

To RCT or not to RCT? The ongoing saga of randomised trials in quality improvement

Gareth Parry,¹ Maxine Power²

¹Institute for Healthcare Improvement, Cambridge, Massachusetts, USA
²Salford Royal NHS Foundation Trust, Salford, UK

Correspondence to

Dr Gareth Parry, Institute for Healthcare Improvement, 20 University Road, 7th Floor, Cambridge, MA 02445, USA; gparry@ihi.org

Received 28 September 2015
Accepted 15 October 2015

Williams *et al*¹ describe a well-conducted cluster randomised trial of a stroke quality improvement (QI) initiative, which aimed to improve two inpatient stroke indicators with strong evidence linking them to improved patient outcomes. They randomised five hospitals to receive a QI intervention, and six to receive only indicator feedback. In aggregate, they found evidence of improvement in one indicator, in the intervention group, relative to the control, but this was not sustained once the intervention period ended. The design, execution and analysis of the study were textbook for a cluster randomised controlled trial (RCT) design, aligning well with the CONSORT statement, the gold standard for RCT execution.²

There is much debate within the improvement field about the value of RCTs in determining the effectiveness of improvement interventions. In 2007, Donald Berwick's monologue 'eating soup with a fork' provided a convincing argument for why the RCT was necessary for evidence-based medicine, but inadequate for evaluating complex social interventions such as collaboratives and campaigns. Since then, there has been an apparent 'cooling' in the appetite of improvement practitioners to adopt RCT methods in attempts to understand the overall impact of improvement initiatives. Against this backdrop, we applaud the authors in their attempt, which goes against the trend, but disappointingly, once again, offers conflicting and weak evidence of beneficial effect despite adherence to rigorous method. So what does this study teach us about whether or not to embrace RCTs in improvement?

It is becoming widely recognised that improvement is a complex function of *the what*—the changes that are being sought, *the how*—the method, rationale

or theory for the wider improvement initiative, the *motivation and capability* of the organisations (both the improvement teams and the delivery team) involved and *the context* within which the improvement work is occurring. Moreover, it is clear from the literature that the policy context, the financial climate, the attention of regulators and professional bodies, the voice of patients and the fabric of the organisational ecosystem are critical to the likelihood of success. Improvement designers are starting to use tools such as the Model for Understanding Success in Quality (MUSIQ) to understand context.³ In the study by Williams *et al*,¹ it is unclear how much 'up front' work was done to understand, influence or modify the context. Indeed, this could be argued as one of the blind spots of the RCT. In addition to the ecosystem, the behaviours and motivations of teams entering improvement collaboratives appear to be as important as the environmental context. For example, in Stroke 90:10 negative and positive behaviours from teams were present in equal measure with observations of social loafing and freeriding.^{4 5} These behaviours go some way to explaining the likelihood of a team to 'benefit' from the improvement intervention as designed and currently are best investigated through qualitative research methods such as ethnography and semi-structured interview.

We know from previous studies that the nature or *the what* of an improvement intervention is often complex, adapts over time and can be poorly described, leading to inapt designs and ambiguous results.⁶ In the study by Williams *et al*,¹ *the what* is simple and well described; yet, the performance on two of the primary outcomes suggest that deep vein thrombosis (DVT) prophylaxis and the implementation of



► <http://dx.doi.org/10.1136/bmjqs-2015-004188>

To cite: Parry G, Power M. *BMJ Qual Saf* Published Online First: [please include Day Month Year]
doi:10.1136/bmjqs-2015-004862

swallow screening are at different gestational stages in their journey towards reliability. Understanding the complexity of the proposed changes and pairing this to the skills within the team delivering the change remains a ‘dark art’ and one that is constantly underestimated in improvement practice. Moreover, social complexity also relates to *the how* of the broader improvement initiative. For example, the Matching Michigan programme illustrated how underestimating the impact of a complex social environment led to disappointing implementation of a checklist.⁷ In the stroke study,¹ *the how* may be the most important feature of the intervention. *The how* includes QI coaching, but overall is poorly articulated, leading to confusion on how to interpret the study, for example do these results mean QI is not effective, or maybe coaching is not effective—and if so, not effective at what?

Classic RCTs assume all of this complexity can be ignored in the apparent search for generalisable solutions that apply everywhere. The reality is that there is usually substantial variation in context and readiness within and across organisations, contributing to variation in how interventions are implemented, and what their impact will be. Analysing heterogeneous sites as a group has been shown time and time again to demonstrate a secular trend towards improvement but no significant benefit in the intervention over control. Consequently, RCT results often tend to leave us with little practical learning on what worked, where and how, that can be used in the future.

We can embrace heterogeneity, and design and analyse studies to ask where and how an initiative works or can be amended to work. For example, we would understand more about barriers or facilitators faced by individual sites, if Williams *et al* described the variation in initial experience of improvement methods with existing stroke facilities and on how the initiative aligned with other activities in their organisations. We would understand more about how these factors impacted on outcomes if the authors presented data from each site. Data on a site-by-site, indicator-by-indicator basis are becoming standard practice in the improvement field; yet, presentation of the data in this way in the research literature is rare. We must move beyond this.

Importantly, embracing heterogeneity does not rule out randomisation. Revisiting the work of Ronald A Fisher, one of the pioneering figures in experimental design, suggests applying approaches, such as factorial designs on a more local level.⁸ That is, if we want to know what works in what contexts, the focus of experimentation and learning should be at the local level. Combining iterative, planned experiments with improvement tools and applying the learning directly in the local system will move us beyond the issues of generalisation that often thwart large-scale RCTs.

Attempts at generalisation to wider contexts are helpful, and will benefit from the use of theory, rather than statistical inference. As heterogeneity is embraced, the improvement field is likely to accelerate adoption of social science methods and designs that are better suited to studying the interaction of context.

The improvement field is evolving in this direction. For example, the new Standards for Quality Improvement Reporting Excellence (SQUIRE) guidelines ask authors to consider the underlying theory and the context.⁹ Others have called for greater clarity in describing the full components of an improvement initiative, recommending the use of theory.^{10–13} In addition, the study of context is developing, and approaches for measuring readiness or capability are being developed.¹⁴

Bringing these frameworks and approaches together, although challenging, offers an agenda for Improvement Science research methods to focus on. Moreover, the improvement field will also benefit from dialogue with funders and publishers so that improvement-related research can be funded and published in a way that will maximise the impact and learning from increasingly complex improvement initiatives in the future.

Contributors GP and MP co-wrote the editorial.

Competing interests None declared.

Provenance and peer review Not commissioned; internally peer reviewed.

REFERENCES

- Williams L, Daggett V, Slaven JE, *et al*. A cluster-randomized quality improvement study to improve two inpatient stroke quality indicators. *BMJ Qual Saf* 2015. Published Online First 24 Aug 2015. doi:10.1136/bmjqs-2015-004188
- Moher D, Schulz KF, Altman DG, Consort Group. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001;357:1191–4.
- Kaplan HC, Provost LP, Froehle CM, *et al*. The Model for Understanding Success in Quality (MUSIQ): building a theory of context in healthcare quality improvement. *BMJ Qual Saf* 2012;21:13–20.
- Power M, Tyrrell PJ, Rudd AG, *et al*. Did a quality improvement collaborative make stroke care better? A cluster randomized trial. *Implement Sci* 2014;9:40.
- Carter P, Ozieranski P, McNicol S, *et al*. How collaborative are quality improvement collaboratives? A qualitative study in stroke care. *Implement Sci* 2014;9:32.
- Shojania KG. Conventional evaluations of improvement interventions: more trials or just more tribulations? *BMJ Qual Saf* 2013;22:881–4.
- Bion J, Richardson A, Hibbert P, *et al*. Matching Michigan Collaboration & Writing Committee. ‘Matching Michigan’: a 2-year stepped interventional programme to minimise central venous catheter-blood stream infections in intensive care units in England. *BMJ Qual Saf* 2013;22:110–23.

- 8 Fisher RA. *Statistical methods, experimental design, and scientific inference*. Oxford University Press, 1935.
- 9 Davies L, Ogrinc G. New SQUIRE publication guidelines: supporting nuanced reporting and reflection on complex interventions. *BMJ Qual Saf* 2015;24:184–5.
- 10 Parry GJ, Carson-Stevens A, Luff DF, *et al*. Recommendations for evaluation of health care improvement initiatives. *Acad Pediatr* 2013;13:S23–30.
- 11 Davidoff F, Dixon-Woods M, Leviton L, *et al*. Demystifying theory and its use in improvement. *BMJ Qual Saf* 2015;24:228–38.
- 12 Reed JE, McNicholas C, Woodcock T, *et al*. Designing quality improvement initiatives: the action effect method, a structured approach to identifying and articulating programme theory. *BMJ Qual Saf* 2014;23:1040–8.
- 13 Goeschel CA, Weiss WM, Pronovost PJ. Using a logic model to design and evaluate quality and patient safety improvement programs. *Int J Qual Health Care* 2012;24:330–7.
- 14 Scaccia JP, Cook BS, Lamont A, *et al*. A practical implementation science heuristic for organizational readiness: R= MC2. *Journal of Community Psychology* 2015;43:484–501.