EDUCATION & DEBATE

Quality control: an application of the cusum

Sheila M Williams, Bryan R Parry, Martin M T Schlup

Implicit in the idea of medical audit are two notions: the quantifying of medical practice in terms of treatment, procedures, outcomes, and complications; and the use of these data to exact some form of accountability. Therefore some sort of objective monitoring, or quality control, of practices or procedures is needed so that periods of suboptimal performance in relation to an agreed standard can be recognised and, ideally, remedied.

Objective monitoring or quality control has been used for surveillance: methods have been developed to detect increases in rare events, such as birth malformations and to detect minimal epidemics.[1] A review of the concepts and definitions of quality control procedures that are commonly used in clinical chemistry have also been presented.[2] The use of the cumulative sum (cusum) has been suggested for both surveillance and quality control. Its use for examining sequential measures or for looking for changes over time has recently been described.[3] It has also been used for plotting temperature charts for assessing antimicrobial treatment in neutropenic patients.[4]

Because the cusum shows changes over time it can be used by individual practitioners to monitor their own performance as a form of quality control to give proof of ongoing competence in a particular skill. It can also be used to show progress in mastering a new technique. One advantage of this sort of self assessment is that an acceptable level of attainment must be defined so that how well it is met can be quantified. It has been suggested, for instance, that the completion rate for colonoscopic examination should be above 90%.[5]

The cusum has been described as a useful graphical tool for discerning trends.[3] For a series of observations $X_1, X_2, \ldots X_n$, the cusum can be defined as

$$S_n = \Sigma(X_i - X_0)$$

where $X_i = 0$ for a success and $X_i = 1$ for a failure. X_0 is a reference or target value. The cusum is interpreted by looking at its gradient or slope. For a series of observations which meet the specified criterion the cusum should be relatively level. If the reference value is specified in terms of an acceptable failure rate, for instance, the cusum will increase and consequently will have a positive slope when the failure rate is higher than the target value, and it will decrease when the failure rate is lower than that specified. This means that if the failure rate were specified as 10%, for instance, 0·9 would be added to the cusum for a failure and -0·1 for a success. In a series of examinations consisting of a success followed by a failure and four successes the cusum would take the values -0·1, 0·8, 0·7, 0·6, and 0·5.

Davies has described the derivation of the boundary lines for sequential tests for a change in proportions.[6] These can be used to compare the observed proportion of successes or failures with some standard proportion to find out whether an event is occurring in a greater or lesser proportion of cases than expected. It has been

suggested that the cusum can be thought of as a sequential test in reverse,[7] which means that boundary lines can be constructed in a similar way. In the present case, only one line is of interest because the aim is to identify periods where performance has fallen below an acceptable level. This level must be specified so that the boundary line can be constructed; and in effect, it is the alternate hypothesis (see appendix).

Two examples given below are taken from a series of consecutive colonoscopic examinations carried out by an experienced colonoscopist during the period August 1987 to May 1989 and by a novice during the period August 1987 to October 1990. The computerised records and dictated notes held in the gastroenterological unit of Dunedin Hospital were reviewed. A complete, or total, colonoscopic examination was regarded as either reaching the caecum (ileocolic anastomosis where appropriate) or encountering an impassable obstructive lesion.

EXAMPLE 1

The first example concerns the outcome of the 361 examinations done by the experienced colonoscopist; the cusum is shown in figure 1. As it had been
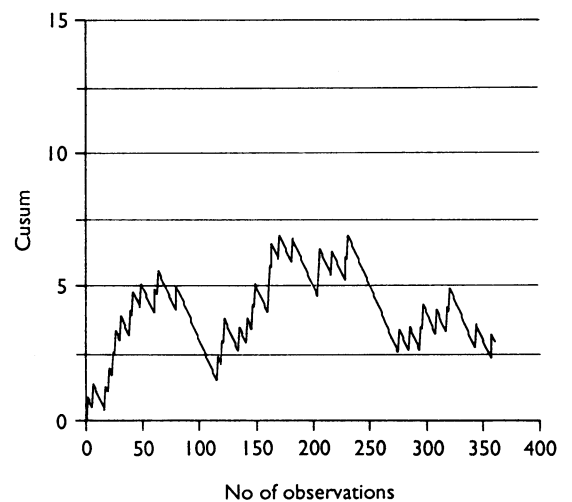

FIG 1—Cusum for experienced colonoscopist (n=361)

suggested that an acceptable failure rate for colonoscopies should be less than 10%[5] this value was used as the reference or target value for the cusum. As a failure rate of 20% was considered unacceptable boundary lines were constructed to identify series of observations in which the failure rate increased to 20% or more. It was found that, with a small amount of rounding, a boundary line 2·5 units above the horizontal axis provided a small risk of accepting an unsatisfactory series of observations and of rejecting a series of observations that were satisfactory. By drawing a series of lines 2·5 units apart it was possible to monitor the overall performance of the colonoscopist; the inter-

University of Otago
Medical School, Box 913,
Dunedin, New Zealand
Sheila M Williams,
biostatistician, department of
preventive and social medicine
Bryan R Parry, senior
lecturer, department of
surgery
Martin M T Schlup, senior
lecturer, department of
medicine

Correspondence to:
Mrs Williams.

section of any of these lines from below marked a period that included more failures than expected. Where this happened the boundary line became the new target line.

The cusum cut the boundary line for the first time after 27 colonoscopies, which included six failures. The overall failure rate for these was 22%. The cusum intersected with the second boundary line after a further 38 examinations, which included six failures (failure rate 16%). It cut the same boundary line again at the 164th observation, the preceding four observations having included three failures. Some sections of the graph represent long runs of successes, the longest of which was 43. After such a long run the target line should be recentered by moving it to the one below.

It is impossible to tell from a retrospective analysis whether the alarms were the consequence of poor operator performance, an unusual run of technically difficult colonoscopies, or instrument malfunction. The control chart is especially sensitive to short runs of failures, and identified them quickly. It could be argued that this was a shortcoming as in a retrospective analysis it can be seen these short runs were of little consequence. Had the failures persisted, however, it would be an advantage to identify such periods as soon as possible.

EXAMPLE 2

The second example, shown in figure 2, describes the progress of the beginner. He clearly did not reach the agreed standard until at least the 67th examination, when he had 22 successes in a row. There were 21 more observations, which included seven failures, before he reached a second plateau completed at examination 129. After this, two periods of suboptimal performance were identified, one at observation 167 after seven further failures (five in the previous 23 observations) and again at 208 after five failures in the previous 17 examinations. His overall failure rate from the beginning of the second plateau was 9·6%. This example is an illustration of what has come to be called the learning curve. The first plateau provides an indication of where the novice colonoscopist's level of skill approached that required.

The overall failure rate of the experienced colonoscopist was 10·8%, which was not significantly different from the target value (binomial test, $z=0·6$, $p>0·05$). The rate for the novice was 18·0%, which was significantly different from the standard value ($z=4·8$, $p<0·05$). The binomial test tests for a difference between the observed value and the expected value but gives no information about the changes in the proportion of failures over time, something that is particularly important in the context of learning. The cusum
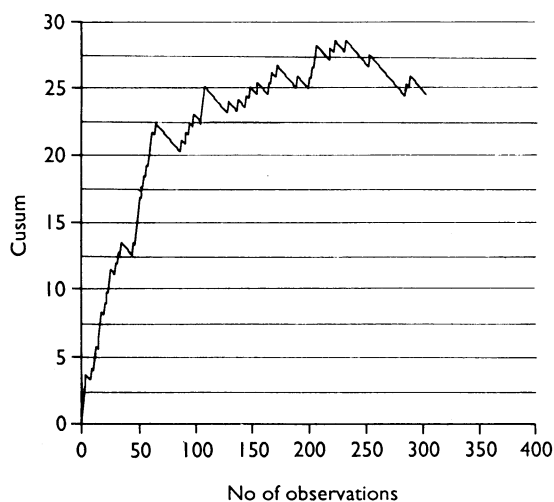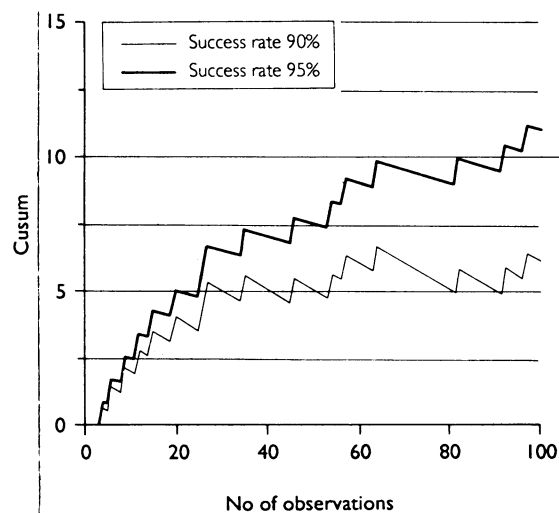

FIG 3—Cusums for two standards in a series of endoscopic retrograde cholangiopancreatography examinations (n=100)

provides an overall picture of progress as well as proving some objective evidence of when the target value has been achieved.

EXAMPLE 3

The data for a third example involve the outcome of the first 100 endoscopic retrograde cholangiopancreatography examinations carried out by one operator in the gastrointestinal unit of Dunedin Hospital between 1986 and 1988. The examination was regarded as successful if the correct duct was cannulated. Again a success rate of at least 90% seems to be acceptable.[8] A failure rate of 20% or more was considered unsatisfactory.

Figure 3 shows the learning process. It seems that only 30 cases were needed for the operator to become proficient.

In another long series of selected endoscopic retrograde cholangiopancreatography examinations a completion rate of 95% was reported.[9] This standard is also shown in figure 3. Because the cusum continued to rise it must be concluded that a 95% success rate was not reached in this series. A success rate as high as 95% may be an unrealistic goal for a series of unselected examinations.

Discussion

The advantage of a quality control procedure such as this is that after each observation it is possible to make one of two decisions: to accept that the level of technical performance is satisfactory or to conclude that some real change has occurred, a decision made if the cusum meets one of the boundary lines from below. The first example shows that the cusum can be used to show that a practitioner can achieve and maintain an acceptable level of success in carrying out a particular procedure. Figure 1 gives a good overview of the performance of the experienced colonoscopist and identifies several periods that were characterised by long runs of successes.

This study also shows that it is possible to identify periods of poor performance. For clinicians serious about the quality control of technical procedures, being able to recognise such periods is important and an explanation for them should always be sought. If monitoring were carried out prospectively then any deviation from acceptable standards would permit early attention to any problem in technique or protocol. Whereas such monitoring makes maintenance of acceptable standards a personal issue it also quantifies them so that comparisons can be made with other operators in the same unit or other centres.

FIG 2—Cusum for inexperienced colonoscopist (n=305)

In both the second and third examples the examiners were beginners, and it is clear that their initial performances did not meet the specified standard. The cusum does, however, suggest that after about 80 examinations the performance of the colonoscopist reached an acceptable level, and it gives a good picture of his experience from novice to skilled examiner. A more complete audit would describe the findings of each examination.[10] The third example shows a shorter learning phase before a 90% success rate was achieved. Again, the cusum indicates how the proportion of successes has changed over time. These examples highlight the advantages of the cusum: its use allows individuals to prove how proficient they have become by quantifying how well their success rate meets an agreed standard. Instead of consisting of an arbitrary number of examinations, training can be continued until an acceptable level of attainment is reached, and supervision and guidance can be tailored to meet individual needs.

Maintaining levels of technical skill is not enough. The findings and the outcome of examinations should also be recorded so that patterns of results can be examined and comparisons made with other examiners, possibly in other centres, and over time. Quality control and objective and quantified recording of the findings meet the recommended criteria for medical audit.[11]

We recommend that this simple cusum technique be considered as a means of monitoring performance, both in the learning phase and in the maintenance of standards. Its potential application to other clincial procedures such as chorionic villus sampling, fine needle aspiration of the pancreas under ultrasound, intubation of the newborn, radial artery catheterisation in neonatology, central vein cannulation of adults, or for surgical procedures either for quality control or to illustrate the learning process should be considered. While meeting the specified success rate for each procedure is a personal issue, the setting of explicit criteria makes comparisons among different operators and different units much more straightforward and provides a quantitative measure for audit. The cusum can also be used when the variable of interest is continuous; again, an agreed target value must be made explicit.

Appendix

This appendix gives the formulas, provided by Davies,[6] on which the boundary lines were based.

$$s = Q/(P+Q) \quad (1)$$
$$h_0 = b/(P+Q) \quad (2)$$
$$h_1 = a/(P+Q) \quad (3)$$

where $P = \ln(p_1/p_0)$, $Q = \ln(1-p_0)/(1-p_1)$, $a = \ln(1-\beta)/\alpha$, and $b = \ln(1-\alpha)/\beta$.

p_0 corresponds to the failure rate under the null hypothesis.

p_1 corresponds to the failure rate under the alternative hypothesis.

α is the risk of accepting a bad batch of observations.

β is the risk of rejecting a good batch of observations.

As s was to be used as the target value for the cusum $s = 0.1$, the failure rate considered to be unacceptable was 0.2. The value of p_0 can be found from (1) using trial and error. In this case $p_0 = 0.04$.

Setting the type 1 and type 2 error rates to be equal means that $h_0 = h_1$; in this example $\alpha = \beta = 0.01$, so that $h_1 = 2.5$.

The average run lengths were 41, 26, and 72 for p_0, p_1, and s respectively.

1 Gallus G, Mandelli C, Marchi M, Radaelli G. On surveillance methods for congenital malformations. *Statistics in Medicine* 1986;5:565-71.
2 Nix AB, Rowlands J, Kemp KW, Wilson DW, Griffiths K. Internal quality control in clinical chemistry: a teaching review. *Statistics in Medicine* 1987;6:425-40.
3 Altman DG, Royston P. The hidden effect of time. *Statistics in Medicine* 1988;7:629-37.
4 Kinsey SE, Giles FJ, Holton J. Cusum plotting of temperature charts for assessing antimicrobial treatment in neutropenic patients. *BMJ* 1989;299: 775-6.
5 Williams CB. Colonoscopy. *Current Opinion in Gastroenterology* 1987;3:36-42.
6 Davies OL. *The design and analysis of industrial experiments.* London: Longman, 1978.
7 Van Dobben de Bruyn CS. *Cumulative sum tests: theory and practice.* London: Griffin, 1968.
8 Cotton PB, Williams CB. *Practical GI endoscopy.* London: Blackwell Scientific, 1982.
9 Vaira D, Dowsett JF, Hatfield ARD, Cairns SR, Poldoron AA, Cotton PB, et al. Is duodenal diverticulitis a risk factor for sphincterotomy? *Gut* 1989;30:939-42.
10 Parry BR, Williams SM. Competency and the colonoscopist: a learning curve. *Aust NZ J Surg* 1991;61:419-22.
11 Shaw CD, Costain DW. Guidelines for medical audit: seven principles. *BMJ* 1989;299:498-9.

Current Issues in Cancer

Management of early breast cancer

R D Rubens

This is the third in a series of articles examining recent developments in cancer

Imperial Cancer Research Fund Department of Clinical Oncology, Guy's Hospital, London SE1 9RT
R D Rubens, *professor*

Series edited by: Dr G M Mead.

BMJ 1992;304:1361-4

Breast cancer is Britain's most common malignancy in women, estimated to affect one in 12 of the female population. About 80% of women present with apparently localised disease which is surgically resectable: so called early breast cancer. Nevertheless, about half relapse despite intended curative treatment. Controversy about the treatment of early breast cancer has existed for at least a century and continues. Two questions have predominated. Firstly, is mastectomy necessary or can the affected breast safely be preserved? Secondly, can adjuvant systemic therapy eliminate micrometastatic disease present, but not identifiable, at diagnosis, and if so how should patients best be selected for such treatment? The available evidence on these issues is considered and a reasonable approach to present good practice is described.

Definition of early breast cancer

The term early breast cancer describes disease which, at presentation and after initial assessment, is apparently confined to the breast and deemed to be technically resectable; axillary lymph nodes may be palpable. The cancer is operable if clinical examination excludes extension of the primary tumour or axillary node metastases to either the skin or the chest wall and confirms that there is no supraclavicular lymphadenopathy. In other words, signs of locally advanced disease are absent. Ancillary investigations, aiming at identifying potential distant metastases, normally include haematological and biochemical tests and chest radiography. If these produce normal results many doctors consider further investigation unnecessary. However, isotopic bone scanning to screen for skeletal metastases is often routinely performed, but the low positive predictive value has called into question the cost effectiveness of this investigation.[1]

The term early to describe operable breast cancer is often a misnomer as the low resolution of staging tests means that occult metastatic disease is not identified in many patients. The risk of distant metastatic disease is particularly high when axillary lymph nodes are