State-of-the-art - Cardiac general

Control charts, Cusum techniques and funnel plots. A review of methods for monitoring performance in healthcare

Luc Noyez*

Heart Center, Radboud University Nijmegen, Department of Cardio-Thoracic Surgery – 677, PO Box 9101, 6500 HB Nijmegen, The Netherlands

Received 5 February 2009; received in revised form 11 May 2009; accepted 14 May 2009

Summary

Quality control in medicine is generating more and more interest. Industrial concepts of quality control have been refined and transformed to be useful in healthcare monitoring. Whereas medical practitioners first reaction to this new concept of quality control was negative ‘we’re treating patients, we’re not a part of an industrial process’, some dramatic cases of inferior medical performance urged the need to adequately monitor healthcare outcomes. To date, several methods have been described, and more and more reports deal with the 'we're treating patients, we're not a part of an industrial process', some dramatic cases of inferior medical performance urged the need to be useful in healthcare monitoring. Whereas medical practitioners first reaction to this new concept of quality control was negative ‘we’re treating patients, we’re not a part of an industrial process’, some dramatic cases of inferior medical performance urged the need to adequately monitor healthcare outcomes. To date, several methods have been described, and more and more reports deal with the subject. Most of us, however, are overwhelmed by the new and different tools in use such as Shewhart control charts, cumulative sum charts and funnel plots. This paper will review the methodology of statistical process control and its application in medical practice.

Keywords: Quality control; Cardiac surgery; Control charts; Cusum; Funnel plots

1. Introduction

Ever since the case of general practitioner Harald Shipman (BRI Inquiry Panel. Learning from Bristol: The Report of the public Inquiry into Children’s Heart Surgery at the Bristol Royal Infirmary 1984–1995. London, UK: The Stationery Office, 2001. Available from http://www.bristol-inquiry.org.uk/final_report/...and the Bristol affaire (Shipman inquiry: The First Report. London, UK: The Stationery Office, 2002. Available from http://www.the-shipman-inquiry.org.uk/reports.asp.), quality control in medicine has generated considerable interest. As a consequence the need for standardized systems to monitor health care quality is growing. Several programs, mostly based on industrial processes, are used though at this point there is no universally accepted method for monitoring performance in health care [1]. However, process charts, cumulative sum techniques and funnel plots are the most commonly used techniques [3–5]. In this paper the methodology and the use of these methods in practice will be reviewed. To clarify our paper a fictitious data set of hospital mortality post cardiac surgery is presented. This data set consists of 1772 adult cardiac operations over a period of 57 months, operated by six surgeons. Mean additional EuroSCORE was 3.00 (range 0–16). Hospital mortality was 3.3% (58/1772). Accepted mortality rate was 2.5% and unaccepted mortality rate 5%. Hospital mortality for the six surgeons: A: 14/376 (3.7%), B: 6/148 (4.0%), C: 9/215 (4.1%), D: 17/409 (4.1%), E: 10/363 (2.7%), F: 2/211 (0.9%).

*Corresponding author. Tel.: +31-24-3613711; fax: +31-24-3540129. E-mail address: l.noyez@thorax.umcn.nl (L. Noyez).

© 2009 Published by European Association for Cardio-Thoracic Surgery

2. Background

2.1. Statistical process control

Statistical process control (SPC) is used to check a process during its run and does so by using control charts [5]. This way, if there are signs indicating a problem with the process, it can be stopped and checked. In other words it is important that, while the process is running, one can constantly check whether it is performed as it should be or whether it is deviating from its normal performance. If the process seems to be going out of control, it can thus be stopped, checked or adjusted before there is an actual problem. An example can be the evolution of the waiting period for cardiac surgery over several years. The most important use of these control charts is in identifying the trend. If the trend suggests that the process is getting worse (longer waiting times) it will be necessary to closely analyze the process. The same goes, however, for a ‘positive’ change. For instance if the trend is steadily improving (shorter waiting times), it is also interesting to identify what the reason for this changes.

2.2. Natural fluctuation within a process

Every process can be seen as a wave-like motion; sometimes it is up, sometimes it is down. These ‘natural’ movements or fluctuations can never be eliminated. This may seem very logical for a process such as a waiting list, but a variation can also be identified in so-called stable processes. As an example, we can look at the bottles of beer in a brewery. Even when they are filled using a machine, they are never equally full, there is always a
variation. This normal variation of distribution can be described by a well known Gaussian curve. Because the normal distribution is the result of probability, the curve is symmetrical: half of the area is on the left side of the mean while the other half is on the right. Using the mean ± 3 standard deviation (S.D.) of the mean, we have 99.7% of the variation, limited by the upper control limit (UCL) (+3 S.D.) and the lower control limit (LCL) (−3 S.D.). If we use ±2 S.D., we cover 95% of the variation.

2.3. Common causes of variation and specific range

The common cause of variation is in fact the natural fluctuation within a process. This fluctuation can never be eliminated, though, it can be reduced. An interesting question is of course which variation is acceptable within a specific process? This range depends on the acceptable variation that can be tolerated in order to guarantee the good performance of a process. If the beer bottles are filled too little, it is not good performance of a process. If the beer bottles are filled too much, it is not good. [lower tolerance limit (LTL)], and the same goes for bottles that are filled too much [upper tolerance limit (UTL)]. The range of how far the bottles must be filled to be good is called the specific range. The widest variation is the normal variation, calculated by mean ±3 S.D. If the specific range is similar to the common cause variation, the average of the process is equal to the midpoint between the UTL and the LTL.

2.4. Assignable-cause/special-cause of variation

An essential question is how to detect whether a process is under control or not. In theory, all points situated outside of the normal variation, UCL and LCL, indicate that the process is possible out of control. However, one should realize that it is never completely certain that the process is out of control because there is always a very small chance (0.3%) that this point is just situated between the 99.7% and 100% limit of the normal distribution. This also means, however, that the chance to fall out of the normal range is minimal (0.3%).

Of course control limits can be placed at any distance from the mean, though the closer these limits are, the higher the chance that a good process is identified as being out of control. For example, if the CL are placed at 2 S.D., the chance is 5%, if the CL are placed at 1 S.D., the chance increases to 32%. To avoid a good process being stopped it is generally accepted that upper and lower control limits are placed at 3 S.D. from the mean.

2.5. Type 1 and type 2 errors

A logical question is of course, why not narrow our limits? It is good to realize that fluctuation outside the UCL and the LCL is the result of factors deviating from the common cause. Control charts use the term ‘out of control’. This, in fact, only means that the variation is statistically greater than that which could be accounted for by the common cause variation. So it is possible (0.3%) that the point is just one exceptional point falling outside the limits. In this situation it would thus be wrong to stop the process: it is a type 1 error (α). On the other hand, by narrowing the UCL and LCL using for example 2 S.D. instead of 3 S.D., the chance of a type 1 error increases to 5%. A type 1 error is thus an error which leads you to make an incorrect decision resulting in a situation that did not warrant it, a false positive conclusion. A type 2 error (β) on the other hand results also in a wrong decision, without actual consequences for the process, it is a false negative conclusion. Ignoring a result which in reality is due to an assignable cause is a type 2 error.

2.6. Binominal data

The above is applicable for numeric data. However, and especially in medicine, a lot of data are binomial, ‘yes’ or ‘no’, ‘alive’ or ‘death’. For these data it is of course impossible to calculate a mean with S.D. In such cases the mean is estimated from the average of the proportion of ‘events’ from A samples with each B items, where A should be at least 30 and B at least 100. In other words, it is simply the percentage (p) of events. The S.D. is then estimated as \(\sqrt{\frac{p(1-p)}{n}}\). UCL and LCL are then set at +3 S.D. and −3 S.D.

2.7. Use and weakness

SPC is based on the idea that process variability indicates whether a process is under control or not. Plots falling outside of the control limits may be out of control and must be investigated.

There are two weaknesses to this SPC. First of all, the SPC assumes that all values within the limits are good and equal, while all values outside the limits are bad. However, it is commonly understood that a value close to the mean is better than one just within the limits. A second point is, that when a process stays within the control limits, we indicate that the process is not deteriorating. We have, however, nothing to improve the process either.

Another minus of the SPC is that, especially for a process with a small natural variation, even variations due to normal process behavior are reported as out of the control limit. On the other hand, for processes with a wide natural variation, natural process variations are masked, but the risk exists that important negative variations are not noticed on time.

3. Quality control chart for surgical (medical) performance

In industry, quality control can be translated as aiming to be ‘on target with minimum variation’. Reduction of variation is also a priority in clinical governance. The medicinal field, however, is characterized by much greater inherent variabilities, case mixes, differences in risk, etc. than most industrial processes.

3.1. The Shewhart control chart

In medical and healthcare literature, the SPCs described above are generally known as Shewhart control charts [2, 6, 7]. Several types of control charts can be used, depending on the type of data, length of hospital stay (continuous), mortality (binominal) or even so-called count data (a sum of complications). For the construction of these charts
values of the mean, the upper and/or the lower control limit (boundary lines, alert lines) must be known. In fact, these charts were designed for monitoring a batch of results, for example the hospital mortality post cardiac surgery over several years (Fig. 1). However, as discussed above, the value of these charts for ongoing monitoring or monitoring of individual results is limited, especially in low-volume situations. An alternative is the Cumulative Sum chart (CUSUM-chart).

3.2. The CUSUM charts

CUSUM charts are based on sequential monitoring of cumulative performance over a period of time. The difference with this analysis is that each procedure can be updated and that there is a real-time monitoring of performance [8–14]. This is very important since these charts identify subtle, slow, sustained degradation in a process that is thought to be under control. However, before using a CUSUM analysis, the event must be defined as a binary variable. This can be a clear binary event, such as, death or alive, or the use of bilateral mammary artery as graft or not. But a composite event is also possible, such as major morbidity, for example defined as sternal wound complications, stroke or renal failure.

3.2.1. Cumulative failure chart

The cumulative failure chart is the simplest form of a CUSUM chart. The chart is constructed by the cumulative number of events (failures), on the vertical axis, plotted against the total number of procedures, patients, on the horizontal axis. With each event the graph rises, resulting against the total number of procedures, patients, on the performance with this analysis is that each procedure can be updated and that there is a real-time monitoring of performance. The difference with this analysis is that each procedure can be updated and that there is a real-time monitoring of performance. The difference with this analysis is that each procedure can be updated and that there is a real-time monitoring of performance.

The cumulative failure chart is the simplest form of a CUSUM chart. The chart is constructed by the cumulative number of events (failures), on the vertical axis, plotted against the total number of procedures, patients, on the horizontal axis. With each event the graph rises, resulting against the total number of procedures, patients, on the performance with this analysis is that each procedure can be updated and that there is a real-time monitoring of performance. The difference with this analysis is that each procedure can be updated and that there is a real-time monitoring of performance.

The choice of the error rates is free, but one should realize that it has an influence on the false-positive and false-negative conclusions. It is thus possible to use lower error rate values for the construction of an alert line and higher values for the construction of an alarm or caution line. In this case it is important that the values of these error rates are clearly specified. In medicine, however, both error rates are usually made equal to 0.1.

It is important to realize that using these cumulative failure charts it is twisted reasoning that if a cusum graph remains between the boundaries, the process is under control. If the graph of cumulative failures exceeds the upper boundary line, one can conclude that the failure rate is higher than the unacceptable rate (p_u). If the graph crosses the lower boundary line, the failure rate is lower than the acceptable rate. An acceptable process shows a graph with a slope towards the lower boundary line.

The following formula presents the calculation for the alert lines of Fig. 2.

\[
\text{Cumulative failure rates} \quad \text{acceptable} = \frac{\text{sn} \pm b}{(p+Q)}
\]

\[
\text{Cumulative failure rates} \quad \text{unacceptable} = \frac{\text{sn} \pm a}{(p+Q)}
\]

\[
\text{Cumulative failure rates} \quad \text{unacceptable} = \frac{\text{sn} \pm a}{(p+Q)}
\]

\[
\text{Cumulative failure rates} \quad \text{unacceptable} = \frac{\text{sn} \pm a}{(p+Q)}
\]

\[
\text{Cumulative failure rates} \quad \text{unacceptable} = \frac{\text{sn} \pm a}{(p+Q)}
\]

\[
\text{Cumulative failure rates} \quad \text{unacceptable} = \frac{\text{sn} \pm a}{(p+Q)}
\]

\[
\text{Cumulative failure rates} \quad \text{unacceptable} = \frac{\text{sn} \pm a}{(p+Q)}
\]

\[
\text{Cumulative failure rates} \quad \text{unacceptable} = \frac{\text{sn} \pm a}{(p+Q)}
\]
3.2.2. Standard non-risk-adjusted CUSUM chart

The standard CUSUM works with a constant risk of failure for each case (procedure, patient). The basic principle of a CUSUM analysis is that of reward or punishment in function of the risk [9, 14]. This is in contrast to the cumulative failure chart, which only takes into account the failures. For example, patients have a risk of 10% to die during an operation. When the patient survives the operation, we are rewarded with 0.10 points. However, when the patient dies we are punished with \(-1\times0.10=-0.10\) points. By summing the values of all patients over a certain period of time, we construct a graph. \(X_t = \max(0, X_{t-1} + W_t)\), where \(t=1,2,3,...\) case number, \(X_0=0\) and \(W_t\) is the score, reward of punishment for each patient.

The horizontal axis represents the cases over time. Above the null-line, the vertical axis indicates the lives saved compared to those expected, while if the graph is below the null-line, it shows the excess deaths. Thus, if the first nine patients survive, the graph is raised by nine times 0.1 points=0.9 points, when as expected the tenth patient dies, the graph is lowered by 0.90 points again. Normally the CUSUM graph should thus form a wave-like pattern around the null line. An upwards slope indicates that the observed deaths are fewer than expected, this means an improvement of the process. A downward slope indicates that the observed mortality is higher than expected, thus a worsening of the process. This means that in these graphs the cumulatively expected mortality minus the observed mortality is plotted on the vertical axis. Therefore the vertical axis can be labeled as ‘lives saved’, Lovegrove called these graphs variable life-adjusted display (VLAD) [15]. We must realize that it is also possible to plot the observed minus the expected mortality, as is done by Novick et al. [14] resulting in a reversal of the graph.

Because the trends in a CUSUM graph are easily recognizable – an upwards slope indicating improvement, a downward slope deteration, most CUSUM curves are constructed without control limits (Fig. 3). However, just like for control charts horizontal control lines can be added. The value of the control lines is defined by ‘h’, which in turn is defined by the accepted (\(p_0\)) and the expected (\(p_1\)), failure rate according with the type 1 error and the type 2 error rates. The spacing between the unacceptable control lines is defined by \(h_0\), and that between acceptable control lines by \(h_1\). Again, however, both error rates are usually set at 0.1, resulting in only one set of control lines. For example, the calculated values for alert lines in a standard CUSUM chart presented in Fig. 3 (\(\ln\) = the natural logarithm)

\[
\begin{align*}
a & = \ln \left(\frac{100\% \text{ type 2 error rate}}{10\% \text{ type 1 error rate}}\right) = \ln \left(\frac{1}{0.1}\right) = 2.1 \\
b & = \ln \left(\frac{10\% \text{ type 1 error rate}}{20\% \text{ type 2 error rate}}\right) = \ln \left(\frac{10}{20}\right) = -0.53 \\
P & = \ln \left(\frac{p_1}{p_0}\right) = \ln \left(\frac{0.05}{0.025}\right) = 0.7 \\
Q & = \ln \left(\frac{1-p_0}{1-p_1}\right) = \ln \left(\frac{0.95}{0.98}\right) = -0.02
\end{align*}
\]

\(H_0\) = defines the space between the acceptable control lines \(= b/(p+Q) = 2.1/(0.7+0.02) = 2.9\)

\(H_1\) = defines the space between the unacceptable control lines \(= a/(p+Q) = 2.1/(0.7+0.02) = 2.9\)

Because type 1 and type 2 error rate are equal (0.01)

\(H_0 = H_1\)

This method can be used when there is a constant risk of failure. However, the individual risk in medicine is seldom constant. Each patient, for instance, has his own risk for mortality. In this situation we construct a risk adjusted CUSUM chart.

3.2.3. Risk adjusted CUSUM chart

The principle of a risk adjusted CUSUM analysis is the same as that of the non-risk adjusted CUSUM analysis [11-13]. The difference in risk adjusted CUSUM charts is that for a patient with a high mortality risk (40%), surviving the operation, we are rewarded 0.40 points and thus more for a low-risk patient. If this high-risk patient dies, however, we are only punished by \(-1\times0.40=-0.40\) points, thus less than for a low-risk patient. For these risk adjusted CUSUM charts control limits can also be constructed. However, because of the different risk for each case, it becomes a complex issue. Sherlaw-Johnson et al. proposed a prediction interval for the last case of the analysis, giving an indication of the inherent variability [16]. Grunkemeijer et al. proposed to add points for each case [11]. In order to do this, they used the standard error (\(S.E.\)) of the risk adjusted CUSUM at time \(t\), \(S.E. = square\ root\ of\ the\ cumulative\ sum\ of\ the\ expected\ *\text{1–expected mortality}\) multiplied by 1.64, 1.96 or 2.58 to obtain the 90%, 95% or 99% two-sided prediction limits (Fig. 4). These point-wise calculated limits give an idea of the variability, but do not account for multiple testing.

Especially when dealing with an ongoing process, however, these point-wise constructed confidence or prediction limits do not suffice because of the multiple comparison.
Spiegelhalter and Steiner have proposed risk-adjusted sequential probability ratio testing [17, 18]. This method uses a cumulative sum, not of intuitive units such as ‘lives saved’ but of units of ‘logarithm of the likelihood ratio’ of the alternative of the null hypothesis. The likelihood ratio is the ratio between the maximum probabilities of a result under two different hypotheses. The null hypothesis (h0) states that the odds ratio (OR0) = 1, meaning that the observed mortality is equal to the excepted mortality. The second alternative hypothesis (h1), can be for example OR1 = 2 to detect a doubling (increase) of deaths than expected, or an OR2 = 0.5 to detect a halving of deaths than expected. To construct the graph we use the formula

$w_t = \ln \left( \frac{OR_t}{1 - p_t + OR_t \cdot p_t} \right)$

when the patient survives, and

$w_t = \ln \left( \frac{1/(1-p_t + OR_t \cdot p_t)}{1 - p_t + OR_t \cdot p_t} \right)$

when the patient dies (Fig. 5).

The advantage of the risk adjusted CUSUM compared to the non-adjusted analysis is, that in case of a series of high-risk patients there is no signal of decrement in the performance. However, we must again realize that every risk adjustment is imperfect and cannot remove all confounding.

3.2.4. Funnel plots

The methods described above are useful for monitoring healthcare processes. Nonetheless, when used for comparisons in a health care context, they lead to ranking tables according to the performance indicator. For example, we can construct a CUSUM chart with six curves, one for each of the six surgeons of our fictitious dataset. This ranking is just a major problem in the evaluation of healthcare processes or individuals. Using a funnel plot, this ranking approach is more or less bypassed [4, 19]. In a funnel plot the observed performance indicator is plotted against a measure of its precision, and the control limits form a ‘funnel’ around the target outcome. The funnel is constructed by using upper and lower control limits, calculated as the 99.8% confidence limit (3 S.D.), as an alert limit. In most cases an additional alarm limit ‘funnel’ based on the 95% confidence limit (2 S.D.) around the mean predicted event using a binomial method is constructed. This mean is recalculated for each new case added in the series and likewise, for the observed event rate (Fig. 6). This provides us with an attractive method for comparison without the problem of ranking.

4. Conclusion

The intention of this article was to review the most commonly used methods for quality control in health care, and especially in cardiac surgery. The most commonly used methods, Shewart charts, CUSUM analysis and funnel plots,
were reviewed. At this moment the CUSUM technique is the most valuable and accepted tool in the assessment and monitoring of a process. Nevertheless, funnel plots are more suitable for comparisons between institutions or surgeons, because these funnel plots avoid the problem of ranking. But no matter what method is used, it is essential to define the studied cohort and the studied event, if possible by using internationally-used definitions, and to avoid vague outcomes as ‘a cardiovascular complication’. Clearly describe the used method, label vertical and horizontal axes, and define the eventually used control limits. Make sure it is clear whether a downward slope indicates a worsening or improving process. Lastly, and probably most importantly, realize that all these methods are based on the statistical principle of process variability and that plots falling outside of the control limits, may be out of control and must therefore be investigated. All these methods result in a warning signal, like the flickering light on your dashboard, when fuel reserve is low.

Acknowledgment

Elise Noyez is thanked for her correction of the English text.

References