

Meeting the challenges of implementing process evaluation within randomized controlled trials: the example of ASSIST (A Stop Smoking in Schools Trial)

Suzanne Audrey^{1*}, Jo Holliday², Nina Parry-Langdon³ and Rona Campbell¹

Abstract

It is increasingly argued that the effectiveness of health promotion interventions should be measured to inform policy and practice. The randomized controlled trial (RCT) continues to be regarded as the ‘gold standard’ of health services research but health promotion practitioners have raised concerns about the RCT’s appropriateness for evaluating their work. A preferred model is currently the pragmatic trial, measuring effectiveness under ‘routine’ conditions, incorporating a process evaluation to examine context, implementation and receipt. This model was chosen by A Stop Smoking in Schools Trial (ASSIST) to evaluate an intervention in which influential Year 8 students (12–13 years old) were trained to encourage non-smoking behaviour through informal conversations with their peers. Outcome data show that the intervention was effective in reducing smoking levels in intervention schools compared with control schools. In this paper we describe the extensive process evaluation embedded within the trial and, rather than focusing on resultant data, we consider the potential for such detailed examination of process to

affect the intervention’s delivery, receipt and outcome evaluation. We describe how some acknowledged challenges were addressed within ASSIST, which have relevance for future similar trials: Hawthorne effects, overlapping roles within the team and distinguishing between the intervention and its evaluation.

Introduction

It is increasingly accepted that evidence of effectiveness is required to inform health promotion practice and assist policy makers in decisions concerning resource allocation [1, 2], but there is less agreement about which methods to employ in developing the evidence base. In health services research, the randomized controlled trial (RCT) is widely regarded as the ‘gold standard’ for evaluating interventions. However, concerns have been raised about the appropriateness of RCTs for evaluating health promotion programmes. These include difficulties in meeting the strict requirements of an RCT, the perceived preoccupation with measuring outcomes rather than the process and the implications for health promotion practice [3–9]. While the ‘model’ RCT requires that a standardized intervention, with statistically measurable outcomes, be implemented uniformly with a specific target audience, these conditions are unlikely to be met in complex health promotion programmes where social and environmental factors vary considerably. Even if strict standardization could be achieved for a trial, producing evidence of efficacy under ‘ideal’ conditions is not sufficient to generalize conclusions to the ‘real world’ in which health

¹Department of Social Medicine, University of Bristol, Bristol BS8 2PR, UK, ²Cardiff Institute of Society, Health and Ethics, Cardiff University, Cardiff CF10 3AT, UK, ³Health Promotion Division, Welsh Assembly Government, Cardiff, CF10 3NQ, Wales

*Correspondence to: S. Audrey.

E-mail: suzanne.audrey@bristol.ac.uk

promotion activities are ultimately required to be effective.

The constraints imposed by trial designs make it almost impossible to use the flexible and participatory approaches considered fundamental to good practice by many health promotion specialists [4, 6, 7, 10]. It has been argued that specific research standards should be developed for health promotion rather than being adapted from other disciplines [5, 11]. However, others have looked for ways in which the RCT could be adapted to achieve a better 'fit' with complex interventions. Recommendations have included gathering contextual and process data; employing a mixture of qualitative and quantitative research methods; using multiple research methods, investigators and data sources and undertaking translational and participatory research involving 'coalitions' which include practitioners, clients and the wider community [1, 4, 9–14]. Methods have been proposed to blend 'evidence-based' interventions with theory-based and experience-based knowledge [11, 15, 16], while others argue that the solution is to 'allow the form to be adapted while standardising the process and function' [17].

While the debate about appropriate methods to evaluate health promotion interventions goes on, the RCT continues to be widely heralded as the gold standard. In recent years a preferred model has been that of the pragmatic trial, which aims to measure the effectiveness of an intervention in routine practice [18, 19]. Pragmatic trials can adopt a 'black box' approach [20], allowing the intervention to take its course and relying on quantitative research methods to measure outcomes. However, the importance of collecting process data to avoid Type III errors (evaluating an intervention that was inadequately implemented) was identified some time ago [21, 22]. It is now widely accepted that outcome data should be illuminated by an integral process evaluation providing information about how an intervention was implemented and received, what its strengths and weaknesses were and what activities occurred under what conditions [23–26] (Table I). This was the model used for A Stop Smoking in Schools Trial (ASSIST).

Table I. *Adapting the RCT to evaluate health promotion interventions: characteristics of trials*

Type of trial	Characteristics
Explanatory trial	(i) Standardized intervention implemented in ideal research conditions (ii) Measuring outcomes (efficacy) (iii) Quantitative research methods
Pragmatic trial	(i) Standardized intervention implemented in routine conditions (ii) Measuring outcomes (effectiveness) (iii) Quantitative research methods
Pragmatic trial with process evaluation	(i) Standardized intervention implemented in routine conditions (ii) Measuring outcomes (effectiveness) (iii) Examining process (iv) Quantitative and qualitative research methods

Data from the ASSIST process evaluation provide valuable insight, for example, into the views and activities of participants [27]. However, the purpose of this paper is not to present analysis of the resultant data but to consider some of the methodological difficulties inherent in implementing process evaluation within an RCT and give practical examples of how these were managed within ASSIST. Others have discussed problems of implementing process evaluation within a school-based RCT in relation to data collection activities, response rates, communication and organizational pressures [28]. Some of the ways in which these issues were addressed within ASSIST have been discussed in previous papers concerning the implementation of the intervention [29] and the study design [30]. In this paper we are concerned with some of the tensions inherent in attempting to meet the differing requirements of the intervention and the outcome and process evaluations: the potential for increased Hawthorne effects, overlapping roles within the team and the importance of distinguishing between the intervention and its evaluation. We

begin with an outline of the trial followed by a more detailed description of the process evaluation.

Methods

A Stop Smoking in Schools Trial (ASSIST)

Smoking among young people continues to concern health practitioners, politicians and policy makers. Given the opportunity for access to the majority of young people over prolonged periods, it is perhaps not surprising that school-based health promotion programmes have been heralded as a way forward. But, despite the factors favouring this approach, the acknowledged lack of sufficiently evaluated studies, particularly in the United Kingdom, has called into question the effectiveness of school-based programmes [31–33]. Peer education has enjoyed considerable popularity as a health promotion approach with young people but, again, there is mixed evidence about its effectiveness [24, 34–36].

In ASSIST, peer education was combined with diffusion of innovation theory [37, 38]. A team of health promotion specialists and researchers developed a school-based, peer-led intervention in which Year 8 students, identified as influential by their peers, were trained to act as ‘peer supporters’ whose role was to discourage smoking through informal conversations with other students in their year. A promising feasibility study [39] led to the funding of a full-scale trial involving 10 730 students at baseline in 59 schools in south-east Wales and the west of England, the majority of which were co-educational, state-funded comprehensive schools. Thirty schools were randomly assigned to receive the intervention and the remaining 29 constituted the control group. Key elements of the trial design are outlined in Table II and the full design is described in detail elsewhere [30].

Smoking behaviour questionnaires and saliva samples were collected at baseline from all Year 8 students in the 59 trial schools. Students were also asked to complete a peer nomination questionnaire to identify other Year 8 students whom they respected, looked up to and regarded as good

leaders. In intervention schools, the 17.5% most nominated students were asked to attend a recruitment meeting at which they were invited to undertake training to become peer supporters. This ensured a desired 15% of the year group undertaking the role after attrition [40]. Nominated students who agreed to train as peer supporters, and had parental consent to participate in the intervention, were taken out of school for a two-day training event run by health promotion trainers. The training aimed to give the students the information, skills and confidence to intervene in informal situations and encourage other Year 8 students not to smoke.

During the following 10 weeks, while the peer supporters undertook their role and recorded their experiences in a diary, the health promotion trainers conducted four school-based follow-up visits to support and encourage the peer supporters in their role. A detailed description of the intervention has been published elsewhere [29]. After 10 weeks the intervention ended, the diaries were collected and the peer supporters were presented with certificates and gift vouchers to acknowledge their efforts. The trial continued, however, and outcome data were collected through smoking behaviour questionnaires and saliva samples immediately post-intervention, and when the students were in Years 9 and 10. Results from the trial are promising. Data from the smoking behaviour questionnaires indicate that the risk of students who were occasional or experimental smokers at baseline going on to report weekly smoking at 1-year follow-up was 18.3% lower in intervention schools. This promising result was supported by analysis of salivary cotinine [27].

The process evaluation

The purpose of the process evaluation was to examine the context, implementation and receipt of the intervention by providing ‘snapshots’ at different points throughout the study, as well as providing ongoing monitoring over time.

Development

Piloting of ASSIST was conducted in three schools geographically removed from the final trial location. At this stage the design and content of the

Table II. ASSIST trial design: key elements of outcome evaluation

Randomization	Cluster: because peer groups within a school year group (Year 8) were the focus of the intervention, schools were the unit of randomization Stratification criteria: (i) private, fee-paying schools (ii) Welsh-medium schools (iii) state schools State schools, additional criteria: (i) in Wales or England (ii) year size greater or less than the median (200 students) (iii) greater or less than the median proportion (19%) of students entitled to free school meals
Outcome measures	(i) smoking prevalence among 'high-risk' group [those who at baseline had experimented with cigarettes, were ex-smokers, or were occasional (less than weekly) smokers] (ii) smoking prevalence among the entire year group
Data collection	Smoking behaviour questionnaire + saliva sample (for cotinine assay and to minimize reporting bias) at: (i) baseline (Year 8) (ii) post-intervention (Year 8) (iii) 1-year follow-up (Year 9) (iv) 2-year follow-up (Year 10)
Data analysis	To measure intervention effect: (i) at student level using random effects logistic regression models with school as random effect, including school-level stratifying variables as covariates (ii) at school-level using appropriately weighted multiple regression analysis of the logarithm of each school's smoking prevalence at baseline and follow-up Additional multilevel modelling to identify interactions between school-level factors and student-level effects

process evaluation were developed and agreed. A sub-group of the research team took primary responsibility for the process evaluation but all members of the team, which included statisticians, social scientists and health promotion specialists, were encouraged to comment before the process evaluation was finalized. Similarly, process data

from the pilot study were used to refine aspects of the intervention and the outcome evaluation. These refinements were discussed at meetings of the multidisciplinary team.

Two researchers were allocated to work part-time on the process evaluation, constituting one whole-time equivalent. Because of the size of the trial, limited time and resources, and the desire to collect process data from key participants in all intervention schools at each stage of the intervention, it was agreed that the health promotion trainers would have a significant role in providing and collecting process data. The trainers were asked to administer short, self-complete questionnaires to peer supporters and teachers throughout the recruitment, training and support of the students involved. The trainers were also asked to complete similar questionnaires about their own experiences. This allowed the two researchers who were primarily responsible for implementing the process evaluation to undertake more detailed study in eight trial schools (see below). A training session was organized before the main trial started at which the content of the process evaluation and the role of the trainers in collecting and providing routine process data were discussed and agreed.

Data collection

The feasibility study [39] and pilot phase were important formative stages in the development of the ASSIST intervention, during which pertinent process data were collected and analysed. However, for the purpose of evaluating the ASSIST intervention as it was implemented during the trial, a training event for the team of health promotion trainers was regarded as the beginning of the intervention. Process data were collected systematically from 'training the trainers' until the end of the trial (Table III).

The process evaluation adopted a broad scope of enquiry but particularly focused on the intervention's context, implementation and receipt. Contextual information was gathered from all schools during the outcome data collection sweeps through questionnaires asking teaching staff and Year 8 students about their understanding of, and attitudes

Table III. *ASSIST process evaluation: key stages and issues*

Outcome data collection	Stage of intervention	Key process information
—	Training the trainers	(i) General arrangements at each stage (venue, timing, staff ratios, etc.)
Baseline data collection (Year 8): (i) smoking behaviour questionnaire (ii) saliva sample	Peer nomination	(ii) Whether stated aims and objectives were met (iii) Variations in content and style of delivery (iv) Interactions between participants (v) Response of participants (vi) Issues/concerns raised
	Peer supporter recruitment	(vii) Extent to which peer supporters were carrying out their role
	Peer supporter training	
	Four follow-up sessions	(viii) Understanding of and attitudes towards school smoking policies and practices
	Presentation of certificates/vouchers	
Post-intervention data collection (Year 8): (i) smoking behaviour questionnaire (ii) saliva sample	—	(i) Extent to which peer supporters carried out their roles (ii) Understanding of and opinions about the intervention (iii) Perceived impact of the intervention
One-year follow-up (Year 9): (i) smoking behaviour questionnaire (ii) saliva sample	—	(i) Peer supporters' longer term views of the intervention (ii) Understanding of and attitudes towards school smoking policies and practices
Two-year follow-up (Year 10) (i) smoking behaviour questionnaire (ii) saliva sample	—	Understanding of and attitudes towards school smoking policies and practices

towards, school smoking policies and practices. Relevant school staff and health promotion agencies were asked to provide information about current and planned anti-smoking initiatives. Data were collected to examine variations in the level of staff interest and commitment to the intervention as well as how schools responded to the recruitment, training and support of peer supporters. An important area of enquiry concerned the experiences of the students involved, including their reaction to the nomination process and the experience of being selected, or not selected, to train as a peer supporter. For those who were selected, the process evaluation explored their perceptions of the training and support given, and how they undertook their peer supporter role. Other questions related to the consistency and intensity of the intervention. Were there variations in the way the intervention was delivered in practice? If so, why did these

variations occur and what impact might they have had on its effectiveness?

The main sources, methods and stages of process data collection are shown in Table IV. A detailed record of the process evaluation design and rationale, including illustrations of all data collection methods used, has been compiled [41].

In all intervention schools relevant school staff, peer supporters and health promotion trainers completed short questionnaires about the recruitment, training and follow-up sessions. To gain more detailed understanding, 'in-depth' process evaluation was conducted in four intervention schools, matched with four control schools. These were all state schools purposively selected on the basis of deprivation levels, size and geographical location (see Table II for stratification criteria).

Designated contact staff at the four in-depth control schools were interviewed about the conduct

Table IV. ASSIST process evaluation data collection: main sources and methods

	Source	Data collection tool	Stage of the trial
Students	Year 8 students in all intervention and control schools	Self-complete smoking behaviour questionnaires	Baseline ($n = 10\ 261$) Post-intervention ($n = 9897$) One-year follow-up ($n = 10\ 043$) Two-year follow-up ($n = 9747$)
	Peer supporters in all intervention schools	Self-complete questionnaires	1st PS follow-up session ($n = 759$) 4th PS follow-up sessions ($n = 733$)
	Peer supporters in four intervention schools selected for in-depth study	33 semistructured interviews, 10 focus groups	Post-intervention ($n = 106$)
	25% random sample of non-peer supporters in four intervention schools selected for in-depth study who indicated they had conversations about smoking with peer supporters	Semistructured interviews	Post-intervention ($n = 32$)
School staff	Teachers supervising data collection in all intervention and control schools	Self-complete 'smoking policy' questionnaires	Baseline ($n = 282$) One-year follow-up ($n = 265$) Two-year follow-up ($n = 291$)
	Supervising teachers in intervention schools	Self-complete questionnaires	PS recruitment ($n = 27$) PS training ($n = 31$)
	Contact teachers/key staff in four intervention schools selected for in-depth process evaluation	Semistructured interviews	Baseline ($n = 8$) Post-intervention ($n = 10$)
	Contact teachers in four control schools selected for in-depth process evaluation	Semistructured interviews, self-complete questionnaire	Baseline ($n = 9$) Post-intervention ($n = 5$)
ASSIST team	Health promotion trainers at all intervention schools	Self-complete questionnaires	PS recruitment ($n = 82$) PS training ($n = 128$) 1st follow-up session ($n = 93$) 2nd PS follow-up session ($n = 95$) 3rd PS follow-up session ($n = 93$) 4th PS follow-up session ($n = 92$) Presentation of certificates/vouchers ($n = 31$)
	Health promotion trainers	Semistructured interviews	Post-intervention ($n = 11$)
	Researchers in four intervention schools selected for in-depth process evaluation	Observation pro forma	Training the trainers ($n = 2$) PS recruitment ($n = 3$) PS training ($n = 6$) 1st PS follow-up session ($n = 4$) 2nd PS follow-up session ($n = 3$) 3rd PS follow-up session ($n = 3$) 4th PS follow-up session ($n = 3$)

PS, peer supporter.

of the trial in school, the perceived effect of data collection on students and staff and the value of peer-led health promotion. Considerably more data

were collected in the four in-depth intervention schools. Individual interviews were conducted with key school staff at baseline and post-intervention,

and non-participant observations of the recruitment, training and support of peer supporters were undertaken. It was not possible to observe the peer supporters 'at work' since they had been asked to have informal conversations with their friends in everyday situations. Instead, data concerning their activities were collected immediately post-intervention through interviews and focus groups with peer supporters, and interviews with a 25% random sample of Year 8 students who reported having conversations with peer supporters in their smoking behaviour questionnaires. In addition, interviews with the health promotion trainers covered each stage of the intervention across all schools.

Analysis

The development of a clear analysis plan was a key element of the process evaluation design. A data analysis group was formed, consisting of team members with a range of qualitative and quantitative research skills, to develop the plan, and manage and analyse the multiple data sources. Results are not presented here but will form the basis of other papers. In this paper we go on to consider strategies employed to overcome some of the challenges of implementing the process evaluation within the trial.

Challenges and solutions

Despite arguments in favour of embedding process evaluation within RCTs, it is acknowledged that there are challenges involved. Here we consider three issues relevant to ASSIST that are likely to be of relevance to other health promotion interventions: Hawthorne effects, overlapping roles and distinguishing between the intervention and its evaluation.

Hawthorne effects

The Hawthorne effect refers to the potential impact of the research process on outcomes when subjects respond to special attention from researchers [42–44]. Two examples are discussed here, one relating to the peer supporters and the other concerning the health promotion trainers.

Qualitative data concerning the response of peer supporters to their involvement in the intervention

were sought through observations, interviews and focus groups. This had the potential to heighten awareness among the peer supporters of the important part they were playing in the research, and consequently to influence their commitment and performance as peer supporters. Focusing the in-depth process evaluation in four of the 30 intervention schools helped to limit this effect across the whole trial. As an additional safeguard, interviews and focus groups, during which young people were asked to consider their experiences in detail, were undertaken when the intervention had finished. This was when memories were fresh enough for the young people to reflect upon each stage, but after they had carried out their peer supporter role.

A more complex example of the Hawthorne effect concerns the impact of the research on the health promotion trainers' behaviour. Process evaluation can have a formative purpose in providing feedback as an intervention progresses so that changes can be made to improve implementation [45]. This approach is favoured by health promotion trainers who welcome the opportunity to adapt their practice in the light of experience. But process evaluation can also exert a 'quality control' effect [46] where, for example, researchers overtly monitor whether an intervention is being implemented according to specific instructions. This approach may be preferred by researchers within an RCT where standardization of delivery is required to link a specific intervention with a measurable effect.

The research priority for standardization was discussed at team meetings throughout the design and piloting phases of ASSIST. When the main trial started, the health promotion trainers were asked to complete short questionnaires in every school at each stage of the intervention, reflecting on whether predetermined aims and objectives had been achieved. In addition, observations of the recruitment, training and follow-up visits were conducted in those schools selected for in-depth process evaluation. These activities inevitably reminded the trainers of the importance of consistency in an RCT (Box 1).

There was some really good ideas on how we could do things better if you like, but that would ultimately mean changing the activities and because we weren't allowed to do that I think that was always going to be an issue from a health promotion point of view that you weren't able to do that. (Trainer 1)

I mean, I think there was always—well in my mind anyway—there was always that sort of thing of 'Oh what are we able to change? What can we do? What is acceptable?' because of the nature of the beast, and that was always, you know, a bit of a tricky thing compared to just running training when training's needed for other initiatives, you know because it was meant to be sort of following a certain model. (Trainer 7)

There were conflicts there and that was a learning experience for us because we really always change and adapt the programmes to suit the groups, even on the ground as we're doing it... But because it was to measure the scientific outputs it had to be done in the same way each time, you know, you had to just get on with it and do it, there were certain things that you couldn't let drop, that inflexibility I found sort of quite hard. (Trainer 9)

Box 1. Trainers' perceptions of the constraints of a standardized intervention

However, the process evaluation also provided opportunities, through the routine questionnaires and during post-intervention interviews, for the health promotion trainers to comment on aspects of the standardized training programme that could be improved. For example, there was some agreement among the training team that the school-based follow-up sessions for peer supporters could be improved (Box 2). Thus, although the need for standardization prevailed, the process evaluation also provided a means by which the complexities of implementation were acknowledged and professional judgement could be expressed.

Overlapping roles

Although ASSIST was relatively well funded, the size and complexity of the trial required team members to adopt a variety of roles and the health promotion trainers undertook a significant role in providing and collecting process data. Problems can arise from overlapping roles in this way, including the potential for reporting bias and a tendency towards positive appraisal. A systematic review of different study designs of peer-delivered health promotion for young people found that, while evidence from outcome studies was equivocal as to the effectiveness of the peer approach, the process evaluations overwhelmingly reported highly positive appraisals by young people [24]. This tendency may be compounded where health promotion trainers are asked, as was the case with ASSIST, to report on how well they had imple-

mented an intervention that they were instrumental in designing.

This issue was confronted within ASSIST through dialogue about the central aims of the research and the best ways in which these could be achieved. During team discussions it was emphasized that the performance of individual trainers was not being assessed but that data were being sought about the complex issue of how the intervention might operate in the 'real world'. Preliminary analysis of the post-intervention interviews has revealed willingness among the health promotion trainers, including those who were instrumental in designing the intervention, to discuss shortcomings and suggest improvements, for example, in relation to the original 'training the trainers' event and the school-based follow-up visits (Box 2).

At the same time, through team discussions and data collection activities in school, the researchers grew to appreciate some of the practical problems of delivering health promotion interventions in school settings, the difficulties inherent in requiring trainers to deliver a standardized programme in line with the requirements of an RCT and the implications for wider implementation should outcome data suggest that the intervention was successful.

Distinguishing between intervention and evaluation

Just as difficulties may arise through overlapping roles within the team, problems can occur in

I think the programme [*training the trainers*] attempted to do too much in too short a time and I think taking things in the wrong order did mean that they weren't quite sure. I mean at the end of the second training day, I felt almost all of the new trainers who hadn't been involved in the planning process were confused about what they were trying to do. Those of us who had been part of the planning process had it in our heads so well that we had gone a stage ahead if you like, and sort of expected them to catch up, and I don't think that worked. (Trainer 4)

Where we were trying to get them [*peer supporters*] to practice having conversations; and on how, telling them how, to get conversations started; we weren't successful enough and there wasn't enough on that. And there certainly wasn't enough on that in the follow-ups. (Trainer 2)

What I would have liked to have done was to hold information in pockets and bring it to them so that they had new things to bring into their conversations ... so we'd do something about the money aspects of smoking on one of the follow-ups, so they've got a whole new load of information that can go into their conversations. And then the environment another week. And then perhaps the impact on the children, new babies, unborn babies, I don't know, those different things. (Trainer 5)

Box 2. Trainers' perceptions of how to improve the ASSIST intervention

maintaining a distinction between activities that are part of the intervention and those that are part of the process evaluation. Two examples are given here.

At the first and fourth follow-up visits, all peer supporters completed process evaluation questionnaires asking them to reflect on their feelings about being selected for training, their conversations with other students and whether the training and support had adequately prepared them for their role. It is possible that completing these questionnaires encouraged peer supporters to be more reflective, and consequently more committed, throughout the 10-week intervention than if they had only filled in evaluation forms at the end of the training event as is common practice in health promotion interventions. Since all peer supporters in all intervention schools were asked to complete these questionnaires, it can be argued that they should be regarded as an integral part of the intervention. We are now, in fact, unable to evaluate how effective the intervention would have been without the process of completing these questionnaires. However, it was felt that the information was potentially valuable and similar questionnaires could be easily incorporated into the intervention if it proved successful and was implemented on a wider basis.

In a second example, the concern is whether data collected as part of the intervention should be used to inform the process evaluation. At the development stage, it was agreed that peer supporters would be asked to complete a simple diary specifying when and where they had conversations

with other Year 8 students about smoking, the length of conversations and whether conversations had gone well or badly. Discussions took place within the team about whether the purpose of diary completion was to gather information for the process evaluation or, as part of the intervention, to keep the students focused on their peer supporter role and monitor their progress at follow-up sessions. It was agreed that diary completion should be considered part of the intervention and the process evaluation would not rely upon data contained within the diaries to examine peer supporters' activities. Instead, process data would be collected using specifically designed questionnaires, in-depth interviews and focus groups. Subsequent qualitative data analysis suggests that this was appropriate. The diaries appeared to function well as a prompt for peer supporters, but cannot be relied upon to give an accurate indication of the number or quality of conversations undertaken since some peer supporters admitted during interviews and focus groups that they had made up some diary entries and forgotten to include others (27).

Conclusion

A preferred model for evaluating health promotion interventions is currently that of the pragmatic trial, measuring effectiveness under 'routine' conditions, incorporating a process evaluation. This was the

model chosen to evaluate the ASSIST school-based, peer-led smoking intervention. But while the advantages of embedding process evaluation within RCTs are increasingly accepted, there are challenges involved. A pragmatic trial aims to provide an unbiased estimate of an intervention's effectiveness as it would be routinely implemented. However, conducting process evaluation involves additional research activity with the potential to influence outcomes.

In the case of ASSIST, additional interest in the activities of the peer supporters may have resulted in greater commitment to the role than would be the case if the intervention were to be implemented without such detailed attention to process. Similarly, researcher observations and the requirement to complete short questionnaires in relation to all intervention schools at each stage of the intervention are likely to have exaggerated the consistency with which the trainers implemented the intervention. Factors such as commitment and consistency are liable to influence the reach and intensity of the intervention and its potential impact on smoking levels. While it is unlikely that Hawthorne effects can be eliminated from any evaluation, awareness of their potential impacts should inform data collection and interpretation. Some practical measures may also be taken to minimize them. In ASSIST these included focusing in-depth process evaluation in a purposive sample of schools, and conducting interviews and focus groups at the end of, rather than during, the intervention.

Other difficulties may arise, and objectivity may be compromised, when team members are asked to take on multiple roles or where the boundaries between the intervention and its evaluation are not clear. These potential problems were confronted within ASSIST through encouraging a critical and reflective research environment particularly during the development and piloting phases. Clarity of research agenda and open debate within the multi-disciplinary team were important in highlighting difficulties (such as the potential for reporting bias, or where data collection activities impinge upon the implementation of the intervention) and reaching an agreement about ways forward.

The process evaluation revealed tension between the research requirement for a standardized intervention (in the attempt to link cause and effect) and the health promotion requirement for flexibility when implementing complex interventions. In practice, the ASSIST process evaluation occupied a dual role: reinforcing standardization and providing a means by which the complexities of implementation were acknowledged and professional judgement could be expressed within the constraints of the trial. In this way we believe that the process evaluation went some way towards meeting the evaluation requirements of both researchers and health promotion practitioners.

Nevertheless, it should be acknowledged that if the process evaluation reduced variation during the trial, it consequently had the potential to increase the variation between the intervention as implemented within the RCT and the way it might be conducted outside of the trial context. There is an acknowledged lack of rigorously evaluated, peer-led interventions that have been shown to be effective in reducing smoking levels among adolescents. Bearing this in mind, the promising results from the ASSIST study have led to discussions about the wider implementation of the intervention in secondary schools in Wales. In acknowledging the 'dual role' of the process evaluation, the forthcoming ASSIST training manual will include some modifications to the intervention, particularly in relation to training the trainers and the school-based follow-up sessions, based on the recommendations of the health promotion trainers. This addresses, albeit belatedly, their desire to shape the intervention as a result of lessons learned during its implementation. At the same time, since the intervention was shown to be effective as trialled, the manual will also emphasize the importance to any subsequent trainers of implementing core features of the intervention consistently and in their entirety.

It is broadly accepted that RCTs evaluating health promotion programmes should include an examination of process as well as outcomes, and it is understandable that researchers tend to focus subsequent publications on the resultant data. However, it is also important to acknowledge the

tensions inherent in the process of implementing process evaluation within trials. Here we have outlined some issues pertinent to ASSIST which we believe have relevance for similar trials.

Acknowledgements

This study is on behalf of ASSIST (A Stop Smoking in Schools Trial). We would like to thank all the students and school staff who participated in this research so willingly. Thanks are due to Professor Laurence Moore, Fenella Starkey and the anonymous referees for their helpful comments on drafts of this paper. The project has been made possible by funding from the Medical Research Council (grant no. G9900538). The core ASSIST team comprised the following—lead investigators: Professor Laurence Moore, R.C., N.P.-L., Professor Mick Bloor and Professor Gareth Williams; Bristol evaluation team: Fenella Starkey and S.A.; Cardiff evaluation team: Mark Sidaway and J.H.; health promotion trainers: Kathleen Cordall, Lin Cooper, Nicola Hewer, Heather Anderson-Paine, Rob Sage and Lorna Coombes; and clerical support: Valerie Karatzas, Lisa Baker, Linda Esprit and Zoe MacDonald. The team was supported by additional sessional workers and we are grateful to them all.

References

1. Mukoma W, Flisher A. Evaluations of health promoting schools: a review of nine studies. *Health Promot Int* 2004; **3**: 357–68.
2. Thorogood M, Coombes Y (eds). *Evaluating Health Promotion: Practice and Methods*, 2nd edn. Oxford: Oxford University Press, 2004.
3. Glasgow RE, Lichtenstein E, Marcus AC. Why don't we see more translation of health promotion research to practice? Rethinking the efficacy-to-effectiveness transition. *Am J Public Health* 2003; **93**: 1261–7.
4. Judd J, Frankish CJ, Moulton G. Setting standards in the evaluation of community-based health promotion programmes—a unifying approach. *Health Promot Int* 2001; **16**: 367–80.
5. McQueen DV. Strengthening the evidence base for health promotion. *Health Promot Int* 2001; **16**: 261–8.
6. Nutbeam D. Evaluating health promotion—progress, problems and solutions. *Health Promot Int* 1998; **13**: 27–44.
7. Speller V, Evans D, Head MJ. Developing quality assurance standards for health promotion practice in the UK. *Health Promot Int* 1997; **13**: 215–24.
8. WHO. *Health Promotion Evaluation: Recommendations to Policymakers*. Working Party on Health Promotion Evaluation, Copenhagen, 1998.
9. Wolff N. Randomised trials of socially complex interventions: promise or peril? *J Health Serv Res Policy* 2001; **6**: 123–6.
10. Koelen MA, Vaandrager L, Colomé C. Health promotion research: dilemmas and challenges. *J Epidemiol Community Health* 2001; **55**: 257–62.
11. Rootman I, Goodstadt M, Hyndman B *et al.* (eds). *Evaluation in Health Promotion. Principles and Perspectives*. WHO Regional Publications, European Series, No 92, 2001.
12. Green LW, Ottoson JM. From efficacy to effectiveness to community and back: evidence-based practice vs practice-based evidence. In: *From Clinical Trials to Community: The Science of Translating Diabetes and Obesity Research*. Maryland: National Institutes of Health, 2004, pp 15–18.
13. Nutbeam D. The challenge to provide 'evidence' in health promotion. *Health Promot Int* 1999; **14**: 99–101.
14. Twinn S. The evaluation of the effectiveness of health education interventions in clinical practice: a continuing methodological challenge. *J Adv Nurs* 2001; **34**: 230–7.
15. Bartholomew LK, Parcel GS, Kok G. Intervention mapping: a process for developing theory- and evidence-based health education programs. *Health Educ Behav* 1998; **25**: 545–63.
16. Greenhalgh T, Robert G, Macfarlane F *et al.* Diffusion of innovations in service organisations: systematic review and recommendations. *Milbank Q* 2004; **82**: 4.
17. Hawe P, Sheill A, Riley T. Complex interventions: how 'out of control' can a randomised controlled trial be? *Br Med J* 2004; **328**: 1561–3.
18. Godwin M, Ruhland L, Casson I *et al.* Pragmatic controlled trials in primary care: the struggle between external and internal validity. *BMC Med Res Methodol* 2003; **3**: 28.
19. Roland M, Torgerson D. Understanding controlled trials: what are pragmatic trials? Clinical review. *Br Med J* 1998; **316**: 285.
20. Wight D, Obasi A. Unpacking the 'black box': the importance of process data to explain outcomes. In: Stephenson J, Imrie J, Bonnell C (eds). *Effective Sexual Health Interventions: Issues in Experimental Evaluation*. Oxford: Oxford University Press, 2003, 151–66.
21. Basch C, Slipevich E, Gold R, Duncan D, Kolbe L. Avoiding type III errors in health education program evaluation: a case study. *Health Educ Q* 1985; **12**: 315–31.
22. Dobson D, Cook T. Avoiding type III errors in program evaluation: results from a field experiment. *Eval Program Plann* 1980; **3**: 269–76.
23. Pawson R, Tilley N. *Realistic Evaluation*. London: Sage, 1997.
24. Harden A, Oakley A, Oliver S. Peer-delivered health promotion for young people: a systematic review of different study designs. *Health Educ J* 2001; **60**: 339–53.
25. Linnan L, Steckler A. Process evaluation for public health interventions and research. In: Steckler A, Linnan L (eds). *Process Evaluation for Public Health Interventions and Research*. San Francisco: Jossey-Bass, 2002, 1–23.

26. Platt S, Gnich W, Rankin D *et al.* Applying process evaluation: learning from two research projects. In: Thorogood M, Coombes Y (eds). *Evaluating Health Promotion: Practice and Methods*, 2nd edn. Oxford: Oxford University Press, 2004, 73–89.
27. Audrey S, Holliday J, Campbell R. It's good to talk: adolescent perspectives of an informal, peer-led intervention to reduce smoking. *Soc Sci Med*, 2006 Feb 2; [Epub ahead print].
28. Oakley A, Strange V, Stephenson JM *et al.* Evaluating processes: an example from a randomized controlled trial of sex education. *Evaluation* 2004; **10**: 440–62.
29. Audrey S, Cordall K, Moore L *et al.* The development and implementation of a peer-led intervention to prevent smoking among secondary school students using their established social networks. *Health Educ J* 2004; **63**: 266–84.
30. Starkey F, Moore L, Campbell R *et al.* Rationale, design and conduct of a comprehensive evaluation of a school-based peer-led anti-smoking intervention in the UK: the ASSIST cluster randomised trial [ISRCTN55572965]. *BMC Public Health* 2005; **5**: 43.
31. Rooney B, Murray D. A meta-analysis of smoking programs after adjustment for errors in the unit of analysis. *Higher Educ Q* 1996; **23**: 48–64.
32. Sowden A, Arblaster L, Stead L. Community interventions for preventing smoking in young people. In: *The Cochrane Library*, 4. Oxford: Update Software, 2002.
33. Thomas R. School-based programmes for preventing smoking. In: *The Cochrane Library*, 3. Oxford: Update Software, 2003.
34. Mellanby A, Newcombe R, Rees J *et al.* A comparative study of peer-led and adult-led school sex education. *Health Educ Res Theory Pract* 2001; **16**: 481–92.
35. Milburn K. A critical review of peer education with young people with particular reference to sexual health. *Health Educ Res Theory Pract* 1995; **10**: 407–20.
36. Turner G, Shepherd J. A method in search of a theory: peer education and health promotion. *Health Educ Res Theory Pract* 1999; **14**: 235–47.
37. Kelly JA, Murphy DA, Sikkema KJ *et al.* Randomised, controlled, community-level HIV-prevention intervention for sexual-risk behaviour among homosexual men in US cities. *Lancet* 1997; **350**: 1500–5.
38. Rogers EM. *Diffusion of Innovations*, 3rd edn. New York: The Free Press, 1983.
39. Bloor M, Frankland J, Parry-Langdon N *et al.* A controlled evaluation of an intensive, peer-led, schools-based, anti-smoking programme. *Health Educ J* 1999; **58**: 17–25.
40. Kelly J. Popular opinion leaders and HIV prevention peer education: resolving discrepant findings, and implications for the development of effective community programmes. *Aids Care* 2004; **16**: 139–50.
41. Audrey S, Holliday J, Parry-Langdon N. *A Stop Smoking in Schools Trial (ASSIST): Process Evaluation Manual*. Bristol: Cardiff University and University of Bristol, 2003.
42. De Amici D, Klersy C, Ramajoli F *et al.* Impact of the Hawthorne effect in a longitudinal clinical study: the case of anaesthesia. *Control Clin Trials* 2000; **21**: 103–14.
43. Mangione-Smith R, Elliot MN, McDonald L *et al.* An observational study of antibiotic prescribing behaviour and the Hawthorne effect. *Health Serv Res* 2002; **37**: 1603–23.
44. Mayo E. *The Human Problems of an Industrial Civilization*. New York: MacMillan, 1933.
45. Helitzer D, Davis S, Gittelsohn J *et al.* Process evaluation in a multi-site, primary obesity-prevention trial in American Indian schoolchildren. *Am J Clin Nutr* 1999; **69**: S816–24.
46. McGraw SA, Stone EJ, Osganian SK *et al.* Design of process evaluation within the Child and Adolescent Trial for Cardiovascular Health (CATCH). *Health Educ Q* 1994; **S2**: S5–26.

Received on May 13, 2005; accepted on March 30, 2006