



Monitoring surgical and medical outcomes: the Bernoulli cumulative SUM chart. A novel application to assess clinical interventions

G Leandro, N Rolando, G Gallus, K Rolles and A K Burroughs

Postgrad. Med. J. 2005;81;647-652
doi:10.1136/pgmj.2004.030619

Updated information and services can be found at:
<http://pmj.bmjournals.com/cgi/content/full/81/960/647>

These include:

References This article cites 28 articles, 12 of which can be accessed free at:
<http://pmj.bmjournals.com/cgi/content/full/81/960/647#BIBL>

Rapid responses You can respond to this article at:
<http://pmj.bmjournals.com/cgi/eletter-submit/81/960/647>

Email alerting service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article

Topic collections Articles on similar topics can be found in the following collections
[Transplantation](#) (254 articles)

Notes

To order reprints of this article go to:
<http://www.bmjournals.com/cgi/reprintform>

To subscribe to *Postgraduate Medical Journal* go to:
<http://www.bmjournals.com/subscriptions/>

ORIGINAL ARTICLE

Monitoring surgical and medical outcomes: the Bernoulli cumulative SUM chart. A novel application to assess clinical interventions

G Leandro, N Rolando, G Gallus, K Rolles, A K Burroughs

Postgrad Med J 2005;81:647–652. doi: 10.1136/pgmj.2004.030619

See end of article for authors' affiliations

Correspondence to: Professor A K Burroughs, Liver Transplantation and Hepatobiliary Medicine, Royal Free Hospital, Pond Street, Hampstead NW3 2QG, UK; andrew.burroughs@royalfree.nhs.uk

Submitted 14 November 2004
Accepted 27 March 2005

Background: Monitoring clinical interventions is an increasing requirement in current clinical practice. The standard CUSUM (cumulative sum) charts are used for this purpose. However, they are difficult to use in terms of identifying the point at which outcomes begin to be outside recommended limits.

Objective: To assess the Bernoulli CUSUM chart that permits not only a 100% inspection rate, but also the setting of average expected outcomes, maximum deviations from these, and false positive rates for the alarm signal to trigger.

Methods: As a working example this study used 674 consecutive first liver transplant recipients. The expected one year mortality set at 24% from the European Liver Transplant Registry average. A standard CUSUM was compared with Bernoulli CUSUM: the control value mortality was therefore 24%, maximum accepted mortality 30%, and average number of observations to signal was 500—that is, likelihood of false positive alarm was 1:500.

Results: The standard CUSUM showed an initial descending curve (nadir at patient 215) then progressively ascended indicating better performance. The Bernoulli CUSUM gave three alarm signals initially, with easily recognised breaks in the curve. There were no alarms signals after patient 143 indicating satisfactory performance within the criteria set.

Conclusions: The Bernoulli CUSUM is more easily interpretable graphically and is more suitable for monitoring outcomes than the standard CUSUM chart. It only requires three parameters to be set to monitor any clinical intervention: the average expected outcome, the maximum deviation from this, and the rate of false positive alarm triggers.

Systems to monitor clinical performance have gained increasing importance in medical practice. In the United Kingdom, since the Bristol Royal Infirmary Inquiry¹ there has been a new impetus to monitor outcomes of surgical and medical interventions. Assessing medical or surgical practice can take advantage of techniques studied for controlling a manufacturing process in industry to create a product. Maintaining an optimum is necessary.² Medical audit aims to monitor the variation around a predetermined optimum and ideally alarm signals should be set at the point when performance goes beyond acceptable standards. However, most audit procedures are retrospective analyses of outcome such as for mortality or important morbidity. This differs from manufacturing industry where real time assessment of the production process occurs.

Any process, whether in a biomedical or non-biomedical field can be subject to inherent deviations from the optimum or from designated limits. These deviations may lead to defective end products or in the medical field defective patient care. Monitoring, which is a process able to identify deviation, and then being able to act if this deviation exceeds certain limits, plays an important part in avoiding adverse consequences and maintaining optimal performance. In production processes the techniques most often used to monitor and control are referred to as statistical process control (SPC), which today is fully accepted as a key element in modern manufacturing industry.

In biomedical science, statistical evaluation usually considers only the first or the last point of a dataset. These statistical tests do not take into account what has happened in the past—that is, data between the first and the last point of a dataset are not considered. In contrast the SPC permits

use of information at more than one time point—that is, sequential testing. SPC has been used in biological applications for surveillance of rare events, but only rarely has it been used in biomedicine, an example being the monitoring of congenital defects by one of us (GG).³ This is mainly because it has proved difficult to bring together the underlying theories with appropriate procedures of calculation, as these are somewhat cumbersome and not easily applied to clinical processes. Among the sequential tests, the cumulative sum (CUSUM) test has had the widest use, in both surveillance and quality control, and in evaluating sequential measures or changes over time.⁴ Williams *et al*⁵ used the simple CUSUM to illustrate its use in monitoring success with colonoscopy and endoscopic retrograde cholangiopancreatography, and its application in monitoring individual clinicians. The method assumes a fixed and defined limit of probability of success of the endoscopic procedure. The authors felt that this simple CUSUM technique should be used as a means of monitoring performance, both for training in medical or surgical techniques and in monitoring to ensure the maintenance of standards.

Lovegrove *et al*⁶ evaluated the probability of death after surgical cardiothoracic procedures, which varied for every patient added cumulatively to a series, according to various risk factors such as comorbidity and surgical expertise.

Both these studies have cumulative series, detailing success or failure. However, an evaluation of the process over time is in practice made by a visual examination of the graphical representation. There are boundary lines traced on

Abbreviations: CUSUM, cumulative sum; SPRT, sequential probability ratio test; ELTR, European Liver Transplant Registry

the CUSUM graph, which are placed to define when the slope changes beyond acceptable limits, but the methodology used is cumbersome and is not sufficiently precise, thus giving the probability of too many false alarms.

What is missing is a formal and robust statistical test to assess medical or surgical performance over time and to have a monitoring system that gives very few false alarms. While further refinement has led to evaluating risk adjusted outcomes, particularly for mortality⁷⁻⁹ these as yet have not entered clinical practice. This is partly because not all datasets contain the relevant risk factors, or datasets are felt to be unreliable in terms of recording risk factors¹⁰ and partly because a consensus needs to be reached on the validity of the risk factors for example in cardiothoracic surgery¹¹⁻¹⁴ where there is a debate. Thus analysis of unadjusted outcome still has a role in clinical medicine and is the starting point of monitoring outcomes in clinical fields.

Monitoring methodology

Recently monitoring protocols of biomedical processes have benefited from developments in computation of statistical procedures, which are specifically designed to evaluate alarm signals.¹⁵⁻¹⁶ These result in a highly accurate and comparatively simple computation, far more precise than hitherto. When monitoring a process, the field of interest is the proportion of defective items (defective fraction),¹⁷ or in a medical scenario the proportion of patients who undergo suboptimal care. Until recently in industrial processes, taking samples of size "n" at regular intervals, and plotting the values of the proportion of defective items on a control chart, such as a Shewart p chart, is no longer used, as it has been shown to have a poor performance.¹⁸

A better approach is to use sequential test theory that uses information from the past dataset at each sample point, such as the sequential probability ratio test (SPRT). This is applicable when the sample size of the selected items of n, varies at each regular sampling point during a process. However, this is not useful in the medical context in which each patient should be evaluated. In medicine the best approach of the use of sequential test theory is provided by the CUSUM (cumulative sum) control charts.

When information from every item in the whole dataset is available, such as mortality data in a consecutive series of patients, the Bernoulli CUSUM chart can be used. Thus every patient can be accounted for, as there is a 100% inspection rate that is particularly suited for the evaluation of processes that have a very low defective rate (low p) as is generally the case in medicine—that is, low morbidity and mortality rates.

This evaluation of all data points is akin to the evaluation of every patient who undergoes a surgical or medical intervention in any particular cohort.

The aim of this study was to apply sound statistical methodology in the form of the Bernoulli CUSUM control chart, as a type of statistical process control, to a clinical process for which outcomes are known—that is, the number with suboptimal outcome—and for which the limits for an unacceptable deviation from the baseline can be defined, so that alarm signals regarding unacceptable deviation beyond the specified limits can be detected.

METHODS

The details of the theory, terminology, and calculations are shown in the appendix. As an example of a clinical intervention, we chose a consecutive cohort of adult patients undergoing their first liver transplant at a single centre, with death at one year as the primary outcome measure. The reasons for this choice are the following: (1) high intrinsic mortality rate and thus a high event rate, compared with other surgical procedures; (2) a comparatively standardised procedure within a single centre with less surgical variability; (3) the existence of a large international published registry with validated data, the European Liver Transplant Registry (ELTR),¹⁹⁻²⁰ which has been published with one year mortality figures for patients with and without associated risk factors for mortality. The total population was 19 370 adults; (4) similarities between the reference registry cohort and the single centre cohort at the Royal Free Hospital. The database covered a similar time period and there were similar proportions of the following characteristics respectively: orthotopic liver transplants (99.7% v 99.9%), fulminant liver failure (11% v 10%), cancer (10% v 11%), re-transplantation (12% v 10.5%) the last three being risk factors for one year mortality.

RESULTS

The consecutive series of first liver transplants at the Royal Free Hospital between 1 October 1988 (the start of the programme) and 31 August 2002 were 674. The median age was 49 years and 60% were male. There were 71 re-transplants (eight of these had a third transplant) and 12 had a combined liver and kidney transplant.

The expected one year mortality was considered to be 24% for the whole population as this was the actual mortality for adults in the ELTR in the report published in 2000.¹⁹ Although our cohort extended to 2002, and mortality after

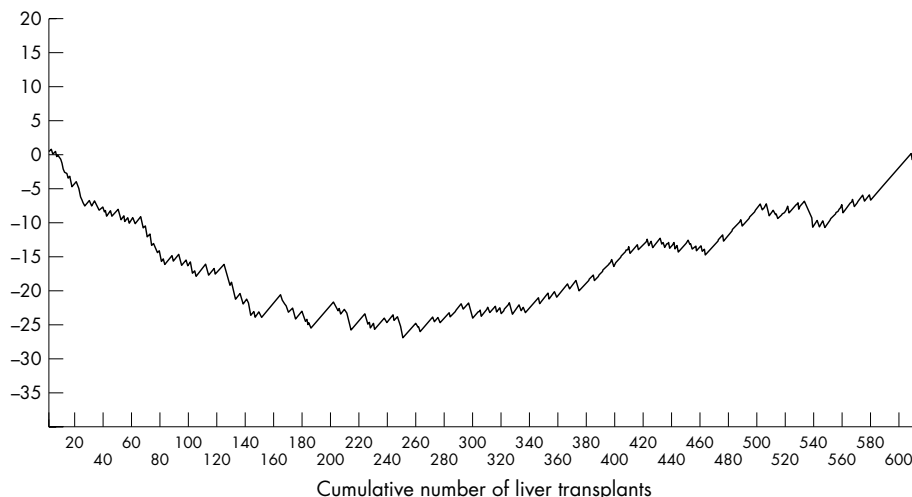


Figure 1 Descriptive CUSUM chart for mortality at one year in a consecutive series of first liver transplants in adults at the Royal Free Hospital from the start of the programme in 1988 to 2002. 0.24 was added for a success and 0.76 was deducted for a failure. These represent 24% (death) and 76% (survival) respectively at one year averages derived from the European Liver Transplant Registry data.¹⁹

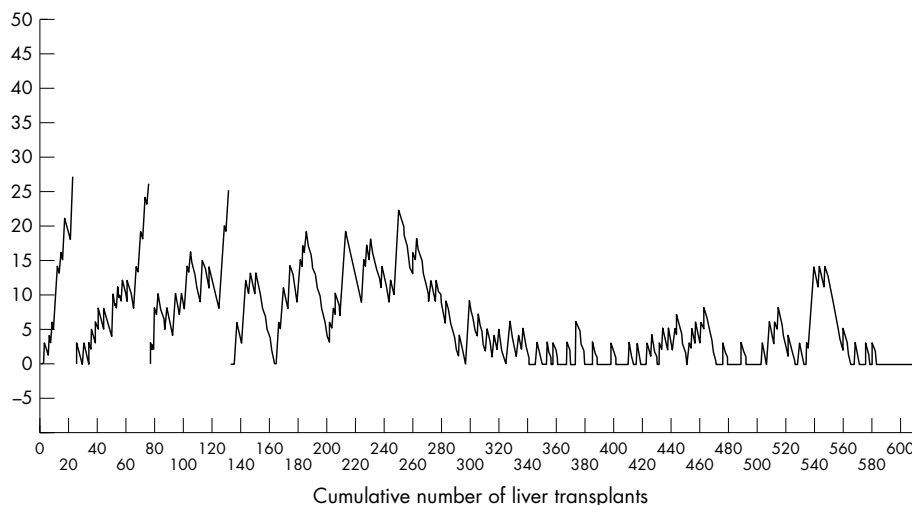


Figure 2 Bernoulli CUSUM chart for mortality at one year in a consecutive series of first liver transplants in adults at the Royal Free Hospital from the start of the programme in 1988 to 2002. Parameters were derived from the European Liver Transplant Registry data.¹⁹ Average probability of death after transplantation $p_0 = 0.24$; the failure rate—that is, mortality rate considered unacceptable in clinical practice $p_1 = 0.30$; the probability of a false alarm signal during monitoring $\text{ANOS}(p_0) = 1:500$. Three alarm signals are shown by breaks in the line graph.

liver transplantation is improving,¹¹ the average 24% mortality is used to give a working example with a real dataset. Therefore, following the suggestion of Williams *et al*⁵ for every patient alive at one year the probability of death ($p_0 = 0.24$) was added to the CUSUM, and for every death within one year, the probability of being alive ($1 - p_0 = 0.76$) was subtracted from the CUSUM. Figure 1 shows the cumulative course of the Royal Free cohort according to the above calculations. A declining course can be seen in the first part and an ascending course in the second.

The second chart (fig 2) displays the same dataset using the Bernoulli CUSUM control chart. The parameters used are: probability of death = $p_0 = 0.24$ (in control value of p). The maximum value of p , that is the maximum permitted mortality at one year, beyond which the process is defined as being outside of the specified parameters was considered to be 30%—that is, $p_1 = 0.30$ (out of control value of p). The average number of observations to signal, $\text{ANOS}(p_0)$ was set to the value 500. The $\text{ANOS}(p_0)$ is the number of observations that would contain one false alarm—that is, a false positive—and is explicitly stated a priori. The alarm signal means that the process is out of control, in other words that there is a factor or several factors that are leading to an unacceptable deviation from the specified limits. In manufacturing industry, the production process would be stopped

and machines would be re-gauged. In medical audit this would be the time to evaluate all potential factors that could lead to an out of control result and to proceed accordingly.

In figure 2 three alarm signals (given when the sum reaches the value 24) are represented by interruptions in the line in the first 138 transplants. In our series the out of control process was probably due to several factors: case mix, surgical technique, case volume, and learning curve in integrating the multidisciplinary teams involved in this complex surgery and postoperative care.²¹ Clearly in the early years of the programme the expected average mortality at one year would have been lower than the average 24%, and in the later years higher. However, what the earlier or later expected average mortalities are, cannot be derived for the earlier period from the ELTR database¹⁹ nor estimated for the later period. Nevertheless the fact that alarm signals are only seen in the early period and then results progressively improve to above the average suggests improving performance in common with many transplant centres.¹⁹

When assessing the patients from the Royal Free cohort without risk factors (ignoring centre size effect based on ELTR data that give a risk ratio of 1.46 compared with baseline mortality), which were 42, and evaluating one year mortality with $p_0 = 0.15$ derived from the ELTR dataset, and considering p_1 as equal to 0.20 and $\text{ANOS}(p_0) = 500$, then

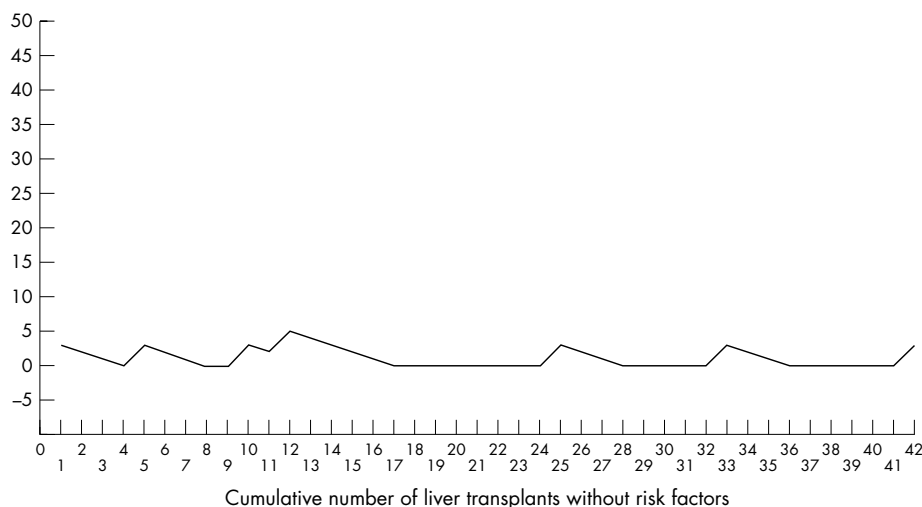


Figure 3 Bernoulli CUSUM chart for mortality at one year in adult patients receiving a first liver transplant without risk factors, parameters derived from the European Liver Transplant Registry data¹⁹: average probability of death after transplantation $p_0 = 0.15$; the failure rate—that is, mortality rate considered unacceptable in clinical practice $p_1 = 0.20$; the probability of a false alarm signal during monitoring $\text{ANOS}(p_0) = 1:500$. No alarm signals have occurred.

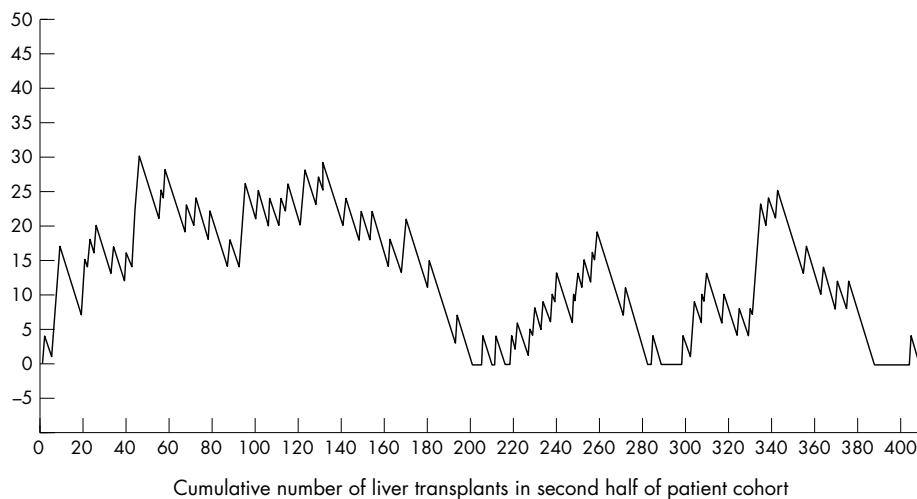


Figure 4 Bernoulli CUSUM chart for mortality at one year of the second half of the first liver transplant cohort at the Royal Free Hospital parameters derived from the European Liver Transplant Registry data¹⁵: average probability of death after transplantation $p_0=0.17$; the failure rate—that is, mortality rate considered unacceptable in clinical practice $p_1=0.20$; the probability of a false alarm signal during monitoring ANOS(p_0) at 1:500. No alarm signals have occurred.

there are no alarm signals at any time (fig 3). This suggests that the initial alarm signals in the whole cohort were less likely to be attributable to surgical techniques or initial perioperative care as part of the learning curve, as one might expect a similar trend in patients without known risk factors. Recipient selection or use of marginal donors may have been reasons for the poorer results and thus the alarm signals. In addition as the average probability of death would have been higher in the early years—1988–1992—it would seem that increasingly improved results with time in our centre and in others (19) may be attributable to better management of worse risk patients with more risk during and after liver transplantation—an effect of increasing experience.

The difference between figure 1, using the CUSUM graphical representation,⁵ which gives only a visual assessment, and the Bernoulli CUSUM approach of figure 2, is that the graphical representation in figure 1 suggests the nadir was reached at number 215 and then there was a continuous improvement, whereas in figure 2 it can be seen that no alarm signals occurred after transplant 148—that is, the process had been under control from an earlier time point.

Moreover in the phase of the descending slope using the simple CUSUM representation, there are no fixed points where one can decide to audit the problem to establish the causes of the deviation from the given parameters.

DISCUSSION

Medical and surgical interventions need a monitoring system, which can signal when the process is getting out of control, as is the case in non-medical fields. How to devise such monitoring is a matter of current and topical interest, particularly in the UK, given the results of the Bristol Royal Infirmary Inquiry,¹ which has resulted in applying graphical CUSUM control charts⁵ and derived risk adjusted models (variable life adjusted display) in cardiothoracic procedures,⁶ although there are difficulties in setting thresholds.⁷ There also have been questions regarding the reliability of the risk scores^{11–14} as well as comparisons of crude and risk adjusted mortality.²² Currently crude mortality rates are still being evaluated. The quality and credibility of the UK cardiac surgery database²³ has been assessed as needing improvement¹⁰ making accurate risk adjusted mortality more difficult to assess. For other surgical procedures mortality control charts may not be the best way to assess or compare performance.^{24–25} Thus hospital mortality charts for common surgical procedures (and medical conditions) have been considered crude and at worst misleading reflections of quality of care.²⁶ Despite these problems cardiac surgeons in

the UK will be assessed according to bypass surgery success²⁷ a decision that is a by product of the momentum generated to demonstrate quality control “in action”. Thus the improvement in statistical computation that has occurred recently as well as improved graphical display are useful, even if solely applied to evaluations of unadjusted mortality.

However, the publications using descriptive CUSUM charts, and derived variable life adjusted display (VLAD) charts have the disadvantage of not displaying the precise point of the alarm signal, and often an assessment can only be made by eye-balling the shape of the curves. Only recently has statistical methodology been applied to monitor medical or surgical outcomes when risk adjustment cannot be manipulated adequately^{28–29} or to evaluate optimal performance with pre-set thresholds.³⁰

In our study we have evaluated a new monitoring system, the Bernoulli CUSUM chart, which has a 100% inspection rate, using as an example of a clinical intervention, a surgical procedure—liver transplantation. We considered mortality at one year as an outcome in a consecutive series of adult patients from the start of the programme in 1988 to 2002.

This monitoring system requires the setting of three values: the average probability of the failure of the procedure (P_0), the percentage of the failure rate considered unacceptable in current practice (p_1), and the probability of having a false alarm in the monitoring system (ANOS(p_0)). With these parameters in place, the statistical procedure determines the number to consecutively accumulate for every success, and for every failure, and the threshold for the alarm signal. If the latter is reached, the system resets itself and starts again from baseline, displaying this on a control chart. Depending on the importance of the outcome measured, the action taken when an alarm is reached will vary. This needs to be planned in advance, but at the very least, in clinical contexts would entail an internal audit. There may be other such contexts where temporary suspension of a procedure is indicated pending review by an external panel of experts, which may act as a hospital monitoring group for mortality or other outcomes.⁹ Clearly if a clinical monitoring system is adopted plans must be in place to follow a particular protocol if an alarm signal is generated.

We set our thresholds for average mortality (P_0) from the large ELTR database¹⁵ of 19370 first liver adult transplants, and set the maximum acceptable one year mortality at 6% higher (30%), and the false positive alarm (ANOS(p_0) at 1:500. Using the same average mortality rate, a comparison of the standard CUSUM chart (fig 1) with the Bernoulli CUSUM chart (fig 2), shows firstly a more easily interpretable graphic

display, with the interrupted lines in the Bernoulli CUSUM chart, representing that the alarm signals have been reached. The first alarm signal occurs before the nadir in the standard CUSUM chart. In addition the chart shows an earlier return to a system under control.

Secondly, the statistical methodology behind the Bernoulli CUSUM control chart is validated and accepted in fields outside medicine and indeed is the most applicable to interventions in medicine as it uses a 100% inspection rate. We believe that this monitoring system, as it permits simple setting of thresholds for performance and error in the alarm signals, overcomes many of the disadvantages of current systems²⁵ and is a step forward in assessing outcome of surgical and medical procedures.

In addition the use of this methodology has several potential applications over and above monitoring a procedure to establish if the outcomes are maintained within set thresholds based on average performance published in the literature (as in our example). One application could be to delineate periods of worse performance (even before alarm signals are reached) permitting a detailed evaluation of a particular series of patients, to try to understand factors that may have led to this. A second application could be that if the monitoring shows a prolonged period of stability or improvement, the thresholds (p_0 and p_1) could be modified to evaluate what performance level has been reached for the particular procedure under scrutiny. In our example, evaluating the more recent, second half of the cohort (fig 4), the alarm signals are not reached until thresholds are set at an average one year mortality of 16% (p_0) and maximum acceptable one year mortality of 20% (p_1). Even with this example, periods of worse performance could be examined before alarm signals are reached.

The Bernoulli CUSUM chart as we have shown applied to a surgical procedure, is simple to use, and applicable to the numerous and various outcomes surrounding any intervention in medicine or surgery. It would be very useful to evaluate a prospectively monitored cohort of patients, using this methodology. In addition, outcomes other than survival should be assessed, as this is not necessarily the best outcome measure in many clinical interventions.^{24 25} As the Bernoulli CUSUM has a robust statistical framework, it should contribute to a more sophisticated surgical, medical, or regulatory audit, as it can easily be modified according to defined thresholds. In the future computation of risk adjusted outcomes using this methodology will further improve its utility in every day clinical practice as has been evaluated with the standard CUSUM chart,¹⁴ resetting sequential probability ratio charts,⁸ and cumulative risk adjusted mortality charts.⁹

Authors' affiliations

G Leandro, Ospedale Gastroenterologico, Castellana Grotte, BA, Italy
G Leandro, N Rolando, K Rolles, A K Burroughs, Royal Free Hospital, Liver Transplantation and Hepatobiliary Medicine, London, UK
G Gallus, Institute of Biometrics and Medical Statistics, University and Istituto Nazionale Tumori di Milan, Italy

Funding: none.

Conflicts of interest: none.

APPENDIX

If p is the probability of a defective, then 0 is the best value for p but, in practice, we use the currently acceptable value, in the absence of any specific cause of process variation, as the reference value of the process, p_0 .

The Bernoulli CUSUM chart is designed to detect an increase in p and for this the Bernoulli CUSUM control statistic is

$$B_k = \max(0, B_{k-1}) + (X_k - \gamma_B) \text{ for } k = 1, 2, \dots$$

Where $\gamma_B > 0$ is called *the reference value*.

The CUSUM chart will signal an increase of p if $B_k > h_B$, where h_B is the *control limit*.

To calculate the value of γ_B it is necessary to specify a value (p_1) which represent the out-of-control value of p . p_1 is also the value of p which should be detected as soon as possible.

Given the in-control value p_0 and the out-of-control value p_1 we can obtain the constants r_1 and r_2 by

$$r_1 = -\text{Log}\left(\frac{1-p_1}{1-p_0}\right) \quad r_2 = \text{Log}\left(\frac{p_1(1-p_0)}{p_0(1-p_1)}\right)$$

and then the value of γ_B as

$$\gamma_B = \frac{r_1}{r_2} = \frac{1}{m}$$

where m is an integer.

The average number of observations to signal (ANOS) is obtained by the so called *corrected diffusion* approximation and is:

$$\text{ANOS}(p_0) \approx \frac{\exp[h_B^* r_2] - h_B^* r_2 - 1}{|r_2 p_0 - r_1|}$$

where h_B^* is an adjusted value of h_B and was obtained by the formula

$$h_B^* = h_B + \varepsilon_{(p_0)} \sqrt{p_0 q_0}$$

where $q_0 = 1 - p_0$ and

$$\varepsilon_{(p)} \approx 0.41 - 0.0842(\log(p)) - 0.0391(\log(p))^3 - 0.00376(\log(p))^4 - 0.000008(\log(p))^7$$

if $0.01 \leq p \leq 0.5$ or

$$\varepsilon_{(p)} \approx \frac{1}{3} \left(\sqrt{\frac{1-p}{p}} - \sqrt{\frac{p}{1-p}} \right) \text{ if } 0 < p < 0.01$$

The ANOS (p_0) is the number of observations that would contain one false alarm and is explicitly stated a priori—that is, a false positive.

To know how fast a shift from p_0 to p_1 will be detected, the CD approximation to the ANOS when $p = p_1$ is

$$\text{ANOS}(p_1) \approx \frac{\exp[h_B^* r_2] + h_B^* r_2 - 1}{|r_2 p_1 - r_1|}$$

REFERENCES

- 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.
- Sergeant P**, Meyns B. La critique est aisée mais l'art est difficile. *Lancet* 1997;**350**:1114-15.
- Gallus G**, Mandelli C, Marchi M, et al. On surveillance methods for congenital malformations. *Stat Med* 1986;**5**:565-71.
- Altman DG**, Royston P. The hidden effect of time. *Stat Med* 1988;**7**:629-37.
- Williams SM**, Parry BR, Schlup MMT. Quality control: an application of the cusum. *BMJ* 1992;**304**:1359-61.

- 6 **Lovegrove J**, Valencia O, Treasure T, *et al*. Monitoring the results of cardiac surgery by variable life-adjusted display. *Lancet* 1997;**350**:1128–30.
- 7 **Steiner SH**, Cook RJ, Farewell VT, *et al*. Monitoring surgical performance using risk-adjusted cumulative sum charts. *Biostatistics* 2000;**1**:441–52.
- 8 **Grigg OA**, Farewell YT, Spiegelhalter DJ. Use of risk adjusted CUSUM and RSPRT charts for monitoring in medical contexts. *Stat Methods Med Res* 2003;**12**:147–70.
- 9 **Poloniecki J**, Sismanidis C, Bland M, *et al*. Retrospective cohort of false alarm rates associated with a series of heart operations: the case of hospital mortality groups. *BMJ* 2004;**328**:375–9.
- 10 **Fine LG**, Keogh BE, Cretin S, *et al*. How to evaluate and improve the quality and credibility of an outcomes database: validation and feedback-study on the UK cardiac surgery experience. *BMJ* 2003;**326**:25–8.
- 11 **Wynn-Jones K**, Jackson M, Grotte G, *et al*. Limitations in the Parsonnet score for measuring risk-stratified mortality in the north west of England. *Heart* 2000;**84**:71–8.
- 12 **Roques F**, Michel P, Goldstone AR, *et al*. The logistic EUROSCORE. *Eur Heart J* 2003;**24**:881–2.
- 13 **Sergeant P**, Blackstone E, Meyns B. Can the outcome of coronary bypass grafting be predicted reliably? *Eur J Cardiothor Surg* 1997;**11**:2–9.
- 14 **Nashef SAM**, Roques F, Michel P, *et al*. The EuroSCORE study group. European system for cardiac operative risk evaluation (EUROSCORE). *Eur J Cardiothorac Surg* 1999;**16**:9–13.
- 15 **Stoumbos ZG**, Reynolds MR. Control charts applying a general sequential test at each sampling point. *Sequent Anal* 1996;**15**:159–83.
- 16 **Reynolds MR**, Stoumbos ZG. A cusum chart for monitoring a proportion when inspecting continuously. *Journal of Quality Technology* 1999;**31**:87–108.
- 17 **Woodall WH**. Control charts based on attribute data: bibliography and review. *Journal of Quality Technology* 1997;**29**:172–83.
- 18 **Ryan TP**, Schwertman. NC Optimal limits for attributes control charts. *Journal of Quality Technology* 1997;**29**:86–98.
- 19 **Adam R**, Cailliez V, Majno P, *et al*. Normalised intrinsic mortality risk in liver transplantation: European Liver Transplant Registry study. *Lancet* 2000;**356**:621–7.
- 20 **Karam V**, Gunson B, Roggen F, *et al*. Quality control of the European Liver Transplant Registry: results of audit visits to the contributing centres. *Transplantation* 2003;**75**:2167–73.
- 21 **Belle SH**, Detre KM, Beringer KC. The relationship between outcome of liver transplantation and experience in new centres. *Liver Transpl Surg* 1995;**6**:347–53.
- 22 **Bridgewater B**, Grayson AD, Jackson M, *et al*. Surgeon specific mortality in adult cardiac surgery: comparison between crude and risk stratified data. *BMJ* 2003;**327**:13–17.
- 23 **Keogh B**, Kinsman R. *National adult cardiac surgical database report 2000–2001*. London: Society of Cardiothoracic Surgeons of Great Britain and Ireland, 2002.
- 24 **Tekkis PP**, McCulloch P, Steger A, *et al*. Mortality control charts for comparing performance of surgical unit: validation study using hospital mortality data. *BMJ* 2003;**326**:786–8.
- 25 **Treasure T**, Utley M, Bailey A. Assessment of whether in hospital mortality for lobectomy is a useful standard for the quality of lung cancer surgery. *BMJ* 2003;**327**:73–5.
- 26 **Jacobson B**, Mindell J, McKee M. Hospital mortality league tables. *BMJ* 2003;**326**:777–8.
- 27 **Dyer O**. Heart surgeons are to be rated according to bypass surgery success. *BMJ* 2003;**326**:1053.
- 28 **Aylin P**, Best N, Bottle A, *et al*. Following Shipman: a pilot system for monitoring mortality rates in primary care. *Lancet* 2003;**362**:485–91.
- 29 **Thompson W**, Hutwagner L, Kolczak M. Monitoring mortality events associated with individual physicians and practices. *Lancet* 2003;**362**:417–18.
- 30 **Spiegelhalter D**, Grigg O, Kinsman R, *et al*. Risk adjusted sequential probability ratio tests: applications to Bristol, Shipman and adult cardiac surgery. *Int J Qual Health Care* 2003;**15**:7–13.