Institutional and Individual Learning Curves for Focused Abdominal Ultrasound for Trauma

Cumulative Sum Analysis

Freda D. McCarter, MD, Fred A. Luchette, MD, Mark Molloy, MD, James M. Hurst, MD, Kenneth Davis, Jr., MD, Jay A. Johannigman, MD, Scott B. Frame, MD, and Josef E. Fischer, MD

From the Division of Trauma and Critical Care, Department of Surgery, University of Cincinnati College of Medicine, Cincinnati, Ohio

Objective

To evaluate both institutional and individual learning curves with focused abdominal ultrasound for trauma (FAST) by analyzing the incidence of diagnostic inaccuracies as a function of examiner experience for a group of trauma surgeons performing the study in the setting of an urban level I trauma center.

Summary Background Data

Trauma surgeons are routinely using FAST to evaluate patients with blunt trauma for hemoperitoneum. The volume of experience required for practicing trauma surgeons to be able to perform this examination with a reproducible level of accuracy has not been fully defined.

Methods

The authors reviewed prospectively gathered data for all patients undergoing FAST for blunt trauma during a 30-month period. All FAST interpretations were validated by at least one of four methods: computed tomography, diagnostic peritoneal lavage, celiotomy, or serial clinical evaluations. Cumulative sum (CUSUM) analysis was used to describe the learning curves for each individual surgeon at target accuracy rates of 85%, 90%, and 95% and for the institution as a whole at target examination accuracy rates of 85%, 90%, 95%, and 98%.

Results

Five trauma surgeons performed 546 FAST examinations during the study period. CUSUM analysis of the aggregate experience revealed that the examiners as a group exceeded 90% accuracy at the outset of clinical examination. The level of accuracy did not improve with either increased frequency of performance or total examination experience. The accuracy rates observed for each trauma surgeon ranged from 87% to 98%. The surgeon with the highest accuracy rate performed the fewest examinations. No practitioner demonstrated improved accuracy with increased experience.

Conclusions

Trauma surgeons who are newly trained in the use of FAST can achieve an overall accuracy rate of at least 90% from the outset of clinical experience with this modality. Interexaminer variations in accuracy rates, which are observed above this level of performance, are probably related more to issues surrounding patient selection and inherent limitations of the examination in certain populations than to practitioner errors in the performance or interpretation of the study.

Prompt identification of significant intraperitoneal hemorrhage in injured patients remains a diagnostic challenge. Until recently, diagnostic peritoneal lavage (DPL) and computed tomography (CT) have been considered the standard modalities for the evaluation of blunt abdominal trauma.[1–3] DPL is overly sensitive, with reported false-negative rates after laparotomy as high as 30%.[1] CT scanning provides precise information about the severity of solid organ injuries but it requires a relatively hemodynamically stable patient. Another diagnostic tool, ultrasonography, has been extensively used in Europe and Japan to identify hemoperitoneum.[4–6] This imaging modality allows rapid performance of focused abdominal sonography for trauma (FAST) in the

initial evaluation of the injured patient. It has been used with increasing frequency in American trauma centers during the past decade.

As surgeons begin to use a diagnostic modality that has traditionally been considered to be a radiologic examination, what constitutes appropriate training for satisfactory performance of the study by nonradiologist practitioners remains the source of considerable debate. Radiologists have stated that 3 months of formal training and 300 examinations are necessary to become proficient at ultrasonography.[7] Although such a volume of experience may be required to develop comprehensive ultrasound evaluation skills, this estimate does not take into account either the focused nature of FAST or the previous training and experience of trauma surgeons in correlating radiographic and clinical findings in injured patients. High levels of sensitivity and specificity have been reported for FAST examinations performed by surgeons with much more modest levels of experience with the technique.[8–10]

The extent of examiner training in ultrasonography has varied in most published studies. Tso et al[11] used a minimum training of only 1 hour of didactic instruction and 1 hour of practical experience. They prospectively studied 163 patients and reported a sensitivity of 91% and a specificity of 96%. Rozycki et al[12] developed a curriculum totaling 32 hours of instruction. The group of surgeons studied after completion of this course performed FAST examinations on 1,540 patients with a sensitivity of 83% and a specificity of 99%.

Several studies have sought to identify a learning curve for FAST. Ma et al[13] proposed that 50 examinations were necessary. Smith et al[14] recommended that only 25 examinations were required for competency in their study examining sensitivity and specificity in four 6-month increments. A recent study by Shackford et al[15] proposed that as few as 10 procedures are necessary to establish competency. All these studies define "competency" by different methods of analysis. Shackford et al used error rates, while others looked at a plateau of sensitivity and specificity.

Cumulative sum analysis (CUSUM) is used by engineers to monitor the performance of machinery and identify possible malfunctions. It provides a visual, objective demonstration of performance accuracy on a case-to-case basis. CUSUM has been used to evaluate a practitioner's initial and continued successful performance of procedures including colonoscopy,[16] endoscopic retrograde cholangiopancreatography,[17] the technical skills of surgical residents,[18] and completion of a cholangiogram during laparoscopic cholecystectomy.[19] After establishing a target success rate, practitioners can plot their consecutive successes and failures. Failure to maintain a predetermined success rate may indicate the need for additional training.

The purpose of this study was to estimate the amount of experience required for a trauma surgeon to be able to achieve an acceptable, reproducible level of accuracy performing FAST examinations. To accomplish this goal, CUSUM analysis was used to assess both institutional and examiner-specific accuracy rates with FAST as a function of experience performing the examination.

METHODS

This study was conducted with the approval of the Institutional Review Board at the University of Cincinnati College of Medicine. The trauma service began using FAST for the evaluation of hemoperitoneum in blunt trauma at our level 1 trauma center on Feb. 1, 1996. Interpretation of all FAST examinations performed by the trauma surgeons were recorded and prospectively gathered by the trauma registrar. All examinations were performed according to the techniques described by Rozycki et al.[12] Briefly, this included a pericardial, right upper quadrant, left upper quadrant, and pelvic view. Injury severity score (ISS),[20] abdominal abbreviated injury score,[21] revised trauma score,[22] and the presence of shock were used to determine selection bias among individual examiners. For purposes of this study, only FAST examinations performed or supervised by one of five attending trauma surgeons were evaluated. An abdominal CT scan, DPL, findings at celiotomy, or observation with serial physical examinations was used to validate the ultrasound readings for the presence or absence of hemoperitoneum. DPL was performed using a semiopen technique. The result was considered positive if there was aspiration of bile, enteric content, or more than 10 mL blood; drainage of lavage fluid by a chest tube or urinary catheter; a red blood cell count greater than $100,000/mm^3$; or a white cell count more than $500/mm^3$.[23]

Surgeon Training

All surgeons successfully completed the postgraduate course for FAST sponsored by the Eastern Association for the Surgery of Trauma before examining patients. Four of the five attendings participated in additional on-campus training in ultrasonography, which included 8 hours of classroom instruction from a staff radiologist, a practice session using a patient with renal failure receiving continuous ambulatory peritoneal dialysis, and patient scans in the emergency care center with proctoring by full-time radiology staff. The first 50 examinations performed by each surgeon were recorded and reviewed by a board-certified radiologist masked to the clinical history and surgeon's interpretation.

Definitions

For purposes of comparing sensitivity and specificity, FAST observations were classified as true negative, true positive, false negative, or false positive. A true-negative result was defined as a negative FAST accompanied by a CT or DPL that did not detect hemoperitoneum. Lack of hemoperitoneum at celiotomy or absence of peritoneal signs

on serial physical examinations also confirmed a true-negative FAST. Intraperitoneal fluid identified by FAST accompanied by a positive confirmatory test was deemed a true-positive result. A false-negative result was defined as a negative FAST with a positive confirmatory test. Finally, when the FAST study was interpreted as positive for free fluid and the confirmatory test was negative, the FAST was classified as a false-positive result.

As previously described, three areas (right upper quadrant, left upper quadrant, and pelvis) were examined for hemoperitoneum. For purposes of analyzing false-negative FAST studies, hemoperitoneum on a confirmatory CT scan was quantified as mild (one area), moderate (two areas), and severe (all three areas). All solid organ injuries identified at celiotomy or CT scan were graded.[24]

Imaging Equipment

An Advanced Technology Laboratories (ATL) Ultramark V (Bothell, WA) sonography unit with a monochrome monitor located in the emergency room was used for all FAST examinations. Real-time images were obtained with an ATL CLA 3.5-mHz phase array probe. Each surgeon's initial 50 examinations were recorded with a video recorder.

All CT examinations were performed after administration of both diatrizoate sodium oral contrast material (Hypaque Sodium oral powder 5 g/350 mL water; Nycomed Inc., Princeton, NJ) and lopamidol nonionic intravenous contrast material (Isovue-370 Injection 150 mL; Bracco Diagnostics, Inc., Princeton, NJ). All scans were helically acquired using 10 mm of collimation with a pitch of 1.0:1 to 1.5:1 from the diaphragm to the iliac crest. Thinner collimation was used for the assessment of a pelvic fracture. CT scans were performed on a GE Hi-Speed (Milwaukee, WI) or Picker PQ 6000 (Cleveland, OH) scanner.

CUSUM Analysis

The CUSUM (S_n) for a series of observations is defined as: $S_n = \Sigma X_i - X_o$, where $X_i = 0$ for success, $X_i = 1$ for a failure, and X_o = a predetermined acceptable failure rate. If the target success rate is 95%, then the acceptable failure rate is 5%. CUSUMs are derived from consecutive observations. For example, surgeon X performs five consecutive procedures. The first, second, and fifth are successes; the third and fourth are failures. The CUSUMs for this set of observations are −0.05, −0.1, 0.85, 1.80, and 1.75. These values are plotted against the number of observations to produce a curve. A positive slope of the curve indicates that the acceptable failure rate is exceeded. In contrast, when the curve assumes a negative slope, the failure rate is less than the selected acceptable failure rate.

For purposes of CUSUM analysis, each observation was recorded as a success or failure. Success was defined as a positive or negative ultrasound observation validated by one or more of the confirmatory evaluations (true-positive or

true-negative FAST). A failure was defined as a positive or negative ultrasound observation invalidated by one of the validating tools (false-positive or false-negative FAST). CUSUM functions were calculated for targeted success rates of 85%, 90%, 95%, and 98% for the institution and 85%, 90%, and 95% for the individual surgeons. These functions were plotted against the number of observations to produce CUSUM graphs.

A typical learning curve produced by CUSUM analysis demonstrates an initial upward slope, corresponding to a period when the incidence of failure exceeds the acceptable failure rate. This upward slope is followed by either a plateau (indicating that the observed incidence of failure is equal to the acceptable failure rate) or a downward slope (indicating that the observed failure rate is lower than the acceptable failure rate). The point at which the curve either plateaus or acquires a consistent negative slope indicates the number of attempts required to learn to perform the process under study with the target rate of success under study.

Statistical Analysis

The ISS, abbreviated injury score, and revised trauma score are reported as mean ± standard error of the mean. Overall differences were analyzed by the nonparametric Kruskal-Wallis test. Pairwise comparisons were tested by the Wilcoxon test. $P < .05$ was regarded as significant.

RESULTS

Demographics

A total of 3,396 patients were admitted after blunt trauma during the study period of February 1, 1996, to August 30, 1998. Six hundred seventy-three (19.8%) patients had a FAST examination performed, of whom 546 (81%) met the study criteria. The ISS for surgeon E was significantly greater than the mean ISS for surgeons C and D. The abdominal abbreviated injury score and revised trauma score for each surgeon's patient population were not significantly different (Table 1). The mechanisms of injury included motor vehicle crash (n = 380, 70%), falls (n = 59, 11%), pedestrian struck by vehicle (n = 43, 8%), motorcycle crash (n = 29, 5%), assault (n = 23, 4%), and crush (n = 12, 2%). There were 339 men and 207 women in the study population, with an average age of 36 years. Only 24 (4.3%) patients were in shock (blood pressure <90) at admission. The number of specific confirmatory studies performed was as follows: CT scan, 375 (68%); DPL, 28 (5%); and celiotomy or serial physical examination, 171 (30%). The frequency of FAST examinations consistently increased throughout the study period (Fig. 1).

Surgeon	ISS	Abd. AIS	RTS	Shock (% of patients)
A	14.0 ± 1.49	0.9 ± 0.20	7.0 ± 0.22	17
B	15.8 ± 1.63	0.6 ± 0.15	6.9 ± 0.22	22
C	11.9 ± 1.34	0.5 ± 0.16	7.5 ± 0.14	13
D	13.2 ± 0.75	0.7 ± 0.08	7.0 ± 0.11	17
E	16.8 ± 1.08*	0.8 ± 0.12	6.8 ± 0.16	18
Departmental	14.3 ± 0.51	0.7 ± 0.06	7.0 ± 0.07	17

ISS, injury severity score; Abd. AIS, abdominal abbreviated injury score; RTS, Revised trauma score.
* $P < .05$ vs. surgeons C and D.

Sensitivity, Specificity, Positive Predictive Value, and Negative Predictive Value

Evaluation of the institutional data yielded 439 true-negative, 60 true-positive, 25 false-negative, and 22 false-positive FAST results. The institutional sensitivity was 71% and the specificity was 95%. The positive and negative predictive values for all the trauma surgeons were 73% and 95%, respectively. The sensitivity for each surgeon varied from 44% to 100%. Specificity exceeded 92% for all surgeons. The positive predictive value for the group ranged from 57% to 83%, the negative predictive value from 91% to 100%. The overall departmental accuracy was 91%; it ranged from 87% to 98% for individual surgeons. An individual's performance did not appear to be affected by the number of studies completed (Table 2).

There were 25 false-negative FAST examinations with hemoperitoneum identified by CT scan (Table 3). More than half of the CT scans identified a mild amount of blood (n = 13). A moderate hemoperitoneum (two areas) was present in 10 patients. Only two patients had blood in all three areas (severe). Four patients with a mild hemoperitoneum also

had a pelvic ring fracture. None of the false-negative studies were explained by the presence of subcutaneous emphysema or spinal injuries. Seven of the patients with false-negative examinations (mild, n = 2; moderate, n = 5) underwent celiotomy based on a positive confirmatory test (DPL, n = 3; CT, n = 4). All were hemodynamically stable. Injuries accounting for the free fluid on CT scan included two small bowel, one mesenteric rent, and four splenic lacerations requiring splenectomy. Eighteen hemodynamically stable patients with a false-negative FAST result were found to have one or more solid organ injuries. The injuries included nine splenic lacerations (grade 1, n = 3; grade 2, n = 6) and eight liver lacerations (grade 1, n = 3; grade 2, n = 3; grade 3, n = 2). One patient had both a grade 1 splenic injury and a grade 3 liver injury. All 18 patients were successfully managed without surgery.

Of the 22 false-positive results, only two patients underwent celiotomy. In both, neither a CT scan nor a DPL was performed. In one patient, the physical examination was consistent with peritonitis, raising concern about a possible bowel injury. No identifiable reason for the false-positive FAST result was found at celiotomy. The second patient had a large retroperitoneal hematoma with renal avulsion. The remaining 20 patients were managed without surgery. Associated injuries were examined in those patients. None had evidence of subcutaneous emphysema, retroperitoneal hemorrhage, or spinal injuries. Only two patients had a pelvic fracture.

CUSUM Analysis

For the purpose of CUSUM analysis, the success or failure for each FAST examination interpretation was determined by agreement of the confirmatory study. A success was a true-positive or true-negative result, a failure a false-positive or false-negative result. These observations were then subjected to CUSUM analysis in chronologic order for target success rates of 85%, 90%, 95%, and 98% (Fig. 2). Upward slopes indicate failures exceeded the acceptable
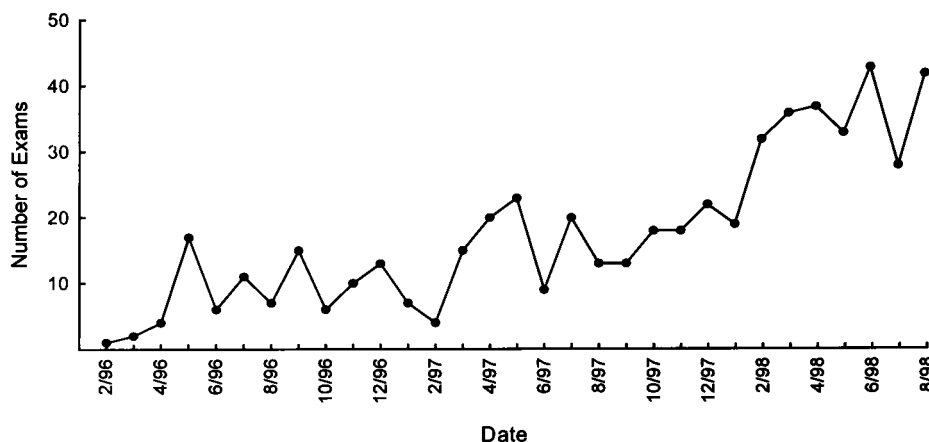

Figure 1. Number of focused abdominal ultrasound for trauma (FAST) examinations performed each month of the study. A total of 546 examinations were performed. The maximum number of examinations performed in any month was 43.

Table 2. CLINICAL UTILITY AND PERFORMANCE

Surgeon	No. of Exams	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy (%)
A	54	55	98	83	92	91
B	63	44	94	57	91	87
C	48	100	98	75	100	98
D	250	80	95	76	96	93
E	131	67	93	70	93	89
Departmental	546	71	95	73	95	91

PPV, positive predictive value; NPV, negative predictive value.

failure rate; downward slopes indicate failures occurring at less than the acceptable rate.

There were 499 successes and 47 failures for the institutional analysis. The first failure occurred with patient 24. For the 98% target success rate, the slope of the curve becomes positive at this patient, with one failure, and remains positive throughout the following patients. The final CUSUM for the 98% target rate is 36.8. With an acceptable failure rate at 2% (-0.02 for each success), more than 1,500 successful examinations without any failures would be necessary to achieve a negative slope again. A 95% target rate produces a positive slope at patient 62, with four failures. The final CUSUM for the 95% target rate is 19.7. With an acceptable failure rate at 5% (-0.05 for each success), nearly 400 successful examinations without any failures would be necessary to achieve a negative slope. The slope for the 90% target success rate remains negative (failures never exceed the target rate) throughout 546 examinations despite 47 failures. Additional examinations never improve the 90% rate, and attainment of this rate of success is met at the onset of clinical practice. The first failure is patient 24, but this does not change the plateau of the curve. Target success rates of less than 90% were examined, and the analysis revealed slopes that exhibited sharp successive decreases for each 5% increment studied.

The CUSUM analysis for individual surgeons was carried out in the same manner as previously described. For surgeon A, target success rates of 95% and 90% were not achieved (Fig. 3A). Surgeon A had 5 failures and 49 successes. Each failure brings the slope sharply upward. At a 95% success rate, the slope becomes positive at patient 7. The CUSUM

at this point is 0.65, and at least 15 consecutive successes were needed to return to a negative slope. For the 90% target rate, the slope becomes positive at patient 21, and it remains positive thereafter. The 85% target rate is easily obtained for surgeon A. The slope for this target rate and lower target rates remains negative throughout the entire caseload.

Surgeon B had 8 failures and 55 successes (see Fig. 3B). The failures occurred at patients 2, 6, 14, 30, 37, 45, 55, and 58, as indicated by a change in the slope of the curve. For a target success rate of 95%, the slope becomes positive at the second patient with a CUSUM of 0.90 ($S_n = -0.05 + 0.95$). From that point, more than 20 consecutive successes are needed for the performance rate to equal the 5% failure rate. The slope of the 90% success rate is similar. The success rate for surgeon B falls between 80% and 85%. The 85% target curve does not produce a negative slope until patient 20. The 80% (not shown) target rate remains negative after patient 11, despite failures.

Surgeon C demonstrates high performance curves despite performing the fewest FAST studies (see Fig. 3C). Negative slopes are maintained for 95% through 85%. Surgeon C only had 1 failure out of 48 examinations; it occurred at patient 21. Because this surgeon maintained a 95% target success rate from the onset, the 98% success rate (not shown) was also examined. For a 98% success rate, the slope is initially negative and becomes positive after the only failure. Despite the 34 successful cases, the 98% slope never returns to a negative value. Hence, the success rate of this surgeon falls between 95% and 98%.

Surgeon D performed more examinations than any other

Table 3. ANALYSIS OF FALSE-NEGATIVE RESULTS BY SURGEON AND THE AMOUNT OF HEMOPERITONEUM

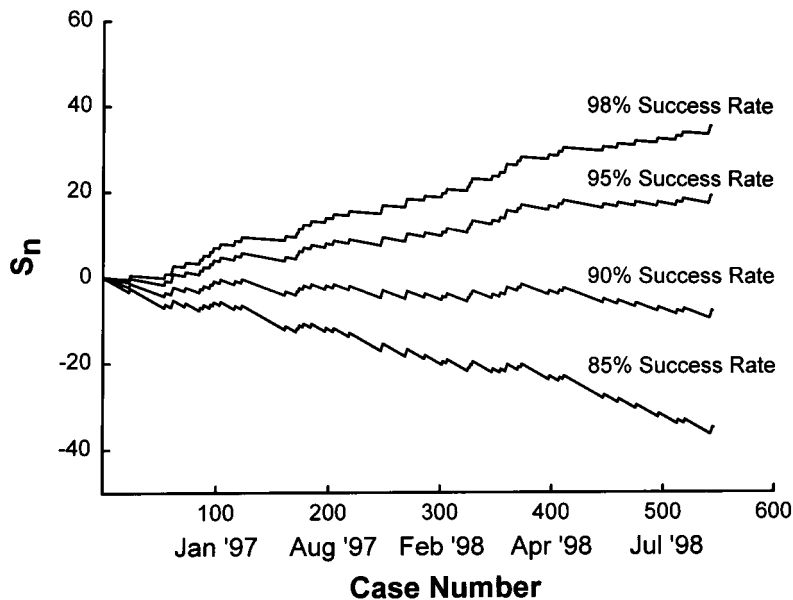|	Surgeon A	Surgeon B	Surgeon C	Surgeon D	Surgeon E	Total
Number of false-negative exams	4	5	0	8	8	25
Mild (1 area)	1	3	—	2	7	13
Moderate (2 areas)	3	1	—	5	1	10
Severe (3 areas)	0	1	—	1	0	2

Figure 2. The cumulative sum analysis (S_n) of the institutional results of focused abdominal ultrasound for trauma (FAST). The cumulative sum was calculated for target success rates of 98%, 95%, 90%, and 85%. Analysis of the curves reveals a plateau of the success rate between 85% and 90%. Achievement of this success rate at the outset of clinical use indicates there is no learning curve.

surgeon (see Fig. 3D). He had 18 failures and 232 successes. A target success rate of 95% is initially met but is never maintained despite the large number of examinations. The CUSUM for surgeon D at the 95% success rate is 5.5 after 250 examinations. More than 100 successes would be necessary to return to a negative slope. The 90% target rate is maintained at the acceptable failure rate (10%) throughout the caseload. The 85% target is also maintained after the first FAST examination.

Surgeon E had 15 failures and 116 successes (see Fig. 3E). Initially, surgeon E demonstrates negative slopes for all target rates. The first failure occurs at patient 16 and is followed by failures at patients 17, 18, and 23. This cluster of failures pushes the slopes for the 95% and 90% target rates irretrievably upward. The CUSUM after 131 examinations is 8.45 for the 95% target rate and 1.9 for the 90% target rate. The 85% success rate is maintained (negative slope) after patient 27.

DISCUSSION

Ultrasonography has been successfully used by nonradiologist physicians such as obstetricians and cardiologists for many years without the benefit of formal radiology training. General surgeons and emergency physicians have added ultrasound to their diagnostic armamentarium more recently.[6,9,14] Several studies have proven the efficacy of ultrasonography in the evaluation of blunt abdominal trauma.[5,6,9] Surgeon-performed FAST has been shown to be as accurate as CT and DPL in detecting hemoperitoneum.[12,13,15,16]

As with any form of technology that is introduced into routine clinical use, the increased use of FAST has raised questions about how to resolve credentialing issues, demonstrate examiner competency, and conduct evaluations of

sustained examiner performance. These issues have been partially addressed in several studies. Thomas et al[25] prospectively studied the de novo implementation of a trauma ultrasound program to establish a learning curve. They found that accuracy improved after the first 100 patients. In a 24-month retrospective review by Smith et al,[14] it was determined that learning curves could not be constructed. When performance data after the initial 6 months of practice were compared with performance data after the last 6 months, there was no significant difference in sensitivity, specificity, or accuracy. They concluded there was no evidence of a learning curve during the 24-month study. Shackford et al[15] prospectively examined the initial FAST experience of 12 surgeons and emergency physicians using ultrasonography to examine 241 injured patients. They identified the initial error rate as 17% and noted that this error rate fell to 5% after 10 patients. Both studies suggest that extensive didactic training and high volumes of examinations are not necessary for trauma surgeons to overcome the learning curve associated with FAST. However, no studies have analyzed the performance data to evaluate the learning curve for individual examiners. Demonstration of competent performance of FAST by individual physicians is necessary to make rational decisions regarding credentialing to perform the study. Performance curves derived from CUSUM analysis of the results of individual examiners, especially when compared with curves for other practitioners conducting the study under similar conditions and on similar populations, can be used to guide such decisions.

Proper analysis of data regarding performance requires an evaluation of individual training, experience, and practice styles. All surgeons in this study participated in a standardized curriculum and practical experience before clinical use, thus eliminating differences in training as a variable when
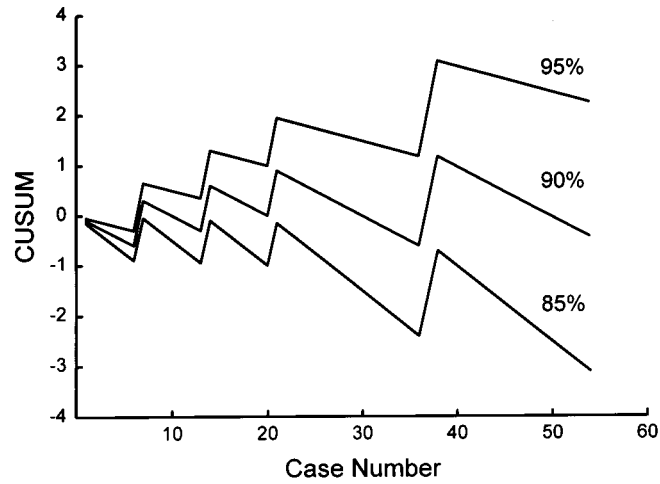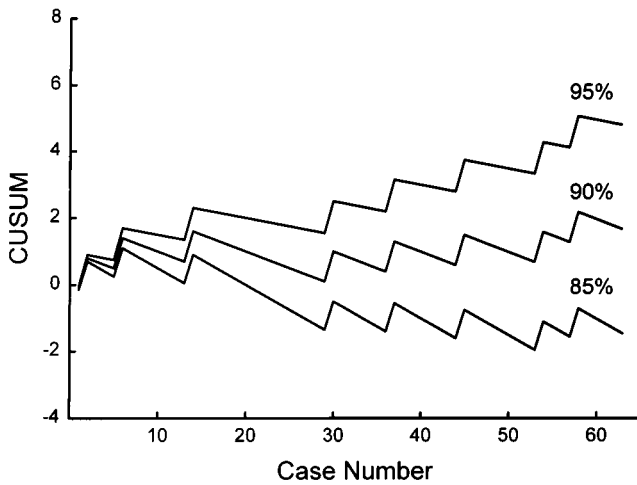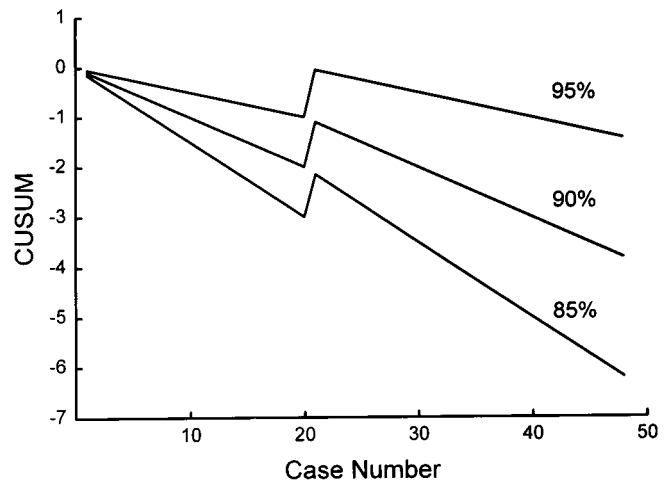
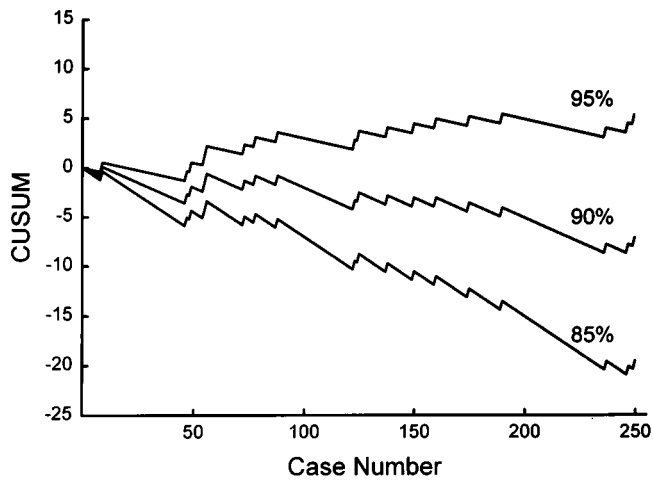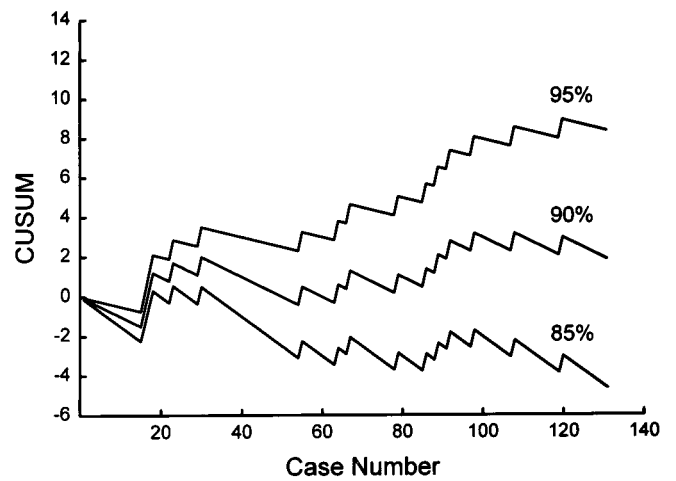**Figure 3.** The cumulative sum analysis of focused abdominal ultrasound for trauma (FAST) results for individual surgeons. Performance rates of 95%, 90%, and 85% are shown for surgeons A through E.

evaluating performance. Only 17% of all patients with blunt trauma admitted during the study period underwent FAST examination. In an effort to control for individual surgeon variation in the use of FAST and patient volume, the number of FAST examinations performed by the attending was divided by the number of trauma patient admissions for each surgeon during the study period. This number allowed us to determine the surgeon's utilization rate of FAST, which ranged from 10% to 41% (10%, 41%, 18%, 36%, and 16% for surgeons A through E, respectively). This individual variation in clinical practice did not affect performance. The overall sensitivity of the department was 71%, and the specificity was 95%. For individual surgeons, these values were 44% to 100% for sensitivity and 93% to 98% for specificity, comparable to previously cited results.[26–28]

Specificity, sensitivity, and predictive values can be influenced by patient selection. There was a significant difference in the mean ISS for patients examined by individual surgeons (see Table 1). These differences could not be attributed to any particular pattern in sensitivity, specificity, or predictive values. Performance numbers can also be affected by patient selection, as demonstrated by surgeon C. He had the highest accuracy, sensitivity, specificity, positive predictive value, and negative predictive value but a significantly lower ISS compared with surgeon E. Although the patients that surgeon E examined with FAST had the highest ISS, his patient selection may explain the lower sensitivity (67%) and specificity (93%) for his performance.

In CUSUM analysis, each result is classified as a success or failure. As a performance tool, each surgeon must reexamine the cause or contributing factor involved in failures. However, failures were subdivided into false-negative and false-positive results in an effort to construct a clinically relevant analysis of the data in this study. Nine percent of the FAST examinations in this study had either a false-negative or a false-positive result. Fifty-two percent (13/25) of the failures to diagnose hemoperitoneum occurred in patients with free fluid present in only one area (mild hemoperitoneum). Previous studies have identified occult hollow viscus injuries, subcutaneous emphysema, and pelvic ring fractures as contributing factors for false-positive results.[29] We identified subcutaneous emphysema, pelvic fractures, spinal injury, and renal injury as a possible cause in the misinterpretation of only seven FAST examinations. Two pelvic fractures and one renal avulsion possibly led to three false-positive results. Four false-negative results were associated with a pelvic ring fracture. Another contributing factor may be the overreading of results with initial use of FAST, which would be suggested by a concentration of incorrect interpretations at the beginning of a surgeon's experience. In this study, false-positive and false-negative results occurred randomly throughout each surgeon's experience for the entire study period.

Our CUSUM analysis revealed that the learning curve for FAST was nonexistent for these attending trauma surgeons after completion of a relatively short program of didactic instruction and simulated clinical examinations. CUSUM analyses of the departmental FAST results were derived for target success rates of 85% to 98%. The 90% target success rate was the highest success rate that maintained a negative slope. The group of surgeons demonstrated immediate performance at or above this target accuracy rate. With the exception of surgeon C, individual surgeons displayed remarkable similarity in their performance despite variations in frequency of use and patient selection. All surgeons immediately achieved a target success rate of at least 85%; surgeons C and D attained higher performance rates of 98% and 90%, respectively. The number of procedures performed by a surgeon did little to improve or diminish performance.

This information is applicable to current debates concerning performance evaluation and credentialing for FAST. Several recommendations have been made with regard to FAST education and credentialing for surgeons.[30,31] Our CUSUM analysis provides a performance curve for determination of competency. Competency is achieved by a plateau of the CUSUM curve at a predetermined target success rate, rather than by an arbitrary number of examinations. Competency is maintained at a certain success rate when the CUSUM curve has plateaued or assumed a negative slope. A positive slope in the CUSUM curve detects a false-positive or false-negative FAST result (a failure). A persistently positive slope with continued use may indicate the need for additional training or retraining. At a target success rate of 90%, all surgeons in this study displayed early competency and maintained that competency throughout the study period.

The advantage of CUSUM analysis is that it allows us to study performance over time, with experience as a continuous variable. A few random failures have little influence on performance, but several failures will become graphically apparent and will change the slope of performance. The CUSUM plot provides a visual output for quality control and compliance. CUSUM analysis is a unique and accurate method of evaluating the process of learning a new skill, as well as monitoring compliance with quality improvement goals after mastery of a skill. Using this method of performance analysis, we conclude that with a modest educational and didactic experience, surgeons use FAST accurately in clinical practice, and additional experience does not appear to affect this performance level.

Acknowledgments

References

1. Fabian TC, Mangiante EC, White TJ, Patterson CR, Boldreghini S, Britt LG. A prospective study of 91 patients undergoing both computed tomography and peritoneal lavage following blunt abdominal trauma. J Trauma 1986; 26:602–608.

2. Meyer DM, Thal ER, Weigelt JA, Redman HC. Evaluation of computed tomography and diagnostic peritoneal lavage in blunt abdominal trauma. J Trauma 1989; 29:1168–1172.

3. Meredith JW, Ditesheim JA, Stonehouse S, Wolfman N. Computed tomography and diagnostic peritoneal lavage. Ann Surg 1992; 58:44–48.

4. Von Kuhn F-P, Schreyer T, Schild H, et al. Sonographie beim stumpfen Bauchtrauma. Fortschr Geb Roentgenstr Nuklearmed 1983; 139:310–313.

5. Wening JV. Evaluation of ultrasound, lavage, and computed tomography in blunt abdominal trauma. Surg Endosc 1989; 3:152–158.

6. Förster R, Pillasch J, Zielke A, et al. Ultrasonography in blunt abdominal trauma: influence of the investigator's experience. J Trauma 1993; 34:264–269.

7. American Institute of Ultrasound Medicine. Official statement: guidelines for physicians who evaluate and interpret diagnostic ultrasound examinations, 1997.

8. Kimura A, Otsuka T. Emergency center ultrasonography in the evaluation of hemoperitoneum: a prospective study. J Trauma 1991; 31:20–23.

9. Hoffman R, Nerlich M, Muggia-Sullam M, et al. Blunt abdominal trauma in cases of multiple abdominal trauma evaluated by ultrasonography: a prospective analysis of 291 patients. J Trauma 1992; 32:452–458.

10. Liu M, Lee CH, P'eng FK. Prospective comparison of diagnostic peritoneal lavage, computed tomographic scanning, and ultrasonography for the diagnosis of blunt abdominal trauma. J Trauma 1993; 35:267–270.

11. Tso P, Rodriquez A, Cooper C, et al. Sonography in blunt abdominal trauma: a preliminary progress report. J Trauma 1992; 33:39–44.

12. Rozycki GS, Ochsner MG, Jaffin JH, Champion HR. Prospective evaluation of the surgeons' use of ultrasound in the evaluation of trauma patients. J Trauma 1993; 34:516–527.

13. Ma OJ, Mateer JR, Ogata M, et al. Prospective analysis of a rapid trauma ultrasound examination performed by emergency physicians. J Trauma 1995; 38:879–885.

14. Smith RS, Kern SJ, Fry WR, et al. Institutional learning curve of surgeon-performed trauma ultrasound. Arch Surg 1998; 133:530–533.

15. Shackford SR, Rogers FB, Osler TM. Focused abdominal sonogram for trauma: the learning curve of nonradiologist clinicians in detecting hemoperitoneum. J Trauma 1999; 46:553–562.

16. Williams SM, Parry BR, Schlup MMT. Quality control: and application of the CUSUM. Br Med J 1992; 304:1359–1361.

17. Schlup MMT, Williams SM, Barbezat GO. ERCP: a review of technical competency and workload in a small unit. Gastrointest Endoscopy 1997; 46:48–52.

18. Van Rij AM, McDonald JR, Pettigrew RA, et al. CUSUM as an aid to the early assessment of the surgical trainee. Br J Surg 1995; 82:1500–1503.

19. Molloy M, Bower RH, Hasselgren PO, Dalton BJ. Cholangiography during laparoscopic cholecystectomy—cumulative sum analysis of an institutional learning curve. J Gastrointest Surg 1999; 3:185–188.

20. Baker SP, O'Neil B, Haddon W, et al. The Injury Severity Score: a method for describing patients with multiple injuries and evaluating emergency care. J Trauma 1974; 14:187–196.

21. American Association for Automotive Medicine. The Abbreviated Injury Scale (AIS)—1990 Revision. Des Plaines, IL.

22. Champion HR, Sacco WJ, Copes WS, et al. A revision of the trauma score. J Trauma 1989; 29:623–629.

23. Committee on Trauma. Advanced Trauma Life Support Course: Instructor Manual. Chicago: American College of Surgeons; 1998.

24. Moore EE, Cogbill TH, Jurkovich GJ, et al. Organ injury scaling: spleen and liver. J Trauma 1995; 38:323–324.

25. Thomas B, Falcone RE, Vasquez D, et al. Ultrasound evaluation of blunt abdominal trauma: program implementation, initial experience, and learning curve. J Trauma 1997; 42:384–388.

26. Kern ST, Smith A, Fry WR, et al. Sonographic examination by senior surgical residents. Am Surg 1997; 63:669–674.

27. McKenney MG, Martin L, Lentz K, et al. 1000 consecutive ultrasounds for blunt abdominal trauma. J Trauma 1996; 40:607–612.

28. Rozycki GS, Oschner MG, Schmidt JA, et al. A prospective study of surgeon-performed ultrasound as the primary adjuvant modality for injured patient assessment. J Trauma 1995; 29:492–500.

29. Ballard RB, Rozycki GS, Newman PG, et al. An algorithm to reduce the incidence of false-negative FAST examination in patients at high risk for occult injury. J Am Coll Surg 1999; 189:145–151.

30. Shackford SR, Ricci MA, Hebert JC. Education and Credentialing. Probl Gen Surg 1997; 14:126–132.

31. Cushing BM, Chiu WC. Credentialing for the ultrasonographic evaluation of trauma patients. Trauma Q 1997; 13:205–213.

Discussion

DR. GRACE S. ROZYCKI (Atlanta, Georgia; read by Dr. David Feliciano): I am pleased and honored to have the opportunity to discuss this manuscript that addresses the important issue of training and accuracy in the performance of the FAST exam. Developed for the evaluation of injured patients and performed by surgeons, the FAST readily and accurately detects hemopericardium and/or hemoperitoneum in patients with potential truncal injuries. Recently, the American Board of Surgery has redefined the scope of surgical practice to include a working knowledge of ultrasonography, underscoring the importance of proper training so that institutional guidelines can be developed for verification and credentialing in this modality.

In this manuscript, the authors evaluated the institutional and individual learning curves of surgeons for the FAST by analyzing the incidence of diagnostic inaccuracies as a function of the examiner experience. The surgeons underwent training and the examinations were videotaped and reviewed by board-certified radiologists; 546 patients underwent FAST examinations, yielding 428 true-negative, 60 positive, 36 false-negative, and 22 false-positive examinations. Although the sensitivity was only 63%, the cumulative sum analysis of the surgeons' experiences with the performance of the FAST showed that they exceeded an 85% accuracy rate at the outset of the study. Based on their data, Dr. McCarter and colleagues concluded the following: one, with a modest educational and didactic experience, surgeons can accurately use the FAST in clinical practice and, two, experience in the performance of the FAST does not improve the accuracy.

The authors have examined this important issue and developed a mathematically sound method of assessing this learning curve. There are concerns, however, about the relevance of this model in the clinical arena because there is no analysis of the "failures," that is, the false-positive and false-negative examinations. To improve the accuracy of ultrasound, it is important to analyze the false-positive and false-negative examinations so that their etiologies can be determined and they can be avoided in the future. Such an analysis yields information that identifies, one, problems with poor imaging related to an incomplete understanding of the principles of ultrasound physics and, two, problems with the expectations of what ultrasound can diagnose in certain clinical scenarios.

In terms of an incomplete understanding of the principles of ultrasound physics, a small amount of blood in the subphrenic space in a patient with blunt abdominal trauma may be obscured by a rib shadow, in the hands of an inexperienced surgeon-sonographer. A small manipulation of the transducer while the patient breathes allows identification of the blood despite the attenuation of the ultrasound waves at the bone/tissue interface. Hence, the recognition of an artifact and knowledge of its etiology allows a more accurate image to be obtained. With only 60 true-positive examinations in this entire study, about 12 per surgeon, there are too few true-positives to provide enough experience with this examination. Even though the surgeons performed FAST examinations on patients who underwent peritoneal dialysis, those patients frequently do not provide the subtleties of small amounts of intraperitoneal fluid.

In terms of expectations for this exam, hollow viscus perforation may be present with a very small amount of fluid that is not detected by the FAST examination in the hands of the best surgeon-sonographer; however, as indicated, a diagnostic peritoneal lavage in the high-risk patient with a "seatbelt sign" would likely show positive results. In similar fashion in our recent study presented at the American College of Surgeons, we determined that a patient with a pelvic fracture has an increased incidence of occult intraabdominal injury and is more likely to have a false-negative FAST examination. Therefore, although FAST examinations may be performed accurately, the test itself may not be able to detect the small amount of hemoperitoneum found in certain patients who may have significant intraabdominal organ injuries, especially in the retroperitoneum.

Considering the overall low sensitivity, 63%, of the FAST in this study, it is even more important to examine these issues to determine if, at Cincinnati, there is opportunity for improvement in the accuracy of surgeon-sonographers. From our analysis of 1,540 patients presented at the American Surgical Association 2 years ago, we found that a careful analysis of our data did yield valuable findings about the accuracy and limitations of the FAST, as well as its role in the evaluation of the complex multisystem-injured patient.

As a registered diagnostic medical sonographer, I have found that my ultrasound technique and that of the partners in our group at Grady has improved considerably over 9 years. I continue to reinforce the principles of ultrasound physics as I teach these skills to other surgeons. As we consider developing guidelines for educating surgeons in this modality, these issues should be carefully examined because, similar to the development of surgical skills, each ultrasound technique learned reinforces the surgeon's knowledge of ultrasound principles, encourages rapid learning of new ultrasound techniques, and extends the surgeon's diagnostic armamentarium.

I have three questions. First, the wide range of accuracies among surgeon-sonographers in this study is interesting. Was any attempt made to analyze the false-negative and false-positive examinations to allow for corrective education and a better accuracy in the future?

Second, of the 36 false-negative studies, how many were related to pure retroperitoneal injuries from pelvic or spinal fractures that, in our published algorithm, are automatically evaluated by follow-up CT, recognizing the limitations of the 3.5-MHz ultrasound probe?

Third, instead of excluding equivocal exams, why did you not consider a repeat FAST within 15 to 30 minutes of the first exam by the same surgeon-sonographer? Most experienced centers do these follow-up exams, particularly if the patient remains tachycardiac and/or complains of persistent abdominal pain.

DR. TIMOTHY C. FABIAN (Memphis, Tennessee): The authors contend that with a modest course of training, surgeons can attain a high level of competency in performing FAST examinations, a level such that there is essentially no learning curve. In analyzing their data, I am less confident in that conclusion. I think a lot of my observations are going to be similar to what Dr. Rozycki made just previously.

I believe there may be a problem with the cumulative sum analysis as applied to this situation. In their methodology, they used accuracy, i.e., true-positive or true-negative, to determine success. At first glance, that appears appropriate; however, it must be noted that 82% of exams were negative. Therefore, without ever examining a patient, given the population selected, one would be accurate 82% of the time if they always state the FAST was negative and would in fact achieve similar results with cumulative sum analysis as reported here.

It would seem to me that much more emphasis should be placed on the true-positive exams, 18% of the study group. Thus, perhaps sensitivity should be used as the determiner for the analysis, i.e., consider only results from positive patients, true-positives plus false-negatives.

The authors report sensitivities of 40% to 75% among the five participating surgeons. Those numbers illustrate two points. First, the sensitivities are lower than the 80% to 90% that the authors quote from other reported series. Second, the intrainstitutional variance which is twofold among the surgeons clearly illustrates the important issue of operator dependency for this technology.

At a 40% sensitivity, meaning that six of ten patients with blood in the abdomen are missed, it is difficult to justify using this diagnostic modality. There was also a wide discrepancy among the surgeons relative to frequency of utilization. Of the five, surgeon D performed 46% of the studies. Interestingly, that surgeon was the only one to achieve a 90% target rate, which occurred after 170 cases were performed. That result would seem to contradict the overall conclusion. Could you please address that issue?

Thirty percent of the cases had confirmation of the FAST exam by serial clinical evaluations. While this may be relevant to significant clinical outcomes, can the data be stacked up with CT and DPL to determine accuracy of ultrasound?

Finally, along the line of clinical relevance, how many of the patients required therapeutic laparotomies to manage significant injuries?

DR. MICHAEL ROTONDO (Greenville, North Carolina): I'd like to commend Dr. Fischer and his trauma section, led by Drs. Jim Hurst and Scott Frame, on their continued contribution to the field of trauma and critical care, and I'd like to thank him and Fred Luchette for the kind invitation to discuss the paper.

The rich history in the development of the techniques to detect hemoperitoneum in injury management can be traced simply by looking around the room and at the roster of the members of this Association. Dr. Root was the father of diagnostic peritoneal lavage in the 1960s and it continued to be the gold standard in the 1970s. Through the 1980s and 1990s, Dr. Fabian and others produced some very insightful papers that compare CT scanning to diagnostic peritoneal lavage in hemoperitoneum. More recently,

Dr. Rozycki and others have introduced focused abdominal sonography for detection of hemoperitoneum.

Dr. Fischer and colleagues have added yet another important consideration as we try to bring ultrasound into clearer focus. The CUSUM approach may take us a long way toward unravelling the mysteries surrounding the issues of credentialling and competency for a whole myriad of procedures that we perform as surgeons. The approach is similar in many ways in utility to control charts, an industry quality management method that is designed to monitor performance variances in a way that relates the timing and sequencing of those variances against performance standards. That's an important concept. This technique may truly be valuable over the long term.

Having said that, I have four comments and questions for the authors. The first has to do with the sensitivity and a need for the authors to analyze false-positives.

The second is in relation to the false-negative studies and the degree of hemoperitoneum. Could the hemoperitoneum be below the threshhold limits of ultrasound? And would a follow-up study have altered the results?

The third is where do you think ultrasound fits into the management algorithm for the CUSUM-credentialed surgeon? For example, with the high negative predictive value of ultrasound, can we call it quits after one negative study?

And finally, how do you recommend that we blend the CUSUM approach into our performance improvement in credentialling programs?

Dr. James Evans (Shreveport, Louisiana): I think the single most important thing is to have placed FAST into its proper perspective with regard to being a modality used by surgeons. There are several comments I would make and then I have one or two questions.

First, as we make the translation from industry to healthcare industry, and most of this has been used in the chemical industry, one of the things that the engineering aspect of this statistical program has been used for has been the delineation of time sequence variability. As you know, the variability of productivity of engineering systems prior to the introduction of robotics was especially variable with regard to shift work. It's very nice to see this statistic applied in this particular aspect, and I would ask the question as to whether or not you think that that is in fact a variable? I would suspect it is not.

The other thing is if you look at the applications in the previous statistical manipulations, you would expect that this is exactly what the outcome would be. And I think the statistics exactly fit the outcome expectations. The reason for that is you are using a dichotomous variable. If you look at those aspects of the engineering world that have used CUSUM for continuous variables where there are more places for having a greater gradation, you are using just all positive or all negative as your outcome, this is actually what you would expect.

I think the other thing that this manuscript highlights for us is that when surgeons come to a new technology, they collectively bring the knowledge that has already been acquired. It is not the first time a surgeon is going to see an ultrasound film. This is not a de novo undertaking, and so this is not an unexpected finding to say that you bring your cumulative experience to it. I think that's an important aspect to look at.

The other thing is to look at the introduction of the timing of this particular clinical report. One of the great advantages that it has is that it comes at a time when we have reached a more user-friendly device.

The first question I would have would be what particular device did you use? There has been a serial progression, and even though the megahertz probes may be the same, the refinement of the display, as well as the ability to interpret it—and I suspect that we are talking about bright people who are using new, good technology, and there is no reason to expect, using a dichotomous variable, that you would have anything else.

My other question would be sort of an extension of the previously asked question, and that is, we are not, I assume, proposing that this would be an eliminative test but an adjunctive test. So the real question is not an all-or-none phenomenon but, wherein will we integrate this into the algorithm? What they have not done is say discard something—they have in fact expanded the algorithm. My difficulty is figuring out exactly where they have expanded the algorithm. But clearly, they have expanded it very strongly on the negative side. It's no different than any other clinical exam that has a dichotomous outcome, exactly what you would have expected and very elegantly presented.

Dr. Fred A. Luchette (Closing Discussion): Thank you for the privilege of presenting our data this afternoon. I also want to thank my Chairman, Dr. Joe Fischer, for sponsoring this work and for mastering a new concept and data in a few hours, leading to a succinct presentation. A few years ago, Dr. Fischer appointed me to our hospital's Credentials Committee. After the first couple of meetings, the turf battles going on in various specialty practices became very apparent. Three years ago, I headed up the efforts of our Division of Trauma/Critical Care to begin using FAST for detection of hemoperitoneum in our trauma population. This is the foundation of and motivation for performing the study presented today.

The discussants had questions and concerns about the relatively low sensitivity of 71% for our group's cumulative data, as well as two practitioners with unacceptably low sensitivity. The goals of this study were to 1) review our experience and accuracy with FAST, and 2) present a new tool, CUSUM analysis, which can be easily used to track individual practitioners' successive performance with a new diagnostic test. This may be useful for credentialing and quality assurance.

All three discussants had major criticisms and questions about the low sensitivity for the group and unacceptably low sensitivity for two practitioners. This is clearly related to the false-negative studies. When we analyzed the number of false-negative studies for each practitioner, they ranged from a low of 2.1% for surgeon C to a maximum of 9.5% for surgeon B. Interestingly, surgeon D had a 6% false-negative rate, while surgeon E had a 7.6% false-negative rate. If one recognizes that surgeon C's patient population had a significantly lower ISS and frequency of shock, then the other four surgeons had a false-negative rate ranging from 6% to 9.5%. We recognize Dr. Rozycki's expertise and mastery of ultrasonography in the initial evaluation of trauma patients, and we congratulate her as a Certified Examiner by the American Registry of Diagnostic Medical Sonographers. In contrast, however, all of the practitioners in this study have added qualifications in Surgical/Critical Care and practice trauma care on a daily basis. We do not have any additional training in ultrasonography other than the curriculum I described, and I would suggest that we represent the majority of trauma surgeons attempting to use FAST to evaluate the peritoneal cavity for hemoperitoneum. Although these false-

negative rates are higher than previous studies in the literature, they do represent our experience. Only seven of 25 patients who had a false-negative study required celiotomy. Injuries accounting for the free fluid on CT scan included two small bowel, one mesenteric rent, and four splenic lacerations requiring splenectomy.

Dr. Rozycki also asked why we excluded equivocal exams. Equivocal exams do not lend themselves to classification as success or failure for CUSUM analysis.

Drs. Rozycki and Rotundo wanted to know whether we were utilizing FAST in a standardized algorithm. We are considering the development of a protocol, but did not use one at the time of this study. During the study, each individual practitioner utilized FAST in the evaluation of their patients at their discretion. There was no protocol or algorithm for utilization of FAST in the initial management of the patient.

Dr. Fabian suggested that 82% of our exams were true-negatives and thus by law of probability, a practitioner could have a fairly high accuracy rate just by calling all studies negative. If one reviews the literature, the majority of reports to date have a preponderance of true-negatives. I am not aware of any series that focuses an analysis of sensitivity, specificity, and positive predictive value by using just true-positive studies. In fact, our study has one of the largest groups of true-positives (N = 60). Only McKenney reported a larger group of 70 true-positives.

Dr. Rotundo asked a question about the role of CUSUM in the credentialing process. This is exactly the reason for this presentation. The Australians have used CUSUM analysis to evaluate surgical trainees' performance. They set time limits for the performance of an appendectomy and inguinal herniorrhaphy, and showed various learning curves for each of the trainees in their series. Other investigators have found CUSUM analysis beneficial for tracking practitioner performance in the successful completion of colonoscopy, with inspection of the entire colon to the level of the cecum as a successful study. Clearly, this tool will allow us to trend individual practitioners' performance of relatively simple studies that lend themselves to easy outcome analysis, such as we have done with FAST in determining whether the study was a success or failure. The success rate can be determined by credentialing bodies, certifying bodies, or individual institutions. For our series, all practitioners performed above an 85% success rate using CUSUM analysis. Two practitioners, surgeons C and D, reached a 90% success rate. Surgeon D only demonstrated this performance level after completing 170 studies. In contrast, surgeon C, who completed 48 FAST examinations during the 30-month study, performed at the 90% level from the outset of his clinical practice with FAST. Thus, we can see that CUSUM analysis will demonstrate improvement for some practitioners with extensive experience, and others will perform at a high success rate with minimal rate or increased usage.

Dr. Evans raised a question about shift work as a variable in the application of FAST exams. All the practitioners in this study take in-house call, which begins at 5:00 PM and concludes at 7:00 AM. We did not look at the effect of our weekend 24-hour call on performance. The equipment we are using for these studies is an outdated unit that our Department of Radiology was considering removing from their daily practice. This machine has less-than-ideal imaging quality compared to today's standard equipment. We would agree with Dr. Evans's comment that the CUSUM analysis does take into consideration the surgeon's experience and understanding of anatomy, which lends itself to a very modest learning curve and good accuracy with minimal experience.