

## CHAPTER 9

Source: Lloyd. R. *Quality Health Care: A Guide to Developing and Using Indicators* 2nd Edition, Jones & Bartlett Learning, 2019.

# Understanding Variation with Shewhart Charts

## ► Run Charts versus Shewhart Charts

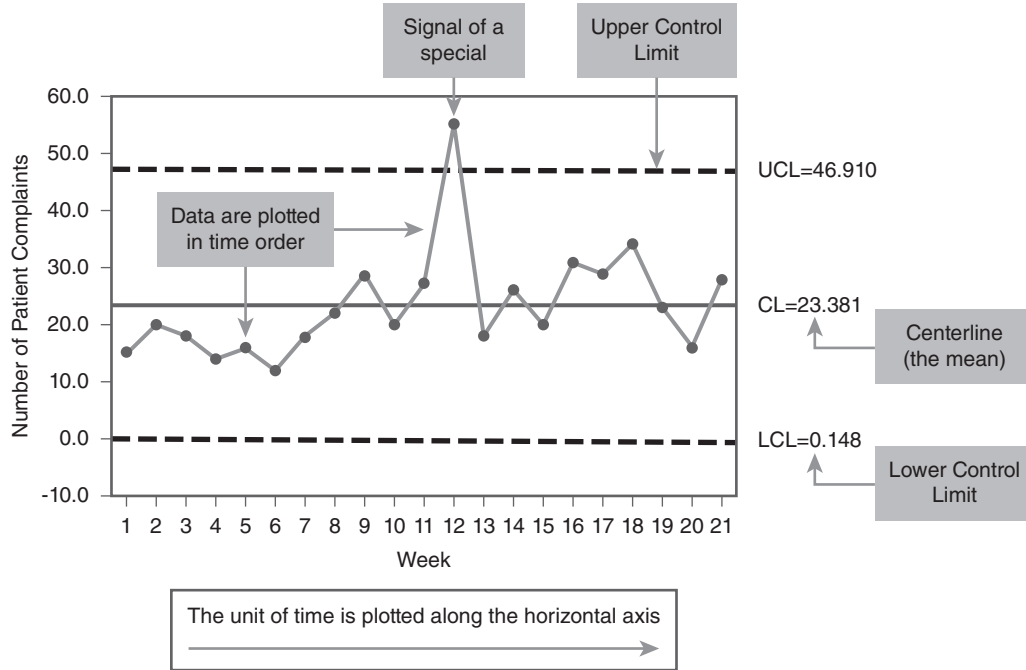
For many teams just beginning their quality measurement journey (QMJ) the run chart provides an excellent starting point. It is easy to construct with paper and pencil, it does not require a software package in order to make one, and it can be used with any type of data (i.e., time, money, counts of errors, percentages, rates, scores, or days between adverse events). Also, the four run chart rules are easy to understand and apply. So, why would I want to use a Shewhart chart instead of a run chart?<sup>1</sup> There are basically three reasons why Shewhart charts are preferable over run charts:

1. Shewhart charts are more sensitive than run charts.
2. Shewhart charts have the added feature of control limits and zones, which run charts do not have.
3. Shewhart charts allow us to more accurately predict process behavior, future performance, and process capability than do run charts.

The details related to these differences are discussed in the remaining sections of this chapter.

## ► What Is a Shewhart Chart?

Like run charts, Shewhart charts are graphic displays of process variation as it lays itself out over time. **FIGURE 9-1** shows the basic elements of a Shewhart chart and one of the tests to identify a special cause (i.e., a data point exceeded the upper control limit [UCL], signaling too much variation in the data, which, by the way, you should recognize as an astronomical data point on the run chart). A run chart and a Shewhart chart look similar in that the indicator of interest and its values are plotted on the vertical or y axis and the chronological order of the data are organized by what are called subgroups (e.g., by individual patients, by day, week, or month) along the horizontal or x axis. The data points are then connected by a line and the mean of the data points is then plotted as the centerline (CL) on the Shewhart chart. The presence of control



**FIGURE 9-1** Elements of a control

limits on Shewhart charts are major points that separate it from a run chart.

Shewhart charts are more sensitive than run charts because the run chart cannot detect special causes that result from point-to-point variation. This is because the CL on the run chart is the median (i.e., the 50th percentile). The run chart basically allows you to classify the data points as being only above or below the median. The actual distance a data point is from the CL is not an issue on a run chart. Therefore, if one data point is 2 units above the median and another point is 22 units above the median, they will both be treated the same because they are both on the same side of the median. The logic for this decision is related to the definition of the median and of a run (i.e., one or more data points on the same side of the median). If these same two data points (i.e., 2 and 22) were placed on a Shewhart chart, however, you would notice a discernable difference because the CL on the control chart is the mean or average of all the

data point actual values. By using the mean we are ensuring that the absolute value and the distance of each data point from the CL will be considered in determining the variation in the indicator and if special cause variation exists.

Another reason why Shewhart charts are more sensitive than a run chart is that Shewhart charts have the added feature of control limits, which run charts do not have. The control limits are properly referred to as the UCL and the lower control limit (LCL). They are also referred to as sigma limits. You will probably hear someone refer to control limits, however, as confidence intervals, confidence limits, or even standard deviation (SD) limits, which they are not (Blalock, 1960; Carey, 2003; Daniel & Terrell, 1989; Provost & Murray, 2011).

The UCL and LCL basically define the boundaries of process variation around the mean. The developer of the chart does not set or define the UCL and LCL. These are determined by mathematical formulae and the width of these limits is dependent on the inherent variation

that lives within the data. The only thing the developer of the chart can place on the Shewhart chart is a target or goal and annotations as to when improvements were introduced.

The control limits enable the Shewhart charts to have increased precision over the run chart. A run chart will miss certain nonrandom patterns that would be detected on a Shewhart chart as special causes. According to Perla, Provost, and Murray (2011, p. 47), “The three probability-based (run chart) rules are used to objectively analyze a run chart for evidence of non-random patterns in the data based on an  $\alpha$  error of  $p < 0.05$ .” This means that run charts could miss a nonrandom pattern in the data approximately 5% of the time. Shewhart chart rules, on the other hand, will not miss detecting a special cause. This is why it is recommended that the terms special and common cause as well as stable or unstable should be reserved for use only with Shewhart charts and that the terms random and nonrandom patterns be applied to run charts.

Shewhart charts also allow us to more accurately predict process behavior and future performance than do the run charts. On a run chart, if the variation is random the best prediction of the future performance of an indicator is the median value. For example, if a team is trying to improve the wait time to see a doctor and have plotted the data on a run chart the median is the best estimate of future performance. Let’s say that the median wait time is 27 minutes. If you were presenting this data to a team or a committee all you could say would be, “Ladies and gentlemen, the median wait time is 27 minutes. The process reflects only random variation. Therefore, if we do nothing to change the current process we can expect to have patients wait about 27 minutes to see the doctor.” On a Shewhart chart, however, because we have the UCL, LCL, and the mean as the CL, we have more precision. In this case, when you present the data to the team or a committee you would be able to say, “Ladies and gentlemen, the average wait time to see a doctor is 35 minutes, the

UCL is 47 minutes, the lower control limit is 23 minutes and there are no special causes detected. This means that the process is a stable and predictable. Therefore, if we do nothing to change how this process works we can predict that patients will wait on the average 35 minutes with the possibility that the wait time could go up as high as 47 minutes or as low as 23 minutes. In light of the target of having all patients seen by their doctor within 20 minutes or less, however, you can see that we have our work cut out for us!”

This scenario provides a summary of how process capability for the wait time in a clinic can be based on the parameters calculated for a Shewhart chart (i.e., the UCL, LCL, and mean). Classically, process capability is defined as, “The calculated inherent variability of a characteristic (indicator) of a product or service. It represents the best performance of the process over a period of stable operation” (ASQ, 2005, p. 78). Process capability is essentially aimed at determining whether under current operating conditions the process can meet the predetermined specifications or achieve the target or goal we have established (Blank, 1998; Carey, 2003; Kume, 1985; Provost & Murray, 2011; Western Electric, 1985; Wheeler & Chambers, 1992).

Besides a verbal summary of the Shewhart chart parameters using the UCL, LCL, and mean as described previously, process capability can also be defined statistically by “a single number assessment of the ability of the process to meet specification limits on the quality characteristic(s) of interest (ASQ, 2005, p. 78). When you move to the statistical indices that capture process capability it is necessary to have an upper specification limit (USL) and a lower specification limit (LSL), which are then compared to the performance of the process as defined by the UCL, LCL, and the mean.<sup>2</sup> Although these indices have not been used extensively in healthcare settings I believe that they have great utility. We have many physiological tests that have upper and lower preferred levels of performance (i.e., specification limits). These include such indicators as temperature, blood

pressure, hematocrits, neutrophils, white and red blood cell counts, platelets, and clotting factors.<sup>3</sup>

There are many useful books and articles on the statistical theory behind Shewhart charts, how to construct them, and how to interpret the results. I have provided only a brief introduction to the key principles behind the Shewhart charts. This is a very rich field of study that has been developed over the past 100 years. Readers interested in the detailed aspects of statistical process control (SPC) and in particular Shewhart charts should consult the rich variety of books and articles on this topic. Ones I have found particularly useful include Benneyan (2001); Benneyan, Lloyd, and Plsek (2003); Blank (1998); Carey (2003); Carey and Lloyd (2001); Duncan (1986); Ishikawa (1989); Mohammed, Worthington, and Woodall (2008); Montgomery (1991); Provost and Murray (2011); Pyzdek (1990); Western Electric (1985); Wheeler (1993, 1995); Wheeler and Chambers (1992); and Woodall (2006).

If you wish to build a firm foundation in Shewhart charts and SPC in general, I would recommend that you read widely on this topic and read what different authors have written. If one item you read seems too academic or mathematical, read another author's description and use of SPC. As you read more of the literature and different authors at some point there will be a moment when you say, "Okay, I get it." Furthermore, if you do not have a reasonably solid working knowledge of the theory and mechanics of Shewhart charts and how they are constructed, it will be rather difficult to successfully apply them to your improvement work. This becomes even more problematic when people say "No problem with the charts. We have software that makes the charts for us." This orientation creates several problems. Although it is easy to push a few buttons on your computer and "make a chart" this does not necessarily mean it is the most appropriate chart for the indicator you are tracking. More important, the SPC software does not help at all with interpreting what the chart is trying to tell you. The chart can come from the machine in front of you with a keyboard but

the interpretation of the chart and what can be learned from it must come from the dialogue that emerges when people with subject matter knowledge interpret the chart. This requires knowledge not a computer.

## ► Key Questions about Shewhart Charts

There are three basic questions people typically ask as we start them on the road to using Shewhart charts:

1. How many data points do I need to make a Shewhart chart?
2. What is a sigma limit? And, why do I need three of them?
3. Do I apply the run chart rules to Shewhart charts?

Each of these is discussed next.

### How Many Data Points Do I Need to Make a Shewhart Chart?

As soon as the team begins to gather data they should start plotting the data points (dots) on a chart. At first this will simply be a line graph. A run chart requires less data because the median as the CL is not as sensitive to point-to-point variation as is a Shewhart chart. Also the run chart rules start to come into play with different amounts of data. The trend rule can be detected when you have five or six data points. As Provost and Murray point out (2011, p. 87) "a trend will remain a trend no matter the amount of additional data added to the graph." The run chart rules related to a shift and too many or too few runs, however, require more data to be detected. The general rule is that a minimum of 10 data points is necessary to properly determine whether a shift has occurred on the run chart or whether too many or too few runs are present.

When we move to using Shewhart charts more data are usually required because (1) the mean is now used as the CL and the absolute value of

any data point enters into the calculations of the UCL and LCL, and (2) the rules for detecting special causes of variation are more rigorous and precise than the four rules for run charts. But, the simple answer to the question of how much data I need to make a Shewhart chart is . . . it depends. I know some readers will be thinking, “What kind of a lame answer is this? Just tell me how many data points I need to make a Shewhart chart!” Because there are many types of Shewhart charts, which are discussed in the next section, it must be realized that the different charts can be produced with differing amounts of data. The subgroup, that is how you have organized your data along the x or horizontal axis of the chart, is key to determining how much data you need to make a particular type of Shewhart chart. For example, if you want to track the wait time of each patient at a family practice clinic to see the doctor then the subgroup is one patient and the one bit of data for this patient will be her wait time to see the doctor. If, on the other hand, you decide that you want to track wait time by day then the horizontal axis of your chart will have Monday, Tuesday, Wednesday, etc. rather than patient 1, patient 2, patient 3, etc. as the subgroup. Selecting day as the subgroup for a clinic could now provide upwards of 30–40 patients’ wait times as possible observations (bits of data) within a single day. Having multiple data points in a subgroup or only one will play a major role in deciding which Shewhart chart you can make. This is why it is very important to make sure you have a well-thought-out data collection plan. Again, more will be said about these issues in the next section when I discuss the types of Shewhart charts.

All this being said, I do know that many people still want to have at least some general guidelines for organizing their data, so here are a few that I offer to improvement teams as we begin to work on developing Shewhart charts:<sup>4</sup>

- It is usually recommended that you have 20–30 subgroups of data before constructing a Shewhart chart. These data should be pulled from the process when it is stable and predictable.
- If you have less than 20–30 subgroups of data you can still create a Shewhart chart but the UCL and LCL should be referred to as “trial” control limits (Carey, 2003; Carey & Lloyd, 2001; Provost & Murray, 2011). The trial limits can be used for learning but the use of the word “trial” is to remind those using the chart that these limits may change as more complete data are obtained and make the limits more reliable and stable. The issue here is that when you have less than the recommended amount of data (i.e., 12–15 data points) the control limits and CL (i.e., the mean) can change rather quickly and dramatically with the addition of each new data point. With a fewer number of subgroups you also run the risk of committing a type II error (i.e., concluding that the chart indicates no special causes when in fact one or more special causes do exist). When you start to have more than the recommended 20–30 subgroups of data, say 40–50, you run the risk of committing a type I error (i.e., finding special causes by chance alone). Additional detail on the theory and use of the type I and type II error concepts can be found in Carey and Lloyd (2001), Carey (2003), and Provost and Murray (2011). In summary, the underlying question here is how much data do you need to create a reasonably stable distribution? Different disciplines recommend different amounts of data needed to form a distribution (e.g., from only a few data points to over 500) but generally speaking a reasonably stable distribution of data for improvement purposes occurs when you have 20–30 subgroups of data.
- As a general rule I also recommend not using quarterly data for your improvement efforts. There is just too much variation being aggregated in quarterly data to be useful for improvement efforts. A quarter consists of 3 months approximately 90 days and over 2100 hours. During this time a

great deal of variation can occur. So when someone begins making conclusions about the quarterly average or SD you should ask them to provide the actual variation that produced these summary statistics by day, week or at a minimum by month.

## What Is a Sigma? And Why Do I Need Three of Them?

These two questions probably pose the most challenging technical aspects of Shewhart chart construction. Some of you will really enjoy this issue and want to learn more whereas others of you will say, “I really don’t care about this statistical distinction just make the chart that is most appropriate for my indicator and tell me what it means.” Either position is fine. I do not intend to go into great detail on this topic but I do want to frame it properly so that you can decide if you want to learn more or accept the fact that these statistical principles have been discussed, debated, and written about extensively for many decades.

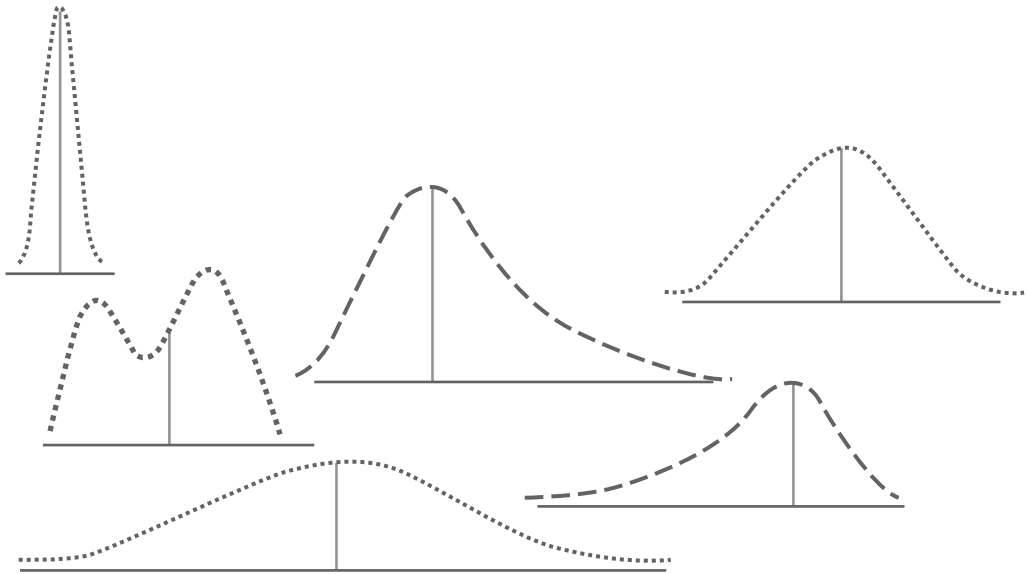
Let’s start with the basics. Whenever you have an array of data you need to consider three characteristics of the distribution these data create: the central tendency of the distribution, the dispersion or spread of the distribution, and the shape of the distribution. You were acquainted with these characteristics when you took your first statistics class, which was probably a number of years ago. So, this should all sound rather familiar even if you have not used the concepts in a while. Measures of central tendency include the mean (i.e., the arithmetic average), the median (i.e., the midpoint of the distribution or 50th percentile), and the mode (i.e., the most frequently occurring number). Measures of dispersion include the minimum and maximum values, the range (i.e., the absolute difference between the min and max values), the sum of the deviations, the mean deviation, the sample variance, and the SD. The shape of a distribution is determined by skewness and kurtosis. A distribution can be symmetrical

(i.e., the normal bell-shaped curve), skewed to the left, or skewed to the right in which case the distribution will have a tail that is longer on one side than the other. A distribution with a long tail skewed to the right will have a mean that is greater than the median whereas a left-skewed distribution will have the mean be less than the median. Kurtosis, on the other hand, refers to how spread out or peaked the distribution is. For additional details on measures of central tendency, dispersion, and distribution shape you can consult any basic stat book. Some of the books I have on this topic come from my undergraduate days and are just as relevant as a statistics book published last year (Blalock, 1960; Daniel & Terrell, 1989; Gonick & Smith, 1993; Levine & Stephan, 2005). These principles are fundamental and have not changed over the years.

**FIGURE 9-2** provides examples of distributions with different characteristics of center, dispersion, and shape. Note that the normal (or Gaussian) distribution, which is popularly referred to as the bell curve, is typically not found in the real world of data collection and analysis. In a theoretical normal distribution, the mean, median, and mode are all at the same position and the data are distributed randomly and symmetrically about the mean. But it needs to be pointed out that not all symmetrical bell-shaped curves are normal (Blalock, 1960, p. 80). You can have, for example, three normal curves that have the same SDs but different means. Similarly, you could have several curves that have the identical means but very different SDs that in turn create different shapes for the distributions.

It is important that you have a comfort level with the characteristics of distributions so that you can more fully understand the characteristics of the data you have collected and their potential limitations. Although the Shewhart charts can accommodate both normally and nonnormally distributed data (Wheeler, 1995) having knowledge of the data you have gathered is the first step toward creating and interpreting Shewhart charts.

With a few of the basics about distributions in hand it is now time to address in a little more



**FIGURE 9-2** Examples of distributions with different centers, spreads and shapes

detail the topic of sigma limits. As was mentioned earlier in this chapter, the UCL and LCL are properly referred to as sigma limits or alternatively, estimates of the SD (Provost & Murray, 2011, p. 115; Wheeler, 1993, 1995). Although some writers will call the UCL and LCL SD limits and not even reference the term sigma, I think it is important to use the term sigma to refer to the estimates of variation in a Shewhart chart rather than SDs for several reasons:

- This is how Dr. Shewhart (1980, 1986) originally described the limits on the charts he developed.
- The SD of a distribution is calculated differently than a sigma. The SD is a single number that represents the average distance any individual data point in a distribution is from the mean. It cannot be a negative number and it will go, theoretically, from zero to a rather large positive value (but generally speaking the SD usually does not go much beyond double digits). A sigma unit, on the other hand, is a “measure of scale for the data” (Wheeler and

Chambers, 1992, p. 60). Wheeler also points out that the common dispersion statistic for a distribution (i.e., the SD) needs to be converted into sigma units by the use of specific formulas. He concludes: “By shifting from measurement units (i.e., SD of a distribution) to sigma units, it is possible to characterize how much of the data will be within a given distance on either side of the average. Thus, the sigma units express the number of measurement units which correspond to one standard unit of dispersion” (p. 61)

A final point to acknowledge is that if you calculate the SD of a distribution using the traditional formula that you will find in many software packages, multiply this number by 3, and then add and subtract this value (i.e., 3 SDs) to the mean, you will get the incorrect UCL and LCL for a Shewhart chart. This becomes even more important when you realize that each type of Shewhart chart has its own formula to compute sigma values and that none of these formulae use

the traditional SD formula for a sample based on the following formula:

$$S_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

$n$  = The number of data points

$\bar{x}$  = The mean of the  $x_i$

$x_i$  = Each of the values of the Data

Note that if you were calculating the SD for a defined population the formula would use  $N$  as the denominator (i.e., the total number of observations in the population being observed) rather than  $n - 1$ , which is typically used when calculating a SD for a sample.

If you do not use the appropriate formula for computing a particular chart's limits, you will produce limits that are too wide or too narrow. This will then lead you to make the wrong decision about the variation in your data (i.e., you will see special causes when they do not exist and miss them when they are actually present).

- The third reason I prefer using the terms sigma units and sigma limits with Shewhart charts rather than SD units or SD limits is to avoid confusion. A majority of healthcare professionals have been exposed to the concept of a SD but not to a sigma. We have all been in meetings where someone presenting data has proudly said, "We have analyzed the data from last month for procedure X and discovered that the average length of stay is 4.3 days and the SD is 2.6 days." When this occurs most participants in the meeting either nod their heads and say nothing or mumble a few words to the person sitting next to them about this is just what we heard last month. The SD is a very popular statistic presented in healthcare management meetings. People hear the number but most could not explain what it is, how it was calculated, or how to interpret it. But, it is a regular part of healthcare management meetings. So, when the word

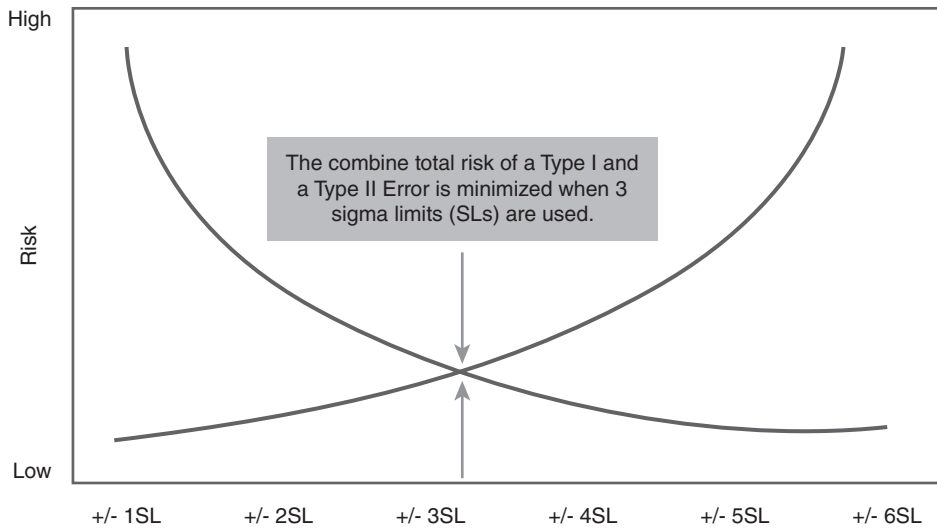
standard deviation is used to describe the units of variation on a Shewhart chart or the UCL and LCL in healthcare settings confusion occurs conceptually and statistically. I would strongly encourage you, therefore, to use the proper terms when constructing and explain Shewhart charts.<sup>5</sup>

Now that we know a little more about what a sigma is, the next question is, "Why do we need three of them?" The answer to this question is found partially in statistical theory and partially in practicality. According to Wheeler (1995, p. 14) Shewhart's use of 3 sigma limits (i.e., three above the mean and three below the mean for a total of 6 sigma units) as opposed to any other multiple of sigma did not stem from any specific mathematical computation. Rather Shewhart said that three "seems to be an acceptable economic value," and that the choice of 3.0 was justified by "empirical evidence that it works" Provost and Murray (2011) provide a succinct summary of the rationale for using Shewhart's 3 sigma limits:

- The limits have a basis in statistical theory.
- The limits have proven in practice to distinguish between special and common causes of variation.
- In most cases, use of the limits will approximately minimize the total cost due to overreaction and underreaction to variation in the process.
- The limits protect the morale of the workers in the process by defining the magnitude of the variations that has been built into the process.

Provost and Murray's point about overreaction and underreaction to variation deserves a few more comments. There are basically two mistakes or errors that you need to avoid when interpreting data. The first mistake (a type I error) is a risk of concluding that a data point requires special action when it is actually reflecting common cause (random) variation. This leads to tampering with a process that is in fact stable and predictable. Tampering (i.e., reacting one way a data point then reacting another way to the next data point when they are part of a process that is stable and predictable) leads to increases variation in a





**FIGURE 9-3** Balancing the risk of a Type I and Type II error

nonlinear manner making things worse. (See the reference to Rule 4 of the funnel demonstration in Chapter 7 for more detail on tampering.) Type I errors happen most often when you decide to use sigma limits on a Shewhart chart that are less than three. On the other hand, the second mistake (a type II error) occurs when you basically do the opposite of a type I error. In this case, you would conclude that a data point indicates no need for action when it fact it reflects a special cause. Type II errors lead to under controlling or what Provost and Murray call underreacting. This happens most often when you decide to use sigma limits that are wider than plus and minus 3 around the mean. As Carey and Lloyd (2001, p. 67–68) point out, “The challenge, therefore, is to balance the risk of tampering against the risk of under controlling. In the first case, you will see special causes when they do not exist, and in the second case, you will miss special causes when they are present. The combined total risk of type I and type II errors is minimized when 3 sigma limits are used.” **FIGURE 9-3** provides a visual of how the total combined risk of two types of error is minimized when the limits are set at  $\pm 3$  sigma. Those of you interested in exploring these issues further should refer to the works of Blumenthal (1993), Deming (1994), Grant

and Leavenworth (1988), Montgomery (1991), Shewhart (1931), Wheeler (1995), and Wheeler and Chambers (1992).

## Do I Apply the Run Chart Rules to Shewhart Charts?

The simple answer to this question is no. The four run chart rules (a shift in the data, a trend in the data, too many or too few runs in the data, an astronomical data point) should be applied only to the run charts. Shewhart charts have their own rules to determine whether special causes are present. These rules are explored next.

## ► Deciding Whether a Special Cause Is Present

Much of the beauty of the Shewhart charts lies in their simplicity. They require just enough data (about 20 data points) to construct a reliable chart, are easy to read, and allow you to determine very quickly whether special cause variation is present in your data. Shewhart charts, according

to Pyzdek (1990, p. 90), “are an operational definition of a special cause,” which I think is a very appropriate way to summarize the purpose of the charts. Shewhart (1931, p. 6) also captured the purpose of the charts nicely when he wrote, “A phenomenon will be said to be controlled when, through the use of past experience, we can predict, at least within limits, how the phenomenon may be expected to vary in the future.” This acknowledges the fact that no one can predict the exact value of the next data point. But, if you understand the differences between a process being in control (i.e., merely random variation) and out of control (i.e., detecting special causes in the data) then you will be well on your way to understanding Shewhart’s notion of prediction within limits. He basically argued that in order to understand the variation in a process you needed to move away from static and aggregated displays of data and look at the process from a more dynamic view by plotting the data over time and understanding the inherent variability in the process. Figure 7-3 (Chapter 7) depicts what Shewhart was recommending.

For decades the Western Electric *Statistical Quality Control Handbook* (1985) has served as the standard reference for the special cause rules. In fact, in many circles and even in several SPC software packages the rules are frequently referred to as the “Western Electric tests for detecting special cause.” Although there are dozens of tests or rules to detect special causes, most experts in the field of SPC maintain that only a few of the tests are essential for a basic understanding of what the charts are trying to tell you. Wheeler (1995) and Wheeler and Chambers (1992) provide excellent summaries and critique of the Western Electric rules and the variations that have been proposed by leading SPC experts. Consider the following passage from Wheeler (1995, p. 139) on this issue:

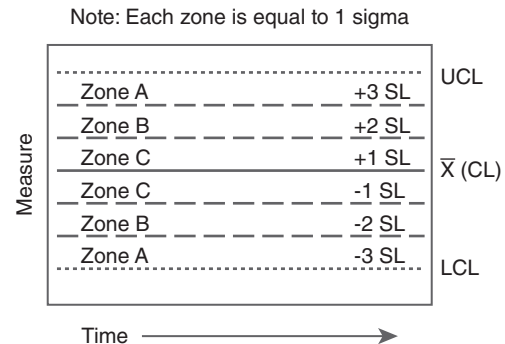
- Shewhart used Detection Rule One.
- David Chambers remarked that “No data set could stand up to the scrutiny of all of the detection rules in the Western Electric Handbook.”

- Irving Burr recommended using no more than Detection Rules One and Four.
- Ellis Ott recommended the use of Detection Rules One, Two, and Four.
- Lloyd Nelson recommends the routine use of Detection Rules One and Four, along with Test 3 (trends) and Test 4 (sawtooth).

The selection of the most appropriate rules, however, should be linked to the subject matter being analyzed, the types of data being collected, and the ability of those who own the processes that produce the outcomes to actually move the relevant indicators in the desired direction.

The application of the rules for special causes to a Shewhart chart begins by dividing the chart into zones. The area between the centerline (CL) (the CL or otherwise known as the mean or average) and the UCL is divided into three equal areas or zones. Because the control limits are referred to as sigma limits, each zone is the equivalent of 1 sigma. The area from the CL to the LCL is divided in a similar manner. These zones are labeled C, B, and A, respectively and emanate outward from the CL. **FIGURE 9-4** provides an example of how a Shewhart chart is divided into six zones. The creation of zones is a very simple process that can be achieved easily with any reputable SPC software program.

A natural or random pattern of data will bounce around across the zones, between the UCL and LCL, and include the following three characteristics:



**FIGURE 9-4** Dividing the Shewhart chart into zones

- Most of the data points are near the CL
- A few of the data points spread out and approach the UCL and LCL
- None of the data points (or at least only a very rare and occasional point) exceeds the control limits (Western Electric, 1985, p. 24)

A natural pattern or random distribution of data will exhibit these three characteristics simultaneously. One of the first signals that a process has special causes, therefore, is the absence of any one of these characteristics.

Because these rules for detecting special causes have grown primarily out of industrial and manufacturing applications, however, we need to evaluate them in light of which rules are most appropriate in health, education, and social services settings. We have done this at the Institute for Healthcare Improvement (IHI) with our colleagues from Associates in Process Improvement (API)<sup>6</sup> and decided that five of the rules for detecting special causes on a Shewhart chart are most appropriate for these disciplines. The five IHI/API rules for detecting special causes on a Shewhart chart are:

Rule 1: 1 point outside the  $\pm 3$  sigma limits

Rule 2: 8 successive consecutive points above (or below) the CL

Rule 3: 6 or more consecutive points steadily increasing or decreasing

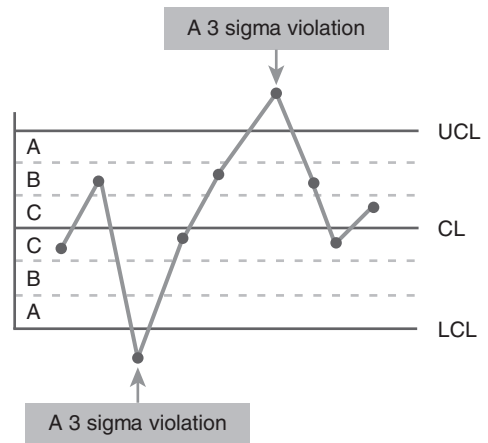
Rule 4: 2 out of 3 successive points in Zone A or beyond

Rule 5: 15 consecutive points in Zone C on either side of the CL

Each of these rules is discussed next.

## Rule 1: 1 Point Outside the $\pm 3$ Sigma Limits (FIGURE 9-5)

This is usually referred to as a 3-sigma violation and is classified as a signal of instability in a process. It is also one of the most easily recognized of all the tests because it is based on a single data point. On the run chart this was a visual determination of the “astronomical data point.” Now we have a



**FIGURE 9-5** Rule #1: A single data point that exceeds the upper or lower control limit

statistical way to determine whether, in fact, it is astronomical. This is the only test that Shewhart used to identify special causes and the reason why Wheeler (1995) stated that “Shewhart used Detection Rule One.” Some texts refer to a single data point that exceeds 3-sigma as a “freak” point (Pyzdek, 1990). Irrespective of the term being used, a 3-sigma violation is a clear signal that the variation of this single point is very different from the variation demonstrated by the rest of the data points on the chart.

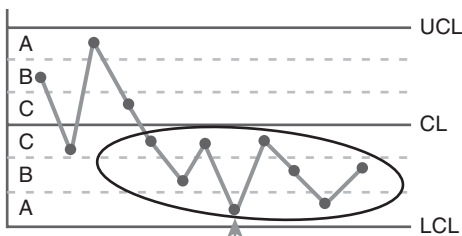
When you detect a 3-sigma special cause do not overreact. The first thing you should do is check the data to make sure that the data point is legitimate. For example, if someone used a different operational definition for this data point it may in fact be a false positive. This data point might also be due to a data collection procedure that sampled the population differently than the other data points. Finally, it could be due to a stratification problem. In this case, data may have been pulled from the afternoon shift when the rest of the chart was based on data sampled from the day shift. My point is that before you see a 3-sigma violation as a true special cause, investigate the methods used to gather that data point. If the data point was based on the same operational definition as the rest of the data and there were no sampling or stratification issues then you do in fact have a special cause that

requires investigation. Why is this data point statistically different from the rest of the data? The presence of a true special cause provides the opportunity for learning.

## Rule 2: 8 Successive Consecutive Points Above (or Below) the Centerline (FIGURE 9-6)

People generally find it easy to detect a 3-sigma violation (Rule 1). But, as the Western Electric *Statistical Quality Control Handbook* points out (1985, p. 26) the data can reflect instability even when all the data points fall between the UCL and LCL. A shift in the process is one such indication of instability. Most writers refer to this rule as “eight consecutive data points on the same side of the centerline.” When such a pattern is observed, it signals that there has been a shift in the process. Another way to think of this rule is that it reflects a run of data that has lingered too long above or below the mean, which indicates a nonrandom pattern. This test is a variation on the run chart shift rule but you will notice that it requires eight data points in a run whereas the run chart rule required six data points to determine a shift.

This test is one of the original Western Electric four primary tests and it is a frequent signal on healthcare charts. As teams work continuously on improvement strategies, their work typically produces results that are not immediate and dramatic in nature (i.e., the 3-sigma rule) but rather



Too many data points in a row below the centerline signals a downward shift in the process.

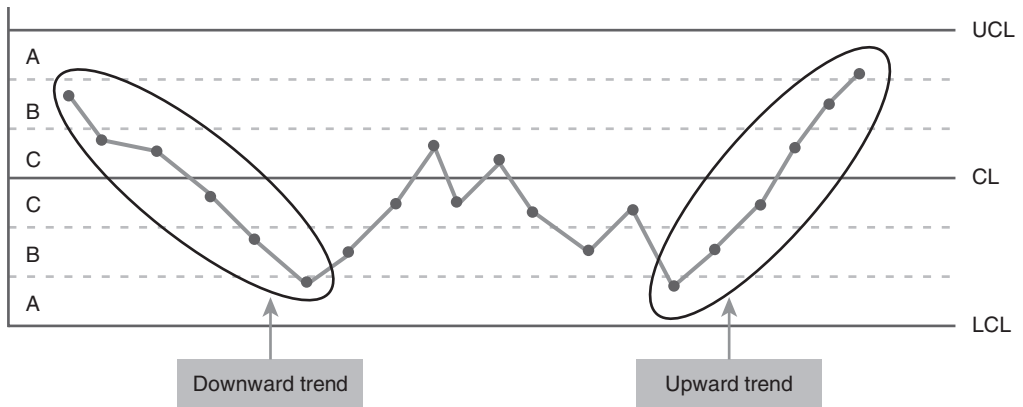
**FIGURE 9-6** Rule #2: 8 successive consecutive points above (or below) the centerline

are observed as a gradual movement of the data over time, which is demonstrated as a shift in the process. Ideally this shift would be in the desired direction but the shift could also be in the opposite direction. The data are neutral. They do not know if they are in the direction of goodness or away from it. This is why it is important that you apply the statistical decision rules that allow you to know when there is a true signal in the data of a special cause and when it is just random variation.

Although the rule of 8 is a classic Western Electric rule you will see other alternatives offered (e.g., 7 in a row, 9 in a row or even one approach that favors a spread of 8 to 10 in a row). Wheeler (1995) lists all the various options defining a shift that have appeared over the years and offers commentary on which ones he has seen used most often. My point in even mentioning these alternatives is that you will hear a variety of opinions on the number of data points used to define a shift and also on what constitutes a trend (rule 3). The challenge is if you define a shift with say seven data points you may see special causes when they do not exist (i.e., a type I error). If, on the other hand, you choose to use 9 or 10 data points as a shift you may fall prone to a type II error, which is missing a special cause when it is present. The rule of 8 has been regarded as a solid practical rule and it is the one I and my colleagues at IHI have decided to use. It is neither too lenient nor too conservative for health and social service application. Unless you like to get into rather heavy statistical theory debates about which rule is “the best” I’d suggest that you accept a set of rules that are practical and appropriate for your work.

## Rule 3: Six or More Consecutive Points Steadily Increasing or Decreasing (FIGURE 9-7)

This rule detects a trend in the data that Provost and Murray (2011, p. 117) define as “a small, consistent drift in a process.” When deciding if a trend exists, duplicate points (i.e., repeating values) should be ignored. This rule engenders considerable debate. First, there is the popular definition of a trend. We



**FIGURE 9-7** Rule #3: 6 or more consecutive points steadily increasing or decreasing

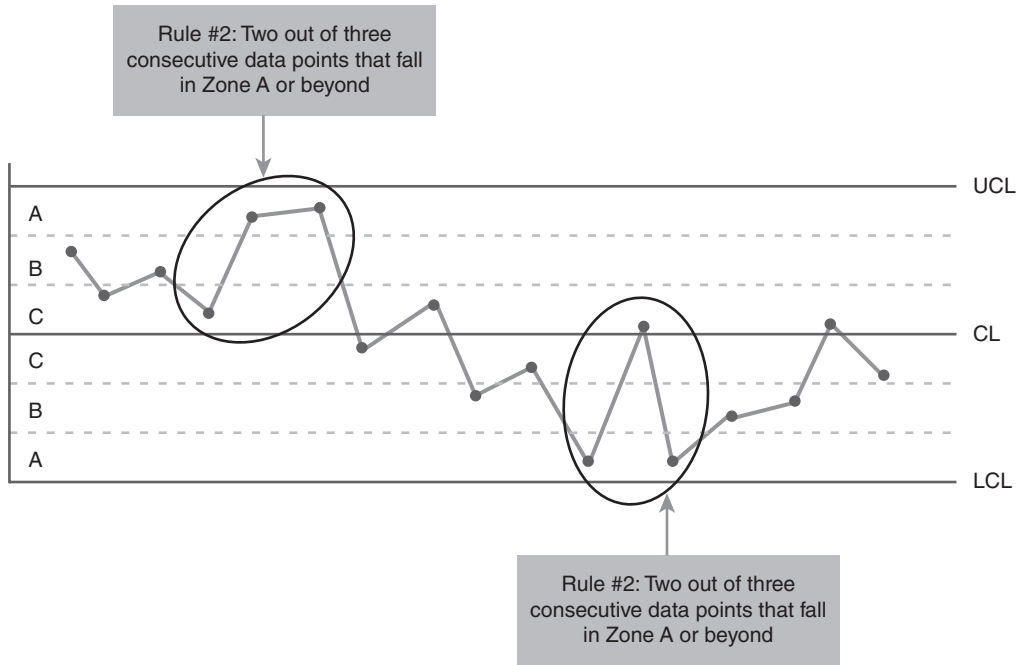
see a trend in fashion, a trend in food, a trend in the stock market, which is usually referring to the fact that the stock market closed higher than it started the day. I regularly hear the weather reporter on the Chicago TV stations referring to a “trend in the temperature.” In this case, the trend is usually a comparison of today’s temperature to the average temperature for the past week or month or the comparison of today’s temperature to the temperature on the same day a year ago. The point is that there are very popular usages of the word trend and then there are statistical definitions. As we analyze Shewhart charts we definitely should be using a statistical definition. But, I have been in many meetings where people interpreting either static or dynamic displays of data have devised their own definitions of a trend. Over the next week make a mental note of how often you hear your coworkers or people in the media look at data and declare a trend is present. People will conclude that there is a “trend” in the data when in fact they are merely comparing two data points. If the second data point is higher than the first and in the direction of goodness then this gets labeled as an upward trend. Deming (1992) had a very good bit of guidance: when you have two data points, “it is very likely that one will be different from the other.”

The Western Electric handbook does not specify how many data points are needed in order to identify a trend. They merely indicate that a trend

is detected when there is “a series of consecutive data points without a change in direction.” At the IHI we have decided to use the rule of 6 as initially defined by Nelson (1985) and then by Pyzdek (1990) as a common practical basis for detecting an upward or downward trend in the data. The final point I will make about this particular rule is that like Rule 2 this rule engenders considerable debate. Wheeler (1995, p. 137), for example, states that “all of these tests (for a trend) are problematic.” He offers a number of reasonable statistical principles as to why he maintains this perspective. Others will argue with you about a trend because they are (1) wanting to see a trend, (2) are using a popular definition of a trend, or (3) have some other statistical reference that says their trend is preferred over the one you propose. So, once again, unless you are ready for these debates I would suggest that you accept the rule of 6 as a trend and see how well it fits with your analysis of the data.

### **Rule 4: Two out of Three Successive Points in Zone A or Beyond (FIGURE 9-8)**

Another of the classic Western Electric rules for instability is when two out of three consecutive data points are more than 2 sigmas away from the CL. In this particular case, the single data point not in Zone A or beyond can be anywhere on the



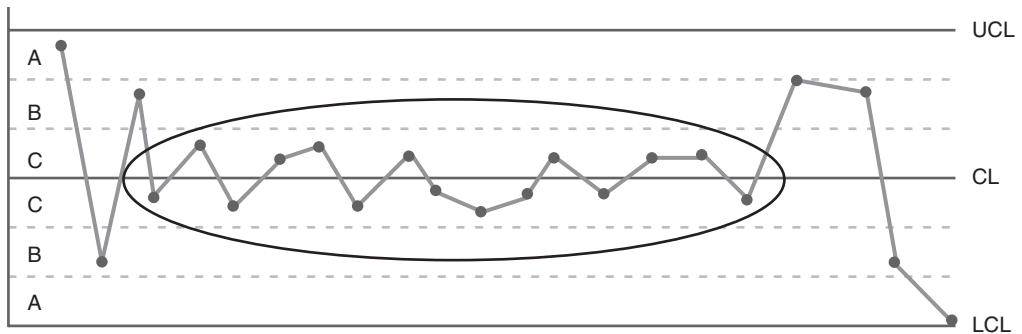
**FIGURE 9-8** Rule #4: 2 out of 3 successive points in Zone A or beyond

chart. The deciding criterion is whether two out of the three successive data points are in Zone A or beyond on the same side of the CL. This is one of the rules that is more difficult to explain in words than pictures. Observing Figure 9-8 will help in understanding this rule. The primary question I get with this rule, however, is “so what?” “Why is it that two out of three data points in Zone A constitutes a special cause?” First, envision the static normal curve. Slightly over two thirds of the data (68.26%) will fall within  $\pm 1$  SD of the mean. When you go out to  $\pm 2$  SDs of the mean you will find 95.46% of the data. This means that by the time you are out to  $\pm 3$  SDs from the mean you should be observing 99.73% of all the data in the distribution. But, because the normal curve theoretically extends infinitely in either direction you do not account for 100% of the data. Now let’s get back to the two out of three data points in Zone A of a Shewhart chart. As you go out the tails of the normal distribution you should expect to see less data. The two out of three rule, therefore, is signaling that you are observing too much data bunching in the

tail(s) of the distribution when you should in fact be observing less and less the further you go out. There certainly are more complex statistical explanations of why this rule detects a special cause. But as Wheeler points out (1995, p. 135) “this rule provides a reasonable increase in sensitivity without an undue increase in the false alarm rate.”

### Rule 5: 15 Consecutive Points in Zone C on Either Side of the Centerline (FIGURE 9-9)

This test is generally described as reflecting an issue with stratification. Stratification usually indicates that two or more different causal systems are present in every subgroup. This pattern of stratification is also known as “hugging the centerline” because there is a run of 15 or more data points within 1 sigma of the CL (i.e., in Zone C above or below the CL) and the variation is relatively small for these data points compared to the width of the UCL and LCL. Stratification



**FIGURE 9-9** Rule #5: 15 consecutive points in Zone C on either side of the centerline

occurs most often because the data collection plan was flawed. For example, you will find a stratification pattern when two separate distributions of data have been collected (e.g., day shift turnaround time was combined with night shift turnaround time) or the sample of data points was drawn from two different distributions of data.

Even though we have these rather specific statistical rules for determining special cause variation that are grounded in decades of testing in manufacturing settings, I think it is important that when we apply these rules to healthcare situations we apply them with a serious dose of common sense. For example, if we are trying to improve food tray delivery time we may be willing to fully accept six data points constantly increasing as a trend. But, on the other hand, if we are dealing with wrong site surgeries we may not want to wait until we have six occurrences of wrong site surgery to declare a trend and then take action. As my colleague Dr. Ray Carey (2003, p. 19) wrote: “When the well-being of patients is at risk, a case can be made for using 2-sigma limits as ‘early warning limits’ or for using 6 rather than 8 points to detect a shift.” In these situations, clinicians would still be looking for signals of special cause so they do not over react to a single data point. But, they would use the data not necessarily to justify changing the system but rather as a basis to investigate potential instability in their process that could cause harm to patients. Wheeler and Chambers (1992) refers to this as having a process on the “brink of chaos.” Statistical decisions must be moderated with and filtered through a healthy dose of common sense and rational thinking.

## ► Deciding Which Shewhart Chart Is Most Appropriate

Although there is only one way to make a run chart, there are numerous ways to make a Shewhart chart. The basic design and look of any Shewhart chart is essentially the same as shown in Figure 9-1 (i.e., data plotted over time, the mean of the indicator as the CL and the calculation of the UCL and LCL). Furthermore, the charts are all grounded in established statistical theory and are all interpreted in terms of the fundamental ideas related to common and special causes of variation. But, there are many different types of Shewhart charts and the user needs to know which one is most appropriate for the indicator being studied. The variety of Shewhart charts is summarized by Benneyan et al. (2003, p. 16):

There are at least a dozen different types of control charts in common use in manufacturing and other industry, with three or four new types being developed each year. The various types differ by the statistic plotted (e.g., averages, percentages, counts, moving averages, cumulative sums, etc.) and the distribution assumed (e.g., normal, binomial, Poisson, geometric, etc.). All have different formulae for calculating centerlines and control limits.

If there are all these different types of Shewhart charts how do you decide which one is the most appropriate for your data? The decision involves two rather simple steps: (1) deciding on the type of data you have collected and (2) deciding which of the various Shewhart charts is the most appropriate for this type of data.

## ► Types of Data

The first step in selecting a Shewhart chart is to determine the type of data you are collecting. There are basically two types of data: (1) variables data (also known as continuous, interval, ratio, or measurement data depending on your background and training) and (2) attributes data (also known as classification or count data). The term used to identify the type of data is a matter of taste and preference. Most SPC books will use the terms variables or continuous data and attributes, classification, or count data. What is more important than the terms you choose to use are the concept the terms are capturing and how you apply them to your data. In this text, I use the terms variables and attributes data as the primary categories. **FIGURE 9-10** provides examples of these two types of data.

**Variables data** can be measured along a continuous scale. In Figure 9-10 this type of data is depicted as money, time, weight, length, and temperature. Consider the ruler as a form of variables data. It has equal appearing intervals that can be divided into as many subdivisions as your calibration instruments will permit. With variable data you can perform all the mathematical function. Data measured this way can be either counts of whole numbers or they can have decimals or fractional parts. Examples of variables data include:

- Wait times in the emergency department (ED)
- Turnaround time for a lab test
- Blood pressure readings
- Newborn weight (measured in grams or pounds and ounces)
- international normalized ratio (INR) and prothrombin times (PTs)

- Blood glucose readings
- The number of procedures or tests performed
- The number of surgeries done each day
- Financial measures such as revenue, operating margin, or expenses
- Duration of a surgical procedure in minutes or hours

**Attribute data** are essentially counts of events that can be placed into discrete categories. Unlike measuring a patient's weight on a continuous scale, attribute data are looking at characteristics that can be classified and placed into categories or "buckets." For example, any time you are measuring mortality you are using attribute data (the patient is either alive or dead). Similarly, pregnancy is an attribute indicator. There are only two outcomes: the woman is either pregnant or not pregnant. A woman does not proudly announce that she is 53.9% pregnant. This is essentially a binomial outcome. Attribute data can be further divided into two subdivisions, *defectives* and *defects*.

**Defectives** (also known as nonconforming units) require that you have a count of the total number of items or events being observed or produced and the number of items from this total that were not acceptable. The unacceptable items or events become the numerator and the denominator is the total number of items or events observed. When you know how many items out of the total are unacceptable you can either plot the number of defective items on your Shewhart chart or you can compute the percent of defectives. When we compute a percentage, therefore, we are basically determining what proportion or percentage the numerator is of the denominator. The standard terminology used in most SPC books to define defectives is that you know both the *occurrence* of the defective product or service (the numerators) and the *nonoccurrences* (defectives plus nondefectives which when added together form the denominator). Knowing these two pieces allows you to calculate a percentage or proportion of defectives. Keep in mind that when you use percentages you are comparing the same types of items, products, or services. If you are looking at the percentage of food trays delivered late to the



## Variables Data



© Ultrashock/Shutterstock



© Pedjami/Shutterstock



© Butterfly Hunter/Shutterstock



© Paul Velgos/Shutterstock



© Lipskiy/Shutterstock

## Attributes Data

**Defectives**  
(occurrences plus  
non-occurrences)  
Nonconforming Units

© HeinzTeh/Shutterstock



**Defects**  
(occurrences only)  
Nonconformities

**FIGURE 9-10** Examples of variables and attributes data

patient, for example, you will have the number of late food trays as the numerator and total number of food trays produced as the denominator. In this case, you have trays divided by trays—like divided by like. The only attribute that is different for classification purposes is whether the food tray was delivered late to the patient. This is an important distinction to keep in mind because as we move next to define defects, this condition will not hold. In summary, data classified as defective can be divided into one of two categories (i.e., a binomial situation) when you know both the occurrences and the non-occurrences of an event. Examples of this form of classification include conforming/not conforming to standards, harm/no harm, go/no-go, pass/fail, OK/not OK, complete/incomplete or present/absent.

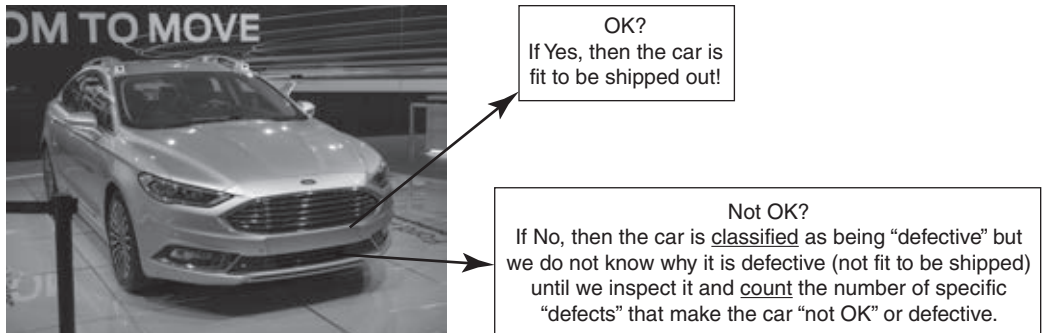
**Defects** pose an interesting challenge. Defects occur and can be counted. But, how do you count all the nondefects? Stated differently, you know when a defect occurs (the occurrence of an event) but you do not know when the nondefects or nonoccurrences happen. I know, at this point you are thinking, “This makes no sense.” When I first heard this statement it did not make a lot of sense to me either. Examples should help to clarify this concept. Look down at the rug in your office or in your family room. How many spots, stains, or snags do you find in the carpet? For argument’s sake I’ll imagine that you found three dirt spots, two coffee stains, and four snags on the carpet. Now, count the number of nonstains on the carpet? How did you do? You cannot count the nonstains or blemishes on the carpet. This is an unknown. Similarly, when the highway department records traffic accidents they do the same thing. They can count how many accidents occurred today on a particular segment of the highway but they have no idea how many nonaccidents there were today. There are times, therefore, when we know the occurrence of an event when the nonoccurrences are unknown and unknowable.

In health care, we experience this situation with patient falls, needle sticks, nosocomial infections, medication errors, and liability cases. We know only when the event happens. Think of needle sticks. A staff member comes into your office to

report that she just stuck herself. After you try to calm her down and explain the next steps you will take, you do not say, “Oh by the way, how many times didn’t you stick yourself today?” Similarly, if a nurse asked a patient, “How many times didn’t you fall today?” she would probably get a rather confused look from the patient.

When you are dealing with defects you need to remember that a count of the number of falls, needle sticks, or medication errors gives you a numerator but you do not have a denominator. So you cannot compute a percentage. So, you either just count the number of defects as whole numbers (e.g., the total number of falls today was eight) or you create a falls rate. A rate is a ratio (i.e., it has a numerator and a denominator just like a proportion or percentage) but the two numbers you are using to form this ratio are not alike. For example, when we compute an inpatient falls rate by month we have the number of inpatient falls (including multiple falls) for the month as the numerator and the denominator is usually the total number of inpatient days for the month. Now we have falls divided by days, two unlike things. The resulting number is reported as so many falls *per* 1,000 patient days. Any time you report that there are so many defects *per* 1,000, 10,000 or 100,000 units (e.g., inpatient days, medication orders, lab tests, or surgeries) you have just created a rate. Note that when you see the little word *per* included with the name of a measure you know that it is a rate and not a percentage. Most of the patient safety indicators as well as epidemiology indicators are constructed as rates (e.g., patient fall rate, restraint rate, surgical site infection rate, ventilator-associated pneumonia (VAP) rate, needle stick rate, or medication error rate).

The other characteristic of a rate is that the numerator of a rate can be larger than the denominator. For example, if you had a 20-bed unit and each patient fell two times you would have 40 falls for 20 patients. What do you call this? Is it 200% falls? No. If you wanted to use a percentage you would have to make a different indicator, which would be the percent of patients who fell *once or more* while they were with us. In this case, we do not care about the total number of falls, which



**FIGURE 9-11** Defectives versus Defects

Ed Aldridge/Shutterstock

includes duplicates. All we are concerned with is at the time of discharge did this patient fall once or more, yes or no? Basically if the indicator is the percentage of patients who fell we do not care if they fell more than once. But, because patients can fall more than once, and we are concerned about this problem, we generally do not use the percentage of patients who fell as a binomial indicator (i.e., fell/did not fall). If we are concerned about the magnitude of the falls and severity we typically track all falls, which means that we have the possibility of having a numerator that is larger than the denominator. When this can occur, we normalize the total number of falls by creating a rate (e.g., 3.2 falls *per* 1,000 inpatient days). A final point about defects is that they usually occur less often than indicators measured by a percentage.

**FIGURE 9-11** provides an easy way to remember the differences between defectives and defects. When cars come off the assembly line they get inspected. If the car is determined to be acceptable by the inspectors it is fit to be shipped to a dealer. But, if the inspectors find one or more things wrong with the car it is not fit to be shipped. In this case, the car would be classified as being defective. This fit to ship determination is a binomial decision: the car is okay or the car is not okay to be shipped. At the end of the shift the inspectors take all the defective cars and provide a summary of why each car was classified as defective (i.e., as a nonconforming unit). This is where the defects come into the picture. The first defective car has a summary report pasted to the windshield. It reads that this defective car has three

defects: one headlight does not come on (defect 1), the driver's side door does not close flush with the body (defect 2), and the driver's seat moves backwards but not forward (defect 3). The entire car is classified as defective but three defects have been discovered. The next car is also classified as defective but it has only one defect (the oil pressure warning light on the dashboard does not go out after the specified period of time). In summary, defects or nonconformities are the specific things that make a product or service defective. Once you understand the distinctions between defectives and defects you will be well on your way to selecting the most appropriate Shewhart chart. To help you in building your skills in differentiating between defectives and defects refer to **EXERCISE 9-1**. For each indicator listed decide if it is describing a defective or defect. The answers to this exercise can be found at the end of this chapter.

## ► Types of Shewhart Charts

After determining whether your data are variables or attributes, the next step is to decide which Shewhart chart is most appropriate for the type of data you have collected. Seven basic control charts are regularly described in the literature and taught in most classes and seminars on SPC. After working with the charts for over 15 years, however, I have found that five of the seven charts are the most

**EXERCISE 9-1** Defective or defect? You make the call!

Indicator	Defective (Classification)	Defect (Count)
1. Number of accidents per 1,000 employee days		
2. Number of errors per 25 food trays		
3. Percentage of acute myocardial infarction (AMI) patients receiving aspirin within 24 hours of arrival in ED		
4. Percentage of inpatient deaths each month		
5. Number of surgical complications per 1,000 surgeries performed		
6. Proportion of hand hygiene observations done incorrectly		
7. Number of falls per 1,000 patient days		
8. Number of medication errors per 10,000 doses dispensed		

relevant and frequently used with healthcare, social services, and educational indicators. I focus on these five Shewhart charts but encourage you to explore the full range of charts as discussed in the Western Electric *Statistical Quality Control Handbook* (1985), Wheeler (1993, 1995), Wheeler and Chambers (1992), Carey and Lloyd (2001), Carey (2003), Duncan (1986), Pyzdek (1990), Kume (1996), and Provost and Murray (2011).<sup>7</sup>

**FIGURE 9-12** presents the Shewhart chart decision tree with the five control charts that have the most relevance to health care, social service and educational indicators. Two of the five charts are used with variables data (i.e., X-bar and S chart and the XmR chart) and three of the charts are used with attributes data (i.e., p-chart, u-chart, and c-chart). Each of the five charts is described next and examples of how to apply the charts are offered. I would suggest, however,

that you read beyond what I summarize in this chapter. As I mentioned earlier in this chapter, by reading the explanations of different authors describing Shewhart charts and their uses you will build knowledge on how to use them with your own improvement efforts.

## ► Defining the Key Terms

Before addressing the details related to each of the reviewing the five basic Shewhart charts shown in Figure 9-12, however, it is necessary to review three key terms that play a critical role in helping you work your way successfully through the Shewhart chart decision tree shown in Figure 9-12. These key terms are subgroup, observation, and area of opportunity and are summarized in **FIGURE 9-13**.

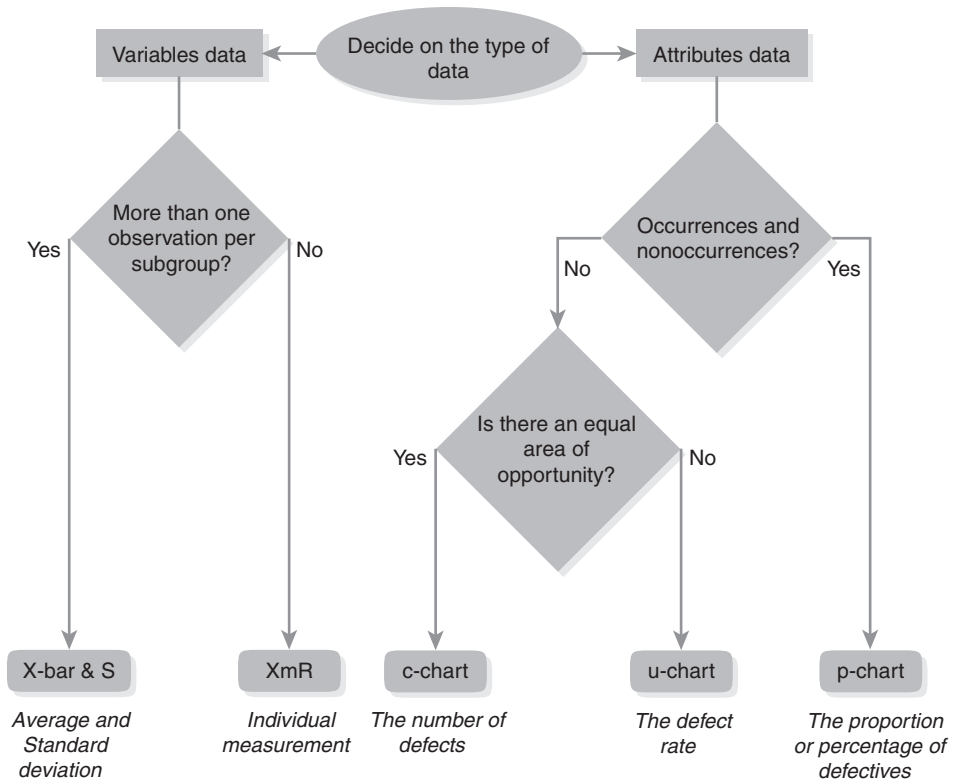


FIGURE 9-12 The Shewhart chart decision tree

Subgroup	Observation	Area of Opportunity
How you organize your data (e.g., by day, week, or month) The label of your <i>horizontal axis</i> Can be patients in chronological order Can be of equal or unequal sizes Applies to all the charts	The actual value (data) you collect The label of your <i>vertical axis</i> May be single or multiple points Applies to all the charts	Applies to all attributes or counts charts Defines the area of frame in which a defective or defect can occur Can be of equal or unequal sizes

FIGURE 9-13 Defining Subgroup, Observation, and Area of Opportunity

## Subgroup

The subgroup defines how you have organized your data and usually captures some dimension of time such as when patients show up for an appointment, day of the week, week, or month. The subgroup will be the label you place on the horizontal or x

axis of the chart. The subgroups will be arranged in chronological order of occurrence. When deciding on a subgroup you should strive to select them so that if special causes exist the chances for differences *between* subgroups will be maximized, whereas the chances for differences due to special causes *within*

a subgroup will be minimized (Duncan, 1986; Montgomery, 1991). The traditional subgroups for Shewhart charts have been:

- An *individual patient* as the subgroup in which case you would order the patients along the x axis of the chart in the order that they presented themselves in the office. Patient 1 arrived at 9:00 a.m., patient 2 at 9:25 a.m., patient 3 at 9:50 a.m. and so on.
- A *day* as your subgroup in which case you would have Monday, Tuesday, Wednesday, etc. across the x axis. Then each day you would select either all of the patients or a sample of them and record their wait times.
- A *week* the subgroup and you would label the x axis as Week 1, Week 2, Week 3, etc. You would then have to decide if you were going to track the wait time for all the patients in a week or just a sample. Usually when a week or even a 2-week period is selected as the subgroup and patient wait time is the indicator of interest you would probably want to draw a sample of the patients. A total enumeration would probably provide more data than you need and create data collection challenges.
- A *month* is a frequently used period of time for a subgroup. But it is not always the best block of time in which to think about improvement or understanding variation. Remember that the Shewhart charts are designed to help you understand the variation in a key process indicator as close to the production of the indicator as possible. In manufacturing, they evaluate products and services on an hourly, shift, or daily basis. Although they may aggregate the key indicators for management reports by month or quarter the ability to improve quality and insure reliability does not come by looking at monthly or quarterly averages. It comes by looking at production almost as it happens. In health care and many social services, there is a strong tendency to track indicators by month or even quarter. This is what I refer to as the tyranny of monthly data. Think of it this way: in a

month we have 30 or 31 days, upwards of 90 shifts in a hospital, and approximately 720 hours in which to deliver care. Why would we want to aggregate all this activity into a monthly average or monthly total? Monthly data frequently lead people down the path of judgment or accountability not quality improvement (QI). In my view, a primary reason we have so many health-care indicators structured around monthly subgroups is that this is how financial and resource allocation systems are organized. In health care, work is being produced every minute of the day not in monthly blocks. Patients are waiting to be seen, have tests performed, or surgery started. Their focus is on minutes or possibly hours not months. Administrators and managers think in terms of months but patients think about the here and now not in monthly aggregates.<sup>8</sup> The other challenge with using month as a subgroup is that the variation in the indicator of interest is usually not visible because the data is aggregated into an average. Although administrators, managers, and policy makers frequently rely on aggregated data and summary statistics to make decisions, no customer, patient, or service user cares about the average. They are concerned about why they or their loved one are not getting service or treatment now. A patient takes no comfort in being told that the average wait time to see the doctor last month was only 49 minutes. Or a physician waiting for her stat lab result to come back will rightfully be irritated if she is told “We don’t have your result quite yet but don’t worry, the average time to get a result last month was only 63.5 minutes.” My point is that although we have a tendency to fall back on making month the subgroup for many healthcare indicators there is no reason to do so. I have made many charts for teams that are collecting monthly data. But each time this happens I make sure we have a discussion about what is the smallest unit of time that we

could gather data on. Using month as a subgroup should be a fallback option not the first choice.

- Finally, it is probably quite evident at this point that it is not advisable to use *quarters* as a subgroup choice because quarterly data represent a very long time period and the variation you are interested in understanding has been aggregated and therefore severely dampened. Quarterly data can lead only to judgment not improvement.

## Observation

As the term implies this is the actual piece of data or the measurement that you record or observe (e.g., turnaround time for a lab test or medication order, blood glucose readings for a patient, or time to administer beta blockers to heart attack patients in the ED). The vertical axis label on the chart defines the observation and the units of measurement along the y axis show the potential distribution of these values. An observation can be classified as either variables or attribute data (e.g., time, money, weight, a percentage of defectives, a count of defects, or a defect rate). For example, if your indicator is wait time to see the doctor in a clinic your observation will be the actual wait time in minutes that occurs from when the patient checks in at the registration desk until she is seen by the doctor. This amount of time will be what gets plotted on the chart. Therefore, the dot on the chart, or the “doink” as I like to refer to it, represents the quantitative aspect of the indicator you are observing during the defined period of time (i.e., the subgroup).

## Area of Opportunity

All Shewhart charts must have a subgroup and an observation clearly defined or the chart cannot be constructed. When we move to the right side of the decision tree (Figure 9-12) and consider the attributes charts, discussed in detail later, a third term comes into play. Notice that the first decision point when dealing with attributes charts is determining whether you have both the

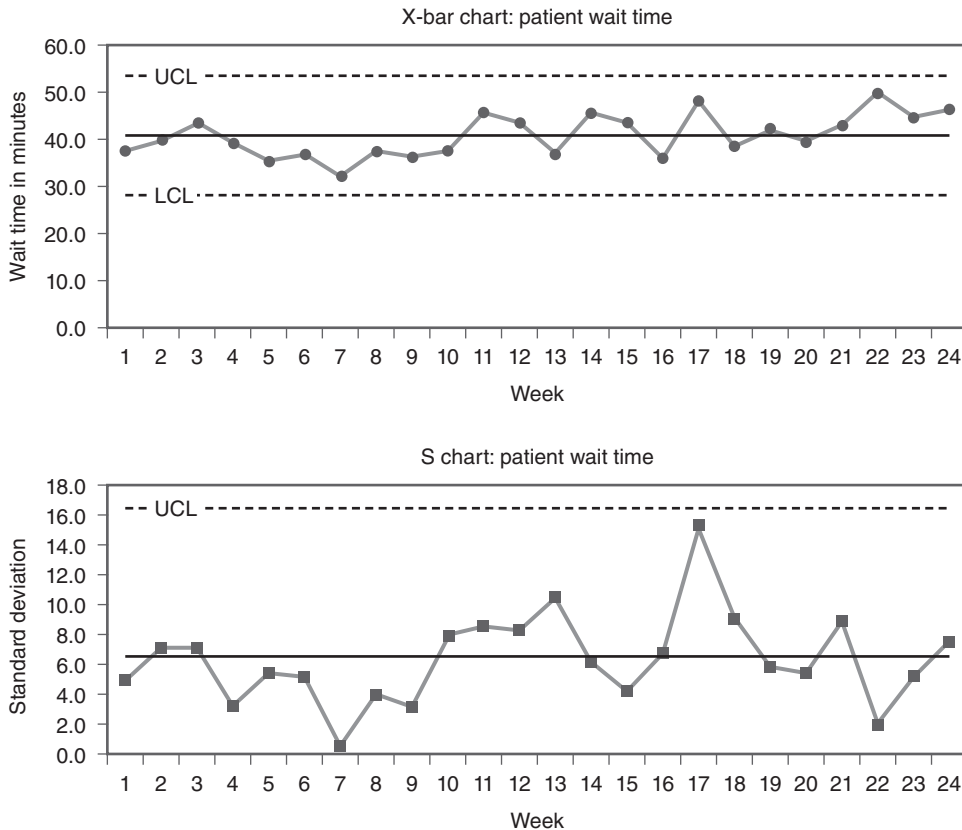
occurrence and nonoccurrences of the events being studied. If you answer “yes” to this question you will be able to calculate either the proportion or percentage of defectives and proceed to make a p-chart. If you respond that you do not have the occurrences and nonoccurrences you are left with having only the occurrence of an event when the nonoccurrences are not known. As was mentioned previously, this gives you a count of the defects and you will make either a c-chart or a u-chart. The decision to make a c- or u-chart is based on your answer to the following question: “Is there a relatively equal area of opportunity for the defect to occur?” If you respond “yes” to this question you will make a c-chart, which is a plot of the number of defects occurring within each subgroup (e.g., a count of the number of falls occurring each day). If you respond “no” to this question (i.e., there is not an equal area of opportunity for a fall to occur) then you would make a u-chart, which would be a plot of the defect rate by subgroup (e.g., 3.2 falls per 1,000 patient days). So, it really does not matter if you respond “yes” or “no” to the area of opportunity question. Consider it essentially a filtering question that will help you select the correct chart for your indicator. As each chart type is explained the use of these terms is demonstrated.

The terms used in the Shewhart decision tree (Figure 9-12) and summarized in Figure 9-13 are not only central to understanding SPC theory but also from a practical perspective, understanding the terms subgroup, observation, and area of opportunity are essential in the operation of SPC software packages. Many of the SPC software packages I have used explicitly ask you to identify the subgroup and the observation or some variant of these terms. Although most software application do not ask you the “area of opportunity” question, understanding this concept is critical to selecting the most appropriate chart for your indicator. With these basic terms in mind, we can start using the Shewhart chart decision tree to understand the conditions that will lead us to select each chart. We will start on the left side of the decision tree and address the variables data charts then move over to the attributes charts.

## X-Bar and S Chart

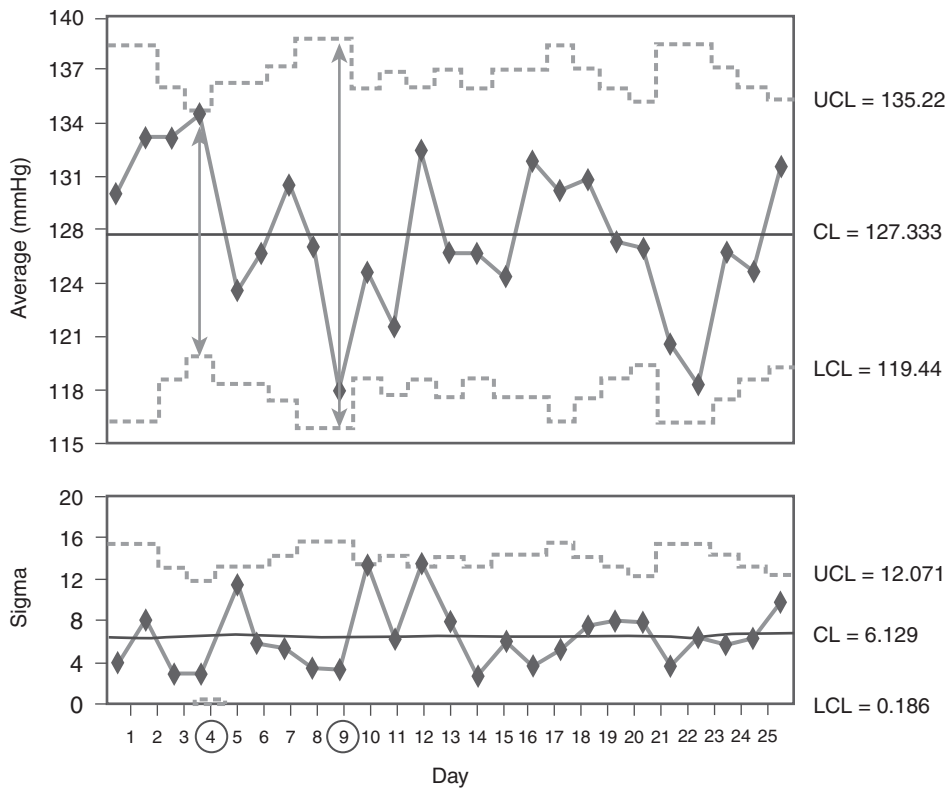
The left side of the Shewhart decision tree (Figure 9-12) follows a pathway to two charts. The first one is referred to as the average ( $\bar{X}$ -bar) and SD (S) chart. It is the most powerful of the five Shewhart charts because it has multiple observations of continuous data that have been organized into subgroups. In this case, the “doink” on the chart (i.e., the plotted dot) has multiple “doinkettes” (i.e., observations) going into it. For an  $\bar{X}$ -bar and S chart the subgroups can be of equal size or unequal size. If the subgroups are of equal size (e.g., a stratified random sample of 15 patients is selected each day and their wait times to see the doctor are recorded) then the UCL and LCL on the chart will be straight lines as shown in **FIGURE 9-14**.

If the subgroups are of unequal size, however (e.g., on Monday, we sample 10 patient wait times, on Tuesday, we had 15, and Wednesday, we collected 20 wait times) the UCL and LCL will not be straight lines. Instead, they will be what are called “stair-step” control limits as shown in **FIGURE 9-15**. With an unequal subgroup size the amount of data varies within each subgroup and so the dots on the chart (i.e., the observations) each have their own individual UCL and LCL calculated. With more data the limits are tighter and with less data in each subgroup the limits are wider as shown in Figure 9-15. Day 4 in Figure 9-15, for example, has tighter limits indicating that there is more data being collected on this day. Day 9, on the other hand, has wider limits due to less data being collected on this day.



**FIGURE 9-14** X-bar and S chart with straight control limits due to equal subgroup sizes





**FIGURE 9-15** X-bar and S chart with stair-step control limits due to unequal subgroup sizes

When you make an X-bar and S chart most software programs will give you the option of producing two charts as shown in Figures 9-14 and 9-15. The top chart is the X-bar chart or average chart and the bottom chart is the S chart or SD chart. The X-bar chart is considered to be the primary chart. Both charts have three main components: (1) the CL or average, (2) the UCL, and (3) the LCL. The X-bar chart will show the average of the data within each subgroup and the lower chart (the S chart) shows you the SD for each subgroup (i.e., each dot) plotted on the X-bar chart. In Figure 9-14, for example, Week 1 on the X-bar chart has an average wait time of 38 minutes and an SD of 5 minutes (seen on the bottom chart). On Week 2 the average wait time is 39.7 minutes and the SD is 7.2 minutes. So for each week we can see what the average wait time is and the amount of spread around that average for this week as measured by the SD.

In Figure 9-15 the indicator of interest is a patient's systolic blood pressure. The patient recorded several blood pressure readings each day (a minimum of three and a maximum of five each day). As a result the subgroups are of unequal sizes, which produces the stair-step control limits. If the patient had recorded *exactly* the same number of blood pressure readings (observations) each day (e.g., four) then the UCL and LCL would be straight lines. As you will see in subsequent examples, stair-step control limits will also be found on p- and u-charts. As was mentioned above, on those days when more data were collected (e.g., Day 4 in Figure 9-15) the control limits will be tighter. On days when fewer data were collected (e.g., Day 9) the limits will be wider.

The upper chart in Figure 9-15 reveals the average systolic blood pressures by day and the overall average. The lower chart shows the SD for each day as well as the average SD across all 25 days.<sup>9</sup> The average systolic blood

pressure shown in the top chart for the patient in Figure 9-15 is 127. Note that the decimal places on the chart can be ignored in this case because this is entirely too finite a reading for blood pressure results. The degree to which you can control the decimal places on a chart will depend on the software being used. The average UCL is 135 whereas the average LCL is 119. Because this chart reveals only common cause variation, the way to describe the performance of this patient's systolic blood pressure this chart is as follows: "On the average this patient's systolic blood pressure is 127. It could go up as high as 135 on any given day or down to 119 and that is the natural rhythm of this patient's systolic blood pressure process."

The lower chart is the S-chart. This chart has two primary purposes. First, it helps you to understand the variation that exists within each subgroup (i.e., day). For example, the SD for Day 4 is around 4. On Day 9 the standard deviation is about 3 mmHg. As you look at each day, therefore, you will see that there is a different average and standard deviation which reflects the variation in this patient's blood pressure over time. The second purpose of the S-chart (i.e., the bottom chart) is that the average SD (the CL) is used in the calculation of the control limits for the average (upper) chart. In this case, the average systolic blood pressure (i.e., CL) is 127. What is important to realize is that if the SD chart reveals wide variation, then the average SD will likewise be large. Because the average SD is used to compute the UCL and LCL, a large SD will also contribute to making the control limits of the top chart (the average chart) wider. The relationship of the two charts must be understood together.

I do not intend to elaborate on the statistical formulae for the calculation of the control limits. It is important to realize, however, that each Shewhart chart has a different set of formulae to calculate the chart's UCL and LCL. For additional details on calculating the statistical parameters for the various Shewhart charts readers should refer to Wheeler (1995), Wheeler and Chambers

(1992), Western Electric (1985), and Provost and Murray (2011).

## XmR chart or the I-chart

The XmR chart is also known properly as the Individuals and Moving Range chart. But it can also be referenced as the Individuals chart or simply the I-chart. The key characteristic of this chart is that each subgroup contains *one and only one individual observation or bit of data* (i.e., the "doink" on the chart has only 1 bit of data and no "doinkettes" as we discovered in the X-bar and S chart). In the Shewhart decision tree (Figure 9-12), this decision point is identified by the question "More than one observation for each subgroup?" When the answer is "no" then the chart of choice is the XmR chart. Like the X-bar and S chart, you will typically get two charts when you request this type of chart from your SPC software. The X chart shows the values for the individual data points as well as the average for all the individual data points. The mR chart documents the "moving range." The XmR chart is typically used when you are interested in answering questions such as:

- How many surgeries do we do each week?
- What is the cost of each knee replacement surgery?
- How long does each patient wait before being seen by the doctor?
- How many home care visits do we conduct each week?
- How many calories do I eat each day?
- What is the length of stay for each coronary artery bypass graft (CABG) patient?

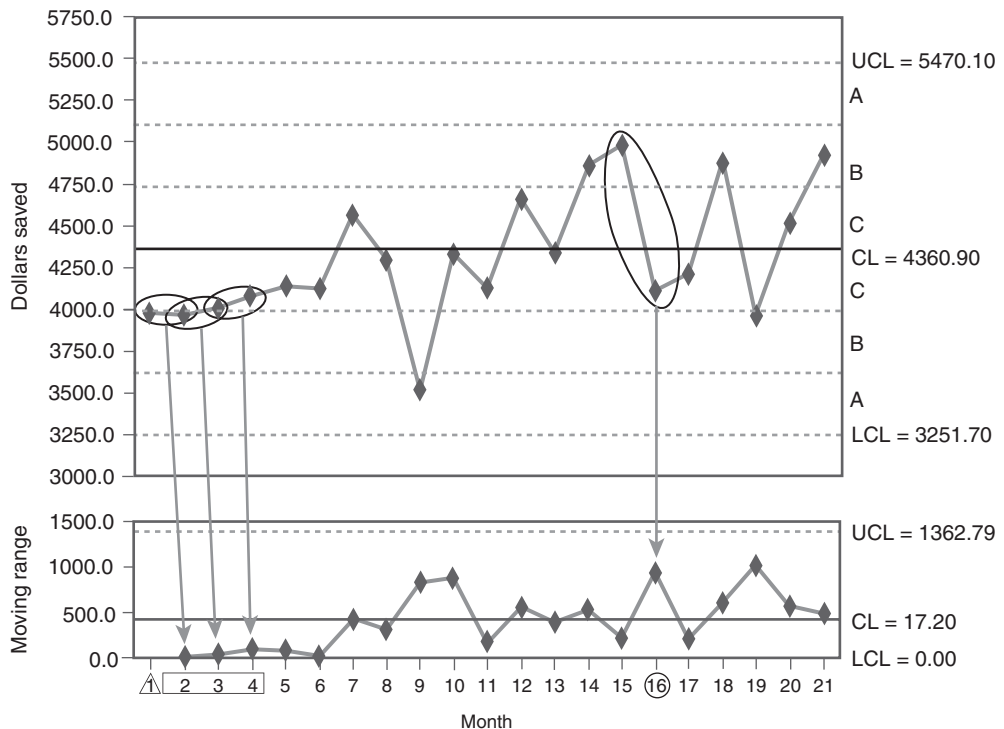
This chart is used frequently to address questions related to volume, frequency of events, or financial issues. Note that you are not interested in finding out what percentage of surgeries started late (this would be considering a late surgery start classified as a defective, which would require a p-chart), but rather you merely want to know how many surgeries are done in the course of a day or a week. In this case, the day or week becomes the

subgroup (x axis label) and the total number of surgeries completed each day or week becomes the individual observation for that week (i.e., the dot on the chart). In short, the XmR chart can be used in many situations. Remember, however, that the indicator being placed on the XmR chart is not being classified as a defective or defect. When you use the XmR chart you will usually be asking a more neutral question such as “How many of procedure X do we do?”

**FIGURE 9-16** provides an example of an XmR chart. In this particular example, the indicator of interest is the total number of U.S. dollars saved each month as a result of implementing a new transcription system for radiology. Note that like the X-bar and S chart there are two charts. The top portion of the chart provides a plot of the individual data points along with the average of all the data points and the UCL and LCL. This chart also has the *zones* identified. These are the

dashed lines that divide the chart into three areas above and below the CL. As was discussed earlier in this chapter, the zones are used to assist in identification of special causes. Typically, and you will see exceptions to this point, the zones are used when you have a chart with equal subgroup sizes.

The bottom chart is referred to as the *moving range chart*. The moving range is derived by calculating the simple difference between two successive data points on the Individuals chart and then plotting this difference on the mR chart. This is also referred to as creating an artificial subgroup of 2 since each subgroup on the chart initially contains only 1 bit of data. These steps are highlighted in Figure 9-16 by the circles drawn around each neighboring data point on the top chart and the corresponding arrows that point to the mR value between the coupled data points on the lower chart. Notice that the first three data points on the mR chart (Months 2–4) are relatively close together. This is due to the fact



**FIGURE 9-16** XmR chart for the total amount of dollars saved each month in radiology transcription

that there are small differences between the first four data points that have been coupled together on the Individuals chart. If you look at data points for Months 15 and 16, however, you see a very different picture. The difference (i.e., the range) between Months 15 and 16 is much larger (\$938 to be exact) than any of the ranges found when the first four data points were compared. In short, the individual values when coupled together produce an artificial subgroup of two, which you must have in order to calculate a range and subsequently the moving ranges. One final thing to note about the XmR chart. The mR chart will always have one less data point on it than the Individuals chart. This is due to the fact that the first data point (Month 1)

does not have a neighboring data point to be compared with until the data for Month 2 is posted. Therefore, there is no moving range for Month 1 on the mR chart, which is identified by the triangle surrounding Month 1. This will always be the case when you use an XmR chart.

**EXERCISE 9-2** (You make the call: Is it an X-bar and S chart or XmR chart?) will test your ability to determine whether a particular indicator should be placed on an X-bar and S chart or an XmR chart. Answers to this exercise may be found at the end of this chapter.

When we move to the Attributes side of the Shewhart chart decision tree (Figure 9-12), we need to address two questions:

<b>EXERCISE 9-2 Is it an XmR (I) or X-bar and S? You make the call!</b>		
<b>Indicator</b>	<b>X-Bar and S Chart</b>	<b>XmR (I Chart)</b>
Time to clean an inpatient room (in minutes)		
Patient satisfaction scores for subgroups of 15 patients in the outpatient clinic		
Average turnaround time for all STAT labs done each day		
Cost for each normal delivery		
A diabetic patient's 3x a day blood sugar readings		
Average length of stay for a subgroup of 20 intensive care unit (ICU) patients		
The distance (in feet) that a sample of 10 knee replacement patients can walk in 15 seconds		

- Do we have the occurrence and nonoccurrences of an event? If “yes” then we make a p-chart (i.e., a percentage chart)
- If the answer is “no” meaning that we have only the occurrence of an event (i.e., a defect when we do not know the non-defects) then we need to ask, “Is there an

equal opportunity for a defect to occur?” If we have an equal opportunity for a defect to occur we make a c-chart. If not, then we make a u-chart. The details are explained next.

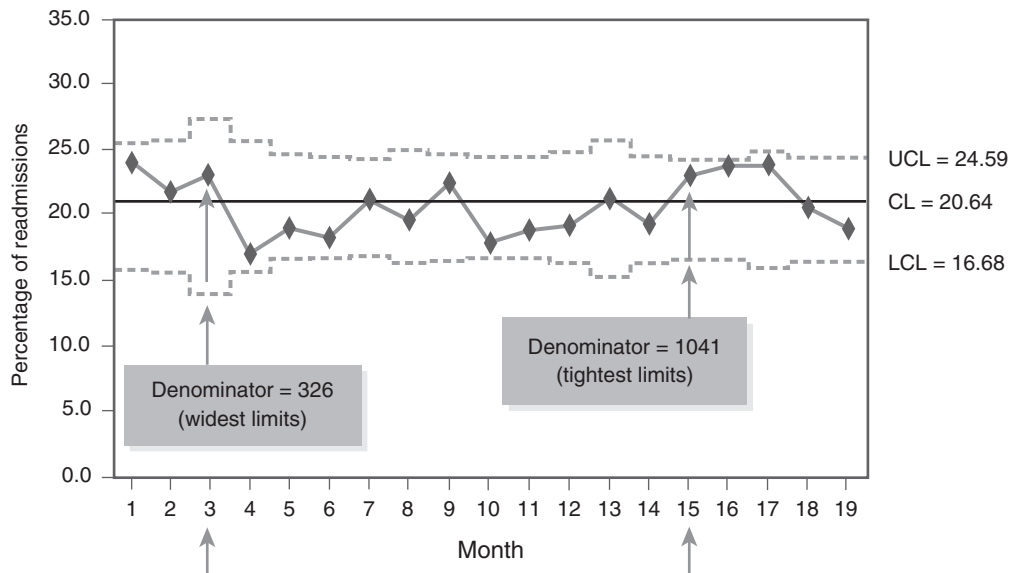
We start with the first question and address the use of the p-chart.

## P-Chart

The p-chart derives its name from the fact that either a percentage or a proportion is what you actually are plotting on the chart. Most of the time the percentage will be the statistic of interest rather than the proportion. When you make a p-chart, or any other attributes chart, you will get only one chart (unlike the variables charts which gives you two charts). The p-chart is used to monitor the proportion or percentage of defectives when you know the occurrence of the defective product, unit, event, or service (the numerator) and the nonoccurrences (the denominator, which is the total being observed). This chart is used frequently in healthcare settings because we track many indicators that look at accuracy, completeness, errors, or the percentage of something done or not done (e.g., cesarean sections, completed history and physical reports, proper hand washing, or compliance with a standard protocol). **FIGURE 9-17** provides an example of a p-chart with stair-step control limits. In this case, the indicator is the percentage of hospital readmissions for home healthcare patients. The

denominators for each dot on the chart are not equal and spread from a minimum of 326 cases (Month 3) to a maximum of 1,041 (Month 15). Notice that the smallest denominator (326 for Month 3) has the widest control limits whereas the largest denominator (1,041 for Month 15) has the tightest set of control limits. The numerators go from a low of 75 readmissions (Month 3) to a high of 249 in Month 16. If the distance or spread between the stair-step limits is relatively small this means that the denominators are relatively close in size.

Finally, it should be noted that if the subgroups (the denominators) were of equal size, the control limits on the p-chart would be straight. But because most healthcare indicators that are defined as percentages differ from one subgroup (i.e., time period) to another (e.g., we do not have the same number of deliveries each month, produce the same number of food trays each day, or have the same number of patients visit at clinic each day) we usually do not have equal subgroups when calculating percentages or proportions. Therefore, most p-charts will generally have stair-step control limits.



**FIGURE 9-17** p-chart on the percent of hospital readmissions for home healthcare patients

## C-chart

The c-chart and the u-chart are the Shewhart charts of preference when you are tracking *defects*. The c-chart is the appropriate chart when you have an *equal area of opportunity* for a defect to occur. As shown in Figure 9-13, an area of opportunity has the following characteristics:

- It applies to all attributes or count charts (c, u, and p)
- It defines the frame or area in which a defective (i.e., a nonconforming unit) or a defect (i.e., a nonconformity) can occur
- It can be of equal or unequal sizes.

A manufacturing example may help clarify this concept. Imagine that you work on a paint line at an automobile manufacturing plant. If you were assigned to paint the hoods of a single model of a car (e.g., a Ford Taurus) there would be an equal area of opportunity for a paint blemish because all Ford Taurus hoods are the same size. In this case, we would make a c-chart and plot the total number of paint blemishes (defects) on each hood you paint. Because each hood has the same number of square inches of surface area there is a constant area of opportunity for a paint blemish.

The challenge now becomes determining when this equal area of opportunity condition exists in healthcare settings. One of the more frequently used examples of how this might occur is with monitoring patient falls. If you conclude that there is basically an equal opportunity for a patient to fall each day of the week at your hospital, rehabilitation facility, or long-term care facility, then you would merely count the number of falls occurring each day, week, or month and plot the number of falls on a c-chart. Other indicators that could be placed on a c-chart if the equal area of opportunity assumption was met include the number of:

- Patient restraints
- Lawsuits
- Patient complaints
- Needle sticks

- Medication errors
- Central line infections

Figure 9-1 was used to show the elements of a Shewhart chart. It also provides an example of a c-chart. In this example, the customer service manager of a large medical group is interested in charting the total number of patient complaints received each week at 17 sites of care. Because each patient could file more than one complaint, complaints are viewed as defects. The alternative, considering the registration of a complaint as a defective, is not selected because this would preclude counting more than one complaint from an individual patient. Remember that defectives are based on the binomial distribution (i.e., the patient complained or did not complain). If you approached this indicator as a defective you would not be concerned with the magnitude of the complaint problem (i.e., the total number of complaints) but rather with the fact that a patient complained or did not complain and you do not care if a patient complained more than once. Measuring complaints as a defective would produce a percentage of patients who complained (a p-chart). As a defect, however, we are concerned with the magnitude of the problem so we count the total number of complaints, including multiple complaints from the same patient. The c-chart is selected because the volume of patients seen at the 17 clinics Monday through Friday remains fairly constant each week and the number of sites included in the study does not change. These two conditions allow the manager to assume an equal area of opportunity for a complaint to occur. She merely counts the total number of complaints received each week and plots this number on a c-chart. The chart produces a CL (average number of complaints) and a UCL and LCL.

A frequent challenge with using the c-chart for healthcare applications is that the condition of equal area of opportunity may not be met. Frequently in healthcare settings there are few indicators that have equal areas of opportunity. The severity of a patient's condition can change

quickly, the census can show fluctuations, the clinic is not a 7 days a week operation, the volume of orders may change rapidly, and the ED may have to go on bypass because there are no more inpatient beds available on a Friday or Saturday night. So if the assumption of an equal area of opportunity is violated, what do we do next? The answer is simple—you make a u-chart.

## U-Chart

This chart is used frequently in health care, especially now that there has been a more concentrated effort to track patient safety indicators. The u-chart, like the c-chart, is used to track defects. The difference is that the u-chart is selected when you conclude that there is *not an equal area of opportunity for the defect to occur*. Let us return to the paint line at the Ford plant for a moment. Although you have in the past painted one model of car at a time, today you have been told that the line will have a mixture of cars and a mixture of hood sizes. So, how do you count the paint blemishes on the hoods of a Ford Escort, a Taurus, a Mustang, and an Expedition? Each hood has a different number of square inches, takes a different volume of paint to cover the surface of the hood, and has a varying probability of experiencing a paint blemish. The u-chart takes care of this problem very quickly by computing a defect *rate*. The number of paint blemishes is used as the numerator and the number of square inches of the hood's surface is used as the denominator. The resultant ratio provides the number of blemishes *per* so many square inches of hood area. The rate essentially normalizes the differences in denominator size (i.e., the area of opportunity for a blemish to occur).

One technical point about rates. Explaining a rate-based statistic can be a little challenging. It is much easier to say, "This past month we had 21 inpatient falls" than to say, "This past month we experienced an inpatient falls rate of 4.4 falls per 1,000 inpatient days." Some in your audience may struggle with what this rate-based statistic

actually means. An example should help clarify how these are opposite sides of the same coin. If we have 21 inpatient falls, this number becomes the numerator of the rate-based statistic. When we place this count of 21 falls over the total number of patient days for the month (e.g., 4,775) we have a ratio of two different numbers that produces a result of 0.00439 (i.e.,  $21/4775 = 0.00439$ ). Because the number of inpatient days is in the thousands we multiple the resultant value of 0.00439 by 1000 to produce the inpatient falls rate of 4.4 falls per 1,000 inpatient days. The number of spots you slip the decimal point on the resultant ratio depends on how large your denominator is. In this case, we had 4,775 inpatient days so we slip the decimal point three places to the right by multiplying the value of 0.00439 by 1000. If you had patient days in the tens of thousands you would slip the decimal point four places to the right and have 43.9 falls per 10,000 inpatient days. Or you could go out to 100k inpatient days and say "I'm sorry but we had 439 inpatient falls per 100,000 inpatient days." Or if you really wanted to depress the senior management team or board you could report 4,397 inpatient falls per 1 million inpatient days. You can adjust the result of the ratio of  $21/4,775$  very easily for any value you place in the denominator position. The general rule for rates, however, is that the denominator you use should be based on the volume you are observing on a regular basis. In the case of inpatient falls, most hospitals are dealing with inpatient days that are in the thousands so this is what should be used to calculate the final rate-based statistic of 4.4 falls per 1,000 inpatient days. If, on the other hand, you were tracking medication errors you would most likely be justified in making the number of errors per 10,000 doses dispensed or scripts written because an average-size hospital will general dispense 10,000 or more doses each month. Finally, if your measure was the neonatal death rate for a state, province, or region then the proper denominator size might be per 100,000 live births.

Because it is an extremely rare to have the same number of medication orders each week,

the same patient census, or the same number of central line days in the ICU, the u-chart is used more often in healthcare settings than the c-chart. Furthermore, because epidemiologists frequently produce rate-based statistics (e.g., the neonatal death rate or the VAP rate) the use of terms associated with the u-chart should sound familiar to many healthcare professionals. Examples of u-chart applications are provided in the case study chapter (Chapter 10).

**TABLE 9-1** provides an overview of the five charts just described and offers examples of indicators that could be placed on each type of chart. Other useful tables that summarize how charts should be set up and their various uses can be found in *Statistical Quality Control Handbook* (Western Electric, 1985) and Benneyan (2001).<sup>10</sup> Readers wishing to gain additional insights about the selection of control charts should consult Wheeler (1995), Montgomery (1991), Pyzdek (1990), Ishikawa (1989), Duncan (1986), Carey and Lloyd (2001), Carey (2003), and Provost and Murray (2011).

## ► You Make the Call

Now that you are familiar with the basic ideas behind the Shewhart charts, the next step is to apply this knowledge to your own indicators. The study questions in **BOX 9-1** will serve as a quick overview of some of the central issues related to Shewhart chart development and as a test of your current knowledge. If you struggle with some of the questions you can review the material presented in this chapter and then explore some of the listed references for additional explanations. Another way to enhance your knowledge base is to attend workshops

on SPC. The ASQ, for example, offers public seminars on SPC. You may want to check with your local ASQ chapter to see when such courses will be offered. The IHI also offers workshops on building effective measurement systems and SPC. The various program offerings that I and my colleagues teach throughout the year can be reviewed on the IHI home page ([www.ihl.org](http://www.ihl.org)). Finally, if you have the opportunity to attend a local or national quality conference (e.g., the IHI National Forum on Quality Improvement in Healthcare or the IHI-BMJ International Forum on Quality and Safety in Healthcare), make sure that you sit in on sessions that are discussing Shewhart charts and SPC. Hearing about control charts from multiple sources will be very beneficial.

You can also test your knowledge of the various charts by completing the *You Make the Call* exercise found in **EXERCISE 9-3**. When I teach my classes on Shewhart chart applications, I give the participants this exercise at the end of the class to provide a final test of their understanding of the selection of appropriate Shewhart charts. It gives them a chance “make the call!” and tests their control chart knowledge. The indicators listed in this exercise are taken from actual teams I have had the opportunity to facilitate or coach. Start the exercise by determining the subgroup. Remember that the subgroup is the label for the horizontal axis and reflects how you have organized your data (e.g., by day or week). Next decide if you have variables or attributes data. Finally, list the chart you think is most appropriate for this situation. You may want to refer to the Shewhart Decision Tree shown in Figure 9-12 to assist you in thinking through the chart options. The answers to the You Make the Call exercise can be found at the end of this chapter.



**TABLE 9-1** Shewhart chart summary

Type of Control Chart	Type of Data and Data Collection Issues	Examples of Indicators Used on This Type of Chart
<p><b>X-bar and S chart</b></p> <p>This is known as the Average (X-bar) and Standard Deviation (S) chart. Most SPC software programs will give you two charts when you select this chart: one for the X-bar portion and one for the S portion. This is considered to be the most statistically powerful of all the charts. The X-bar and S chart can have straight or stair-step control limits.</p>	<p><b>Variables data</b></p> <p>The X-bar and S chart usually involves drawing a sample of observations (e.g., 3–10 per subgroup). Rational subgrouping is frequently used with this chart. The statistical principles behind this chart are based on the assumptions of the normal (Gaussian) bell-shaped distribution.</p>	<ul style="list-style-type: none"> <li>■ Actual turnaround time for five lab tests or three pharmacy orders each day</li> <li>■ Blood pressure readings (e.g., three to five per day)</li> <li>■ Diabetes monitoring (e.g., three fasting blood sugar readings each day)</li> <li>■ Anesthesia time for a sample of cases each day</li> <li>■ Patient satisfaction scores</li> </ul>
<p><b>XmR chart</b></p> <p>This chart is known as the Individual values (X) and moving range (mR) chart. Sometimes it will be referred to as the Individuals or I-chart. It does not have the statistical rigor or power of the X-bar and S chart because each dot on the chart is representing only one observation. This chart is used frequently to answer questions related to volume, for example, “How many surgeries did we do this week?” The XmR chart does not address the question as to whether these surgeries were started on time (this would require a p-chart). Instead, the XmR chart is answering a neutral question, “How many?” or “How much?” The XmR chart will always have straight control limits.</p>	<p><b>Variables data</b></p> <p>The XmR chart is used when you have a single observation for each subgroup (i.e., <math>n = 1</math>). Sampling typically is not done but might be if the process being monitored has an extremely large volume. Because this chart frequently uses aggregates as the plotted number (e.g., days in accounts receivable this month), it is important to make sure that the data are consistently collected from one time period to the next. This chart is used to evaluate questions related to process outcomes (volumes), with no concern as to whether the outcomes of the process are acceptable or not acceptable.</p>	<ul style="list-style-type: none"> <li>■ Patient wait time to see the physician or to be seen in the ED</li> <li>■ The number of days to mail a patient bill after discharge</li> <li>■ The number of calls coming into a clinic each day</li> <li>■ Average length of stay by week for a particular diagnosis-related group (DRG)</li> <li>■ The number of surgeries done each week</li> <li>■ Operating margin by month</li> <li>■ Pounds of laundry each day</li> <li>■ Average turnaround time by day</li> <li>■ The number of food trays produced</li> <li>■ Patient satisfaction score</li> </ul>

*(continues)*

**TABLE 9-1** Shewhart chart summary*(continued)*

Type of Control Chart	Type of Data and Data Collection Issues	Examples of Indicators Used on This Type of Chart
<p><b>p-chart</b></p> <p>The p-chart is used frequently in health care to compute the percentage (or proportion) of defective products or services. The p-chart requires being able to count both the numerator and the denominator. The p-chart is the weakest of the attributes charts because it is based on the binomial distribution (i.e., there are only two outcomes such as yes/no, acceptable/not acceptable, or complete/not complete. The p-chart can have straight or stair-step control limits.</p>	<p><b>Attributes data</b></p> <p>These data are classified as defectives or nonconforming units because they reflect the percentage (or proportion) of things or events that do not meet specifications or criteria (the numerators). The denominators usually (but not always) are of varying sizes, which produce stair-step control limits. Data of this type reflect the binomial distribution. The denominators need to be sufficiently large (e.g., usually greater than 12) to enable a reasonable percentage to be calculated yet not too large (e.g., over 5,000).</p>	<ul style="list-style-type: none"> <li>■ Percentage of cesarean sections</li> <li>■ Percentage of late food trays</li> <li>■ Percentage of incomplete charts</li> <li>■ Percentage of late surgery starts</li> <li>■ Percentage of bills that are inaccurate</li> <li>■ Percentage of mortality</li> <li>■ Percentage of staff turnover</li> <li>■ Percentage of patients responding “Very Good” to a survey question</li> <li>■ Percentage of x-rays that had to be redone</li> <li>■ Percentage of did not attends (DNAs) at an outpatient clinic</li> </ul>
<p><b>c-chart</b></p> <p>The c-chart is used to count the number of defects that occur within an equal area of opportunity when the nondefects are unknown. In this case, each observed unit (e.g., a patient) can have multiple defects (e.g., falls). Generally speaking, defects are the specific reasons why a product or service is classified as defective (i.e., a defective product or service will suffer from one or more defects). Generally speaking, indicators appropriate for a c-chart should be considered “rare events.” The c-chart will always have straight control limits.</p>	<p><b>Attributes data</b></p> <p>The key to using a c-chart is that there should be an equal area of opportunity for a defect to occur. This condition frequently makes it difficult to use this chart in health care because the conditions under which we provide care do not always remain constant. One way to address this inequality in the area of opportunity is to apply stratification. For example, if the conclusion is that there is not an equal area of opportunity for an inpatient fall because the hospital functions differently on weekends than weekdays then separating the data by weekdays versus weekends may be sufficient to conclude that there is a relatively equal area of opportunity for a fall during each of these periods. The c-chart is based on the Poisson distribution.</p>	<ul style="list-style-type: none"> <li>■ The number of falls</li> <li>■ The number of restraints</li> <li>■ The number of needle sticks</li> <li>■ The number of lawsuits filed</li> <li>■ The number of ventilator-associated pneumonias</li> <li>■ The number of nosocomial infections</li> <li>■ The number of medication errors</li> <li>■ The number of returns to surgery</li> <li>■ The number of surgical site infections</li> <li>■ The number of violent events in a mental health ward</li> <li>■ The number of central line infections</li> </ul>

**TABLE 9-1** Shewhart chart summary*(continued)*

Type of Control Chart	Type of Data and Data Collection Issues	Examples of Indicators Used on This Type of Chart
<p><b>u-chart</b></p> <p>The u-chart is used to track defects when the area of opportunity is not equal. For this reason, the u-chart is typically used more often in health care than the c-chart. This chart is based on rates rather than simple counts. The u-chart can have straight or stair-step control limits.</p>	<p><b>Attributes data</b></p> <p>The Poisson distribution is also used as the frame of reference for this chart. The u-chart presents rates (e.g., so many falls per 1,000 patient days). Knowledge of how to collect data to form rates is essential.</p>	<ul style="list-style-type: none"> <li>■ Medication errors per 10,000 doses dispensed</li> <li>■ VAP per 1,000 vent days</li> <li>■ Total falls per 1,000 patient days</li> <li>■ Total readmissions per 1,000 discharges</li> <li>■ Bloodstream infections per 1,000 line days</li> </ul>

**BOX 9-1** Shewhart charts study questions

- When is it appropriate to use Shewhart charts? Should I use them in place of descriptive statistics?
- What is the relationship between Shewhart charts and tests of significance?
- How many data points do I need to make a Shewhart chart? What do I do if I do not have enough data?
- Which is better, attributes or variables data?
- What is a subgroup? Do I have to have one to make a Shewhart chart?
- Can I make a Shewhart chart with only single data points?
- Do my subgroups have to be of equal size when I make Shewhart charts?
- Much of the data I get does not have the date on it. So, does it really matter if the data points are not in chronological order?
- I still don't get this distinction between a SD and a sigma limit. Why aren't they the same? Does it really matter? My spreadsheet software will give me a SD. Why can't I just multiple this number by 3 and then add and subtract this product from the mean to get the control limits?
- Why do I have to use 3 sigma control limits? Why can't I use two or maybe 1.5 sigma limits?
- Do defects add up to make defectives or is the other way around?
- When I make a p-chart, does the size of the denominator make a difference? Can I have, for example, 4 or 5 in my denominator?
- What is the difference between a proportion, a percentage, and a rate?
- Should I view common cause variation as "good" variation and special cause variation as "bad" variation?
- Do I really have to investigate a special cause? Can't I just remove the data point from the chart and get on with making changes?

**EXERCISE 9-3** You make the call! Selecting the right chart

Situation	Subgroup?	Type of Data?	Type of Chart?
1. Each day you record the number of films processed in the radiology department.		<b>V or A</b>	
2. Each day you record the number of films requested and the number that cannot be found in the radiology library.		<b>V or A</b>	
3. The number of inpatient restraints each month is placed over the total inpatient days each month.		<b>V or A</b>	
4. Each day you pull a stratified random sample of 15 complete blood counts (CBCs) and record the turnaround time (in minutes) for each CBC.		<b>V or A</b>	
5. The number of minutes it takes to get a stat med order administered to the patient (order time to administration time).		<b>V or A</b>	
6. Every 2 weeks you pull a sample of 30 medication orders and count the total number of orders that have one or more errors.		<b>V or A</b>	
7. The wait time in the ED (door to discharge) is tracked for each patient.		<b>V or A</b>	
8. The clinic receptionist notes the time of check-in for each patient. The physician notes the time when he/she first sees the patient in the exam room. An analyst compiles the data daily and reports the percentage of patients who had to wait more than 30 minutes.		<b>V or A</b>	
9. The director of surgery keeps track of the total number of surgical procedures performed each week.		<b>V or A</b>	
10. The dietary department records the number of food trays that come back uneaten each day and the total number of trays they produced for that day.		<b>V or A</b>	
11. You are interested in the average time patients spend in your waiting area, so every day a student randomly picks eight patients and measures their actual waiting time in whole minutes.		<b>V or A</b>	

*(continues)*

**EXERCISE 9-3** You make the call! selecting the right chart*(continued)*

12.	The ICU nurses want to evaluate the ventilator-associated pneumonia (VAP) rate. So every 2 weeks they record the total number of pneumonia episodes and the total number of vent days.	V or A	
13.	Each week patient satisfaction scores for three units are compiled and an average is calculated for the three units.	V or A	
14.	The finance department tracks the total number of business days it takes to process a vendor's request for payment. Process time starts when the request for payment is received in the finance department and ends when the payment is sent (electronically or posted in the mail) to the vendor.	V or A	
15.	Every week each medication order is checked against five potential types of errors. The total number of errors for the week is divided by the total number of orders submitted that week.	V or A	
16.	You know the number of people who come to the ED complaining of chest pain and the number who are actually diagnosed with an AMI or unstable angina.	V or A	

## ► Additional Shewhart Charts

In addition to the five basic Shewhart charts described previously, there are many other charts that have their roots in manufacturing but have proven to be very useful in certain healthcare situations. Some of these alternative Shewhart charts include:

- Median chart
- t-chart
- g-chart
- Moving average chart
- Cumulative Sum chart (CUSUM)
- Exponentially weighted moving average chart (EWMA)
- Standardized Shewhart Chart

- Multivariate Shewhart-type Charts
- P primed chart ( $p'$ -chart)
- U primed chart ( $u'$ -chart)

Provost and Murray (2011) do a very good job of not only describing these alternative charts and provide examples of their use. I do not intend to go into depth about these various charts but I do want to make a few comments about the t- and g-charts that are being used more and more in healthcare improvement work.

The t- and g-charts are designed to address the occurrence of rare events. I know, you are wondering, "What is the operational definition of a rare event?" When I was first learning about these charts the instructor used a simple example. He would ask, "What is the probability of looking out the window and seeing a car go by?" Everyone would respond, "High." Then he

would ask, “What is the probability of looking out the window and seeing an accident occur?” Everyone would respond, “Low.” He then would proudly announce, “You now understand a rare event.” Now this is a pretty casual explanation of a rare event but I think it helps to set the context for thinking about rare events. If you wish to get very statistical about rare events you can study what is called the “rare event rule for inferential statistics.” Within this body of statistical theory you will be reacquainted with probability theory that you were exposed to relatively early in your statistical training. Most of the time this is explained by using the probabilities associated with rolling various combinations on dice or getting a particular combinations of cards while playing blackjack or poker or betting on a roulette wheel.

At the IHI we use a practical approach to defining rare events that is grounded in statistical theory but does not require detailed computations. Simply stated, if you have more than 25% of the data on a p-, c-, or u-chart at zero (or conversely at 100%) you need to consider moving to a t- or g-chart. With 25% or more of the data points at zero the use of the traditional rules for detecting special causes on a Shewhart chart become questionable (Provost & Murray, 2011). It also is a practical issue. If you do not have sufficient nonzero data for an attribute chart the LCL may not exist, which makes the interpretation of the chart difficult. In these situations, you should consider moving to the time between chart (t-chart) or the cases between chart (g-chart). The t-chart (the t part of the name refers to “time”) or time between chart shows you how much time has gone by since the last adverse event. Nelson (1994) provided the details on how this chart is constructed. When you use this chart you have to reorient the way in which you explain the chart. For example, the horizontal axis on the t-chart is a discontinuous time sequence. If you start next Monday to begin tracking patient falls but a fall does not occur until Wednesday then you would place a dot on the chart’s vertical axis at 2 (i.e., 2 days have gone by before a fall occurred). When an event

happens (e.g., a fall, a pressure ulcer, a surgical site infection) you basically reset the counter and begin counting the number of days again until the next fall occurred. This is the same approach that factories use to track the number of days that have gone by without an accident in the factory. If you never had an event you would never have a dot on the chart. Because you are counting the number of days that have gone by since the last event (i.e., a defective or a defect) the horizontal axis will not have Monday, Tuesday, Wednesday or January, February, March, etc. marked. When an event occurs this is when you place the date of the event on the horizontal axis, which will not be occurring in equal periods of time. The indication of improvement on a t-chart is when you observe an ever-increasing run of days without the occurrence of an event.

As healthcare providers have become more focused on safety indicators and reducing harm the t-chart has become increasingly popular. But a word of caution is in order. The statistical basis for properly calculating the limits on a t-chart are a little involved. First, you need to realize that a distribution of rare events does not follow a normal Gaussian bell curve. A Poisson distribution is a better referent for rare events. The Poisson distribution is appropriate as a referent for the c- and u-charts as well as the time between chart. In the case of the t-chart, however, the form of the Poisson distribution is actually an exponential distribution, which is in turn highly skewed. Second, the skewness of the exponential distribution is not a major problem and is addressed by transforming the time between events (i.e., days gone by) into a quasi-normal or symmetric distribution by performing what is called a Weibull transformation. Third, once the data are more or less approximating a normal distribution the UCL and LCL can be calculated by using the formulae for the XmR chart. Finally, after the limits and CL are calculated they are transformed back to their original state for plotting on the chart. I know, this all sounds rather complicated. The detailed steps for constructing the t-chart

are clearly discussed by Provost and Murray (2011) and Nelson (1994). Also take heart in knowing that any reasonably good SPC software package will do all the calculations for a t-chart quickly and easily once you create the time between data file. Your biggest challenge will be to explain how to interpret the t-chart. But because its use is growing in popularity in healthcare setting, it is well worth your time to gain more knowledge of the time between chart.

The g-chart (or geometric chart) is similar in principle to the t-chart. It too is a chart for rare tracking events except that instead of plotting the amount of time (e.g., days) between a rare event, the g-chart plots the number of cases that are regarded as being successful against cases considered to be failures. A failure in this situation might be a surgical site infection, patients experiencing a medication error, or a return to surgery within 24 hours. Like the t-chart success is determined by having a long run of successful cases with no failures or adverse events. Although the t-chart is modeled after an exponential distribution the g-chart referent is a geometric distribution. Again, the steps for computing the limits and the CL on a g-chart are nicely laid out in Provost and Murray (2011). There is also a considerable body of literature on both the exponential and geometric distributions that can be found in the ASQ's *Journal of Quality Technology*. The g-chart is also becoming a standard offering in most SPC software packages. Once again, however, the challenge is making sure you have at least a moderate foundation in being able to explain the chart and how to interpret it.

## ► Using Shewhart Charts Effectively

At some point after reading various books on run and Shewhart charts and listening to others explain control chart theory, you will all of a

sudden proclaim, “I think I get it!” Once you reach this point it is now time to start applying this knowledge to actual improvement opportunities. But be careful. I have seen some people become so enthusiastic about the various charts that they start making graphs on any process that produces data. It is at this stage that I remember an old adage—if you give a child a hammer, the whole world looks like a nail! The charts play a valuable and central role in all QI efforts. It is important to realize, however, what they can do and what they cannot do.

First, appreciate the fact that the charts do not answer the following questions:

- What is the reason for a special cause?
- Should a common cause process be improved?
- What should I do to improve the process?

The answers to these questions do not come from the charts or statistics. They come from the will, ideas, and ability of the team to execute tests of change. I have seen too many teams feel that once they have created a chart their work is finished. I think that this occurs because the chart is a tangible thing that can be pointed to and shown to others. Improvement strategies, on the other hand, are not as finite or discrete. Developing improvement strategies is actually much more difficult than mastering control chart theory and construction because you are dealing with people, behaviors, and culture not numbers.

Second, after you make a chart and decide whether the process exhibits common or special cause variation, you then need to decide how you are going to approach the variation you have identified. Do you need to merely reduce variation in the process or fundamentally redesign the process and change the way in which work is envisioned and delivered? All improvement strategies emanate from an understanding of variation. If the process exhibits special cause variation the appropriate decision is to investigate the special cause(s) and determine why they have made the process unstable and unpredictable. Just as we would investigate a patient safety event

(i.e., a sentinel event) by conducting a root cause analysis, we also need to do the same thing when a special cause is detected on a run or Shewhart chart. Ignoring a special cause will guarantee that it will rear its ugly head at some point in the future. We cannot predict exactly when a special cause will occur but you can be sure that it will pop up again if you choose to ignore it.<sup>11</sup>

The other aspect of a special cause is that not every special cause is negative and undesirable. Remember that special causes are not bad and common causes are not good. The key point is the special causes make a process *unstable* and *unpredictable*. It is very likely that you will observe a special cause that you want to emulate (e.g., when lab turnaround time is much faster than it has been or the past 15 days). In this case, you want to investigate why the process worked so well on those days and see whether these conditions can be replicated. Common causes on the other hand are not inherently good. Common cause variation merely means that the process is *stable* and *predictable* (i.e., predictable within the boundaries of the UCL and LCL). Just as you can have a special cause that you might want to emulate, you can also have common cause variation that is unacceptable (e.g., when a patient's blood pressure is running at a very high level and staying there or when the wait time to see your family physician is consistent and predictable but it is at such a high level that it is predictably bad and unacceptable).

QI starts with making the correct decision about the variation that lives in your data. Walter Shewhart introduced the control chart and the notions of common and special causes of variation in 1924 (Hare, 2003). Since then SPC has become the foundation for all successful QI initiatives. It is a key component of the Baldrige criteria, Six Sigma, Lean, and International Organization for Standardization (ISO). Without a clear understanding of variation and its causes, however, individuals and organizations will continue to suffer from numerical illiteracy.

## Notes

1. Historically these charts have been known as control charts. Shewhart himself even referred to them as control charts as have many writers since Shewhart's time. But as Blank (1998, p. 1) points out, "It is important to understand that SPC does not control processes. People control processes. SPC is merely a tool that provides you with information you need to reduce variation and tell you whether or not your processes can meet the customer's expectations." In more recent times, the charts have been referred to more and more as Shewhart charts (Provost & Murray, 2011, p. 113) to emphasize their use primarily in understanding variation and to facilitate learning about process capability rather than conveying images of "control." The term Shewhart chart is also used to recognize the significant contributions of Dr. Shewhart to the field of SPC. A final note on the use of the word "control." The ASQ was originally called the American Society for Quality Control (ASQC). In 1997, the membership voted to drop the word "control" from the organization's name. This was to recognize that quality was becoming a broader concept and used in many other fields besides manufacturing where initially in the early 1900s control was used as a key operative word. Shewhart's book, *Economic Control of Quality of Manufactured Product* (1931) provides a classic reference to the initial use of the term "control." So for a variety of reasons I use the term Shewhart chart (s) in this text rather than control charts.
2. The USL and LSL are frequently referred to in manufacturing as "tolerance limits" and are also frequently referred to as the voice of the customer (VOC, i.e., what the customer wants, needs, or expects from the product or service). There are many different types of indices that have been developed to capture statistically process



- capability. The three basic process capability indices are the process capability index ( $C_p$ ), the minimum process capability index ( $C_{pk}$ ), and the process capability index to the mean ( $C_{pm}$ ). The traditional statistical use of process capability ( $C_p$ ) is to indicate whether or not the process can meet the predetermined specifications (Blank, 1998). There are numerous variations on the  $C_p$  statistic, all of which are designed to help the quality control (QC) researcher investigate special causes and get the process to perform as closely as possible to the expectations of the customer (i.e., the specifications).
3. I have calculated a  $C_p$  and  $C_{pk}$  statistics only once for a healthcare indicator. It was when I was helping to set up an outpatient clinic designed to manage patients on anticoagulation medication (i.e., warfarin sodium). Several key indicators are used in assessing clotting issues. The PT, along with its derived measures of prothrombin ratio (PR) and INR, are assays evaluating the extrinsic pathway of coagulation. This test is also called “ProTime INR” and “PT/INR” (MedlinePlus medical encyclopedia, <https://medlineplus.gov/ency/article/003652.htm>). Because there are defined therapeutic limits associated with these measures they can be regarded as USL and LSL. These values would be set on the Shewhart charts as reference lines. Then the patient’s actual results on the PT and INR would be plotted on the chart, and the UCL and LCL of the patient would then be compared to the USL and LSL. Because we had both an USL and a LSL and control limits the capability statistics could be calculated to determine how well the patient was conforming to the therapeutic limits (USL and LSL) of the drug. But in most instances in healthcare settings, there is only a single target or goal rather than the USL and LSL. Healthcare simply does not currently function at the same level of precision for many of its indicators as the manufacturing industry. So, when I use the term process capability I am using it in a general sense to describe the variation in the process as defined by the mean (CL) and the UCL and LCL. These numbers define how well the process is performing relative to the target or goal.
  4. Some of my colleagues may disagree with these guidelines. I have found over the years that there are two general issues that need to be balanced against each other: statistical purity and practicality. The science of improvement (SOI) is as Shewhart referred to it an “applied science.” Therefore, in my work I have always tried to balance the precision of statistical requirements with a heavy dose of practicality. For example, I have worked with wonderful people in the National Health Service (NHS) of Scotland for over 12 years. During this time we have developed a variety of health and social service measurement systems. Most of the data are collected monthly and many of the indicators were not collected historically. So, we were starting out with no data on selected indicators and had to build charts as we went along. In this case, trail control limits were essential. We also made a very practical decision to use the first 6 months of data as baseline for indicators that had no history. Again some would argue that this is not enough data to establish a baseline but it was sufficient to get us started on the road to improvement.
  5. In one of my measurement workshops, a few years ago this confusion was highlighted very clearly. A young woman near the front of the room raised her hand after I was done explaining that a sigma was not equivalent to a SD. She had a bit of a wrinkled brow and looked concerned. She said, “I was told that the UCL and LCL were calculated as SD. Is this not correct?” I drew the formula for

the SD on a flipchart and asked her if this is what she used to calculate the three sigma control limits. She said “Yes, that is what I was told to use.” I then proceed to politely tell her that the control limits on her charts were wrong. The UCL and LCL would either be too wide or too narrow if she used the SD of the data. She got a very strange look on her face, was quiet for a moment, then burst out, “But this means I have been giving the charts to the senior management team and to the board! What am I going to do?” I asked her if anyone had ever noticed or commented on the fact that the limits were not properly calculated. She responded, “No.” I suggested that she learn how to make the charts correctly with SPC software (she was merely using Excel with no SPC add-on software that properly computes sigma limits) and then submit the correct charts to the senior management team and board the next time around. She still looked a little perplexed, however. She was concerned that she would lose credibility with the management team when they found out the charts were wrong. I told her that unless she tells them that her original charts had the wrong limits it does not sound like anyone on the board or the senior management team had sufficient grounding in Shewhart charts to actually discern that the charts were different. I told her to let me know how it went when she showed them the correct charts. She wrote back and said that no one asked any questions.

6. API develops methods, works with leaders and teams, and provides education and training to help organizations improve their products and services and to build their capability for ongoing improvement. The principals of API have worked in industrial, educational, health, and social service settings. They have worked extensively with Dr. Deming and provided support

for his 4-day seminars. They have written a number of key books on QI including *The Improvement Guide* (Langley, Moen, Nolan, Nolan, Norman, & Provost, 2009), *Quality Improvement Through Planned Experimentation* (Moen, Nolan, & Provost, 2012), and *The Health Care Data Guide* (Provost & Murray, 2011).

7. When Dr. Ray Carey and I first started teaching control chart applications to healthcare professionals in 1992, we taught the traditional list of seven control charts. In 1995, we wrote a book that described these seven charts and their use (Carey & Lloyd, 2001). In December, 2002, Dr. Carey and I taught a minicourse and two workshops on control charts at the 14th National Forum on Quality Improvement in Health Care sponsored by the IHI. This was the first time in the 12 years that we had been teaching for the IHI that we reduced the number of charts we taught from seven to five. The sessions were well received, and the participants found the more simplified approach to be appealing. The two charts we dropped were the X-bar and R chart and the np-chart. Our reasoning for doing this was that the X-bar and S chart can be used in any situation that calls for the X-bar and R chart (when the subgroup is greater than 2). The np-chart, which is a count of the number of defectives, requires equal subgroup sizes (i.e., the denominators), which do not happen very often in healthcare settings. The p-chart can be used effectively, however, in any situation where an np-chart could be used. If there are equal subgroup sizes then the p-chart will have straight control limits. If, on the other hand, the measure has unequal subgroups then the p-chart will have what is known as “stair-step” control limits. In this case, the control limits are different for each data point. The closer in size the denominators the smaller the

- “steps” between each of the control limits. If there are large differences between the denominators the “steps” will be greater between the individual data points.
8. I wrote a commentary in JAMA a few years ago titled “A Matter of Time” (Lloyd & Goldmann, 2009) highlighting how clinicians, researchers, patients, and improvement specialists all have very different views of time. To these four categories I could add management time, which focuses on monthly aggregates of data.
  9. The control chart examples presented in this chapter have been developed to demonstrate the five different charts. The substantive importance of the various charts is not the focus of this chapter. The charts have been developed for heuristic purposes, and the clinical or operational impacts of the indicators presented on the charts are not the primary objective in this chapter. Analysis and interpretation of control charts are addressed in Chapter 10.
  10. The idea for creating this table came from Dr. James Benneyan of Northeastern University in Boston. In a paper titled “Design, Use, and Preferences of Statistical Control Charts for Clinical Process Improvement” (September 16, 2001), he used a table to summarize the various charts. After reading this paper, I realized that the table was something I had not used to summarize the control charts. I believe a table format works nicely to augment the utility of the decision tree shown in Figure 9-12 and the textual details. Dr. Benneyan has written extensively on the topic of control charts in health care and I would encourage readers to review his work. He can be reached at the following address: MIME Department, 334 Snell Engineering Center, Northeastern University, Boston, MA 02115; phone 617-373-2975; email benneyan@coe.neu.edu.
  11. There are many good examples of how people have ignored special causes when they first occurred and then decided to deal with them when they popped up again. The terrorist attacks on our nation on September 11, 2001 provide a classic example. Several years prior to 9/11 the World Trade Center was bombed by terrorists. Although this seemed to draw the nation’s attention for a while, interest in this special cause soon faded into the “old news” category and steps were not taken to extricate the factors that led to the special cause. The conditions for 9/11 were still existing within our system. The September 11 special cause, however, generated a completely different reaction. Our nation mobilized not only to investigate the special cause but also take steps to literally try to eliminate the origin of the special cause. Every day there are stories in the news that should prompt a discussion as to whether the event is a special cause or part of a common cause system. All too often, however, we overreact to a special cause and want to change the system without fully investigating the reasons why it occurred. Other times, however, we ignore a special cause and “hope” that it will not happen again. Hope is not a plan. Knowing how to appropriately react to common and special causes is a much better approach than hoping a special cause will not pop up again.

## Answers to the Chapter 9 Exercises

This section provides the answers to the exercises presented earlier in this chapter. The first **(EXERCISE 9-1)** deals with differentiating defectives from defects. **EXERCISE 9-2** provides indicators that could be placed on either an X-bar and S chart or an XmR chart. The answers to these two exercises are shown here.

**EXERCISE 9-1** Defective or defect? You make the call! (Answers)

Indicator	Defective (Classification)	Defect (Count)
1. Number of accidents per 1,000 employee days		*
2. Number of errors per 25 food trays		*
3. Percentage of AMI patients receiving aspirin within 24 hours of arrival in the ED	*	
4. Percentage of inpatient deaths each month	*	
5. Number of surgical complications per 1,000 surgeries performed		*
6. Proportion of hand hygiene observations done incorrectly	*	
7. Number of falls per 1,000 patient days		*
8. Number of medication errors per 10,000 doses dispensed		*

**EXERCISE 9-2** You make the call: Is it an X-bar and S chart or XmR chart? (Answers)

Indicator	X Bar and S Chart	XmR (I Chart)
Time to clean an inpatient room (in minutes)		*
Patient satisfaction scores for subgroups of 15 patients in the outpatient clinic	*	
Average turnaround time for all STAT labs done each day		*
Cost for each normal delivery		*
A diabetic patient's 3x a day blood sugar readings	*	
Average length of stay for a subgroup of 20 ICU patients		*
The distance (in feet) that a sample of 10 knee replacement patients can walk in 15 seconds	*	

The final (**EXERCISE 9-3**) brings together the key issues related to selecting the most appropriate Shewhart chart for different measurement situations. In this exercise, the subgroup, type of data, and type of chart all need to be specified. Depending on how you interpret the wording describing the situations in Exercise 9-3, you might think that a type of chart other than that I have listed could be selected. A key learning point for this exercise is that slight changes to the wording of the situation could lead you to selecting a different chart. For example, take a close look at situations 8 and 11 in Exercise 9-3. The wording for situation 8 points you to select a p-chart because they decided to focus on patients who had to wait more than 30 minutes. Even though they had variables data (i.e., time) they basically turned it into attributes data because of the 30-minute target. They have taken the more powerful form of data (variables data) and relegated it to a binomial condition, over 30 minutes and under 30 minutes. They will never understand the true variation in wait time. What is the longest wait? You have no idea. All we know is that a certain percentage of patients had to wait more than 30 minutes. The longest wait could be 31 minutes or 13,184 minutes. The more appropriate approach is found in situation 11. Here they are taking a sample of eight patients each day and recording their actual wait times. The chart of preference in this situation is the X-bar and S chart. We will now have the average wait time for a given day and the SD from this average. We can lay a separate line of the chart showing the target of 30 minutes. This gives us much more information about the process variation and how capable it is of achieving the target, which cannot be determined by using the p-chart. The final aspect

of this situation is that if the target is to have all patients be seen in 30 minutes or less, the 30-minute target actually needs to be the UCL of the X-bar and S chart not the average. If 30 minutes is the average on the chart you will naturally have some patients waiting more than 30 minutes and some waiting less. A target is useful on a chart but it needs to be understood in light of the actual variation in the process and the capability of the current process to achieve the target. The Shewhart chart can help you determine the magnitude of improvement needed to achieve the target and but in the case of improving wait time, this is best accomplished by not turning variables data into attributes.

The most appropriate chart for each situation described in Exercise 9-3 is shown here. Note that situation 16 is a trick question. Did you determine that a chart cannot be identified? Why? Because there is no subgroup identified in the situation description. Remember, a Shewhart chart must have a subgroup and an observation as minimum requirements. In this situation, there is no subgroup. But if the situation had been worded as follows then we would have a subgroup: "You know the number of people who come to the ED complaining of chest pain EACH MONTH and the number who are actually diagnosed with an AMI or unstable angina." Now you would be able to determine which chart is most appropriate. In this situation, the Shewhart chart of choice would be the p-chart because we know the denominator (i.e., the number of people coming to the ED complaining of chest pain) and the numerator (i.e., the number who were actually diagnosed with an AMI or unstable angina). Without a subgroup, however, we cannot make a decision about which chart is most appropriate.

**EXERCISE 9-3** You make the call!: Selecting the right chart (Answers)

Situation	Subgroup?	Type of Data?	Type of Chart?
1. Each day you record the number of films processed in the radiology department.	Day	V	XmR
2. Each day you record the number of films requested and the number that cannot be found in the radiology library.	Day	A	p-chart
3. The number of inpatient restraints each month is placed over the total inpatient days each month.	Month	A	u-chart
4. Each day you pull a stratified random sample of 15 CBCs and record the turnaround time (in minutes) for each CBC.	Day	V	X-bar & S
5. The number of minutes it takes to get a stat med order administered to the patient (order time to administration time).	Stat med order	V	XmR
6. Every 2 weeks you pull a stratified sample of 30 medication orders and count the total number of orders that have one or more errors.	Two weeks	A	p-chart
7. The wait time in the ED (door to discharge) is tracked for each patient.	Patient	V	XmR
8. The clinic receptionist notes the time of check-in for each patient. The physician notes the time when he/she first sees the patient in the exam room. An analyst compiles the data daily and reports the percentage of patients who had to wait more than 30 minutes.	Day	A	p-chart
9. The director of surgery keeps track of the total number of surgical procedures performed each week.	Week	V	XmR
10. The dietary department records the number of food trays that come back uneaten each day and the total number of trays they produced for that day.	Day	A	p-chart
11. You are interested in the average time patients spend in your waiting area, so every day a student randomly picks eight patients and measures their actual waiting time in whole minutes.	Day	V	X-bar & S

**EXERCISE 9-3** You make the call!: Selecting the right chart (Answers)

(continued)

Situation	Subgroup?	Type of Data?	Type of Chart?
12. The ICU nurses want to evaluate the ventilator-associated pneumonia (VAP) rate. So every 2 weeks they record the total number of pneumonia episodes and the total number of vent days.	Two weeks	A	u-chart
13. Each week patient satisfaction scores for three units are compiled and an average is calculated for the three units.	Week	V	XmR
14. The finance department tracks the total number of business days it takes to process a vendor's request for payment. Process time starts when the request for payment is received in the finance department and ends when the payment is sent (electronically or posted in the mail) to the vendor.	A request for payment	V	XmR
15. Every week each medication order is checked against five potential types of errors. The total number of errors for the week is divided by the total number of orders submitted that week.	Week	A	u-chart
15. You know the number of people who come to the ED complaining of chest pain and the number who are actually diagnosed with an AMI or unstable angina.	Unknown*	A	Unknown*

**\*NOTE:** Item 16 is a trick question. A subgroup is not specified. Without a subgroup you cannot make a decision about the most appropriate chart. If this description indicated that "You know the number of people who come to the emergency department EACH MONTH . . ." you would have a subgroup. The chart of choice would then be a p-chart.

## References

ASQ Statistics Division. *Glossary and Tables for Statistical Quality Control*, 4th ed. Milwaukee: ASQ Press, 2005

Benneyan, J. "Design, Use and Performance of Statistical Control Charts for Clinical Process Improvement." Unpublished paper, Northeastern University, September 2001. For access to this paper contact Professor James C. Benneyan, Ph.D., MIME Department, 334 Snell Engineering Center, Northeastern University, Boston MA 02115; tel: 617-373-2975; fax: 617-373-2921; e-mail: [benneyan@coe.neu.edu](mailto:benneyan@coe.neu.edu).

Benneyan, J., R. Lloyd, and P. Plsek. "Statistical Process Control as a Tool for Research and Health Care Improvement." *Journal of Quality and Safety in Healthcare* 12, no. 6 (December, 2003): 458–464.

Blalock, H. *Social Statistics*. New York: McGraw-Hill, 1960.

Blank, R. *The SPC Troubleshooting Guide*. New York: Quality Resources, 1998.

Blumenthal, D. "Total Quality Management and Physicians' Clinical Decisions." *Journal of the American Medical Association* 269 (1993): 2775–2778.

Carey, R. *Improving Healthcare with Control Charts*. Milwaukee: Quality Press, 2003.

- Carey, R., and R. Lloyd. *Measuring Quality Improvement in Healthcare: A Guide to Statistical Process Control Applications*. Milwaukee: Quality Press, 2001.
- Deming, W. E. *The New Economics*, 2nd ed. Cambridge, MA: Massachusetts Institute of Technology, Center for Advanced Studies, 1994.
- Deming. "Quality, Productivity and Competitive Position." Indianapolis, Indiana, August 11–14, 1992.
- Duncan, A. J. *Quality Control and Industrial Statistics*, 5th ed. Homewood, IL: Irwin Press, 1986.
- Gonick, L., and W. Smith. *The Cartoon Guide to Statistics*. New York: Harper Perennial, 1993.
- Grant, E., and R. Leavenworth. *Statistical Quality Control*. New York: McGraw-Hill, 1988.
- Hare, L. "SPC: From Chaos to Wiping the Floor." *Quality Progress* (July 2003): 58–63.
- Ishikawa, K. *Guide to Quality Control*. White Plains, NY: Quality Resources, 1989.
- Kume, H. *Statistical Methods for Quality Improvement*. Tokyo: Association for Overseas Technical Scholarship, 1996.
- Langley, G. L., R. Moen, K. M. Nolan, T. W. Nolan, C. L. Norman, and L. P. Provost. *The Improvement Guide: A Practical Approach to Enhancing Organizational Performance*, 2nd ed. San Francisco, CA: Jossey-Bass, 2009.
- Levine, D., and D. Stephan. *Even You Can Learn Statistics: A Guide for Everyone Who Has Been Afraid of Statistics*. Upper Saddle River, NJ: Pearson-Prentice Hall, 2005.
- Lloyd, R., and D. Goldmann. "A Matter of Time." *Journal of the American Medical Association* 302, no. 8 (2009): 894–895.
- Moen, R., T. Nolan, and L. Provost. *Quality Improvement Through Planned Experimentation*, 3rd ed. New York: McGraw-Hill, 2012.
- Mohammed, M., P. Worthington, and W. Woodall. "Plotting Basic Control Charts: Tutorial Notes for Healthcare Practitioners." *Quality and Safety in Health Care* 17 (2008): 137–145.
- Montgomery, D. C. *Introduction to Statistical Quality Control*, 2nd ed. New York: John Wiley & Sons, 1991.
- Nelson, L. "Interpreting Shewhart Average Control Charts." *Journal of Quality Technology* 17 (1985): 114–116.
- Nelson, L. "A Control Chart for Parts-Per-Million Nonconforming Items." *Journal of Quality Technology* 26, no. 3 (1994): 239–240.
- Perla, R., L. Provost, and S. Murray. "The run chart: a simple analytical tool for learning from variation in healthcare processes" *British Medical Journal Quality and Safety in Healthcare* 20, no. 1 (2011): 46–51.
- Provost, L. P., and S. Murray. *The Health Care Data Guide*. San Francisco: Jossey-Bass, 2011.
- Pyzdek, T. *Pyzdek's Guide to SPC*. Vol. 1: *Fundamentals*. Milwaukee: Quality Press, 1990.
- Shewhart, W. *Economic Control of Quality of Manufactured Product*. New York: D. Van Nostrand, 1931. Reprint, Milwaukee: Quality Press, 1980.
- Shewhart, W. *Statistical Method from the Viewpoint of Quality Control*. New York: Dover Publications, 1986.
- Western Electric Company. *Statistical Quality Control Handbook*. Indianapolis: AT&T Technologies, Inc., 1985.
- Wheeler, D. *Understanding Variation: The Key to Managing Chaos*. Knoxville, TN: SPC Press, 1993.
- Wheeler, D. *Advanced Topics in Statistical Process Control*. Knoxville, TN: SPC Press, 1995.
- Wheeler, D., and D. Chambers. *Understanding Statistical Process Control*. Knoxville, TN: SPC Press, 1992.
- Woodall, W. "The Use of Control Charts in Health-Care and Public-Health Surveillance." *Journal of Quality Technology* 38, no. 2 (2006): 89–104.