
COMMENTARIES

A statistical approach to measuring the competence of anaesthetic trainees at practical procedures

I. G. KESTIN

Summary

Cusum analysis is a statistical technique to distinguish deviations from an acceptable failure rate. The progress of anaesthetic trainees learning four practical procedures (obstetric extradural anaesthesia, spinal anaesthesia, central venous cannulation and arterial cannulation) was monitored from their first attempt using cusum analysis. Suitable acceptable and unacceptable failure rates for each procedure were chosen by consultant anaesthetists. For obstetric extradural anaesthesia, four trainees eventually achieved acceptable failure rates (5%) and the number of attempts required to demonstrate this statistically ranged from 29 to 185; three trainees had an unacceptable failure rate (10%) and five trainees had inconclusive records. For spinal anaesthesia, two trainees achieved an acceptable failure rate (10%) and the number of attempts required to demonstrate this statistically ranged from 39 to 67; two trainees had an unacceptable failure rate (20%) and four trainees had inconclusive records. One trainee demonstrated statistically an acceptable failure rate in arterial cannulation (20%) after 14 attempts and four trainees had inconclusive records. Two records of central venous cannulation were inconclusive. Some records showed variable failure rates which were sometimes associated with lack of practice or a change in technique. Cusum analysis can be used to monitor training in practical procedures and as a continuous audit of quality of clinical practice. (*Br. J. Anaesth.* 1995; 75: 805–809)

Key words

Anaesthesia, audit. Anaesthetist, training. Statistics, cusum.

After the introduction of structured training, trainees may be required to demonstrate that they have attained minimum standards of competence. This would be to ensure uniform standards of training, allow comparisons between different hospitals and for audit. Cusum analysis is a statistical technique used in industry as a method of quality control, analogous to the sequential design for clinical trials [1]. The potential applications in anaesthesia are two-fold: to determine when a trainee is proficient in a new practical procedure, and as a continuous audit of quality of practice for experienced clinicians.

With each successive failure or success at the procedure, positive or negative increments are added to a cumulative score, the cusum, which increases

with failure and decreases with success. The cusum starts at zero, so success is indicated by a declining trend of the cusum, and failure by an increasing trend of the cusum. Acceptable and unacceptable failure rates at the procedure and the desired magnitude of the type 1 and type 2 errors (α and β) are chosen, and two boundary limits to the cusum, h_0 and h_1 , are calculated. When the cusum declines below the lower boundary limit h_0 , then the true failure rate is not statistically different from the acceptable rate (the null hypothesis), with the risk of a type 2 error equal to β ; if the cusum exceeds the upper boundary limit h_1 , then the true failure rate is statistically significantly higher than the acceptable rate (the alternative hypothesis), with a risk of a type 1 error equal to α ; if the cusum stays between the two boundaries, then observations must be continued.

The advantage of plotting the cusum as a graph is that periods of acceptable and unacceptable performance and trends in performance are immediately obvious to the trainee and trainer. For the experienced clinician, the cusum graph is a continuous audit of quality and a measure of the effects of any change in technique, for example, using new equipment or a different anatomical approach. The cusum analysis has been used to monitor anaesthetic trainees learning four common practical procedures.

Methods

All senior house officers (SHOs) and registrars who joined the Anaesthetic Directorate in Derriford Hospital between 1994 and 1995 were asked to plot the cusum as they learnt four practical procedures; insertion of arterial and central venous cannulae, subarachnoid anaesthesia and extradural analgesia or anaesthesia in obstetric patients. Records were kept if the trainees had no previous experience of the procedure, unless the trainees had a complete sequential record of their successes and failures from a previous hospital using the same definitions. This record was used to plot the cusum, which was continued during their training in Plymouth. If no record had been kept by trainees with previous experience, the cusum was not plotted for this procedure. Four separate graphs were supplied to the trainees to plot the cusum with successive attempts at each of the procedures.

I. G. KESTIN, BA, MB, BS, FRCA, Derriford Hospital, Derriford Road, Plymouth PL6 8DH. Accepted for publication: June 14, 1995.

This article is accompanied by Editorial 1.

Failure was defined as: arterial cannulation; attempt at another artery or another anaesthetist attempting the same site; central venous cannulation; using a different approach to the same vein, another anaesthetist attempting the same vein or changing to another site; extradural anaesthesia in obstetric patients or subarachnoid anaesthesia; failure to obtain satisfactory analgesia or anaesthesia for any reason.

A consensus was obtained from consultant anaesthetists in Derriford Hospital about acceptable and unacceptable failure rates for these procedures. Consultants were asked "What failure rate would be acceptable in a trainee who has just learnt this procedure, but lacks experience?" and "What failure rate would be unacceptable in a trainee who has just learnt this procedure, and would indicate a further period of supervised training is required?"

The α and β errors were chosen to be 0.1, and using the acceptable and unacceptable failure rates, the three values required to plot the cusum were calculated (appendix). These are h_0 and h_1 , the boundary limits, and s , where the cusum increases by $1-s$ for a failure, and decreases by s for a success. The values for h_0 , h_1 and s were multiplied by 10 and rounded to the nearest integer. When α and β are equal, then h_0 and h_1 are of equal magnitude, and these limits were marked on the graph. The X-axis is the attempt number and the trainees plotted the cusum on the Y-axis with successive attempts. The trainees were instructed on how to use the graph and to come and see the author if the cusum crossed a boundary. In those trainees whose cusum exceeded h_1 (an unacceptable failure rate), further lines were drawn on the graphs at heights $2h_1$, $3h_1$ and $4h_1$ as required. These lines are the boundary limits to accept or reject the null hypothesis for further attempts when the cusum is continued after the first boundary line (appendix).

The records were collected from the trainees before they left Derriford Hospital or at the end of 1994, when the project finished. When the records had been collected, the cusum analysis was repeated to check for differences between the exact values and the integer values used by the trainees for the graph.

Results

The acceptable and unacceptable failure rates for the four procedures and the values to calculate the cusum are shown in table 1. The number of

Table 1 Acceptable and unacceptable failure rates for the four procedures (as defined by a consensus of consultants), the values of s and the boundary limits for the cusum

	Acceptable failure rate	Unacceptable failure rate	s	h_0, h_1
Obstetric extradurals	5%	10%	0.07	2.94
Spinal anaesthesia	10%	20%	0.14	2.71
Central venous cannulation	5%	15%	0.09	1.81
Arterial cannulation	20%	40%	0.29	2.24

Table 2 Outcome for trainees learning four practical procedures (number of records and range of number of attempts)

	Acceptable failure rate (number of records (range of attempts to statistical significance))	Unacceptable failure rate (number of records (range of attempts to statistical significance)).	No statistical significance (number of records (range of attempts))
Obstetric extradurals	4(29-185)	3(11-21)	5(21-119)
Spinal anaesthesia	2(39-67)	2(13-23)	4(10-91)
Central venous cannulation	0	0	2(3-9)
Arterial cannulation	1(14)	0	4(7-29)

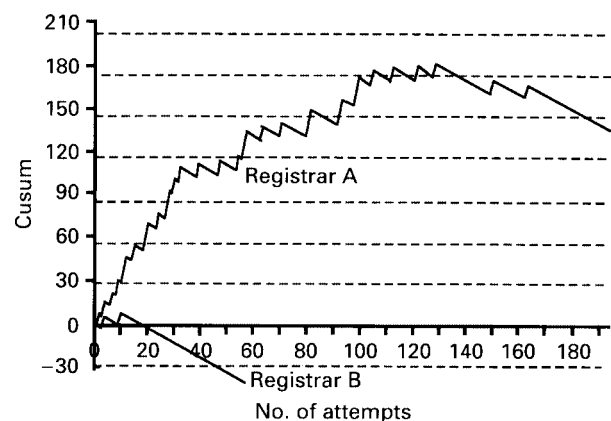


Figure 1 The cusum for obstetric extradurals from two anaesthetic registrars. For registrar A, the cusum increases through six successive boundary lines in the first 99 attempts; the true failure rate for this series is 27%. From attempt 100 onwards, the cusum is stable and then declines through two boundary lines; the failure rate for this series (6.7%) is not significantly different from the acceptable rate (5%). For registrar B, a lower boundary line is crossed after 47 attempts; the true failure rate is not significantly different from the acceptable rate.

completed records ranged from 2 to 12 (table 2). Comparing the graphs with the cusum analysis using the exact values from table 1, there were small differences in the number of procedures done before crossing a boundary, but no difference whether the null hypothesis was rejected or accepted. The results and figures given are those kept by the trainees.

An upper boundary line was crossed (denoting an unacceptable failure rate) in five records; three for obstetric extradurals and two for spinal anaesthesia. The number of attempts made before crossing the boundary ranged from 11 to 21 for obstetric extradurals and from 13 to 23 for spinal anaesthesia. Seven trainees eventually achieved a failure rate not statistically different from the acceptable failure rate; the number of attempts made before crossing the boundary ranged from 29 to 185 for obstetric extradurals, from 39 to 67 for spinal anaesthesia and 14 attempts for arterial cannulation. In 15 records, the cusum remained between the boundaries (table 2). There were only two records of central venous cannulation. All but two trainees has previous

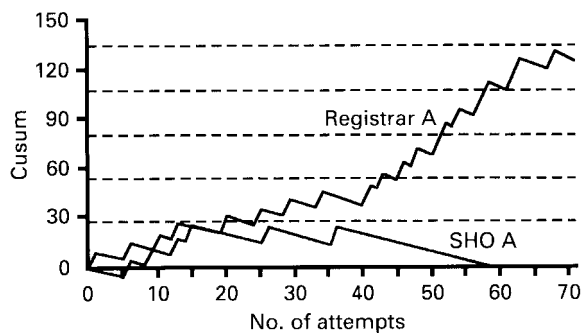


Figure 2 The cusum for spinal anaesthesia from an anaesthetic SHO and a registrar. Registrar A (upper plot) has an unacceptable failure rate of 28 % with no evidence of improvement over 71 attempts; this is the same registrar as registrar A in figure 1. The cusum for SHO A (lower plot) has not crossed a boundary; the null hypothesis cannot be accepted or rejected. The failure rate for this series is 10 %.

experience of this procedure, and none had any record of their failure rate.

Some typical records are shown in figures 1 and 2. Figure 1 is two plots of the cusum for obstetric extradurals from two registrars who had learnt obstetric anaesthesia before coming to Plymouth. The first 62 extradurals by registrar A were given as an SHO at the previous hospital, followed by 18 months of non-anaesthetic training before coming to Plymouth, where the failure rate continued to be unacceptably high. The record can be divided into three sections: for the first 32 attempts, the cusum rises steeply, crossing three boundary limits; the failure rate is 44 %. For the next 67 attempts the overall rise is not as steep and there are periods when the cusum is relatively stable, for example between attempts 33 and 53. The failure rate for this series is 19 %. The run of failures beginning after attempt number 91 is attributed by the trainee to lack of practice because of taking the FRCA examination, and changing the technique of locating the extradural space from loss of resistance to air to loss of resistance to saline. After 99 attempts, the cusum oscillates and then steadily declines until a boundary limit is crossed from above, and the null hypothesis (an acceptable failure rate) can be accepted. The failure rate for this series is 6.7 %. This trainee required 185 attempts to achieve and demonstrate statistically a satisfactory failure rate. In contrast, registrar B has only two failures in the first few attempts, followed by a steady run of successes until the cusum falls below the lower boundary limit after 47 attempts. The null hypothesis can now be accepted; the true failure rate is not significantly different from the acceptable failure rate of 5 %.

Figure 2 shows two plots of the cusum for spinal anaesthesia from an anaesthetic SHO and a registrar. The performance of the registrar (the same registrar as registrar A in fig. 1) is consistently worse than the acceptable failure rate, and shows no evidence of improvement over 71 attempts; the failure rate is 28 %. The positive slope of the cusum increases after attempt number 36, which implies performance has worsened. This occurred after 18 months of non-anaesthetic medical experience, and then re-entering anaesthesia as a registrar, and possibly attempting

spinal anaesthesia in patients who were more technically difficult than those encountered as an SHO. This registrar had difficulty with both extradural and subarachnoid anaesthesia; there may have been a basic fault in knowledge of the anatomy of the spinal column or lack of knowledge of different approaches that may be used in difficult patients.

The other cusum in figure 2 is from a novice SHO learning spinal anaesthesia (SHO A). The cusum approaches h_1 , but does not cross it. The cusum then oscillates around the value of 20, and after attempt number 37, the cusum steadily declines with a sequence of successes. So far, the null hypothesis cannot be accepted or rejected. The failure rate for this series of 59 attempts is 10 %.

Discussion

Cusum analysis has been used to monitor the quality of investigative procedures in medicine, but not for assessing training [2]. The advantages of cusum analysis can be illustrated by the example of the records from registrar A in figures 1 and 2. This registrar was aware that his success rate at obstetric extradurals was not good, but the department at the previous hospital was unaware of any problem. The consultants in Plymouth with an interest in obstetric anaesthesia reported to the author (as college tutor) that this trainee may have a problem with obstetric extradurals, but were not sure of the magnitude of the problem. The cusum analysis clearly demonstrated that the failure rate was unacceptable, and the trainee was given some additional advice and supervision (there was no obvious benefit from this). It was not until 185 attempts had been made that an acceptable failure rate was demonstrated statistically. The second feature illustrated is that success may not be maintained; an increase in the failure rate occurred after 91 attempts, associated with a change in technique and lack of practice. The poor performance in spinal anaesthesia (fig. 2) was unrecognized by everyone. The advantage of the graphical record is that this poor performance would have been recognized at a much earlier date as soon as an upper boundary line had been crossed.

No attempt was made to modify the rather *ad hoc* method of training in these practical procedures in Plymouth. Better training may allow higher success rates to be obtained earlier. Earlier detection of difficulties should avoid the circumstances illustrated in figures 1 and 2; poor technique can be corrected before the trainee becomes demoralized by repeated failures. The wide variation in the speed that trainees become proficient can be seen by comparing the two records in figure 1. It will be difficult to organize structured training programmes to cope with this wide variability. Satisfactory competence cannot be assured by defining a minimum caseload or a minimum number of successful procedures; the registrar performed 71 spinal anaesthetics without achieving the acceptable failure rate and no evidence of improvement, and there were sequential series of 10 successful extradurals within an overall pattern of unacceptable performance.

Simple graphical methods that sequentially dis-

play success or failure as diagonal upward or downward movement of the graph have been reported as measures of proficiency [3]. These are simple to use, but have no statistical method of testing the failure rate against a minimum standard. The other method of testing proficiency is by an observer watching a clinical procedure by a trainee, and scoring performance according to predefined criteria. This has been developed for spinal and extradural anaesthesia [4, 5]. The advantage of this technique is that a whole range of criteria can be assessed objectively at the same time, including safety, communication with the patient, sterile technique and technical competence. This principle is used in the objectively structured clinical examination in the FRCA part 3 examination, but is too time-consuming as a routine method of assessing trainees' progress. The data required for cusum analysis are easy to obtain and allow statistical decisions to be made with reference to minimum acceptable standards. The disadvantages are that it relies on the honesty of trainees, their consistent interpretation of the definitions of success and failure, and does not assess other important aspects such as safety.

Another feature illustrated by this study is that relatively large numbers of procedures are required before the null hypothesis can be accepted. In figure 1, registrar B needed to perform 47 extradurals before the null hypothesis of an acceptable success rate was accepted, despite having only two failures in this sequence. It may be that our definitions of acceptable and unacceptable failure rates are too stringent and that these success rates may be expected only from experienced anaesthetists. The unacceptable and acceptable failure rates used in this study were obtained from a consensus of consultant anaesthetists in Plymouth. There are other methods of defining the standards, for example, from a survey of the literature. These standards could be altered to suit the patient population (e.g. obstetric compared with non-pregnant patients, intensive care compared with operating theatre procedures) or altered according to the experience of the trainee. If training is shorter, it may be impossible to provide sufficient experience of these procedures to demonstrate statistically acceptable success rates. It would be important to know the failure rates of experienced practitioners if national standards are to be specified for structured training. This information is lacking, but should not be difficult to obtain.

There were two difficulties using this graphical method. Integer values are convenient to construct a graphical record because the original fractional values in table 1 are not practical. Therefore, the values of s and h were multiplied by 10 and the result rounded to the nearest integer. Several anomalies result as a consequence of this rounding to an integer. First, the graphs for spinal anaesthesia and obstetric extradurals were essentially the same, despite using different acceptable and unacceptable failure rates. In practice, this did not matter in this small study; it was obvious when a trainee was not meeting the standard, and exact analysis of the cusum did not alter any of the statistical decisions.

Second, the discrepancy between the cusum analysis using the integer and fractional values increases as the number of attempts increases. Third, using the integer values may lead to incorrect analyses in certain circumstances. For example, with the integer values used for obstetric extradurals, the cusum increases by 9 for a failure and decreases by 1 for a success. The cusum of a trainee who starts with a failure and then regularly fails every 10th extradural thereafter would remain between 0 and 9 and never cross a boundary line, despite having a true failure rate of 10% equal to the unacceptable failure rate. If the cusum analysis is done with the exact fractional values, this does not happen; the cusum slowly trends upwards until the null hypothesis is rejected after 98 attempts. For accuracy, the original variables ought to be used to calculate the cusum. The other problem with the graphs was practical; the graphs were easily damaged during daily use, and had to be rewritten on occasion. For these reasons, trainees are no longer required to keep four contemporaneous records on paper. Currently, the trainees in Plymouth keep a separate log book for practical procedures. A computerized system of keeping a personal log book would make cusum analysis simple and practical.

In summary, cusum analysis is suitable for assessing some aspects of training and quality of anaesthetic practice, which may be required for structured training and continuing medical education.

Appendix

Choose the values for: p_0 , the unacceptable failure rate; p_1 the acceptable failure rate; α , the risk of a type 1 error; β , the risk of a type 2 error. The null hypothesis is: the true failure rate is not different from the acceptable failure rate. The alternative hypothesis is: the true failure rate is equal to or exceeds the unacceptable failure rate.

The formulae to calculate the variables for the cusum are:

$$a = \ln \left[\frac{1 - \beta}{\alpha} \right]$$

$$b = \ln \left[\frac{1 - \alpha}{\beta} \right]$$

$$P = \ln \frac{p_1}{p_0}$$

$$Q = \ln \left[\frac{1 - p_1}{1 - p_0} \right]$$

$$h_0, \text{ the lower boundary limit, } = \frac{-b}{P + Q}$$

$$h_1, \text{ the upper boundary limit, } = \frac{a}{P + Q}$$

$$s = \frac{Q}{P + Q}$$

The starting value for the cusum is 0, and the cusum is increased by $(1 - s)$ for a failure, and decreased by s for a success. If the cusum exceeds h_1 , then the null hypothesis is rejected; if the cusum decreases below h_0 , then the alternative hypothesis is accepted. If the cusum stays between h_0 and h_1 , then no decision can be made, and observations are continued. As in the example in this article, when α and β are equal, then $h_0 = h_1 (=h)$. If a graphical record is to be used as a continuous record of training

and proficiency, then lines drawn parallel to the X -axis at heights $2h$, $3h$, $4h$ and $-2h$, $-3h$, $-4h$ as required, are the boundary limits to accept or reject the null hypothesis for further attempts after a boundary limit has been crossed. If the cusum increases from one boundary line to the next boundary line above, then the null hypothesis can be rejected for that series of attempts; if the cusum decreases from one boundary line to the boundary line below, then the null hypothesis can be accepted for that series of observations.

References

1. Davies OL. *The Design and Analysis of Industrial Experiments*, 2nd Edn. London: Longman, 1978; 57–93.
2. Williams SM, Parry BR, Schlup MMT. Quality control: an application of the cusum. *British Medical Journal* 1992; **304**: 1359–1360.
3. Lawler PG, Patla VR, Garcia E, Puttick N. Assessment of training in anaesthesia and related skills. *Anaesthesia* 1991; **46**: 597.
4. Sivarajan M, Lane PE, Miller EV, Liu P, Herr G, Willenkin R, Winter P, Hardy CA, Mulroy MF. Performance evaluation: continuous lumbar epidural anesthesia skill test. *Anesthesia and Analgesia* 1981; **60**: 543–547.
5. Sivarajan M, Miller E, Hardy C, Herr G, Liu P, Willenkin R, Cullen B. Objective evaluation of clinical performance and correlation with knowledge. *Anesthesia and Analgesia* 1984; **63**: 603–607.