

Understanding what works—and why—in quality improvement: the need for theory-driven evaluation

KIERAN WALSH

Centre for Public Policy and Management, Manchester Business School, University of Manchester, Booth Street West, Manchester, M15 6PB, UK

Clinicians who are asked to participate in quality improvement programmes in healthcare organizations are often heard to ask for the evidence that they ‘work’. By that, they often mean they want randomized controlled trials, which show that accreditation, or credentialing, or criterion-based audit, or adverse event monitoring, or continuous quality improvement programmes, or whatever approach is being used cause meaningful and worthwhile improvements in the quality of care [1]. When they learn that there are relatively few experimental studies of quality improvement interventions [2], and those which do exist often show weak or moderate effects at best, this state of affairs is sometimes used to argue that it is not worthwhile investing time and effort in quality improvement. After all, the argument goes, we should not embark on using a new clinical intervention such as a drug or a surgical procedure without solid experimental evidence of its effectiveness, so why should we have a lower threshold for the adoption of organizational interventions like quality improvement programmes? Surely, they too should be proven to ‘work’ before they are adopted or implemented widely?

This point of view needs to be challenged. It is founded on a fundamental misunderstanding of the place of experimental methods in investigating and understanding complex social interventions, which is commonplace particularly among clinicians and biomedical researchers and which can seriously hamper those both researching and implementing quality improvement in healthcare.

In all research, our starting point should be to match our research methods to the questions or issues being investigated (and not the other way around). Before selecting an experimental research design, we should ask whether it fits the intervention we want to study. In particular, we need to consider the context in which the intervention is used, the content of the intervention itself, the process by which it is applied and the nature of its results or outcomes. In each of these domains, we may find either low variance (homogeneity) or high variance (heterogeneity). Table 1 offers some examples by way of illustration.

When an intervention has low variance across these domains, then the experimental method elegantly eliminates

all potential biases and confounders, proves causality beyond dispute and quantifies effect. The theoretical basis for the intervention is a secondary consideration—our focus is on its empirical performance. However, when we find high variance in one or more domains, the value of the experimental method is less clear, because the variance reduces or even eliminates our ability to generalize empirically about the impact of an intervention from any specific experiment, done with a certain context, content and application combination, and to draw conclusions about how it might work in a different context, with a different content and different application. Then, the theoretical basis for the intervention (why and how it works) becomes more important than its empirical performance (whether it works) in any particular study.

As the natural heterogeneity of an intervention increases, experimental methods become progressively less helpful in understanding its effectiveness. We see this in the findings from experimental research into some clinical interventions, such as controlled trials of surgical procedures where application variance can be high, or of behavioural therapy interventions where content, application and outcome variance are significant. In such studies, the findings are more difficult to interpret and apply to clinical practice than those for other clinical interventions where variance is low—for example, many pharmaceutical therapies.

But quality improvement initiatives are complex social interventions, for which high levels of variance in context, content and application are often inherent and desired characteristics of the initiative [3]. For example, the responses of different healthcare organizations to a continuous quality improvement programme, or a system for adverse event reporting and investigation, will be quite different—and the programme or system will rightly be tailored or modified to make it work better in the individual organizational setting. Attempts to ‘standardize’ or control such interventions to make them fit an experimental paradigm completely miss the point—that their multiple outcomes are a complex co-product of context, content and application variables. A different approach to evaluating effectiveness is needed.

Address reprint request to Kieran Walshe, Manchester centre for Healthcare Management, University of Manchester, Manchester, UK; Tel: (0161) 275 3852. Fax: (0161) 273 5245. E-mail: kieran.walshe@man.ac.uk

Table 1 Variance in context, content, application and outcome for interventions

| | Low variance (homogeneity) | High variance (heterogeneity) |
|--|--|---|
| Context—the situation, setting or organization in which the intervention is deployed | All contexts are the same or similar—for example, the functioning of the human body and its physiological response to disease follow deterministic patterns that can usually be presumed not to vary between study populations or over time. | Context varies widely—for example, significant differences exist between organizations or communities, between social cultures, or between health system delivery and funding mechanisms. |
| Content—the nature or characteristics of the intervention itself | The content is clearly specified and standardized and highly repeatable—for example, the delivered dose of a pharmaceutical agent. | Content varies widely—for example, an intervention may be tailored to an individual or to an organization, or modified to fit organizational or other characteristics, or redesigned and changed when in use. |
| Application—the process through which the intervention is delivered | The application process is the same or similar—for example, protocol-driven therapeutic regimes for a given condition. | The application process varies—depending, for example, on the skill and experience of the people involved, or the response or behaviour of recipients or other actors. |
| Outcomes—the results of the intervention | There is a single, clearly measurable outcome—such as a physiological function, or survival for a given time period. | There are multiple and less directly measurable outcomes that cannot be easily quantified and aggregated—such as learning, development or behavioural changes at an individual and organizational level. |

In researching healthcare quality improvement, we should learn from those in other settings—such as education [4] and criminal justice [5]—who have decades of experience in researching complex social interventions. Programmes designed to teach children to read, to rehabilitate offenders or to mentor disaffected teenagers are just as challenging to evaluate. Researchers in these fields have largely abandoned the experimental method, in favour of theory-driven approaches to evaluation [6]. In brief, theory-driven evaluation first attempts to map out the programme theory lying behind the intervention and then designs a research evaluation to test out that theory. The aim is not to find out ‘whether it works’, as the answer to that question is almost always ‘yes, sometimes’. The purpose is to establish when, how and why the intervention works, to unpick the complex relationship between context, content, application and outcomes, and to develop a necessarily contingent and situational understanding of effectiveness. The researchers seek theoretical rather than empirical generalizability—the ability to transfer theories from the research setting and bring them to bear in often quite different combinations of context, content and application [7].

In conclusion, I do not argue that there is no place for experimental research methods in testing the effectiveness of quality improvement interventions. Far from it, I fully acknowledge the power of experiments in demonstrating causality and quantifying effect, but I also assert the need for a theoretically driven approach to understanding complex social interventions and their effects, and the greater explanatory power of a more contingent and situationally sensitive approach to research.

The last word can be left to the researchers who, over 60 years ago, undertook the famous Hawthorne experiments, exploring the effects on the productivity of workers at a Western Electric plant of various changes in working conditions. Although the rigour of these studies has since been criticized, the researchers acknowledged at the time the methodological problems they faced. They found that their efforts to use controlled experiments to test for the effects of single variables (like pace of work, or rest periods) were practically impossible, and resulted in them altering the very thing they were trying to research. The researchers found that experimentation created ‘[not] an ordinary shop situation, but a socially contrived situation of their own making. With this realisation, the inquiry changed its character. No longer were the investigators interested in testing for the effects of single variables. In the place of a controlled experiment, they substituted the notion of a social situation which needed to be described and understood as a system of interdependent elements.’ [8] Similarly, healthcare quality improvement programmes are complex social interventions that can only be properly evaluated if their interconnected context, content, application and outcomes are understood.

References

1. Grimshaw J, McAuley LM, Bero LA *et al.* Systematic review of the effectiveness of quality improvement strategies and programmes. *Qual Saf Health Care* 2003;**12**:298–303.

2. Perneger T. Ten reasons to conduct a randomized study in quality improvement. *Int J Qual Health Care* 2006; **18**:395–6.
3. Walshe K, Freeman T. Effectiveness of quality improvement: learning from evaluations. *Qual Saf Health Care* 2002; **11**:85–87.
4. Brown AL. Design experiments: theoretical and methodological challenges in creating complex interventions in classroom settings. *J Learn Sci* 1992; **2**:141–78.
5. Pawson R, Tilley N. What works in evaluation research. *Br J Criminol* 1994; **34**:291–306.
6. Pawson R, Tilley N. *Realistic Evaluation*. London: Sage Publications, 1997.
7. Pawson R, Greenhalgh P, Harvey G, Walshe K. Realist review – a new method of systematic review designed for complex policy interventions. *J Health Serv Res Policy* 2005; **10**:21–34.
8. Roethlisberger FS, Dickson WJ. *Management and the Worker*. Cambridge, MA: Harvard University Press, 1939.

Accepted for publication 15 January 2007