RESEARCH METHODS & REPORTING

Evaluating service delivery interventions to enhance patient safety

Celia Brown, Richard Lilford

Celia Brown and **Richard Lilford** describe the recommendations of an MRC sponsored network to improve the quality of research into patient safety

Department of Public Health and Epidemiology, University of Birmingham, Edgbaston, Birmingham B15 2TT

Correspondence to: C Brown c.a.brown@bham.ac.uk

Accepted: 7 October 2008

Cite this as: *BMJ* **2008;337:a2764** doi: 10.1136/bmj.a2764 The quantity of research into patient safety has risen substantially over the past decade.¹ Nevertheless, concerns remain over the quality of much of this research² and there are disagreements about methods.³ As a result, the Medical Research Council sponsored a cross council research network in patient safety research to provide methodological guidance on evaluation of patient safety interventions. This resulted in publication of a series of articles in *Quality and Safety in Health Care.*⁴⁻⁷ Here we summarise two main themes developed in the series: study design and determining what observations should be made.

It is not possible to cleanly separate safety interventions from quality improvements.⁴ That said, the term safety is typically used in the context of rare incidents where there is a rapid and strong link between an error and its associated outcome. Our comments therefore apply to both safety and quality improvement but take into account the rarity of many safety incidents. Safety interventions are directed at the system in which care is delivered. They are thus service delivery interventions not new health technologies. Such interventions are often complex and should be evaluated before implementation



Fig 1 | Mean scores for clinical quality at practice level for coronary heart disease, asthma, and type 2 diabetes, 1998 to 2005. Reproduced with permission from Campbell et al¹⁰

(alpha testing), as advocated by both the MRC⁸ and the tenets of safety science. The methods of such evaluation are discussed elsewhere.⁴ However, even the most careful evaluation before implementation is no substitute for evaluation once interventions are rolled out in practice, and it is with these "in practice" evaluations that this article is concerned.

Study design

Before and after studies without concurrent controls

Much quality and safety improvement research involves before and after studies in single institutions. Such studies may provide convincing evidence of effectiveness, particularly when interventions have large effects, as in the Michigan based evaluation of a multifaceted intervention to reduce central line infections in intensive care.⁹

If the effects are small, these studies are a relatively weak method to distinguish cause and effect, since any observed change might plausibly be attributed to secular trends (for example, due to other service developments) or regression to the mean. Nevertheless, this design may be the only feasible option in some circumstances-for example, when service managers wish to evaluate a local initiative, policy makers introduce an intervention across an entire service, or safety incidents are very rare. If possible, a series of observations (time series, control chart) should be used for before and after studies, since a significant interruption in this series will be better evidence of cause and effect than differences in single before and after observations. Such an approach was used to track the quality of primary care in England before and after the introduction of a system of financial incentives in 2003 and showed that quality was already improving before the intervention (figure 1).¹⁰ However, we cannot be sure what would have happened had the intervention not been put in place as such inference requires concurrent controls.

Studies with concurrent controls

The interaction of two design variables (the timing of data collection and whether allocation to control or intervention is randomised or not) generates four types of controlled study design. The table shows the potential effect of each design on study quality. Controlled before

Advantages and disadvantages o	of controlled study designs	
	Allocation	
Timing of data collection	Randomised	Non-randomised (natural experiment)
After intervention (cross sectional)	Any baseline differences (more likely with small trials) could bias the results	Risk that comparisons will be confounded by differences between departments or organisations at baseline
Before and after	Allows for specific comparison of change net of any baseline differences. Enables comparisons to be made between sites that change most or least	Controls for baseline differences possible; rates of change are less confounded than cross sectional data

and after designs are much stronger than study cross sectional (post-intervention only) designs.⁵ The chance of bias is reduced because comparisons can control for any differences at baseline. This is particularly important if the study is not randomised. In addition, most service delivery interventions have to be compared at the organisational level (hospital, practice, ward) rather than the individual (patient, doctor) level. Before and after evaluations are less sensitive to the tendency for observations to correlate within their clusters than post-intervention comparisons.¹¹

The stepped wedge design is a type of before and after controlled study design that is useful for evaluating interventions aimed at patient safety and service delivery.¹² In this design, an intervention is rolled out sequentially to participants so that all participants have received the intervention by the end of the study (fig 2). This design may be particularly appropriate when the intervention cannot be implemented in all sites simultaneously or there is a strong prior belief that the intervention will do more good than harm. The evaluation of a critical care outreach service in York provides a good example of the use of a stepped wedge design.¹³ The Audit Commission and English Department of Health had recommended an outreach service, and hence it was assumed that the service would do more good than harm. However, the intervention could not be introduced simultaneously across the whole hospital because staff had to be trained in the new procedures. The intervention reduced inhospital mortality without affecting the length of stay.

Selecting the study design

In balancing the need for rigorous methods with the limited time and money available for research into patient



Fig 2 | The stepped wedge design in which the intervention is rolled out to individuals or clusters sequentially over the study period (from blank cells (control) to shaded cells (intervention)) safety, we propose that the choice of study design should be influenced by four key factors:

Target audience—Before and after designs are useful for studies that are part of local quality improvement projects (audits, "plan do act" cycles, or action research)

Logistical or practical constraints—When an intervention has already been implemented, concurrent controls are possible only if retrospective information can be obtained from case notes or computerised databases

Frequency of the safety problem—For rare problems, a study with concurrent controls is likely to be underpowered and a before and after study must suffice

Face validity, cost, and potential side effects of the intervention—The value of information from prospective controlled studies will turn on these variables, and it is more efficient to concentrate research resources on interventions whose effects are uncertain, that may plausibly cause harm or that are expensive.

Figure 3 shows an algorithm to help select study design. We have discussed the importance of baseline observations (ideally taken repeatedly before the intervention is put in place), whether or not concurrent controls can be used. It may be possible to construct this baseline retrospectively using routinely collected data from case notes or computer systems, but information specific to a study requires prospective planning and data collection. Thus commissioners of applied research need to liaise with those responsible for delivering new services so that the evaluation can be initiated before the service redesign is implemented. The National Institute for Health Research in England has established several collaborations where researchers work with local health services to evaluate service change prospectively.

Observations, measurement, and qualitative data Causal chain linking interventions to outcomes

Donabedian described a causal chain that links structure, process, and outcome.¹⁴ We have developed this chain to create a conceptual model of the healthcare system (fig 4). Like Donabedian, we start with the structure within which an organisation operates—national and regional health systems and the building blocks of care such as buildings and staff to patient ratios. Policy makers, rather than the managers or clinicians delivering health care in particular organisations, control these factors. Describing structure provides useful information on the context within which an intervention is being implemented.

Next in the chain are processes. We divide processes into two broad classes: managerial processes at the organisational level and clinical processes at the clinicianpatient interface. Managerial processes include human resource policies (such as a system of appraisal or development review for staff), how staff rotas are organised, and time allocated to continuing professional development. Interventions at the managerial level are intended to reduce the chance of latent errors—problems buried deep in the system. These generic interventions often

RESEARCH METHODS & REPORTING

Fig 3 | Framework for selection of study design



act through intervening variables such as staff morale, knowledge, and sickness absence.

Clinical processes cover the tenets of safe and evidence based care, such as washing hands between patients, maintaining normothermia under anaesthesia, and responding to signs of a deteriorating patient. Interventions aimed at clinical processes are intended to reduce active errors—for example, "forced function" engineering solutions to prevent anaesthetic tubes being misconnected or alarms built into equipment to alert staff to signs that a patient's condition is deteriorating. Such interventions may be expected to have a large effect on a small number of errors whereas management interventions will have a smaller effect across many clinical processes and outcomes.

At the end of the causal chain comes actual patient outcomes (including patient reported outcomes) and throughput (number and types of patients treated).

Using the framework in evaluation

The causal chain linking interventions to outcomes highlights the opportunities for making multiple observations as part of an evaluation:

• The effectiveness of an intervention should be observed at all points to the right of the intervention point

- Observations at the point of intervention should be used to assess the fidelity with which the intended intervention was implemented and to monitor how the intervention may have been adapted over time
- Observations to the left of the intervention point enable a description of the context within which an intervention has been implemented.

Making observations across the causal chain provides information on context that enables factors that affect the success or failure of an intervention to be identified. It also provides multiple measures of effectiveness. A given outcome (safety culture or morale) may be an effectiveness measure in one type of study (where the intervention affects management processes) while providing information on context in another (where the intervention pertains to clinical processes). We briefly consider observations at four levels in the chain below. Further discussion is available in our paper in *Quality and Safety in Health Care.*⁶

Patient outcomes

A focus on clinical outcomes (morbidity and mortality) may seem ideal, but as the signal to noise ratio tends to be low there is a high risk of false negative results.¹⁵ For example, Mant and Hicks report that even if all care standards were followed in some sites and none in

RESEARCH METHODS & REPORTING



Fig 4 | Causal chain linking interventions to outcomes. Observations can be made at all points in the chain to provide information on context, fidelity, and effectiveness and safety

others, this could account for only half of the observed differences in heart attack survival across sites.¹⁶ Patient reported outcomes are increasingly used but also seem relatively insensitive to most safety and service delivery interventions.¹⁷ We therefore advocate buttressing outcomes by observing surrogates for patient outcomes within the causal chain in figure 4:

Fidelity

Fidelity measures whether an intervention was implemented as planned and is a necessary, but not sufficient, condition to prove that the intervention has improved care. A positive result for fidelity therefore shows that positive results further down the causal chain are plausible, while a negative result can help to explain a null result further down the chain.

Intervening variables

Management interventions such as human resource policies that aim to strengthen an organisation generically often achieve their effects through intervening (mediating) variables such as knowledge, beliefs, fatigue, morale, and safety culture. As with fidelity, improvements in intervening variables do not prove that downstream clinical benefits will be realised: measurements of both intervening variables and clinical outcomes in the same study can, however, test whether the intervening variable is valid as a surrogate outcome and also help explain the findings. For example, Landrigan and colleagues found that reducing interns' working hours reduced both physiological measures of staff fatigue and clinical errors.¹⁸

Clinical process or error

We define error as the failure to apply the correct standard of care, the failure to carry out an action as intended, or the application of an incorrect plan.⁴ Collecting data on error can help improve the signal to noise ratio if errors are more common than the corresponding adverse outcome; medication error is an example.⁶ Error can be identified using reporting systems and trigger tools, but these methods are not suitable for calculating error rates, which are necessary to make inferences about effectiveness. Furthermore, the problems associated with case mix bias can be reduced (but not eliminated) by using the opportunity for error rather than the number of patients as the denominator for error rates.⁶ Error rates can be measured prospectively (such as in the UK myocardical ischaemia national audit project study of cardiac care) but are usually obtained by case note review.⁶

Assessment of case notes may be explicit (based on predefined criteria) or implicit (a holistic method where the reviewer uses clinical knowledge to assess the quality of care). Each method identifies a different spectrum of errors and hence the two methods may be regarded as complementary. Explicit methods are more reliable (repeatable), whereas implicit methods require more highly skilled and experienced reviewers.¹⁹ In comparative studies the reviewers should be independent of the organisations being compared to avoid observer bias.

Qualitative studies

Qualitative information in the form of interviews, perhaps augmented by observations of behaviour and social interactions, enable the subjective experiences of staff and patients to be explored. These qualitative observations should be made at different levels in the causal chain. Such qualitative data provide a more complete picture than quantitative data alone, explaining findings and contributing to theory. For example, the quantitative finding that maternity care became safer after dissemination of national evidence based guidelines was enriched by the results of qualitative interviews showing that change in practice was influenced by endorsement from influential clinicians rather than by local management initiatives.²⁰ Qualitative research may also identify new hazards introduced by interventions that can then be observed as part of the evaluation.

Putting it all together

We advocate mixed methods research and have provided a conceptual framework for systematic application of these methods across the causal chain. The results of evaluations using this framework should provide a rich picture to inform answers to questions such as: How well did the intervention work? What were the costs and side effects? Why did it work (or not)? What factors affected how well it worked? What theories can help to explain how it works and why it may work better in some places than others? In turn these questions inform judgments about the future: Should the intervention be rolled out? How should it be implemented? Should it be adapted? What are the next research questions? Moreover, the various observations at different points of the causal chain strengthen each other. If an intervention was implemented with high fidelity, resulted in positive changes in intervening variables, reduced errors, and improved outcomes, this tells a story even if the effect on outcomes itself was not statistically significant.

Integration (synthesis) of the diverse observations obtained by mixed methods research requires judgment, which is inevitably a subjective process. However the creation of scientific meaning is always subjective: extrapolating the results of a randomised controlled trial of a drug to a new place requires an inductive (and hence

SUMMARY POINTS

The choice of study design and determining what observations should be made are important when evaluating quality and safety interventions

A controlled before and after design (including the stepped wedge design) should be considered when summative evaluations are required

A modified form of Donabedian's causal chain (structure, process, and outcome) provides a framework for evaluation

Whenever possible, mixed methods (qualitative and quantitative) should be used to make observations at different levels across the causal chain

subjective) step. Subjective synthesis of diverse information is normally an entirely intuitive (and hence opaque) process, but it can be made more transparent by using bayesian statistical models.²¹

Applying an evidence based approach to service delivery interventions in general, and to patient safety interventions in particular, is not straightforward. However, scientific principles do not lose their relevance when we move out of the laboratory, and we hope we have shed some light on how these principles may be applied within the practical and logistical constraints imposed by real life settings.

We thank all the contributors to the patient safety network and in particular our coauthors on the four papers in *Quality and Safety in Health Care* on which this article is based. We would also like to thank Andrew Stevens, Alan Girling, and the reviewers for helpful comments and the NHS Research Methodology Programme for its support.

Contributors: RL conceived the idea for the MRC Patient Safety Network and led the work of the network. CB and RL drafted the article. RL is guarantor. **Funding:** The research on which this article is based was funded by an MRC cross-council network grant.

Competing interests: None declared.

Provenance and peer review: Not commissioned; externally peer reviewed.

- Lilford R, Stirling S, Maillard N. Citation classics in patient safety research: an invitation to contribute to an online bibliography. *Qual Saf Health Care* 2006;15:311-3.
- 2 Ovretveit J. Which interventions are effective for improving patient safety? A review of research evidence. Karolinska, Sweden: Karolinska Institute Medical Management Centre, 2005.

- 3 Leape LL, Berwick DM, Bates DW. What practices will most improve safety? Evidence-based medicine meets patient safety. JAMA 2002:288:501-7.
- Brown CA, Hofer T, Johal AJ, Thomson R, Nicholl J, Dean Franklin B, et al. An epistemology of patient safety research: a framework for study design and interpretation. 1. Conceptualising and developing interventions. *Qual Saf Health Care* 2008;17:158-62.
- 5 Brown CA, Hofer T, Johal AJ, Thomson R, Nicholl J, Dean Franklin B, et al. An epistemology of patient safety research: A framework for study design and interpretation. 2. Study design. *Qual Saf Health Care* 2008;17:163-9.
- 6 Brown CA, Hofer T, Johal AJ, Thomson R, Nicholl J, Dean Franklin B, et al. An epistemology of patient safety research: a framework for study design and interpretation. 3. End points and measurement. Qual Saf Health Care 2008;17:170-7.
- 7 Brown CA, Hofer T, Johal AJ, Thomson R, Nicholl J, Dean Franklin B, et al. An epistemology of patient safety research: a framework for study design and interpretation. 4. One size does not fit all. Qual Saf Health Care 2008;17:178-81.
- Medical Research Council. Developing and evaluating complex interventions: new guidance. London: MRC, 2008.
- 9 Pronovost P, Needham D, Berebholtz S, Sinopoli D, Chu H, Cosgrove S, et al. An intervention to decrease catheter-related bloodstream infections in the ICU. N Engl / Med 2006;355:2725-32.
- 10 Campbell S, Reeves D, Kontopantelis E, Middleton E, Sibbald B, Roland M. Quality of primary care in England with the introduction of pay for performance. N Engl / Med 2007;357:181-90.
- 11 Murray DM, Blitstein JL. Methods to reduce the impact of intraclass correlation in group-randomised trials. *Eval Rev* 2003;27:79-103.
- 12 Brown CA, Lilford RJ. The stepped wedge trial design: a systematic review. *BMC Med Res Methodol* 2006;6:54.
- 13 Priestley G, Watson W, Rashidian A, Mozley C, Russell D, Wilson J, et al. Introducing critical care outreach: a ward-randomised trial of phased introduction in a general hospital. *Intens Care Med* 2004;30:1398-404.
- 14 Donabedian A. Explorations in quality assessment and monitoring. In: Griffith JR, ed. The definition of quality and approaches to its assessment. Washington, DC: Health Adminstration Press, 1980:4-163.
- Lilford RJ, Brown CA, Nicholl J. Use of process measures to monitor the quality of clinical practice. *BMJ* 2007;335:648-50.
- 16 Mant J, Hicks N. Detecting differences in quality of care: The sensitivity of measures of process and outcome in treating acute myocardial infarction. *BMJ* 1995;311:793-6.
- 17 Lowe DB, Sharma AK, Leathley MJ. The CareFile project: a feasibility study to examine the effects of an individualised information booklet on patients after stroke. *Age Aging* 2007;36:83-9.
- 18 Landrigan CP, Rothschild JM, Cronin JW, Kaushal R, Burdick E, Katz JT, et al. Effect of reducing interns' work hours on serious medical errors in intensive care units. N Engl J Med 2004;351:1838-48.
- 19 Lilford RJ, Edwards A, Girling A, Hofer T, Di Tanna GL, Petty J, et al. Interrater reliability of case-note audit: a systematic review. J Health Service Res Policy 2007;12:173-80.
- 20 Wilson B, Thornton JG, Hewison J, Lilford RJ, Watt I, Braunholtz D, et al. The Leeds University maternity audit project. *Int J Qual Health Care* 2002;14:175-81.
- 21 Lilford RJ, Braunholtz D. The statistical basis on public policy: a paradigm shift is overdue. *BMJ* 1996;313:603-7.

Dangerous fresh fruit: un hommage to Monty Python

Limiting occupational needlestick and sharps injuries is a priority for infection control in hospitals. Percutaneous injuries are often caused by inappropriate disposal of sharps,¹ and full, blocked sharps bins are a common factor predisposing to such injuries. Blocking up of sharps bins is commonly the result of improper disposal of blunt items into the bins.

During an audit of waste disposal practices that lead to needlestick injuries at a hospital in Cape Town, I came across a sharps bin whose contents included a viciously sharp banana skin and an apple core whose stalk could cause a nasty eye injury were it to be disposed of in a normal bin. These decommissioned weapons provide a salient reminder that sharps bins are for needles and other sharps, not syringes, plastic wrappings, and the remains of people's lunch.

Marc Mendelson head, Division of Infectious Diseases and HIV Medicine, Groote Schuur Hospital, University of Cape Town, South Africa marc.mendelson@uct.ac.za



1 Centers for Disease Control and Prevention. Updated US Public Health Service guidelines for the management of occupational exposures to HIV and recommendations for post-exposure prophylaxis. *MMWR* 2005;54(No RR-9).

Cite this as: BMJ 2009;338:a1106