were there any \(more\) active ingredients?](#)), results from earlier follow-ups ([Earlier findings](#)) and for a subsample of the former pupils at age 19 ([Effects in early adulthood](#)), and the degree to which the interventions the teams selected from really were evidence-based ([Was the intervention menu evidence-based?](#)).

Commentary in brief

This section summarises points explored and substantiated later, and introduces a concern over this trial and many others – the possibility that researchers committed to an intervention will lean towards finding and declaring it works.

The first question raised above – generalisability to other types of communities – can only be settled by further studies, but the need for well-developed university-community partnerships, the small, distinct nature of the communities in the trial, and their selection and self-selection into the study, mean similar interventions cannot be assumed to be feasible or to work elsewhere.

On the second question, the weight of the findings favours the interventions, but generally only modestly and without meeting conventional criteria for ruling out chance findings. Based on these criteria – more stringent than those to which the study's authors subjected their findings – probably only one of the nine measures of current/recent use at the final follow-up of the full sample were significantly in favour of PROSPER – a small difference in the average frequency of cannabis use. Despite other positive findings, judged in terms of how the pupils ended up as they emerged into adulthood, overall the programme was not *shown* to be a success – though the possibility remains that actually it was.

Across the entire PROSPER follow-up period, findings of reduced use at different ages were consistently strongest for cannabis and at times strong too for inhalants and methamphetamine, but consistently non-significant for measures related to drinking and usually also smoking. Consistently too, a measure combining ever having illicitly used a range of substances was lower in PROSPER communities, probably, mainly reflecting fewer youngsters having tried cannabis, but this measure could reflect a single episode of use several years before rather than any current behaviour of concern.

By age 19, normally a year after leaving school, significant results were confined to a slightly lower average frequency of smoking and a lower frequency of cannabis use (about 11 v. 15 times in the past year), and slightly fewer drug-related problems, but none may have survived an adjustment for the many chances given PROSPER to register a statistically significant difference. At this age, in no case was the proportion of youngsters who had recently used various substances significantly lower in PROSPER communities, and neither were any of the alcohol-related measures.

'High risk' pupils did in some respects seem to benefit most from PROSPER, and at least did no worse, but methodological concerns limit confidence in these findings.

Attending the family programme and choosing Life Skills Training as the school programme emerged as the most active components. Since both were choices rather than random allocations, the findings can only be considered suggestive. Outside the context of the PROSPER trial, the options on its intervention menu have not been found consistently effective, particularly in trials conducted by research teams with no stake in the intervention's success – though it might justifiably be claimed that they are at least *more* evidence-based than most preventive programmes of their kind.

A further important consideration is whether PROSPER's generally small and (when subjected to stringent tests) patchily significant longer term gains are worth the investment it took to produce them. Perhaps the most hopeful finding was a reduction in drug-related problems when the former pupils were aged 19 ([▶ below](#)), but how much importance to attach to what seems a small difference is unclear, as is how representative the sample was by that stage.

These conclusions were reached on the basis of more stringent criteria for success than those considered appropriate by the researchers. The conditions for a 'researcher allegiance' effect ([▶ panel](#)) were present in the study, and some of the methods used to analyse the results were of the kind associated with such an effect – in particular the tendency of evaluators committed to an intervention to take the stance of trying to prove its success, rather than testing whether the findings would survive a rigorous *dis*-proval attempt.

This phenomenon is not just endemic, but integral to programmes of work of the kind which culminated in the PROSPER trial. Researchers not committed to the issue they are researching would be unlikely to sustain a research and development effort over decades, and that easily shades into commitment to 'their' particular way of addressing the issue. Without this commitment, the sustained effort needed to develop effective approaches might never happen, but with it, the results may not be reliable or replicable.

Would the system work elsewhere?

PROSPER was developed for small rural communities with the required infrastructure, willingness and capacity. "Results are primarily expected to generalize to the type and size of communities selected for this study," [acknowledged](#) the researchers. Their reports [highlight](#) the importance of local school and university personnel willing to engage in a collaborative effort to prevent substance use problems in their communities. Support from universities mean the local teams are not left on their own, but still they

Advocate and examiner

A difficult but (in substance use prevention) unavoidable issue is the [conflict of interest](#) between the developer of a programme whose motivation may be to show it works and promote its dissemination – perhaps for laudable reasons to do with advancing public health – and those of an evaluator, whose motivation should be to 'stress-test' the intervention by subjecting effectiveness findings to rigorous scrutiny – also for the laudable reason of not wasting resources on unproven interventions. If it survives this scrutiny, then the intervention has a strong claim to evidence-based status. Claims which emerge from an attempt to prove rather than disprove effectiveness risk being based on a less rigorous examination more friendly to the intervention being tested.

In substance use prevention, the [very common](#) overlaps between the developers of a programme and its main evaluators [lie at the heart](#) of the so-called 'researcher allegiance' effect. As in other social and medical research areas ([1 2 3](#)), there is concern that researchers with an interest in a programme's success record more positive findings than fully independent researchers. Possible reasons include implementation quality unachievable without the developer's inputs, transmission of optimistic expectations to the interventionists and in turn to their pupils or clients, and the relaxing of accepted research practices intended to prevent bias and minimise the risk of falsely declaring an intervention a success.

A prominent substance use prevention researcher [has themselves](#) highlighted the pervasiveness of developer-led evaluations, and the consequent risk of biased reporting of findings: "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." Rigorous evaluation can amount to asking someone who has invested a lifetime's work in an intervention to be prepared to test it to destruction. Counter-measures the researcher called for included "full transparency" (entailing the publication of specific research plans before the data has been collected) and making all the data available to others for re-analysis.

In the featured trial the conditions for an allegiance effect were plainly present. The lead author [led the project](#) which initially developed the only intervention chosen by all the PROSPER teams. The [institute](#) he directs is dedicated to promoting university/community partnerships. Among its "values and core beliefs" are that "Partnerships with communities – schools, Extension, other community agencies – are essential to long-term, positive

have to do the bulk of the work. Inevitably this genesis is likely to limit the applicability of the PROSPER system – but arguably limit it less than delivery systems which require a research/expert team to do all the work.

Whether the PROSPER model would be able to engage communities to implement the interventions was a major focus of the research. In most

trial communities, that test largely seems to have been passed: all the interventions [were well implemented](#), with the notable exception that the family programme engaged few of the targeted families.

However, trial communities were enrolled after an extensive selection process. By design, they were not necessarily typical in their resources or interest in preventing substance use. [Required](#) was both a school district administration and a university extension educator willing to be involved in the PROSPER programme. Of 68 otherwise eligible districts, 20 did not meet staffing requirements, and five refused to join the study. The 28 which did join had volunteered to do so rather than having been selected at random or for their representativeness. Even in these same types of small communities, their performance may not be replicated more broadly. The chances become slimmer yet if a region or nation decided to mandate or strongly incentivise such a programme across its schools – yet to fulfil a significant public health role, widespread dissemination is essential. Nevertheless, during the first year of the project two of the 14 allocated to PROSPER [dropped out](#), signalling what may in those areas have been major implementation problems.

The final sample [consisted](#) of small towns and rural communities, almost all of whose residents and team members were variously described as White or Caucasian. How a community-based programme would be engaged with and implemented in these perhaps isolated, small, and [homogenous](#) communities is, the researchers acknowledge, not necessarily a guide to what would happen in places like London and New York were a similar university-community collaboration attempted.

From the communities which did join the study and were allocated to the PROSPER arm emerged indications of where that kind of approach is most likely to succeed. On the negative side, strongest of these indicators was poverty. Findings from the trial (for details [unfold](#)  [supplementary text](#)) suggest that programmes like PROSPER flourish best in communities most willing and able to help themselves, which also tend to have less serious substance use and other problems. Poor, fractured and disheartened neighbourhoods most in need of an antidote to substance use problems seem least likely to profit, and most likely to need intensive and ongoing support. It was, however, promising that across all the PROSPER teams, members' knowledge of how to implement and evaluate prevention programmes [increased significantly more](#) than among people in similar positions in the control schools, perhaps heralding the generation of a core community network able and willing to continue these attempts.

[Close supplementary text](#)

Interviews with PROSPER team members two months after the teams were formed [revealed](#) that communities with more families below the poverty level were less ready to initiate community partnerships. In turn, six months later these communities registered lower ratings of team culture and team leadership. Beyond poverty, the degree to which from the start team members saw their communities as cohesive and ready and able to mount new collective initiatives, and the priority they afforded to preventive activities, were positively related to good team functioning six months later.

The authors of this report saw a vicious cycle developing, of poorer communities having greater substance use and other problems, leading to lower expectations for change and a sense of disempowerment, in turn undermining the ability to combat those problems, a cycle aggravated by relative lack of resources and consequent stress on public services.

Similar relationships remained apparent a year later after the teams had implemented the first of the interventions – a universally applicable version of the [Strengthening](#)

outcomes for youth, adults, families and communities.” The PROSPER trial was the culminating test of these values and beliefs, and of a programme of work dating back to the early 1990s which has attracted serial government funding. Three of his co-authors on the featured report also worked at the institute and the remaining two are described as collaborators. With so much to lose from a trial which might contradict core beliefs and values and undermine decades of work, the “stakes” were indeed high.

Families Program. At this stage teams [were re-interviewed](#), and local prevention coordinators, school heads and staff from other organisations linked to PROSPER were also asked how the project was going. For the analysts the implications of their findings were that “communities that have low level levels of resources, both economic and social, may require greater levels of support in successfully adopting collaborative prevention approaches”. One way this played out in the implementation of the interventions was that areas with better functioning teams [tended to recruit](#) higher proportions of targeted families to the family programme, though relationships with the *quality* of implementation of this and the other interventions [were equivocal](#).

[Close supplementary text](#)

Were the interventions really shown to be effective?

Answering this question is hampered by there being no publicly available study plan published or registered before data had been collected. Without this it is not known which (if any) of the many measures taken were considered primary yardsticks of the success of the PROSPER model, whether any measures have not been reported on, and which of the analyses were planned in advance of knowing what the data showed, making it impossible to rule out selective reporting.

After tightening the criteria for what counts as a statistically significant finding, only a small difference in the frequency of cannabis use over the past year is likely to remain statistically significant among the nine current/recent measures at the 12th grade follow-up. The ‘tightening’ entailed accounting for the possibility that some of the many differences could breach conventional criteria for statistical significance by chance, and not assuming that a negative result was unthinkable. These and other issues are expanded on below.

Specify the make-or-break measure

Pre-selecting a primary yardstick of success [is considered critical](#) to clinical trials. Without this, it is unclear whether the findings indicate the intervention has been successful, and the way is open for studies to take so many measures that [purely by chance](#) at least one would meet the conventional criterion (differences which would happen fewer than 1 in 20 times by chance) for statistical significance, permitting the intervention falsely to be declared a success. Specifying this yardstick in a publicly available pre-data study plan also makes it more difficult to cherry-pick or construct other measures known to produce a significant result, or to choose from among the many ways to subdivide samples those which cast the intervention in the best light.

Some change in measures is to be expected over such a long follow-up at a time when substance use will be changing due to maturation and changing patterns in US society, and also when national drug misuse priorities change. However, this does not seem to account for some of the changes ([below](#)), and it would have remained possible to specify primary measures at each new age-point but before the data had been collected.

To cater for the risk of finding ‘significant’ differences by chance, studies which test a hypothesis across multiple outcomes are advised to consider [\(1 2\)](#) adjusting the significance bar upwards. This safeguards against chance variations seeming to vindicate an intervention, but at the cost of more possibly real effects being dismissed as non-significant.

PROSPER chose not to do this, yet the risk of a chance finding was substantial. By gathering or constructing many measures, the featured analysis [provided](#) at least 18 chances for PROSPER to prove effective at the 12th grade, and another 18 at the 11th grade, plus 34 across the whole follow-up – 70 in all. In [one of the reports](#) on the trial the research team [explained](#) why they nevertheless chose not raise the bar, instead testing each outcome against the usual ‘1 in 20 by chance’ criterion. Effectively, the argument seems to have been that risking many effects being discounted might have “masked any interpretable patterns of findings” – yet the pattern of findings could still have been interpreted [by laying aside](#) for the time being the requirement that each be statistically significant, and instead focusing on the size and (in its everyday meaning) significance of the differences, expressing any conclusions in suitably tentative language.

Was a negative finding unthinkable?

Another major methodological concern is the use of so-called ‘one-tailed’ tests of statistical significance in the featured analysis and in other PROSPER follow-ups except

the first and last. Effectively these assume that a negative finding (in this case, that the PROSPER programme made things worse relative to control schools) must be a meaningless fluke and can be ignored, roughly doubling the chance of finding a significant positive effect. The argument for this assumption rested on the past record of the interventions in this and other trials, but that was and remains by no means so universally positive or so predictable that a negative outcome was unthinkable. In any event, the point of the trial was to test the new PROSPER delivery system, the effects of which could not be predicted in advance; [unfold](#)  [supplementary text](#) for details.

 [Close supplementary text](#)

Prevention programmes have either solely or mainly been tested in certain specific conditions, or [been found](#) to vary in effectiveness when conditions vary. School drug education in the USA [is evolving](#), meaning any past record of evaluated programmes improving on usual education cannot be relied on to predict the future. Some of the comparison schools [also implemented](#) drug education programmes of the PROSPER kind, while others implemented interventions described as “not evidence-based”. But if lack of an evidence base is due to the lack of research, the researchers could not be sure these programmes would fail to better the interventions chosen by PROSPER communities; before they were evaluated and the results published, PROSPER’s interventions too were “not evidence-based”. The researchers’ knowledge of what was done in control communities was insufficient (“specific information concerning the numbers of families and youth served by those interventions, and the quality of intervention delivery, was not available”) to rule out finding PROSPER had done worse than whatever was happening in the control communities, so insufficient to justify one-tailed tests.

At least two of the school-based programmes in the trial have produced some statistically significant *negative* results (1 2). One of these programmes was an early version of Life Skills Training, which when taught incompletely [has been shown](#) to produce a pattern of nil or negative findings. More on the interventions’ records [later](#).

Even if in prior trials the interventions had proved universally successful, PROSPER [trials a new delivery system](#) for them, which could not be known in advance never to produce results worse than in comparison communities. In the researchers’ own words, “No effectiveness trials of evidence-based, multicomponent family and school interventions selected from a menu could be found. The present study addresses this gap.”

To appeal to the earlier results from the PROSPER trial is to let the data from the study determine the hypothesis being tested – in this case, changing from testing for the existence of any difference to testing only for positive differences. In effect, the earlier results from the trial were used to justify doubling the chances of later findings proving significant, loading the dice in favour of discovering the long-term effects the researchers [hoped for](#). Yet [with](#) “only very limited experimental study of long-term substance use outcomes in the context of effectiveness trials,” they could not know that the directional trend of earlier results would be sustained.

Finally, in cases where one-tailed testing is justified, it [has been argued](#) that the significance level should be halved to make the test as stringent as if a two-tailed test had been used.

 [Close supplementary text](#)

Perhaps appreciating that their rationale for one-tailed tests would not be universally accepted, the researchers provided exact significance levels so these could be doubled to approximate results of a two-tailed test. But their preference for one-tailed tests gave an impression of robustness in the findings which does not stand up to a more appropriate analysis.

Missing communities and pupils

The featured report claims ‘intent-to-treat’ status for its analysis of the PROSPER trial, an important signifier that the findings were not biased by drop-out from the study or the exclusion of communities which inadequately implemented the interventions or pupils/families who inadequately attended them. Without this assurance, the randomisation intended to create a level playing field between PROSPER and control districts cannot be assumed to have been sustained, and the outcomes [may be biased](#).

The claim is made on the basis that pupils were included regardless of whether they attended the substance use education lessons or their families attended the family

intervention, an important safeguard. But this is only one aspect of a true intent-to-treat analysis. To meet this standard, every school district and pupil randomly allocated to PROSPER versus control arms of the trial [would have to have been](#) included in the outcome analyses, even if they dropped out from the study early.

Yet in the first year of the project two of the school districts allocated to PROSPER [dropped out](#) and were replaced. Especially because the trial was a trial of the PROSPER model rather than the interventions as such, these communities should have been included in the outcome analyses. This would have incorporated in the evaluation of PROSPER the possibility that even communities seemingly willing and able to undertake it will later find themselves unable or unwilling to do so, a possibility effectively excluded from the analysis. [Several analyses](#) were undertaken which showed that the replacement sites did not upwardly bias estimates of PROSPER's effectiveness. Among these were excluding the replacement sites and their paired control districts, omitting them and the two 'worst' control districts ([panel below](#)), comparing trajectories of substance use for each individual site, and comparing PROSPER-control differences within pairs of sites; none meets the standard of including the missing districts in the analysis. However, the re-analyses which were done suggest that the impacts of this departure from the standard were probably small.

The [first report](#) of outcomes from the trial says 12,022 pupils were assessed at the start, while [elsewhere](#) we learn that 11,960 provided "valid information.". With about 7784 completing 12th grade assessments, it means about 35% of the pupils intended to be included in the study were missing, amounting to 27% of those who completed baseline measures. Though par for the course for long-term follow-ups, assessors [were concerned](#) that this loss might lead to inaccurate estimates of PROSPER's effects.

The analyses tried to account for the missing data, and checks found that on known variables there were no significant differences between the kinds of pupils not followed up in PROSPER versus control districts, leading the researchers to argue that loss to follow-up was unlikely to have biased the findings. However, this is not the same as saying that the pupils who contributed data remained representative of the starting sample, and such checks [cannot eliminate](#) the possibility that unmeasured factors introduced a bias in the comparison between PROSPER and control pupils. Sophisticated statistical procedures like those employed in the featured analysis [help relatively little](#) when the reason why the data is missing may be related to the outcomes being assessed – in this case perhaps, that pupils more likely to use substances are also more likely to have missed follow-up assessments. Even when this is not the case, the degree of missing data in the featured analysis [can result](#) in biased estimates of effects.

Missing data is of greatest concern when a PROSPER-control difference was due to so few youngsters that a missing third could have easily have tipped the balance – the case, for example, in respect of the 12th-grade finding on methamphetamine use. Across the entire sample this yielded the largest relative reduction in proportions, in this case of users in the past year. But in absolute terms the reduction was from 4% to 3%, meaning that about 78 pupils held the balance in a sample missing over 4000 of all those in the sampled school years and over 3000 who completed baseline measures.

Focus on the final in-school follow-up

Once the considerations explained above are taken into account, how do the findings from PROSPER look? The short answer is, much less encouraging. Across all measures, generally the programme would not have registered a statistically significant advantage. Had two-tailed tests ([above](#)) been used, 31 of 70 of substance use outcomes recorded in the featured article or in its supplementary data would have registered a statistically significant advantage for the PROSPER communities. Add in an adjustment for multiple tests and take steps to conservatively adjust for the two replaced PROSPER communities, and at the 12th grade [just two](#) of 13 measures remain significantly in favour of PROSPER. Focus on those reflecting how the pupils ended up at the 12th grade rather than their history of having ever used substances, and just one of the nine assessments significantly favoured PROSPER – a small difference in the average frequency of cannabis use. Even with other measures of use and earlier follow-ups, it is questionable whether this single current/recent use outcome warrants the "long-term ... robust" description given their findings by the researchers. This summary of findings is expanded on below, along with the justification for focusing on the final in-school follow-up.

Before the 12th-grade follow-up there had been three assessments of the effects of

PROSPER at younger ages, and additionally assessments of trends across all school follow-ups. But if how the pupils *ended up* does not significantly differ, it matter less that there were positive findings from earlier years and positive trends – and the most recent of the two follow-ups **might have** produced inaccurate findings due to sizable differential drop-out of PROSPER and control pupils.

This is **not to dismiss** the possible value of preventing substance use during the early teens, even if there are no lasting impacts on later substance use. A clear example is the retardation in the uptake of inhalant use associated with PROSPER, any one episode of which could prove fatal, and similarly **the possibility** of long-term damage due to accidents or incidents while the teenager is drunk, generally not significantly retarded by PROSPER.

But the **main argument** for school-years prevention is as a cost-beneficial way of forestalling substance use problems and related costs and consequences in later life. PROSPER authors **themselves argue** that early findings from the project must be given full weight because delayed initiation of substance use could prevent substance use disorders later in life and/or prevent lasting developmental damage.

Reviewing this literature is beyond the scope of this commentary, but we can perhaps do enough to show that there is not necessarily any direct or proven causal link between early onset of substance use and later problems, and that the evidence for such a link is weak. There may be a lasting effect, but it cannot be assumed and has in each case to be demonstrated. The analysis of some of this literature which you can **unfold**  **below** includes three prevention-study reports **cited by** PROSPER authors in support of their argument that delaying age of onset of substance use has lasting benefits.

Close supplementary text

Evidence for a link between early use and later substance use problems rests largely on observations that children who start to use (or to regularly use) substances early are more likely to develop substance use problems later. But no matter how the findings are adjusted for other factors possibly related to later substance use, such studies cannot eliminate the possibility that early onset is merely a symptom of broader risk factors, not a causal factor in itself. The same limitation applies to studies relating early use to psychosocial difficulties in later life.

An example is a **national US survey** which found that among current and former drinkers, the chances of later drink problems “showed a striking decrease with

An alternative reckoning

PROSPER's **web site** forefronts the fact that the trial “has been recognized by two of the most rigorous review panels for prevention programs, the Coalition for Evidence-Based Policy and Blueprints for Healthy Youth Development”. Blueprints **rated** PROSPER as “promising”, while the Coalition **judged** that it only narrowly failed to meet its “top tier” standard, perhaps due to the need to replicate the findings in other types of areas.

Both sets of assessors recalculated statistically significant differences on a two-tailed basis ( **above**), apparently rejecting the explanation given for one-tailed tests. Additionally, the Coalition's analysis reported **adjustments** for multiple tests to reduce the concern that some will have produced a statistically significant difference purely by chance ( **above**). It also offers the most detailed analysis of the featured article's 12th grade results, and based these on a slightly different sample of communities, one the assessors considered less likely to produce ‘false positives’ favouring PROSPER.

The Coalition was concerned that the drop-out of two PROSPER school districts undermined the randomisation intended to assure a level playing field, so instead of all 28 districts, they relied on a re-analysis by the study's researchers which “omitted the two replacement communities in the PROSPER group along with the two control communities with the highest overall rates of substance use at 6.5-year follow-up”. Excluding the ‘worst’ of the comparison districts would have tilted the analysis against PROSPER, yet the estimated effects were only marginally smaller than across the full sample. Nevertheless, this analysis still contravened intent-to-treat principles ( **above**), and it remains possible that the two of the 14 districts allocated to PROSPER which dropped out would have registered worse outcomes than even the worst of the comparison communities.

In the re-analysis for the Coalition, given two-tailed tests the only statistically significant results relating to current/recent use at the final follow-up were a reduction in the likelihood of having smoked cigarettes in the past month from 36% in control districts to 31% in PROSPER districts, and a reduction in the frequency of cannabis use over the past year amounting to the difference between 2.1 and 1.8 on a scale scored from 0 to 7. Only the cannabis use frequency difference remained statistically significant at the conventional ‘1 in 20 by chance’ level once an adjustment had been made for multiple tests – also the only difference which across all 28 districts had a chance of surviving these stricter criteria.

increasing age at onset of use” – apparently evidence that early onset causes later problems. Strongest was the association between starting to drink aged 16 or younger and later alcohol abuse or dependence. In an attempt to eliminate factors which might give a false impression of a causal relationship between early onset and later problems, the results had been adjusted for sex, race, age, duration of drinking, family history of alcoholism and current drinking status.

Nevertheless, the authors were cautious about translating their findings into a recommendation to try to prevent early onset of drinking, because it was not known whether this was “a critical and potentially modifiable risk factor in the development of alcohol use disorders, or alternatively ... a marker or early indicator of the inevitable, and perhaps unmodifiable development of alcohol use disorders” – in other words, whether early onset was in itself a cause of later drink disorders, or a symptom of the deeper causes of those disorders which would not necessarily be changed by delaying onset.

Their caution was well founded according to authors who looked for more detailed evidence in a sample of young people to test their conclusion from adult studies that “early onset of any substance use is simply a marker of a constellation of behavior problems that already exist contemporaneously during adolescence”. Able to adjust for more relevant factors than studies of adults, their [report](#) on a national US sample of 12–16-year-olds led them to the conclusion that very early onset of regular smoking (also a surrogate for early drinking) “did not have an independent effect on current drinking or other drug use. Neither is it associated with depression, school problems, or sexual activity as studied herein. This suggests that age of onset may be more important in adult research where it serves as a retrospective marker of earlier vulnerability for a problem that has already occurred and appears to imply causality.”

An [editorial](#) in the *Addiction* journal raised similar doubts, citing “evidence which suggests that early onset and alcohol dependence are both manifestations of a general vulnerability to problem behavior and ... that attempts to prevent the development of alcohol dependence by delaying drinking onset are unlikely to be successful. Because much of the literature on earlier onset is based on secondary analyses in which many critical variables are absent, the complete picture awaits further study.”

These views seemed justified by the [findings](#) of reviewers who systematically sought and analysed studies which followed-up adolescents for at least three years to assess the relationship between early onset and later drinking, a securer methodology than asking adults to recall their early drinking and other risk factors. Published in 2014, it found only five relevant studies. Some did find a relationship between early onset of drinking and later drinking or drink problems, but “these effects attenuate or disappear with more rigorous control for confounding”. Analyses which more adequately adjusted for underlying vulnerabilities found little or no evidence of a causal link between early onset and later drinking. Despite some evidence to the contrary, “finding of a weak or absent association between [age of first drink] and later alcohol problems, after observed associations are attenuated with rigorous control for confounders, provides some support for a marker hypothesis. Put simply, those who possess an underlying vulnerability to alcohol problems may also be likely to start drinking early ... There is no strong evidence for delaying [age of first drink] as a means of preventing future adult alcohol problems.”

More convincing but still not definitive evidence could come from studies of school-age prevention programmes which are aimed at and succeed in delaying uptake of substance use. If in later years these have prevented substance use or psychosocial problems, it could be that despite underlying vulnerabilities remaining, the very fact of retarding uptake has been able to counter these. This kind of long-term follow-up is rare, [leaving](#) the evidence, at least in relation to alcohol, very thin.

PROSPER authors cite three reports of long-term follow-ups of prevention programmes as supporting a link between delayed onset and later use or problems. Two (1 2) are from the same study, the [Midwestern Prevention Project](#) in Kansas City. An immediate problem is lack of clarity over whether these results derive from a randomised trial capable of establishing an unbiased intervention effect, because other reports on what [seems](#) the same set of eight schools (co-authored by the lead researcher on the project) say the schools were not allocated at random to implement the intervention or act as control schools. Casting further doubts on the evidenced ability of the intervention to delay initiation of substance use is the [absence](#) of

formally published data on the whole sample from a much larger randomised trial in Indianapolis. A researcher who worked on the study has admitted this reporting gap “leaves many questions unanswered and reflects negatively” on the Midwestern Prevention Project as a whole.

On this somewhat confusing foundation, the long-term follow reports (including those cited by PROSPER authors) focused on cannabis and amphetamine use, begging the question of why of the substances assessed these were selected and not others, and leaving open the possibility that it was because they offered the desired results.

The [report on cannabis use](#) derived from 1606 pupils followed up individually in what are said to have been schools randomly allocated to the intervention or to act as control schools, but which other reports (1 2) suggest were not allocated at random. It related allocation to the Midwestern Prevention Project’s school/community intervention to cannabis use later in high school when pupils were aged 14–17, and both to having at age 27–30 recently received treatment for mental health problems. Data derived from 961 of the 1388 (of the 1606) former pupils randomly selected for long-term follow-up, nearly 70%. On conventional criteria there was a small but statistically significant reduction in having been treated among pupils allocated to the intervention versus those who had been in control schools, which may have been mediated by the programme leading to fewer pupils engaging in at least weekly cannabis use aged 14–17.

The [report](#) concerned with amphetamines also derived its data from 1606 pupils followed up individually in what are said to have been schools randomly allocated to the intervention or to act as control schools. The report identifies the sample with that in [another report](#) which numbered the longitudinal panel sample in the eight schools at 1264, not 1606. The amphetamine report says 1002 of the 1606 were “randomly selected to be tracked and followed-up through early adulthood”, though the [cannabis report](#) described above says of the same sample that “1388 were randomly selected for follow-up into adulthood”. The amphetamines report combined findings for methamphetamine and amphetamine use into a single measure of ever having user amphetamines, finding that up to age 18 it remained the case that former pupils in schools allocated to the intervention were less likely ever to have used amphetamine since starting high school. However, this effect was largely a historical legacy of amphetamine uptake differences already apparent in school rather than the intervention’s continued suppression of uptake in later life.

Both reports made no adjustment for the clustering of pupils and former pupils in the same schools, likely to be particularly important for the panel sample since they were selected on the basis that their middle schools fed clearly defined high schools, making it more likely that the same groupings of pupils will be sustained after the transfer. It [is explained](#) that in practice the school the child was in had little systematic influence on the measures taken in the study, but the [report](#) cited in support seems to make no mention of such a calculation, and at 1264, its panel sample was different at least in number. The other reasons given for not adjusting for clustering (especially the splitting of the children into different classes at high school) [do not seem](#) sufficient justification for omitting this important step in an unbiased outcome analysis, and may have led to a false finding that the intervention was effective.

A third report cited by PROSPER authors derived from a [trial](#) of the US Communities That Care (CTC) prevention system. Like PROSPER, rather than a specific intervention, CTC offers a process aiming to generate science-based, effective prevention initiatives led by community coalitions. The 24 communities randomly allocated to CTC or to act as controls were also reminiscent of PROSPER – small, self-contained communities with relatively socially homogenous populations. They were supported by the study and half had been selected by their states to be leading the way in prevention activities. Each community planned on average three annual prevention policies/programmes aimed at 10–14-year-olds and their families.

Just over 90% of the surviving 4388 children recruited to the study when aged 10–11 completed the [final follow-up assessment](#) at age 19, a year after normal school-leaving age. Many measures of how substance use and delinquency differed between youngsters in CTC versus control areas were taken. To counter the risk of some meeting conventional criteria for significance purely by chance, they were combined into a few overall metrics reflecting lifetime and current substance use and delinquency and other outcomes including substance use problems. None of these

significantly favoured CTC communities, and current substance use and delinquent behaviour were not significantly lower in those communities than in control areas. Among the many individual measures, across the whole sample significantly fewer youngsters from CTC areas had ever committed a delinquent act, but significantly more had recently used ecstasy, though these may have been chance findings. Among the boys only, there was a significant reduction on a combined measure of lifetime substance use and delinquency in CTC areas (and on individual measures of delinquency and ever having smoked cigarettes) but not on measures of current substance use or delinquency or a measure including substance use problems. Accepting the authors' focus on the combined measures to avoid chance findings, on this basis across the entire sample there were no significant benefits of CTC by age 19.

 [Close supplementary text](#)

For PROSPER as for other programmes, the proof of long-term benefits has to be found in the pudding of long-term findings, not assumed. For this reason, of greatest interest in the featured report is current/recent substance use at the 12th-grade follow-up, the last to include the full sample. At this stage, how many of the pupils could as they emerged into adulthood be considered 'substance users'? Relaxing the significance threshold by using one-tailed tests and not adjusting for multiple measures meant success was recorded on all but the alcohol-related measures – an exception attributed to the normative and relatively widespread nature of drinking, despite there being plenty of room for the PROSPER programme to have reduced drunkenness and drink-driving, and combating normative perceptions was what some of the interventions were partly about.

Had two-tailed tests ([▶ above](#)) been used, the picture would have been very different. In respect of recently having engaged in these behaviours, there would have been no significant results for getting drunk, driving after drinking, smoking tobacco or cannabis, or inhaling solvents, leaving PROSPER to register a statistically significant advantage only in relation to methamphetamine – the difference between about 3% of the PROSPER sample having used this drug in the past year versus 4% in control school districts. Had an adjustment also been made for multiple tests, this finding would almost certainly not have survived, leaving no significant impacts on proportions of current/recent users.

Additionally, of the three frequency of use measures over the past month or year, the one relating to cannabis was significantly lower among PROSPER pupils, though it is unclear whether this finding would have remained significant after adjusting for multiple tests. It did when the Coalition for Evidence-Based Policy [recalculated](#) the 12th-grade results, but they had also excluded the two PROSPER communities which dropped out of the study early, leaving it unclear whether the sample reported on in the featured analysis would have produced the same result. Also, different ways of adjusting for multiple tests produce different results. The way chosen by the Coalition [depends](#) on the choice made of the proportion of outcomes one is prepared falsely to find statistically significant. Vary that choice, and which findings remain significant will also vary.

If by 12th grade the interventions had little significant effect, this might have been related to a waning in their impacts on the factors thought to account for effects on substance use. Though more consistently apparent at earlier follow-ups, by the 9th grade when pupils were typically aged 14–15, just eight of the 21 factors [were significantly different](#) among PROSPER versus control pupils. Among those which were not were the pupils' confidence in their abilities to refuse an offer of drugs, their attitudes regarding substance use, and perhaps most closely related to actual use, their intentions to use a range of substances over the coming year.

On the basis of the published report, only one of the eight significant findings would certainly have passed a two-tailed test, and it is unclear whether any would have remained after adjustment for multiple measures. Whether significant or not, differences were generally very small.

'High risk' pupils benefit most

The ambition to offer a universally applicable prevention option of public health significance means the main interest is in PROSPER's impacts across all the children in the sample. It is, however, reported that on certain measures effects were significantly greater among the [roughly 30%](#) of pupils who at the start had already used cannabis, alcohol or tobacco, than among "lower risk" pupils. By the 12th grade, of the individual

substances, this applied only to cannabis, but non-significant trends were in the same direction for the other substances.

As the authors say, this means PROSPER was at least no less effective among pupils who raise greatest concerns. This is, however, not the same as saying that it had statistically significant impacts *among* these pupils – that it has been shown to significantly restrain their tendency to more rapidly and more commonly engage in substance use.

The main methodological concern is whether this way of dividing the sample into higher and lower risk was planned before it was known to produce desirable results. With no publicly available pre-data analysis plan, the possibility cannot be excluded that the division criterion was chosen not on the basis of its theoretical or practical salience, but in order to best demonstrate that PROSPER would not fail to protect children at higher risk, a concern reinforced by a shift in the dividing line from that used in an earlier study; [unfold !\[\]\(2020723f97c3fe13d8ecf52b30807736_img.jpg\) supplementary text](#) for details.

[Close supplementary text](#)

The featured study defined pre-intervention use of any one of cannabis, alcohol or tobacco as placing pupils at higher risk, the same criterion as used in [the first PROSPER follow-up](#). However, an [earlier study](#) by the same research team and in similar communities gave a cogent rationale for its choice of *two* out of three substances: “[I]ndividuals who initiate more than one substance progress from alcohol and/or tobacco as the first substance initiated to more ‘illicit’ drugs, such as marijuana. Further, research has consistently documented that a greater degree of substance involvement indicated by use of more than one substance leads to more problematic outcomes.” There is no explanation of why this rationale was abandoned in the featured study. One possibility is that it was because pupils were a year older at the start of the trial. Later the criteria were changed again [below](#).

[Close supplementary text](#)

Were there any (more) active ingredients?

Even if a varied set of combined interventions has no overall effect on a particular measure, certain variants might. With non-universal attendance at the family intervention and three different school interventions, the PROSPER trial offered a chance to investigate this possibility. The answers given ([1 2](#)) were that attending the family intervention and, among the school programmes, Life Skills Training, were the most active ingredients.

Since both were choices rather than the result of random allocation, the findings can only be considered suggestive. Choice brings with it other factors associated with the reasons for making or not making those choices which could have affected the outcomes, regardless of the impact of the interventions, most clearly in the choice only a perhaps atypical minority made to attend the family intervention. It is also a concern that the two analyses selected very different outcomes via which to test the interventions, raising the possibility of ‘cherry-picking’ to be able to report desirable results.

The difference choice to attend or not might make can be gauged from the [apparent boost](#) PROSPER found given to substance use prevention by supplementing Life Skills Training with the family programme. No such boost was apparent in a [trial](#) which eliminated choice through random allocation. [Unfold !\[\]\(f30959f11bdeaf28e16bccc4e6d35d33_img.jpg\) supplementary text](#) for details.

[Close supplementary text](#)

It was no surprise that the intervention which demanded most voluntary effort from community members was the one least well attended – by fewer than 2 in 10 of the families to whom it was offered. The [Strengthening Families Program](#) asks parents and children to attend seven sessions, typically conducted in the evenings. Though [seen by the researchers](#) as good attendance compared to alternative programmes, in [other trials](#) in similar communities, the attendance rate at one or more sessions has been about twice as high. Ironically, PROSPER teams chose Strengthening Families [because](#) the other two family programmes on the menu did not adequately involve the children, yet the children of over 8 in 10 of the families allocated to this option did not attend a single session.

Since Strengthening Families was universally chosen, there was no option to compare communities who had versus had not chosen it. Instead, across all the PROSPER districts an [attempt was made](#) to assess the added value of the sessions which *were*

attended compared to the school-based programmes alone. Without random allocation, it had to try to adjust for the possibility that families whose children were in any event least likely to develop into substance users chose to attend the sessions, and would have registered better outcomes regardless. Taken to indicate the intervention's impacts, children in the families who attended at least five of the seven sessions were 7% less likely to have drunk alcohol by the 11th grade of schooling. Among the remainder who attended fewer sessions, it seemed their attendance made no significant difference compared to not attending at all. For the authors, they had "demonstrated the added value of receiving a family-based [evidence-based intervention] in addition to a school-based [evidence-based intervention] and emphasized the importance of program attendance."

However, whether attending had any impact at all cannot securely be established by this kind of analysis, which can never be sure of having taken into account all the relevant factors affecting outcomes. In this case it seems a major gap was the inability to gather data directly from parents. Incidentally, if it was the case that just 13% of families attended sufficiently to register an impact, this analysis seems to ask much of the theory that children and families who attend the Strengthening Families Program spread its influence to others in the community.

Similar limitations affected another analysis which attempted to differentiate between the impacts of the three school-based programmes on ever having used certain prescription-only painkillers not prescribed by a doctor, a current major concern in the USA. The conclusion was that Life Skills Training was the most effective, both among the over 8 in 10 families who did not attend the family programme, and among the minority who did. The combination of attending the family programme and being allocated to Life Skills Training was associated with 16% of pupils having tried these drugs by their 12th grade compared to 26% in control schools, and 20% among PROSPER pupils allocated to Life Skills Training but whose families did not attend the family programme. On their own, the other two school programmes – Project Alert and All Stars – had no significant impact, despite having been added to PROSPER's menu on the basis of evidenced effectiveness.

The problem was that this analysis had to forfeit the reassurance of a level playing field created by randomising communities to PROSPER. For example, the four PROSPER communities which chose Life Skills Training were compared with all the control communities, yet perhaps there was something different about these communities which chose to commit to the most extensive of the school programmes. Similarly, there was almost certainly something very different about the minority of families who chose to or were able to attend the family programme, most of whom attended most of the sessions. The analysis adjusted for known differences between pupils, schools and PROSPER teams which might have affected their choice of or response to the programmes, but a *post-hoc* attempt to create virtual equivalence will always be a distant second-best to ensuring equivalence beforehand.

The apparent boost the above analysis found given to substance use prevention by supplementing Life Skills Training with the family programme was not apparent in a trial which did ensure equivalence beforehand. By allocating schools at random to either carry on as normal (the control schools), to Life Skills Training alone, or to this supplemented by the Strengthening Families Program, it avoided the possibility in the analysis above that the findings were due to certain types of families choosing to attend or not attend the family programme. Despite earlier findings, by the time pupils were typically aged 17–18, adding the family programme had not improved on the school lessons. As in the featured trial, this trial took place in secondary schools in the rural US mid-west and was led by the same researcher. It was speculated that the reason for the findings was the delayed (a year later than in the featured trial) implementation of the Strengthening Families Program. Nevertheless, these results and those from PROSPER leave it unclear whether adding the family programme improves on well-structured, school-based drug education.

 [Close supplementary text](#)

Earlier findings

In PROSPER the first (free source at the time of writing) post-intervention follow-up was conducted in the 7th grade when pupils would typically have been aged 13, 18 months after baseline measures. Of the 14 measures assessed, in eight cases fewer children in PROSPER districts had used or started to use substances, results reached on the

appropriately stringent basis of two-tailed tests of significance and meeting the conventional 'fewer than 1 in 20 times by chance' criterion, though not adjusted for the multiple chances PROSPER was given to prove effective.

A harbinger of later findings, results were strongest for cannabis and non-significant for alcohol- and tobacco-related measures. Since the baseline measures were taken, about 3.7% of PROSPER-district pupils had started to use cannabis compared to 6.1% in control districts, and 2.8% v. 4.8% of baseline non-users had used the drug in past year. Past-year use of inhalants ('glue sniffing') was also significantly less common among PROSPER-district pupils, but not past-month smoking. Recent methamphetamine use was not reported on because so rare, effectively meaning that at this stage PROSPER had no effect. The upshot is that of the six measures of recent use of different substances which were or might have been reported on, only the two related to cannabis and inhalants showed statistically significant benefits, and these may not have survived stricter significance criteria to take account of the multiple chances give PROSPER to prove effective.

Among other significant results was a retarding in the uptake of inhalant use, a lower 'lifetime illicit substance use index' score combining ever-use of methamphetamine, ecstasy, cannabis (by smoking), drugs or medications prescribed for someone else, or certain prescription-only painkillers, and a lower score on an index combining the same cannabis use measure with ever-use of alcohol or cigarettes. Both may primarily reflect differences in cannabis use.

Of the 14 measures reported on across the whole sample, seven concerned 'uptake' of substance use among pupils who at the start of the trial had never used the substance concerned. In effect these were subgroup analyses, involving only prior non-users. Of the 12,022 pupils who started the trial and the 10,781 to supply 18-month follow-up data, between 7766 and 8385 pupils were included in the analyses. Subgroup analyses [weaken](#) the reassurance of equivalence of starting point given by the randomisation of the whole sample, though in this case probably only slightly given the size of the subgroup (non-users [comprised](#) at least 90% of the full sample for each measured substance).

Perhaps more seriously, the measure used was the absolute difference in proportions of new users between PROSPER and control districts, not the [recommended](#) relative measure. Why this matters can be appreciated from a hypothetical example. Assume 100 pupils in each arm of the trial of whom 90 PROSPER pupils had not used cannabis before the intervention and 80 control pupils. If 20 of the 90 PROSPER non-users have started use 18 months later the uptake proportion is 22%. If 30 of the 80 control non-users have started use 18 months later the uptake proportion is 38%. It was these kinds of calculations which were compared in the PROSPER trial report. But in *relative* terms the number of PROSPER users had increased from 10 to 30 or by 300%, and the number of control users from 20 to 50, just 250% – demonstrating that the two calculation methods can give opposing impressions of the effectiveness of an intervention. Whether this might have been the case in the PROSPER trial is impossible to say, because there seems no table detailing the baseline characteristics of the pupils in PROSPER v. control districts, including the proportions who had already used the various substances, though it seems none of the differences were statistically significant. Similar issues arise in respect of the five measures of rates of recent use of substances, which also excluded baseline users and reported absolute differences in proportions.

At this early stage there were consistent signs – some statistically significant – that PROSPER had greater effects among higher-risk pupils – [roughly 30%](#) of the total – who at the start had already used cannabis, alcohol or tobacco, than among lower risk; pupils.

Three years later and 4½ years after the baseline measures had been taken, when the 10th grade pupils would typically have been aged 15–16, [again](#) ([free source](#) at the time of writing) there was no adjustment for the multiple chances given PROSPER to prove effective, and the report used one-tailed tests, questionably doubling the chances of a 'statistically significant' finding.

Recalculating to two-tailed tests suggests seven of the 15 measures of use or uptake of use at 10th grade were statistically significant, and most [were probably](#) strong enough (it depends on the method used and in some cases on correlation between the measures or choice of acceptable error rate) to survive adjustment for multiple tests. The same concerns about the 'new user' measures as described above apply, but this time there is

no mention of the recent-use measures also being based on the subgroup of baseline non-users. Uptake of illicit substances was most strongly affected by PROSPER, while on a two-tailed basis there were no statistically significant effects on tobacco or alcohol use either in terms of numbers of new users or recent use. Past-year use of cannabis and (more strongly) methamphetamine were each significantly less common among PROSPER pupils, and use of inhalants narrowly missed being added to this list. Of the measures combining use of several substances, one reflecting ever-use of illicit drugs was significantly lower in PROSPER districts, again probably primarily reflecting differences in cannabis use. A similar measure combining ever-use of alcohol, tobacco and cannabis just missed significance. Additionally, trends in the growth of substance use over the follow-up period were often significantly less steep in PROSPER districts. There was no analysis of whether PROSPER's effects were greatest among high-risk pupils.

Finally, two year later the featured report gave findings for the 11th grade when pupils were aged 16–17. On a two-tailed basis, at this stage eight of 18 measures registered significantly lower past or recent substance use among PROSPER v. control pupils. All the eight related to either inhalants or illicit substances, none to tobacco or alcohol. Significant findings including the combined measure of ever having ever having illicitly used a range of substances. Nine of the measures reflected recent use. Of these, three involving cannabis or methamphetamine registered statistically significant reductions among PROSPER v. control pupils, though the methamphetamine finding would probably not have survived stricter significance criteria to take account of the multiple chances given PROSPER to prove effective. Unaffected to a significant degree were recent drunkenness, drink-driving, smoking or inhalant use.

Effects in early adulthood

In 2017 [further results](#) from the study took us on another year to when the former pupils had left school and were aged about 19, a welcome and rare test of the persistence of school-age prevention efforts. The authors saw their findings as supporting “the utility of the PROSPER delivery system to provide preventive health benefits into emerging adulthood.” As a year before, the overall pattern of the findings did favour PROSPER, especially in relation to the recent frequency of cannabis use, a difference which may have partly accounted for youngsters from PROSPER districts experiencing fewer drug-related problems.

Two-tailed tests were used to assess statistical significance, an appropriate strategy. But if the bar had been raised to account for the multiple chances given PROSPER, none of the recent-use measures may have proved statistically significant, including the frequency of cannabis use. The fact that an unknown but perhaps substantial proportion of pupils asked to provide data did not do so, also means the results may be less applicable to the entire population than those gathered while pupils were still in school and easier to reach. For details [unfold](#)  [supplementary text](#).

 [Close supplementary text](#)

For the analysis a random sub-sample was selected consisting of 1985 of the pupils who had completed the sixth-grade baseline assessments, remained in the same school district three years later, and responded to the phone- or web-based survey at age 19. For the same reasons given above in relation to the 12th grade findings, the focus here is on the measures of recent use rather than those which might have related to use several years before.

Without adjusting for the many chances given PROSPER to register a significant results, of the 12 measures of how common recent (past-month or past-year) substance use was, or how frequent or problematic, three indicated a statistically significant advantage for the youngsters educated in PROSPER districts. Though all the results favoured PROSPER, in no case was the proportion of youngsters who had recently used various substances significantly lower in PROSPER areas, and neither were any of the alcohol-related measures. The three significant results were a slightly lower average frequency of smoking cigarettes (1.8 v. 2 times in the past [month or year](#)), a lower frequency of cannabis use (about 11 v. 15 times in the past year), and slightly fewer drug- (but not alcohol-) related problems – a difference of from just over to just under 1 on a range which probably runs from about 0 to 15, equating to an [effect size](#) conventionally considered small. Unlike the earlier report on pupils aged 18, no findings are presented on trends in substance use across the entire follow-up period. There were no significant differences on the measures taken of risky sex or

recent antisocial and illegal behaviours.

An unplanned analysis also assessed how many youngsters had started to use the various substances after leaving school who had never previously used them. There were no statistically significant differences, but of the seven substance types assessed, for six (cocaine, ecstasy, methamphetamine, LSD, non-prescription narcotics, and non-prescription amphetamine) initiation rates were lower in PROSPER districts. The exception was cannabis, which a non-significantly higher proportion of PROSPER-district participants had started to use.

The explanation given for the cannabis finding is that for “gateway substances (eg, alcohol, cigarettes) ... lifetime rates tend to approach ceiling levels that are eventually observed for both intervention and control groups”. In other words, for these most commonly use substances, the expectation is that the proportion of users will even out as youngsters whose use was delayed by prevention initiatives ‘catch up’. If this is the case, it means any effects in relation to the two substances of greatest public health concern – tobacco and alcohol – will dissipate, undermining the “potential public health benefit” expected from the PROSPER model.

A [previous report](#) on the trial had sought to justify not raising the bar for statistical significance by saying this would “have masked any interpretable patterns of findings across individual substance outcomes”, but in this age-19 report, no such justification is offered. Instead readers are [obliquely warned](#) that with so many outcomes tested, some might have reached the conventional criterion for statistical significance purely by chance, but then left to take this into account themselves – an adjustment many practitioners will not appreciate the importance of nor be able to conduct. With [28](#) substance use measures reported, and probably others taken but not reported, arguably the bar should have been raised ([▶ above](#)) substantially. Had that been done, of the five measures in respect of which differences between PROSPER and control districts breached statistical significance, none of the recent-use measures [may have](#) survived as significant, and fewer relating to lifetime use.

As earlier, intervention effects seemed greater (but generally not significantly so) among pupils judged to have been at high risk at the start of the trial, though the criteria determining high risk changed from earlier reports to include not just early substance use, but also “conduct problems”, an indicator of the relative poverty of the family, lower family cohesion, and living with only one or no biological parents. About 85% of the classification into low v. high risk would have remained the same had the earlier system been used. To improve the validity of the findings on the high-risk pupils, they were over-sampled, raising their proportion from 29% to 37% – in the process slightly biasing the overall sample in favour of PROSPER, since its effects were concentrated among the high-risk youngsters.

Reports at age 18 and age 19 included a total of 28 lifetime and recent-use measures of substance use, of which 17 seem to have been common to both ages. Several which were not shared make sense in terms of maturation and altered US priorities (in particular the recording at age 19 of misuse of prescription drugs and non-prescribed narcotics), but it is not clear why, for example, lifetime amphetamine and LSD misuse and the frequency of recent smoking and drinking were relevant at age 19, but not 18, nor why the extent of recent methamphetamine use was relevant at age 18, but not 19.

This variation in the gathering or reporting of outcomes [has attracted criticism](#) because it opens up the possibility that measures were selected which showed PROSPER in the best light, criticism which in turn received a [rejoinder](#) from the researchers, explaining that “since we were no longer limited to collecting data in one class period [this] allowed for a developmentally appropriate expansion of the measurement items”.

There was probably considerable selectivity – not by the researchers, but by the participants – in the former pupils included in the analysis, [creating doubt](#) over whether the findings were an artefact of the kinds of youngsters who could be contacted and who responded versus those who did not. The sample of 1985 young people followed up at age 19 were drawn from the [10,849](#) who completed pre-intervention baseline questionnaires, narrowed down to [the 7734](#) also in the same districts three years later, meaning that about 1 in 4 of the eligible population was sampled. What proportion of the approached youngsters did not complete the survey is not recorded, but it was likely to be high. The eligible population was divided into different sectors before being sampled, and the former pupils recruited to the

follow-up until the required number was reached for that sector. In a few cases all the potential participants were tried without reaching the required number, suggesting that refusal/no-contact rates could have been very high, perhaps 3 in 4. This impression seems confirmed by the considerable variability in the proportion of pupils recruited in different districts, ranging from about 16% to 54%. It is not known how representative the respondents were of the total population being sampled.

 [Close supplementary text](#)

For the first time the age-19 report alluded to the “primary substance misuse outcomes targeted by PROSPER”. Among them were those for which “evidence for positive intervention effects was most pronounced”. It seems that these primary outcomes related mainly to illegal drug use and non-prescribed use of opioid painkillers – six measures of which five were reported to be significantly lower among participants from PROSPER districts. With no evidence that these were specified as primary in advance of this analysis, nor any hints from previous years that they were the most important, the possibility cannot be ruled out that they were selected at least partly because they provided the most convincing evidence. It seems surprising that the frequency of relatively serious events like drunkenness and drink-driving were not among the selection.

Was the intervention menu evidence-based?

Studies of ingredients of PROSPER's intervention menu outside the context of the PROSPER delivery system have registered some statistically significant preventive effects, but overall do not offer strong support for their consistent success, either delivered on their own or in combination.

Most relevant was an earlier trial from the same research team of the combination of programmes (Life Skills Training and the Strengthening Families Program) which on one measure at least was judged the most effective in the featured trial. It was conducted in the same type of communities as those recruited to PROSPER. Had the more appropriate two-tailed tests been used, at the 12th grade follow-up it seems that only the initiation of cannabis smoking, and a measure combining this with initiation of drinking and tobacco smoking, would have been significantly lower after the combined programme. Analyses for current use over the past month and other more serious patterns of substance use produced no significant results favouring the interventions. It seems unlikely that the cannabis use finding would have survived an adjustment for the multiple chances given the interventions to prove significant.

The same study also tested Life Skills Training on its own, the programme chosen by four of the PROSPER teams as their school-based programme. As in other trials of this extensive curriculum, long-term effects were most convincing in respect of smoking. For cannabis they were only partially significant and not at all for alcohol. [Unfold the supplementary text](#)  for more on this trial and other studies of the family programme and school-based programmes used in the featured trial.

 [Close supplementary text](#)

Life Skills Training and the Strengthening Families Program Earlier apparently positive findings from a trial of the combination of Life Skills Training and the Strengthening Families Program derived from inappropriate one-tailed tests, and in one paper from the highly irregular practice of not using the baseline measures as the baseline for assessing change, but measures taken after the initial phases of the interventions had been completed. This unusual practice was seen as “clearly indicative of the presence of a flexible approach to data analysis” – or, in plainer language, being prepared to move the goalposts to obtain significant results. Overall, the results from this trial fall short of a convincing demonstration of the combined programme's public health potential.

Life Skills Training The same study also tested Life Skills Training on its own, the programme chosen by four of the PROSPER teams as their school-based programme. As in other trials of this extensive curriculum, long-term effects were most convincing in respect of smoking. For cannabis they were only partially significant and not at all for alcohol. Overall the evidence does not indicate that Life Skills Training will prevent substance use, only that rather patchily it has sometimes done so, most consistently in respect of the substance for which it was first developed – tobacco. Even then, the shifting criteria for success used in the different studies leaves a question over whether had consistent measures been used, the findings would have been as

consistently positive. Few significant long-term effects [have been documented](#).

Strengthening Families Program Across all studies, the version of the Strengthening Families Program used in the featured trial has generally produced non-significant impacts on substance use when these impacts have been assessed in trials not conducted by researchers who developed the intervention, or when all the trials have been [closely scrutinised](#) by an independent expert. The [registry](#) of prevention projects hosted by the European Monitoring Centre for Drugs and Drug Addiction delivered an “**Avoid**” [verdict](#) on the programme based on negative or inconclusive findings from European trials.

Among the European trials was [one in Cardiff](#) in Wales, which by June 2012 had recruited 715 families comprising those with and without “challenges” such as children not attending school or suffering from attention deficit disorder, low literacy or learning difficulties. Families were randomly allocated to normal support alone or supplemented by Strengthening Families. The primary yardsticks of success were the frequency of drinking and drunkenness among the children two years after baseline measures. Overall, no benefits were found, though there was a “Consistent pattern of benefit for families with challenges but poorer outcomes for general population.”

Possibly a lack of positive European evidence is due to few studies, to the adaptations made to the programme, and the lack of longer term follow-ups – though in US studies, effects [have been evident](#) within one and two years, as they were in the PROSPER evaluation [above](#). Possibly too, European studies did not find the programme effective due to a cultural specificity to the US context, or because its apparent effectiveness in US studies were due to methodological shortcomings or the researcher allegiance effect.

The most positive results [emerged](#) from the first of the US studies. Impressive as they were and apparent across cannabis, smoking and drinking, they derived from analyses based on just over a third of the families invited to join the study, the ones prepared or able to participate and whose children completed the follow-up assessment. Reports on studies led by the researcher who also led the project which developed the intervention [have raised concerns](#) “about flexible data analysis, multiplicity of outcomes and selective reporting of results”. The recommendation was that future evaluation plans “should be described in detail in a clinical trials registry and published study protocol. These should include specification of the primary and secondary outcome variables and exactly how these will be measured and entered into data analyses.”

For a taste of the debate on methodology readers can consult the commentaries and replies ([1](#) [2](#) [3](#) [4](#)) associated with a [report](#) on two studies including the Strengthening Families Program among their interventions.

Family-based preventive programmes Referring not just to the Strengthening Families Program but other family-based preventive programmes, analysts for the Cochrane Collaboration [argued](#) that the positive findings were unlikely to be chance occurrences, but also judged it plausible that they reflected a “prevailing bias” embodied in very small effects, poor design or conduct of studies, multiple outcomes, and few independent replications.

Project ALERT Project ALERT, chosen for their schools by four of the 14 PROSPER teams, [has not proved itself](#) in evaluations conducted by research teams not led by the programme’s developer, and therefore with less of an interest in proving it successful. The studies done by its developers found little evidence of statistically or practically significant long-term preventive effects, and some of these emerged from trials which reported so many measures that some significant outcomes were only to be expected ([1](#) [2](#)).

 [Close supplementary text](#)

Thanks for their comments on this entry in draft to research author Richard Spoth of Iowa State University in the USA and his colleagues, Dennis Gorman of the Texas A&M University System in the USA, and David Foxcroft of Oxford Brookes University in England. Comments from research author Professor Spoth and colleagues [are available](#) on his university’s web site. Some of the comments have been incorporated in the Summary and Commentary above, but there remain differences of opinion. Commentators bear no responsibility for the text including the interpretations and any remaining errors.

Last revised 10 January 2018. First uploaded 29 August 2017

[▶ Comment/query to editor](#)

[▶ Give us your feedback on the site \(one-minute survey\)](#)

► [Open Effectiveness Bank home page](#)

► Add your name to the [mailing list](#) to be alerted to new studies and other site updates

Top 10 most closely related documents on this site. For more try a [subject or free text search](#)

STUDY 2008 [Substance use outcomes 5½ years past baseline for partnership-based, family-school preventive interventions](#)

STUDY 2010 [The effectiveness of a school-based substance abuse prevention program: 18-month follow-up of the EU-Dap cluster randomized controlled trial](#)

STUDY 2000 [Education's uncertain saviour](#)

STUDY 2010 [One-year follow-up evaluation of the Project Towards No Drug Abuse \(TND\) dissemination trial](#)

REVIEW 2014 [Interventions to reduce substance misuse among vulnerable young people](#)

STUDY 2011 [Effects of a school-based prevention program on European adolescents' patterns of alcohol use](#)

STUDY 2010 [Long-term effects of the Strong African American Families program on youths' alcohol use](#)

HOT TOPIC 2017 [It's magic: prevent substance use problems without mentioning drugs](#)

STUDY 2004 [Doing it together strengthens families and helps prevent substance use](#)

REVIEW 2013 [The comparative effectiveness of outpatient treatment for adolescent substance abuse: A meta-analysis](#)