

Applications and Experiences of Quality Control to Surgical and Interventional Procedures

David Biau^{1,3}, Lionel Dessolle² and Raphaël Porcher^{1,3}

^{1,3}*Département de Biostatistique et Informatique Médicale, Hôpital Saint-Louis, Assistance Publique Hôpitaux de Paris, Paris*

³*UMR-S 717, Inserm; Université Paris Diderot – Paris 7, Paris*

²*Service de gynécologie obstétrique, Hôpital Tenon, Assistance Publique Hôpitaux de Paris, Paris France*

1. Introduction

Statistical process control have emerged in the medical literature after a wide expansion in the industry. In the nineteen-twenties, Walter A. Shewart, with a group of experts, established the scientific ground for quality control at Bell Telephone Laboratories Shewhart (1931). It is only in the nineteen fifties that quality control procedures were employed to ensure the precision of hospital laboratory machines used in biology, nuclear medicine, drug industry and other medical domains Anderson (1982); Batson et al. (1951); Brookeman (1978); Hollinger & Lansing (1956); Loynd (1962); Pribr & Shuster (1967); Waid & Hoffmann (1955). Later, in the nineteen seventies, the use of these methods shifted to the monitoring of the effect of treatments on patients Kinsey et al. (1989); Morgan et al. (1987); Robinson & Williamson (1974); Walters & Griffin (1986); Wohl (1977), and then to other more complex levels such as the performance of departments Chamberlin et al. (1993); The Inquiry into the management of care of children receiving complex heart surgery at the Bristol Royal Infirmary (2001), hospitals Sellick (1993), regions Tillett & Spencer (1982) or nation-wide processes Hand et al. (1994). It is only recently that these methods have been used to monitor the performance of physicians themselves in various disciplines. In this case, the success of the procedure is directly imputed to the the competency of the care provider. Statistical process control methods have been used in surgery de Leval et al. (1994); Novick & Stitt (1999); Van Rij et al. (1995), obstetrics Lane et al. (2007), endoscopy Williams et al. (1992), anaesthesiology Kestin (1995); Lagasse et al. (1995), etc.

Recently the Institute of Medecine, Washington DC, USA, in its report “Crossing the Quality Chasm: A New Health System for the 21st Century” stated that “*reducing waits and sometimes harmful delays for both those who receive and those who give care*” was one of the six aims proposed to adress key issues of today’s health care system of Medicine (2001). The implementation of quality control procedures are meant for that purpose: detecting inadequate performance, should it occur, as soon as possible so that corrective actions may be undertaken and patient care improved.

The use of quality control methods in medicine presents some specificity and challenges when compared to that of the industrial context. When one monitors the performance of a care provider, the manufactured good is biologic: the patient. First, the raw product onto

which the procedure will be performed, the patient, presents characteristics whose only a very limited subset are usually known. For instance when a physician is interested in monitoring endoscopic cholecystectomy, the patient undergoes a CAT scan and a few blood tests which provide very limited information on the product. Second, since some significant characteristics are retrieved before an intervention, we know that based on this limited information only, patients are nonetheless very different from one another. These differences may lead to variation in the outcome which should not always be attributed to a change in the performance of a surgeon for instance. Third, the process under scrutiny is poorly contained and under the effect of numerous unknown or unanticipated variables, either seen as chance or assignable causes. A surgeon who desires to monitor the implantation of total knee replacements to ensure that his or her performance is correct, may see failures because of poor bone quality, because of a traumatic fracture, or because of an infection, all being unrelated to his or her surgical skills. Last, the measure of quality is equivocal, difficult and subject to variation. For instance, measuring the success of an interventional procedure such as fiberoptic intubation may include the actual intubation per se, the time required to complete the procedure, the occurrence of a complication, or a composite of these criteria.

Another very specificity with using quality control methods in medicine is that often physicians are interested in determining when one can say, with some confidence, that a trainee has reached a certain level of performance. In that case, we are monitoring a process that we know is out of control at the beginning of the observation period and we want to determine when this process reaches an in control state. In standard control chart usage this corresponds to the phase I of control chart application where engineers gather data, define control limits, identify and correct assignable causes and repeat this cycle until the process is brought to a stable, in control performance state Montgomery (2005). In medicine, this situation corresponds to the learning curve of a trainee and a new control chart has been recently developed for that purpose Biau & Porcher (2010).

In the sequel, we will present the main statistical process control methods used to monitor surgical and interventional procedures with their specificity such as the possibility for case-mix adjustment. We will also present a new method specifically designed to monitor the learning curve Biau & Porcher (2010). Some historical papers will be presented and an example will be detailed.

2. Statistical process control methods

2.1 Monitoring a process from an in control to an out of control state

2.1.1 Control charts

Walter A. Shewart developed a control chart in the nineteen-twenties while working for Bell Telephone Laboratories Shewart (1931). A control chart is a graphical representation of the measure of some characteristics of a product, or of a sample of products, against the sequence of these measures. Usually a control chart has a central line which corresponds to the expected value of the characteristics when the process monitored is in control and one or two control limits which define in and out of control zones. Different control charts exist: variables control charts when we are interested in a continuous variable or attributes control charts when we are interested in a dichotomous variable; control charts without memory such as the Shewart chart or with memory such as the exponentially weighted moving average (EWMA) that averages the data in a way that gives less and less weight to data as they are further removed in time, etc Page (1954); Roberts (1959); Shewart (1931); Wald (1945). But for all, the objective is to display the variation over time of the process under scrutiny.

The choice of control limits will determine the capacity of the chart to detect a change in the process, and therefore the user will require fine adjustment of these limits before the start of monitoring. Typically, given a period of monitoring the user will have to set the risks he or she is willing to take not to detect a change in the performance of the process when it occurs or, on the contrary, to emit a false alarm when no significant change in the performance of the process has occurred. The upper and lower control limits, UCL and LCL have a general definition of the form:

$$\begin{cases} UCL = \mu_w + L\sigma_w \\ LCL = \mu_w - L\sigma_w \end{cases} \quad (1)$$

where μ_w and σ_w are the mean and standard deviation of the statistic used to monitor the process, and L the distance of the limit to the mean. A well known limit used in quality control is the 3 sigma limit where $L = 3$; in that case, when the process is in control and if the statistic used follows a normal distribution, 99.7% of the measures will be found in between the upper and lower control limits. Numerous control charts have been developed over time such as the cumulative summation (CUSUM) chart, the EWMA, etc.

2.1.2 the CUSUM test

2.1.2.1 Definition

The CUSUM test was developed in the nineteen-fifties by Page Page (1954) after Wald's work on sequential tests Wald (1945). It was originally developed to monitor the quality of manufactured products and later was found to be attractive to the medical context. Let us note X_1, X_2, X_3, \dots , the sequence of observations of the process under surveillance and assume no serial correlation. We denote by μ the mean of the process. These could be the times to complete appendectomies, the outcomes of neonatal arterial switch operations, or any other quantity that one would wish to monitor. The CUSUM sequentially tests after each observation $X_i (i > 0)$ the following hypothesis $H_0: \mu = \mu_0$, ie the process is in control, versus $H_1: \mu \neq \mu_0$, ie the process is out of control. The value μ_0 is often referred as the target. In its one sided formulation, where $H_0: \mu = \mu_0$ is tested against $H_1: \mu = \mu_1 (\mu_1 > \mu_0)$ the test is based on the statistic S_i computed after each observation X_i as:

$$S_i = \max(0, S_{i-1} + W_i), \quad S_0 = 0 \quad (2)$$

where the sample weight W_i depends on the observation X_i , μ_0 , and μ_1 . Moustakides Moustakides (1986) showed that optimal choices for W_i are proportional to the log-likelihood ratio. The test statistic S_i is compared to a predefined limit h . If S_i equals or exceeds h , the null hypothesis is rejected. In quality control wording the CUSUM test is said to emit an alarm indicating that the process is out of control. As long as S_i remains below h , the null hypothesis cannot be rejected and monitoring continues under the assumption that the process is in control.

2.1.2.2 Normally distributed data

The monitoring of normally distributed data is uncommon in the medical literature. Examples are the monitoring of the alignment of knee prosthetic replacements Nizard et al. (2004) and the laxity of anterior cruciate ligament reconstructions Biau et al. (2010). For normally distributed data of known standard deviation σ , because deviation may occur above or below the target value, two CUSUM tests are used in parallel to test H_0 against $H_1: \mu = \mu_1 = \mu_0 + g\sigma$ and $H_{-1}: \mu_{-1} = \mu_0 - g\sigma$ (with $g > 0$). The tests statistics S_i^+ and S_i^- used are defined as:

$$\begin{cases} S_t^+ = \max(0, S_{t-1}^+ + W_t^+) \\ S_t^- = \min(0, S_{t-1}^- + W_t^-) \end{cases} \quad (3)$$

with the following sample weights,

$$\begin{cases} W_t^+ = (X_t - \mu_0)/\sigma - g/2 \\ W_t^- = (X_t - \mu_0)/\sigma + g/2 \end{cases} \quad (4)$$

With such scores, S_t^+ and S_t^- only build up if X_t deviates from μ_0 by $g\sigma/2$ or more. The null hypothesis is rejected if $S_t^+ \geq h$ or $S_t^- \leq -h$. In practice σ is not known but as monitoring is prospective and parameters defined a priori, σ represents the expected standard deviation of the series and not the estimated standard deviation.

2.1.2.3 Binomial data

Medical data monitored are most often expressed in a success/failure dichotomy and therefore, the CUSUM test for binary data is the most frequently used Biau et al. (2007). Examples include monitoring the quality of implantation of hip prostheses Biau et al. (2009), or monitoring the 30-day mortality after complex cardiac surgery Steiner et al. (1999). In this case the mean of the process μ is equal to the probability of the event monitored (usually the failure), which will be noted p . The CUSUM tests after each procedure the following hypothesis $H_0 : p = p_0$ vs $H_1 : p = p_1$, where p_0 is the probability of failure when performance is adequate, and p_1 the probability of failure when performance is inadequate. Inadequate performance failure rates are usually 2 to 5 times higher than the target rates Bolsin & Colson (2000); de Oliveira Filho (2002); Kestin (1995). In this setting, the log-likelihood ratio score in equation 2 is obtained from the Bernoulli distribution as:

$$W_t = \log[p_1^{X_t}(1 - p_1)^{1-X_t}/p_0^{X_t}(1 - p_0)^{1-X_t}] \quad (5)$$

where X_t , the outcome, is equal to 0 for success and 1 for failure.

2.1.2.4 Time to event data

CUSUM test have been developed for monitoring time to event data such as the failure of a prosthesis, or that of kidney and liver transplants, over time Biswas & Kalbfleisch (2008); Hardoon et al. (2007); Sego et al. (2009). The reason for using time to event data is that compared to a CUSUM test based on the Bernoulli distribution, one does not have to wait the end of a period of time to assess failures and successes. Therefore, the outcome of each procedure contributes to the test statistic instantaneously. Depending on the situation, the test is based on the deviation of the process from a constant hazard or from one that varies over time such as that from a Weibull model Hardoon et al. (2007); Sego et al. (2009). The method may be adjusted and based on a discrete time Hardoon et al. (2007) or continuous time detection Biswas & Kalbfleisch (2008); Sego et al. (2009). However, in practice it should be noted that eventually the responsiveness of the CUSUM test developed will be limited by the technical possibilities of collecting, analysing, and reporting the data on a regular and frequent basis. For instance the CUSUM test based on the likelihood ratio statistic will have the following sample weight Hardoon et al. (2007):

$$W_j = O_j \log(\text{HR}) - (\text{HR} - 1)E_j \quad (6)$$

where O_j and E_j represent the number of failures observed and expected on interval j under the null hypothesis, and HR represents the hazard ratio of departure from target that one wants to detect. For instance, the number of failures expected on interval j based on the Weibull distribution where the hazard function is $h(t) = \lambda\gamma t^{\gamma-1}$ Collett (2005), with $\lambda > 0$, $\gamma > 0$, and $0 \leq t < \infty$, is:

$$E_j = \lambda_0 \sum_{i=1}^n \{(t_{2i} - t_{0i})^\gamma - (t_{1i} - t_{0i})^\gamma\} \quad (7)$$

where λ_0 is the scale parameter of the Weibull distribution under the null hypothesis, γ is the shape parameter under both the null and alternative hypothesis, t_{0i} is the time of realisation of the procedure on a time scale, t_{1i} is the maximum between t_{0i} and the lower boundary of interval j , and t_{2i} is the minimum between the time of failure of the procedure and the upper boundary of interval j .

2.1.2.5 Adjusted CUSUM

To avoid unduly penalizing a care provider or a center owing to disadvantageous case-mix selection, adjusted CUSUM tests have been proposed Sego et al. (2009); Steiner et al. (2000). In that case, each procedure is attributed a risk of failure based on the values of a set of covariates, and the statistic used is weighted on these covariates. Therefore, say a surgeon after some years in practice sees his or her practice shifts towards operating more and more complex cases. Although this surgeon has gained experience over the years, his or her results may tend to be pulled downwards and a standard monitoring method may detect inadequate performance when it is in fact the opposite. Adjusting the CUSUM on the difficulty of the cases which the surgeon operates on will avoid unnecessary alarms and audits. As an illustration, let us consider the monitoring of a binary (success or failure) outcome. Assume that for patient t the pre-operative surgical risk may be determined from a set of covariates, and is noted p_t . The null and alternative hypotheses are now defined in terms of an odds ratio (OR), since each patient has a different baseline risk. The CUSUM test repeatedly tests $H_0: OR = OR_0$ versus $H_1: OR = OR_1$, and one may use the following sample weights Steiner et al. (2000):

$$W_t = \begin{cases} \log \left[\frac{(1-p_t+OR_0 p_t)}{(1-p_t+OR_1 p_t)} \right] & \text{if } X_t = 0 \quad (\text{success}) \\ \log \left[\frac{(1-p_t+OR_0 p_t)OR_1}{(1-p_t+OR_1 p_t)OR_0} \right] & \text{if } X_t = 1 \quad (\text{failure}) \end{cases} \quad (8)$$

There are caveats in using adjusted CUSUM tests though. First, the set of covariates used for the adjustment must be unambiguous and consensual, and cannot be determined from the sample being monitored. Second, adjusted CUSUM tests explicitly infer that the results expected are different from one patient to another, and at some point, this goes against the philosophy of quality control which is to get rid of variability; indeed, as Douglas Montgomery puts it "*Quality is inversely proportional to variability*" Montgomery (2005). The variability seen should not necessarily be accounted for in the statistical method, and one may have to deal with it in the care of the patients.

2.1.2.6 Performance of the test

The CUSUM test has the particularity to display a holding barrier at 0 and to never accept the null hypothesis. Eventually this latter hypothesis is rejected regardless of the true performance of the process. Therefore, when the number of observations tends to infinity the CUSUM test has type I error rate and power of 100% in terms of traditional hypothesis tests. Performances

of CUSUM tests are thus expressed differently. Average run lengths (ARL) are commonly used to report the performance of a CUSUM test: the ARL_0 and ARL_1 are the mean number of observations before an alarm occurs if the process runs under the null and alternative hypotheses respectively. Ideally, one wants to choose a limit so as to yield an important ARL_0 and a low ARL_1 ; therefore when the process is in control, false alarms are rare, and when the performance shifts, the test is responsive. However, both ARLs vary in the same direction as the limit h is moved upwards or downwards and there is a necessary trade-off between the probability of false alarms under the null hypothesis and the speed at which alarms are emitted under the alternative. Concerns have been raised in the literature regarding the use of average run lengths. First, run lengths have a skewed distribution that renders the average an inadequate measure of the underlying performance of CUSUM tests Woodall (1983). Second, run lengths are sometimes misleading and lead to the choice of inadequate limits Biau et al. (2009) and physicians usually prefer dealing with type I and type II errors Frisen (1992). In that case the performances of CUSUM tests are expressed in terms of true and false discovery rates (TDR and FDR), namely the probability of an alarm to be emitted under the alternative and null hypotheses within a defined number of observations Marshall et al. (2004). We now prefer this latter approach.

In some medical applications no decision boundaries are used and the sample weight W_t is simply plotted against the sequence of observations; this is called a CUSUM graph. In that case performance is informally judged by looking at the slope of the line plotted and compared to the theoretical slope expected under H_0 .

2.1.3 Other methods commonly used

Numerous statistical methods have been developed to monitor the performance of a process and, consecutively, have been used in surgery and interventional procedures. The sequential probability ratio test (SPRT) was developed by Wald and is sometimes used in the medical domain Wald (1945). The SPRT is based on the following test statistic:

$$Y_t = Y_{t-1} + W_t, \quad Y_0 = 0, \quad t = 1, 2, 3, \dots \quad (9)$$

with

$$W_t = \log \left\{ \frac{l_{1t}}{l_{0t}} \right\}, \quad (10)$$

where l_{1t} et l_{0t} correspond to the likelihood of observation t under H_1 and H_0 , respectively. W_t is thus defined similarly as for the CUSUM tests. However, one major difference with the CUSUM test is that the SPRT has no holding barrier and presents two decision limits, one to accept the null hypothesis, and the other one to reject the alternative hypothesis. The ability of the SPRT to accept the null hypothesis, i.e. to reject the alternative, has led physicians to use it to monitor the learning curve Biau et al. (2007). However, there are caveats in using such a procedure. We have found no practical interest so far in detecting, at the same time, a shift towards inadequate or adequate performance. It is one or the other that is of interest. If the performance monitored is assumed to be in control, then we are interested in detecting when it becomes inadequate and there is no need in detecting adequate performance only to continue the monitoring. If the performance is assumed to be out of control, such as for a trainee for instance (see below), there is no need for a test that indicates that the performance is indeed inadequate and we are only interested in knowing when it reaches a certain level of adequacy. If we don't know the status of the process being monitored, then we should assume it is not adequate and use a LC-CUSUM test (see below) and once an in control status has been reached, a conventional CUSUM test can be used.

The Exponentially Weighted Moving Average (EWMA) was developed by Roberts in the late fifties and, later, developed further by other authors Hunter (1986); Roberts (1959). In this method, the older the observation the less weight it conveys. The test is based on the following statistic:

$$Z_t = \lambda X_t + (1 - \lambda)Z_{t-1}, \quad Z_0 = \mu_0, \quad 0 < \lambda \leq 1, \quad t = 1, 2, 3, \dots \quad (11)$$

With the following limits:

$$\begin{cases} h^+ = \mu_0 + L\sigma\sqrt{\frac{\lambda}{2-\lambda}[1 - (1-\lambda)^{2i}]} \\ h^- = \mu_0 - L\sigma\sqrt{\frac{\lambda}{2-\lambda}[1 - (1-\lambda)^{2i}]} \end{cases} \quad (12)$$

where L is the difference between the limits. The parameters L and λ , usually chosen as $0.05 \leq \lambda \leq 0.25$ and $L = 3$, allow to define the test performance. Overall, EWMA and CUSUM charts have been reported to have similar efficiency Lucas & Saccucci (1990), but choosing the appropriate exponential function remains however difficult in the absence of previous information. An example of the use of this method in medicine may be found in an intensive care unit Pilcher et al. (2010).

2.2 Monitoring a process from an out of control to an in control state: the LC-CUSUM test

Often physicians are interested in determining the time at which point we can assume a trainee has reached a certain level of performance so that supervision, or even training, may stop and resources be redirected to another trainee Biau et al. (2007). The LC-CUSUM test has been developed for that purpose Biau & Porcher (2010); Biau et al. (2008). The LC-CUSUM test was designed in a similar manner to the two one-sided tests procedure for equivalence clinical trials Chow & Shao (2002). This time, the null hypothesis H_0 is that the process is not in control, or in the present case that performance is not adequate. In terms of measurements, the null hypothesis corresponds to a process that is not centered on a target μ_0 , but deviates on average at least by δ . This can be rewritten as $H_0: |\mu - \mu_0| \geq \delta$. The alternative hypothesis is that the process does not deviate from μ_0 by more than δ , namely $H_1: |\mu - \mu_0| < \delta$. The null hypothesis is the union of the two one-sided hypotheses $H_{01}: \mu - \mu_0 \geq \delta$ and $H_{02}: \mu - \mu_0 \leq -\delta$, and H_0 has to be rejected when both H_{01} and H_{02} are rejected. In practice, simple hypotheses are set to parameter the test: $H_{01}: \mu = \mu_0 + \delta$ and $H_{02}: \mu = \mu_0 - \delta$ under the null hypotheses and $H_1: \mu = \mu_0$ under the alternative hypothesis. A CUSUM test statistic is then constructed for each hypothesis in a traditional way (see eq. 2). When the process can only deviate on one side, the corresponding hypothesis H_{01} or H_{02} only is tested. For instance, when monitoring binary data (success/failure) one is generally interested in testing only whether the failure rate of a trainee does not deviate by more than an acceptable amount above adequate performance. In that situation, the hypothesis H_{02} is ignored. This situation can be regarded as akin to non-inferiority trials. The purpose of the CUSUM test is to reject an inadequate performance level; the purpose of the LC-CUSUM test is to signal when performance is sufficiently far away from an inadequate performance level to consider the performance as adequate.

For normally distributed data, with $\delta = g\sigma$ ($g > 0$), the following sample weights are used:

$$\begin{cases} W_t^1 = (X_t - \mu_0)/\sigma - g/2 \\ W_t^2 = (X_t - \mu_0)/\sigma + g/2 \end{cases} \quad (13)$$

and for dichotomous data:

$$W_t^1 = \log \left[\frac{p_0^{X_t} (1 - p_0)^{1 - X_t}}{(p_0 + \delta)^{X_t} (1 - p_0 - \delta)^{1 - X_t}} \right] \quad (14)$$

Given the fact that one may start monitoring with an LC-CUSUM test and then, once adequate performance has been shown, continue monitoring with a CUSUM test, confusion may arise in the definition of the adequate and inadequate performance levels under the null and alternative hypotheses for the CUSUM and LC-CUSUM tests. Therefore, it is preferable to determine the adequate and inadequate performance levels and the acceptable deviation based on clinical information such as previous data, literature, and opinions of expert and then decide what levels will be chosen under the null and alternative hypothesis for both tests. For instance, in endoscopic retrograde cholangiopancreatographies one will decide that the adequate performance is 10% failure, that 25% failure is inadequate and that 5% is an acceptable deviation from the adequate performance level. Therefore, for the LC-CUSUM test, the performance considered under the null hypothesis is $H_0: \mu = \mu_0 + \delta = 0.15$ and under the alternative is $H_1: \mu = \mu_0 = 0.1$; and for the CUSUM test, the performance considered under the null hypothesis is $H_0: \mu = \mu_0 = 0.1$ and under the alternative is $H_1: \mu = \mu_1 = 0.25$.

Obviously the acceptable deviation from target, δ , considered for the LC-CUSUM may vary between 0 and $|\mu_1 - \mu_0|$. As proposed by Jones et al. (1996) for equivalence trials, choosing this acceptable deviation to be $\delta \leq (\mu_1 - \mu_0)/2$, will reasonably allow a buffer between what is deemed as acceptable for the LC-CUSUM test and the performance considered as unacceptable for the CUSUM test. Performances of LC-CUSUM tests may also be expressed in terms of true and false discovery rates (TDR and FDR).

3. Applications

3.1 Applications to surgical procedures

Marc R. de Leval, a cardiothoracic surgeon, was the first to report the monitoring of his own performance in the nineteen-nineties de Leval et al. (1994). In his landmark paper, he used a sequential probability ratio test to monitor the 30-day mortality rate after 104 neonatal arterial switch operations. During this period, he experienced a cluster of surgical failures with 7 deaths from the 53th to the 68th observation. The probability of observing 7 deaths or more in 16 patients when the expected mortality rate is 10% is very small (0.0005, by the binomial distribution), and this cluster was attributed to inadequate performance. He concluded that control charts were useful to detect unfavorable trends in performance and correct surgical technique if needed. He also introduced the concept of near-miss, a surrogate for a failed procedure, which was used in civil aviation to monitor the skills of pilots, and that could play a role in monitoring the performance of surgeons.

The Bristol Royal Infirmary case was another major determinant in the interest that quality control procedures would raise at the beginning of the 21st century. The Inquiry into the management of care of children receiving complex heart surgery at the Bristol Royal Infirmary (2001). A public inquiry was conducted between 1998 and 2001 on the management of the care of children receiving complex cardiac surgery at the Bristol Royal Infirmary between 1984 and 1995. The mortality rate at that center proved to be twice the rate in England and did not demonstrate the decrease seen over time in the rest of the country. Despite having enough information from the late 1980s onwards to raise significant questions about the mortality rate, there was no systematic mechanism for monitoring the clinical performance of care providers or of the center. Therefore, it took many years before clinical activity was stopped and an audit conducted. The report therefore recommended that *“for the future there must be effective*

systems within hospitals to ensure that clinical performance is monitored. There must also be a system of independent external surveillance to review patterns of performance over time and to identify good and failing performance". An application of quality control methods to the data collected during the inquiry showed that inadequate performance could have been detected as early as 1991 Spiegelhalter et al. (2003). Numerous applications have then been published around the world with various control charts, adjustment on patient risk factors, and use of near-misses Albert et al. (2004); Novick et al. (2001); Poloniecki et al. (2004); Rogers et al. (2005); Tsang et al. (2009). The use of quality control procedures has spread to other surgical disciplines. Although the enthusiasm is not as strong as that seen in cardiac surgery, there are initiatives using these methods and improving patient care in other areas. In general surgery these methods have been used to monitor the quality of, among other procedures, laparoscopic cholecystectomy Bartlett & Parry (2001), parotidectomies Sharp et al. (2003), ileal pouch-anal anastomosis Tekkis, Fazio, Lavery, Remzi, Senagore, Wu, Strong, Poloneicki, Hull & Church (2005), laparoscopic colorectal resections Tekkis, Senagore, Delaney & Fazio (2005), or thyroidectomy Connolly & Watters (2010). In orthopaedics, these methods have been used to monitor the quality of implantation of knee and hip replacement Biau et al. (2009); Nizard et al. (2004), the failure of hip implants Hardoon et al. (2006) and reconstruction of the anterior cruciate ligament Biau et al. (2010).

3.2 Applications to interventional procedures

Numerous applications of quality control methods have been reported in interventional disciplines. Parry Parry & Williams (1991) was one of the first to report the monitoring of his experience of 305 unassisted colonoscopies using a CUSUM chart (no limits were used in this report). Success was defined as completion of the procedure and a target performance of 10% failure rate was aimed at. Overall, 250 procedures (82%) were completed. The analysis showed that there was a steep improvement in performance during the first 100 procedures and that the target performance was reached around the 200th procedure; however, as noted earlier, the CUSUM chart does not allow a precise determination of when the performance can be considered as adequate as opposed to the LC-CUSUM test.

Anaesthesiologists who perform many critical and highly technical interventional procedures for their practice have also shown a particular interest in using quality control procedures to monitor their performance. Kestin Kestin (1995) reported in the mid-nineties the use of the SPRT to monitor residents performing obstetrics extradurals, spinal anaesthesia, central venous and arterial cannulations. The levels of performance varied from one procedure to another from 5% to 20% and 20% to 40% for the adequate and inadequate levels respectively. The same year Lagasse Lagasse et al. (1995) reported the use of a Shewart chart to monitor 13 clinical indicators of inadequate care, such as the unplanned admission to intensive care unit of a patient or the occurrence of a postural headaches following spinal or epidural analgesia. 13,389 procedures were monitored over a 12 months period, and 116 complications were reported. Only one indicator revealed inadequate performance. Afterwards, other authors have monitored the outcome of various interventional procedures such as labour epidurals Naik et al. (2003), nerve blocks Schuepfer & Jahr (2005), or orotracheal intubation in a mannequin Correa et al. (2009).

3.3 Applications to the learning curve

From the beginning, statistical process control methods have been used to monitor the performance of trainees because physicians were not satisfied with the usual methods available for evaluation. Traditionally, either a predetermined number of cases was prescribed or direct observation by the tutor was necessary. In the former case, the method was not felt

to be individualized to the trainee with some trainees having completed the required number of procedures but who were not proficient yet and others showing competence early and thus rendering supervision unnecessary Dagash et al. (2003). In case a tutor was used to evaluate the trainee, subjectivity was a problem Sloan et al. (1995). Therefore, most studies using statistical process control methods to monitor the performance of surgical or interventional procedures have been directed at monitoring the learning curve Biau et al. (2007). The idea with using statistical process control methods is that one would be able to determine when a trainee has reached proficiency in an objective and individual manner at the same time. Numerous methods have been used: some, such as Parry Parry & Williams (1991) for instance, have used a CUSUM chart and based on the slope of the curve they decide approximately when an adequate level of performance has been reached. This method provides a visual help but it is not accurate. Some, such as Kestin Kestin (1995), have used a sequential probability ratio test and use one of the boundaries to detect adequate performance whilst the other boundary detects inadequate performance. Although this test is attractive, it is inadequate to monitor the learning curve because of the inconsistency in the emission of alarms when processes with similar performance are monitored. It is for that reason that we have designed the LC-CUSUM test as described above. To date it has been used to monitor the competence of an endoscopist performing endoscopic retrograde pancreatography Biau et al. (2008), that of a gynecologist performing vitrification of embryo Dessolle et al. (2009) and embryo transfers Dessolle et al. (2010), and that of radiologist in the diagnosis of endometriomas by transvaginal ultrasound Bazot et al. (2010).

4. Worked example

4.1 Monitoring the ongoing performance of a surgical procedure

Say a surgeon wants to monitor the quality of implantation of total hip replacements. She knows that the longevity of the implant is associated with the technical precision with which she will perform the procedure. She defines ten criteria (see table ??) to assess the overall quality of the procedure. Some, such as the cup abduction angle, are a direct measure of the technical aspect of the procedure, whilst others, such as a revision during hospital stay, are surrogates for the whole process; however, she feels that all criteria are of importance. In the absence of previous information on case-mix she decides not to use an adjusted method and to weight each criterion identically. Therefore, the procedure will be considered as failed if any of the ten criteria is failed. Based on previous information Biau et al. (2009) and on consensus among other surgeons she decides that 18% failure represents an adequate level of performance (target performance) and that 36% failure rate would indeed indicate inadequate performance. She decides to use a CUSUM test and based on computer simulation she sets the decision limit $h = 5.6$, yielding a TDR of 99.8% and FDR of 3% for 200 observations. Figure 1 shows the CUSUM test for the monitoring period. The test is based on equation 5, with $p_0 = 0.18$ and $p_1 = 0.36$. Five meetings were held during that year (indicated by the arrows). Overall 106 of the 500 total hip replacements (21.2%) were considered as failed. The score increases with each failure and decreases otherwise. The holding barrier at 0 prevents the score from drifting away from the decision limit with the accumulation of successes; therefore, the test remains responsive to the occurrence of inadequate performance at all times. The CUSUM test emitted an alarm after 87 procedures. After an audit, she identifies the main reason for failures and corrects her surgical technique accordingly. Monitoring is restarted and no alarm is emitted afterwards.

4.2 Monitoring the learning curve of a trainee

Say a program director in obstetrics and gynecology desires to assess the competence of trainees wishing to learn the use of fetoscopic-directed laser therapy for the treatment of twin-twin transfusion syndrome. Given the syndrome is rare only few centers around the world perform this procedure and the resources required to train a gynecologist are limited. Also, the procedure is technically complex and the life of fetuses at risk. Therefore, he would like to make sure that each trainee is only allowed to perform the procedure with no supervision once he or she is competent and that, given the limited number of cases available for teaching, as soon as a trainee is competent the resources may be directed to another trainee. After a literature search and consensus among senior practitioners in the country they decide to set a nation-wide program that will allow fellows to be trained and during which their performance will be monitored with a LC-CUSUM test. An adequate performance level of 10% failure rate, inadequate performance of 25% and an acceptable deviation from adequate performance of 5% will be considered. Because the outcome of a poorly performed procedure threatens the life of the fetus, they decide to allow no more than 5% FDR. Given the limited number of cases they cannot provide more than 100 procedures per trainee during training. Therefore, they choose a limit $h = 1.25$ so that the TDR is 85% over these 100 procedures. Three fellows decide to enter the program. All procedures they perform are under supervision and at the end of each procedure a debrief is conducted. The procedure is considered as failed if none of the fetus survives at birth or if the trainee was not able to complete the procedure without any assistance. Therefore, in case the tutor considers the procedure is not performed adequately by the trainee, and that this puts the fetus life at risk, he or she takes on the case and completes the procedure in an adequate manner so that the level of care provided to patients remains adequate. The LC-CUSUM test for the three trainees is depicted figure 2. One trainee demonstrated competency at the 52nd procedure, another one at the 83rd procedure, and the last one did not demonstrate competency during the allotted number. The case of the last trainee should be discussed among seniors and with the trainee before he or she can resume monitoring. One of the reason being that statistically, if the trainee is given more cases with the same limit, the risk of a false detection increases, and eventually, regardless the true performance of the trainee, an alarm will always occur.

5. Conclusion and perspectives

To date, numerous reports using quality control methods have been published in the surgical and interventional literature. These reports provide good evidence that these methods are useful in ensuring that adequate performance level is maintained over time. Also, and more importantly, they provide some assurance that should inadequate performance occur, it will not go undetected for an undue period of time. Ensuring that the quality of care provided is adequate at all time is paramount from the patient, physician and health authorities perspectives. However, there are non statistical issues that deserve more attention before we see these methods unveil their full potential.

First, these methods are underused because of the lack of knowledge of most physicians that such methods exist. These methods were transferred only recently to the medical domain and are quite separated from the natural understanding of physicians. They are not taught in curriculums and it is most of the time by an incidental discovery that a surgeon or an endoscopist will encounter them. As quality improvement initiatives make their way into the medical curriculum hopefully the awareness of physicians will increase accordingly Boonyasai et al. (2007).

Second, there is a genuine fear that these methods, which can measure performance at an individual level, will be misused. Physicians may feel these methods will be used to rank and compare them and ultimately be used in a coercive way to impose changes they do not agree with. This is not what these methods should be used for. These methods are not meant for deterrence but for improvement. Quality control methods should be part of a quality improvement program and, as such, it is paramount that they remain in the hands of those who are under observation to improve the quality of care they deliver. They should be aimed at detecting under-performing units in order to improve them, not in order to blame them, and at all time they should be centered on the patient's need. It is only by stepping in and implementing these methods first that physicians will limit their use to what they want so that these methods are not imposed blindly by healthcare authorities or management. Also, the use of such methods to detect inadequate performance does not mean that only a small subset of units will benefit from quality improvement methodology; ideally the modification of the process under observation should benefit every care provider, or every center, not just the lower end of them.

Third, there is a lack of consensus among centers and practitioners regarding numerous aspects of monitoring which restrain the spread of these methods. Processes are not always clearly defined when, for instance, it may involve the technical act *per se* or include the perioperative period. For a similar process, the outcome measures may vary from one center to another or from one practitioner to another. The levels of performance may also be different. If we agree that discrepancies are expected with surgeons or interventionists willing to see different aspects of a process being monitored or different levels being targeted, there is an urge for those who spearhead the implementation of quality control methods to meet and discuss their fundamentals.

Last, although the statistical methods of quality control are constantly developed, the major difficulty physicians will encounter is the practical aspect to implement these methods. It is time and resource consuming to collect, analyse and report data. Ideally, these methods should be functioning in an almost automated manner, smoothly, in the background so as not to interfere with clinical practice. And it is only at the time of regular reports that physicians should come aware of them and, if necessary, change the process that was under observation. Only a close relation between information technology and clinical practice will allow that.

6. Figures and tables

Subset	Inadequate if	Method of measurement
Technique	1 fracture of the acetabulum requiring fixation or change in the technique	as per operative note
	2 fracture of the femur requiring fixation or change in the technique	as per operative note
	3 -1 cm > leg length difference > 2 cm	X-ray AP pelvis
Stem	4 -4° > stem inclination angle > +4°	X-ray AP pelvis
	5 stem cement mantle modified Barrack C or D class	X-ray AP pelvis
Cup	6 +30° > cup abduction angle > +55°	X-ray AP pelvis
	7 0° > cup version angle > +30°	X-ray AP pelvis and hip
	8 lucent line ≥ 1mm in one zone or more for cemented cups or a gap ≥ 3mm for uncemented cups	X-ray AP pelvis
Clinical	9 Dislocation or reoperation for any reason during hospital stay	as per hospital record
	10 Death regardless the reason during hospital stay	as per hospital record

Table 1. Performance criteria considered for the success of total hip replacement

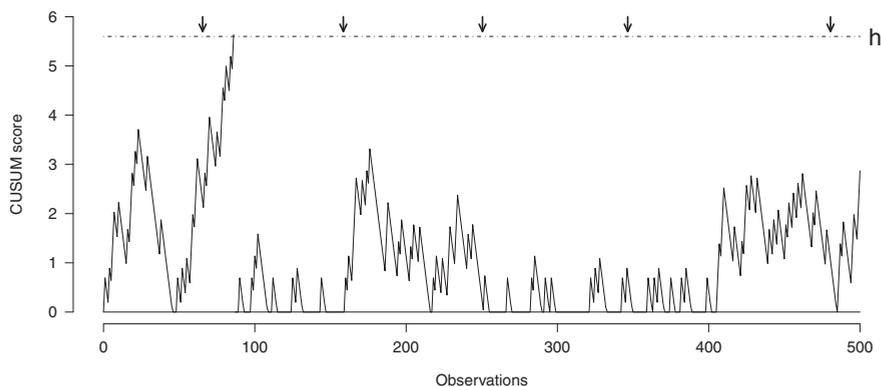


Fig. 1. CUSUM test for a surgeon performing 500 total hip replacements with $p_0 = 0.18$, $p_1 = 0.36$, and $h = 5.6$. Arrows indicate meetings where the results were presented. The CUSUM test emitted an alarm after the 87 procedures.

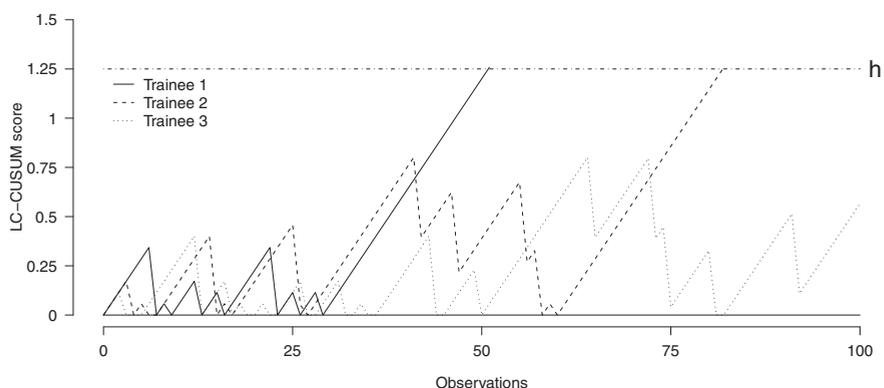


Fig. 2. LC-CUSUM test for 3 trainees performing fetoscopic-directed laser therapy for the treatment of twin-twin transfusion syndrome with $p_0 = 0.1$, $\delta = 0.05$, and $h = 1.25$. The LC-CUSUM test emitted an alarm after 52 and 83 procedures for trainee 1 and 2 respectively. No alarm was raised for trainee 3.

7. References

- Albert, A. A., Walter, J. A., Arnrich, B., Hassanein, W., Rosendahl, U. P., Bauer, S. & Ennker, J. (2004). On-line variable live-adjusted displays with internal and external risk-adjusted mortalities. A valuable method for benchmarking and early detection of unfavourable trends in cardiac surgery, *Eur J Cardiothorac Surg* 25: 312–319.
- Anderson, D. J. (1982). The quality control of medical gases, *Hosp Eng* 36: 18–19.
- Bartlett, A. & Parry, B. (2001). Cusum analysis of trends in operative selection and conversion rates for laparoscopic cholecystectomy, *ANZ J Surg* 71: 453–456.
- Batson, H. C., Brown, M. & Oberstein, M. (1951). An adaptation of quality control chart methods to bacterial vaccine potency testing, *J Bacteriol* 61: 407–419.
- Bazot, M., Darai, E., Biau, D. J., Ballester, M. & Dessolle, L. (2010). Learning curve of transvaginal ultrasound for the diagnosis of endometriomas assessed by the cumulative summation test (LC-CUSUM), *Fertil Steril*.
- Biau, D. J., Landreau, P., Gravelau, N. & Gravelau, N. (2010). Monitoring surgical performance: an application of industrial quality process control to anterior cruciate ligament reconstruction, *Knee Surg Sports Traumatol Arthrosc* 18: 1263–1268.
- Biau, D. J., Milet, A., Thevenin, F., Anract, P. & Porcher, R. (2009). Monitoring surgical performance: an application to total hip replacement, *J Eval Clin Pract* 15: 420–424.
- Biau, D. J. & Porcher, R. (2010). A method for monitoring a process from an out of control to an in control state: Application to the learning curve, *Stat Med* 29: 1900–1909.
- Biau, D. J., Resche-Rigon, M., Godiris-Petit, G., Nizard, R. S. & Porcher, R. (2007). Quality control of surgical and interventional procedures: a review of the CUSUM, *Qual Saf Health Care* 16: 203–207.

- Biau, D. J., Williams, S. M., Schlup, M. M., Nizard, R. S. & Porcher, R. (2008). Quantitative and individualized assessment of the learning curve using LC-CUSUM, *Br J Surg* 95: 925–929.
- Biswas, P. & Kalbfleisch, J. D. (2008). A risk-adjusted CUSUM in continuous time based on the Cox model, *Stat Med* 27: 3382–3406.
- Bolsin, S. & Colson, M. (2000). The use of the Cusum technique in the assessment of trainee competence in new procedures, *Int J Qual Health Care* 12: 433–438.
- Boonyasai, R. T., M., W. D., Chakraborti, C., S., F. L., Rubin, H. R. & Bass, E. B. (2007). Effectiveness of teaching quality improvement to clinicians: a systematic review., *JAMA* 298(9): 1023–1037.
- Brookeman, V. A. (1978). Computers and quality control in nuclear medicine, *Semin Nucl Med* 8: 113–124.
- Chamberlin, W. H., Lane, K. A., Kennedy, J. N., Bradley, S. D. & Rice, C. L. (1993). Monitoring intensive care unit performance using statistical quality control charts, *Int J Clin Monit Comput* 10: 155–161.
- Chow, S. C. & Shao, J. (2002). A note on statistical methods for assessing therapeutic equivalence, *Control Clin Trials* 23(5): 515–520.
- Collett, D. (2005). *Modelling Survival Data in Medical Research*, New York: Chapman & Hall/CRC.
- Connolly, T. M. & Watters, D. A. (2010). Monitoring performance in thyroidectomy: cumulative sum analysis of outcomes, *Thyroid* 20: 407–412.
- Correa, J. B., Dellazzana, J. E., Sturm, A., Leite, D. M., de Oliveira Filho, G. R. & Xavier, R. G. (2009). Using the Cusum curve to evaluate the training of orotracheal intubation with the Truview EVO2 laryngoscope, *Rev Bras Anesthesiol* 59: 321–331.
- Dagash, H., Chowdhury, M. & Pierro, A. (2003). When can I be proficient in laparoscopic surgery? A systematic review of the evidence, *J. Pediatr. Surg.* 38: 720–724.
- de Leval, M. R., Francois, K., Bull, C., Brawn, W. & Spiegelhalter, D. (1994). Analysis of a cluster of surgical failures. Application to a series of neonatal arterial switch operations, *J Thorac Cardiovasc Surg* 107: 914–923.
- de Oliveira Filho, G. R. (2002). The construction of learning curves for basic skills in anesthetic procedures: an application for the cumulative sum method, *Anesth. Analg.* 95: 411–416.
- Dessolle, L., Biau, D. J., de Larouziere, V., Ravel, C., Antoine, J. M., Darai, E. & Mandelbaum, J. (2009). Learning curve of vitrification assessed by cumulative summation test for learning curve (LC-CUSUM), *Fertil. Steril.* 92: 943–945.
- Dessolle, L., Freour, T., Barriere, P., Jean, M., Ravel, C., Darai, E. & Biau, D. J. (2010). How soon can I be proficient in embryo transfer? Lessons from the cumulative summation test for learning curve (LC-CUSUM), *Hum. Reprod.* 25: 380–386.
- Frisen, M. (1992). Evaluations of methods for statistical surveillance, *Stat Med* 11(11): 1489–1502.
- Hand, R., Piontek, F., Klemka-Walden, L. & Incauskis, D. (1994). Use of statistical control charts to assess outcomes of medical care: pneumonia in Medicare patients, *Am J Med Sci* 307: 329–334.
- Hardoon, S. L., Lewsey, J. D., Gregg, P. J., Reeves, B. C. & van der Meulen, J. H. (2006). Continuous monitoring of the performance of hip prostheses, *J Bone Joint Surg Br* 88: 716–720.
- Hardoon, S. L., Lewsey, J. D. & van der Meulen, J. H. (2007). Continuous monitoring of long-term outcomes with application to hip prostheses, *Stat Med* 26: 5081–5099.

- Hollinger, N. F. & Lansing, R. K. (1956). Standards, control charts, and reliability in the determination of serum protein, *Am J Med Technol* 22: 365–375.
- Hunter, S. J. (1986). The exponentially weighted moving average, *J Qual Technol* 18: 203–210.
- Jones, B., Jarvis, P., Lewis, J. A. & Ebbutt, A. F. (1996). Trials to assess equivalence: the importance of rigorous methods, *BMJ* 313: 36–39.
- Kestin, I. G. (1995). A statistical approach to measuring the competence of anaesthetic trainees at practical procedures, *British Journal of Anaesthesia* 75(6): 805–809.
- Kinsey, S. E., Giles, F. J. & Holton, J. (1989). Cusum plotting of temperature charts for assessing antimicrobial treatment in neutropenic patients, *BMJ* 299: 775–776.
- Lagasse, R. S., Steinberg, E. S., Katz, R. I. & Saubermann, A. J. (1995). Defining quality of perioperative care by statistical process control of adverse outcomes, *Anesthesiology* 82: 1181–1188.
- Lane, S., Weeks, A., Scholefield, H. & Alfirevic, Z. (2007). Monitoring obstetricians' performance with statistical process control charts, *BJOG* 114: 614–618.
- Loynd, H. J. (1962). Quality control in the drug industry, *J Indiana State Med Assoc* 55: 236–238.
- Lucas, J. & Saccucci, M. (1990). Exponentially weighted moving average control schemes: properties and enhancements, *Technometrics* 32(1): 1–12.
- Marshall, C., Best, N., Bottle, A. & Aylin, P. (2004). Statistical issues in the prospective monitoring of health outcomes across multiple units, *J Roy Stat Soc A* 167: 541–559.
- Montgomery, D. C. (2005). *Introduction to Statistical Quality Control, Fifth edition*, New York: John Wiley and Sons.
- Morgan, G. J., Thomas, S., Cavill, I. & Bentley, D. P. (1987). Detection of relapse in acute lymphoblastic leukaemia by cusum analysis of peripheral blood-count, *Lancet* 2: 1274–1275.
- Moustakides, G. V. (1986). Optimal stopping times for detecting changes in distributions, *Ann Statist* 14: 1379–1387.
- Naik, V. N., Devito, I. & Halpern, S. H. (2003). Cusum analysis is a useful tool to assess resident proficiency at insertion of labour epidurals, *Can J Anaesth* 50: 694–698.
- Nizard, R. S., Porcher, R., Ravaud, P., Vangaver, E., Hannouche, D., Bizot, P. & Sedel, L. (2004). Use of the Cusum technique for evaluation of a CT-based navigation system for total knee replacement, *Clin. Orthop. Relat. Res.* pp. 180–188.
- Novick, R. J., Fox, S. A., Stitt, L. W., Swinamer, S. A., Lehnhardt, K. R., Rayman, R. & Boyd, W. D. (2001). Cumulative sum failure analysis of a policy change from on-pump to off-pump coronary artery bypass grafting, *Ann. Thorac. Surg.* 72: S1016–1021.
- Novick, R. J. & Stitt, L. W. (1999). The learning curve of an academic cardiac surgeon: use of the CUSUM method, *J Card Surg* 14: 312–320.
- of Medicine, I. (2001). *Crossing the quality chasm: a new health care system for the 21st century*, National academy press, Washington, DC.
- Page, E. S. (1954). Continuous inspection schemes, *Biometrika* 41: 100–115.
- Parry, B. R. & Williams, S. M. (1991). Competency and the colonoscopist: a learning curve, *Aust N Z J Surg* 61: 419–422.
- Pilcher, D. V., Hoffman, T., Thomas, C., Ernest, D. & Hart, G. K. (2010). Risk-adjusted continuous outcome monitoring with an EWMA chart: could it have detected excess mortality among intensive care patients at Bundaberg Base Hospital?, *Crit Care Resusc* 12: 36–41.
- Poloniecki, J., Sismanidis, C., Bland, M. & Jones, P. (2004). Retrospective cohort study of false alarm rates associated with a series of heart operations: the case for hospital mortality monitoring groups, *BMJ* 328: 375.

- Pribr, H. C. & Shuster, M. (1967). Improving quality control in the hospital laboratory, *J Med Soc N J* 64: 213–217.
- Roberts, S. W. (1959). Control chart tests based on geometric moving averages, *Technometrics* 42: 97–102.
- Robinson, D. & Williamson, J. D. (1974). Letter: Cusum charts, *Lancet* 1: 317.
- Rogers, C. A., Ganesh, J. S., Banner, N. R. & Bonser, R. S. (2005). Cumulative risk adjusted monitoring of 30-day mortality after cardiothoracic transplantation: UK experience, *Eur J Cardiothorac Surg* 27: 1022–1029.
- Schuepfer, G. & Johr, M. (2005). Psoas compartment block (PCB) in children: Part II—generation of an institutional learning curve with a new technique, *Paediatr Anaesth* 15: 465–469.
- Sego, L. H., Reynolds, M. R. & Woodall, W. H. (2009). Risk-adjusted monitoring of survival times, *Stat Med* 28: 1386–1401.
- Sellick, J. A. (1993). The use of statistical process control charts in hospital epidemiology, *Infect Control Hosp Epidemiol* 14: 649–656.
- Sharp, J. F., Cozens, N. & Robinson, I. (2003). Assessment of surgical competence in parotid surgery using a CUSUM assessment tool, *Clin Otolaryngol Allied Sci* 28: 248–251.
- Shewhart, W. A. (1931). *Economic Control of Quality of Manufactured Products*, New York: Van Nostrand.
- Sloan, D. A., Donnelly, M. B., Schwartz, R. W. & Strodel, W. E. (1995). The Objective Structured Clinical Examination. The new gold standard for evaluating postgraduate clinical performance, *Ann. Surg.* 222: 735–742.
- Spiegelhalter, D., Grigg, O., Kinsman, R. & Treasure, T. (2003). Risk-adjusted sequential probability ratio tests: applications to Bristol, Shipman and adult cardiac surgery, *Int J Qual Health Care* 15: 7–13.
- Steiner, S. H., Cook, R. J. & Farewell, V. T. (1999). Monitoring paired binary surgical outcomes using cumulative sum charts, *Stat Med* 18: 69–86.
- Steiner, S. H., Cook, R. J., Farewell, V. T. & Treasure, T. (2000). Monitoring surgical performance using risk-adjusted cumulative sum charts, *Biostatistics* 1: 441–452.
- Tekkis, P. P., Fazio, V. W., Lavery, I. C., Remzi, F. H., Senagore, A. J., Wu, J. S., Strong, S. A., Poloneicki, J. D., Hull, T. L. & Church, J. M. (2005). Evaluation of the learning curve in ileal pouch-anal anastomosis surgery, *Ann. Surg.* 241: 262–268.
- Tekkis, P. P., Senagore, A. J., Delaney, C. P. & Fazio, V. W. (2005). Evaluation of the learning curve in laparoscopic colorectal surgery: comparison of right-sided and left-sided resections, *Ann. Surg.* 242: 83–91.
- The Inquiry into the management of care of children receiving complex heart surgery at the Bristol Royal Infirmary (2001). The Report of the Public Inquiry into children's heart surgery at the Bristol Royal Infirmary 1984–1995, *Technical report*.
URL: <http://www.bristol-inquiry.org.uk/>
- Tillett, H. E. & Spencer, I. L. (1982). Influenza surveillance in England and Wales using routine statistics. Development of 'cusum' graphs to compare 12 previous winters and to monitor the 1980/81 winter, *J Hyg (Lond)* 88: 83–94.
- Tsang, V. T., Brown, K. L., Synnergren, M. J., Kang, N., de Leval, M. R., Gallivan, S. & Utley, M. (2009). Monitoring risk-adjusted outcomes in congenital heart surgery: does the appropriateness of a risk model change with time?, *Ann. Thorac. Surg.* 87: 584–587.
- Van Rij, A. M., McDonald, J. R., Pettigrew, R. A., Putterill, M. J., Reddy, C. K. & Wright, J. J. (1995). Cusum as an aid to early assessment of the surgical trainee, *Br J Surg* 82: 1500–1503.
- Waid, M. E. & Hoffmann, R. G. (1955). The quality control of laboratory precision, *Am J Clin Pathol* 25: 585–594.

- Wald, A. (1945). Sequential Tests of Statistical Hypotheses, *Ann Math Stat* 16: 117–186.
- Walters, S. & Griffin, G. E. (1986). Resolution of fever in *Staphylococcus aureus* septicaemia—retrospective analysis by means of Cusum plot, *J Infect* 12: 57–63.
- Williams, S. M., Parry, B. R. & Schlup, M. M. (1992). Quality control: an application of the cusum, *BMJ* 304: 1359–1361.
- Wohl, H. (1977). The cusum plot: its utility in the analysis of clinical data, *N Engl J Med* 296: 1044–1045.
- Woodall, W. H. (1983). The distribution of the run length of one sided cusum procedures for continuous random variables, *Technometrics* 25: 295–301.



Applications and Experiences of Quality Control

Edited by Prof. Ognyan Ivanov

ISBN 978-953-307-236-4

Hard cover, 704 pages

Publisher InTech

Published online 26, April, 2011

Published in print edition April, 2011

The rich palette of topics set out in this book provides a sufficiently broad overview of the developments in the field of quality control. By providing detailed information on various aspects of quality control, this book can serve as a basis for starting interdisciplinary cooperation, which has increasingly become an integral part of scientific and applied research.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

David Biau, Lionel Dessolle and Raphaël Porcher (2011). Applications and Experiences of Quality Control to Surgical and Interventional Procedures, Applications and Experiences of Quality Control, Prof. Ognyan Ivanov (Ed.), ISBN: 978-953-307-236-4, InTech, Available from: <http://www.intechopen.com/books/applications-and-experiences-of-quality-control/applications-and-experiences-of-quality-control-to-surgical-and-interventional-procedures>

INTECH

open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](#), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.