The Sector Chart

A New Engineering Graph for Pharmaceutical Processes

Lynn Torbeck

The author proposes a new analytical graphic, the sector chart, which presents data that cannot be adequately presented with current graphs. This chart combines features of zone charts with the basic principle of precontrol charts. It addresses engineering control and is superior for representing data such as beneficial and adverse trends.

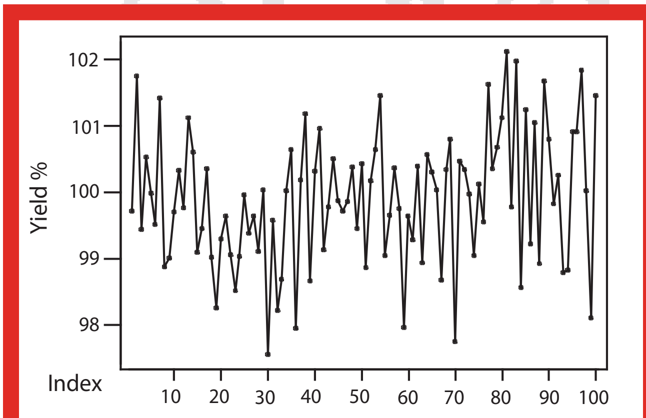

Figure 1: Stationary time series; $\mu = 100$ and $\sigma = 1.0$

Mr. Lynn Torbeck is a statistician, international consultant, and member of *Pharmaceutical Technology*'s Editorial Advisory Board, Torbeck and Assoc., Inc., 2000 Dempster Plaza, Evanston, IL 60202, tel. 847.424.1314, Lynn@Torbeck.org.

A simple analytical graphic is the time plot. The reportable value for the response variable is scaled on the vertical axis and time in some metric is represented on the horizontal axis. Time plots are powerful tools for observing and understanding process changes or lack of change over time. These plots make no assumptions about the data except that the data are collected sequentially over time. Time plots, when used by a knowledgeable and experienced analyst, are the best first step in an analysis of data collected over time.

Beyond a time plot, the classic Shewhart statistical control chart is best known and most widely used. First presented by Walter A. Shewhart, PhD, in 1924 in an internal memorandum at Barr Labs, it has become, all too often, the default choice for any type of data taken over time. Shewhart charts are excellent for their intended use but inadequate for many situations.

Data

Data taken over time—*i.e.,* a time series—make up a large portion of the data collected and analyzed in all industries. These data can be continuous or discrete and can be taken over any span of time from microseconds to years or even decades. Ideally, that data conform to some known statistical model that can be used to evaluate the time series and any changes it may exhibit.

The usual model is a homogeneous stationary time series of independent values from a normally distributed population with an unknown mean, μ, and variance, σ^2. The population mean is estimated by the average of a selected sample, and the population variance is estimated by the variance of the sample. Because the data are expected to be stationary, there should not be any upward or downward long-term movement. The variability is assumed to be the cumulative summation of dozens if not hundreds to thousands of small sources of variation, all increasing and decreasing at random as the data are collected. These sources of variation are usually called *common causes.*

Departures from this ideal can come in many forms. The simplest departure is a single cause, an event or change, that results in a value (or values) not typical of the other values. This has always been called *the special cause*. Upon reflection, it should be recognized that an unusual value could be the result of two or more causes, an interaction effect known in the terminology of designed experiments. But, textbooks continue to promote the false dichotomy of a single special cause versus the hundred common causes. In reality, we should expect the magnitude of the causes to follow a Parteo distribution.

Other departures include a constant change, increasing or decreasing, over time, a sudden shift in the mean, or an increase or decrease in the variance. Cycles over time can occur depending on the source of the variation.

Time plots. A time plot presents the data over time but does not provide a basis for comparison or evaluation, as shown in the graph provided in Figure 1.

Statistical control charts. A time plot becomes a Shewhart control chart if control limits are added (1). These limits are in theory the $\mu \pm 1$, 2, and 3σ lines. As previously noted, however, the population mean is estimated by the sample average and the population variance by the sample variance. Thus, the graph plots the average ± 1, 2, and 3 $\hat{\sigma}$. The average of the data is represented by the center line. Shewhart charts can be plotted for individual values, averages, standard deviations, variances, ranges, moving ranges, fraction defective, number of defectives, defects per unit, and the number of defects. Other less-common charts exist because any summary statistic with a known distribution can be used as a control chart. An example of an individual value and moving range chart is shown in Figure 2.

The control limits permit objective analytical evaluation of the time series. There are eight common rules for comparison. The most recognized is a single point outside of the three sigma lines. Six points in a row all increasing or decreasing is another rule. In all cases, the underlying model is assumed to be a stationary time series with a known statistical distribution.

Zone charts. Zone charts continue the Shewhart model with the average ± 1, 2, and $3\hat{\sigma}$ lines (see Figure 3) (2). Instead of eight rules, one rule is used on the basis of an accumulated tally. Successive values are given a cumulative score depending on how far they are from the center line plus the cumulative score of the previous value. There are four zones, based on the three control lines. The tallies in the zones can be selected depending on the perceived sensitivity required. When the accumulated tally is equal to or greater than the score in zone 4, a flag is raised. The tally is reset to zero when a value falls in the center two zones.

Cumulative summation chart. A cumulative summation chart continues the stationary time series model but is designed to detect one or more shifts in the average (3). It is, in fact, quite sensitive to a small shift in the average (see Figure 4). Like the zone chart, this chart plots an accumulated tally. But, here the sum of the differences from the grand average of all of the data is accumulated. If the average has shifted in the series, the plot will show a dramatic change in direction, as in Figure 4.

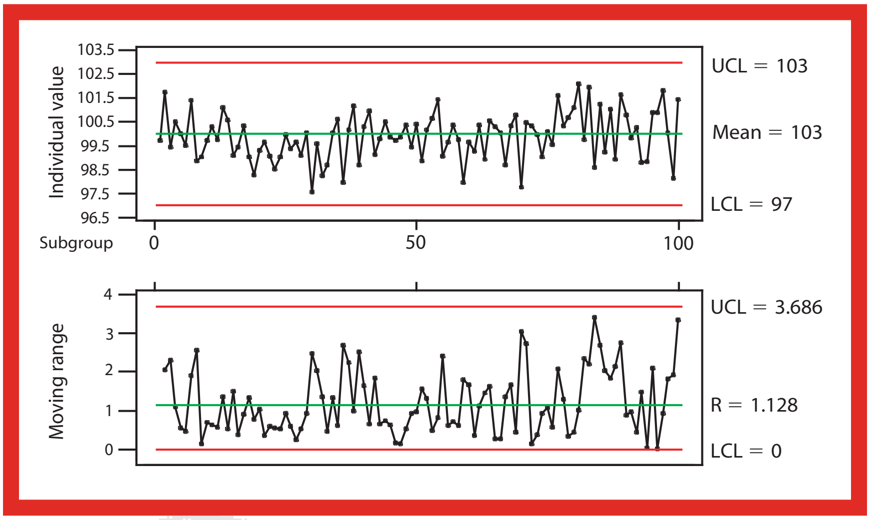

Figure 2 Individual value and moving range charts for yield %.

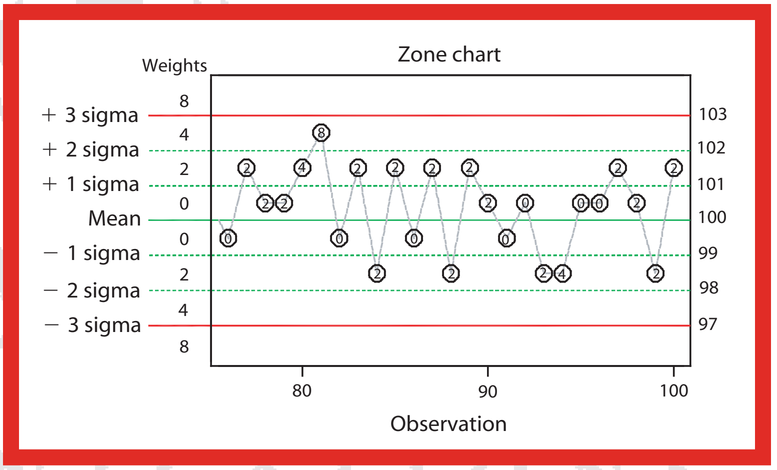

Figure 3: A zone chart.

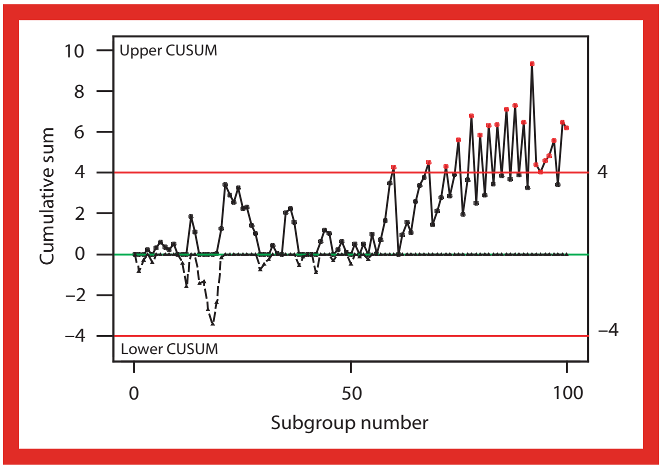

Figure 4: A cumulative sum chart.

Precontrol charts. All of the graphs discussed above have limits calculated from the natural variability of the data itself and without any reference to the specifications. Precontrol charts introduce the use of the predetermined specification limits to

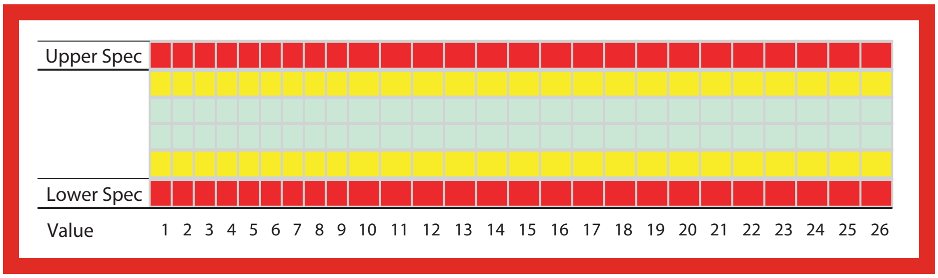

Figure 5: Precontrol chart.

find the control limits (see Figure 5) (4). The specification limits are divided into four equal zones. The center two zones are green, the two outer zones are yellow. Anything outside of the specification limits are red.

No assumption about the distribution of the data is needed to use the precontrol chart, and no summary statistics are calculated. It is assumed, however, that the process is easily adjusted and can be well-controlled once adjusted. This is not always a reasonable assumption. Precontrol is used for both initial setup and routine production. The rule for setup requires that five consecutive values lie within the green zones. If so, begin production. If any of the five are outside of the green zones, adjust the process and take five more values. Continue until all five are in the green zone. After production begins, take two consecutive values from the process at periodic intervals. If both are in the green zone, continue production. If one is green and one is yellow, continue. If two are yellow in the same zone, adjust the process. If two are yellow in different zones, stop production and do a root-cause investigation. If one or two are red, stop production and do a formal investigation. If production stops, restart after the investigation and adjustment with five in the green zone again. As noted, the limits are based on the specification limits and not the variability of the data. Precontrol is an engineering control chart and not a statistical control chart. The distinction is an important one.

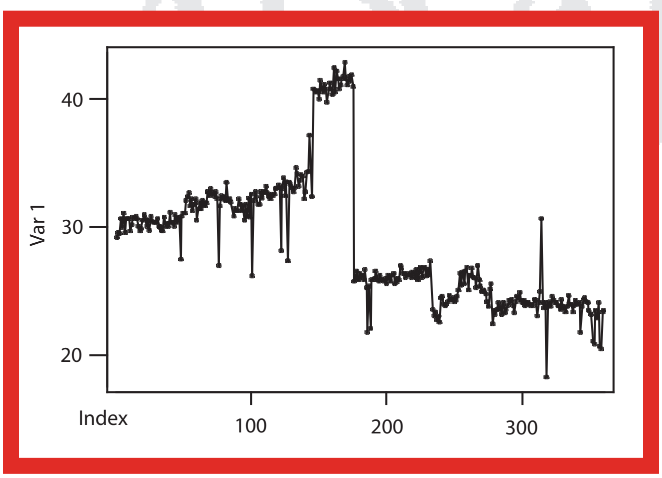

Figure 6: Data not in statistical control.

Engineering control versus statistical control

Some processes never reach a state of Shewhart statistical control despite heroic efforts. This is particularly true of biological processes. But, often the average level and the variability of the data are so far inside the specification limits that acceptable product is being made and distributed. If a manufacturing process can operate well within its specifications, it need not achieve a state of statistical control. As in the precontrol situation, if all of the data are in the green zones or better, the actual nature of the data is not really an issue. Normality, or the lack of it, shifts in the average, changes in the variance and cycles are not cause for concern, action, or investigation. Such a process can be said to be in a state of engineering control. It is not predictable with a statistical model, but it has shown historically that it is more than capable of meeting the specification limits and the perceived subjective probability of producing out-of-specification product is acceptably low even if not calculable from the data. In Figure 6, the specification limit is ≤ 60. Clearly, this process is not in a state of statistical control and probably never will be. However, it is meeting its specification limit and in all likelihood will continue to meet its specification limit in the future. If, however, the data start getting close to the specifications, then an objective decision rule would be needed to evaluate the data. Because statistical assumptions about the data cannot be made, an approach other than a statistically based graph is needed.

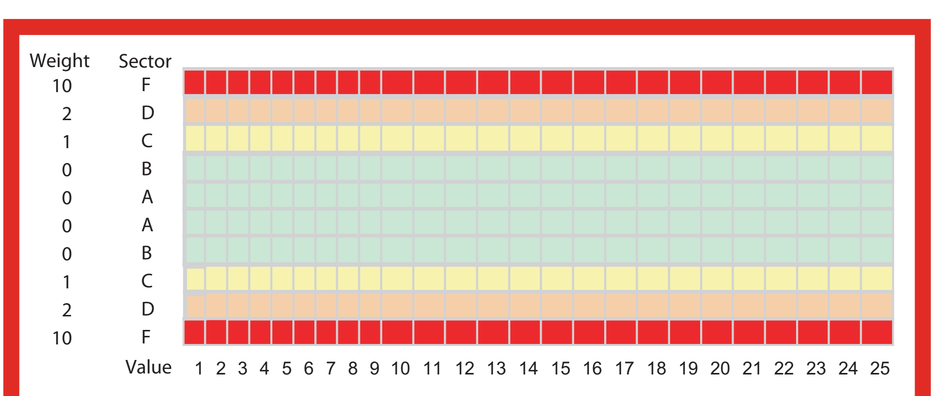

Figure 7: A two-sided sector chart.

Sector charts

Sector charts are proposed and defined here for the first time to address this issue (5). (A literature search has not found this type of graph to date.) Sector charts incorporate concepts from the charts previously described. This new chart is an engineering control chart and not a statistical control chart. It is nonparametric in that no assumptions are made about the data and no statistical summary statistics are needed (see Figure 7).

It is somewhat like a zone chart but with the zones, here called *sectors*, as

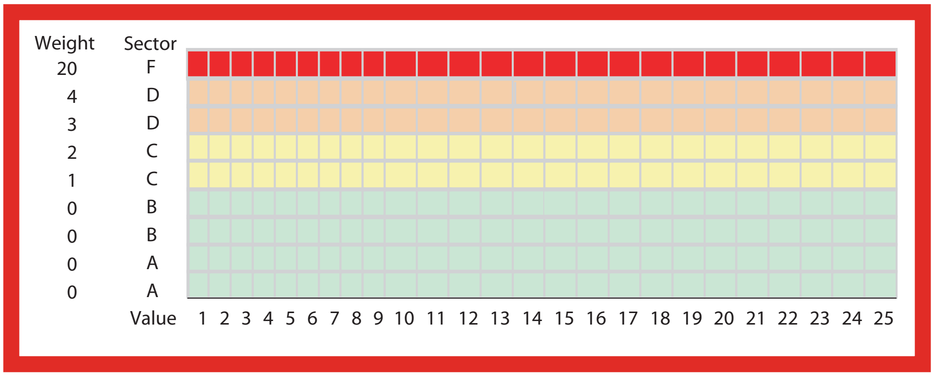

Figure 8: A one-sided sector chart.

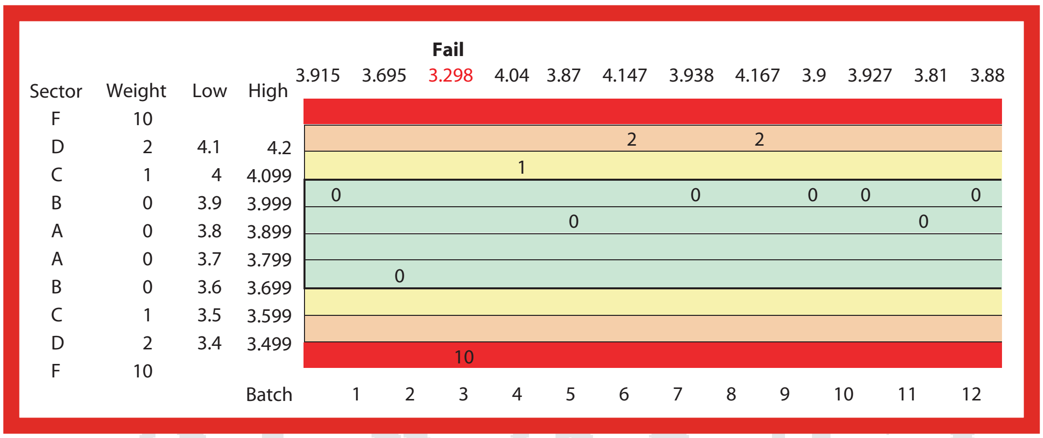

Figure 9: A sector-chart example.

divisions of the specification range as in precontrol and not as multiples of the estimated standard deviation. The sector chart is started by dividing the specification range into sectors. The sectors are usually of equal size but would not have to be. The specification can be two-sided as in Figure 7 or one-sided as in Figure 8. An example with data is shown in Figure 9.

Although four zones can be used as in precontrol, it is probably advantageous to have more. Ten or more could be considered in some situations. Unlike precontrol, the sectors are assigned colors for ease of visual inspection rather than as decision rules. The decision rule is like that for the zone chart: a tally is accumulated. Successive values are given an accumulative score depending on the sector they are in, plus the cumulative score of the previous value. Sector weights can be set depending on the needed sensitivity. If a value falls inside the green sectors, usually the center 50% of the specification range, the tally is reset to zero. Once the accumulated tally reaches a trigger point, action is taken. This could be a simple root cause investigation or major research project with production halted. The trigger point need not be the value for the red sector as in the zone chart, but it typically is. If a point falls outside of the specification in the red sector, action is taken.

Conclusion

A sector chart has several advantages. First, it is an engineering control chart and does not depend on the data or any assumptions about the statistical nature of the data to determine the limits. The focus is on avoiding an out-of-specification result, not on achieving and maintaining statistical process control, as desirable as that may be. It works equally well with continuous and discrete data. It does not require a minimum number of data values. It works with one value and a decision can be made after the addition of each data value. It is flexible in that the number of sectors, the values for the weights, and trigger point are selectable. Sector weights and trigger points are determined by expert selection and should reflect the best scientific knowledge available. Like a zone chart, it can detect an improved process. If a process has reduced variability and stays within one or two sectors other than green, it would eventually trigger action, thus highlighting the improvement. It is forgiving, given that it resets to zero when the values are in the green sectors. Finally, the rules are simple and easily understood.

References

1. W. Shewhart, *Statistical Method from the Viewpoint of Quality Control* (Dover Publications, New York, NY, 1939, reprinted 1986).
2. A. Jaehn, "The Zone Control Chart," *Quality Progress* **24** (7), 65–68 (1991).
3. L. Alwan, *Statistical Process Analysis* (Irwin McGraw-Hill, New York, NY, 2000).
4. K. Bhote, *World Class Quality* (Amacon, American Management Assoc., New York, NY, 1991).
5. L. Torbeck, first drawn 17 October 2003. **PT**