InnovAiT, 10(1), 51–56 DOI: 10.1177/1755738016674494

Measurement for improvement

In our first article on an introduction to quality improvement we discussed the philosophy and principles of quality improvement, and our second article discussed the model for improvement. In this article we discuss measurement for improvement, and how it differs from research and audit. We explore the selection, and types, of measures before explaining how to understand and analyse variation in your data in order to answer the question 'How do we know a change is an improvement?' posed by the model for improvement.

The GP curriculum and quality improvement

Professional module 2.2: Patient safety and quality of care states that understanding how and when to apply tools and metrics to improve the quality of care is a key skill that can and should be learnt during GP training, as well as enhanced in lifelong learning. Relevant knowledge and skills learning outcomes include:

- Describe the uses and abuses of clinical indicators and metrics such as benchmarking
- Understand the concept of variation in clinical care, how it is determined and measured and what actions might need to be taken to address inappropriate variation, for example in referrals, prescribing, admissions
- Demonstrate an understanding of the power of reporting from clinical systems for personal/practice audit and data analysis; and for comparisons with other practices that assist in setting the agenda for improving quality of care and recording of care

> If you cannot measure it, you cannot improve it.

This is a quote from Lord Kelvin, a mathematician and physicist, who developed the Kelvin scale of temperature measurement. The philosophy behind the quote has much relevance to healthcare. It is based on the precept that if you do not attempt to study (often quantitatively but not always) and ask questions of that data, then improvement is more challenging. A study of high-performing primary care practices in the United States found that a key foundation building block is data-driven improvement, with data systems that track clinical, operational and patient experience metrics (Bodenheimer, Ghorob, Willard-Grace, & Grumbach, 2014). A separate study found that one of the four habits of high-performing organisations is measurement and oversight (Bohmer, 2011).

This should not be a surprise to clinicians who value the role of measurement in clinical practice and in research, however, they often do not extend the same principles to health service delivery, and when they do, they apply research frameworks. However, measurement for improvement is different (albeit with some overlaps) to measurement for quantitative research and to measurement for clinical audit (Table 1) (Solberg, Mosser, & McDonald, 1997).

To assess whether an improvement project has led to the desired aim requires a selection of measures. The driver diagram was a tool we introduced in our introductory article to help bring drivers and change ideas together, to demonstrate how they contribute to the aim of the improvement project. It is also a very useful tool to consider what to measure as illustrated in Fig. 1.

This is adapted from a project to improve the quality of care in acute kidney injury published in *BMJ Quality Improvement Reports* (Brady et al., 2015). The project team identified a range of drivers and change ideas that were combined to form a driver diagram. Metrics for measuring each of the drivers can be overlaid on the same diagram. Sometimes, selecting measures requires a number of plan, do, study, act (PDSA) cycles and

Table 1. Differences between measurement for improvement, audit and research.

	Improvement	Audit	Research
Goal	To understand the process and evaluate a change (i.e. develop knowledge in a particular setting)	Often for comparison or to encourage change	To develop new know-ledge, but not necessarily its application
Hypothesis	Multiple and flexible	None	Fixed
Sample	Small	Variable, often large and 100% of available data	Large
Measures	Few	Very few	Many
Time period	Short and current (as close to real-time as possible)	Long, past	Long, past
Confounders	Consider, but rarely measured	Describe and try to measure	Measure or control
Testing strategy	Sequential tests	No tests	One large test
Determine if change is an improvement	Run charts or statistical process control charts	No change focus or sometimes ranking or percentage change	Statistical testing e.g. *t*-tests, Chi squared, *p*-values
Risks in improvement settings	Sometimes incorrectly perceived as 'inferior statistics'	Ignores time-based variation Over-reaction to natural variation	Ignores time-based variation

Based on Solberg et al. (1997).

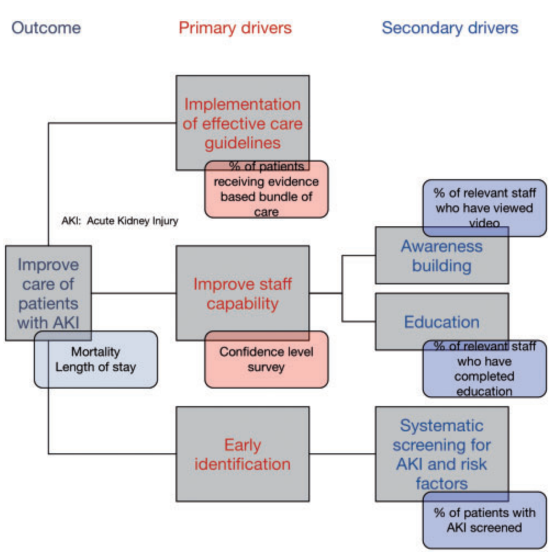

Figure 1. Driver diagram and measurement.
Adapted from Brady et al. (2015).

pilots before deciding on the final data to measure. General guidance in this area recommends that improvement projects should look to have a selection of outcome, process and balancing measures.

Outcome measures reflect the impact of the system on patients. Examples may include mortality or complication rates, e.g. the amputation rate in a practice's diabetic population. They may include proxy outcome measures, such as the glycosylated haemoglobin (HbA1c) level in a practice's diabetic population. It may also include patient-reported outcome measures in addition to more technical outcome measures as in the NHS Outcomes Framework.

Process measures assess how the different steps or processes in a healthcare delivery system are operating. In a diabetic population, the percentage of patients that have had their HbA1c checked in the previous year would be a process measure (whereas the percentage that has an HbA1c level below a target level is a proxy outcome measure). Other examples of process measures in relation to diabetic care may include the percentage of diabetic patients that have a recall, the proportion that have had a foot check or retinal screening performed. In the example in Fig. 1 the percentage of patients who have received the evidence-based bundle of care or number of staff who viewed the video are both examples of process measures. In a driver diagram, process measures are more likely to be towards the right of the diagram and outcome measures towards the left of the diagram.

Balancing measures determine the unintended consequences of an improvement. For example, trying to improve in one part of a system may lead to a knock-on consequence in another part of the system. Consider an intervention that requires the GP to go through a checklist during the consultation. The unintended consequence of this might be an unacceptable increase in consultation length, leading to patients waiting longer and a reduction in quality of patient experience.

Once the data is collected, an analysis is necessary. In research, data is often collected on a sample of patients who are assigned either to an intervention or control group. The data in the two groups is aggregated and then they are compared to one another; tests of statistical significance are undertaken to see if any difference between the two groups is significant or if any observed difference is purely due to chance (null hypothesis). In measurement for improvement, the analyst is trying to understand the type of variation in the data. Deming was one of the gurus of quality improvement in manufacturing and his work has been extremely influential in quality improvement in healthcare. He stated improvement occurs 'by understanding the variation that lives within your data and by making good management decisions on this variation'. There are two types of variation, 'common cause' variation and 'assignable (or special) cause' variation (Table 2).

Common cause variation arises from a host of factors that invariably impact any process. For example, if you write your name 10 times with your dominant hand, each time it will be slightly different and will be affected by the paper, the pen, how tired you are, how much time you have, and so on.

Assignable cause variation, on the other hand, is variation that occurs due to an identifiable cause. For example, if you write your name with your non-dominant hand, it is likely to be significantly different from the 10 times with your dominant hand.

In most cases in quality improvement, you will want to see assignable cause variation after implementing your quality improvement interventions. In some occasions, you may also want to see a reduction in common cause variation.

In order to tell the difference between common cause variation and assignable cause variation, the data should be presented visually either in the form of a run chart or a statistical process control chart (see Fig. 2). A run chart is a chart that plots the observed data points in a time sequence. A statistical process chart extends a run chart with limits that are drawn above and below the median. Similar to the standard deviation, these limits usually extend for a value of three times sigma (but in this case accounting for the temporal relationship in the data). The disaggregation of data and a subsequent understanding of its temporal nature is an important principle in measurement for improvement; it often involves smaller sample sizes, but more frequent measurements.

Consider the care of patients with epilepsy in general practice. There are many indicators of good quality of epilepsy care (Fountain et al., 2015), one of which is recording seizure frequency, so that a change in treatment to improve seizure control may be instigated if necessary. In a typical research project to improve quality of epilepsy care, one approach would be to randomly select some practitioners to be part of the intervention group; the control group would receive the usual practice and the two groups might be matched. At the end of the study the results in the two groups would be aggregated and compared to each other and statistical tests conducted on the aggregated results to see if any difference is a true difference or not.

If an audit was being conducted to improve the quality of care of epilepsy patients then typically there would be a baseline collection, some analysis of that baseline data,

Table 2. Types of variation and characteristics.

	Common cause variation	Assignable cause variation
Attributes	• Many factors, some known and others 'unknowable' • Described as 'noise in the system' • Affects all processes most of the time • Is inherent in the design of the process • Variation is predictable and the process is stable	• Usually few, not many • Can usually be identified • Not part of the process • Intermittently apparent • Unpredictable variation
Potential actions when found	• Do not react to individual results • Look at the average and process limits or spread of variation • Improve the whole process if the variation is not acceptable • Or continuously improve quality	• Look for the assignable cause and do something about it • Almost always something to find • Opportunities to learn

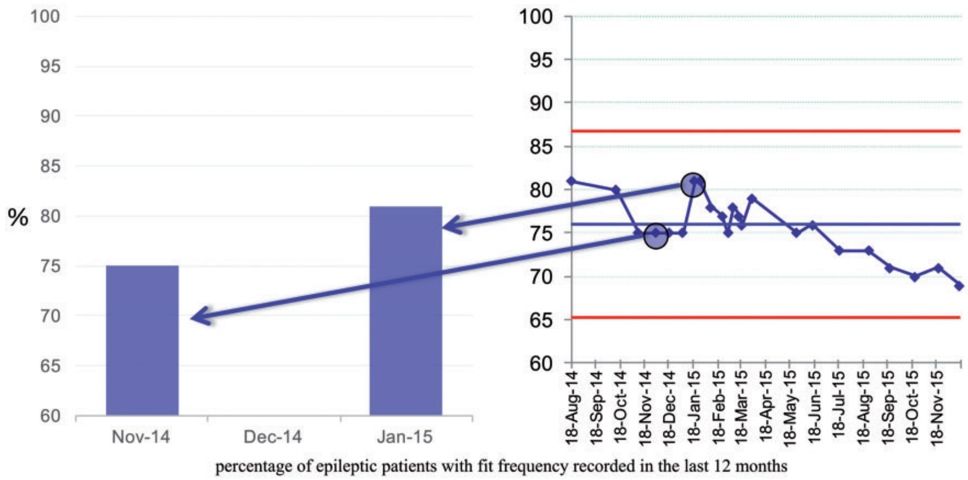

Figure 2. Recording of fit frequency in epileptic patients in the previous 12 months.

usually with a comparison to a standard or benchmark, the introduction of interventions and then a repeat data collection a few months later. However, in the case of measurement for improvement those data points would be collected more frequently and plotted as a run chart. The RCGP has produced guidance of tools in quality improvement including such charts.

Run charts allow for a better appreciation of the variation in the data. Figure 2 shows two charts for the percentage of epileptic patients with the frequency of their fitting episodes recorded. The column chart on the left is as in a typical audit, and the inference from this chart alone would be that there has been an improvement. However, if the chart on the right is observed, it actually suggests a deteriorating level of care, and that the two data points on the column chart were just showing common cause variation. Experts in healthcare quality improvement have suggested a series of rules for interpreting run charts, to decide whether there has been a change in the type of variation from common cause to assignable cause variation (Perla, Provost, & Murray, 2011). Four of these rules are illustrated with primary care examples (Fig. 3).

Rule 1: *Six or more consecutive points with all above or all below the median*.

A practice with a stable population was looking to improve the number of patients on long- term angiotensin-converting enzyme inhibitors (ACE-Is) who had had their electrolytes checked annually. They did a catch-up exercise by organising blood tests on patients that had not had a test and then instigated a regular search conducted by the data clerk every month to maintain the improvement. The red arrow indicates the point at which their improvement was implemented. The existence of six or more consecutive points above the median demonstrates their improvement.

Rule 2: *Five or more consecutive points all going up or down*.

In this improvement project the team members were looking to improve the number of patients identified and recorded as suffering with atrial fibrillation. The team made an improvement at the position of the red arrow, where they started screening for irregular pulses in patients over the age of 65 years and then conducting an electrocardiogram in those with irregular pulses to confirm atrial fibrillation. They had five or more subsequent points all going up.

Rule 3: The number of times the chart crosses the median line is compared to a table of critical values.

In this improvement project a practice team was looking to improve their referral rate per 1000 patients, which is represented on the *y*-axis. They introduced a change where they had regular team discussions about potential referrals for the five specialties that they most commonly referred (represented by the red arrow). Visually it appears as if the variation has reduced. A table of critical values states that the range of runs for a chart with 35 data points should be between 12 and 24. A run is a series of points in a row on one side of the median and is easily determined by the number of times the graph crosses the median line and adding one. This is represented by the green circles on the chart, which add up to 27 and adding one makes it 28. It is above the range in the table of critical values, suggesting that this change in data is non-random and assignable in nature.

Rule 4: Astronomical data point

A practice was measuring the percentage of patients with chronic kidney disease with hypertension and proteinuria who were treated with an ACE-I or angiotensin receptor blocker. They made a change indicated by the red arrow

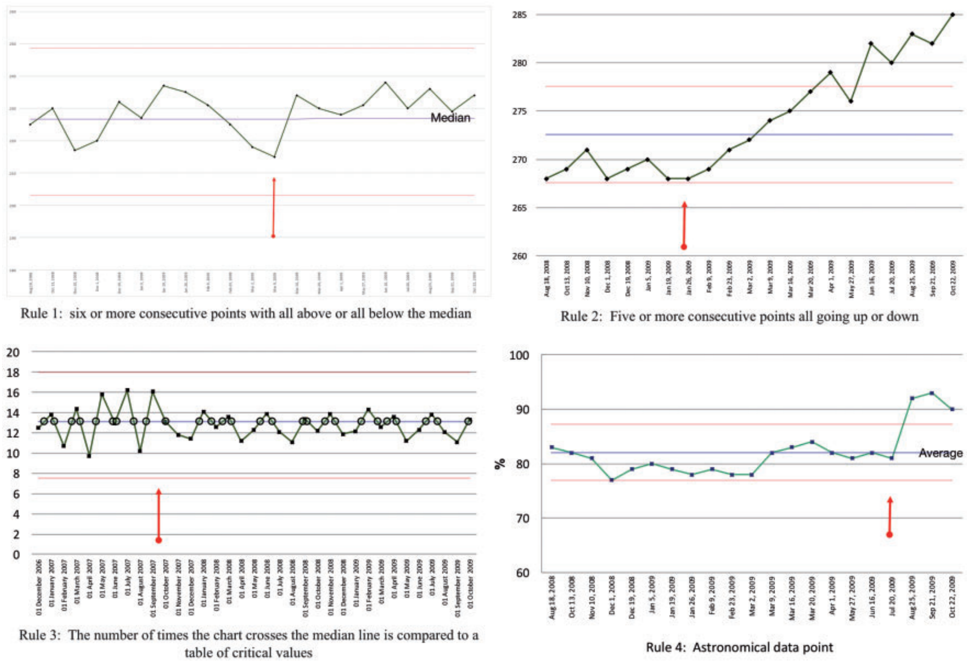

Figure 3. Examples of the four rules.

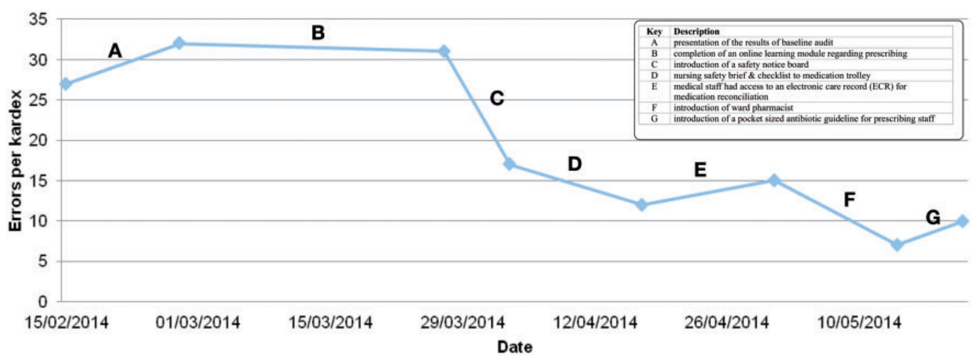

Figure 4. Annotated run chart.
Adapted from Donnelly, Lawson, and Watterson (2015).

with the subsequent data point being significantly different from the other data points (outside the limit of three-times the sigma value from the mean).

Charts help the analysis of the data, but if taken further and annotated, can be a very useful communication and engagement tool. Figure 4 shows an annotated run chart adapted from a project to improve paediatric prescribing practice (Donnelly, Lawson, & Watterson, 2015). There were a number of sequential interventions or PDSA cycles. After each intervention the team randomly sampled 10 drug charts to obtain their measure for improvement. One of the interventions in this project was the creation of a patient safety notice board, which the team used to highlight the results from each PDSA cycle. They regularly updated the run chart to demonstrate progress. Interestingly the graph suggests that this intervention itself led to an improvement.

Deming also said 'in God we trust, all others must bring data'. There is a reluctance to perform measurements in healthcare quality improvement efforts, and often it is perceived as a chore. However, there are some simple approaches to measurement for improvement that help answer the question 'Is the change an improvement?' More importantly, those same approaches are useful to engage and motivate others. By continuing to measure regularly they also give confidence that quality care is being delivered in a sustainable and ongoing basis.

Key points

- Measurement for improvement is different from measurement for research or accountability
- Driver diagrams can help to select what to measure and pull together a measurement framework
- Measures should ideally include outcome, process and balancing measures
- Measurement for improvement aims to distinguish between common cause and assignable cause variation using run charts and other similar charts
- Annotating charts is a powerful mechanism to drive motivation and engagement

References and further information

- Bodenheimer, T., Ghorob, A., Willard-Grace, R., & Grumbach, K. (2014). The 10 building blocks of high-performing primary care. *Annals of Family Medicine*, *12*(2), 166–171. doi:10.1370/afm.1616
- Bohmer, R. M. (2011). The four habits of high-value health care organizations. *New England Journal of Medicine*, *365*(22), 2045–2047. doi: 10.1056/NEJMp1111087
- Brady, P., Gorham, J., Kosti, A., Seligman, W., Courtney, A., Mazan, K., . . . Juniper, M. (2015). "SHOUT" to improve the quality of care delivered to patients with acute kidney injury at Great Western Hospital. *BMJ Quality Improvement Reports*, *4*(1), u207938. w3198. doi: 10.1136/bmjquality.u207938.w3198
- Donnelly, P., Lawson, S., & Watterson, C. (2015). Improving paediatric prescribing practice in a district general hospital through implementation of a quality improvement programme. *BMJ Quality Improvement Reports*, *4*(1), u206996.w3769. doi: 10.1136/bmjquality.u206996.w3769
- Fountain, N. B., van Ness, P. C., Bennett, A., Absher, J., Patel, A. D., Sheth, K. N., . . . Stecker, M. (2015). Quality improvement in neurology: Epilepsy update quality measurement set. *Neurology*, *84*(14), 1483–1487. doi: 10.1212/WNL.0000000000001448
- Perla, R. J., Provost, L. P., & Murray, S. K. (2011). The run chart: A simple analytical tool for learning from variation in healthcare processes. *BMJ Quality and Safety*, *20*(1), 46–51. doi: 10.1136/bmjqs.2009.037895
- RCGP. Profesional module 2.2: Patient safety and quality of care. Retrieved from www.rcgp.org.uk/training-exams/gp-curriculum-overview/online-curriculum/working-in-systems-of-care/2-02-patient-safety-and-quality-of-care.aspx
- RCGP. Quality improvement in general practice run charts. Retrieved from www.rcgp.org.uk/clinical-and-research/our-programmes/~/media/Files/CIRC/Quality-Improvement/A%20practical%20guide%20on%20how%20to%20build%20a%20run%20chart.ashx
- Solberg, L. I., Mosser, G., & McDonald, S. (1997). The three faces of performance measurement: Improvement, accountability, and research. *Joint Commission Journal on Quality Improvement*, *23*(3), 135–147

Dr Paresh Dawda
GP and Regional Medical Director, Ochre Health, Australia; Adjunct Associate Professor, University of Canberra, Australia; ACT Clinical Director, National Home Doctor Service, Australia; Honorary Associate Professor, Department of Health Services Research & Policy, Australian National University
Email: dr.paresh.dawda@gmail.com
Twitter Handle: @pareshdawda via @RCGP_InnovAiT

Dr Mareeni Raymond
General Practitioner and Clinical Advisor at BMJ Quality; Managing Editor BMJ Quality Improvement Reports