

Increasing the generalisability of improvement research with an improvement replication programme

John Øvretveit,¹ Laura Leviton,² Gareth Parry^{3,4}

¹Medical Management Centre (MMC), Karolinska Institutet, Stockholm, Sweden

²The Robert Wood Johnson Foundation, New Jersey, USA

³Institute for Healthcare Improvement, Cambridge, Massachusetts, USA

⁴Harvard Medical School, Boston, Massachusetts, USA

Correspondence to

Gareth J Parry, Research Scientist, Institute for Healthcare Improvement, 20 University Road, 7th Floor, Cambridge, MA 02138, USA; gparry@ihi.org

Accepted 12 October 2010

ABSTRACT

If a quality improvement is found effective in one setting, would the same effects be found elsewhere? Could the same change be implemented in another setting? These are just two of the 'generalisation questions' which decision-makers face in considering whether to act on reported improvement. In this paper, some of the issues are considered and a programme of research for testing improvements in different settings is proposed to build theory and practical guidance about implementation and results in different settings.

INTRODUCTION

'Will findings from quality improvement studies apply in my situation?' 'How can we repeat this quality intervention in our organisation?' Better answers to these practical questions would facilitate faster and more widespread implementation of effective quality improvement changes (QICs) and could possibly reduce the spread of ineffective or poorly adapted ones. These are two of the questions this paper addresses in considering how to enable decision-makers to better decide whether and how to implement a QIC which others have applied and which is proven or promising. It describes the practical and scientific challenges which make it more difficult for decision-makers to repeat similar interventions or predict effects in situations other than where the study was made. The paper presents a design for a programme of research to provide more useful information and to accumulate knowledge about which results could be expected in which situations from a QIC.

THE CHALLENGE OF GENERALISABILITY

Generalisability of research is not a new issue, and discussion of the subject can usefully

draw on resolutions devised by many different disciplines.^{1 2} As regards medical treatments, many practitioners try to assess whether improved outcomes found in a randomised controlled trial (RCT) of a treatment will apply to their patients. The practitioners consider how their patients differ from those selected to participate in the RCT (eg, comorbidities) or the likelihood of their patients following the treatment.

The extent to which the effect seen in a RCT is associated with the treatment is related to the concept of 'internal validity' (defined in [box 1](#)).³ Whether the treatment is likely to have similar effects in other settings is related to the concept of 'external validity' or 'generalisation'.^{1 2} The purpose of this paper is not to give an overview of validity or to repeat design issues that are well-discussed elsewhere.² Conventionally, generalisation is discussed in relation to sampling, guided by a theory. In the case of QI, however, theory needs development which is usefully informed by a better understanding of practice. And any sampling frame for QI is inadequate without better information about the characteristics of patients, providers, organisations and notably—treatments—to be sampled.⁴

Where 'the subject' of the study is not a patient but an individual provider or organisational unit, and the 'treatment' is an intervention to change behaviour or organisation, uncertainty about generalising from an external study is even greater. If the study design is a randomised or matched controlled trial of a quality intervention to providers, then generalisability can be increased by describing the intervention and the context in more detail. The science of quality improvement is in the early stages of recommending how these details should be reported.⁵

Compared with controlled trials, findings reported from the field have less strong



This paper is freely available online under the BMJ Journals unlocked scheme, see <http://qualitysafety.bmj.com/site/about/unlocked.xhtml>

Box 1 Definitions used in this paper

- A quality improvement change (QIC) is an individual improvement change addressing a specific problem or opportunity, carried out in a specific context: a specific healthcare organisation and time period, with a specific patient population and improvement activities. *It may or may not be described in a research study.*
- A class of QICs is the set of underlying improvement principles or concepts from which individual QICs are derived and to which we would like to generalise from the individual QICs that are observed.
- Theory of a QIC is the use of a class of QICs, its outcomes, and related mediators and moderators that permit better understanding of why and how a QIC would produce an outcome in a given situation.
- Internal validity of a QIC study is the validity of a claim that a QIC caused a change in outcome.³
- External validity about the outcomes of a QIC study has a two-part definition.²
 - Generalisation about a class of QICs: Validity of the claim that the causal relationship holds across variation in organisations, patient populations, time periods, individual QIC activities and measurement of outcomes
 - Generalisation to a new QIC: Decision-makers' degree of certainty about whether a QIC, or a class of QICs, will produce similar outcomes for their organisations, patients and situation.
- Generalisability describes the range of situations or units of study to which findings or methods from elsewhere can be applied.
- Decision-maker: clinician, manager, policy maker or others who decide about whether or how to implement a quality improvement reported to be effective in one or more places.

internal and external validity. Because many such reports have no control group, the intervention effects at the study sites may be uncertain.^{2 6} In such reports, assessment of internal validity requires a description of the context factors 'surrounding' the intervention in order to understand the possible influences of these factors on the outcomes. Both the intervention and the context are often not well enough described for others to make a judgement about whether the QICs might be implementable with the same results in their situation. Similar points about descriptions of interventions for RCTs have also been made recently.⁷

Thus adequate description of context and intervention are necessary.^{5 7} Yet there is no consensus about which context factors should be reported or assessed, both to help practitioners and to build theory about which factors are important for which interventions. We also do not have a standard vocabulary for describing the features of QIC interventions. Many quality improvements are not detailed prescriptions that can be described in a fixed protocol. Rather they are concepts and ideas or objectives to be achieved, which provide

latitude to practitioners to adapt them during their implementation.

For example, the central line blood stream infection (CLBSI) bundle contains a hand hygiene component. One change concept is that having hand cleaning material easily available will lead to its greater use, increased hand hygiene and lower rates of CLBSI.⁸ But few studies describe well the details of how this change concept and other aspects of the bundle are interpreted locally or the details of implementation.⁹ Moreover, many changes are made during the implementation using a Plan–Do–Study–Act cycle or equivalent.¹⁰

DESIGNING PROGRAMMES TO DEVELOP MORE GENERALISABLE QUALITY AND SAFETY KNOWLEDGE

One way to provide more useful information for generalising from QICs is to provide better descriptions of the intervention and context.¹¹ For both controlled outcome studies and uncontrolled case studies, more detailed descriptions are required both for decision-makers and for researchers to build theory.^{12 13}

Another way is to examine how QICs are amended in their implementation and to develop a programme of research to study a particular QIC in different settings. Such a programme is proposed in the four steps shown in figure 1 to assess the implementation and outcomes of a QIC in different settings. The aim is to develop 'small theory' of QIC (small theory of improvement) about ways to deal with a specific problem across a variety of settings.¹⁴

Small theory here refers to a theory about which changes will lead to changes in provider behaviour and organisation (implementation theory) as well as how their changed behaviour or organisation will result in better outcomes for patients and/or lower costs. This

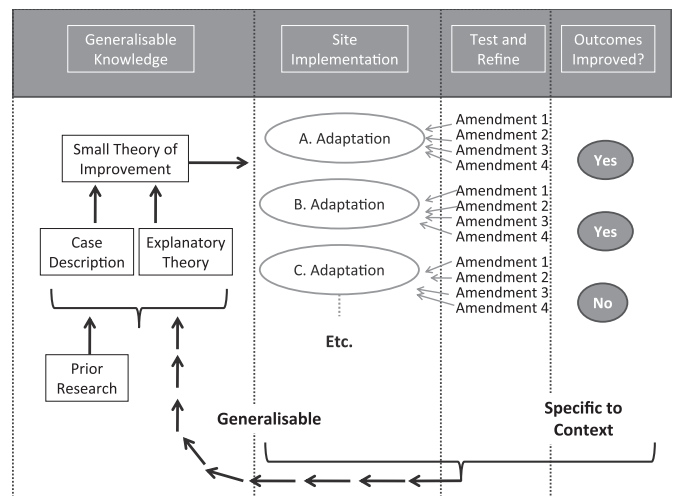


Figure 1 Improvement replication programme to increase the generalisability of improvement research.

small theory of improvement is not a protocol which requires very specific prescribed actions. Instead it is a set of principles or functions that together are hypothesised to produce quality improvement in the particular situation. These principles or functions might be enacted by practitioners in several different ways, but they would all achieve the same function and intermediate objectives.¹⁵ For example, there are many ways that cleaning material can be placed so that healthcare workers will see it in the course of their standard line insertion and monitoring work and the function of ensuring it is placed there and is used can also be assigned to a variety of people.

The programme design

The design of the programme shown in [figure 1](#) is for a number of studies of different implementations of one QIC. The aim is to understand adaptation in different contexts in order to produce knowledge about implementation and outcomes which is generalisable beyond one setting.

Step 1. Generalisable knowledge

The design starts with previous knowledge of the intervention to be investigated. The ‘small theory of improvement’ is the theory about the sequences of actions and intermediate effects which may lead to improvement in the ultimate clinical outcomes. Any previously reported explanatory theory and case descriptions could inform this small theory of improvement. Or it could be making explicit the assumptions made by implementers about how the intervention will produce the desired results (one type of ‘programme theory’). In the above CLBSI bundle example, a ‘small theory of improvement’ may be that having hand cleaning material placed in positions where clinical staff will see it in the course of their standard line insertion and monitoring work will lead to greater use of the materials, increased hand hygiene and lower rates of CLBSI.

Step 2. Site implementation

In this stage, each site interprets and applies the small theory of improvement according to what is possible and feasible in their context. The implementers need to consider how to accommodate differences in the population being treated, the constraints on the staff, other aspects of the setting and resources available, and the time period over which the QIC is to be attempted.¹ In quality improvement research, there is little systematic analysis and classification of these initial interpretations of the QI idea or principles, but there is a need to assess whether the interpretations still reflect the same underlying principles or theory that drove the original concept

of the particular QIC. In the CLBSI example, placing a poster on the doorways of care areas in order to remind healthcare workers to clean their hands can be seen as a very different ‘small theory of improvement’ to actually placing the hand cleaning material at the doorway. At a minimum, such amendments need to be studied in sufficient detail to be described to provide guidance to other implementers. While this example (described in more detail in [box 2](#)) seems straightforward, many QICs are more complicated, so descriptions and classification require reflection.

Step 3. Test and refine

‘Test and refine’ is where a detailed change or amendments to the initial interpretation are tested by practitioners, which may be one of many later ‘iterations’ of the original change. Each principle of the intervention will need to be interpreted for the situation, which we term here as ‘amendments’. Each amendment may work, or it may not work, or it may be rejected before it is tried. In the ‘Model for Improvement’, the Plan–Do–Study–Act cycle formalises this type of trial and error learning.¹⁰ This gives practitioners ‘permission’ to think actively about how a principle needs to be implemented in their situation, and it is here that their resourcefulness, creativity and detailed practical knowledge are the most needed.

By understanding each and every amendment, whether it is retained, abandoned, reconsidered or re-amended, we will better understand ‘what worked’ and why at an individual site. This, however, is context-specific knowledge that may not be relevant for other practitioners—even for those with similar situations. In the CLBSI bundle example above, for the placement of hand cleaning material, an organisation may test a number of settings for the cleaning material, including first placement near the entrances to the care area and then in other places physically nearer to the place where a line insertion is carried out or as part of the line-insertion kit. The usefulness of these placements is likely to be specific to the local context.

Step 4. Outcomes improved

The fourth part shows information about intermediate and final outcomes at each site. Provost describes in more detail how appropriate measurement approaches can be used to indicate whether the local amendments to the ‘small theory of improvement’ are resulting in improvements at a local level.¹⁶

Information from each part of the programme design can then be gathered in the feedback loop to provide a better basis for exploration of the intervention in a further set of situations. It can be used to provide detailed case descriptions as well as a revised small

Box 2 Example of the application of the improvement replication programme framework**Application of the improvement replication programme**

The central line blood stream infection (CLBSI) bundle contains five key components, including an element that relates to proper hand hygiene.⁸ We use this element here as an example of how the improvement replication programme framework can be applied.

Generalisable knowledge

The hand hygiene element of the CLBSI bundle was originally developed from a combination of *Prior Research* and *Explanatory Theory*. A group of local practitioners, who aim to apply the CLBSI bundle to their setting, may view a '*small theory of improvement*' as: 'Having hand cleaning material placed in positions where clinical staff will see it in the course of their standard line insertion and monitoring work will lead to their greater use, increased hand hygiene and lower rates of CLBSI'.

Site implementation

The local practitioners may start by using their local knowledge to make an initial adaptation based on the *small theory of improvement*; of placing hand cleaning materials at the doorways of care areas in order to remind healthcare workers to clean their hands.

Test and refine

In the CLBSI bundle example above, an organisation may test a number of settings for the position of the cleaning material, first placement near the entrances to the care area and then a number of other settings, including near universal precautions equipment used in the course of line insertion. The usefulness of these placements is more likely to be specific to the local context than to be generalisable to other settings. The practitioners may find that only by placing the hand cleaning material directly with the universal precautions equipment used in the course of line insertion do they achieve improved hand hygiene.

Outcomes improved

The data on hand hygiene and central line infections will be important to help local practitioners understand whether their local application of the small theory of improvement related to improving hand hygiene has led to decreased central line infections.

Synthesis

By synthesising across a number of sites, the descriptions of the initial *Site Implementation*, what was done under *Test and Refine* and whether *Outcomes Improved*, the initial *small theory of improvement* can be refined and if necessary updated.

theory of improvement specific to the QIC to inform implementation and other studies.

Possible steps forward

By designing a programme to report how the same QIC, or rather 'small scale theory', is interpreted and

amended to the situation, it is possible to strengthen both internal and external validity of studies which do not use controls. The design could be used retrospectively to investigate successful and less successful implementations, for example, in the Michigan state keystone CLBSI collaborative.¹⁷ It could be used piecemeal over time, as individual sites contribute information about their QI efforts. Or it could be used as a basis for a larger scale, more planned and organised prospective study, which would be easier to interpret using a common protocol for measurement and analysis. For example, the Transitional Care Model for older adults has been tested and found to be effective in three different randomised trials and is now being implemented throughout the Aetna system in the USA. Implementation in new settings has established the need for specific adaptations to address common barriers and opportunities of the settings.¹⁸ Developers are studying the degree to which these adaptations improve outcomes in order to guide how the model can best be used in different settings.

This programme design is also compatible with three other possible activities of importance to QI and QI research. The first are multi-site cluster randomised trials, which could be designed using these ideas. The second type of activity is pedagogy in QI and QI research, which can use the programme diagram to show how QI is applied in different settings. The third possibility is an online forum, blog or wiki for participatory research, in which individual practitioners can describe their efforts to colleagues, and where knowledgeable professionals and researchers in QI may be able to identify and analyse the larger patterns at work and to abstract the principles so necessary to further elaborate a theory of quality improvement.

Research funders also might be more willing to finance research or reports about QI which are part of an overall programme like this one where individually funded studies can contribute to an accumulation of knowledge that is more useful to decision-makers. An entity similar in idea to the Cochrane Collaborative might also help provide leadership and expert synthesis of the results of quality improvement research that follows the design described here, and help to build and refine practical theories of QIC that will be more readily available for adaptation by practitioners. If QI researchers working with practitioners can provide clear details of the adaptation and amendment processes, in addition to the outcomes undertaken in single or multiple site studies, then this entity could synthesise this evidence and integrate it with the existing evidence in order to update the 'small theory of improvement' associated with this particular quality improvement area.

CONCLUSIONS

A practitioner can never be certain that a quality or safety intervention reported to be effective elsewhere can be implemented successfully in their setting with the same positive results. However, QI researchers can provide better descriptions of implementation and context, especially in studies without controls or comparisons, which would help decision-makers assess relevance to their local situation. Another approach is a programme of research to repeat the intervention and the studies in different places to discover how it is adapted, why and with what results. Such programmes would require researcher coordination and one or more funders to support it because of its potential advantages for useful knowledge and for building theory cumulatively about quality improvements. Finally, to increase the practicality and application of the results from this QI research programme, a synthesising and disseminating approach similar to the highly influential Cochrane Collaborative might also be considered.

Funding The Health Foundation.

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. Leviton LC. External validity. In: Smelser NJ, Baltes PB, eds. *The International Encyclopedia of the Social and Behavioral Sciences*. Amsterdam, The Netherlands: Elsevier, 2001:5195–200.
2. Shadish WR, Cook TD, Campbell DT. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin, 2001.
3. Campbell DT, Stanley J. *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand-McNally, 1963.
4. Leviton LC. Reconciling complexity and classification in quality improvement research. *BMJ Qual Saf* 2011;20(S1):i28–i29.
5. Davidoff F, Batalden PB, Stevens DP, et al; SQUIRE Development Group. Development of the SQUIRE Publication Guidelines: Evolution of the SQUIRE Project. *Jt Comm J Qual Patient Saf* 2008;34:681–7.
6. Hand R, Plsek P, Roberts HV. Interpreting quality improvement data with time-series analyses. *Qual Manag Health Care* 1995;3:74–84.
7. Glasziou P, Chalmers I, Altman DG, et al. Taking healthcare interventions from trial to practice. *BMJ* 2010;341:c3852.
8. How-to guide: improving hand hygiene. <http://www.ihl.org/IHI/Topics/CriticalCare/IntensiveCare/Tools/HowtoGuideImprovingHandHygiene.htm> (accessed Jul 2010).
9. Shekelle PG, Pronovost PJ, Wachter RM, et al; and the PSP Technical Expert Panel. *Assessing the Evidence for Context-Sensitive Effectiveness and Safety of Patient Safety Practices: Developing Criteria (Prepared under Contract No. HHSA-290-2009-10001C)*. Rockville, MD: Agency for Healthcare Research and Quality, Jan 2010.
10. Langley GL, Nolan KM, Nolan TW, et al. *Provost LP The Improvement Guide: A Practical Approach to Enhancing Organizational Performance*. 2nd edn. San Francisco, CA: Jossey-Bass Publishers, 2009.
11. Department of Health. *Report of the High Level Group on Clinical Effectiveness*. London: Department of Health, 2007.
12. Baker GR. The Contribution of case study research to knowledge of how to improve quality of care. *BMJ Qual Saf* 2011;20(S1):i30–i35.
13. Øvretveit J. Understanding the conditions for improvement: research to discover which context influences affect improvement success. *BMJ Qual Saf* 2011;20(S1):i18–i23.
14. Lipsey M. Theories as methods: small theories of treatment. In: Sechrest L, Perrin E, Bunker J, eds. *Research Methodology: Strengthening Causal Interpretation of Non-Experimental Data*. Washington DC: US Department of Health and Human Services, 1990:33–51.
15. Hawe P, Shiell A, Riley T. Complex interventions: how out of control can a randomised controlled trial be? *BMJ* 2004;328:1561–3.
16. Provost LP. Analytic studies: a framework for quality improvement design and analysis. *BMJ Qual Saf* 2011;20(S1):i92–i96.
17. Pronovost P. Interventions to decrease catheter-related bloodstream infections in the ICU: the Keystone Intensive Care Unit Project. *Am J Infect Control* 2008;36:e171–5.
18. Naylor MD, Feldman PH, Keating S, et al. Translating research into practice: transitional care for older adults. *J Eval Clin Pract* 2009;15:1164–70.