OPEN ACCESS

Value of small sample sizes in rapid-cycle quality improvement projects 2: assessing fidelity of implementation for improvement interventions

Edward Etchells,[1,2] Thomas Woodcock[3]

[1]Department of Medicine, Sunnybrook Health Sciences Centre, University of Toronto, Toronto, Ontario, Canada
[2]Centre for Quality Improvement and Patient Safety, University of Toronto, Toronto, Ontario, Canada
[3]NIHR CLAHRC for Northwest London England, London, UK

Correspondence to
Dr Edward Etchells, Department of Medicine, Sunnybrook Health Sciences Centre, Centre for Quality Improvement and Patient Safety, University of Toronto, Toronto, ON M5S, Canada;
edward.etchells@sunnybrook.ca

Received 29 May 2017
Revised 3 October 2017
Accepted 15 October 2017
Published Online First
2 November 2017

To cite: Etchells E, Woodcock T. *BMJ Qual Saf* 2018;**27**:61–65.

INTRODUCTION

Fidelity is the degree to which a change is implemented as intended. Improvement project teams should measure fidelity, because if the change is not implemented, nothing will change. However, measurement resources are usually limited, especially in the early stages of implementation. A frequent problem in quality improvement is that people waste time collecting too much data. A previous paper[1] showed how to demonstrate local gaps in care with very small samples of 5–10 patients. In evaluative clinical trials, the goal is to detect small differences between groups with precise estimates of these differences. By contrast, local quality improvement is often asking whether local performance meets a specific standard, such as 80% compliance with a guideline. If local performance is poor, small samples of 5–10 patients may be large enough to demonstrate a gap in care. In this paper, our goal is to offer some general guidelines to measuring fidelity of implementation on small samples in the face of constrained measurement resources.

Our target audience is healthcare improvers who have:

- Identified a local gap in care.
- Analysed the causes of this gap.
- Developed a change theory to address the gap.
- Created an initial change concept to be tested and refined locally.

A hypothetical scenario

A hospital-based improvement team is focused on medication reconciliation. Medication reconciliation refers to efforts to avoid unintentional changes to medication regimens at transition points such as hospital admission and discharge.[2] The team has conducted several small audits showing important gaps in the local system of medication reconciliation. The team has designed a new medication reconciliation form and is keen to start Plan-Do-Study-Act (PDSA) rapid improvement cycles. The hospital director wants the new medication reconciliation form implemented broadly as soon as possible. One member of the team wonders about conducting a randomised controlled trial of the new form. The team has about four person-hours per week to devote to the project. The team is unsure how to use their measurement resources to the best effect.

WHY IS IT IMPORTANT TO MEASURE FIDELITY OF IMPLEMENTATION?

We suggest four reasons to measure fidelity of implementation. The obvious reason is that if changes are not implemented, nothing changes. Fidelity is analogous to treatment adherence in controlled clinical trials. Ensuring excellent adherence is a major focus during clinical trials, because clinical trials tend to be negative if participants do not take the treatment. In the Diabetes Control and Complications Trial, 97% of patients adhered to a very tight regimen of insulin, likely because of very careful attention to adherence issues in a detailed feasibility study.[3]

Second, successful early implementation usually involves a committed local opinion leader, working with convenience samples of enthusiastic early

Table 1 Impact of fidelity of implementation on required sample size for an evaluative study

Fidelity of implementation (%)	Sample size required
100	100
95	110
90	123
85	139
80	156
75	178
70	204
60	278
50	400
40	625
30	1111

adopters. Broader implementation and dissemination, in the absence of committed leaders and enthusiastic adopters, may not necessarily be so successful. Monitoring fidelity allows for timely detection of implementation problems as the project moves from ideal early conditions to broader dissemination.

Third, it is important to optimise fidelity of implementation prior to undertaking evaluative studies. Low fidelity of implementation leads to smaller effect sizes, which means that a larger sample is needed to detect the effects of the change (table 1). Suppose that the estimated sample size to detect an effect is 100 patients. This sample estimate assumes that the fidelity of implementation is 100%. If the fidelity of implementation is only 70%, then the required sample size to detect the same effect doubles to 204.[4 5] If the fidelity of implementation is 70%, and only 100 patients are enrolled, then the study will yield a negative result. The corollary is that small improvements in fidelity can reduce the required sample size and increase the chance of demonstrating an effect. Improving the fidelity of implementation from 70% to 80% reduces the required sample size from 204 to 156.

Finally, a measure of fidelity of implementation is needed to interpret the result of an evaluative study. Suppose an evaluative study has a negative result. One possibility is that an ineffective change was correctly implemented. Another possibility is that a potentially effective change was not adequately implemented. A measure of fidelity will help determine which possibility is most likely.[6]

We assume that resources for measuring fidelity are constrained. We assume that if the local improvement efforts are initially successful, there may be efforts made to obtain additional resources for further dissemination and evaluation. Therefore, there should be enough data from the initial phase to support and inform further dissemination and evaluation. We focus on quantitative measurement, but we emphasise that qualitative measurement is important. Quantitative fidelity measures inform whether a change was implemented, but qualitative data inform why the change was, or was not, implemented.

MEASURING FIDELITY OF IMPLEMENTATION

We suggest five steps:

1. Choose fidelity measures based on a change theory.
2. Establish a minimum acceptable fidelity for each measure.
3. Establish a sampling strategy.
4. Choose a practical sample size.
5. Create run charts.

Choose fidelity measures based on a change theory

Change efforts should be based on a change theory that describes how each element of the change is expected to lead to improved processes and outcomes. A sound change theory makes it easier to choose fidelity measures. The initial steps of the change theory are logical places for fidelity measurement.

The initial fidelity measures can be rudimentary signals that change is occurring during early PDSA cycles. These initial measures can be considered PDSA-level fidelity measures. More robust fidelity measures should give confidence that the change has been implemented *as intended*, based on the change theory. Robust fidelity measures can subsequently be used as project-level fidelity measures during broader dissemination or evaluation. Interventions are often multifaceted, in which case a group of fidelity measures is appropriate.

The medication reconciliation team's change theory includes the following:

1. Clinicians will complete a form that outlines the patient's home medications and the plan for each medication (continue, change, stop) at the time of admission.
2. The form will be placed on the patient's chart.
3. This form will lead to fewer medication errors during the hospitalisation.
4. Fewer medication errors will lead to fewer preventable adverse drug events during the hospitalisation.

The project team considers the rudimentary ('PDSA level') fidelity measure of 'form is found on the patient's chart'. (If the form is not on the patient's chart, nothing is changing!) The project team recognises that the appearance of the form on the chart is insufficient evidence the change has been implemented as intended. The project team decides that 'forms correctly completed' is a more robust fidelity measure that could ultimately be a project-level fidelity measure. The team is willing to commit a little extra effort to measure 'forms correctly completed'. The project team supplements the fidelity measurement with qualitative feedback from users: What did the user like about the form? What could be improved? Why didn't the user complete the form?

Establish a minimum acceptable fidelity

We suggest a minimum acceptable fidelity of 70%. If fidelity is less than 70%, then the effect of any change will be attenuated, broader dissemination will be difficult to achieve, and the required sample size for evaluation may become prohibitively large (table 1). If fidelity is less than 70%, then a reasonable next step is to identify and ameliorate barriers to implementation, with ongoing monitoring of fidelity.[7] Our suggestion of a minimum acceptable fidelity of 70% is arbitrary. We are unaware of a 'correct' minimum fidelity value. A lower fidelity might be acceptable for changes that are simple, inexpensive and for which formal evaluation is not planned.

Establish a sampling strategy

The following are the two useful guiding principles for sampling in improvement projects, as previously outlined by Perla *et al*[8]:

1. obtain just enough data to guide next steps
2. make full use of local subject matter expertise in selecting the most appropriate samples.

In the early stages we suggest small convenience samples to test and refine the change. These convenience samples could include enthusiastic early adopters in ideal conditions. We propose a first arbitrary milestone of two consecutive convenience samples above minimum acceptable fidelity. If this simple milestone cannot be reached, broader dissemination is unlikely to be successful. The change should be refined or abandoned. After reaching this first milestone, we suggest moving to purposive (or judgement) samples that are more reflective of typical system performance. Purposive sampling can draw on local expert knowledge to ensure that implementation is tested more broadly, including situations when implementation can be expected to be difficult. We propose a second arbitrary milestone of two consecutive purposive samples above minimum acceptable fidelity.

As discussed in the prior paper on small sample sizes,[1] data quality is important on small samples. The five important steps to data quality are to (1) define the eligible sample, (2) establish exclusion criteria, (3) state the study period for each cycle/sample, (4) keep a reject log and (5) ensure complete data collection. Aim to enrol consecutive eligible patients. Random sampling is ideal but usually not practical.

The project team establishes the eligible sample and exclusion criteria, plans to keep a log of patients enrolled in each sample, and commits to complete data collection for all samples. The team decides to start with convenience samples of patients admitted to the enthusiastic physician on the project team during regular working days. The team also consults with local experts, who suggest that implementation will be a challenge for patients admitted to physicians uninvolved with the project, especially on weekends and nights. The team decides that the purposive samples will comprise patients admitted under these more challenging conditions.

Choose a practical sample size (table 2)

We assume that resources for measurement are constrained. Suppose that reality dictates that a maximum of 10 patients per cycle can be sampled. The minimum acceptable fidelity is 70%, so there must be at least 7/10 successes every cycle. If there are four failures in a cycle, then the cycle cannot achieve 7/10 successes. The cycle can be stopped, the failures can be studied qualitatively, and necessary adjustments can be made.

The medication reconciliation team's first PDSA implementation cycle (table 2) yields two correctly completed forms and four incomplete or missing forms on the convenience sample of six patients. There are already four failures, so the cycle can be stopped, because there is no chance of achieving 7/10 successes. The team studies the qualitative feedback and takes the necessary actions. The second cycle yields a similar experience. The third and fourth cycles are above the minimum acceptable fidelity of 70%, so the first milestone is achieved. The team moves to purposive samples of patients admitted to other

Table 2 Hypothetical results of early Plan-Do-Study-Act implementation cycles (steps 2–4)

Cycle	Sample	Form correctly completed	Qualitative feedback	Action
1	Convenience	2/6=33%	Hard to understand Hard to find	Make form more clear Make form easier to find
2	Convenience	4/8=50%	Hard to understand	Make form more clear
3	Convenience	7/10=70%	Hard to find	Make form easier to find
4	Convenience	8/10=80%	Time-consuming	Make form shorter First milestone achieved; change to purposive sampling
5	Purposive	0/4=0%	Unaware of the form	Improve awareness
6	Purposive	8/10=80%	Hard to understand Time-consuming	Make form more clear Make form shorter
7	Purposive	9/10=90%	Unaware of the form	Improve awareness Second milestone achieved; prepare for run charts

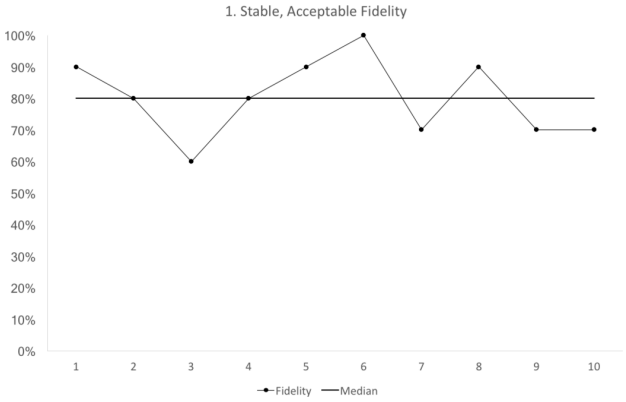

Figure 1 Run chart 1 shows the median fidelity above 70%, with no evidence of instability. Each data point represents 10 observations.

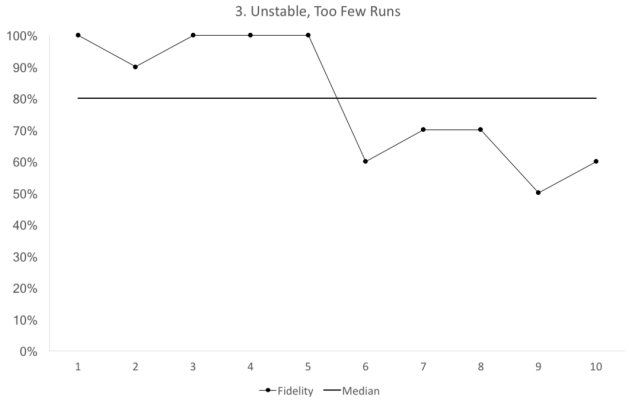

Figure 3 Run chart 3 shows a median fidelity above 70%, but the run chart is not stable. The run chart line crosses the median only once. For a run chart where there are 10 points that do not fall on the median, we would expect to see the median crossed at least three times[8]. Each data point represents 10 observations.

physicians, including nights and weekends. Cycle 5 (table 2) shows poor fidelity, so the team obtains feedback and makes the necessary improvements. Cycles 6 and 7 both achieve the minimum acceptable fidelity of 70% on two consecutive purposive samples, so the second milestone is achieved.

Create a run chart (figures 1–3)

A run chart is an efficient method for further enhancing a team's confidence in fidelity of implementation. Provost and Murray[9] recommend that a run chart needs 10 data points, with at least 10 observations per data point, and a consistent sampling approach. Do not use the data from step 4 for the run chart because the change, the implementation method and the sampling approach are all in flux during step 4.

We propose that a third milestone is a stable run chart with a median fidelity above the minimum acceptable value (figure 1). Run charts have some simple rules for stability.[10] If any rule is broken then the run chart is not stable.

1. trend: five consecutive points up or down (figure 2)
2. shift: six consecutive points above or below the median value (not including points that touch the median)

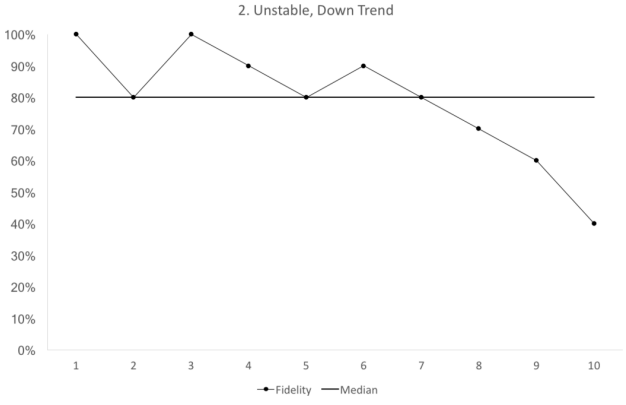

Figure 2 Run chart 2 shows a median fidelity above 70%, but the run chart is not stable. There is a downward trend starting with cycle 6. Each data point represents 10 observations.

3. runs: the values cross the median too often, or not often enough (figure 3).

One exceptional situation would be a run chart showing five consecutive increases (unstable but improving trend) and a median fidelity above target. In this case, the third milestone would be achieved.

Run charts cannot quantify the level of variation in the fidelity of implementation. Control charts, using larger samples and longer sampling periods, are needed to quantify the variation in fidelity of implementation. If ample data and resources are available, then control charts could be used earlier in a project.[11] Otherwise, we suggest that a stable run chart with a median value above the minimum acceptable fidelity of implementation is an important milestone before undertaking additional dissemination or evaluation.

Project-level measures that reflect the ultimate targets of change can also be developed and tested during this step. This additional measurement effort can:

1. Enhance confidence that the change theory is correct. If implementation is acceptable, then there should be a trend to improvement in the other project-level measures, with no expectation of demonstrating a statistically significant difference. If there is an unexpected worsening, then the change theory, the implementation plan and the measurement methods should be re-examined.
2. Plan the sample size of an evaluative study. A sample size estimate requires an estimate of fidelity, and an estimate of the baseline values for the other project-level measures. For many projects, change is targeted at providers, but the impact of the change is measured on patients. For example, in the medication reconciliation example, the target of the change is physicians filling out the form, but the downstream impacts (medication errors and preventable adverse drug events) are measured on patients. In such cases, the correlation between the targets of change and downstream measures is needed for a sample size estimate.[12]

These additional measurement activities might not be needed if the relationship between the change and downstream effects is already incontrovertible and generally agreed on, or if there is no plan for broader dissemination or evaluation.

The project team commits to measuring 'forms correctly completed' on 10 purposive samples (each with sample size of 10) over the next 10 weeks. The team also begins to measure medication errors and preventable adverse drug events. These are project-level measures that reflect the ultimate target of the medication reconciliation project, as outlined in the team's change theory. At the end of the 10 weeks, the run chart (figure 1) is stable and above the minimum acceptable fidelity. The team also has baseline measures for medication errors and preventable drug events, and an estimate of the correlation between physicians completing the form, medication errors and preventable adverse drug events. The team meets with a statistician to plan an evaluative study and obtain a sample size estimate. The team begins efforts to secure additional resources for broader dissemination and evaluation.

SUMMARY

The following are our take home points:

1. It is essential to ensure high fidelity during early implementation prior to broader dissemination and evaluation.
2. We propose five steps to measuring fidelity during early implementation.
   - choose a fidelity measure based on a change theory
   - establish a minimum acceptable fidelity
   - establish a sampling strategy
   - choose a practical sample size
   - create a run chart.
3. We suggest three measurement milestones during early implementation:
   - two consecutive convenience samples above minimum acceptable fidelity
   - two consecutive purposive samples above minimum acceptable fidelity
   - one run chart of 10 purposive samples, each with at least 10 observations, showing stable implementation with median value above the minimum acceptable fidelity.

Contributors EE and TW both conceived the review, contributed to initial drafts, revised the paper for critical content and approved the final version.

Funding This work was supported by the Health Foundation and the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care (CLAHRC) Northwest London.

Disclaimer The views expressed in this publication are those of the authors and not necessarily those of the Health Foundation, the NHS, the NIHR or the Department of Health.

Competing interests None declared.

Provenance and peer review Commissioned; externally peer reviewed.

REFERENCES

1 Etchells E, Ho M, Shojania KG. Value of small sample sizes in rapid-cycle quality improvement projects. *BMJ Qual Saf* 2016;25:202–6.
2 Kwan JL, Lo L, Sampson M, *et al*. Medication reconciliation during transitions of care as a patient safety strategy: a systematic review. *Ann Intern Med* 2013;158(5 Pt 2):397–403.
3 Implementation of treatment protocols in the diabetes control and complications trial. *Diabetes Care* 1995;18:361–76.
4 Chuang JH, Hripcsak G, Heitjan DF. Design and analysis of controlled trials in naturally clustered environments: implications for medical informatics. *J Am Med Inform Assoc* 2002;9:230–8.
5 Friedman LM, Furberg CD, DeMets DL. *Fundamentals of clinical trials*. 3rd edn. St Louis, Missouri: Mosby, 1996.
6 Gillespie BM, Marshall A. Implementation of safety checklists in surgery: a realist synthesis of evidence. *Implement Sci* 2015;10:137.
7 Hulscher M, Laurant M, Grol R. Process Evaluation of Implementation Strategies. In: Grol R, Wensing M, Eccles M, eds. *Improving patient care: the implementation of change in health care*. 2nd edn. Oxford UK: Wiley-Blackwell, 2013:444–70.
8 Perla RJ, Provost LP, Murray SK. Sampling Considerations for Health Care Improvement. *Qual Manag Health Care* 2013;22:36–47.
9 Provost LP, Murray SK. *The Health Care Data Guide. Learning from data for Improvement*. San Francisco, California: John Wiley and Son, 2011:48–9.
10 Perla RJ, Provost LP, Murray SK. The run chart: a simple analytical tool for learning from variation in healthcare processes. *BMJ Qual Saf* 2011;20:46–51.
11 Mohammed MA, Worthington P, Woodall WH. Plotting basic control charts: tutorial notes for healthcare practitioners. *Qual Saf Health Care* 2008;17:137–45.
12 Grimshaw J, McAuley LM, Bero LA, *et al*. Systematic reviews of the effectiveness of quality improvement strategies and programmes. *Qual Saf Health Care* 2003;12:298–303.