Process evaluation: the new miracle ingredient in public health research?

Qualitative Research

http://qrj.sagepub.com
vol. 10(6) 699–713

ALISON MUNRO
Glasgow Caledonian University
MICHAEL BLOOR
University of Glasgow

ABSTRACT Good evaluation practice in public health research has become equated with the inclusion of a mixed-methods 'process evaluation' alongside an 'outcome evaluation' to gather data on how and why interventions are effective or ineffective. While the incorporation of process evaluations in randomized controlled trials is to be welcomed, there is a danger that they are being oversold. The problematic position of process evaluations is illustrated by data from an evaluation of an unsuccessful schools health promotion intervention. The process evaluation data (designed to 'explain' the outcome evaluation results) must be collected before the outcome evaluation results are typically available: unanticipated outcomes cannot always be addressed satisfactorily from prior process data. Further, qualitative process data draw inductively general inferences from particular circumstances and the generalizability of those inferences is therefore uncertain: qualitative data can deepen our understanding of quantitative data, but the commensurability of the two classes of data remains problematic.

KEYWORDS: *complex interventions, drugs, mixed methods, process evaluation, public health, schools, smoking, triangulation*

Introduction

It has been suggested that, after the long-standing 'paradigm wars' between qualitative and quantitative methods, 'paradigm peace' has broken out (Bryman, 2006). One seeming manifestation of this paradigm peace is found in public health research where qualitative methods have been increasingly incorporated into outcome studies, that may utilise a variety of study designs including the randomized controlled trial (RCT), as part of a mixed-methods 'process evaluation': a quantitative 'outcome evaluation' is conducted to determine

DOI: 10.1177/1468794110380522

whether or not a public health initiative has been effective, while a mixed-method 'process evaluation' is conducted to explain why the intervention was successful or unsuccessful. Such outcome-evaluation-plus-process-evaluation RCTs correspond to what Moran-Ellis et al. (2006) have termed 'integrated methods', where different methods retain their paradigmatic natures but are inter-meshed with each other to deepen understanding of the phenomenon under study. Such integrated methods designs have clear advantages which need not be reiterated here, but there are also dangers that the burden of expectation on process evaluations is too great.

COMPLEX INTERVENTIONS

Much social research, including evaluations of educational interventions and public health initiatives, involves the assessment of 'complex interventions' (Giannakaki, 2005; Oakley et al., 2006; Victora et al., 2004; Young et al., 2008), that is to say that the interventions are multifaceted, organizationally elaborate, and socially mediated. These complex interventions sit uncomfortably with the classic positivist model of health services research, with its 'hierarchy of evidence' and the randomized controlled trial (RCT) at its apex (Maynard and Chambers, 1997): a simple RCT of a complex intervention will be wholly lacking in explanatory power (Bradley et al., 1999). The outcome of the RCT will tell the evaluators whether or not the intervention has an effect, but they will have no idea why: the processes of the delivery of the intervention and its reception remain in a 'black box'. This lack of explanatory power is of course particularly unfortunate when the outcome evaluation shows the intervention to have been ineffective, because not only are there no data available on why the complex intervention failed, there are also no data available to suggest how that ineffective intervention might be successfully modified. Many an experienced investigator has thus been reduced to embarrassing flights of wild speculation in accounting for his or her disappointing RCT results.

PROCESS EVALUATION

It has therefore become commonplace, and considered as part of good research practice, for a 'process evaluation' to be paired with an 'outcome evaluation' in the design of studies of complex interventions (Oakley et al., 2006; O'Cathain, 2009). Advice on the conduct of RCTs by bodies such as the United Kingdom's Medical Research Council lays particular stress on the importance of a process evaluation component in the pilot or early stages (Phases I and II) of an RCT, both to contribute to the decision about whether to proceed to a full-scale (Phase III) trial, and to inform final changes in the design of both the intervention and the outcome evaluation. Process evaluations gather data on the social processes involved in the delivery of the intervention, the reception of the intervention and the setting of the intervention. They frequently entail mixed methods (Oakley, 2005) involving both survey questions and semi-structured interviews, sometimes focus groups, sometimes direct observation,

and sometimes (in the collection of data on the setting of the intervention) documentary methods.

The advocacy of process evaluations seems to be, if not admirable, then at least unexceptionable. Until, that is, one begins to examine in detail the explanatory burden being placed on the process evaluation components of these evaluation studies. Consider, for example, this statement from an early widely-cited paper on process evaluation design:

> Process evaluation complements outcome evaluation by providing data to describe how a program was implemented, how well the activities delivered fit the original design, to whom the services were delivered, the extent to which the target population was reached, and the factors external to the program that may compete with the program effects. (McGraw et al., 1994: S5)

The reader may think that McGraw et al. are looking for an awful lot of bangs for their buck. But Oakley et al.'s (2006) influential 'analysis and comment' paper on process evaluations in the *British Medical Journal* is, if anything, even more demanding:

> [Process evaluations] ... may aim to examine the views of participants on the intervention; study how the intervention is implemented; distinguish between the components of the intervention; monitor dose to assess the reach of the intervention; and study the way the effects vary in subgroups. (Oakley et al., 2006: 413)

A subsequent paper from the same team argued that process evaluations were also essential to allow an assessment of the generalizability, or external validity, of complex interventions (Bonnell et al., 2006). Consider also that these complex interventions are frequently multi-site trials, involving perhaps a score or more of clinics (or communities, or schools, etc.) and perhaps an equal number of intervention delivery teams. Furthermore, the effect of the intervention may be time-limited and data may have to be collected in these various sites longitudinally. The reader will no doubt now be forming the notion that any aspiration to collect systematic data on all these different topics or themes, at all these different sites, at a number of different points in time, will inevitably lead the researcher into that nightmare study mis-design, the intensive large-scale study. Enormous effort expended, mountains of complex data collected, and no earthly chance of making any sense of it all. Indeed difficulties associated with the collection of 'too much' data have been reported by other process evaluators (e.g. Hong et al., 2005).

Consider too that funders are not just being expected to fund process evaluations: there is also the outcome evaluation to fund (does the intervention have an effect or not?). Outcome evaluations can also come with a hefty balance sheet, particularly if, as well as survey data, they involve physical measures – lung function tests, oral fluid samples for testing exposure to illicit drugs or tobacco, and so on. The funder may also be being asked to fund the intervention itself, including salaries and training for the team delivering the intervention. And there may be other costs too: for example, every trial

application to the UK's Medical Research Council is required either to incorporate a health economics evaluation or an explanation of why such an evaluation is not required (Medical Research Council, n.d.). There are therefore budgetary reasons as well as sound study design reasons for limiting the scope of data collection for a process evaluation. Hence, the contention has been made that there is a need to prioritize both the 'type and amount of data collected' (Linnan and Steckler, 2002).

The wise Principal Investigator therefore does not design his/her process evaluation to collect comprehensive data on *all* aspects of the delivery, reception and context of the intervention. Rather, the optimum design will involve some element of selective sampling in respect of the more labour-intensive qualitative data collection and analysis. A typical design might involve the following comprehensive data components: first, comprehensive survey data (cross-sectional or longitudinal) at all intervention and control sites (clinics, communities, schools, etc.), usually as part of the outcome evaluation survey instrument; second, attempted comprehensive gathering and analysis of documents concerning contextual and confounding factors relating to the possible differential reception of the intervention at the different sites and to changes in outcome measures at the control sites; third, additional interview data on the same topic collected from key respondents (leading clinician, community leader, senior teacher, etc.) at the intervention sites; and fourth, attempted comprehensive interview data, gathered from those responsible for delivering the intervention at the various sites and seeking information how variously the intervention was delivered and received. While the typical selective data components might be: observation of the delivery of the intervention at a selection of the sites; and qualitative interviews or focus groups with a selection of the recipients of the intervention at a selection of the intervention sites together with comparable data from a selection of control sites. See Parry-Langdon et al. (2003) for an extended description of one such process evaluation design.

However, while the above may be a sound process evaluation design in that it blends comprehensive and selective elements in a practicable manner, it does not necessarily meet the demands of positivistic health services research in opening up the 'why questions' black box – it does not explain unproblematically why a complex intervention worked or did not work. This paper uses data from a process evaluation of a complex intervention that did not work as a particular case study to illustrate a number of general reasons why there should be a necessary degree of indeterminancy to process evaluation explanations of outcome evaluation results. Process evaluations are not a miracle ingredient.

Methods

The process evaluation reported here was conducted as part of a feasibility study (Phase I and II) for a full-scale trial (Phase III) of a schools-based, peer-led,

drugs prevention programme. The aims of the study were threefold: to develop and deliver a training package for S2 school pupils; to assess the feasibility of a future (Phase III) randomized controlled trial; and to conduct a process evaluation of the delivery and impact of the interventions. Based on the successful ASSIST (A Stop Smoking In Schools Trial) programme for reducing the uptake of cigarette smoking amongst adolescents (Campbell et al., 2008), the intervention aimed to train a proportion of 'influential' S2 school pupils (second year of Scottish secondary school) to intervene with their peers, i.e. to have conversations in informal settings such as break times, within schools to prevent cannabis smoking. The influential pupils were nominated anonymously by their peers at the first (of three) data sweep(s) and then trained to become 'peer supporters' (n=107). Since it was thought possible that an intervention which was already known to reduce pupil smoking prevalence could, of itself, also reduce cannabis prevalence, the feasibility study followed a three-arm design with two intervention arms and a control arm. A total of six schools participated in the study.

In one intervention arm, two schools received the original ASSIST programme which involved two days training for 'peer supporters' at a venue away from school, plus follow-up visits, over a 10-week period. In the second intervention arm, two schools received the ASSIST training and follow-up plus an extra day's training devoted exclusively to training on cannabis prevention. The training was delivered by experienced health promotion trainers, some of whom had been involved in the delivery of the training in the original ASSIST study. Since the 'ASSIST' acronym had already been used in Scotland for an adolescent suicide prevention programme, the acronym CASE (Cannabis And Smoking Education) was used instead; those schools receiving the renamed ASSIST intervention we designate here as CASE schools, while those schools receiving ASSIST and an additional third day's training on cannabis, we designate CASE+. The two control schools were asked to continue with their usual programme of health education and therefore received no additional intervention.

Survey data and saliva samples (as an encouragement to truthful reporting) were collected at pre-intervention baseline, immediately post-intervention and three-months post-intervention. From a potential 1128 pupils, 896 participated at the first data sweep (achieving a 79% response rate), and data were collected from 732 pupils at all three sweeps. Outcome evaluation data were collected on cigarette and cannabis smoking, but because it was not expected that differences between intervention and control schools in these behavioural measures would necessarily be evident within this deliberately 'under-powered' feasibility study (as opposed to a full-scale trial involving many more schools), additional outcome measures were piloted and used, designed to elicit from pupils their future intentions on cannabis smoking, both in six months time and aged 16. See Munro and Bloor (2009) for a full report of the outcome evaluation.

The design of the process evaluation largely followed that of the earlier ASSIST trial (Audrey et al., 2006a; Parry-Langdon et al., 2003). Survey data were collected at each of the three data collection sweeps on numbers of conversations (as reported by both peer supporters and their fellow pupils) about both cigarette smoking and cannabis. Post-intervention, focus groups (see, for example, Bloor et al., 2001) were conducted with the peer supporters (in one intervention school it only proved possible to conduct a focus group with the female peer supporters), qualitative interviews (see, for example, Gubrium and Holstein, 2002) were conducted with the trainers, and qualitative interviews were also conducted with key staff contacts in the intervention and control schools. Observational data (see, for example, Atkinson et al., 2001) were also collected on the training. Semi-structured interviews and focus groups were transcribed and systematically analysed along with the fieldnotes. Ethical approval was granted by the UK National Health Services' local research ethics committee.

The results of the feasibility study were uneven. Measures of the dose, reach and fidelity (Young et al., 2008) of the interventions were all found to be good. In other words, the training was delivered in the way it was intended, and was received well by the target group (the pupils). In addition, the peer supporter training was viewed positively by the key school personnel. Furthermore, no major harms or negative consequences were experienced by pupils: in two of the schools, the key contacts stated that the peer supporters had gained in confidence and were making positive contributions to school life; 73 per cent of the peer supporters agreed with the statement that 'being a peer supporter made me feel more confident'. And useful information was also gained that was relevant to the design of a future full-scale trial (Munro and Bloor, 2009). However, although the expected absence of an effect on reported cigarette and cannabis smoking was confirmed, there was also no effect on intentions to smoke cannabis in the future between the intervention schools and the controls. Further, and rather alarmingly, there was actually a significant increase over time among the peer supporters in their expectations that they would be smoking cannabis when they were 16. And, in respect of the survey data on pupil conversations about cannabis, there was a significant difference (p=0.03) between the two CASE+ intervention schools: in the immediately post-intervention survey, 27 per cent of pupils in one school reported having had a conversation with a peer supporter about cannabis, but only 9 per cent of pupils reported such a conversation in the other CASE+ intervention school. This difference between the CASE+ schools was not evident in pupils' reports of conversations with peer supporters about cigarette smoking (Munro and Bloor, 2009).

In what follows, we report in detail on our attempts, using the process evaluation data, to interpret the effects of the feasibility study and – we hope – to illustrate thereby the necessary degree of indeterminacy in process evaluation findings in general.

Problem 1: data gathering without the benefit of hindsight

The prevalence of cannabis use (in the last month) among Scottish 13 year-old pupils is only 2 per cent (Maxwell et al., 2007). The questions about pupils' intentions or expectations about future cannabis use (e.g. 'Do you think you will take cannabis [marijuana, dope, hash, blow, joints] when you are 16 years old?') were therefore chosen as alternative outcome measures for the feasibility study because it was thought that questions about future intentions would produce higher proportions of would-be cannabis smokers than questions about current cannabis smoking, and thus would offer a greater chance of showing an intervention effect in this relatively small population. In the event, and despite prior piloting, the questions proved doubly problematic. In the first place, the proportion of would-be cannabis smokers (at 16 years) at baseline was low – 4.2 per cent in the four intervention (CASE and CASE+) schools, so the questions' supposed utility as more sensitive outcome measures was undermined. And, in the second place, regression analyses of changes in intentions at 16 years between pre-intervention baseline and three-months post-intervention showed that peer supporters were 4.3 times (95% confidence interval: 1.6–11.7) more likely over the period of the study to think they would be smoking cannabis at 16. Being a peer supporter proved to be a similarly significant factor in regression analyses conducted on changes in expectations of smoking cannabis in three months time. This was an unexpected finding. It would be a matter of concern if it were to be repeated in a full-scale trial. It would not be the first time that a drugs education intervention had produced an unintended negative effect, increasing the propensity to future drug misuse (cf. Palinka et al., 1996), but the point we wish to stress here is that the regression analysis findings were unexpected.

The significance of the unanticipated nature of the findings lies in the fact that study timetables do not allow the luxury of the postponement of process evaluation data collection until the results of the outcome evaluation are known. In planning the process evaluation data collection it is good practice, even necessary, to *anticipate* the results of the outcome evaluation. In this study (as in many similar studies), much of the process evaluation data is collected in advance of the final sweep of outcome evaluation data collection (in this case, three months after the end of the intervention being monitored in the process evaluation). By the time the regression analyses were completed, implying an increased propensity for peer supporters to expect that they would go on to smoke cannabis, the focus groups with the peer supporters had long since been completed (and indeed a new school year had started) and the opportunity to quiz the peer supporters on these unanticipated changes in their expectations had been lost. The focus groups could only be oriented towards addressing anticipated changes because of their place in the study timetable. As it was, the focus groups yielded a very positive view of the peer

supporters' reactions to the training on cannabis and gave no hint that peer supporters might be more likely to use cannabis after the training.

For what it is worth, we strongly suspect that the changes in peer supporters' responses to the intentions/expectations questions pre- and post-intervention had nothing to do with an increased resolution to use cannabis in the future, but rather were indicative of a sense of fatalism among these adolescents about their future vulnerability to drug misuse, a sense of fatalism that may have been inadvertently increased by their increased awareness of cannabis as a result of their peer supporter training. This would be consistent with other findings on a culture of fatalism towards health risks including drug misuse (e.g. Douglas and Calvez, 1990), but we are unable to support this suspicion from focus group data because we did not anticipate the future need to examine this issue. This is not sloppiness on our part, it is a problem inherent in conventional process evaluation designs: to explain all those why and how questions arising from the outcome evaluation, the process evaluation needs to be designed with 20:20 hindsight.

Problem 2: induction and generalizability

All analysis of qualitative data is based upon inductive (rather than deductive) thinking, upon drawing general inferences from particular circumstances (consider for example, methodological writings on analytic induction techniques in qualitative analysis [e.g. Bloor, 1978]). And there is therefore a tension here between one of the alleged purposes of process evaluations, namely to pronounce on external validity or generalizability, and the methods of analysis used, that of inferring general statements from particular instances. This same tension can be found in our analyses in this study.

As previously stated, in one of the CASE+ intervention schools only 9 per cent of the pupils (as opposed to 27% in the other CASE+ school) reported a conversation about cannabis with a peer supporter in the immediate aftermath of the intervention. When it came to conversations about cigarettes, the discrepancy between the two schools was less marked, with 23 per cent of the pupils in the first-mentioned school reporting a conversation with a peer-supporter, compared with 34 per cent in the other CASE+ school. And indeed that 23 per cent of pupils reporting conversations about cigarettes was higher than in either of the two other (CASE) schools that just received peer supporter training about cigarettes and not about cannabis. So we might infer that the peer supporters in the first-mentioned school only had difficulties with their peer supporter role in respect of preventing cannabis smoking, not in respect of cigarette smoking.

There is some support for this inference from the focus group data and also grounds for further development of the inference. There was general agreement, among peer supporters in all the focus groups and across the different schools that cannabis was more difficult to talk about than cigarette smoking.

The following extract is taken from the first-mentioned school, but similar sentiments could be found in transcripts of other groups:

Researcher: Yeah and how did you decide like whether you were going to talk about cannabis? Or *did* you decide whether you were going to look at, talk about cannabis this day or talk about cigarettes?]
Pupil 3: Smoking, because not as many people
Pupil 2: [Use cannabis
Researcher: So what did that mean then for cannabis, was it]
Pupil 5: It was still spoken about but not as much
Pupil 4: Because there's not, not a lot of people do it, aye, like, well not the people we know anyway
Researcher: Uhuh, so what did that mean? Did that mean you tended not to bring it up, or it made it very difficult to bring up, or]
Pupil 3: It made it quite difficult]
Pupil 2: [Quite difficult
Researcher: Right.
Pupil 3: Cos they might think that you thought they were taking it then.

The extract begins with the group suggesting that cigarettes are easier to talk about because cigarette use is more widespread, but it concludes with the suggestion that talking about cannabis to fellow pupils is more interactionally difficult because the fellow pupil may think that the peer supporter believes them to be a cannabis smoker. So we now have some grounds for peer supporters feeling that talking to fellow pupils is more difficult, and evidence that peer supporters in one of the CASE+ schools tended to concentrate their efforts on discussions about cigarettes rather than cannabis.

There is the further question of why this concentration on cigarettes should occur in one of the CASE+ schools and not the other. Some relevant evidence here is supplied by the qualitative interviews with the trainers who both commented that the girls who were peer supporters in the first-mentioned CASE+ school were likely to have difficulties in performing their roles:

Trainer 1: Very, very painfully shy and they seemed to really lack self-esteem and quite a lot of the things that they were asked to do they really struggled with – like with the role play.
Trainer 2: You just think that if they are not confident enough to do things in that safe environment [the training venue], you know, that's the whole idea that you can try things there, you know, and then put them into practice when they are back in school. But they weren't happy to try it really.

Moreover, the male peer supporters in the same school expressed doubts in the focus group of the effectiveness of the cannabis component of the intervention. So we might infer that, in this school, the perception that talking about cannabis was more difficult than talking about cigarettes was wedded to lack of

confidence (girls) and pessimism about effectiveness (boys) to result in a concentration of peer supporters' efforts on cigarettes, rather than on cannabis.

Ostensibly, this seems a valuable finding to emerge from the process evaluation. It suggests that the intervention needs to be re-designed to remove the element of discretion for peer supporters that allows them to concentrate on prevention of cigarette smoking, rather than cannabis smoking. It would seem that, despite the substantial economies involved in training peer supporters for more than one task at a time, peer supporters should only be trained to prevent cigarette smoking in S2 (second year). Training to prevent cannabis uptake could follow separately in S3 (third year).

While this may seem like just the kind of valuable finding that process evaluations are designed to generate, there is nevertheless a potential problem here: we are drawing a *general* inference about intervention design from the *particular* case of one school. But are we justified in generalizing from this one particular case? Possibly, a contrary argument could be constructed that we should be generalizing from the other CASE+ school (where 27% of pupils – as opposed to 9% in the first-mentioned school – reported speaking with a peer supporter about cannabis) – why focus analytic attention on one school rather than another? This particular contrary argument can be addressed by pointing out that peer supporters in the other CASE+, although they too reported that it was more difficult to talk about cannabis than cigarettes, did not appear as unconfident and pessimistic as in the first school.

A second contrary argument could be constructed around the fact that, although the peer supporters in the first CASE+ school reported difficulties in talking about cannabis, the proportion of conversations which pupils reported having about cannabis in that CASE+ school (9%) was in fact almost identical with the proportion of conversations which pupils reported having about cigarettes in the two CASE schools (10%). In this argument, although the peer supporters in the first CASE+ school complained about the difficulty of talking about cannabis, and although the trainers' judgement implied that the girls in particular might 'struggle' with the intervention, in fact they performed adequately in having conversations about cannabis and performed above the norm in having conversations about cigarettes: 23 per cent of pupils in the first CASE+ school reported a conversation with a peer supporter about cigarettes. We are inclined to reject this argument since, in the ASSIST trial (with 29 intervention schools), the proportion of pupils reporting a conversation about cigarettes with a peer supporter was 20 per cent (Audrey et al., 2006b: 326), indicating that the peer supporters in the first CASE+ school were in fact performing at a normative level for conversations about smoking and the peer supporters at the two CASE schools were performing below the expected norm. Nevertheless, it can be seen that our original inference (peer supporters will tend to concentrate on smoking prevention over cannabis prevention and so should be trained in each of these interventions separately rather than simultaneously), although it remains plausible, is certainly contestable. The

generalizability of inductively-generated findings from qualitative research is problematic. Other questions that may be worth exploring in any further feasibility work would be whether a longer training course, to tackle pupils' (often) stereotypical notions of cannabis (and other drug) users would be more effective, and whether directly training teaching staff, along with pupils, may help pupils to feel more 'comfortable', and supported, discussing an illegal drug in and around the school environment.

Conclusion

This is not the first paper to draw attention to some of the difficulties in conducting process evaluations. While commending the inclusion of process evaluations in trial designs, Wight and Obasi (2002) have pointed to the problems that arise where intervention and evaluation functions are not carefully separated, as where members of the team delivering the intervention may be asked to also collect process evaluation data. They also point to problems of interpretation and suggest that the problems of bias in the interpretation of qualitative data are such that it is best to complete the analysis of the process data before the analysis of the outcome evaluation, so as to identify the key process factors likely to affect outcomes uninfluenced by prior knowledge of what those outcomes are. It might be objected that this is not always possible logistically. But we have raised here a second difficulty in the integration and sequencing of process and outcome evaluations, namely that the decisions on the topical foci of the process evaluation data collection need to anticipate the results from the outcome evaluation, and such anticipation is not always going to be wholly successful. As a consequence, some results of the evaluation may remain only accounted for in a speculative manner, as is the case with our tentative suggestion that the greater propensity for the peer supporters to expect that they will be using cannabis at 16 years is due to a combination of the peer supporters' greater exposure in the training sessions to the risks of cannabis smoking alongside a sense of adolescent powerlessness and fatalism – we never anticipated the need to collect process data on the latter topic.

The use of multiple methods has long been the hallmark of good research design (see, for example, Barbour, 1999). But it is a mistake to think of qualitative and quantitative findings as commensurate in some straight-forward fashion. Elsewhere, Bloor (1997) has argued that triangulation, in the sense of using findings produced by one method to validate the findings produced by a second method, is a chimera: each method will produce findings that are separately distinctive in respect of their degree of specificity/abstraction and their topical focus. Although use of multiple methods can deepen analytic understanding of a specific issue, straightforward replication is an impossibility.

The use of qualitative methods in process evaluations alongside quantitative data (whether it be process or outcome data) is rarely straight-forward. Data generated by qualitative methods may be made to bear on the interpretation of

data generated by quantitative methods, they may deepen and enrich our understanding of quantitative findings, but that deepened understanding will always be nuanced and qualified and rarely determinate. The qualitative data reported earlier – on the low levels of confidence of some female peer supporters, the low level of perceived self-efficacy of some male peer supporters, the perceived greater interactional difficulty of discussing cannabis rather than tobacco, and a consequent preference for discussing tobacco rather than cannabis – these data all appear to deepen and enrich our understanding of quantitative data on numbers of reported conversations about cannabis and tobacco. But they do not necessarily explain those quantitative data in some determinate fashion: other explanations remain possibilities.

Relatedly, the interpretation of qualitative data involves making general inferences from particular instances. In the case study reported above, we drew general inferences from data on pupils in a particular school where numbers of reported cannabis conversations were low. But in a second CASE+ school, reported cannabis conversations were much more numerous. It has seemed to us that a properly cautious inference to draw from this feasibility study is that the intervention should be redesigned so that peer supporters efforts are directed to address cigarette and cannabis smoking sequentially rather than simultaneously, but we have no analytic grounds for giving more weight to data from one school rather than another. The generalizability of inferences from process evaluation data remains problematic.

Many public health and health promotion interventions are a far cry from the simple dose-response models which randomized controlled trials typically address. In adapting trial designs to the evaluation of complex interventions, researchers have sought to complement outcome evaluation components in their designs with process evaluations which may provide answers to many questions on which outcome evaluations are silent. Process evaluations do indeed enrich our understanding of the social processes involved in the delivery and reception of complex interventions. But process evaluations do not slot comfortably into evidence-based medicine's 'hierarchy of evidence', providing interpretations to which a degree of indeterminancy is always attached that cannot be expressed in probabilistic terms. While it is possible and desirable to integrate qualitative and quantitative methods pragmatically within a research design (Moran-Ellis et al., 2006), the continuing tension between positivist and interpretative paradigms within that research design pose problems of understanding and reporting. Not so much 'paradigm peace' as 'paradigm truce'.

Hong et al. suggest that the development of interventions 'is both an art and a science' (Hong et al., 2005: 9). And Professor Sir Michael Rawlins, the chair of the UK's National Institute of Clinical Excellence (NICE), which summarizes the evidence-base for clinical interventions and advises National Health Service managers and practitioners on services effectiveness, has recently argued that in health research 'hierarchies of evidence should be replaced by

accepting – indeed embracing – a diversity of approaches' (Rawlins, 2008: 34). Process evaluations, including qualitative data, can and should be part of that diversity, but it is important that process evaluations are not oversold: they are not a miracle ingredient.

ACKNOWLEDGEMENTS

This study was funded by the Medical Research Council (G0601006) with additional support from the local Alcohol & Drugs Action Team and ASH Scotland. We wish to thank the Project Advisory Group: Lesley Armitage, David Craig, Emma Cepok, Louise Kane, Brian Pringle, Maria Reid and Mary Turley. We also wish to thank fellow project worker Sarah Welsh, and grantholders Rona Campbell, Candace Currie, James McIntosh, Neil McKeganey and Laurence Moore. We are deeply grateful to the staff and pupils of the participating schools for their help and participation. An earlier version of this paper was presented at the BSA Medical Sociology conference in Manchester 05/09/09.

REFERENCES

Atkinson, P., Coffey, A., Delamont, S., Lofland, J. and Lofland, L. (eds) (2001) *Handbook of Ethnography*. London & Thousand Oaks, CA: Sage.

Audrey, S., Holliday, J., Parry-Langdon, N. and Campbell, R. (2006a) 'Meeting the Challenges of Implementing Process Evaluation within Randomised Controlled Trials: The Example of ASSIST (A Stop Smoking in Schools Trial)', *Health Education Research* 21: 366–77.

Audrey, S., Holliday, J. and Campbell, R. (2006b) 'It's Good to Talk: An Adolescent Perspective of Talking to their Friends about Being Smoke-Free', *Social Science and Medicine* 63: 320–34.

Barbour, R. (1999) 'The Case for Combining Qualitative and Quantitative Approaches in Health Services Research', *Journal of Health Services Research & Policy* 4: 39–43.

Bloor, M. (1978) 'On the Analysis of Observational Data: A Discussion of the Worth and Uses of Inductive Techniques and Respondent Validation', *Sociology* 12: 545–52. Reprinted in A. Bryman and R. Burgess (1999) (eds) *Qualitative Research*. London: Sage. And reprinted also in P. Johnson and M. Clark (eds) (2006) *Business and Management Research*. London: Sage.

Bloor, M. (1997) 'Techniques of Validation in Qualitative Research: A Critical Commentary', in G. Millar and R. Dingwall (eds) *Strategic Qualitative Research*. London: Sage. Reprinted in R. Emerson (ed.) (2001) *Contemporary Field Research: Perspectives and Formulations*. Prospect Heights, IL: Waveland.

Bloor, M., Frankland, J., Thomas, M. and Robson, K. (2001) *Focus Groups in Social Research*. London: Sage.

Bonnell, C., Oakley, A., Hargreaves, J., Strange, V. and Rees, R. (2006) 'Assessment of Generalisability in Trials of Health Interventions: Suggested Framework and Systematic Review', *British Medical Journal* 333: 346–9.

Bradley, F., Wiles, R., Kinmonth, A-L., Mant, D. and Gantley, M. (1999) 'Development and Evaluation of Complex Interventions in Health Services Research: Case Study of the Southampton Heart Integrated Care Project (SHIP)', *British Medical Journal* 318(13): 711–15.

Bryman, A. (2006) 'Paradigm Peace and the Implications for Quality', *International Journal of Social Research Methodology* 9: 111–26.

Campbell, R., Starkey, F., Holliday, J., Audrey, S., Bloor, M., Parry-Langdon, N., Hughes, R. and Moore, L. (2008) 'An Informal School-based Peer-Led Intervention for Smoking Prevention in Adolescence (ASSIST): A Cluster Randomised Trial', *Lancet* 371: 1595–602.

Douglas, M. and Calvez, M. (1990) 'The Self as Risk-Taker: A Cultural Theory of Contagion in Relation to AIDS', *The Sociological Review* 38: 445–64.

Giannakaki, M.-S. (2005) 'Using Mixed-Methods to Examine Teachers' Attitudes to Educational Change: The Case of the Skills for Life Strategy for Improving Adult Literacy and Numeracy Skills in England', *Educational Research and Evaluation* 11(4): 323–48.

Gubrium, J. and Holstein, J. (eds) (2002) *The Handbook of Interview Research*. Thousand Oaks, CA: Sage.

Hong, Y., Mitchell, S.G., Peterson, J.A., Latkin, C.A., Tobin, K. and Gann, D. (2005) 'Ethnographic Process Evaluation: Piloting an HIV Prevention Intervention Program among Injecting Drug Users', *International Journal of Qualitative Methods* 4: 1–10.

Linnan, L., and Steckler, A. (2002) 'Process Evaluation for Public Health Interventions and Research – An Overview', in A. Steckler and L. Linnan (eds) *Process Evaluation for Public Health Interventions and Research.* San Francisco, CA: Jossey Bass.

McGraw, S.A., Stone, E.J., Osganian, S.K., Elder, J.P., Perry, C.L., Johnson, C.C., Parcel, G.S., Webber, L.S. and Luepker, R.V. (1994) 'Design of Process Evaluation within the Child and Adolescent Trial for Cardiovascular Health (CATCH)', *Health Education Quarterly* Suppl. 2: S5–26.

Maxwell, C., Kinver, A. and Phels, A. (2007) *Scottish Schools Adolescent Lifestyle and Substance Use Survey: Smoking, Drinking and Drug Use among 13 and 15 Year-Olds in Scotland in 2006*. URL: http://www.drugmisuse.isdscotland.org/publications/local/SALSUS_2006.pdf

Maynard, A. and Chambers, R. (1997) *Non-Random Reflections on Health Services Research*. London: BMJ Publishing Group.

Medical Research Council (n.d.) Trial Grant Annex. URL (consulted 10 February 2009): http://hwww.mrc.ac.uk/consumption/groups/public/documents/content/mrc001738.pdf

Moran-Ellis, J., Alexander, V., Cronin, A., Dickinson, M., Fielding, J., Sleney, J. and Thomas, H. (2006) 'Triangulation and Integration: Processes, Claims and Implications', *Qualitative Research* 61: 45–59.

Munro, A. and Bloor, M. (2009) 'A Feasibility Study for a Schools-Based, Peer-Led, Drugs Prevention Programme, Based on the ASSIST Programme: The Results', Centre for Drug Misuse Research occasional paper. Glasgow: University of Glasgow.

Oakley, A. (2005) 'Design and Analysis of Social Intervention Studies in Health', in A. Bowling and S. Ebrahim (eds) *Handbook of Health Research Methods: Investigation, Measurement and Analysis*, pp. 73–81. Maidenhead: Open University Press.

Oakley, A., Strange, V., Bonell, C., Allen, E., Stephenson, J. and the RIPPLE Study Team (2006) 'Process Evaluation in Randomised Controlled Trials of Complex Interventions', *British Medical Journal* 332: 413–6.

O'Cathain, A. (2009) 'Editorial: Mixed Methods Research in the Health Sciences: A Quiet Revolution', *Journal of Mixed Methods Research* 3(1): 3–6.

Palinka, L., Atkins, C., Miller, C. and Ferreira, D. (1996) 'Social Skills Training for Drug Prevention in High-Risk Female Adolescents', *Preventive Medicine* 25: 692–701.

Parry-Langdon, N., Bloor, M., Audrey, S. and Holliday, J. (2003) 'Process Evaluation of Health Promotion Interventions', *Policy & Politics* 31(Suppl.): S25–S34.

Rawlins, M. (2008) *De Testimonio – on the Evidence for Decisions about the Use of Therapeutic Interventions: The Harveian Oration of 2008*. London: Royal College of Physicians.

Victora, C., Habicht, J-P. and Bryce, J. (2004) 'Evidence-Based Public Health: Moving Beyond Randomized Trials', *American Journal of Public Health* 94(3): 400–5.

Wight, D. and Obasi, A. (2002) 'Unpacking the Black Box: The Importance of Process Data to Explain Outcomes', in J. Stephenson, J. Imrie and C. Bonell (eds) *Effective Sexual Health Interventions: Issues in Experimental Evaluation*, pp. 151–66. Oxford: Oxford University Press.

Young, D.R., Steckler, A., Cohen, S., Pratt, C., Felton, G., Moe, S.G., Pickrel, J., Johnson, C.C., Grieser, M., Lytle, L.A., Lee, J.S. and Raburn, B. (2008) 'Process Evaluation Results from a School- and Community-Linked Intervention: The Trial of Activity for Adolescent Girls (TAAG)', *Health Education Research* 23(6): 976–86.

ALISON MUNRO is a research fellow at Glasgow Caledonian University (School of Health). She has a PhD in the area of alcohol studies, and has conducted a number of research studies related to alcohol and drug misuse. *Address*: School of Health, Glasgow Caledonian University, Cowcaddens Rd, Glasgow G4 0BA, UK. [email: alison.munro@gcal.ac.uk]

MICHAEL BLOOR is a medical sociologist and is a part-time professorial research fellow at the University of Glasgow (Centre for Drug Misuse Research) and Cardiff University (Seafarers International Research Centre). *Address*: Centre for Drug Misuse Research, University of Glasgow, 89 Dumbarton Rd, Glasgow G11 6PW, UK. [email: m.bloor@socsci.gla.ac.uk]