ORIGINAL ARTICLE

Cumulative sum (Cusum) analysis provides an objective measure of competency during training in endoscopic retrograde cholangio-pancreatography (ERCP)

Hayley M. Waller & Saxon J. Connor

Department of Surgery, Christchurch Hospital, Christchurch, New Zealand

Abstract

Background and Aim: Endoscopic retrograde cholangio-pancreatography (ERCP) is a complex and technically challenging procedure with a risk of significant complications. Although recommendations exist as to how many ERCPs should be performed before a trainee is deemed competent they lack objective measures of outcome. Cumulative sum (Cusum) analysis allows continuous monitoring of a proceduralist's performance to ensure acceptable outcomes are being achieved. The aim of the present study was to assess a trainee's progress with ERCP using Cusum analysis.

Method: A single trainee performed one supervised ERCP list per week between June 2005 and May 2008. Prospective data were collected and analysed by intention to treat using Cumulative failure charting and Cusum analysis. The binary outcome measure was successful cannulation in patients with a native sphincter. Acceptable and unacceptable cannulation failure rates were set at 20% and 35% respectively, with Type I and II error rates set at 0.10.

Results: A total of 290 ERCPs were performed. Successful cannulation was achieved in 168 of 238 (70.6%) patients with intact biliary sphincters. Although cumulative failure charting suggested the trainee had not yet achieved satisfactory performance, Cusum analysis with sequential probability ratio testing indicated that an acceptable outcome had been achieved for the last 179 consecutive ERCPs.

Conclusion: Cusum analysis enables sensitive and continuous monitoring of a trainee's performance to objectively determine competency. Wider and systematic use may enable appropriate benchmarks to be identified and more objective assessment of a trainee's experience.

Correspondence

Saxon J Connor, Department of Surgery, Christchurch Hospital, Christchurch, New Zealand. Tel: 0064 3 3640640; Fax: 0064 3 3640352; E-mail: saxon.connor@cdhb.govt.nz

Introduction

Endoscopic retrograde cholangio-pancreatography (ERCP) is an important procedure both in the diagnosis and management of pancreaticobiliary diseases. It is a technically demanding procedure that carries with it an incidence of complications ranging from 4% to 30%.[1]

The American Society for Gastrointestinal Endoscopy has developed an ERCP core curriculum that outlines recommendations for ERCP training. The currently accepted benchmark for deep bile duct cannulation in individuals seeking credentialing for independent practise is >80% with some suggesting that 90% is a more appropriate standard.[2,3] In 1996, guidelines stated that fellows must complete at least 100 ERCPs before they are assessed for competence. These guidelines were later revised and it is now felt that most trainees need to complete at least 180 procedures, the majority of which are therapeutic, before they are able to achieve competency.[2] Other guidelines (Table 1) are in keeping or similar to the American Society for Gastroenterology guidelines including those developed for the British Society of Gastroenterology and the Gastroenterological Society of Australia.[4,5]

Table 1 Recommendations from national societies regarding training requirements for endoscopic retrograde cholangio-pancreatography (ERCP)

Variable	Data recorded
British Society of Gastroenterology: Joint Advisory Group on Gastrointestinal Endoscopy[4]	• No specific number of procedures prescribed
	• Emphasis on the number of procedures a centre performs with the recommended number being 250 annually (or fewer if trainees are assured adequate exposure)
	• At the time of assessment for credentialing the trainee must be:
	– aware of the indications, limitations and complications of diagnostic and therapeutic ERCP
	– able to identify the papilla in >95% of patients who have not undergone gastric surgery.
	– competent in those manoeuvres which facilitate cannulation of the biliary tree and the pancreatic duct
	– able to independently perform sphincterotomy, stone extraction and stenting and display a safe and potentially successful approach to these technical skills.
American Society for Gastrointestinal Endoscopy[2]	• Must perform at least 180 procedures
	• At least 50% of procedures should be therapeutic
Gastoenterological society of Australia: Conjoint Committee for the Recognition of Training in Gastrointestinal Endoscopy (CCRTGE)[5]	• Must perform a total of 200 unassisted ERCP examinations in patients with intact papillary sphincters
	• Must include a minimum of 80 supervised, independently performed, sphincterotomies in patients with intact papillary sphincters
	• A minimum of 60 stents

It is recognized that individuals training in any technical procedure acquire skills at differing rates and that the number of procedures completed is alone an inadequate marker of competency in a given procedure.[3]

Cumulative summation (Cusum) is a statistical technique of sequential analysis that allows an observer to judge whether the variation observed in performance is acceptable or if the variation was outside that expected from random variation. It is useful in settings in which the sample size is both unknown and ever increasing.[6] It has been used previously in the technical aspects of medicine for both assessing the competence of trainees and also in the assessment of doctors performing certain procedures who are already deemed competent in those procedures.[7–9]

The aim of this study was to assess a trainee's progress with ERCP using Cusum analysis.

Method

From the period June 2006 to May 2008, a single trainee maintained a prospective database of all ERCPs performed on a weekly supervised list. Data recorded are shown in Table 2.

The binary outcome measure was successful duct (pancreatic or biliary dependant on indication) cannulation in patients with an intact sphincter of Oddi as confirmed by cholangio-pancreatography. Data were analysed by intention to treat.

The current study met the definition of an audit and quality assurance related activity as per New Zealand national ethics committee guidelines. Therefore it did not require specific ethical committee review.[10]

Table 2 Data recorded for endoscopic retrograde cholangio-pancreatography (ERCP) recorded by a single trainee from June 2006 to May 2008

Variable	Data recorded
Date of ERCP/consecutive attempt	Date/sequential attempt number
Age	Years
Indication	
Previous sphincterotomy	Yes/No
Successful cannulation	Yes/No
If trainee cannulation failed, did supervisor succeed	Yes /No
Successful sphincterotomy	Not indicated/No/Yes
Procedure performed	Duct clearance/Stent placement/No procedure indicated
Outcome	Successful/Repeat ERCP required/ Other technique required

The trainee had previously completed training in upper gastrointestinal (GI) endoscopy (minimum 400 endoscopies) and had been briefly exposed to ERCP during their fellowship training (but did not achieve competency).

Statistical method

Cumulative failure charting and sequential probability ratio testing was performed. The number of cumulative failures was charted against the sequential attempt number for the series of patients. Control lines regarding acceptable or unacceptable per-

formance were determined by firstly defining certain parameters. These parameters were: the acceptable outcome rate (p_0), the unacceptable outcome rate (p_1), the Type I (false-positive) error rate (α) and the Type II (false-negative) error rate (β). Acceptable and unacceptable cannulation failure rates (determined by benchmarking against previously published data[1,2]) were set at 0.20 and 0.35, respectively. As standard for Cusum analysis, Type I and Type II error rates were set at 0.10.[6]

Intermediate values (a, b, P, Q and s) were calculated as follows[6]:

$$a = \ln[(1-\beta)/\alpha] \quad \text{i.e.} \quad a = \ln[0.9/0.1]; a = 2.197\,(3\,\text{sf})$$
$$b = \ln[(1-\alpha)/\beta] \quad\quad\quad b = \ln[0.9/0.1]; b = 2.197\,(3\,\text{sf})$$
$$P = \ln(p_1/p_0) \quad\quad\quad\quad P = \ln(0.35/0.20); P = 0.560\,(3\,\text{sf})$$
$$Q = \ln[(1-p_0)/(1-p_1)] \quad Q = \ln[(1-0.20)/(1-0.35);$$
$$Q = 0.208\,(3\,\text{sf})]$$
$$s = Q/(P+Q) \quad\quad\quad\quad s = 0.208/(0.560+0.208);$$
$$s = 0.271\,(3\,\text{sf}).$$

The acceptable and unacceptable cumulative failure (CF) rates were calculated, where n is the attempt number[6]

$$CF_{\text{acceptable}} = sn - b/(P+Q)$$
$$CF_{\text{unacceptable}} = sn - a/(P+Q).$$

These CF rates, or acceptable and unacceptable control lines, were also plotted on the chart. The crossing of either of these lines by the cumulative failure curve indicates acceptable or unacceptable performance in a series and hence conclusions can be drawn from the data.

The sequential probability ratio testing involved the plotting of the Cusum score against the number of attempts. With each procedure, the Cusum score either decreases by the value 's' (0.271) if the attempt is a success, or increases by the value '1 − s' (0.729) if the attempt is a failure.

Control lines are drawn horizontally on the Cusum plot with the line values being defined by 'h'. This value is determined by the acceptable and unacceptable failure rates and the Type I and II error rates. The spacing between the unacceptable control lines (h_0) and the acceptable control lines (h_1) is determined as follows[6]:

$$h_0 = b/(P+Q) \quad \text{i.e.} \quad h_0 = 2.197/(0.560+0.208); h_0 = 2.86\,(2\,\text{sf})$$
$$h_1 = a/(P+Q) \quad\quad\quad h_1 = 2.197/(0.560+0.208); h_1 = 2.86\,(2\,\text{sf}).$$

Because the Type I and II error rates are equal, then h_0 and h_1 become equal and only one set of control lines are plotted.[6] In terms of interpretation, each time the plotted line crosses a control line it can be concluded that the performance since the last control line was crossed was unacceptable if the line trends up [level of importance is deemed importantly divergent i.e. alternative hypothesis (H_1) can be accepted] or acceptable [practice is as expected i.e. null hypothesis (H_0) can be accepted] if the line is trending down.

In a steady-state situation (i.e. monitoring rather than learning phase), the sensitivity of the graph in detecting a detrimental change can be increased by resetting the graph whenever H_0 is accepted.[11] Conversely, in the situation where a learning curve

Table 3 Indications for endoscopic retrograde cholangio-pancreatography (ERCP) performed by a single trainee from June 2006 to May 2008

Indication (n = 290)	n (%)
Choledocholithasis	159 (54.8)
Malignancy	45 (15.5)
Acute pancreatitis	21 (7.2)
Bile leak post-cholecystectomy	7 (2.4)
Other	58 (20.0)

Table 4 Success rate of procedures performed using endoscopic retrograde cholangio-pancreatography by a single trainee from June 2006 to May 2008

Procedure	Total attempted	Successful attempts (%)
Sphincterotomy	186	162 (87.1)
Duct clearance	100	93 (93.0)
Stent placement	48	40

exists, the sensitivity for detecting improved performance can be increased by resetting the graph when H_1 has been accepted.

Continuous data are presented as median (range) whereas nominal data are presented as numerator with percentages (%) when the denominator is >50.

Results

During the study period, 290 ERCPs were performed. The median (range) age was 72 years (21–94). One hundred and twenty-nine patients (44.5%) were male. Indications for ERCP are shown in Table 3.

Of the 238 (82.1%) patients with an intact sphincter of Oddi, successful cannulation was achieved by the trainee in 168 (70.6%) patients. In the 70 (29.4%) patients in whom the trainee was unable to cannulate the appropriate duct, the supervisor was able to achieve cannulation in 25 (35.2%) of these patients. Fifty-two (17.9%) patients had undergone a sphincterotomy previously and the success rate of cannulation by the trainee in this group was 86.5% (45/52).

The success rate of procedures subsequently performed by the trainee is shown in Table 4.

The cumulative failure to cannulate an intact sphincter of Oddi by the trainee is shown in Fig. 1. As can be seen in Fig. 1, the cumulative failure rate crosses the upper control line (unacceptable failure rate) at ERCP attempt number 14 hence indicating unacceptable performance and remains so throughout the entire series ($n = 238$).

Figure 2 shows the Cusum score plotted against the ERCP attempt number for the series.

It can be seen that the unacceptable control line [crossing horizontal grid lines on the y-axis in an upward fashion (positive gradient)] is first crossed at attempt number 14 and again subse-
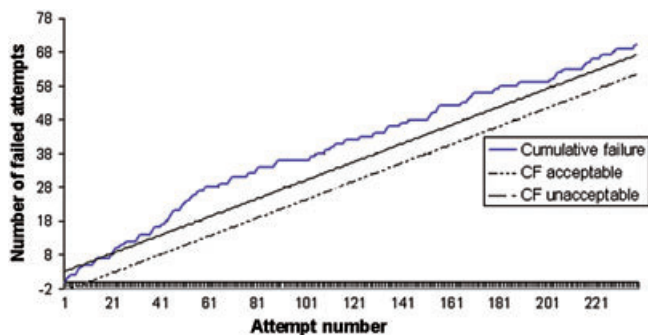
Figure 1 Cumulative failure (CF) rate of cannulation of an intact sphincter of Oddi at 238 sequential endoscopic retrograde cholangio-pancreatographies (ERCP) by a single trainee from June 2006 to May 2008. CF acceptable rate set at 0.20, CF unacceptable failure rate set at 0.35. Type 1 and II error rates set at 0.10
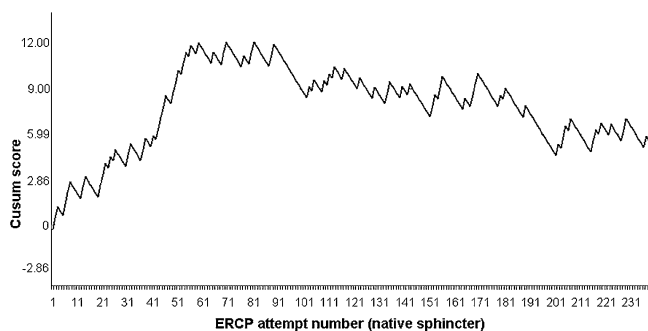

Figure 2 Cumulative sum (Cusum) score for successful cannulation using endoscopic retrograde cholangio-pancreatography (ERCP) on patients with an intact sphincter of Oddi ($n = 238$) by attempt number (ERCP) for a single trainee from June 2006 to May 2008. Unacceptable cannulation failure rate set at 0.35, acceptable cannulation failure rate set at 0.20. Type 1 & II error = 0.10. Each horizontal grid line on the y-axis represents a control line

Figure 3 Cumulative sum (Cusum score) (—) for successful cannulation using endoscopic retrograde cholangio-pancreatography (ERCP) on patients with an intact sphincter of Oddi ($n = 238$) by attempt number (ERCP) for a single trainee from June 2006 to May 2008 with the graph resetting if H0 or H1 is accepted. Unacceptable cannulation failure rate set at 0.35, acceptable cannulation failure rate set at 0.20. Type 1 & II error = 0.10. Each horizontal grid line on the y-axis represents a control line

downward fashion, rate of cannulation can be concluded to be as expected) and the graph resets.

Discussion

In this prospective analysis, the learning experience of a single ERCP trainee was analysed using Cusum analysis. The results of the current study have confirmed the usefulness of such techniques in objectively demonstrating the learning curve associated with the acquisition of new skills. Such analysis allows for a form of 'quality control' during the assessment period and objective conclusions can be drawn regarding whether a trainee's performance is acceptable or not. In addition ongoing analysis provides ongoing documentation of competent practice.

This study also highlights that although conclusions can be drawn from analysing cumulative failure charting, Cusum analysis using sequential probability ratio testing provides a more sensitive analysis of ongoing performance when there is rapid improvement in the level of skill (training situation). Within the current study, the cumulative failure chart (Fig. 1) indicated towards the latter part that the trainee was showing improvements as it trended towards the unacceptable performance line previously crossed at attempt number 14. If the trainee continued along this trend it would have taken over 250–300 attempts from this attempt number in order to cross the upper control line before any further conclusion could be drawn regarding the trainee's performance. In contrast, the sequential probability ratio test (Figs 2,3) could conclude that satisfactory performance had been achieved (at attempt number 100) for the period between ERCP attempt number 60–100. More detailed analysis of the sequential probability ratio test (Fig. 3) also enables continuous analysis of whether

quently at attempt numbers 43, 50 and 59. Between attempt numbers 60 and 99, the trainee's performance begins to improve (as indicated by levelling of curve and no further unacceptable control lines are crossed) and acceptable performance can first be concluded at attempt number 100 when the plotted line crosses the acceptable control line [crossing horizontal grid line from above to below (negative gradient)]. Performance continues to improve crossing a further control line at attempt number 196.

Figure 3 plots the Cusum score against the attempt number with the graph resetting each time H_0 or H_1 is accepted. It can be seen that at ERCP attempt numbers 14, 43 and 50 the graph has been reset when H_1 has been accepted (control line crossed by graph in upward fashion, rate of cannulation deemed significantly divergent from accepted practice). Satisfactory performance is demonstrated for the series between attempt numbers 51 to 196 when the H_0 is accepted (control line crossed by graph in
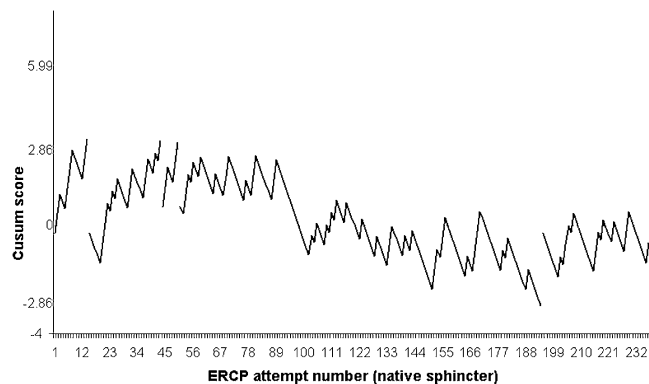
trainees are improving with the expectation that the interval (number of attempts) between the curves crossing the control bands should increase and eventually the slope of the line should change and become negative. This has the advantage of allowing individualization of the training programme to an individual's rate of learning. Equally with time, if a national database was developed, a learning curve with confidence intervals could be constructed which would allow for early identification of non-performing trainees. Such trainees could then be redirected into more appropriate areas of practice. A recent survey of UK trainees has highlighted deficiencies with prospective data collection, with only 21% of trainees maintaining such a database.[1]

This study highlights two important points that have been addressed previously, the first concerns the recognition that individuals follow different learning curves and second that number of procedures alone are an inadequate guideline for assessing a trainees competence and that more objective measures should be assessed.[1,3]

Although to some extent The American Society for Gastrointestinal Endoscopy guidelines suggest that competency is achieved in part by completing at least a set number of ERCPs, the guidelines fail to take into account individual trainees' rates of learning. Competency in ERCP is achieved when an individual can successfully cannulate the bile duct in ≥80% of attempts. It has been shown that the number of ERCP attempts required to achieve such competence is highly variable[2,3,12] and therefore the ability to individualize a training programme is crucial to optimize quality of practice and training opportunities for the maximum number of trainees. In the current study, the trainee had previously had brief intermittent exposure to ERCP but on commencement of the training programme was aware that they were not competent at ERCP. So although the number of attempts required to achieve competency would appear to be less than that previously reported,[2,3,12] this simply reinforces the robustness of Cusum. This is particularly important in the era where trainees are expected to be increasingly mobile and transfer between units during the training programme. Thus the ability to transfer previous experience in an objective fashion is increasingly important if redundancy in training is to be avoided.

Although this study only assessed cannulation as a binary outcome for Cusum analysis the results confirm this was appropriate as most of the other procedures (sphincterotomy, stent placement and duct clearance) can be completed with high levels of success with minimal exposure. It is clear from the current results that cannulation is the technically most demanding part of ERCP. Obviously complication rates are an important measure of quality; however, in the setting of training this is more difficult given the relative infrequent nature of serious adverse outcomes.

The obvious limitation of this study is the fact that it represents only one trainee's experience. Therefore it would be important for this to be replicated by large training centres or national training committees. This would be relatively simple to do with the availability of web-based databases.

In conclusion, Cusum analysis provides a simple but very effective and sensitive objective measure of competency for ERCP training. Further studies using larger numbers of trainees need to be performed to confirm the current studies findings.

Conflicts of interest
None declared.

References

1. Lombard M, Martin D, Riley S, Veitch P, Wilkinson M, Williamson PJ *et al.* (2007) Are we meeting the standards set for endoscopy? Results of a large-scale prospective survey of endoscopic retrograde cholangiopancreatography practice. *Gut* 56:821–829.

2. The ASGE Training Committee. (2006) ERCP core curriculum. *Gastrointest Endosc* 63:361–376.

3. Jowell PS, Baillie J, Branch S, Affronti J, Browning CL, Bute BP. (1996) Quantitative assessment of procedural competence: a prospective study of training in endoscopic retrograde cholangiopancreatography. *Ann Intern Med* 125:983–989.

4. British Society of Gastroenterology (2004) *Guidelines for the training, appraisal and assessment of trainees in gastrointestinal endoscopy and for the assessment of units for registration and re-registration 2004*. prepared by the Joint Advisory Group on Gastrointestinal Endoscopy. [homepage on the Internet]. c2004. Available at: http://www.bsg.org.uk/images/stones/docs/training/jag_recommendations_2004.pdf (accessed 2 June 2008).

5. Conjoint Committee for the Recognition of Training in Gastrointestinal Endoscopy (on behalf of the Royal Australasian College of Surgeons, the Gastroenterological Society of Australia and the Royal Australasian College of Physicians) (2008) *Requirements for CCRTGE recognition*. [homepage on the Internet]. Available at: http://conjoint.gesa.org.au/information.html (accessed 2 June 2008).

6. Yap C-H, Colson ME, Watters DA. (2007) Cumulative sum techniques for surgeons: a brief review. *ANZ J Surg* 77:583–586.

7. Grunkemeier GL, Wu YX, Furnary AP. (2003) Cumulative sum techniques for assessing surgical results. *Ann Thorac Surg* 76:663–667.

8. Kestin IG. (1995) A statistical approach to measuring the competence of anaesthetic trainees at practical procedures. *Br J Anaesth* 75:805–809.

9. Lim TO, Soraya A, Ding LM, Morad Z. (2002) Assessing doctors' competence: application of CUSUM technique in monitoring doctors' performance. *Int J Qual Health Care* 14:251–258.

10. Ministry of Health National Ethics Advisory Committee (2008) *Ethical guidelines for observational studies: Observational research, audits and related activities*. [homepage on the internet]. c2006. Available at: http://www.neac.govt.nz/neac/ (accessed 24 June 2008).

11. Spiegelhalter D, Grigg O, Kinsman R, Treasure T. (2003) Risk-adjusted sequential probability ratio tests: applications to Bristol, Shipman and adult cardiac surgery. *Int J Qual Health Care* 15:7–13.

12. Vitale GC, Zavaleta CM, Vitale DS, Binford JC, Tran TC, Larson GM. (2006) Training surgeons in endoscopic retrograde cholangiopancreatography. *Surg Endosc* 20:149–152.