Journal of Perinatology (2020) 40:962–971
https://doi.org/10.1038/s41372-019-0572-x

QUALITY IMPROVEMENT ARTICLE

Measurement for quality improvement: using data to drive change

Munish Gupta[1] · Heather C. Kaplan[2]

Received: 8 July 2019 / Revised: 1 December 2019 / Accepted: 18 December 2019 / Published online: 8 January 2020

Abstract

Measurement is a core foundation of quality improvement (QI), and analysis of data for QI requires distinct approaches and tools as compared with other areas of healthcare. QI efforts can use structural, process, outcome, and balancing measures, and each measure should have a clear operational definition. Data for improvement should be analyzed dynamically over time, with a focus on understanding the variation present in the data. Distinguishing between common cause and special cause variation is necessary to evaluate and guide improvement efforts. Statistical process control tools such as run charts and control charts can be powerful tools to analyze data over time and help understand variation. This article continues a series of QI educational papers in the *Journal of Perinatology*, and offers a review of the use of data and measures to drive improvement.

Introduction

This article continues a series in the *Journal of Perinatology* designed to offer a broad overview of fundamental quality improvement (QI) methods and tools. Previous articles in the series have reviewed developing a roadmap for a successful QI project [1], identifying a QI project [2], basic QI methods including the Model for Improvement and process mapping [3], and advanced QI methods including Lean Six Sigma and planned experimentation [4].

In this article, we review measurement for improvement, and how data should be used to drive QI. As the previous articles in this series have shown, measurement is a core principle of QI frameworks. For example, the Model for Improvement requires that one identifies measures to evaluate the impact of planned changes before considering the change ideas themselves [3, 5]. Similarly, Lean Six Sigma suggests measuring and analyzing performance before considering steps to improve and control it [4]. The central role of measurement in these frameworks, and QI theory more broadly, has its roots in the landmark work of Shewhart and Deming and their foundational efforts around understanding and measuring variation [6, 7].

Measurement can drive several improvement activities, including: (1) assessing current performance; (2) setting goals for future performance; and (3) monitoring impact of improvement efforts and interventions. This review will address the use of data in the service of all of these goals. Our intent is to offer practical concepts as well as tools for measurement and data analysis that can be quickly incorporated into everyday use by improvement teams.

Measures

Measurement for improvement is distinct from other types of performance measurement in healthcare, including measurement for research [8]. Measurement methods for research are generally designed to capture complete data that allow for the most robust analysis, while measurement for improvement is built around monitoring performance over time using only those data elements necessary for evaluation [9].

Within improvement, several types of measures need to be considered. Most frameworks for QI measures begin with the Donabedian model of structure, process, and outcome measures [10]. Many frameworks also add balancing measures [11].

(1) Structure measures describe the healthcare setting and environment. They can include assessments of the physical environment, human resources, and organizational configuration.

✉ Munish Gupta
mgupta@bidmc.harvard.edu

[1] Harvard Medical School, Beth Israel Deaconess Medical Center, Boston, MA, USA

[2] Perinatal Institute and James M. Anderson Center for Health Systems Excellence, Department of Pediatrics, Cincinnati Children's Hospital Medical Center, University of Cincinnati College of Medicine, Cincinnati, OH, USA

(2) Process measures examine the delivery of care, and describe activities of the healthcare system. Process measures capture what we do as healthcare providers.
(3) Outcome measures focus on the impact of care on patients or populations. They can include health status and disease outcomes as well as patient satisfaction and knowledge. Outcome measures capture what the patient experiences.
(4) Balancing measures assess potential negative or unintended consequences of improvement efforts in other outcomes or other parts of the system.

In the Donabedian model, good structures lead to good processes, and good structures and processes lead to good outcomes [10]. While outcome measures are often considered the most important, as they reflect the end result of care for the patient, they can be complex and multifactorial and thus be slow to change, or they may reflect rare events and thus have limited power to show improvement. Process measures are more precise reflections of care delivery and more amenable to change, and are therefore often the primary targets of improvement efforts; however, in those cases, a strong link between process and outcome is necessary. Ideally, improvement efforts will have a comprehensive set of measures that include outcome, process, structure, and balancing measures. A summary of these measure types as well as an example of a measure set that might be used for an improvement project around neonatal respiratory care are shown in Table 1.

After measures are identified, operational definitions need to be developed that clearly describe each measure. An operational definition of a measure is a specific, detailed description that leads to universal agreement on what that measure captures [12]. It is clear and unambiguous, and brings a measure from a general concept to a precise tool [9]. For example, early milk expression may be the general concept for a measure in a QI project to increase mother's milk use in very low birth weight infants; an operational definition for this measure might be the percent of mothers of very low birth weight infants that pumped or hand-expressed breast milk within 6 h of the infant's birth as documented in the maternal or infant's record. Components of an operational definition commonly include the measure denominator, the measure numerator, the source of data, the sampling plan (if relevant), and the measurement frequency. Table 2 demonstrates how the measure concept of compliance with an admission hypothermia prevention bundle is translated to an operational definition.

Understanding variation

QI requires the ability to monitor performance and to measure the impact of interventions toward achieving improvement goals. Two unique aspects of data for improvement are worth highlighting: (1) the importance of data over time; and (2) the need to understand variation in data.

Data over time

QI is inherently time-oriented, and performance must be examined over time. Numerous designs can be used to assess data over time, from observational retrospective or prospective studies to time-series or quasi-experimental designs. Data analysis for these different designs generally falls into two approaches: static, based on description of aggregated data, and dynamic, based on continuous analysis of data over time.

Traditional healthcare statistics methods are primarily static. A population is described at a single point in time with a measure of central tendency (e.g., mean, median) and a measure of dispersion (e.g., standard deviation, range). Two populations are then compared with statistical tests such as chi-squared, *t*-tests, or regression models, and a *p* value is used to determine statistical differences. These methods are applied to time-oriented data from an improvement project by

Table 1 Measure types in quality improvement.

Type	Description	Examples in improvement project targeting neonatal respiratory care
Structure	Measures of the healthcare setting and environment	· Participation of respiratory therapist on rounds · Availability of bubble CPAP in the delivery room
Process	Measures of delivery of care and activities of the healthcare system	· Percent of very low birth weight infants requiring positive pressure respiratory support whose first mode of support was CPAP · Average time to surfactant administration after intubation among very low birth weight infants receiving surfactant
Outcome	Measures of the impact of care on patients or populations	· Percent of very low birth weight infants with bronchopulmonary dysplasia, defined as need for oxygen or positive pressure respiratory support at 36 weeks post menstrual age
Balancing	Measures of potential negative or unintended consequences of improvement efforts in other outcomes or other parts of the system	· Percent of very low birth weight infants with pneumothorax

Table 2 Operational definition of admission hypothermia bundle compliance measure.

Feature	Operational definition
Measure name	Percent of admissions with hypothermia bundle completed
Type of measure	Process
Included population	Infants ≤ 1500 g at birth or <30 weeks gestation at birth Inborn Checklist completed at birth Admitted directly from labor and delivery to NICU
Excluded population	Comfort care only NICU team not present at delivery
Numerator	Number of admissions ≤1500 g at birth or <30 weeks gestation at birth, with compliance on all of the following elements of the hypothermia checklist completed: · Radiant warmer preheated on arrival of NICU team · Room Temp set at ≥77 °F on arrival of NICU team · Unknown, blank, or Temp <77 °F is noncompliant · Infant placed immediately in plastic bag · Temp probe attached and infant placed on servo within 1 min · Hat placed on infant after drying head · Radiant warmer side rails remain up until infant in transporter · Skin temp checked at 5 min and documented on checklist · Transport incubator pre-warmed to 37–37.5 °C · Warm blankets in transporter prior to leaving OR/LDR
Denominator	Number of admissions ≤1500 g at birth or <30 weeks gestation at birth, with completed checklist
Data source	Hypothermia checklist (manual data collection)
Sampling	None (data collected on all infants, estimated 10–30 per month)
Reporting frequency	Monthly

using a before–after analysis, where performance on a measure is compared before and after an intervention.

Before–after analyses, however, are limited in their ability to describe the dynamic nature of measurement over time. Aggregated data, even when compared before and after an intervention, can fail to show important trends that would be visible with more granular data. Performance on a measure reported annually would not show month-to-month changes that could more accurately reflect the impact of an intervention or show improvement opportunities. Before–after analyses also suggest QI is based on single interventions, rather than capturing the iterative, sequential nature of most improvement efforts. Shewhart, Deming, and others promote the use of dynamic displays of data over time rather than static approaches as the most rigorous methods of data analysis for QI [9, 13, 14].

Understanding variation in data

A second core principle of data analysis for improvement is the need to understand variation. All measures monitored over time will show variation; understanding the causes and types of variation is necessary to understand the system and guide improvement efforts. In their efforts to understand and control variation in the telephone and other industries beginning in the 1920's, Shewhart and Deming developed theories of variation and tools to understand variation that became core components of modern QI. Shewhart and Deming identified two types of variation: common cause and special cause [6, 7, 15]. Common cause variation is the natural variation inherent in any process, and would be expected to continue to occur in a consistent fashion if the parameters of that process are not changed. Special cause variation refers to variation that is not typical, and is due to changes or circumstances that were not previously part of the regular process [14]. Special cause variation is often unintended and reflective of unstable processes that are not standardized; however, it can also be intended, such as when it reflects the introduction of a change in the process (e.g., indicates the impact of implementing interventions in a QI project).

An example of these types of variation is shown in Fig. 1, which displays a graph of the time to get to work measured daily. In period 1, the time to get to work varies between 25 and 34 min, while in period 2, the time varies between 50 and 59 min. Graphically, it is fairly evident that the variation within periods 1 and 2 reflects common cause variation, and is likely due to minor differences in factors such as volume of traffic or timing of lights. It is also evident that the variation between periods 1 and 2 is atypical, or reflective of special cause variation, and is likely due to something new in the process such as construction or a change in route.

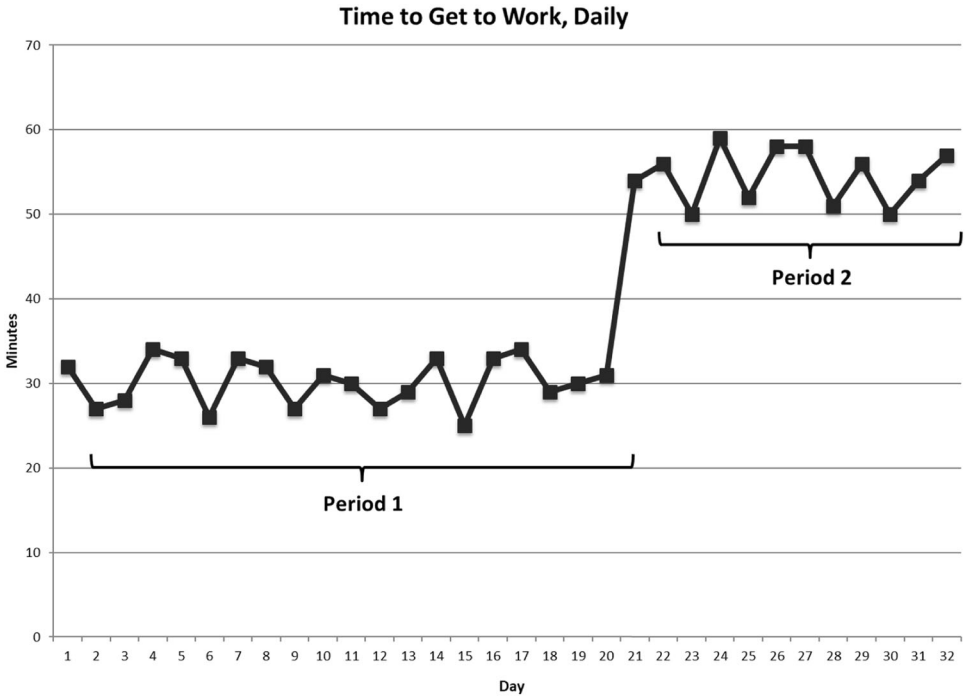

Fig. 1 Example of common cause and special cause variation. Variation within periods 1 and 2 is suggestive of common cause variation; variation between periods 1 and 2 is suggestive of special cause variation.

Importantly, understanding the types of variation present in a process helps shape improvement efforts. Stable processes with only common cause variation can be predicted to produce the same results in the future if that process is maintained; a change in the fundamental process will be needed to achieve a different level of performance. Unstable processes are unpredictable; if the special cause variation is unintended, then efforts should first be directed at identifying and removing the special causes to produce a stable process that can then be managed. If the special cause variation is intended and desired, then efforts should be directed at making that special cause part of the regular process [14].

For most improvement measures, while a graphical display of data over time should be the starting point for analysis, additional tools will be necessary to identify common versus special cause variation. Shewhart and Deming continued their work in variation by developing an approach to data analysis called statistical process control (SPC). SPC focuses on dynamic analysis of time-oriented data, and includes a number of powerful tools for understanding data for improvement. The two most common SPC tools are run charts and control charts.

Run charts

A run chart is similar to commonly used graphs of data over time. The *x*-axis represents time; typically, data are plotted based on a measure of time such as month, week, or day, but data can also be plotted using other time series such as sequential patients or groups of patients. The *y*-axis plots the measure of interest. A run chart also includes a center line, typically the median, which reflects the central tendency of the data. This center line distinguishes a run chart from a simple graph of data over time, and provides the basis for analyzing variation. Run charts also characteristically include annotations of interventions or PDSA cycles and a line indicating the goal of the improvement initiative. An illustrative run chart showing these components is shown in Fig. 2.

Run charts allow for more rigorous and objective interpretations of data then simple graphs over time. Probability-based rules can be used with run charts to identify non-random data patterns. A pattern of data following one of these rules would not be expected in a stable process and is an indication of a signal, similar to Shewhart's concept of special cause variation. If none of the rule-based patterns are noted, then the data likely reflects a stable process with only chance variation or noise, similar to Shewhart's concept of common cause variation [9].

Commonly used rules for detecting signals in run charts are shown in Fig. 3, and are described as follows:

(1) Shift: A shift is six or more consecutive points all above or below the median; values on the median do not count toward a shift, nor do they break a shift.
(2) Trend: A trend is five or more consecutive points all increasing or decreasing; a point that is the same value as the preceding point does not count toward a trend, nor does it break a trend.

(3) Too few or too many runs: A run is a series of points in a row on one side of the median. A point on the median does not break a run. A process with only common or chance variation has a predictable number of runs based on the total number of data points present; widely available probability-based tables provide upper and lower limits for the expected number of runs [16].
(4) Astronomical data point: Astronomical points are ones that are clear outliers from the remainder of the data, such that there is universal agreement that point is an extreme variation. Unlike the first three rules that are objective and probability based, the astronomical data point is a subjective determination, but one on which there should be no disagreement.

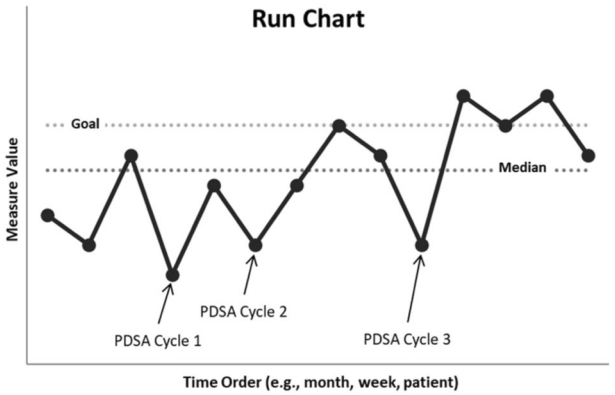

Fig. 2 Components of a run chart.

The presence of a data pattern fitting one of these rules indicates a signal in the data, suggesting atypical variation. This information should then be interpreted by those with knowledge of the system to assess if that signal is a marker of intended process change or if that signal suggests a process that is unstable.

Run charts are simple to create without special software, and can be used with virtually all types of data. They are a powerful QI tool and should be used regularly; however, they are not quite as powerful as control charts.

Control charts

Control charts, also known as Shewhart charts, are the primary tools of SPC. Control charts are similar to run charts, with a few differences. Like run charts, control charts are

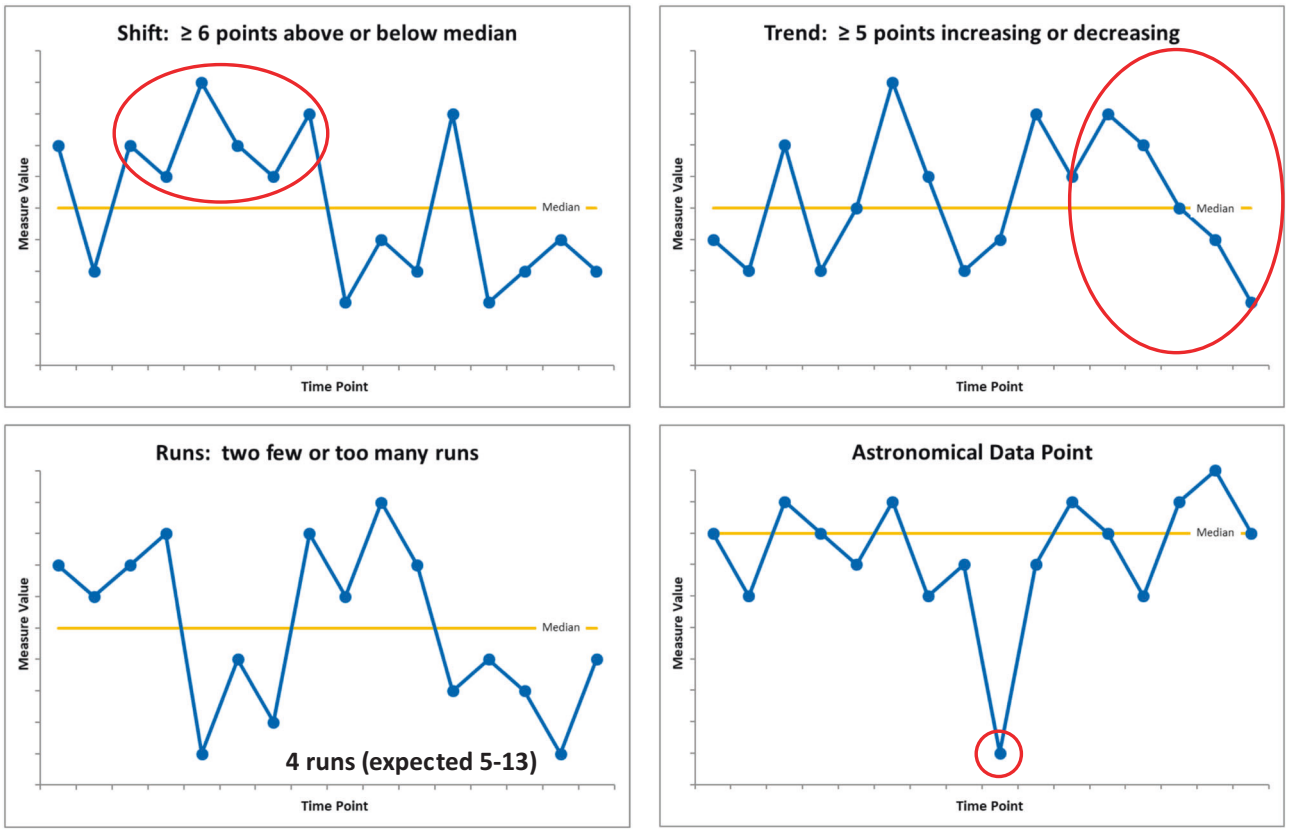

Fig. 3 Rules for detecting signal in run charts.

graphical displays of data over time, with the data plotted on the *x*-axis in time order and the *y*-axis showing the measure of interest. Control charts also have a center line, but rather than the median as used on run charts, the center line is typically the mean. To these elements, control charts add upper and lower control limits. These limits are calculated from the inherent variation in the data; outer control limits are usually calculated at 3 standard deviations from the mean. Inner control limits calculated at 1 and 2 standard deviations from the mean are sometimes shown as well. An illustrative control chart with these features is shown in Fig. 4. Annotations and a goal line can also be added.

Similar to run charts, probability-based rules exist to identify nonrandom patterns in data in control charts. The original rule proposed by Shewhart was one data point outside of the outer control limits, or more than 3 standard deviations from the mean. Over time, as the use of control charts has expanded, additional rules have been developed. At least eight such rules have been proposed, with different authors recommending the use of different combinations of these eight rules [14, 17–19]. The various proposals for individual and sets of rules are designed to balance the risks of misinterpretation of data, with one risk being to falsely detect special cause variation when only common cause variation is present (false positive) and the opposing risk being to fail to detect special cause variation when it is present (false negative). The use of a broader set of rules will minimize false negatives but increase the likelihood of false positives; the use of a more restricted set of rules will minimize false positives while increasing the chance of a false negative. Improvement teams should choose rules prior to the start of an improvement initiative based on the relative importance of minimizing false positives versus false negatives.

The most common set of rules used in healthcare are shown in Fig. 5, and are listed as follows:

(1) One or more data points more than 3 standard deviations from the mean.
(2) Two out of three consecutive points more than 2 standard deviations from the mean.
(3) Eight consecutive points on one side of the mean.
(4) Six consecutive points increasing or decreasing.
(5) Fifteen consecutive points within one standard deviation of the mean.

As compared with run charts, control charts provide greater insight into variation in data, as the control limits offer greater ability to distinguish common cause and special cause variation and identify stable and unstable processes. While signal on a run chart is statistically similar to special cause variation on a control chart, it is more difficult to determine definitively that the absence of signal on a run chart is comparable to lack of special cause variation on a control chart, given that run charts have less power to detect signal or special cause variation than control charts. Thus, determining that a process is stable with only common cause variation is typically done through the use of a control chart.

The creation of a control chart is dependent on accurately calculating the central tendency of the data

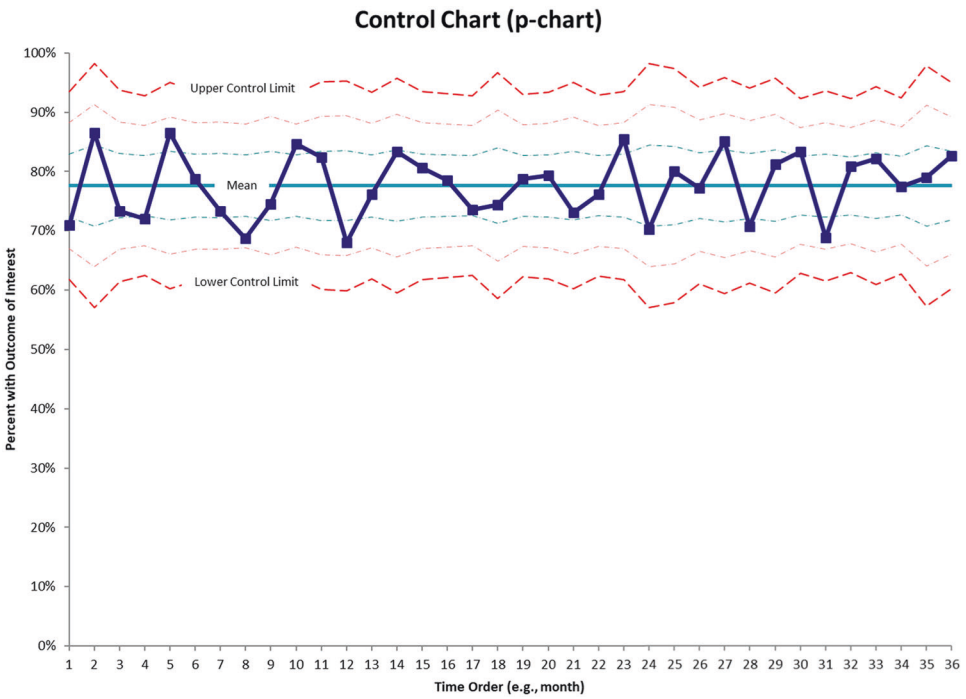

Fig. 4 Components of a control chart.

(the mean) and the dispersion (standard deviation). These calculations are derived from the distribution that underlies the data. Data derived from stable processes can be described by known statistical distributions determined by the type of data. Continuous data are described by a normal (or Gaussian) distribution; discrete dichotomous data are described by a binomial distribution; and discrete count data are described by a Poisson distribution. The type of data, and therefore the underlying distribution, determine the calculations used to derive the control limits. Different types of control charts have been created for each type of data; XbarS and XMR charts are used for continuous data, p-charts are used for discrete dichotomous data, and u-charts are used for discrete count data. For example, admission temperature would be continuous data and could be analyzed with XbarS or XMR charts; percent of infants with bronchopulmonary dysplasia would be discrete dichotomous data analyzed by a p-chart; and rate of unplanned extubations would be discrete count data analyzed by a u-chart. Fig. 6 provides an algorithm to help determine the type of control chart needed for a given type of data [20].

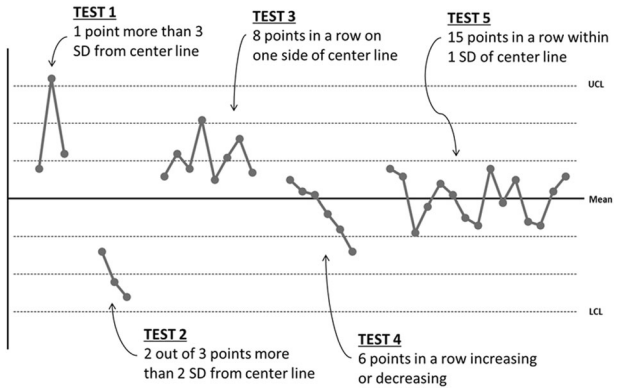

Fig. 5 Rules for detecting special cause variation in control charts.

Of note, while p-charts and u-charts are single graphs of the measure of interest, XbarS and XMR charts are actually two graphs: the Xbar and X graphs show the actual measure of interest, and the S and MR graphs show the variability within each subgroup. Over time, improvement could be reflected by a change in the measure of interest (which would be seen in the Xbar or X graphs) or a decrease in the variability of that measure (which would be seen in the S or MR graphs).

Control charts can be created manually, and references exist that provide appropriate formulas for calculating center lines and control limits [17, 21]. More commonly, statistical software packages are used that can generate all of these types of control charts.

Using run and control charts in QI

Run and control charts, used alone or together, can be powerful tools to guide and drive QI efforts. Control charts, in particular, drive improvement by helping to determine the type of action required based on the type of variation present. A stable process with only common cause variation will continue to perform at the existing level unless fundamental process changes are made. An unstable process with special cause variation is unpredictable; desired special cause variation should be made a standard process, and undesired special cause variation should be identified and removed. Mistaken interpretations of data variation can make performance worse; failure to recognize special cause variation is a missed opportunity to achieve a stable process, and reaction to common cause variation as if it were special cause variation will destabilize a process. Deming called this last result tampering, and considered it one of the greatest dangers of misuse of data for improvement [7].

Below, we offer an example of the use of run and control charts to analyze data for one QI initiative. First, however,

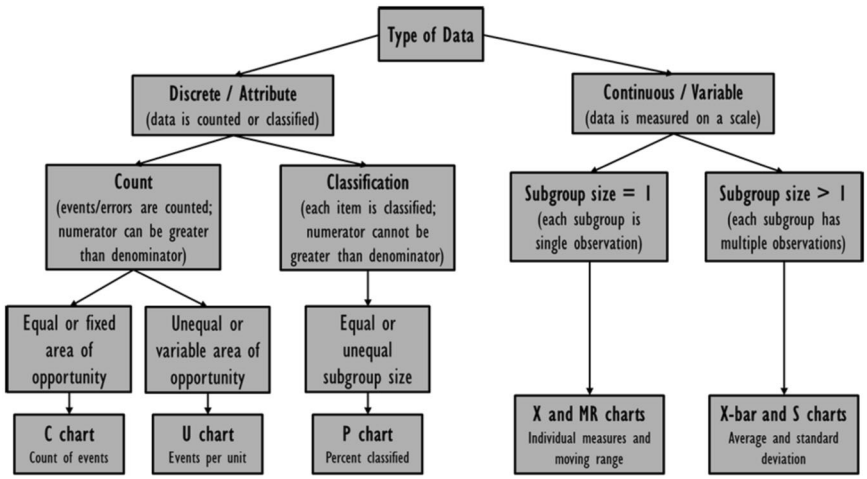

Fig. 6 Algorithm to choose type of control chart based on type of data. Reproduced with permission from Gupta and Kaplan [20]. Adapted from Provost and Murray [17] and Carey [19].

Fig. 7 Run charts and control charts for QI initiative to increase first feedings as mother's milk in NICU infants. a Run chart after 12 months. **b** Run chart after 24 months with signal indicated. **c** Run chart after 24 months with median adjusted and goal line and annotations added. **d** Control chart after 36 months.

several additional concepts related to run chart and control chart use are worth highlighting.

Minimum number of data points

A run chart can be started with just a few data points, or even just with one; however, at least ten data points are needed to establish a reliable median and use the median-based shift and run rules for detecting signal [16]. Control charts require more data points; preliminary "trial" limits can be calculated with as few as 12 data points, but at least 20 points are generally recommended to establish effective control limits [17].

Fixing center line

If a run chart or control chart has adequate data points and suggests a stable process without evidence of signal or special cause variation, and if that data pattern matches knowledge of the system by the improvement team (i.e., the team believes the system has been largely stable), then it can be a good practice to fix or freeze the center line, and compare future performance to this baseline. This approach will allow for detection of signal or special cause variation more quickly than if the center line is recalculated with each new data point. Conversely, if the initial run chart or control chart show an unstable process, or if the improvement team believes their process to be unstable, then center line should be updated and recalculated as new data points are added.

Adjusting center line

When evidence of signal or special cause variation is noted, then it may be helpful to adjust the center line to reflect this change in performance. In order to adjust a center line, at least three factors must be present: (1) evidence of signal or special cause variation on the chart; (2) system knowledge of changes in process that are thought to lead to this change

in performance; and (3) system knowledge that these process changes will persist in the future. A common error in the use of run charts and control charts is to calculate a new center line following the implementation of a change in practice alone, without evidence of change in performance in the data.

Using run charts and control charts in improvement: an example

A QI team aims to increase early expression and early administration of mother's milk for high-risk NICU infants as part of an overall effort to increase mother's milk use. As part of their suite of measures, they are measuring the monthly percent of NICU infants admitted for at least 24 h whose first enteral feeding is their own mother's milk versus donor milk or formula. Their use of run charts and control charts for this effort is shown in Fig. 7.

They begin to collect baseline data on this measure. They start a run chart with their first few data points, recalculating their median with each additional data point, and their chart after 12 points is shown in Fig. 7a. The median on their run chart is just under 55%, there are no shifts, trends, or astronomical data points, and there is an expected number of runs. With no signal on the run chart, and without any notable changes to their process, they fix this median moving forward, and begin testing new change ideas, such as increasing the number of breast pumps available in labor and delivery and asking lactation consultants to meet with high-risk mothers before birth. Their updated run chart after an additional 12 data points is shown in Fig. 7b. They now see evidence of signal with a shift of 9 data points above the median. Believing this is strong evidence of signal that matches their knowledge of system changes, they adjust their center line and recalculate their median to reflect this process change; they also add annotations of their changes and a goal line. This final run chart is shown in Fig. 7c. They recognize that although they have seen improvement, they are not yet at their goal at 70%, and further changes to their process may be necessary to achieve this goal.

Once the team has over 20 data points, they begin to use a control chart to monitor performance and assess for impact of improvement efforts. They determine they need a p-chart because the numerator asks whether each infant received mother's milk as their first feeding or not making this a discrete, dichotomous, variable. Using software, they calculate a p-chart with the first 24 data points and notice evidence of special cause variation in year 2 of the project. They adjust their center line and compare future performance in year 3 to this new baseline. Their final control chart after year 3 is shown in Fig. 7d. Examining this chart, the team determines they saw an increase in their overall performance from 54 to 67% in year 2 of their improvement efforts. If no further changes to their process are made, they predict that around 67% of NICU infants will have their first feeding as mother's milk, with natural month-to-month variation in this number between 50 and 80%.

Conclusions

A famous aphorism in QI is "you can't improve what you can't measure." This quote, with no easily identified attributed source, has been spread widely because the message resonates; measurement is foundational to QI, and all improvement efforts should include a robust measurement framework. Any measurement plan for a QI initiative should address the core concepts discussed in this article, including: an understanding of measures for improvement; selection of a balanced set of outcome, process, balancing, and structure measures, with operational definitions for each; use of graphical displays to show performance over time; and use of run charts and control charts to understand variation in data and use that understanding to correctly guide improvement efforts.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Swanson JR, Pearlman SA. Roadmap to a successful quality improvement project. J Perinatol. 2017;37:112–5. https://doi.org/10.1038/jp.2016.216.
2. Katakam L., Suresh GK. Identifying a quality improvement project. J Perinatol. 2017;37:1161–5. https://doi.org/10.1038/jp.2017.95.
3. Picarillo AP. Introduction to quality improvement tools for the clinician. J Perinatol. 2018;38:929–35. https://doi.org/10.1038/s41372-018-0100-4.
4. Coughlin K, Posencheg MA. Quality improvement methods - part II. J Perinatol. 2019. https://doi.org/10.1038/s41372-019-0382-1.
5. Langley GJ, Moen RD, Nolan KM, Nolan TW, Norman CL, Provost LP. The improvement guide: a practical approach to enhancing organizational performance. San Francisco, CA: John Wiley & Sons; 2009.
6. Shewhart WA. Economic control of quality of manufactured product. Milwaukee, WI: ASQ Quality Press; 1980.
7. Deming WE. The new economics for industry, government, education. Cambridge, MA: MIT Press; 2018.
8. Solberg LI, Mosser G, McDonald S. The three faces of performance measurement: improvement, accountability, and research. Jt Comm J Qual Improv. 1997;23:135–47.

9. Lloyd RC. Navigating in the turbulent sea of data: the quality measurement journey. Clin Perinatol. 2010;37:101–22. https://doi.org/10.1016/j.clp.2010.01.006.
10. Donabedian A. The quality of care: how can it be assessed? JAMA. 1988;260:1743–8. https://doi.org/10.1001/jama.1988.03410120089033.
11. Institute for Healthcare Improvement. Institute for Healthcare Improvement: science of improvement: establishing measures. http://www.ihi.org:80/resources/Pages/HowtoImprove/ScienceofImprovementEstablishingMeasures.aspx. Accessed 18 Jun 2019.
12. Ogrinc GS, Headrick LA, Barton AJ, Dolansky MA, Madigosky WS. Fundamentals of health care improvement: a guide to improving your patients' care. 3rd ed. Oak Brook Terrace, IL: Joint Commission Resources; 2018.
13. Provost LP. Analytical studies: a framework for quality improvement design and analysis. BMJ Qual Saf. 2011;20 Suppl 1:i92–6. https://doi.org/10.1136/bmjqs.2011.051557.
14. Benneyan JC, Lloyd RC, Plsek PE. Statistical process control as a tool for research and healthcare improvement. BMJ Qual Saf. 2003;12:458–64. https://doi.org/10.1136/qhc.12.6.458.
15. Berwick DM. Controlling variation in health care: a consultation from Walter Shewhart. Med Care. 1991;29:1212–25.
16. 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. https://doi.org/10.1136/bmjqs.2009.037895.
17. Provost LP, Murray S. The health care data guide: learning from data for improvement. San Francisco, CA: John Wiley & Sons; 2011.
18. Wheeler DJ, Chambers DS. Understanding statistical process control. Knoxville, TN: SPC Press; 2010.
19. Carey RG, Stake LV. Improving healthcare with control charts: basic and advanced SPC methods and case studies. Milwaukee, WI: ASQ Quality Press; 2003.
20. Gupta M, Kaplan HC. Using statistical process control to drive improvement in neonatal care: a practical introduction to control charts. Clin Perinatol. 2017;44:627–44. https://doi.org/10.1016/j.clp.2017.05.011.
21. Benneyan J. The design, selection, and performance of statistical control charts for healthcare process improvement. Int J Six Sigma Compet Advant. 2008;4:209–39. https://doi.org/10.1504/IJSSCA.2008.021837.