

Analysis and comment

Health services research

Process evaluation in randomised controlled trials of complex interventions

Ann Oakley, Vicki Strange, Chris Bonell, Elizabeth Allen, Judith Stephenson, RIPPLE Study Team

Most randomised controlled trials focus on outcomes, not on the processes involved in implementing an intervention. Using an example from school based health promotion, this paper argues that including a process evaluation would improve the science of many randomised controlled trials

“Complex interventions” are health service interventions that are not drugs or surgical procedures, but have many potential “active ingredients.”¹ A complex intervention combines different components in a whole that is more than the sum of its parts.² Randomised controlled trials (RCTs) are the most rigorous way to evaluate the effectiveness of interventions, regardless of their complexity. Because of their multifaceted nature and dependence on social context, complex interventions pose methodological challenges, and require adaptations to the standard design of such trials.³ This paper outlines a framework for using process evaluation as an integral element of RCTs. It draws on experience from a cluster randomised trial of peer led sex education.

Why evaluate processes?

Conventional RCTs evaluate the effects of interventions on prespecified health outcomes. The main question is, “Does it work?” Process evaluations within trials explore the implementation, receipt, and setting of an intervention and help in the interpretation of the outcome results. They may aim to examine the views of participants on the intervention; study how the intervention is implemented; distinguish between components of the intervention; investigate contextual factors that affect an intervention; monitor dose to assess the reach of the intervention; and study the way effects vary in subgroups.⁴ Process evaluation can help to distinguish between interventions that are inherently faulty (failure of intervention concept or theory) and those that are badly delivered (implementation failure).⁵ Process evaluations are especially necessary in multisite trials, where the “same” intervention may be implemented and received in different ways.

Although “process” and “qualitative” are often used interchangeably, data for process evaluation can be both quantitative and qualitative. For example, in a series of systematic reviews of research on health promotion and young people, 33 (62%) of 53 process evaluations collected only quantitative data, 11 (21%)



Baby talk: the most effective sex education?

collected only qualitative data, and nine (17%) collected both.⁶ Process evaluation in RCTs is common in health promotion and public health, but often data are not collected systematically and are used mostly for illustration.⁵

Process evaluation in the RIPPLE study

The RIPPLE (randomised intervention of pupil peer led sex education) study is a cluster RCT designed to investigate whether peer delivered sex education is more effective than teacher delivered sessions at decreasing risky sexual behaviour. It involves 27 English secondary schools and follow-up to age 19. In the schools randomised to the experimental arm, pupils aged 16-17 years (given a brief training by an external team) delivered the programme to two successive year cohorts of 13-14 year olds. Control schools continued with their usual teacher led sessions. The trial was informed by a systematic review and pilot study in four schools.^{7 8} Its design by a multidisciplinary research team includes an integral process evaluation that has four aims: to document the implementation of

Social Science Research Unit, Institute of Education, University of London, London WC1H 0NR

Ann Oakley
professor of sociology and social policy

Vicki Strange
research officer

Public and Environmental Health Research Unit, London School of Hygiene and Tropical Medicine, London WC1E 7HT

Chris Bonell
senior lecturer in sociology and epidemiology

Centre for Sexual Health and HIV Research, London WC1E 6AU

Elizabeth Allen
statistician

Judith Stephenson
senior lecturer

Correspondence to:
A Oakley
a.oakley@ioe.ac.uk

BMJ 2006;332:413-6

Box 1: Methods for data collection in the RIPPLE trial process evaluation

- Questionnaire surveys of students and peer educators
- Focus groups with students and peer educators
- Interviews with teachers
- Researcher observation of peer led and teacher led sex education

the peer led intervention and sex education in control schools; to describe and compare processes in the two forms of sex education; to collect information from study participants (schools and students) about the experience of taking part in the trial; and to collect data on individual school contexts.⁹

Several methods were used to collect process data (box 1), including questionnaire surveys, focus groups, interviews, researcher observations, and structured field notes. Some methods, such as questionnaire surveys, were also used to collect outcome data. Other methods, such as focus groups and interviews with school staff and peer educators, were specific to process evaluation.

The outcome results by age 16 showed that the peer led approach improved some knowledge outcomes; increased satisfaction with sex education; and reduced intercourse and increased confidence about the use of condoms in girls. Girls in the peer led arm reported lower confidence about refusing unwanted sexual activity (borderline significance). The incidence of intercourse before age 16 in boys and of unprotected first sex, regretted first intercourse (or other quality measure of sexual experiences or relationships), confidence about discussing sex or contraception, and some knowledge outcomes for both girls and boys were not affected.¹⁰

We analysed process data in two stages. Before analysing outcome data (box 2), we analysed the data to answer questions arising from the aims outlined above. The hypotheses generated were tested in statistical analyses that integrated process and outcome data to address three questions:

- What was the relation between trial outcomes and variation in the extent and quality of the implementation of the intervention?
- What processes might mediate the observed relation between intervention and outcomes?
- Did subgroups of students and schools differ in their responses to the intervention?

We used several strategies to combine the different types of data. These included on-treatment analyses, in which results for students who received the peer led intervention were compared with results from the standard intention to treat approach (where allocation to, rather than receipt of, the intervention forms the basis of the analysis). We then carried out regression analyses and, where appropriate, tests for interactions, to examine the relation between key dimensions of sex education, subgroups of schools and students most and least likely to benefit from the peer led programme, and study outcomes (further details are available elsewhere).^{11 12} We used regression analyses because they are best for dealing with many types of

outcome, assessing the impact of mediating factors, and analysing subgroups.^{13 14}

These strategies provided answers to our three questions.

- More consistent implementation of the peer led programme might have had a greater impact on several knowledge outcomes and reduced the proportion of boys having sex by age 16, but it would not have changed other behavioural outcomes.
- There were key interactions between the extent to which sex education is participative and skills based and who provided it: when sex education was participative and skills based, the peer led intervention was more effective, but when these methods were not used, teacher led education was more effective.
- Peer led sex education was less good at engaging students most at risk of poor sexual health (risk was assessed using housing tenure, attitude to school, and educational aspirations); the peer led approach was better at increasing knowledge in schools serving “medium” rather than “low” risk populations (assessed using the proportion of students receiving free school meals).

Integrating process and outcome evaluation

Process evaluation in the RIPPLE trial is an example of the trend to move beyond the rhetoric of quantitative versus qualitative methods. By integrating process and outcome data, we maximised our ability to interpret results according to empirical evidence. Importantly, an on-treatment approach to outcomes using process data made little difference to the results. Using process data to identify key dimensions of sex education and examining these in relation to the trial arm revealed the circumstances in which peer led sex education was most effective, as did analysis of risk for both individual schools and students. The conclusion that the peer led approach is not a straightforward answer to the problem of school disengagement and social exclusion is an important policy message.

Other recent trials of complex interventions in which integral process evaluations helped explain the outcome findings include a trial of peer education for homosexual men that had no apparent impact on HIV risk behaviour; the process evaluation established that the intervention was largely unacceptable to the peer educators recruited to deliver it.¹⁵ In trials of day care for preschool children and of supportive health visiting for new mothers, integral process evaluations revealed unpredicted economic effects of day care and

Box 2: Methodological issues of integrating process evaluation within trials

- Process data should be collected from all intervention and control sites
- Data should be both qualitative and quantitative
- Process data should be analysed before outcome data to avoid bias in interpretation
- Steps should be taken to minimise the possibility of bias and error in interpreting the findings from statistical approaches, such as on-treatment analyses and regression and subgroup analyses

the reasons for low uptake of health visiting in a culturally diverse population.^{16 17} In a trial of a secondary prevention programme for cardiovascular disease, interviews and focus groups showed that because patients viewed heart attacks as self limited episodes, they were less willing to adopt long term lifestyle changes, and that the ability of practice nurses to provide skilled continuity of care was affected by their lack of training and low status in the primary health care team.¹⁸ Such trials involve issues of service and organisation and the frameworks of understanding that inform the behaviour of healthcare users—features of the interactions between intervention and environment that characterise complex interventions.¹⁹

“Good” randomised controlled trials have been defined as those having a high quality intervention, adequate evaluation of the intervention and its delivery, documentation of external factors that may influence outcome, and a culturally sensitive intervention.²⁰ Thus, many RCTs would be enhanced by an integral process evaluation. The additional costs (such as collecting and analysing qualitative data) would probably be balanced by greater explanatory power and understanding of the generalisability of the intervention.²¹ Complexity as a quality of interventions is a matter of degree, not an absolute. The argument that process evaluation is most useful in cluster trials, and where the intervention is non-standardised, also applies to many “pragmatic” clinical RCTs,²² as does the idea that process evaluation in feasibility studies is crucial to developing appropriate and effective interventions.²³

A “social” model of RCTs?

The move towards combining process and outcome evaluation builds on other, related methodological developments. For example, more attention is being paid to the need for pilot and consultation work in the development phase of RCTs,²³ the importance of a more theory based approach to evaluation,²⁴ and the modification of intervention effects by context in community intervention trials.²⁵ Statistical methods to integrate process and outcome data, such as those developed in the RIPPLE study, are a move forward.

Some of the methods we suggest—such as allowing hypotheses about the effectiveness of interventions to be derived from process data collected during a trial, and drawing conclusions (with reservations) from on-treatment and subgroup analyses—go against the conventions of clinical RCTs. Perhaps the effort devoted to developing “how to do it” guidelines for clinical trials could be matched by a similar effort in the field of complex interventions.

Contributors: The RIPPLE study group includes the authors and Abdel Babiker, Andrew Copas, and Anne Johnson. The RIPPLE study was designed by AJ, AO, and JS. JS is principal investigator and is responsible for the overall conduct of the trial. The process evaluation is led by AO and VS. VS, S Forrest, and S Charleston developed methods for collecting and integrating qualitative and quantitative data, and coordinated data collection. EA and AC were the trial statisticians; S Black, M Ali, and AB contributed to statistical analyses. AJ contributed to quantitative data collection methodology. CB led data analyses related to social exclusion and sexual health. AO was the guarantor.

Funding: Medical Research Council.

Summary points

A detailed process evaluation should be integral to the design of many randomised controlled trials

Process evaluations should specify prospectively a set of process research questions and identify the processes to be studied, the methods to be used, and procedures for integrating process and outcome data

Expanding models of evaluation to embed process evaluations more securely in the design of randomised controlled trials is important to improve the science of testing approaches to health improvement

It is also crucial for persuading those who are sceptical about using randomised controlled trials to evaluate complex interventions not to discard them in favour of non-randomised or non-experimental studies

Competing interests: None declared.

Ethical approval: Committee on ethics of human research, University College, London.

- Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, et al. Framework for design and evaluation of complex interventions to improve health. *BMJ* 2000;321:694-6.
- Hawe P, Shiell A, Riley T. Complex interventions: how “out of control” can a randomised controlled trial be? *BMJ* 2004;328:1561-3.
- Gueron JM. The politics of random assignment: implementing studies and affecting policy. In: Mosteller F, Boruch R, eds. *Evidence matters: randomized trials in education research*. Washington, DC: Brookings Institution Press, 2002:15-49.
- Wight D, Obasi A. Unpacking the “black box”: the importance of process data to explain outcomes. In: Stephenson J, Imrie J, Bonell C, eds. *Effective sexual health interventions: issues in experimental evaluation*. Oxford: Oxford University Press, 2002:151-66.
- Rychetnik L, Frommer M, Hawe P, Shiell A. Criteria for evaluating evidence on public health interventions. *J Epidemiol Community Health* 2002;56:119-27.
- Oakley A. Design and analysis of social intervention studies in health. In: Bowling A, Ebrahim S, eds. *Handbook of health research methods*. Maidenhead: Open University Press (in press).
- Harden A, Oakley A, Oliver S. Peer-delivered health promotion for young people: a systematic review of different study designs. *Health Educ J* 2001;60:363-70.
- Stephenson JM, Oakley A, Charleston S, Brodala A, Fenton K, Petrukevitch A, et al. Behavioural intervention trials for HIV/STD prevention in schools: are they feasible? *Sex Transm Infect* 1998;74:405-8.
- Stephenson JM, Oakley A, Johnson AM, Forrest S, Strange V, Charleston S, et al. A school-based randomized controlled trial of peer-led sex education in England. *Control Clin Trials* 2003;24:643-57.
- Stephenson JM, Strange V, Forrest S, Oakley A, Copas A, Allen E, et al. Pupil-led sex education in England (RIPPLE study): cluster-randomised intervention trial. *Lancet* 2004;364:338-46.
- Oakley A, Strange V, Stephenson J, Forrest S, Monteiro H. Evaluating processes: a case study of a randomised controlled trial of sex education. *Evaluation* 2004;10:440-62.
- Strange V, Allen E, Oakley A, Stephenson J, Johnson A. Integrating process and outcome data in a randomised controlled trial of sex education. *Evaluation* (in press).
- Victora CG, Huttly S, Fuchs SC, Olinto MTA. The role of conceptual factors in epidemiological analysis. A hierarchical approach. *Int J Epidemiol* 1997;26:224-7.
- Oxman AD, Gordon H, Guyatt A. Consumers’ guide to subgroup analyses. *Ann Intern Med* 1992;116:78-84.
- Elford J, Sherr L, Bolding G, Serie F, Maguire M. Peer-led HIV prevention among gay men in London: process evaluation. *AIDS Care* 2002;14:351-60.
- Toroyan T, Oakley A, Laing G, Roberts I, Mugford M, Turner J. The impact of day care on socially disadvantaged families: an example of the use of process evaluation within a randomised controlled trial. *Child Care Health Dev* 2004;30:691-8.
- Wiggins M, Oakley A, Roberts I, Turner H, Rajan L, Austerberry H, et al. The social support and family health study: a randomised controlled trial and economic evaluation of two alternative forms of postnatal support for mothers living in disadvantaged inner city areas. *Health Technology Assessment Monograph* 8:32.

- 18 Bradley F, Wiles R, Kinmonth A-L, Mant D, Gantley M. Development and evaluation of complex interventions in health services research: case study of the Southampton heart integrated care programme (SHIP). *BMJ* 1999;318:711-5.
- 19 Wolf N. Randomised trials of socially complex interventions: promise or peril? *J Health Serv Res Policy* 2001;6:123-6.
- 20 Ross DA, Wight D. The role of randomized controlled trials in assessing sexual health interventions. In: Stephenson J, Imrie J, Bonell C, eds. *Effective sexual health interventions: issues in experimental evaluation*. Oxford: Oxford University Press, 2002:35-48.
- 21 Helitzer DL, Davis SM, Gittelsohn J, Going SB, Murray DM, Snyder P, et al. Process evaluation in a multisite primary obesity-prevention trial in American Indian schoolchildren. *Am J Clin Nutr* 1999;69:816S-24S.
- 22 Hotopf M. The pragmatic randomised controlled trial. *Adv Psychiatr Treat* 2002;8:326-33.
- 23 Power R, Langhaug LF, Nyamurera, Wilson D, Bassett MT, Cowan FM. Developing complex interventions for rigorous evaluation—a case study from rural Zimbabwe. *Health Educ Res* 2004;19:570-5.
- 24 Stame N. Theory-based evaluation and varieties of complexity. *Evaluation* 2002;10:58-76.
- 25 Hawe, P, Shiell A, Riley T, Gold L. Methods for exploring implementation variation and local context within a cluster randomised community intervention trial. *J Epidemiol Community Health* 2004;58:788-93.

(Accepted 14 October 2005)

Drugs

Lessons for clinical trials from natalizumab in multiple sclerosis

Abhijit Chaudhuri

The approval of natalizumab and its recall after three months raises questions about the fast tracking of new drugs by the Food and Drug Administration for commercial licensing

Essex Centre for Neurological Sciences, Oldchurch Hospital, Romford RM7 0BE
Abhijit Chaudhuri
consultant neurologist
chaudhuri@gmail.com

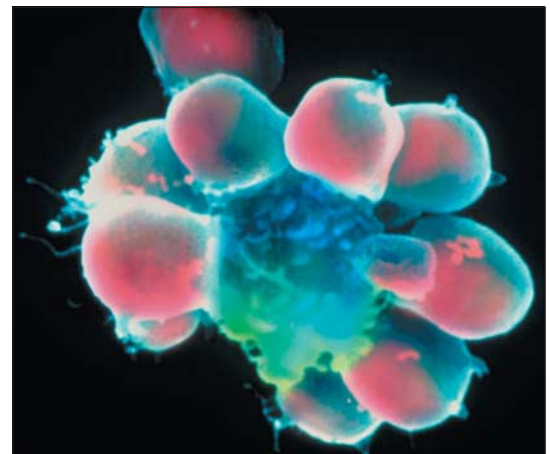
BMJ 2006;332:416-9

On 28 February 2005 Biogen Idec and Elan voluntarily suspended marketing natalizumab (Tysabri or Antegren) for clinical use because two patients with multiple sclerosis developed progressive multifocal leucoencephalopathy (PML) while being treated. Clinicians were advised to suspend all ongoing trials, and commercial distribution of the drug was halted. Three months earlier the US Food and Drug Administration gave natalizumab accelerated approval to treat relapsing multiple sclerosis. Approval was given on the basis of short term (one year) data from two multi-centre, randomised double blind placebo controlled phase 3 trials. Neither trial was published in a peer reviewed journal and the FDA granted approval before final trial and cumulative safety data were available. PML has been confirmed in three patients taking natalizumab.¹⁻³

The unpublished multiple sclerosis trials

Natalizumab is a humanised monoclonal antibody to $\alpha 4$ integrin, which plays a key role in the adhesion and migration of immunocompetent T cells through its interaction with endothelial selective adhesion molecule.⁴ Approximately 3000 patients, mostly with multiple sclerosis and Crohn's disease, were treated with natalizumab in clinical trials, and nearly 5000 patients have been treated in the United States since it became commercially available in 2004. In the United Kingdom, natalizumab was due for appraisal by the National Institute for Health and Clinical Excellence in 2006 for use in multiple sclerosis.

In the two studies that formed the basis of its approval by the FDA, natalizumab was given intravenously every four weeks to patients with multiple sclerosis who had experienced at least one clinical relapse during the preceding year. The primary end point of each study was the annualised relapse rate at one year. In the first trial (the AFFIRM trial) patients were randomised 2:1 to receive natalizumab (n=627) or placebo (n=315). In the second study (the SENTINEL



T cell attacking a cluster of foreign red blood cells: natalizumab prevents the migration of immunocompetent T cells across biological barriers and suppresses T cell mediated immune responses

trial) patients had experienced at least one relapse, despite treatment with interferon beta-1a (Avonex; Biogen Idec). Patients were randomised to receive natalizumab (n=589) or placebo (n=582) in addition to intramuscular injections of interferon beta-1a. In the first study, patients receiving natalizumab had a relapse rate of 0.25 relapses per patient year, compared with 0.74 in the placebo group (66% relative reduction of relapses). In the second study, patients taking natalizumab had 0.36 relapses per patient year compared with 0.78 in the placebo group (54% relative reduction of relapses). The FDA concluded that natalizumab was superior to all available treatments for relapsing multiple sclerosis (three types of interferon beta and glatiramer).⁵

Safety data were available to the FDA for 1617 patients treated for multiple sclerosis in both controlled and uncontrolled studies.⁵ The median exposure time to the drug was 20 months and the most frequent serious adverse events were infection, hypersensitivity reactions, and depression.⁵