



Assessment of procedural skills training and performance in anesthesia using cumulative sum analysis (cusum)

Évaluation de la formation en habiletés techniques et de la performance en anesthésie au moyen de l'analyse des sommes cumulées (cusum)

Tim Starkie, BMBS · Elizabeth J. Drake, BM

Received: 19 April 2013 / Accepted: 25 September 2013 / Published online: 16 November 2013
© Canadian Anesthesiologists' Society 2013

Abstract

Purpose *The current methods (work based assessments and logbooks) used to assess procedural competency and performance have well-documented deficiencies. Cumulative sum analysis (cusum), a statistical method that generates performance graphs over time, is an alternative tool that is not currently widely used. The purpose of this review is to investigate its current role in anesthetic procedural skills training and performance.*

Source *A literature search of MEDLINE®, EMBASE™, BNI, CINAHL®, the Cochrane Library, NHS Evidence, and the Trip database was performed in October 2012. All papers using cusum to investigate performance in anesthetic procedural skills were included. Their references were searched manually to identify any additional studies.*

Principal findings *Thirteen papers were identified. The procedural skills they investigated could be split broadly into three groups: ultrasound skills, airway and cannulation, and regional anesthesia. All papers had small sample sizes (< 30), with most researching novice*

trainee performance. Wide ranges were seen in the number of procedures required to achieve cusum-defined procedural competency. These were due to differences in definitions of success/failure of a procedure, the acceptable and unacceptable failure rates used for the initial cusum calculation, and individual trainee performance.

Conclusion *Cusum can be used to assess procedural competency, but several problems need to be overcome before it can become a universally accepted method. It is ideally placed to be used as a quality control tool for a trained individual and could also be used to assess the impact of new training methods or equipment on performance.*

Résumé

Objectif *Les méthodes actuelles (évaluations du travail accompli et registres) utilisés pour évaluer la compétence pour la réalisation de certaines procédures, ainsi que la performance de réalisation des actes présentent des lacunes bien documentées. L'analyse des sommes cumulées (cusum) – méthode statistique générant des graphiques de performance en fonction du temps – est un autre outil dont l'utilisation est actuellement encore limitée. L'objectif de cette analyse est de mieux connaître sa place actuelle en anesthésie.*

Source *Une recherche documentaire a été menée dans les bases de données MEDLINE®, EMBASE™, BNI, CINAHL®, Cochrane Library, NHS Evidence et Trip en octobre 2012. Tous les articles utilisant le cusum pour évaluer la performance des habiletés techniques en anesthésie ont été inclus. Leurs références ont fait l'objet de recherches manuelles pour identifier des études supplémentaires.*

Constatations principales *Treize articles ont été identifiés. Les habiletés techniques qu'ils étudiaient ont*

Author contributions Both authors have made substantial contributions to this work, and editing of the final drafts was undertaken together. Below is a breakdown of the main author for each section. *Tim Starkie*: implication statement, abstract, introduction, regional anesthesia (majority), airway and cannulation skills (majority), ultrasound skills first paragraph, the pitfalls of cusum (majority), use of cusum for quality control in anesthesia (majority). *Elizabeth Drake*: literature review, regional anesthesia paragraph 4, airway and cannulation skills last paragraph, ultrasound skills (majority), discussion (majority), summary of papers (Table 1) and Table 2.

T. Starkie, BMBS (✉) · E. J. Drake, BM
Department of Anesthesia, Derriford Hospital, Derriford Road,
Plymouth PL6 8DH, UK
e-mail: timstarkie@yahoo.co.uk

été globalement scindées en trois groupes: habiletés en échographie, en gestion des voies aériennes et intubation, et en anesthésie locorégionale. Tous les articles comportaient des échantillons de petite taille (moins de 30 sujets) et la plupart étudiaient la performance de stagiaires débutants. Le nombre de procédures requises pour atteindre l'habileté technique définie par la cusum pouvait varier considérablement. Cela était dû à des différences dans les définitions de réussite ou d'échec d'une procédure, des taux d'échec acceptables et inacceptables utilisés pour le calcul initial de la cusum, et de performance individuelle du stagiaire.

Conclusion *La cusum peut être utilisée pour évaluer la compétence à réaliser des procédures mais plusieurs problèmes doivent être surmontés avant qu'elle ne puisse devenir une méthode universellement acceptée. Elle se situe idéalement comme outil de contrôle de la qualité d'un sujet formé et elle pourrait également servir à évaluer l'impact de nouvelles méthodes de formation ou d'un nouvel équipement sur la performance.*

Anesthesia is a high-risk medical specialty where the ability to perform practical procedures proficiently is essential. In spite of this, training opportunities are under significant pressure from a variety of factors, including the decrease in working hours demanded by the European Working Time Directive and an increasingly time-pressured clinical environment. This has led some in the medical profession to question whether there is time for procedural skills to be learned adequately within current training programs.¹

In 2007, the American Accreditation Council for Postgraduate Medical Education (ACGME) developed an initiative called the "Outcome Project"² which emphasized the importance of the educational outcomes of residency programs rather than only their potential to educate. This requires data on learners' performance to be assessed adequately and their competence to be documented reliably. This ethos is invaluable to ensure that doctors are properly trained and to protect patients from unsafe practice, and it is in line with the international trend towards competency-based training.

A robust system for ensuring competence in procedural skills in anesthesia is therefore required: first, to address concerns regarding the lack of training opportunities, and second, to show that the delivered training is effective. Current methods of evaluating technical skills are logbook summaries and work-based assessments (WBAs). The former is best-suited to detailing the learning cases encountered.³ It does not normally include a record of success or failure and is unable to identify unsafe or poor

practice. The WBAs, designed to document proficiency in specific skills, are also prone to weakness. They assess only single (often favourably selected) episodes; they may be completed only after success; and the assessor can be carefully chosen to avoid poor reports. These problems affect their validity and reliability. It is usually easy for instructors to recognize trainees having extreme difficulties from logbook analysis and WBAs, but it is much harder to identify more subtle performance deficiencies.⁴ The rotational nature of most training programs compounds this problem as it commonly results in trainees working with many different trainers. This results in a lack of continuity in their supervision and, therefore, in their evaluations and assessments, which may mean that episodes of poor performance are dismissed as "one off" mistakes rather than recognized as a pattern of behaviour, repeated failings, or an inability to progress that requires further action.

Assessment of procedural skills in anesthesia is poor compared with other domains of learning and has fallen behind surgical fields.⁵ Owing to the dependency of patients' outcomes on a surgeon's technical skills, research into this area has been pioneered in surgery.⁶ Several assessment tools have now been developed and validated for use on surgical trainees outside the operating room. Examples include: The Objective Structured Assessment of Technical Skills,^{7,8} which involves a task checklist and a global rating score; the McGill Inanimate System for Training and Evaluation of Laparoscopic Skills,⁹ which tests generic laparoscopic skills; and the Imperial College Surgical Assessment Device, which tracks trainees' hand movements via sensors and provides an effective index of technical skill in both laparoscopic¹⁰ and open^{11,12} procedures. Anesthesia knowledge, clinical judgment, and communication skills are all tested in postgraduate exams,¹³ but there is currently no formal evaluation of procedural skills. Given that numerous studies have shown that the time required to achieve competency at specific procedures varies widely depending on the individual learner,^{4,14} there is a great need for a reliable and valid method for demonstrating procedural competency and for identifying struggling trainees who require additional support.

Cumulative sum (cusum), a statistical method that looks at the outcome rather than at the process of performing procedural skills, is an alternative tool that may be used to assess an individual's procedural performance. It produces graphs that allow rapid detection of deviations from a pre-established standard, initially being developed during World War II as a quality control tool in munitions factories.¹⁵ The graphs are generated by relatively simple calculations based on set acceptable and unacceptable failure rates and the degree to which type 1 (α) and type 2 (β) errors (false positive and false negative errors) will be

tolerated (Appendix). The null hypothesis is: the true failure rate is not different from the acceptable failure rate. The calculations produce decision limits (h_0 and h_1) and a value for the cusum - s . The cusum value is plotted on the y-axis, and the number of consecutive attempts is plotted on the x-axis,¹⁶ as shown in Fig. 1. The graphs start at zero and successes cause the cusum to fall by a value equal to s and failures to rise by a value equal to $1-s$. To aid interpretation, the decision limits (and multiples thereof) may be drawn onto the graphs as horizontal boundary lines. When α and β are equal, h_0 and h_1 are of the same magnitude. Crossing the lower decision limit (h_0) from above means that the true failure rate does not differ significantly from the acceptable failure rate, with the probability of a type 2 error equal to β^4 (as occurs in Fig. 1 for doctor A after 39 attempts). This has been taken, in scientific literature, to show competency as defined by cusum. When the upper line (h_1) is crossed from below, the actual failure rate is greater than the unacceptable failure rate (as occurs in Fig. 1 for doctor B after 16 attempts). This shows a process that is out of control. From this position, competency (or the acceptable failure rate) can be achieved only by a falling cusum that crosses two adjacent boundary lines.¹⁷ When the plot is between the decision limits, no statistical inference can be made and performance remains uncertain.⁴

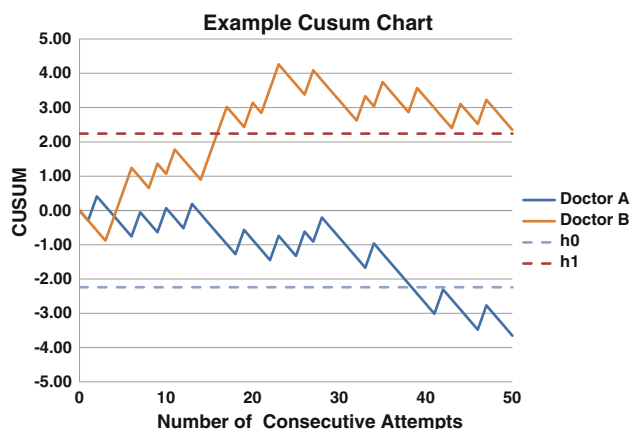


Fig. 1 This is a fictional example of a cusum chart for two doctors serially performing a procedure. Using the formula in Appendix, the decision limits h_0 and h_1 were calculated, along with s (the amount the cusum falls with each success). The unacceptable failure rate (f_u) was set as 0.4 (40%) and the acceptable failure rate (f_a) as 0.2 (20%). A 10% chance for a type 1 or type 2 error was tolerated and therefore alpha and beta were both 0.1. From this h_0 was calculated as -2.240 , and h_1 as 2.240 . These decision limits have been drawn on the graph as horizontal lines. The value for s was 0.293 , so for each success the cusum falls by 0.293 , and each failure rises by 0.707 (s^{-1}). The graph shows doctor A's cusum crossing h_0 after 39 procedures, indicating cusum defined competency with a 10% risk of a type 2 error. Doctor B's cusum crosses h_1 after 18 attempts demonstrating an unacceptably high failure rate

Cusum charts have been used in a variety of specialties (including endoscopy, orthopedics, surgery, and anesthetics) as a quality control method for experienced clinicians and to examine trainee learning curves.¹⁸ Their use, though, is currently limited to research despite having the potential to be a useful tool for providing continuous performance data, for evidence of achieving competency, and potentially for assessing training programs themselves. The aim of this review is to evaluate the available literature on the current use of cusum in anesthetic training with a view to establishing its role.

Literature review

A literature search of MEDLINE[®] (1950 to present), EMBASE[™] (1980 to present), BNI (1985 to present), and CINAHL[®] (1981 to present) was conducted using the key terms: anesthesiology/ed (ed = Education) OR an*esthes* (all fields); cusum (all fields) OR 'Cumulative Sum*' (all fields) OR learning curves (all fields). The Cochrane Library, NHS Evidence, and the Trip database were also reviewed. The last electronic search was performed in October 2012. All papers using cusum to investigate performance in anesthetic procedural skills were included. The search was limited to studies reported in English. Review articles, commentaries, abstracts, and letters were excluded. Thirteen relevant studies were identified and are shown in Table 1.

All 13 studies had small sample sizes (< 30), with most researching novice trainees' performance. The procedural skills they investigated could be split broadly into three groups: regional anesthesia, airway and cannulation, and ultrasound skills. These are dealt with in turn.

Regional anesthesia

Four studies examined cusum charts for epidural insertion.^{4,16,17,19} Naik¹⁷ had a cohort of 11 novices, ten of whom achieved cusum defined competency (i.e., crossing the lower boundary line from above) with between 1-85 attempts. In contrast, Kestin¹⁶ found only 4/12 recruits were competent, needing between 29-128 attempts, with five trainees having an unacceptable failure rate. De Oliveira Filho⁴ had similar results to those of Kestin,¹⁶ with 4/11 trainees achieving the acceptable failure rate.

One reason for the variation in the numbers of trainees reaching competency is illustrated in the paper by Sivaprakasam.¹⁹ They adjusted the acceptable and unacceptable failure rates from 10% and 15% to 20% and 30%, respectively. By doing so, the number of trainees

Table 1 Summary of papers

First Study Author	Study Population	Sample Size	Design Type	Study	Outcomes
de Oliveira Filho ⁴	Residents year 1	11	Prospective observational	<p>CUSUM graphs were drawn for:</p> <ul style="list-style-type: none"> • Peripheral venous cannulation • Orotracheal intubation • Epidural Anesthesia • Spinal Anesthesia <p>Acceptable failure rates for cannulation and intubation arbitrarily set at 20%.</p> <p>The acceptable failure rate for epidurals was 20% and for spinal was 15%. (These were set following data collection of staff anesthesiologists noting their success and failure rates.)</p>	<p><u>Venous Cannulation:</u></p> <p>Mean (SD) number per resident was 168.42 (67.37). All crossed 20% acceptable failure rate line after 56.85 (43.77) procedures (range 19-146).</p> <p><u>Orotracheal intubation:</u></p> <p>Mean (SD) number of procedures per resident was 127 (46.29) (range 50 -190) Only 4 residents crossed the 20% acceptable failure rate line.</p> <p><u>Spinal anesthesia:</u></p> <p>7/11 were below the 15% acceptable line 36 (20.16) (range 13-68).</p> <p>2/11 were above the 30% unacceptable failure rate line after 48 and 66 procedures.</p> <p><u>Epidural anesthesia:</u></p> <p>4/11 were below the acceptable 20% line 21.4 (11.17) procedures (range 9-36 procedures). 30% of subjects achieved competency in Experiment 1 and 11% in Experiment 2.</p>
de Oliveira Filho ¹⁴	Ultrasound novices	30	Prospective observational	<p>Investigated CUSUM curves for acquiring the skill to:</p> <p>Experiment 1: align a needle with an ultrasound beam</p> <p>Experiment 2: follow the needle with ultrasound to reach a target</p> <p>Each subject performed 25 trials.</p> <p>Acceptable and unacceptable failure rates were 5% and 20%, respectively.</p>	<p>Median number of attempts to achieve competency in the technique was 57 (but it could take up to 75 to ensure proficiency).</p> <p>10/11 residents reached competency in a range of 1-85 attempts. One did not reach competency as defined by CUSUM.</p> <p>Two failed to complete the study.</p>
Naik ¹⁷	Residents year 1	13 11 completed study	Prospective observational	<p>All residents had training in insertion of labour epidurals before the start of the study. They then charted their attempts over 6-mth period.</p> <p>Acceptable failure rates were set at 10 %</p> <p>CUSUM graphs drawn</p>	<p>Median number of attempts to achieve competency in the technique was 57 (but it could take up to 75 to ensure proficiency).</p> <p>10/11 residents reached competency in a range of 1-85 attempts. One did not reach competency as defined by CUSUM.</p> <p>Two failed to complete the study.</p>

Table 1 continued

First Study Author	Study Population	Sample Size	Design Type	Study	Outcomes
Kestin ¹⁶	Anesthetic trainees	12	Prospective Observational	<p>CUSUM graphs drawn for:</p> <ul style="list-style-type: none"> • Epidural anesthesia • Spinal anesthesia • Central Venous cannulation • Arterial line insertion <p>Unacceptable and acceptable failure rates were reached by consultant consensus:</p> <ul style="list-style-type: none"> • Epidural 5% acceptable 10% unacceptable • Spinal 10% acceptable 20% unacceptable • Central Venous cannulation 5% acceptable 15% unacceptable • Arterial cannulation 20% acceptable 40% unacceptable 	<p><u>Epidural anesthesia:</u></p> <p>4/12 reached acceptable rates (range 29-185 attempts).</p> <p>3/12 had unacceptable failure rates.</p> <p>5/12 had inconclusive records.</p> <p><u>Spinal anesthesia:</u></p> <p>2/8 reached acceptable rates (range 39-67 attempts.)</p> <p>2/8 had unacceptable rates.</p> <p>4/8 had inconclusive records.</p> <p><u>Central Venous Cannulation:</u></p> <p>Only 2 records of central line insertion during the whole study; results are inconclusive.</p> <p><u>Arterial cannulation:</u></p> <p>One trainee reached 20% acceptable rates after 14 arterial line insertions.</p> <p>4 had inconclusive records.</p>
Sivaprakasam ¹⁹	ST3 /4	6	Prospective observational	<p>CUSUM graphs drawn by trainees for first 50 epidurals. CUSUM graphs drawn from the results using acceptable and unacceptable rates as 10-20% and 15-30%, respectively.</p>	<p>If 10-20% used, then trainees required to insert a median number of 25 epidurals to be competent. (range 25-32 attempts) 4 competent.</p> <p>If 15-30% used, then trainees required to insert a median number of 18 epidurals to become competent. (range 14-33 attempts.) 5 competent.</p>
Schuepfer ²³	Anesthetic department	7	Prospective observational	<p>Psoas compartment block performed on 100 consecutive children using a nerve stimulator.</p> <p>An institution CUSUM graph was drawn using an acceptable failure rate as 10%.</p>	<p>CUSUM analysis showed clear insitutional learning. A learning curve was generated, and a success rate of 70% in a case load of 55 attempts is required.</p>
Komatsu ²⁴	Interns	15	Prospective observational	<p>Standard and risk-adjusted CUSUM charts were generated for 695 bag-mask ventilations (BMV) and 679 orointubations performed by the subjects.</p> <p>Acceptable and unacceptable failure rates were 20% and 40%, respectively.</p>	<p>BMV - 14/15 crossed the 20% acceptable failure rate in mean (SD) 27 (13) attempts. No intern would have succeeded if a 10% failure rate was set.</p> <p>Orointubation - 9/15 achieved the defined competency level after 26 (8) procedures.</p> <p>Mean number of attempts 47 (range 32-64).</p> <p>CUSUM graphs:</p>
Dalal ²⁵	Residents year 1	16	Prospective observational	<p>Simulated nasal fiberoptic endoscopy performed.</p> <p>Proficiency plotted on CUSUM graph; acceptable rate was 10% and unacceptable rate set at 20%</p>	<p>7 were proficient, 3 downward (improving) trends and 6 upward (worsening) trends. Attempts to reach proficiency were median 42 (range 27-58).</p>

Table 1 continued

First Study Author	Study Population	Sample Size	Design Type	Study	Outcomes
Correa ²⁶	Experienced anesthetists	4	Prospective observational	Evaluating the learning curves for intubation when using the Truview EVO2 laryngoscope on a mannequin. 45 sec were allowed for each attempt.	All achieved a failure rate of < 5%, requiring 42-97 attempts.
Margarido ²⁷	Spinal ultrasound novices	18	Prospective observational	Spinal ultrasound was performed on a live model up to 20 times, with the participants' ability to identify correctly the intervertebral spaces, the optimal needle insertion point, and the distance to the ligamentum flavum being assessed. Acceptable and unacceptable failure rates were 20% and 40%, respectively.	Correctly identifying the interspace: 5/18 reached competency with a median of 11 attempts. Identifying the optimal insertion point: 0/18 reached competency. Distance to ligamentum flavum: 0/18 reached competency.
Barrington ²⁸	Novices to US-guided regional anesthesia	15	Prospective observational	Each trainee performed 30 US-guided sciatic blocks in cadavers. Post procedure, a supervisor provided feedback on success / failure and CUSUM curves drawn. Trainees categorized into proficient/ not proficient and undetermined, and logistical regression performed. Mathematical modelling used to predict no. of attempts to reach competency in US visualization of the needle.	Mean trials to reach competency in US needle visualization = 28 Non-proficient group and undetermined groups showed a significant increase in the likelihood of success with each attempt ($P = 0.0243$ and $P = 0.024$, respectively) vs proficient group ($P = 0.076$).
Niazi ²⁹	Novices to US-guided regional anesthesia	20	Unblinded randomized control trial	Trainees were randomized into 2 groups; both received conventional teaching in peripheral nerve blockade using ultrasound, with the intervention group receiving, in addition, an hour long simulation session on needling and proper hand-eye coordination. Both groups then performed peripheral nerve blocks on patients and recorded their successes and failures over a 3-wk period on CUSUM charts	The conventional training group had 98 successful blocks vs simulation group with 144 (51.3% vs 64%, respectively; $P = 0.016$). Proficiency achieved in 40% (4/10) of the conventional training group and 80% (8/10) in the simulation training group ($P = 0.0849$).
Halpen ³⁰	Experienced anesthetists	2	Prospective observational	Two experienced anesthetologists performed ultrasound scans of patients undergoing lumbar CT scans. Markers used to determine which spine level was being scanned. The levels were determined by radiologists post scan, and CUSUM graphs constructed to determine the number of scans to reach success 90% of the time	The two anesthetologists took 36 and 22 scans to reach the reliability criteria.

reaching competency increased from 4/6 to 5/6, thus showing the importance of the initial variables used to construct the cusum graph. The methods used for deciding the failure rates varied widely between the papers: Sivaprakasam's team¹⁹ arbitrarily set the rates; de Oliveira Filho⁴ used rates from a control sample of trained anesthetists; and Kestin¹⁶ and Naik¹⁷ employed departmental consensus to decide. Table 2 shows the differences in set failure rates for the four studies, along with the numbers of participants reaching competency. Lowering the chosen failure rates means more successes and, therefore, attempts are required to reach competency; thus, in order to produce valid results, it is essential that these values are set appropriately for the level of the individual's training.

The definition of success and failure also varied between the studies and was another reason for the differing number of trainees attaining competency. In Naik's trial, any degree of pain relief from the epidural signified a success,¹⁷ but both Kestin¹⁶ and de Oliveira Filho⁴ had stricter criteria. Kestin required satisfactory analgesia/anesthesia, and de Oliveira Filho required technical success at the first interspace chosen and adequate surgical anesthesia. It is, however, well documented that correct anatomical placement of an epidural catheter does not always provide adequate or indeed any analgesia,^{20,21} and therefore, it could be argued that complete analgesia/surgical anesthesia may be too rigid an endpoint by which to judge the technical ability of a trainee inserting an epidural (although obviously it is the ideal outcome for the patient). This highlights the point that appropriate and consistent definitions of success and failure, that are clearly defined and unambiguous, must be used to avoid confusion and ensure meaningful cusum results.

Three studies examined spinal anesthesia. The studies by Kestin¹⁶ and de Oliveira Filho⁴ had similar definitions for success, namely, adequate surgical anesthesia, but the former used less lenient acceptable and unacceptable failure rates in the statistical analysis (10% and 20% vs 15% and 30%, respectively). This accounts for the fact that

64% (7/11) of de Oliveira Filho's trainees were deemed competent vs 25% (2/8) of Kestin's trainees. Again, this stresses the importance of the figures used in the cusum calculations.

One randomized controlled study investigated whether there was a difference in the learning curves of trainees when using two different spinal needles (25G and 27G),²² and results showed no significant difference. Using cusum to evaluate the effect of different equipment on the acquisition of technical skills is novel to anesthetic practice. Given its origins as a quality control tool, it could be employed effectively to assess both this and the impact of changes in equipment on the performance of experts. For example, if an expert practitioner plotted their cusum chart for a particular procedure (e.g., epidural insertions), their performance would be expected to be in steady state, i.e., tracking between the decision limits h_0 and h_1 . If the equipment used (needles/syringes etc.) then changed, the cusum chart would identify any impact this would have on performance. A plot remaining within the decision limits would indicate no significant effect, but if either limit were crossed, it would signify that a statistically significant change in performance had occurred, either an improvement if h_0 were breached or a deterioration if h_1 were crossed (Fig. 2). In a similar manner, other interventions, e.g., different teaching methods, such as simulation-based procedural training, could be investigated by assessing their effect on the cusum plot.

Schuepfer²³ investigated a new technique for performing psoas compartment blocks (PCB) in children using cusum. It was calculated that at least 55 blocks would need to be performed to achieve a success rate of 70% by looking at the learning curves of residents practicing the procedure. Schuepfer concluded that, with a strict definition for success, > 100 PCBs may need to be attempted. This has significant implications: If cusum is used to define competency and average procedure-specific learning curves are known, institutions and training rotations could then be evaluated to determine whether they are likely to provide trainees with enough opportunities to achieve competency. This is an interesting area, and one that demands further research.

Table 2 Summary of regional anesthesia papers

Author	Acceptable and unacceptable failure rates for epidurals	Range of attempts to become competent	Number reaching competency	Acceptable and unacceptable failure rates for spinals	Range of attempts to become competent	Number reaching competency
Sivaprakasam ¹⁹	10-20% initially and then 15-30%	25-32 (10-20% success rate) 14-33 (15-30% success rate)	4/6 5/6	Not studied	Not studied	Not studied
Naik ¹⁷	10-15%	1-85	10/11	Not studied	Not studied	Not studied
de Oliveira Filho ⁴	20-40%	9-36	4/11	15-30%	13-68	7/11
Kestin ¹⁶	5-10%	29-185	4/12	10-20%	39-67	2/8

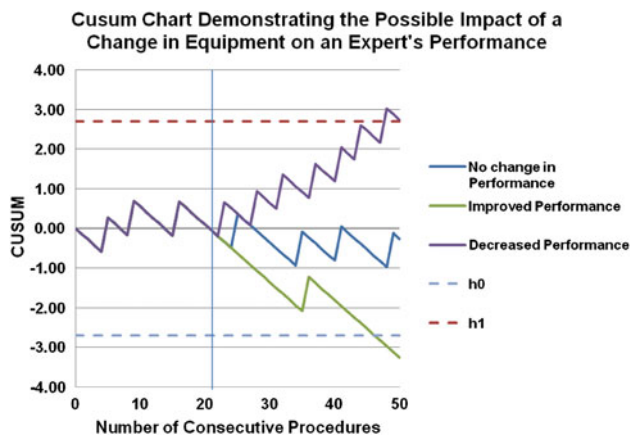


Fig. 2 This is an example of a cusum chart showing the possible effects a change in equipment could have on an expert’s performance. The first 22 procedures were carried out prior to the change and show a steady state performance graph. The plot then splits into the possible outcomes after different equipment has been introduced: either no change in performance (blue line) or a statistically significant improvement in performance (green line crossing h0); or conversely, a statistically significant decrease in performance (purple line crossing h1)

Airway and cannulation skills

The findings of studies plotting cusum curves for basic airway and cannulation skills are summarized in Table 3. Kestin’s¹⁶ study had few recruits for arterial and central line insertion (five and two, respectively) because most of the trainees had prior experience and were therefore excluded. Also, the trainees who were included performed only a small number of the procedures so interpretation of the results is difficult. It is clear, though, that using the cusum method requires novices to perform a large and very variable number of procedures before they are statistically proven to have an acceptable failure rate. This applies even to basic skills like cannulation, as the interns in de Oliveira Filho’s study⁴ required between 19-146 attempts to achieve competency, despite the fact they were allowed to miss 20% of the cannulas!

Komatsu’s group²⁴ performed an interesting additional analysis in their trial. Airway management was risk stratified by grading the likelihood of difficulties in bag-mask ventilation and tracheal intubation. This produced a risk-adjusted cusum. As a single failure has a significant impact on the cusum graph, a few atypically difficult patients and subsequent procedural failures can require learners to perform large numbers of procedures successfully in order to be anywhere near the lower boundary line of statistical significance. Risk-adjusting the cusum score would help account for this, and therefore, this approach is very appealing. Komatsu used this adjusted score to assess trainee performance as either better than expected, given the level of difficulty encountered, or worse than expected. The “expected” level of

Table 3 Summary of airway and cannulation papers

First study Author	Skill	Number of subjects	Competent/unproven/not competent	Range of attempts to become competent
de Oliveira Filho ⁴	Intravenous Cannulation	11	11/0/0	19-146
	Intubation	7	4/7/0	9-88
Kestin ¹⁶	Arterial Line	5	1/0/4	14
	Central Line	2	0/0/2	0
Komatsu ²⁴	Bag-Mask Ventilation	15	14/1/0	12-59
	Intubation	15	9/3/3	28-72
Dalal ²⁵	Fibreoptic upper airway endoscopy	16	7/3/6	27-58
Correa ²⁶	Intubation (specialized equipment trail)	4	4/0/0	42-97

performance was taken from the average performance of all of the interns. Ideally, this would have been derived from a larger external source of performance data.

Two studies investigated the learning curves of trainees performing airway procedures (upper airway endoscopy and orotracheal intubation with the Truview EVO2 laryngoscope) on mannequins.^{25,26} Again, a wide variation of attempts were needed to reach proficiency in both studies, showing that performance is very variable even in a controlled environment with the same training opportunities and teaching. This suggests that training should be individualized for each trainee and should ensure that extended practice of a procedure is possible (when required) to allow each trainee time to become competent.

Ultrasound skills

Five studies investigated ultrasound skills pertinent to anesthesia.^{5,27-30} In one study, cusum was used to determine the amount of training required to achieve competency in spinal ultrasound.²⁷ The conclusion was that 20 attempts and a coaching session were not sufficient to teach the relevant skills, and that this should inform the planning of future educational sessions and workshops. Unlike most other papers, no feedback was given during the attempts, which meant that all learning was experiential once the trial began. This was similar to the study by de Oliveira Filho⁵ that involved needling a phantom. After the trial, 6/26 subjects were deemed competent at following their needle using ultrasound, and only 2/18 were able to follow their needle to a target. The argument for the lack of feedback was that “most

individuals use a type of discovery learning when incorporating ultrasound guidance into their practice".^{31,32} Feedback was given in all the other studies when required and has been shown to be of great value in learning.³³ This might account for the low number of trainees achieving competency at these basic ultrasound skills in these studies.

Barrington *et al.*²⁸ used a bovine cadaver to assess the number of attempts 15 trainees required before they were able to visualize their needles competently on ultrasound during simulated sciatic nerve blocks. The trainers provided feedback after each attempt. The mean number of attempts to achieve competency was 28, but again, the range was wide.

Niazi *et al.*²⁹ used cusum to assess the effect of simulation on the acquisition of ultrasound skills in 20 novices by splitting them into two groups, one that was simulation trained and the other acting as a control. In the simulation group, 8/10 achieved proficiency compared with only 4/10 in the control group. This did not reach statistical significance, but the study was hampered by the low number of subjects. Nevertheless, this highlights another potential use for cusum, i.e., a tool to evaluate the effectiveness of different teaching methods in the development of procedural skills.

Halpern *et al.*³⁰ used cusum to prove that it was possible to learn to identify the lumbar spinous processes using ultrasound. Two experienced anesthesiologists performed an ultrasound scan of the lumbar spine and placed a radio-opaque marker at a designated level. The actual level was then determined by a radiologist after reviewing the patients' computed tomography scans. The results showed that skilled anesthesiologists required a minimum of 22 attempts to become reliable in defining lumbar spine anatomy with ultrasound, but that it was possible and could be used to improve the accuracy of needle placement during neuraxial techniques.

Use of cusum for quality control in anesthesia

Only one study³⁴ has implemented cusum analysis as a quality control system for anesthesiologists, which is surprising given that this was initially the reason for its creation. The study was performed by an experienced consultant who (bravely) published cusum charts for all arterial and central line insertions he performed over a three-year period. He concluded that it was a good practical performance monitor for consultants (and ideal for appraisals), but would need to be adapted to monitor trainees to reflect their level of experience.

Two other studies investigated the use of cusum with experienced doctors.^{35,36} They compared the cusum graphs of a registrar with those of a consultant performing non-

anesthetic medical procedures. The consultant's graphs rapidly produced a steady-state plot with acceptable failure rates, whereas the registrar's graphs were more haphazard and showed a significant learning curve.

Discussion

In anesthetic literature the use of cusum analysis has been limited almost totally to investigating the learning curves of specific procedures. From this type of work, an estimate of the number of procedures required to achieve competency can be made. This information could then be used to inform and evaluate training programs and help guide decisions about the most appropriate hospitals for trainees to rotate to depending on their educational requirements.

The ACGME requires that graduating residents perform a minimum of 50 spinal and 50 epidural techniques for surgical procedures.³⁷ Looking at the published cusum studies, this number may be sufficient for some trainees to acquire competency but certainly not for all. It is difficult, however, to provide an accurate estimate of the actual number needed from the available literature. This is because the existing studies have significantly different results due to the varying definitions for procedural success and failure, the differences in the variables used to construct the cusum graphs (e.g., acceptable and unacceptable failure rates), and also their small sample sizes. It is clear, though, that there is a wide spectrum of learning curves and consequently, the only way to guarantee competency is to tailor training to the individual rather than to focus on minimum numbers.

The accepted meaning of cusum-defined competency in the literature is crossing the h_0 boundary line from above or crossing any two consecutive boundary lines from above.¹⁷ The problem is that the latter criterion may well demand a significantly larger number of successes than the former, as the distance required to travel down the cusum chart is much greater. Indeed, at certain points on the chart, the number of successes required to achieve the acceptable failure rate is almost double that needed when compared with starting at the zero point. This means that novices who have several initial failures (which is to be expected when learning a new skill) will potentially end up at a great disadvantage when trying to prove their competency. Therefore, it would be more appropriate to reset the cusum to zero each time the upper boundary is breached. This approach has also been suggested when the lower boundary line is breached as if the cusum is allowed to continue to fall, a run with an unacceptably high failure rate may go unnoticed.³⁸ In fact, reaching a steady state on the graph may be enough assurance to conclude that the learning curve has settled down.¹⁸

There are several problems with using cusum analysis to assess performance in procedural skills. First, there are no nationally agreed definitions for success or failure at any given procedure, and those used in the literature vary greatly. Also, there is currently no consensus of opinion as to where the acceptable and unacceptable boundaries should be set or to what degree alpha and beta errors should be tolerated. Tight boundaries are important for quality control and for assessing trained individuals, but should these boundaries be much wider for the novice trainee to allow for their learning curve and to provide encouragement and a sense of achievement? The number of competent doctors produced can increase dramatically simply by altering the boundaries.¹⁹ Therefore, if procedural competency is to be defined by cusum, it would be necessary to establish national rates, and these would need to be tailored to the experience of the trainee.

Second, ensuring the accuracy of the recorded data is problematic. Cusum often relies on self-reporting, which introduces a subjective element to the interpretation of a procedural outcome. There is also the potential for recording bias, where favourable results are documented more frequently than unfavourable ones.⁵ If competency is to be defined by cusum, then the consequences of repeated failures are significant for trainees. This will increase the pressure on them to perform, and therefore, potentially give a positive skew to their procedural outcomes.

Third, as trainee seniority increases, so does their exposure to more difficult procedures. This could result in a deterioration in the cusum curve, as failures are more likely because of increasing procedural difficulty despite no change in skill level.⁵ As described previously, Komatsu²⁴ risk adjusted the cusum score for airway management, showing that this can be achieved successfully. It is however a single study, with a small sample population which has not been validated. Therefore, a universally recognized and accepted method is required to stratify the technical difficulty of different procedures.

Finally, cusum graphs can be difficult to construct and interpret. A recent review article suggested that only 17 of the 31 cusum graphs analyzed were drawn correctly.¹⁸ If these problems were overcome, cusum would have a valuable role in assessing trainee procedural performance.

Cusum is a good performance monitor for trained individuals and is a valuable quality control tool that could be used for revalidation and appraisal. It could be employed for rapid detection of medical errors, near misses, and suboptimal clinical performance and to monitor the effects of prolonged periods of time off work. For example, Kestin¹⁸ identified a registrar whose performance at spinal anesthesia fell significantly after an 18-mth period of non-anesthetic medical training. With the introduction of increasingly complex procedures and technologies, it may also be more

sensitive in assessing health care providers' skill than the current available methods.¹⁸ Finally, it could help assess the impact of new equipment on performance and therefore advise on procurement of medical supplies.

In summary, cusum has many potential applications in anesthesia. In its current form, it could be adopted readily to monitor performance in trained individuals. Also, it can produce an objective graph of performance in newly learned techniques, providing trainers with information that is unattainable from logbooks or WBAs. This allows trainees to assess their progress and consequently self-direct their learning, and gives trainers the opportunity to review a trainee's current skills on first contact. Poor performance can be readily identified and rapidly remediated, thus providing high-quality health care.³⁹ There are, however, several hurdles to overcome before cusum can be used reliably as proof of trainee competency. Further work in this area should focus on assessing the failure rates of expert anesthesiologists for individual procedures so informed decisions can be made about the acceptable and unacceptable trainee failure rates. Setting such standards nationally would aid the move towards competency-based residency training and act as a benchmark for future research. This should include investigating ways to adjust cusum scores for predictably difficult procedures, e.g., epidurals in morbidly obese patients, and performing validation studies.

Competing interests Neither author has any disclosures or competing interests.

Appendix: Construct of cusum graphs

The variables required to construct a cusum chart are the acceptable (f_0) and unacceptable (f_1) failure rates and the chosen probabilities for type I and II errors (α and β). From these numbers, decision limits (or boundary lines) h_0 and h_1 and a value for the cusum (s) are calculated using the formula below:

$$h_0 = \frac{b}{P + Q}$$

$$h_1 = \frac{a}{P + Q}$$

Where:

$$P = \ln\left(\frac{f_1}{f_0}\right), \quad Q = \ln\left(\frac{1-f_0}{1-f_1}\right), \quad a = \frac{\ln(1-\beta)}{\alpha},$$

$$b = \frac{\ln(1-\alpha)}{\beta}$$

$$s = Q/P + Q$$

The cusum graphical trend is described as follows:

$S_n = \sum (X_n - f_0)$ where $X_n = 0$ for success and 1 for failure, n is the number attempted, and $f_0 =$ the acceptable failure rate.

The cusum graph is plotted with the cusum value on the y-axis and the number of consecutive attempts on the x-axis. The graphs start at zero, and the cusum falls by s with each success and increases by $1-s$ with each failure. The decision limits h_0 and h_1 are drawn onto the chart as horizontal lines to aid interpretation.

The type I and II errors are frequently set as 0.1. By making these errors equal $h_0 = h_1$ and subsequent boundary lines are multiples of h_0 . A type I error is the risk of declaring competence when it is not achieved, and a type II error is the risk of not declaring competence when it is achieved. If the line crosses the upper decision limit (h_1) from below, then the failure rate is significantly greater than the acceptable failure rate. If a line crosses the lower decision limit (h_0) from above, then the true failure rate does not differ significantly from the acceptable failure rate with the probability of a type 2 error equal to β .⁴ If the cusum remains between two boundary lines, the results are indeterminate and the null hypothesis cannot be accepted or rejected and more observations are required. The null hypothesis is that the true failure rate is not different from the acceptable failure rate.¹⁶ Competency is achieved when the graphical trend falls below 2 adjacent calculated boundary lines ($h_0, 2h_0, 3h_0, 4h_0, \dots$). Competency is lost if the graph ascends and crosses two calculated boundary lines.^{4,17}

References

1. Cooper GM, McClure JH. Anesthesia chapter from saving mothers lives; reviewing maternal deaths to make pregnancy safer. *Br J Anaesth* 2008; 100: 17-22.
2. Swing SR. The ACGME outcome project: retrospective and prospective. *Med Teach* 2007; 29: 648-54.
3. Bould MD, Crabtree NA. Are logbooks of training in anaesthesia a valuable exercise? *Br J Hosp Med (Lond)* 2008; 69: 236.
4. de Oliveira Filho GR. The construction of learning curves for basic skills in anesthetic procedures: an application for the cumulative sum method. *Anesth Analg* 2002; 95: 411-6.
5. Bould MD, Crabtree NA, Haik VN. Assessment of procedural skills in anaesthesia. *Br J Anaesth* 2009; 103: 472-83.
6. Reznick RK, MacRae H. Teaching surgical skills - changes in the wind. *N Engl J Med* 2006; 335: 2664-9.
7. Martin JA, Regehr G, Reznick R, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg* 1997; 84: 273-8.
8. Reznick R, Regehr G, MacRae H, Martin J, McCulloch W. Testing technical skill via an innovative "bench station" examination. *Am J Surg* 1997; 173: 226-30.
9. Fried GM, Feldman LS, Vassiliou MC, et al. Proving the value of simulation in laparoscopic surgery. *Ann Surg* 2004; 240: 518-28.
10. Taffinder N, Sutton C, Fishwick RJ, McManus IC, Darzi A. Validation of virtual reality to teach and assess psychomotor skills in laparoscopic surgery: results from randomised controlled studies using the MIST VR laparoscopic simulator. *Stud Health Technol Inform* 1998; 50: 124-30.
11. Datta V, Mackay S, Mandalia M, Darzi A. The use of electromagnetic motion tracking analysis to objectively measure open surgical skill in the laboratory-based model. *J Am Coll Surg* 2001; 193: 479-85.
12. Darzi A, Mackay S. Assessment of surgical competence. *Qual Health Care* 2001; 10: Suppl 2: ii64-9.
13. Wass V, Van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. *Lancet* 2001; 357: 945-9.
14. de Oliveira Filho GR, Helayel PE, da Conceicao DB, Garzel IS, Pavei P, Ceccon MS. Learning curves and mathematical models for interventional ultrasound basic skills. *Anesth Analg* 2008; 106: 568-73.
15. Bolsin S, Colson M. The use of the cusum technique in the assessment of trainee competence in new procedures. *Int J Qual Health Care* 2000; 12: 433-8.
16. Kestin IG. A statistical approach to measuring the competence of anaesthetic trainees at practical procedures. *Br J Anaesth* 1995; 75: 805-9.
17. Naik VN, Devito I, Halpern SH. Cusum analysis is a useful tool to assess resident proficiency at insertion of labour epidurals. *Can J Anesth* 2003; 50: 694-8.
18. Biau DJ, Resche-Rigon M, Godiris-Petit G, Nizard RS, Porcher R. Quality control of surgical and interventional procedures: a review of the CUSUM. *Qual Saf Health Care* 2007; 16: 203-7.
19. Sivaprakasam J, Purva M. CUSUM analysis to assess competence: what failure rate is acceptable? *Clin Teach* 2010; 7: 257-61.
20. Hermanides J, Hallmann MW, Stevens MF, Lirk P. Failed epidural: causes and management. *Br J Anaesth* 2012; 109: 144-54.
21. Arendt K, Segal S. Why epidurals do not always work. *Rev Obstet Gynecol* 2008; 1: 49-55.
22. Charuluxananan S, Kyokong O, Preamsamran P. Comparison of 25 and 27 gauge needle in spinal anesthesia learning curve for anesthesia residency training. *J Med Assoc Thai* 2005; 88: 1569-73.
23. Schuepfer G, Johr M. Psoas compartment block (PCB) in children: Part II - generation of an institutional learning curve with a new technique. *Paediatr Anaesth* 2005; 15: 465-9.
24. Komatsu R, Kasuya Y, Yogo H, et al. Learning curves for bag-and-mask ventilation and orotracheal intubation: an application of the cumulative sum method. *Anesthesiology* 2010; 112: 1525-31.
25. Dalal PG, Dalal GB, Pott L, Bezinover D, Prozesky J, Bosseau Murray W. Learning curves of novice anesthesiology residents performing simulated fiberoptic upper airway endoscopy. *Can J Anesth* 2011; 58: 802-9.
26. Correa JB, Dellazzana JE, Sturm A, Leite DM, de Oliveira Filho GR, Xavier RG. Using the cusum curve to evaluate the training of orotracheal intubation with the Truview EVO2 laryngoscope (Portuguese). *Rev Bras Anesthesiol* 2009; 59: 321-31.
27. Margarido CB, Arzola C, Balki M, Carvalho JC. Anesthesiologists' learning curves for ultrasound assessment of the lumbar spine. *Can J Anesth* 2010; 57: 120-6.
28. Barrington MJ, Wong DM, Slater B, Ivanusic JJ, Owens M. Ultrasound-guided regional anesthesia: how much practice do novices require before achieving competency in ultrasound needle visualization using a cadaver model. *Reg Anesth Pain Med* 2012; 37: 334-9.
29. Niazi AU, Haldipur N, Prasad AG, Chan VW. Ultrasound-guided regional anesthesia performance in the early learning period: effect of simulation training. *Reg Anesth Pain Med* 2012; 37: 51-4.

30. Halpern SH, Banerjee A, Stocche R, Glanc P. The use of ultrasound for lumbar spinous process identification: a pilot study. *Can J Anesth* 2010; 57: 817-22.
31. Vereijken B, Whiting HT. In defence of discovery learning. *Can J Sport Sci* 1990; 15: 99-106.
32. Tsui BC. "Credentials" in ultrasound-guided regional blocks. *Reg Anesth Pain Med* 2006; 31: 587-8.
33. Issenberg SB, McGaghie WC, Petrusa ER, Lee GD, Scalese RJ. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Med Teach* 2005; 27: 10-28.
34. Runcie CJ. Assessing the performance of a consultant anaesthetist by control chart methodology. *Anaesthesia* 2009; 64: 293-6.
35. Lim TO, Soraya A, Ding LM, Morad Z. Assessing doctors' competence: application of CUSUM technique in monitoring doctors' performance. *Int J Qual Health Care* 2002; 14: 251-8.
36. Williams SM, Parry BR, Schlup MM. Quality control: an application of the cusum. *BMJ* 1992; 304: 1359-61.
37. Panni MK, Camann WR, Tsen LC. Resident training in obstetric anesthesia in the United States. *Int J Obstet Anesth* 2006; 15: 284-9.
38. Grigg OA, Farewell VT, Spiegelhalter DJ. Use of risk-adjusted CUSUM and RSPRT charts for monitoring in medical contexts. *Stat Methods Med Res* 2003; 12: 147-70.
39. Lanigan CB. Cusum scoring theory and practice. *Bulletin of Royal College of Anaesthetists* 2009: 16-9. Available from URL: <http://www.rcoa.ac.uk/document-store/bulletin-54-march-2009> (accessed September 2013).