

Quality improvement science

Measuring for improvement

A Niroshan Siriwardena MMedSci PhD FRCGP

Professor of Primary and Prehospital Health Care, Community and Health Research Unit (CaHRU), University of Lincoln, UK

Steve Gillam MD FFPH FRCP FRCGP

Department of Public Health and Primary Care, Institute of Public Health, University of Cambridge, UK

ABSTRACT

This is the fourth in a series of articles about the science of quality improvement. We examine what to measure, how to measure and some important measurement techniques, such as run charts, control charts and funnel plots. These help us to understand healthcare processes, to assess whether they

are stable or improving and to determine how they can be improved further.

Keywords: control charts, funnel plots, general practice, primary care, quality improvement, run charts, statistical process control

Introduction

This, the fourth in our series of articles on quality improvement tools and techniques, examines what to measure, how to measure it and techniques of measurement for improvement.¹ Previous articles in the series have considered: frameworks for improvement, understanding processes and how to improve them, and most recently leadership and management for improvement.^{2–4} Everything we do can be seen as part of a process. The structure and process of care – for example, how people work, their work methods, the equipment and materials used, the work environment – as well as the outcomes of care are important measures for evaluating quality.⁵ Measurement is itself a process which not only helps us to assess other processes, but which can also be used to drive improvement.⁶ The techniques we discuss will help us understand whether processes are stable, improving or deteriorating and the extent to which they can be improved further.

audit or review criteria), standards and indicators. The latter may include quality indicators, performance indicators or clinical performance indicators, depending on what is being measured and why.

A criterion is a measurable aspect of quality (structure, process or outcome) of care and has been defined as ‘a systematically developed statement that can be used to assess the appropriateness of specific healthcare decisions, services and outcomes’.⁷ An example of a criterion is that every patient diagnosed with hypertension should have had their blood pressure recorded within the previous six months. This is translated into a measure: the proportion of patients with hypertension who have a blood pressure recorded within the previous six months (usually expressed as a percentage). If the criterion is based on research evidence which directly links it to improved outcomes it is sometimes referred to as a ‘review criterion’. For example, patients with hypertension should have a latest blood pressure reading (measured in the preceding six months) of 150/90 mmHg or less.

The level achieved for the criterion is compared with a standard, i.e. what should be achieved. The standard is the threshold of expected compliance for the criterion. Standards are usually derived from consensus opinion (either local or from a wider group) or are based on a previous audit. Less commonly, a standard may be based on published evidence about

What to measure

What we measure depends on what outcome we wish to achieve and therefore which parts of the process we should improve to do this. There are different types of measure. We can select particular criteria (also called

levels of performance that lead to improved outcomes. An example here is the standard required to achieve herd immunity for measles, mumps and rubella vaccination of 95% of the population.

How to measure

Different types of improvement project use different types of measurement, but not all are equally useful or informative. Measurements are generally of three types: before-and-after, continuous or comparative.

Clinical audits characteristically employ before-and-after measurements where standards are compared before and after an intervention. The advantage of this approach is that it provides the analyst with a target to aim for and it is usually simple to analyse and present data. A disadvantage is that the standard may be arbitrary. This may lead to gaming or unintended consequences. Furthermore, and most importantly, a change comparing a single measurement before and after an intervention may be an artefact of measurement rather than demonstrating a real improvement.

By contrast, quality-improvement projects tend to use continuous data-measuring processes, either recorded as counts or proportions (percentages). Rather than two measurements, i.e. before-and-after, multiple measurements are taken before, during and

after the intervention has taken place. Relatively simple statistical methods are then used to analyse whether the process is showing a natural (or random) variation over time, and if so, can demonstrate the extent of this variation and whether real improvement (over and above natural variation) is occurring.

The advantage of this approach is that it helps us understand whether real improvement has taken place and can demonstrate the extent of this improvement. It avoids interpreting natural variation as real change; and it enables us to see the effects of multiple interventions over time. The disadvantage of this method is that measurements need to be taken repeatedly during the process of change and some basic analytical concepts and techniques need to be learned (Table 1).

Measurement, whether using simple counts (e.g. numbers of referrals to hospital), rates (e.g. proportion of patients with a particular condition referred to hospital) or other more complex continuous variables, is also a process that can introduce variation. Therefore, it is important that data using samples or whole populations are gathered in a careful and consistent way.

Finally, we may wish to compare performance for different groups or organisations with the aim of comparing the best performers with the worst. The traditional method of doing this has been to represent the performance of each organisation on a bar chart that ranks the highest with the lowest or vice versa.

Table 1 Comparing measurement approaches in clinical audit and plan–do–study–act cycles

	Clinical audit	Plan–do–study–act
Criteria	Yes	Yes
Standards (target)	Yes	No
Type of measurement	Before-and-after	Continuous or repeated measurements over time (statistical process control)
Change implementation strategy	Change(s) are implemented after first audit, i.e. more rigid	Multiple changes can be introduced in sequence or parallel, i.e. more flexible
Cyclical	Yes	Yes
Ideal outcome	Meet or exceed standard	Improvement from baseline
Advantages	Simple to analyse Aims to achieve standard (target)	Analysis more complex Aims to achieve improvement
Disadvantages	Does not account for natural (common cause) variation (see below) Standard may be arbitrary	Takes common cause variation into account Improvement based on baseline performance rather than arbitrary standard

Unfortunately, this method may prevent a clear differentiation being made between high- and low-performing organisations/groups. If they are aware of the method of presentation, it may lead them to aim for a middle rank in such a table, where they are less likely to be noticed. This pursuit of mediocrity can be prevented by using funnel plots to compare organisations. Funnel plots are a special type of control chart that compare different organisations rather than a single organisation over time (see below).

Understanding variation

Every measure of a process, a combination of processes or an outcome will show variation over time. Variation is therefore part of any process. It is inevitable, and ubiquitous, but is amenable to measurement and control. If we want to demonstrate improvement it is essential that we select the key variables to measure quality in terms of outputs or outcomes that will signify improvement.

The natural variation in a stable process unaffected by the external factors affecting it or attempts to improve it is called ‘common cause variation’. We see common cause variation in, for example, repeated measures of blood pressure. These may be due to changes in the physiological state of the individual, subtle differences in the technique of measurement or in the response to the measuring instrument.

For other measures, such as prescribing rates, these may vary over time due to differences in patient case-mix; between prescribers and in their prescribing behaviour. Similarly, referral rates, vaccination rates, or indeed any other measure of health processes, organisations or systems will also vary over time due to variation in the process itself or in the process of measurement.

Variation which falls outside the ‘common cause variation’ is termed ‘special cause variation’. As its name implies, ‘special cause variation’ is caused by an ‘external’ factor, whether this is planned or unplanned, intended or unintended. Analysing variation over time involves using statistical techniques, but the simplest way of analysing and representing such variation involves a technique called statistical process control (SPC), developed by Walter Shewhart at Bell Laboratories in the 1920s and championed by WE Deming and Joseph Juran, Davis Balestracci and many others since.^{4,8–10} Table 2 summarises the differences between common and special cause variation.

Any improvement in the healthcare process requires a change in a process to reduce the effect of ‘common cause variation’ and to trigger a ‘special cause variation’ which will represent a significant improvement. However, responding to common cause variation as though it is special cause variation has the opposite effect to that which is intended. It may actually increase variation in the system. This is called ‘tampering’. An example of tampering is when an organisation responds to a single reduction (or increase) in a measure *before* checking that the change is due to common cause variation.

Table 3 summarises how we should and should not respond to the different types of variation. A special cause strategy calls for investigation and explanation, which will sometimes lead to specific responses depending on the special cause identified. Common cause variation requires a different approach. A common cause strategy first requires us to explore the variation more closely using stratification to reveal any special causes. Next, one should seek to understand variation through the processes and systems which cause a problem. Finally, we should redesign processes to reduce inappropriate and unintended variation in an agreed measure and a way that is responsive to patients’ needs.¹¹

Table 2 Differences between common and special cause variation

Common cause variation	Special cause variation
Predictable	Unpredictable
Due to ‘chance’ causes	Due to ‘assignable’ causes
Many factors	Usually few factors
Often ‘unknowable’	Can usually be identified
Is part of the process	Not part of the process
Affects process most of the time	Intermittently apparent

Table 3 Responding to common and special cause variation

Common cause variation	Special cause variation
Do not respond to individual results	Respond appropriately to special cause variation
Look at the average and the control limits	Investigate each point representing a special cause
Understand reasons for common cause variation by looking at underlying processes and systems	Try to explain why the special cause has occurred and what factors led to it
Try to find special causes by stratification according to organisational unit	Reinforce special causes leading to improvement while eliminating those leading to deterioration
Improve the whole process if this is not acceptable or go for continuous quality improvement	Improve aspects of the process which have resulted in special cause variation
Redesign processes to reduce inappropriate and unintended variation	Redesign processes to reduce inappropriate and unintended variation

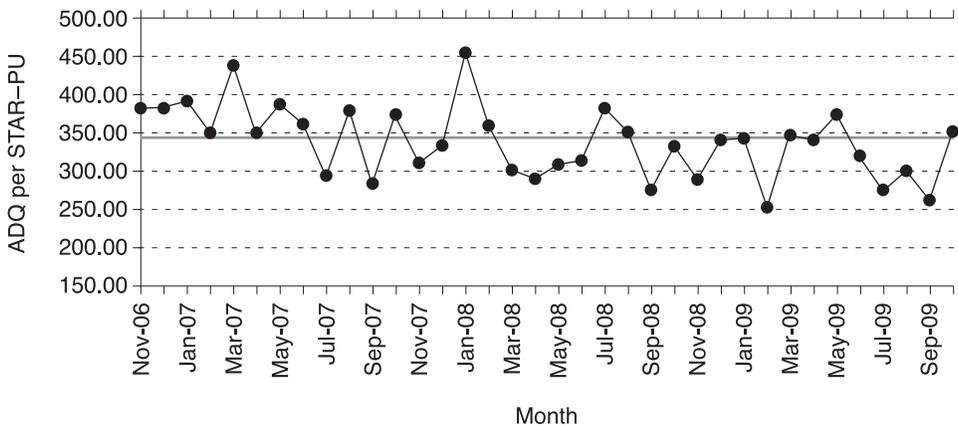


Figure 1 Run chart showing common cause variation. The central line represents the *median* average daily quantity per specific therapeutic group age-sex related prescribing units (ADQ per STAR-PU) for practice benzodiazepine prescribing

Statistical process control

Run charts

Run charts are the simplest way of plotting data over time. Data for a particular indicator are plotted as dots (data points) on a simple graph with time plotted on the *x*-axis and the value of the indicator plotted on the *y*-axis. The time intervals should be ordered and sequential, but not necessarily equal. They are often regularly spaced but need not be. At least 16 dots are usually required to see if a process is stable. The dots are connected by lines and a median line is drawn. Figure 1 is a run chart showing hypnotic prescribing data for a single general practice.

A 'run' is a sequence of dots above or below the median. Common cause variation is represented in a run chart as runs randomly distributed about the median. One way of conceptualising this is that, in

common cause variation, the chance of a run being above or below the median is the same as the chance of throwing heads or tails with a coin. Three simple statistical rules have been developed to show whether there has been a significant change in a measure over time, i.e. a special cause variation.

These rules are helpful because they prevent individuals or groups just 'eyeballing' a chart of measurements over time and misinterpreting them. Following certain rules leads to consistent interpretation of what constitutes a significant change over time. This also prevents an inappropriate response to common cause variation as if it were a special cause.

The three rules that identify the most important types of special cause variation are: shifts, trends and runs (shown in Box 1). A shift is a sequence or 'run' of seven dots above or below the median. A trend is a sequence of seven dots all going upwards or downwards (dots on the same level are excluded from the count).

Box 1 Basic rules for run charts

- Ordered sequence of data.
- At least 16 points (more than 25 does not usually add any further information in a stable process).
- Run: sequence of points above (or below) the median.
- Shift: 8 (or 7) or more points in 20 or more (or < 20 points).
- Trend: 7 (or 6) or more points going up or going down including the median but ignoring repeat values (or < 20 points).
- Runs randomly distributed around the median (see Table 2).

An example of a shift shown in Figure 2 is the rate of hypnotic drug prescribing in another general practice. The run chart shows a sequence of 25 dots. There are 11 dots below the median from January 2008, indicating a shift.

The final rule refers to ‘runs’ which give the ‘run chart’ its name. Runs should be randomly distributed about the median when there is only common cause variation. Therefore, we can calculate whether there are the right numbers of runs (between upper and lower limits) depending on how many dots there are in total in the chart which gives a probability table for runs (see Table 4). In Figure 2 there are only four runs when we would expect between 10 and 16 according to Table 4.

Control charts

A control chart is a more sophisticated form of run chart. The relationship between a run chart and control chart has been described as analogous to that between an X-ray and a magnetic resonance imaging (MRI) scan. The latter is more sensitive at detecting abnormalities, but also more complex and requires

greater resources. The principles of its construction and interpretation are very similar.

Figure 3 is a control chart showing hypnotic prescribing data for a single general practice and corresponds to the run chart in Figure 1. Again, data for a particular indicator are plotted as dots on a simple graph with time plotted sequentially on the *x*-axis and the value of the indicator plotted on the *y*-axis. The time intervals should be sequential. They are often regularly spaced but need not be. The dots are connected by lines but this time a mean line is drawn (Figure 3).

In addition, the control chart has two further lines: the upper and lower control limits. They differ from confidence intervals and should not be confused with them. The control limits are lines representing three standard deviations above and below the mean. The only slight complication here is that the mean and standard deviations should be calculated according to the type of data, i.e. normal distribution for biological variables such as blood pressure, Poisson distribution for count data and binomial distribution for yes/no or percentage performance data.

Confidence intervals are different from control limits. They represent a range of effect sizes around an odds or risk ratio for a study. Confidence levels are usually set at 95%. Stated simply, this means that if a study were to be repeated 100 times, 95 times out of a 100 the effect size would fall between the 95% confidence intervals.

Common cause variation in a control chart is shown in Figure 3 (which shows the same data as the run chart in Figure 1). All the dots fall within the upper and lower control limits and are randomly distributed about the mean. Control charts are more sensitive than run charts in detecting significant change over time. There are more rules for determining significant changes over time in a control chart, but again three basic rules identify the most important types of special cause variation: points outside the control limits, shifts and trends.

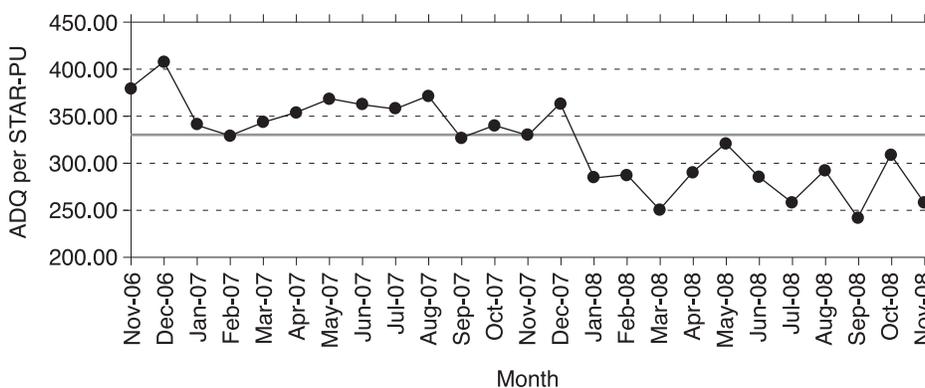


Figure 2 Run chart showing a ‘shift’. The central line represents the *median* ADQ per STAR-PU for practice benzodiazepine prescribing

Table 4 Probability table for runs

Number of observations (or dots)	Too few runs	Too many runs	Number of observations (or dots)	Too few runs	Too many runs
14	4	11			
15	4	12	28	10	19
16	5	12	29	10	20
17	5	13	30	11	21
18	6	13	31	11	21
19	6	14	32	11	22
20	6	15	33	11	22
21	7	15	34	12	23
22	7	16	35	13	23
23	8	16	36	13	24
24	8	17	37	13	25
25	9	17	38	14	25
26	9	18	39	14	26
27	9	19	40	15	26

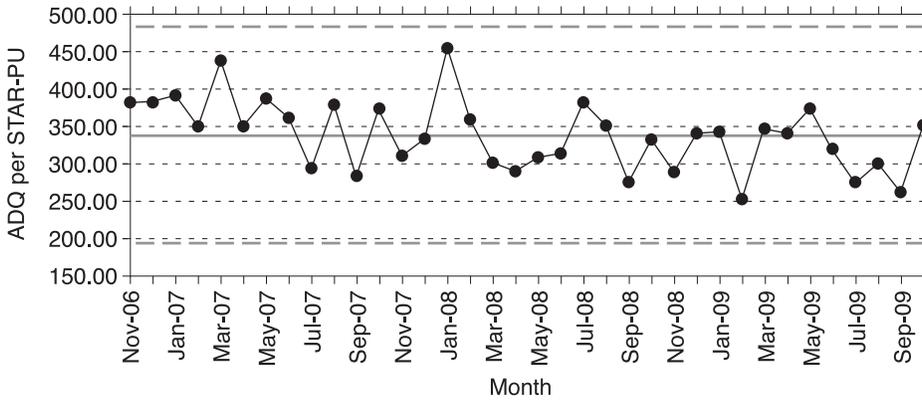
**Figure 3** Control chart showing common cause variation. The central line represents the *mean* ADQ per STAR-PU for practice benzodiazepine prescribing

Figure 4 is a control chart showing hypnotic prescribing data for a single general practice corresponding to the run chart in Figure 2. In Figure 4 the second dot in the sequence falls above the upper control limit. A shift is a sequence of seven dots above or below the mean (as shown in Figure 4). A trend is a sequence of seven dots all going upwards or downwards (dots on the same level are again excluded from the count). The number of dots for a shift or trend varies from six to

eight depending on the text and the total number of data points. A number of other rules can help to provide further signals that a significant change may have occurred. These are additional, more sensitive, rules but they are also more likely to cause false positive signals and need to be interpreted with caution. This practice has significantly reduced its hypnotic drug prescribing since December 2007.

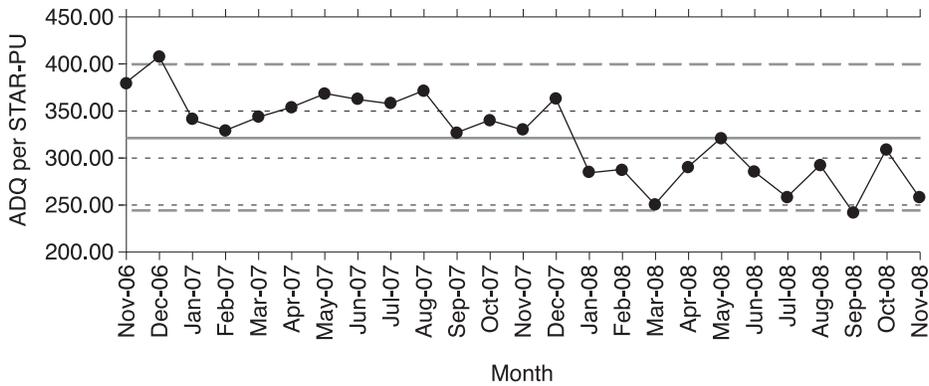


Figure 4 Control chart showing special causes. The central line represents the *mean* ADQ per STAR-PU for practice benzodiazepine prescribing

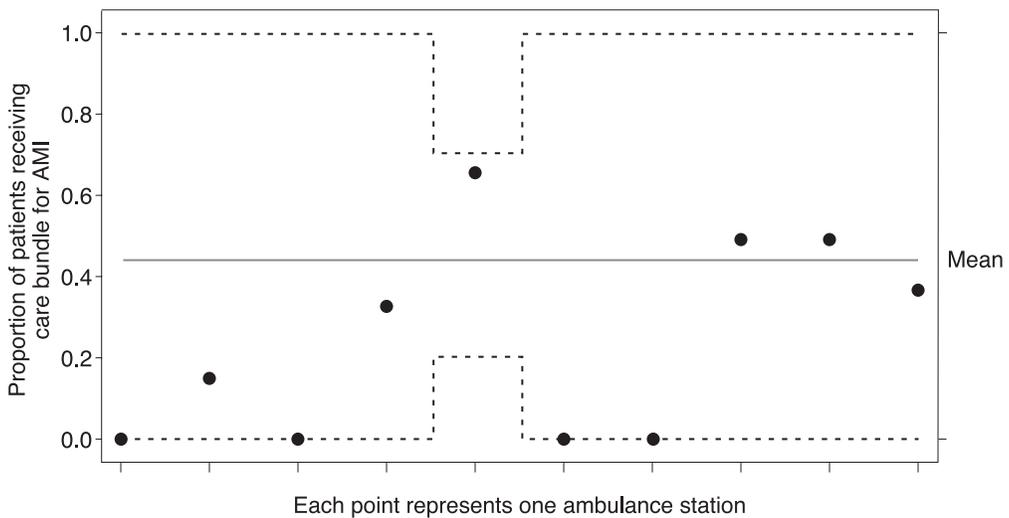


Figure 5 Control chart showing comparative (analysis of mean) performance for provision of acute myocardial care by ambulance station. $P = \text{mean probability} = 0.44$ (or 44%)

Funnel plots

As well as looking at an indicator over time, control charts can also be used to compare organisational units at a single point or during a fixed period. In this type of control chart, organisational units are arranged on the *x*-axis with their performance as a count, rate or proportion (or percentage) on the *y*-axis. The mean is represented and control limits are calculated for each organisational unit based on all the data provided (Figure 5).

In Figure 5 we represent data on the performance of ambulance stations as an organisational unit. In each case, a team of paramedics, delivered care for patients with acute myocardial infarction (AMI) during a single month. Each dot represents an ambulance station. Performance is measured as the delivery of a process 'care bundle' for AMI. A care bundle is an all-or-none measure where every eligible patient with

AMI should receive aspirin, glyceryl trinitrate, pain assessment and analgesia, unless there is a valid exception. The delivery of the care bundle can vary from 0 to 1 (i.e. 0 to 100%). The samples provided by each station were small which led to wide control limits (0 and 1) for most stations. The mean performance was 44.4% which meant that the care bundle was delivered to just over two in every five patients.

The control limits are denoted again by the dashed line vary for each station. If data are arranged according to the size of the sample denominator provided by each organisational unit, this produces a funnel plot (Figure 5). In Figure 6 we see AMI care bundle performance for 12 larger regional ambulance services in England. Each service is represented by a dot labelled 1 to 12. The sample denominator is now greater but varies from a few cases (service 12) to over 200 cases (service 7) in the month that performance is measured. The mean performance across all the services is

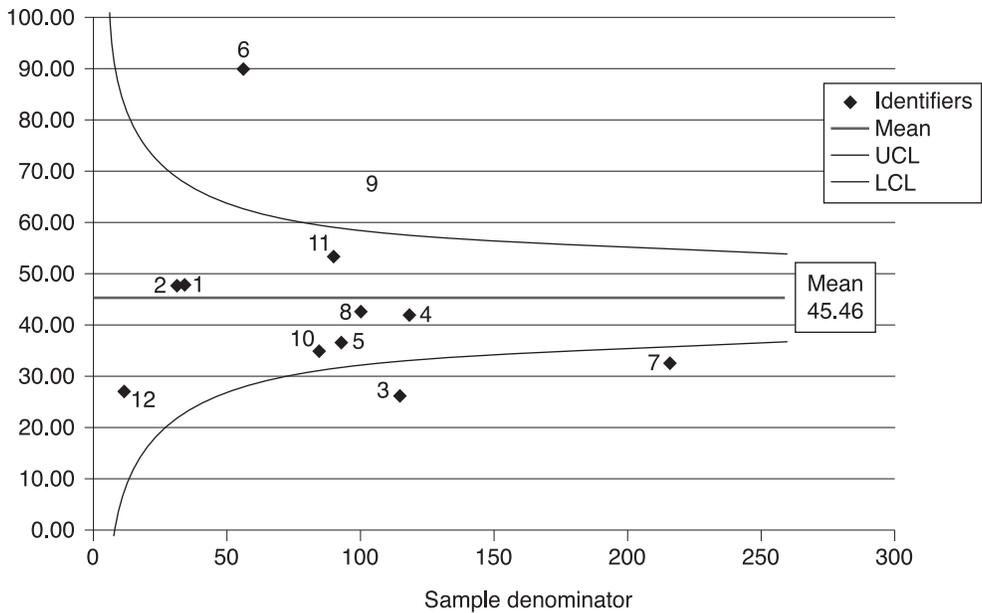


Figure 6 Funnel plot showing comparative performance for provision of acute myocardial care by ambulance services in England

again around 45%. The control limits, which are joined by a smooth line, are wider for services with small samples of AMI and become narrower as the sample size increases. This produces the characteristic funnel shape of the control limits. Because this looks like the bell of a trombone, funnel plots are sometimes referred to as ‘trombonograms’.

In the chart one can immediately see that most trusts are contained within the control limits. The processes in these trusts are delivering ‘average’ care defined by these control limits. Four trusts show performance either above or below the control limits. These trusts show significantly different performance, either higher or lower, than other trusts and this special cause requires further investigation to understand why this might be the case. The investigation may reveal a difference in the system of care providing this outcome.

Conclusion

This article has tried to provide an introduction to measurement for improvement. In it, we have explained how we can measure processes and why repeated measurement over time is the key to understanding variation in a process and how to improve quality. We have also introduced readers to the principles of constructing and interpreting run and control charts and how to respond to common or special cause variation. More information on SPC is available from a number of excellent articles and books.^{4,12–17}

REFERENCES

- Gillam S and Siriwardena AN. A primer for quality improvement in primary care. *Quality in Primary Care* 2013;21:1–4.
- Gillam S and Siriwardena AN. Frameworks for improvement: clinical audit, the plan–do–study–act cycle and significant event audit. *Quality in Primary Care* 2013; 21:123–30.
- Siriwardena AN and Gillam S. Understanding processes and how to improve them. *Quality in Primary Care* 2013;21:179–85.
- Balestracci D. *Data Sanity: a quantum leap to unprecedented results*. Medical Group Management Association: Englewood, CO, 2009.
- Donabedian A. Evaluating the quality of medical care. *Milbank Memorial Fund Quarterly* 1966;44(Suppl):206.
- Plsek PE. Quality improvement methods in clinical medicine. *Pediatrics* 1999;103:203–14.
- Field MJ and Lohr KN (eds); Committee on Clinical Practice Guidelines, Institute of Medicine. *Guidelines for Clinical Practice. From development to use*. National Academy Press: Washington, DC, 1992.
- Mohammed MA, Cheng KK, Rouse A and Marshall T. Bristol, Shipman, and clinical governance: Shewhart’s forgotten lessons. *The Lancet* 2001;357:463–7.
- Deming WE. *Out of the Crisis*. Massachusetts Institute of Technology, Center for Advanced Engineering Study: Cambridge, MA, 1982.
- Juran JM, Gryna FM and Bingham RS. *Quality Control Handbook* (3e). McGraw-Hill: New York, 1974.
- Siriwardena AN and Balestracci D. Using a common cause strategy for quality improvement: improving hypnotic prescribing in general practice within a Quality Improvement Collaborative. *Quality in Primary Care* 2011;19:283–7.

- 12 Balestracci D. Data 'sanity': statistics and reality. *Quality in Primary Care* 2006;14:49–53.
- 13 Balestracci D. Statistics and reality: part 2. *Quality in Primary Care* 2006;14:111–19.
- 14 Mohammed MA, Worthington P and Woodall WH. Plotting basic control charts: tutorial notes for health-care practitioners. *Quality & Safety in Health Care* 2008;17:137–45.
- 15 Mohammed MA. Using statistical process control to improve the quality of health care. *Quality & Safety in Health Care* 2004;13:243–5.
- 16 Carey RG. *Improving Healthcare with Control Charts: basic and advanced SPC methods and case studies*. ASQ Quality Press: Milwaukee, WI, 2003.
- 17 Staphenurst T. *Mastering Statistical Process Control: a handbook for performance improvement using cases*. Butterworth-Heinemann: Oxford, 2005.

PEER REVIEW

Commissioned; not externally peer reviewed.

CONFLICTS OF INTEREST

None.

ADDRESS FOR CORRESPONDENCE

Professor A Niroshan Siriwardena, Community and Health Research Unit (CaHRU), Faculty of Health and Social Sciences, University of Lincoln, Brayford Pool Lincoln LN6 7TS, UK. email: nsiriwardena@lincoln.ac.uk

Received 8 July 2013

Accepted 31 July 2013