An epistemology of patient safety research: a framework for study design and interpretation. Part 4. One size does not fit all

C Brown, T Hofer, A Johal, R Thomson, J Nicholl, B D Franklin and R J Lilford

Qual. Saf. Health Care 2008;17;178-181
doi:10.1136/qshc.2007.023663

Updated information and services can be found at:
http://qshc.bmj.com/cgi/content/full/17/3/178

These include:

References
This article cites 14 articles, 7 of which can be accessed free at:
http://qshc.bmj.com/cgi/content/full/17/3/178#BIBL

4 online articles that cite this article can be accessed at:
http://qshc.bmj.com/cgi/content/full/17/3/178#otherarticles

Rapid responses
You can respond to this article at:
http://qshc.bmj.com/cgi/eletter-submit/17/3/178

Email alerting service
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article

Notes

To order reprints of this article go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to Quality and Safety in Health Care go to:
http://journals.bmj.com/subscriptions/
An epistemology of patient safety research: a framework for study design and interpretation.
Part 4. One size does not fit all

C Brown,¹ T Hofer,² A Johal,¹ R Thomson,³,⁴ J Nicholl,⁵ B D Franklin,⁶ R J Lilford¹

ABSTRACT
This is the final article in the series on the epistemology of patient safety research, and considers the selection of study design and end points during the planning of an evaluation. The key message of this series is that "one size does not fit all": the nature of the evaluation will depend on logistical and pragmatic constraints, a priori assessment of the probability of benefits and harms, the plausible scale of these effects and the target audience for the results. This paper also discusses the advantages of mixed method designs. The strength of any conclusions can be increased if different end points concur and the authors therefore advocate assessment of the effect of the intervention on different end points across the generic causal chain linking structure, process and outcome. The use of both qualitative and quantitative methods is also advocated to help explain findings, generate theory and help contextualise results. We show how a bayesian framework can be used to synthesise evidence from a number of different sources and why this approach may be particularly appropriate for the evaluation of patient safety interventions.

LOGISTIC/PRAGMATIC CONSTRAINTS
The nature of the patient safety problem and the intended ameliorative intervention may impose two constraints on the study design and end points that can be used during an evaluation as discussed below.

Rarity of the patient safety problem
Quantitative designs with contemporaneous controls are feasible for the evaluation of interventions aimed at common, rather than rare, incidents. It is simply logically difficult to organise such studies to evaluate the effects of interventions to prevent rare but fatal events such as inadvertent intra-thecal injections of vincristine, which have occurred only once every few years. However, concurrently controlled prospective designs are eminently feasible at the high-frequency end of the spectrum we described in Part 1 of this series. Techniques to improve adherence with clinical guidelines, for example, have been extensively evaluated by this method.¹⁻³ Rare events need to be studied with before and after studies. Even then, it may be years before an effect on the intended target can be measured with precision. In such cases it may be necessary to rely on service-level process end points such as fidelity and, where relevant, intervening variables. For example, the evaluation of the National Patient Safety Agency (NPSA) national initiative to avoid wrong site surgery was built around monitoring uptake of the prescribed procedures in randomly selected hospitals and the attitude of staff to such procedures.⁴

Timing of the introduction of the intervention
The timing of the introduction of the intervention will influence the study design that can be used. Before and after designs will only be possible where an evaluation can be planned prior to the implementation of the intervention, while concurrent controls will not be possible if an intervention is implemented simultaneously across an entire service. The problem of simultaneous intervention across an entire service can arise for many inter-related reasons. There may be a high risk of "contamination" even between clusters. Policy makers may simply be too unaware of the importance of evaluation or they may feel the political imperative is too strong, perhaps because they expect the intervention to work (ie, they lack equipoise). The British Prime Minister, Gordon Brown, has recently announced a programme of “deep cleaning” for UK hospitals in an effort to reduce hospital-acquired infection. Conversely, it

Building on the first three articles of this four-part series on the epistemology of research into the effects of patient safety interventions, we propose that selection of study design should be influenced by four considerations:

- logistical/pragmatic constraints imposed by the nature of the patient safety problem and the intended intervention;
- a priori assessment of the probability of benefit and harm;
- plausible effects on end points (in relative and absolute terms);
- the target audience for the results: who is the study intended to influence?

We discuss each of these four considerations in turn, identifying their effect on study design, before summarising our findings on “which size fits whom?” We then discuss mixed method research, particularly the advantages of using multiple end points and of combining qualitative and quantitative methods. Such an approach provides a method of combination (triangulation), contributes to the generation of theory and helps to contextualise findings. Our argument rests heavily on the “causal chain” shown in fig 2 of the first part of this series. The opportunities afforded by a bayesian approach to evidence synthesis are also discussed, as a method of incorporating the evidence obtained from the use of different study end points and/or methods of measurement.
has been argued that the effectiveness of such a programme is far from proven. We argued in Part 2 that policy makers might be persuaded to roll out the intervention, using the stepped wedge design for evaluation. Such a design could reconcile the political and scientific imperatives.

A PRIORI ASSESSMENT OF BENEFIT AND HARM

It is also important to consider the risks, costs and benefits of proposed interventions. Specifically, this implies identifying the extent of uncertainty surrounding the outcomes or consequences that might plausibly arise from an intervention, both clinical and financial. If it is anticipated, following careful preimplementation evaluation (PIE), that an intervention is inexpensive and unlikely to do material harm then a controlled comparative quantitative study may be a council of perfection. A study of the fidelity of uptake followed by a before and after comparison of error rates may be sufficient to quantify effectiveness crudely and identify factors that may impact on effectiveness. For expensive interventions or where potential harms may offset benefits, a more robust evaluation involving before and after measurements in both intervention and control groups will be necessary. The need for formal comparative studies can be quantified using value of information economic modelling (see below).

The potential effectiveness of an intervention (preferably one that has undergone PIE) can be captured explicitly and quantitatively in a bayesian prior (a distribution or “density” of prior probabilities, usually elicited from experts in the field).3 This prior reflects a degree of belief before obtaining direct comparative data. Direct data from head to head comparisons can then be used to “update” these priors to provide a posterior distribution of probabilities. This process of updating prior probabilities by means of Bayes Rule effectively answers the question: “Given what I believed before a study, what should I believe now?” The mid-point of a posterior probability density function can then be used to model the cost effectiveness of an intervention.7 The bayesian method thus provides a method to combine data of the same type (in a meta-analysis) and also provides a method to integrate different types of evidence. Such analysis can be contrasted with the dichotomisation of results from single studies as statistically significant/not significant that forms the mainstay of classical (frequentist) statistics.

For example, Roberts and colleagues8 undertook a bayesian synthesis of qualitative and quantitative evidence to identify factors affecting the uptake of childhood immunisations. The authors noted that relying on either qualitative or quantitative data alone could have led to inappropriate formulation of evidence-based policy. Bayesian methods are particularly useful in the world of policy and service delivery where conclusive, generalised, comparative studies are hard to come by, as is often the case in the field of patient safety. The idea of bayesian updating of belief, rather than hypothesis testing, is linked to the idea of triangulation discussed below.

If there are likely to be financial consequences arising from the intervention, if two or more interventions are being compared or if a patient safety intervention is being compared with another type of intervention, then an economic evaluation may be required. Methods for the economic evaluation of healthcare are detailed by Drummond.9 Economic evidence could be used to justify capital expenditure on a patient safety intervention, given the savings that could be made in the medium and long term. Ovetvreit10 provides examples of how research can be used to predict savings from interventions to improve the quality and safety of healthcare. Formal economic evaluations based on bayesian priors and explicit models can be used to model the cost effectiveness of both safety interventions and different designs to study the effectiveness of these interventions. These are referred to as value of information studies,11 which have hitherto been used for health technology assessments but are also applicable to service delivery.

PLAUSIBLE EFFECTS ON END POINTS

When we consider the costs of nationwide interventions to tackle rare but serious incidents we are typically looking for large effect sizes in relative terms. For example, a proposal to reduce by 25% the one or two cases of death from inadvertent bolus injection of potassium chloride concentrate that previously occurred each year in England would seem inadequate. One might hope that the NPSA directive to remove the concentrated form of potassium chloride from open storage on hospital wards should all but eliminate the problem. Hence, not only are contemporaneous comparative studies less feasible since they would inevitably be underpowered, there is also less need for them. A study of fidelity of uptake of the NPSA directive on potassium chloride concentrate along with a before and after census of reported incidents would be fit for purpose and such a study has been undertaken as part of the Patient Safety Research Portfolio.12

On the other hand, when common errors are the target of a focused (specific) intervention, then even small improvements in relative risk may be cost effective and hence it is important that improvements are accurately quantified. For example, improving by 25% the use of influenza vaccine among older people or detection of incipient deterioration in patients with pneumonia would result in important gains that have previously eluded many. Although the effects of interventions in the context of high frequency incidents are often modest in terms of relative effect (eg, 25% improved uptake), such practices may have large benefits at the population level (eg, harm prevented or number of lives saved). Indeed it is precisely in the area of these high frequency incidents that the greatest (utilitarian) population health gains lie. Furthermore it is also in such circumstances—those where relative effects are modest—where use of contemporaneous controls and masking of observers are most important. This follows from the disarmingly simple principle that results will be most distorted when the plausible magnitude of effect is small relative to potential bias.13

In conclusion, prospective comparative studies are important when incidents are frequent and minimisation of bias is crucial, given the more modest relative risk reductions that may be expected.

TARGET AUDIENCE

The level of evidence required to persuade clinicians and managers to persist with a local intervention is probably less than the level required to convince those in a completely different setting to adopt a new intervention.14 For example clinicians and managers in a particular locality may wish to evaluate an improvement strategy by means of a relatively inexpensive before and after study as suggested in the second part of this series. Such a study may provide sufficient information to justify pursuing or adjusting a course of action, but it may be quite insufficient to persuade an external audience who may require more rigorous evidence of cause and effect, as well as on the generalisability of results to other settings, where an intervention may not be implemented under the tight
controls employed during the primary evaluation. The distinction between evaluation to improve a local service versus scientific evaluation to convince outsiders has been made in more detail elsewhere.\textsuperscript{14, 15} The distinction is sometimes captured in the language of sufficing (doing the best with what is available) versus optimising (doing robust scientific studies that may shift international opinion and practice).\textsuperscript{16} In particular, an external audience will often want to see how intervention sites compared with control sites and this brings to the fore all the issues regarding precision and accuracy discussed in this series.

**WHICH SIZE FITS WHOM?**

An important corollary of the above argument is that there is “no one size fits all” methodology for the evaluation of patient safety interventions. Decisions regarding study design and end points should reflect the four considerations outlined above, with the aim of ensuring that the evaluation itself provides sufficient “burden of proof” for the intended audience. We discuss the use of mixed methods—in terms of what end points will be assessed and how data on these end points will be collected in the next section. Before doing so, we present a stylised framework for the selection of study design in fig 1. This framework considers the main factors influencing study design, although we acknowledge that the existence of intervention-specific or local constraints means that this framework does not provide “hard and fast rules” for the selection of the most appropriate study design.

**MIXED METHOD DESIGNS**

The use of mixed methods to evaluate complex interventions is not new, originating from Cook and Reichardt.\textsuperscript{17} By mixed method designs in the context of evaluations of patient safety interventions we mean that:

- different end points are measured across the causal chain (see fig 2 in Part 1 of this series);
- qualitative and quantitative observations are both made (i.e., different methods of measurement are employed).

There are four key advantages of mixed method designs,\textsuperscript{18} which are now discussed in turn:

- triangulation—use of different methods to get at the same underlying truth by seeking corroboration between methods;
- understanding—to elaborate and explain results;
- development—to generate theory and thereby guide generalisation and inform further studies;
- linked qualitative studies may provide evidence of problems with the intervention at an early stage that may lead to revisions of study protocols.

These aims link to the causal chain (fig 2 in Part 1 of this series) as follows:

- Measurement (and triangulation) of end points to the right of the intervention provides evidence of effectiveness.
- Measurement of fidelity of uptake of the components of the intervention, along with qualitative work, helps in understanding.
- Measurement of end points to the left of the intervention provides evidence on context, and this again helps in building understanding.

**Triangulation**

The term “triangulation” is borrowed from sociology (which in turn borrowed it from physics) but the concept is important in the philosophy of science. The conclusions drawn from research findings of one type are reinforced when they are corroborated by findings of a different type. Evolutionary theory was strengthened when the DNA record fitted well with the fossil record. Similarly the effect of fatigue on performance was strengthened by showing both that it results in electrophysiological changes (“mini-sleeps”) and propensity to make errors. Additional credence through triangulation will add to the strength and generalisability of findings. Alternatively, where the results from different methods conflict, triangulation may prevent over-hasty inferences that might have been made had results been obtained from measurement of a single end point.\textsuperscript{19}

Consider a diffuse intervention such as “management walkabouts”. If the intervention is implemented with fidelity, affects an intervening variable (such as morale) in a positive direction, reduces error and improves outcomes, then we will be more confident that we have identified a cause and effect relationship than if some of the end points were not in agreement. Even if changes in some of those end points fail to reach statistical significance, the overall picture may suggest a positive effect if all the results move in a positive direction. The risk of a false-negative study result is high for many safety interventions, particularly diffuse interventions where different end points are considered separately using frequentist (conventional) statistics.

---

**Figure 1 Framework for selection of study design.**

```
<table>
<thead>
<tr>
<th>Eg, wrong site surgery</th>
<th>Are error and harm both rare?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Before and after study, often at national level</td>
</tr>
<tr>
<td>No</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Eg, medication error</th>
<th>Can concurrent controls be used?</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>Before and after study</td>
</tr>
<tr>
<td>Yes</td>
<td>Is an experimental design possible/acceptable?</td>
</tr>
<tr>
<td>No</td>
<td>Natural experiment, allocation may be by cluster Consider stepped wedge design, as no sites will be excluded</td>
</tr>
<tr>
<td>Yes</td>
<td>RCT, allocation may be by cluster Consider stepped wedge design, enables effect of timing to be evaluated</td>
</tr>
</tbody>
</table>
```
We therefore advocate a bayesian approach, where all end points and plausibility of the intervention (based on PIE) can all be integrated in an explicit quantitative framework. A bayesian approach is incremental, yielding changing probability estimates on a continuous scale, and this seems much more suitable to patient safety and service delivery than frequentist methods which toggle the conclusions from positive to negative on a statistical threshold.6 20

Understanding
Oakley and colleagues21 consider how the integration of process and outcome data can also help in interpreting and elaborating results. Such analysis enables researchers to consider why an intervention has or has not been effective. For example, if a method such as a medicines reconciliation protocol was not implemented widely (ie, with low fidelity), with staff saying they were too busy to follow it, then this would explain a null result on error rates. Ideally, data of different types should be analysed independently, or at least those where the observers’ subjectivity may colour the results should be analysed in ignorance of more objective measures.21 An example of a mixed methods approach to both generate and explain findings is reported by Wilson and colleagues.22 The authors assessed the change in compliance with four evidence-based recommendations for safe perinatal care before and after the publication of authoritative guidance (itself based on randomised trial evidence). The quantitative audit was supported by interviews with 88 members of staff across 20 hospitals. Data from the interviews were used to explore differences in rates of compliance with the guidance, and showed that influential clinical leaders were the most important factor driving the improvement process.

Development of theory
Results arising from one research method can be used to guide research questions or designs in subsequent studies. For example, interviews could be used to identify the key issues to be investigated using quantitative methods, or alternatively the results of a quantitative analysis may highlight areas requiring indepth investigation using qualitative approaches.19

Revisions of study protocols
In a randomised controlled trial of a computerised patient decision aid on stroke prevention in patients with atrial fibrillation, one of three arms of a trial was discontinued after the videos of consultations and interviews with participants revealed a problem with application of the more complex decision aid with older patients. In the absence of supportive data from the observational study, the trial arm would have been much harder to discontinue.31

CONCLUSION
In the four parts of this series we have provided an analysis of methods used to evaluate patient safety interventions. We have considered the issues at some depth because we discern some impatience with formal scientific methods among some patient safety practitioners. If there is one over-arching theme to our work, it is that there is no simple formula that applies in all circumstances. Each situation must be considered carefully and judgement will always be required. This does not mean that anything goes. On the contrary, the detailed attention we have given to topics such as bias and measurement is intended to inform such judgement. We hope it is helpful.

Acknowledgements: We would like to acknowledge the support of the National Coordinating Centre for Research Methodology and the Patient Safety Research Programme. The authors would also like to acknowledge the contributions of attendees at the Network meetings and the helpful comments of the peer reviewers.

Competing interests: None.

Authors’ contributions: RL conceived the Network and formulated the first draft of the report and the current paper with assistance from AJ. CB contributed to subsequent drafts of the report and this paper. BDF, TH, RT and JN contributed to the Research Network and provided comments on drafts of the report and papers in their areas of expertise.

This work forms part of the output of a Cross-Council Research Network in Patient Safety Research funded by the Medical Research Council (Reference G0300370). More details of the Research Network can be found at: http://www.pcpoh.bham.ac.uk/publichealth/psrp/MRC.htm

REFERENCES