



3

Measuring to Improve in Quality Improvement

Sopnil Neil Bhattarai

We have heard the phrase, “That which cannot be measured cannot be improved” often attributed to Peter Drucker, the inventor of modern business management. Drucker is considered one of the greatest management thinkers of all time [1], and his lessons about measurement can be applied to quality improvement (QI). In QI, measures serve as pulse checks in a system, based on which a healthcare practitioner can understand how the system is performing. The Model for Improvement asks, “How will we know a change is an improvement?,” and this is where measures come into place. There are three main measure types: (1) outcome, (2) process, and (3) balancing. Before we look at measures, we need to understand how measures provide context, and for this, we look at the Donabedian model of care.

S. N. Bhattarai (✉)

Department of Performance Improvement, Children’s National Hospital,
Washington, DC, USA

e-mail: Sbhattara2@childrensnational.org

Donabedian Model of Care

Avedis Donabedian was a health systems researcher in the University of Michigan. He is considered the founder of the study of quality in healthcare and medical outcomes research. He is most famous for creating the “Donabedian model” of care [2]. There are other frameworks that exist to improve quality, such as the Bamako Initiative and the World Health Organization framework. The Bamako Initiative focuses on economical ways of defining quality of care, such as focus on effectiveness, efficiency, sustainability, and equity. The World Health Organization framework of quality of care focuses on a philosophical understanding that high-quality healthcare is a universal right. The Donabedian model works on a macroscopic and organization and health systems level. The Donabedian quality-of-care framework states that the impact on health status is directly driven by the structure of the health system and the processes in the form of good medical practices.

How Are Structure and Processes Associated with Outcomes?

Jake Shimabukuro is an American ukulele virtuoso, and his rendition of the Beatles’ “While My Guitar Gently Weeps” catapulted his career and brought ukulele into the forefront of Millennial and Gen Z pop cultures. One cannot go past 100 music videos on YouTube without encountering a ukulele cover of a song.

The ukulele is a simple string instrument belonging to the lute family. It generally has four nylon strings and fits nicely into a child’s hand. Less intimidating than a six-string guitar, the ukulele is a great first instrument to pick up for anyone. I first encountered Jake Shimabukuro’s tenor ukulele late 2006 and thought it was incredible. Between many years of following Jake’s career, and countless YouTube ukulele stars coming out, I never actually picked up the instrument. Eleven years later, however, I finally

concluded it was time. Consider my ukulele journey as self-improvement.

In 2017, my wife went to a 7-day conference, and I had my apartment for myself. After cleaning every inch, organizing all paperwork, and completing all my chores, I finally ran out of excuses. It was time. I looked for a ukulele on Craigslist and found one for a reasonable price. This simple act of starting up and actually getting the instrument was establishing the *structure*.

Without the structure, there is no context. Without the ukulele, I could not start learning the instrument. Sure, I could watch YouTube videos, but without actually picking up and trying the instrument, it would be a futile task. In healthcare, similarly, the structure is how a particular improvement journey starts.

If a hospital wants to look at the rates of *Clostridioides difficile* (*C. diff*), there needs to be context where this improvement work is done. This includes the physical facility, equipment, and human resources. Without a laboratory that can test for *C. diff*, it would be a Himalayan task for the hospital to look at its rates. Without staff that could process the samples, the hospital would need another way; for example, it would need to figure out how to contract out the lab samples. The structure also involves the organization's characteristics, such as staff training. If none of the staff members trained to draw lab samples, the hospital's desire to conduct the improvement work is futile.

Structure can also be establishing systems for improvement work. Just because I picked up my ukulele does not mean that I would become a virtuoso overnight. I signed up for courses online that would provide me with some fundamentals and decided I would practice in the mornings so that I do not disturb my neighbors. In our lab analogy, the "time" could be the time allocated during leadership meetings to discuss the project. It could be venues, such as grand rounds or lectures, where individuals working on the project can report out regularly. Hospitals with established quality outcomes meetings have avenues that can help facilitate discussion of the project.

Donabedian Triad

When taken with the context of measurements, the Donabedian triad depicted in Fig. 3.1 makes perfect sense. The triad has two main bases: structure and process with the apex of the triad as the outcome. The structure and process hence hold up the outcomes. With active venues to discuss project or conduct project improvement activities, appropriate investment in equipment and training (structure), and processes actively engaged in improvement, outcomes will follow. A hospital cannot improve *C. diff* rates without investing in structure.

Outcome Measures

The outcome measure refers to your overall goal or aim. My short-term aim was to learn to play three chords in the ukulele by the end of the week. The outcome measure, in this case, would be

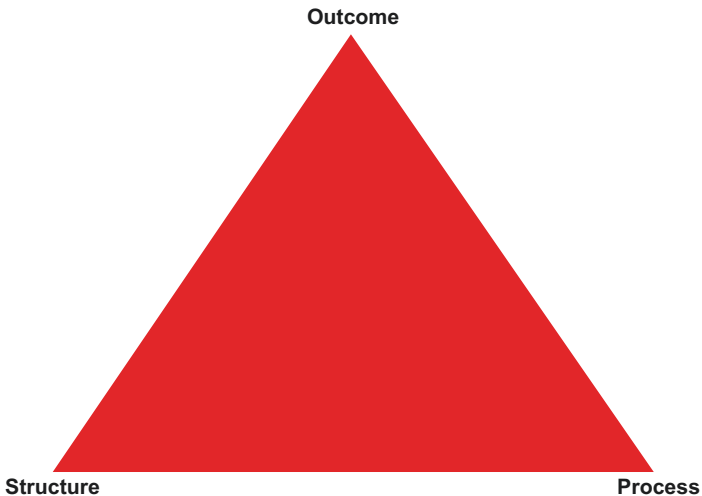


Fig. 3.1 Donabedian triad [2] sets a framework in which outcomes follow structure and processes in systems. Without the structure and processes in place for improvement activities, outcomes are tough, if not impossible, to achieve

whether I learned three chords or not based on standards that I set for myself at the beginning of the project. More discussion about these standards, also known as operational definitions, will be done later in this chapter. In a hospital setting, if you are trying to improve patient and family satisfaction in the emergency department, the outcome measure may refer to satisfaction scores. If measuring overall patient experience in the hospital, one could use a combination of outcome measures, such as health outcomes (morbidity and mortality), with patient satisfaction measures. One must track outcome measures throughout the length of the project, and the sampling strategies will be different than other measures for the project. Our care team at the hospital would need to measure *C. diff* rates every month through the end of the project.

Process Measures

When trying to make system changes, it is helpful to look at the system as a whole. However, a system does not work in isolation. It is essentially a sum of parts working together. An organ system is a network of tissues working together towards the common function. A factory is a system that produces a product. A hospital system works towards improving the healthcare of its patients. In my ukulele example, the system I establish with the structure work together to give me the outcome I desire. However, is learning three chords in ukulele something that happens right away? How do I know that I am headed in the right direction? To understand this, we need to understand the lagging and leading indicators.

Outcome measures are sometimes referred to as lagging indicators. Economists often refer to lagging indicators as “any measurable or observable variable that moves or changes direction after a change has occurred in target variable of interest” [3]. Often, these have occurred after a large shift has occurred in the markets. Some famous economic lagging indicators are the average duration of unemployment, corporate profits, labor cost per unit of output, interest rates, etc. These indicators shift upwards or downwards when some significant events have already occurred. A surge in demand may increase the consumer price index in a

few months after the surge has already occurred. In our healthcare example, the lagging indicator is the *C. diff* rate of the hospital. The final rates of *C. diff*, which could be the number of events normalized by the census, will not be apparent until the end of the month. A hospital's central line-associated bloodstream infection rates will not be apparent until numbers are finalized, which lags the actual event by many months.

In philosophy, there is a type of knowledge called “a posteriori,” which is knowledge that is known by experience. A posteriori is a form of lagging indicator – a change has already occurred, and the knowledge arrives afterward. Process measures are the opposite. Process measures can be referred to as leading indicators. Leading indicators are factors that change before the rest of the economy begins to go in a particular direction [4]. Market observers and policymakers predict significant changes in the economy based on leading indicators. In philosophy, process measures can be referred to as “a priori,” which is knowledge that appeared beforehand, often by reasoning. In my ukulele example, a process measure is an indication of how well I may do in achieving my goals of learning three chords. Process measures help you measure how well your system is functioning to achieve your goals ultimately. In the ukulele example, it would be how often I practice. If I set up a structure (buy a ukulele, sign up for classes, etc.), I will not automatically be proficient in the instrument. I will have to practice in frequent intervals. If I slack and do not practice for a few months, my progress may be diminished. A lack of practice could be an early sign of system failure.

Similarly, in a healthcare setting, a lack of proper antibiotic prescription patterns may give a clue on why patients are developing *C. diff* infections. If a hospital has a protocol, but no way of measuring how the protocol is being followed, i.e., how practitioners are adhering to the *process*, the only signal they would know would be their infection rates. Similarly, if the hospital conducts interventions, such as policy and procedure changes, education, changes to the electronic health system, etc., how would one be able to discern that the practices are changing? If education was effective, and compliance to using new guidelines

is high, there is a high chance that the outcomes will follow. Process measures are a key pillar to the Donabedian framework and serve as pulse checks on how the system is working; thus, it could predict how the system performs in the future.

Balancing Measure

Balancing measures serve as the quantification of unintended consequences in the system. It is important to monitor balancing measures because iterative changes in one part of the system could affect other parts of the system and ultimately could lead to changes in other outcome measures. If a hospital wants to reduce the time patients spend in the recovery room after surgery, leaders should make sure the rates of patients returning to the surgery are not high. In the ukulele example, a balancing measure could be tracking time spent doing other activities. Normally, I work out in the mornings. If playing the ukulele reduces my time working out, this is a negative impact on my overall lifestyle. However, a word of caution: balancing measures do not always have to be negatively impacting the system. A change in the system leads to an increase in patient satisfaction but also impacts staff satisfaction in a positive manner. In this case, increased satisfaction is “good.” Balancing measures are not inherently good or bad.

Operational Definitions

Operational definitions are one of the tenets of quality improvement. If the measurement is not consistent, it is impossible to say whether change over time has occurred. Here are the four rules of operational definitions [5].

1. Gives communicable meaning to the concept.
2. Specifies measurement methods and equipment.
3. Identifies criteria (inclusion vs. exclusion).
4. Has clear and unambiguous wording.

Consider the following example. Hospital-acquired C. diff rates are measured over time as infections per 1000 days of therapy. Days of therapy is a measurement that adds every day that patients are on antibiotics for a time period normalized by 1000. When coming up an operational definition for this measure, you will have to address all points: “we will measure monthly C. diff rates per 1000 days of therapy (concept) in our hospital (inclusion criteria) using CDC criteria (method).”

Data

Healthcare data is a complex topic, and this chapter will only highlight the most important concepts. For detailed discussions on data and how it is used for improvement, I highly recommend *The Health Care Data Guide: Learning from Data for Improvement* [6] by Lloyd Provost and Sandra Murray. This book is essential for further reading.

Creating strong measures sets the stage for an effective data strategy. Data can then be turned into information. Healthcare is no different than other industries in recent times: we are inundated with data. A 2016 estimate suggested that the amount of digital data is set to exceed 1 zettabyte, of which 30% is healthcare. This data represents 30 million times the data in all books that have ever been written [7]. But without context, data becomes useless. Measures help put data into context. In quality improvement work, data shown over time is of utmost importance. Without this tracking, measurement of improvement becomes impossible. Consider Figs. 3.2, 3.3, and 3.4, where a hypothetical shipping company wants to track delays in package delivery before and after implementation of a new software system. Figure 3.2 shows that before the software system, there were a total of 9 hours of delay per package. After the new system, it looks like there were 4 hours of delay per package. This looks like an incredible 56%

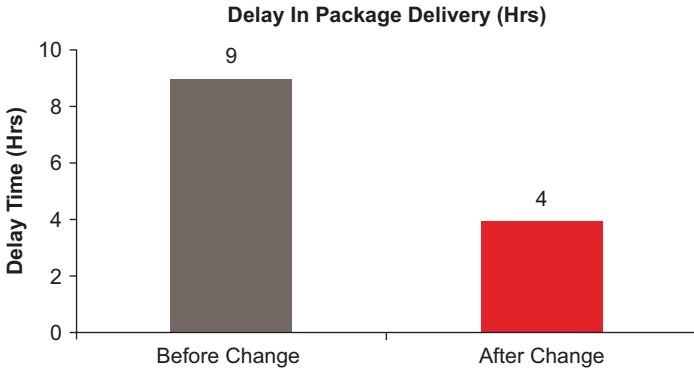


Fig. 3.2 Change is shown as pre- and post-implementation. This practice could be deceptive as they do not show sustained improvement, but rather a snapshot of “cherry-picked” data

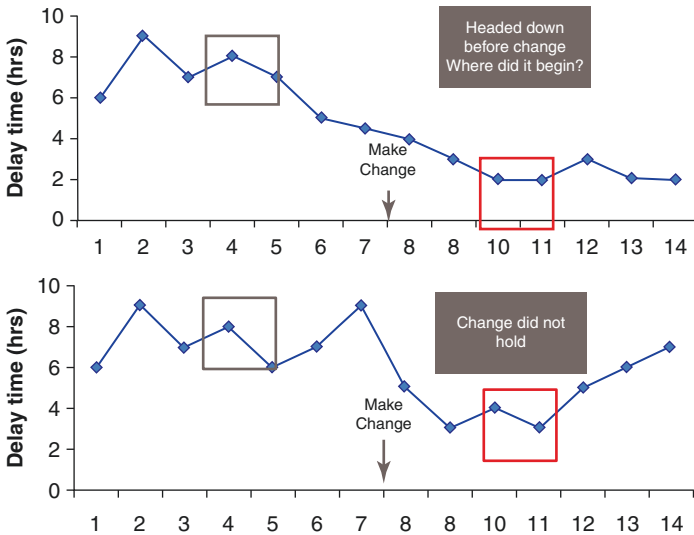


Fig. 3.3 This figure shows data over time and how changes could happen before the implementation of a new process, upon which the system leaders need to analyze whether the improvement was because of the change made or other factors. Alternatively, the lower portion of the data shows that there is an upward swing after the change, which shows sustainment did not occur

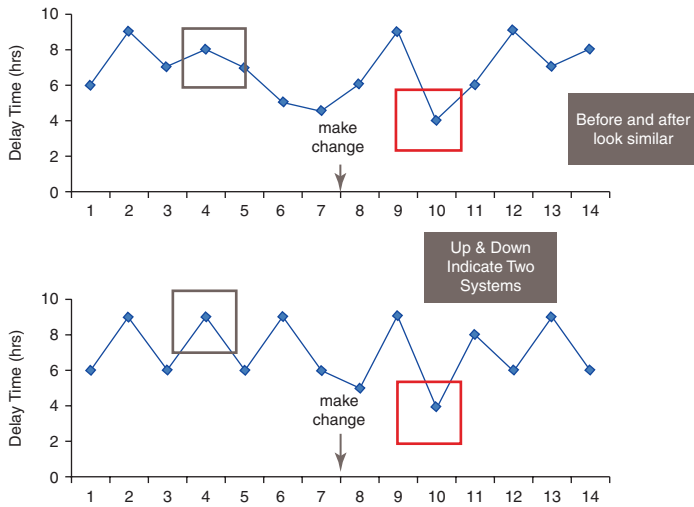


Fig. 3.4 The top image shows that before and after look similar in the data over time, which could mean that the change was not as effective. The second part of the image shows that there could be two distinct systems – the time is lower every other week, which indicates the system is reacting in two different ways

reduction in delays. However, is this the full story? What if we tracked the delay over time? Figure 3.3 shows that when tracked over time, we see two distinct features: (1) the improvement was already headed downwards weeks before the changes were implemented and (2) change did not hold after a few weeks of implementation. In both instances, the company was unable to learn and hence unable to manage nor improve. Data over time shows transparency and can show sustained changes rather than snapshots in systems. In Fig. 3.4 we see other observations: (1) the before and after change data look similar and (2) ups and downs in the data indicate that there could be two systems. When looking at data, as shown in Fig. 3.2, the company can paint the narrative that the software usage successfully decreased the shipping delays. When showing data over time, however, the data becomes more ambiguous. An ideal improvement effort shows data over time and with sustained change.

Sampling Strategies

Effective data collection has a strong sampling strategy – timing is everything. The outcome measure of a project refers to the goal of the project and should, therefore, be tracked throughout the life of the project. Process measures, if applicable to the entire project, should also span the length of the project or at least until there is sustained strong process performance. For our example of *C. diff* reduction, the rates of *C. diff* should be tracked on a regular basis based on what the team agrees upon in the beginning of the project. If the team wants to conduct a small test of change for a subset, then the subset of data should be tracked only until improvement has occurred. For example, if the team wants to track education effectiveness in new fellows, they can test a small group immediately after the education activity was implemented and then a month afterward. The data over time aspect still exists, but the sampling strategy is smaller. For any Plan-Do-Study-Act cycle, the PSDA-level data will always be smaller than the global dataset, which measures the outcome measure.

Common Data Types in Healthcare

There are two main data types in healthcare: attribute data and continuous data. Attribute data can be categorized or be assigned a characteristic. A list of patients with *C. diff* infection is an attribute. Attributes can be as small as a binary characteristic. A list of patients where the appropriateness of their antibiotic regimens is measured have two attribute characteristics: whether they received proper antibiotics (“yes”) or they did not (“no”).

Continuous data include data such as weight, time, and temperature. They can be measured, but they have an infinite number of possible values within a selected range. The easiest way to conceptually think about this is the idea that one cannot “count” continuous data. You would never be able to count weight, time, or temperature.

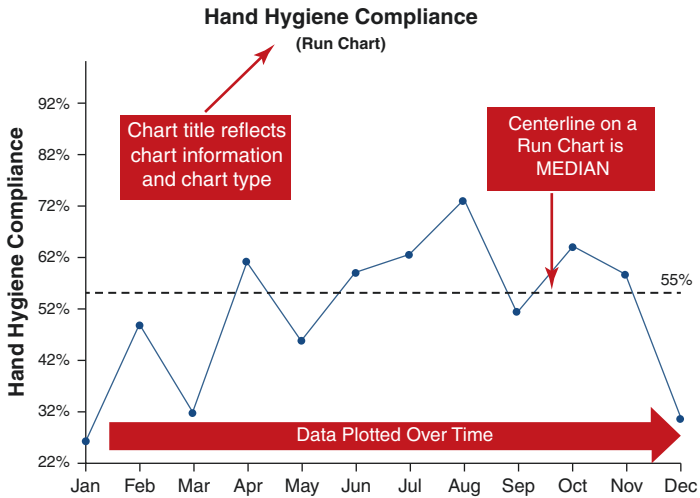


Fig. 3.5 This is a simple run chart. The data are shown over time (monthly). The central tendency or the “normal” of the data are shown as median

The type of data determines what charts one must use. This is especially important in quality improvement, where data over time are measured. The easiest way to display data over time is on a run chart. A run chart and its components are shown in Fig. 3.5. The centerline of the data is the median of all the points in the dataset. The centerline is also known as “central tendency,” which means that any “middle of the road” data for the dataset will fall at or close to the median. The most common type of run chart used in healthcare is known as the control chart. A control chart is an advance mathematical chart that has complex calculations behind it. However, when properly interpreted, control charts can be extremely powerful tools for measuring hospital quality.

Basics of Control Charts

Control charts are based on the assumption that a system’s data are stable. If a hospital fluctuates in its *C. diff* rates every month, there may be a need to stratify, or separate out, individual sections rather than to display the entire data as an aggregate. If the inten-

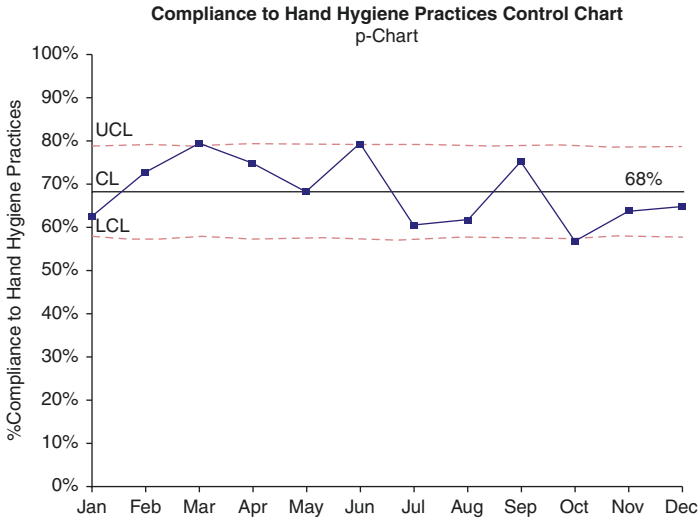


Fig. 3.6 Simple control chart. The central tendency in a control chart is the average of points. The control limits (UCL, or upper control limit, and LCL, lower control limits) fall at 3-sigma or three standard deviations from the middle

sive care units have a higher rate of *C. diff* infections, they could be pulling the central tendency of the entire system higher than usual. Control charts assume the median of the data and the average of the points are close to each other.

Control charts have thresholds known as “control limits.” Control limits are based on the centerline and help project leaders decide what is “normal” for the system. Figure 3.6 shows a basic control chart. The upper control limit is three standard deviations away from the centerline.

Consider the example of men’s shoe sizes. If you were to plot out men’s shoe sizes, the general distribution would say that the shoe size is around 8–8.5. This becomes the central tendency, “middle of the road,” or “average.” On the extremes, however, there are large shoe sizes, such as size 16 or size 5. The extremes of this distribution graph can be considered as probabilities. If you were to randomly pick a man and ask them their shoe size, there is a 68% chance that they would be between 7.5 and 8.5. Similarly, the probabilities of extremely large or extremely small shoe sizes

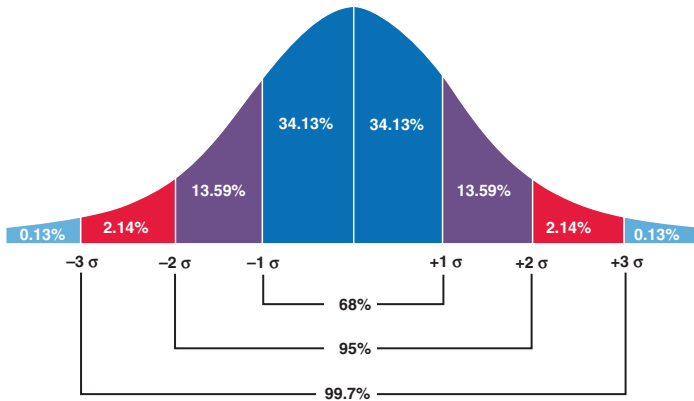


Fig. 3.7 This is a normal distribution graph divided into zones based on colors and standard deviations. The probability of a point landing within the zones is also illustrated. Any points beyond 3-sigma are considered outliers. The probability of those points is low; therefore, if they appear in the data, they must be investigated as “special cause” [8]

are fairly low. If the shoe size is size 24 or size 2, they are considered outliers, extremes, or rare events. Figure 3.7 shows a normal distribution graph.

A control chart is the same data displayed over time. The upper and lower control limits fall at three standard deviations (or 3-sigma) from the centerline. This means that if a data point is beyond 3 sigma, the probability that the data is an outlier is >99% [9]. If that data point occurs, the team must investigate and find out why that has occurred.

Control charts help us understand the concept of common cause and special cause variation. In common cause variation, the processes inherent to the system yield the result the “normal” results for the system. If I survey a normally distributed database of men for their shoe size, I should generally get the shoe size range of 7–8.5. However, what if I surveyed NBA players that are generally taller than the average population? The data would skew in the extremes. In the *C. diff* example, what if there was a dramatic increase in *C. diff* rates in the hospital and it was above the 3-sigma threshold? This would be considered a special cause

variation when data behaves differently than it normally does. Special cause variations can be intended or unintended. Let us say that the team wants to improve compliance to adhering to certain protocols for antibiotics and the team implements forcing functions on the electronic medical system. Providers are unable to select any other options besides the ones chosen by the hospital committee to be the best regimen of antibiotics. This change would be considered an intentional special cause if the adherence dramatically increases beyond 3 sigma.

Special causes can also exist if data over time behave consistently away from “normal.” The most common is a rule referred to as “centerline shift.” A centerline shift implies that the data has deviated enough from the central tendency that the central tendency is no longer useful. Typically a centerline shift happens if eight data points are below the centerline. For example, if your commute to work is typically 45 minutes and never exceeds 1 hour, or is below 35 minutes, you could comfortably say that your average time to commute is around 45 minutes, with an upper limit of 1 hour and lower limit of 35 minutes. If all of a sudden, your commute was closer to an hour every day for 8 days in a row, one could consider that this was a shift. Your new centerline is now around an hour. Why is that the case?

If a coin is tossed once, there is a 50% probability that it falls heads or tails. What if it falls to heads eight times in a row? You would investigate whether this coin was rigged. Similarly, if your commute was an hour instead of 45 minutes every day, there could be something special going on, such as construction or new business opening in your regular route. This example is an unintended variation and a special cause. This is a signal for improvement teams to conduct an investigation.

Control Chart Types

Control charts are based on the data that is being collected. The audience for this chapter is anyone involved in healthcare; as such, it is impossible to be able to adequately address the nuances of

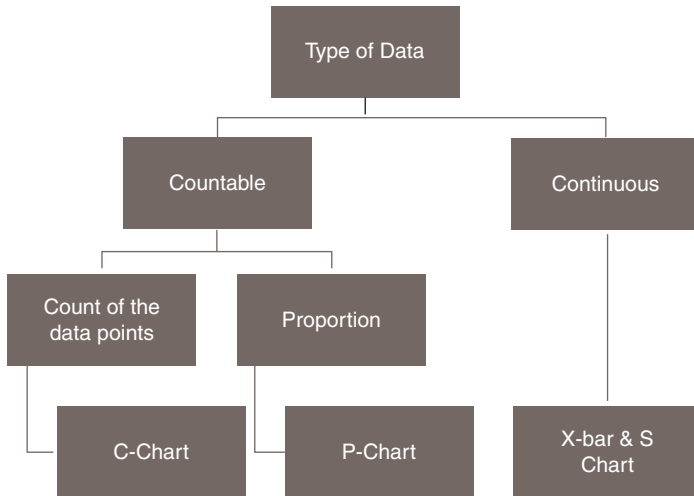


Fig. 3.8 Basic forms of control charts

control charts. The following paragraphs are simple primers to help set the stage for you to be conversant regarding control chart types.

The basic forms of control charts are shown in Fig. 3.8. The most common control chart is a c-chart, which counts data over time. These are helpful when you are just getting started with improvement work. A p-chart stands for *proportion* chart. This chart is typically utilized for compliance. If a protocol is adhered to 60% of the time, a p-chart would reflect such data.

X-charts measure the averages of continuous variables, such as times, temperature, etc. These charts take into account subtle differences between averages within a sample size. The purpose of X-charts is to show stability between points and also between sample of points. X-charts are typically used to measure average turnaround times. S-charts show the differences within standard deviations of each point. It helps users discern how points are different from each other and understand system-level nuances based on that information. If average temperatures fluctuate from 0 °F to 60 °F in 1 day, there may be system instability that an S-chart

could help one determine, since the two data points are so drastically different from each other.

Putting It All Together

The Donabedian triad brings structures and processes to systems to drive desirable outcomes. Structures, processes, and outcomes should be measured, so that leaders can know when improvements lead to change. Using effective data strategies and plotting data over time, leaders can then show that their changes led to measurable improvements in their systems. Using control charts can be an effective way of displaying data over time, and the scientific basis of the control charts is digestible to anyone. Data helps us understand our system and gets us from a posteriori to a priori. The process measures tracked over time serve as leading indicators. If I am successful in learning the ukulele, it is not due to chance, but it is because I implemented structural changes and measure my processes on a continuous basis. In my case, I am no virtuoso, but I can certainly play three chords with ease, which I consider to be baby steps to a YouTube career in case quality improvement does not pan out.

References

1. Forbes. *Forbes.com*. [Online]; 2014 [cited 2020 June 2]. Available from: <https://www.forbes.com/sites/stevedenning/2014/07/29/the-best-of-peter-drucker>
2. Agency for Healthcare Research and Quality. Types of health care quality measures. [Online]; 2015 [cited 2020.10.1]. Available from: <https://www.ahrq.gov/talkingquality/measures/types.html>
3. Chappelow J. Investopedia. [Online]; 2020 [cited 2020 June 2]. Available from: <https://www.investopedia.com/terms/l/laggingindicator.asp>
4. Chen J. Investopedia. [Online]; 2020 [cited 2020 June 2]. Available from: <https://www.investopedia.com/terms/l/leadingindicator.asp>
5. Lloyd RC. Navigating in the turbulent sea of data: the quality measurement journey. *Clin Perinatol*. 2010;37(1):101–22.

6. Provost L, Murray S. The healthcare data guide: learning from data for improvement. San Francisco: Jossey-Bass; 2011.
7. Bhattarai S, RK S. Leveraging real-time data to drive quality improvement in a pediatric health system. *Pediatric Investigation*. 2018;2(3):184–7.
8. NIST/SEMATECH. NIST/SEMATECH e-Handbook of statistical methods. [Online]; 2012 [cited 2020.11.1]. Available from: <http://www.itl.nist.gov/div898/handbook/>
9. Juran JM. Juran's quality control handbook. 3rd ed. New York: McGraw-Hill; 1988.