Perspectives on Quality

Learning about improvement to address global health and healthcare challenges—lessons and the future

JOHN OVRETVEIT

Department Learning Informatics Management and Education (LIME/MMC), 17a Tomtebodavägen 18A. Karolinska Institutet, Stockholm 17177, Sweden

Address reprint requests to: John Ovretveit, Director of Research and Professor of Health Innovation Implementation and Evaluation. LIME/MMC, Tomtebodavägen 18A. Karolinska Institutet, Stockholm 17177, Sweden. Tel: +46703283935; Fax: +468-524 800 00; E-mail: jovretbis@aol.com

Editorial Decision 19 January 2018; Accepted 20 January 2018

Abstract

This perspectives’ paper highlights some of the learning from the seminar that the author considers to have particular relevance for improvement practitioners and for investigators seeking to maximize the usefulness of their investigations. The paper discusses the learning under four themes and also notes the future learning needed to enable faster and lower-cost improvement and innovative methods for this learning. The four themes are: describing and reporting improvement interventions; the theme of increasing our certainty about attributing effects to implemented improvement changes; the theme of generalizing the learning from one investigation or improvement and the theme of learning for sustainment and scale-up. The paper suggests ways to build on what we learned at the seminar to create and enable faster take up of proven improvements by practitioners and healthcare services so as to benefit more patients more quickly in a variety of settings.

Key words: improvement, learning, complex adaptive systems, implementation, delivery

Introduction

Why do we know less about the effectiveness of different improvement methods compared to what we know about the effectiveness of clinical treatments and care practices? To enable faster and lower-cost improvement, what have we learned and do we need to learn about improvement methods?

These are urgent questions: learning in ways that give knowledge that is actionable and leads to more effective improvement change can make the difference between life and death. The motive for the seminar was a view that we can improve the ways we learn as well as how we link this learning to practical action. We used the term ‘learning’ so as to encompass a broad range of ways of gaining knowledge, but with a focus on practical knowledge that directly can guide action in healthcare settings and policy-making.

This paper by one of the participants of the seminar gives reflections in relation to the 35-year history of learning from improvement projects and research into improvement in healthcare. The reflections arise from practical and research experience in four areas: working internationally in high- and low-resource settings; taking a combined improvement and implementation science approach; using different evaluation designs and methods; and choosing investigation methods suited to the knowledge needs of the users of the investigation. I observed four themes running through the seminar and papers: documenting and describing interventions and their contexts, attribution, generalization and sustainment and scale-up for more actionable research. This paper considers issues, debates and conclusions under each of these headings, and then gives some answers to the earlier questions about what have we learned and what do we most need to learn?
‘Learning’, ‘research’ and ‘investigation’

Before this, some reflections about why the need for the seminar and about what certain words meant to different people from the different countries and occupations represented at the seminar. The title ‘How do we learn about improvement?’ invites the observations that there is already much written on the subject and the question what could the seminar could add to this literature? A range of research designs for evaluating improvement interventions are well described [1–5]. This research literature in part overlaps with a more practice-based quality improvement literature describing pragmatic methods for project teams to discover whether their change is an improvement [6, 7]. Less well known by quality improvers and researchers was the literature noted at the seminar from implementation science, program evaluation, public health evaluation and international health [8–12].

A dialogue between researchers and practical improvers is helped by viewing different investigation methods positioned at different points on a continuum of ways to evaluate the effects of an improvement intervention. At one end is a randomized controlled trial design, then different observational designs and annotated time series graphs, with, at the other end, informed-observers assessments of the effects of an improvement intervention. This continuum shows different methods for learning about improvement. Another idea that helped dialogue in the mixed group represented at the seminar was to emphasize that all methods have their strengths and weaknesses and to use the method suited to the questions of the users and to the time and resources available. Using the term ‘learning’ made it easier to discuss a full range of ways of gaining knowledge about improvement, and the strengths and weaknesses of each. It avoided the misunderstandings and unproductive exchanges that can happen in a mixed researcher and practitioner group if the term ‘research’ is used.

‘Quality improvement’

There was confusion in the seminar at times about what the term ‘quality improvement’ referred to: does it refer to a quality tool or a specific improvement change? Does it mean a program including many projects, would it encompass safety improvement methods, or ‘lean methods’? Apart from the hindrances to communication, the different meanings and understanding of quality improvement and other terms make it difficult for reviewers to collect relevant studies, or for improvers to find methods or changes that they are interested in. The seminar noted that the basis of science is precision in terms and concepts which is the first step towards measurement. It is also important to the culmination of knowledge.

Theme 1: documenting and describing improvements

One theme that ran through many sessions and papers is the need to improve how we document improvement changes, and then how we describe the intervention actions and methods in reports. For example, a bundle for preventing central line associated bloodstream infections (a ‘CLABSI bundle’) that were used in the original Johns Hopkins Hospital randomized controlled trial was well described in the published paper [13]. Well-resourced controlled trials like this one use precise intervention protocols to ensure the actual intervention is the planned one, and is well-described. But some investigations do not have these resources to ensure implementation as planned, or to collect data about what was actually implemented. For example, in the subsequent scale-up of this bundle in Michigan State across 102 intensive care units, we know there were large variations in outcomes between the different sites [14, 15]. That there was a significant average improvement across all is an important finding. But we lost the opportunity to discover exactly why some units achieved higher or lower results because we did not have details of how well the bundle was implemented at each site.

Incomplete and imprecise descriptions mean that others cannot copy either the improvement change or the implementation methods: the descriptions do not provide them with enough of the right details about these two parts of an improvement intervention. The seminar noted some of the guides that help investigators to decide which data to gather about the improvement intervention that was actually implemented, as well as reporting guides [16–19]. Knowing exactly what was implemented can help to focus our outcome data collection: if what was implemented is different from what was planned then different outcomes may be expected. Guides also exist for documenting the implementation methods used to enable practitioners or patients to take up the improvement change [20–22].

Part of this ‘documentation and descriptions’ theme also involved the question of how to collect and report data about the context of the improvement intervention. For example, before the original Johns Hopkins’s controlled trial of the CLABSI bundle there had been a 2-year clinical unit safety program that had created a context that made implementation easier, as well as an earlier publicized child death and culture change program at the hospital [23]. The seminar and papers discussed different context data-gathering frameworks and instruments for documenting internal and external context [24–26]. We noted the extra resources needed for gathering data about context. Also, that the data-gathering burden could be reduced by better pre-data collection theorizing about which context influences were most likely to affect implementation for the particular type of intervention being considered: ‘context is not background but one of the star actors in the drama’. There were also different views about whether ‘readiness for change’ is part of context and whether readiness assessment tools were one type of measure of context [26].

Theme 2: strengthening attribution and internal validity

How do we know that a change in outcome data is due to the quality intervention, and not due to something else, such as a change in staffing or types of patients? The seminar noted that learning about attribution in some quality projects could be improved by including a comparison site or group that did not receive the quality intervention. For example, in the Michigan state study, including comparison ICUs that was not involved in the CLABIs improvement project. However such comparison may not be possible and only time series outcome data may feasible for some improvement investigations. There was debate about what degree of certainty of attribution was possible from well-annotated time series graphs, and discussion of the importance of calculating and graphing upper and lower control limits using the right statistical methods [6, 7, 27, 28].

More generally, the seminar discussed whether a new approach for observational research provided one way forward. This approach introduces into quality improvement investigations some of the methods used in program evaluations and implementation science. These designs involved pre-study formulation of a program theory, sometimes depicted in a logic model, that maps out the inputs, change activities and outcomes expected [29–36]. The seminar discussed how these approaches were similar and different to ‘theory of change’ and quality improvement ‘driver diagrams’ [37, 38].
Theme 3: strengthening generalization and external validity

How can we maximize certainty of attribution at the same time as maximizing the generalizability of the findings of an investigation? A controlled trial maximizes our certainty of attributing observed outcomes to the intervention, but only for the particular staff or patients that were exposed to the intervention. Would it work for others? Often research funding provides the resources and support necessary for full implementation: even if the patients or staff were similar, would others be able to copy the intervention exactly?

At the seminar, there was less agreement and less progress made on the set of generalization questions we considered than we made on the attribution questions. Yet, it became clear that valid generalizing of learning about improvements beyond the test site is crucial if others are to use what has been learned from that site. Uncertainty about whether the intervention can be copied exactly in different places, and whether the same results would then be expected hinders others from acting on the research [35]. Researchers eager to produce generalizable knowledge also need to qualify their findings and provide ‘health warnings’ about when and where the same findings might not be observed. Better planning of research designs was needed with a view to external validity as well as internal validity [36].

Theme 4: sustainability and scale-up

Many countries and organizations want to take an improvement that has been successful in one pilot and to ‘scale it up’ or ‘spread’ it to many services and settings. We may get the wrong impression from the research because of a publication bias towards successful scale-up examples: anecdotal evidence suggests many have ‘patchy’ success’, with large variations in take up and results in the many local projects across a program. There is some guidance provided by the literature on the subject, but there is limited empirical evidence about which improvements can be effectively scaled-up in different situations [39–41]. We noted the urgent need for knowledge for more effective scale-up, and the potential to learn from local projects that are part of scale-up programs. We also noted the inevitability of adaption of an improvement change in a scale-up program [42], often forced on local projects by different contexts to those of the original pilot: ‘We make improvements but not always in conditions of our own choosing’.

The practical significance of the above three themes became clear when we considered sustainability and scale-up of improvements at the seminar:

- Without adequate descriptions, others cannot copy the improvement intervention or assess whether they have the conditions to implement it successfully,
- Without careful documentation we cannot assess whether it was copied exactly or how it was adapted, and attribution of outcomes to the intervention is more difficult,
- Without follow-up over longer than 2 years we cannot assess if the change, or the results were sustained, or how the improvement was adapted to adjust to changing circumstances. We cannot test the hypothesis that non-sustainment after initial full implementation is due to changes in context, such as a change in staffing, that undermines the continued viability of an improvement.

Structured reports can be stored in an electronic database and accessible to other sites in a scale-up program. This would allow peers to learn from the sites most similar or near to them, as well as researchers and others to review the scale-up program for learning and management. The internet technology is now sufficiently mature and affordable for global learning communities of practice for specific improvement changes, tied to shared development goals.

Answers to the questions?

Drawing on the discussions above, the following gives a personal view of answers provided by the seminar to the questions raised in the introduction as well as where future development is most needed:

**Do we know as much about the effectiveness of different improvement methods and approaches for improving care as we know about clinical treatments and care practices?**

There is some evidence that improvement methods give an effective way for enabling faster take up of proven treatments. In addition, that improvement approaches help to engage staff in effective activities to reorganize healthcare and make it more efficient. More knowledge about these methods as effective ways to change care, and that focus on clinical practice and organization, is valuable for bringing improvements to most patients in ordinary clinics and hospitals. Yet the amount of money and resources given to evaluating medical treatments is considerably greater than that spent on evaluating improvement methods and changes. One view is that the methods for evaluating medical treatments are more highly developed and that there are supposedly greater difficulties of evaluating improvement interventions. Another view is that this is not the case, and, even if it were, then we would need to invest more to develop better methods for evaluating improvement methods and changes because of the potentially high payback.

**What is the difference in learning between using traditional research methods to evaluate an improvement intervention and using improvement tools to evaluate an improvement change?**

The degree of certainty afforded by controlled trials about associations between changes in outcome data and the improvement is greater. A single controlled trial gives high attribution certainty but is generalizable only to similar patients or providers in similar settings. Annotated improvement time series graphs give less certainty of attribution and are also not generalizable. A third approach discussed at the seminar—theory-based program evaluation methods—give ‘medium certainty’ of attribution and some degree of generalization. The generalization is not in terms of similarity of settings or participants, but by proposing others find ways to implement the same principles in their setting as those that were theorized to underlie the studied change.

At the seminar, there appeared to be a divide between those who used, and believed in, the validity of quality improvement tools for evaluating whether a change is an improvement, and some researchers who viewed these methods as flawed and misleading in their conclusions. The ideas of a continuum of investigation methods helped our discussions of the strengths and limitations of different methods along the continuum, as well as the principles of choosing the methods most suited to the knowledge user’s needs and being scrupulous about reporting the limitations of the study for non-researchers.

**What have we learned and do we need to learn about improvement methods and approaches and what are the best ways for enabling faster and lower-cost improvement?**

In my view, the most urgent learning needed is about how to effectively sustain and scale-up improvements proven to work in one place and
time. Focusing on this can drive advances in learning about description, attribution and generalization. We have learned that understanding more about context and about how to adapt a change proven to work in different settings is one-way forward and we have tools to enable this learning [24–26]. In addition, that new ways to learn about this may include harvesting data about context and adaptation collected by many improvement projects in a scale-up program and to enable sharing of this data within the scale-up community.

Building a global science of improvement
Quality improvement and implementation methods are being used to good effect to address global health challenges, for example for reducing maternal and child mortality [43]. Discoveries about how quality improvement methods can help to implement proven practices to reduce maternal and child mortality in rural India can also be used to save lives in parts of Europe and USA. Part of a ‘global quality perspective’ is a recognition that improvement methods and experience developed in low-resource settings in the two-thirds world are often relevant to settings more wealthy western countries. The seminar showed that the flow of knowledge is not just one way from the wealthy west to low-income countries, but that innovations in both implementation and research in these resource-constrained settings can be used in parts of Europe and USA [44].

Conclusions
The importance of more effective and actionable learning about improvement has never been greater. There is a rapid increase in quality activities around the globe and great opportunities to increase our learning from each other. What became clear from the seminar was the work we need to do to maximize learning of a particular type: learning for actionable knowledge that improvers can apply to find and make improvements that others have tested. This included knowledge about how to adapt an improvement effectively when this is necessary. The internet and learning management systems give us new opportunities to capture, make available and apply the knowledge that others have gained about effective improvement and implementation.

The seminar aimed to find ways to learn more quickly about improvement, and globally, so as to save lives and reduce waste. Investigators are privileged with talents and a position in their societies that now expects more of them: to create and help apply knowledge to reduce the suffering that exists because the right knowledge is not available in the places and in the form it is needed. We welcome debate and further co-operation with those who share our aims to use and generate knowledge about improvement to address these challenges.

Appendix: Definition of terms
Adaptation: modifying a change that has been tested and found to be effective in another care setting or patient population, such as an assessment method, a treatment method, a care practice (e.g. for sterilization), or a service delivery model.

Improvement change: a change to work practices or work organization that results in better patient outcomes and/or less waste, sometimes termed the ‘new better way of working’. Examples: increased compliance with best practice in hand hygiene; or a new way of organizing patient transitions from hospital to home.

Implementation action: actions to invite and enable people to take up and perform the new better way of working, such as training, performance feedback or providing rewards or incentives.

References
1. Mercer S, DeVinney B, Fine L et al. Study designs for effectiveness and translation research identifying trade-offs. Am J Prev Med 2007;33:139–54.
2. Tunis SR, Stryker DB, Clancy CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. JAMA 2003;290:1624–32.
3. Portela M, Pronovost P, Woodcock T et al. How to study improvement interventions: a brief overview of possible study types. BMJ Qual Saf 2015;24:325–36. doi:10.1136/bmjqs-2014-001620.
4. Brown C, Liford R. Evaluating service delivery interventions to enhance patient safety. BMJ 2008;337:a2764. doi:10.1136/bmj.a2764.
5. Fan E, Laupakis A, Pronovost P et al. How to use an article about quality improvement. JAMA 2010;304:2279–87. doi:10.1001/jama.2010.1692.
6. Langley G, Nolan K, Nolan T et al. The Improvement Guide: A Practical Approach to Enhancing Organizational Performance. San Francisco: Jossey-Bass, 1996.
7. Speroff T, O’Connor GT. Study designs for PDSA quality improvement research. Qual Manag Health Care 2004;13:17–32.
8. Brownson RC, Colditz GA, Proctor EK (eds). Dissemination and Implementation Research in Health: Translating Science to Practice. New York: Oxford University Press, 2012.
9. Shadish W, Cook T, Leviton L. Foundations of Programme Evaluation: Theories of Practice. London: Sage, 1991.
10. Habicht JP, Victora CG, Vaughan JP. Evaluation designs for adequacy, plausibility and probability of public health programme performance and impact. Int J Epidemiol 1999;28:10–8.
11. Potvin L, Haddad S, Frohlich KL. Beyond process and outcome evaluation: a comprehensive approach for evaluating health promotion programmes. WHO Reg Publ Eur Ser 2001;92:45–62.
12. Green LW, Kreuter MW. Health Promotion Planning: An Educational and Environmental Approach, 2nd edn. Mountain View, CA: Mayfield, 1991.
13. Pronovost P, Needham D, Berenholtz S et al. An intervention to decrease catheter-related bloodstream infections in the ICU. N Engl J Med 2006;355:2725–32.
14. Watson S, George C, Martin M et al. Preventing central line–associated bloodstream infections and improving safety culture: a statewide experience. Jt Comm J Qual Patient Saf 2009;35:593–7.
15. Pronovost P, Berenholtz SM, Goeschel C et al. Improving patient safety in intensive care units in Michigan. J Crit Care 2008;23:207–21.
16. Ogrinc G, Davies L, Goodman D et al. SQUIRE 2.0 (Standards for Quality Improvement Reporting Excellence): revised publication guidelines from a detailed consensus process. BMJ Qual Saf 2016;25:986–92. doi:10.1136/bmjqs-2015-004411.
17. Goodman D, Ogrinc G, Davies L et al. Explanation and elaboration of the SQUIRE (Standards for Quality Improvement Reporting Excellence) guidelines, v.2.0: examples of SQUIRE elements in the healthcare improvement literature. BMJ Qual Saf 2016;25:e7. doi:10.1136/bmjqs-2015-004480.
18. Strinam S, Miller R, Toder K et al. Development of a framework and coding system for modifications and adaptations of evidenced-based interventions. Implement Sci 2013;8:65. http://www.implementationscience.com/content/8/1/65.
19. Pinnock H, Harwick M, Carpenter C et al. for the StaRI Group. Standards for Reporting Implementation Studies (StaRI) Statement. BMJ 2017;356:j6795. 10.1136/bmj.j6795.
20. Proctor EK, Powell BJ, McMillen JC. Implementation strategies: recommendations for specifying and reporting, *Implement Sci* 2013;8:1–11.
21. Powell B, Waltz T, Chinman M et al. A refined compilation of implementation strategies: results from the Expert Recommendations for Implementing Change (ERIC) project. *Implement Sci* 2015;10:21. DOI:10.1186/s12110-015-0209-1.
22. Michie S, Richmond M, Johnston M et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. *Ann Behav Med* 2013;46:81–95. doi:10.1007/s12160-013-9486-6.
23. Pronovost P, Weast B, Rosenstein B et al. Implementing and validating a comprehensive unit-based safety program. *J Patient Saf* 2005;1:33–40.
24. Carroll C, Patterson M, Wood S et al. A conceptual framework for implementation fidelity, *Implement Sci* 2007;2:40. doi:10.1186/1748-5908-2-40.
25. Helfrich C, Li Y, Sharp N et al. Organizational readiness to change assessment (ORCA): development of an instrument based on the Promoting Action on Research in Health Services (PARIHS) framework. *Implement Sci* 2009;4:38. doi:10.1186/1748-5908-4-38.
26. Weiner BJ, Amick H, Lee SD. Review: conceptualization and measurement of organizational readiness for change: a review of the literature in health services research and other fields. *Med Care Res Rev* 2008;65:379–436.
27. Wheeler D. *Understanding Variation*. Tennesse: SPC Press, 1993.
28. Carey R, Lloyd R. *Measuring Quality Improvement in Healthcare*. New York: Quality Resources, 1995.
29. Birckmayer J, Weiss C. Theory-based evaluation in practice: what do we learn? *Eval Rev* 2000;24:407–31.
30. Patton M. The program’s theory of action: conceptualizing causal linkages. In: Patton M (ed). *Utilization-Focused Evaluation*. Thousand Oaks: Sage Publications, 1986.
31. Funnell S. Program logic: an adaptable tool for designing and evaluating programs. *Eval News Comment* 1997;6:5–17.
32. Davidoff F, Dixon-Woods M, Leviton L et al. Demystifying theory and its use in improvement. *BMJ Qual Saf* 2015;24:228–38.
33. Goeschel C, Weiss W, Pronovost P. Using a logic model to design and evaluate quality and patient safety improvement programs. *Int J Qual Health Care* 2012;24:330–7.
34. Kellogg K. Logic model development guide. K. Kellogg Foundation One East Michigan Avenue East Battle Creek, Michigan 49017-4058; 2004. www.wkkf.org (17 November 2017, date last accessed).
35. Leviton L, Trujillo M. Interaction of theory and practice to assess external validity. *Eval Rev* 2016: 1–36. DOI:10.1177/0193841X15625289.
36. Overtveit J, Leviton LC, Parry GJ. Increasing the generalisability of improvement research with an improvement replication programme. *BMJ Qual Saf* 2011;20:87–91.
37. Clark H, Taplin D. *Theory of Change Basics: A Primer on Theory of Change (PDF)*. New York: Actknowledge, 2012. http://www.theoryofchange.org/wp-content/uploads/toco_library/pdf/ToCBasics.pdf.
38. Bennett B, Provost L. What’s your Theory? Driver diagram serves as tool for building and testing theories for improvement, July 2015 *Associates of Process Improvement*. 2015. Available from: http://www.apiprojects.org/qp_what’s-your-theory_201507.pdf (17 November 2017, date last accessed).
39. Barker PM, Reid A, Schall MW. A framework for scaling up health interventions: lessons from large-scale improvement initiatives in Africa. *Implement Sci* 2016;11:12.
40. Massoud MR, Nielsen GA, Nolan K et al. A framework for spread: from local improvements to System-Wide Change. IH Innovation Series white paper. Cambridge, Massachusetts: Institute for Healthcare Improvement; 2006. Available on www.IHI.org (17 November 2017, date last accessed).
41. Massoud MR, Donohue KL, McCannon CJ. Options for large-scale spread of simple, high impact interventions. technical report. Published by the USAID Health Care Improvement Proj. Bethesda, MD: University Research Co, 2010.
42. Cohen D, Crabtree B, Etz R et al. Fidelity versus flexibility: translating evidence-based research into practice. *Am J Prev Med* 2008;35:5381–9.
43. Singh R, Kumar R, Taneja DP et al. Scaling up quality improvement to reduce maternal and child mortality in Lohardaga District, Jharkhand, India; 2015. USAID ASSIST/University Research Co., LLC (URC) from https://www.usaidassist.org/sites/assist/files/india_scaling_up_improvement_lohardaga_district_june_2015_a4.pdf (11 December 2016, date last accessed).
44. Ramaswamy R, Barker PM. Quality improvement in resource poor countries. In: Sollecito W, Johnson J (eds). *Continuous Quality Improvement in Health Care*, 4th edn. Sudbury MA: Jones and Bartlett Publishers, 2010: 537–65.
45. Overtveit J. *Evaluating Improvement and Implementation for Health*. Milton Keynes, UK: McGraw Hill/Open University Press, 2014.