Monitoring Personalized Learning Curves for Emergency Ultrasound With Risk-adjusted Learning-curve Cumulative Summation Method

Olivier Peyrony, MD, Léa Legay, MD, Ivonne Morra, MD, Anne Verrat, MD, Hélène Milacic, MD, Jessica Franchitti, MD, Jihed Amami, MD, Ariane Gillet, MD, Kouchiar Azarnoush, MD, Arben Elezi, MD, Adélia Bragança, MD, Pierre Taboulet, MD, Pierre Bourrier, MD, Jean-Paul Fontaine, MD, and Matthieu Resche-Rigon, MD, PhD

ABSTRACT

Background: Ultrasound (US) has been a regular practice in emergency departments for several decades. Thus, train our students to US is of prime interest. Because US image acquisition ability can be very different from a patient to another (depending on image quality), it seems relevant to adapt US learning curves (LCs) to patient image quality using tools based on cumulative summation (CUSUM) as the risk-adjusted LC CUSUM (RLC).

Objectives: The aim of this study was to monitor LC of medical students for the acquisition of abdominal emergency US views and to adapt these curves to patient image quality using RLC.

Methods: We asked medical students to perform abdominal US examinations with the acquisition of 11 views of interest on emergency patients after a learning session. Emergency physicians reviewed the student examinations for validation. LCs were plotted and the student was said proficient for a specific view if his LC reached a predetermined limit fixed by simulation.

Results: Seven students with no previous experience in US were enrolled. They performed 19 to 50 examinations of 11 views each. They achieve proficiency for a median of 9 (6–10) views. Aorta and right pleura views were validated by seven students; inferior vena cava, right kidney, and bladder by six; gallbladder and left kidney by five; portal veins and portal hilum by four; and subxyphoid and left pleura by three. The number of US examinations required to reach proficiency ranged from five to 41 depending on the student and on the type of view. LC showed that students reached proficiency with different learning speeds.

Conclusions: This study suggests that, when monitoring LCs for abdominal emergency US, there is some heterogeneity in the learning process depending on the student skills and the type of view. Therefore, rules based on a predetermined number of examinations to reach proficiency are not satisfactory.

Recently, the French Society of Emergency Medicine and the American College of Emergency Physicians published clinical ultrasound (US) guidelines.[1,2] American guidelines recommend a minimum of 150 (or 25–50 for each core) US examinations for general competence.[1] Nevertheless, the learning process differs from each individual student based on his or her skills.[3] Thereby, criteria based exclusively on quantity of examinations are not fully comprehensive and one suggest the promotion of more personalized learning curves (LCs).

The majority of the studies assessing medical students' LCs for emergency US[4–6] show average results but only a few monitor the individual learning process.[7] Cumulative summation (CUSUM) methods have been used within the industrial context to monitor stabilized processes and it has been used in medical and surgical settings to monitor proficiency in procedures.[8] Biau et al.[9] adapted it to learning processes and LCs (LC-CUSUM). Due to heterogeneity among patients in medical and surgical settings possibly entailing a variable risk of failure, Steiner et al.[10] proposed to adapt the CUSUM test to the patient's characteristics (RA-CUSUM) to monitor surgical procedures. As a result, a poor outcome in a high-risk patient does not penalize the performance of the surgeon. Similarly, US examinations of some obese patients with increased sound attenuation and poor image quality may influence the learning process.

The RA-CUSUM has been briefly described for learning processes as the risk-adjusted LC CUSUM (RLC).[9] Nevertheless, it has never been applied. The aim of this study is to monitor LCs of medical students for the acquisition of abdominal emergency US and to adjust these curves to patients' characteristics and image quality using RLC.

METHODS

RLC Method

RLC tests, after each procedure, the null hypothesis that a process is "out of control" against the alternative hypothesis that a process is "in control" (when the student reaches proficiency). To plot the LCs, a score is attributed to each procedure: positive when associated with success (curve rises) and negative when associated with failure (curve decreases). This score is adjusted to patient inherent risk of failure, which depends on its characteristics.[10] Hereby, the score can take multiple values; thus a failure with a high-risk patient will be less penalizing for the student than a failure with a low-risk patient. The chart has a holding barrier at 0, which prevents the score from being negative. The score will increase from 0 as soon as the trainee's performance improves. When the score equals or exceeds a proficiency limit h, the null hypothesis is rejected and the test indicates that the process has reached an "in control state"; thereby the student has achieved proficiency. Limit h is settled through a simulation study according to the percentage of students with a specific performance the examiner wants to validate.[9,11]

Study Setting and Protocol

The first step consisted in establish a predictive model of failure (impossible US image acquisition by an expert) based on patients' characteristics. Nonetheless, we were unable to do it, due to the low rate of events (<10%). Consequently, the probability of poor image quality was chosen as a surrogate of failure. Indeed, the probability of success when a student performs an US examination is driven by image quality. Two emergency physicians performed US examinations on 100 consecutive patients and assessed their image quality ("good" or "poor"). Patients' characteristics were collected and a multivariable logistic regression was performed to investigate the association between patient characteristics and poor image quality. Increasing age and body mass index were the only characteristics associated with poor image quality in our model.

After a 10-hour learning session, medical students with no previous experience in abdominal US were asked to perform US examinations to all consecutive patients admitted in our emergency department (ED) with a pocket-size US device (VSCAN, GE Healthcare). During each examination, the student recorded video loops of 11 views of interest: subxyphoid, aorta, inferior vena cava, portal veins, portal hilum, gallbladder, kidneys, pleuras, and bladder (see Data Supplement S1, available as supporting information in the online version of this paper, which is available at http://onlinelibrary.wiley.com/doi/10.1002/aet2.10073/full).

An emergency physician with more than 2 years of emergency US practice, blinded to the student's identity and examination number, reviewed their video loops and assessed each of the 11 views of interest as a "success" (if the view allowed detection or exclusion of anomalies) or as a "failure" (e.g., absence of, not centered, truncated, or reverted view).

RLC score was then computed and LCs plotted for each student and for each of the 11 views. The student was considered proficient if his curve reached limit h.

Simulation study was performed and figures were plotted with R software version 3.1.2 (R code is provided in Data Supplement S2, available as supporting information in the online version of this paper, which is available at http://onlinelibrary.wiley.com/doi/10.1002/aet2.10073/full). The study was approved by the Biomedical Research Ethics Committee (Comité d'Evaluation de l'Ethique des projets de Recherche Biomédicale), n° 16-023.

RESULTS

Seven medical students were recruited during their emergency traineeship. After the learning sessions, they performed 237 emergency US examinations on ED patients. Each student performed a median of 35 (31–35) abdominal US examinations, which included all 11 views.

The seven students achieved proficiency for a median of 9 (6–10) views. Aorta and right pleura views were validated by seven students; inferior vena cava, right kidney, and bladder by six; gallbladder and left kidney by five; portal veins and portal hilum by four;

and subxyphoid and left pleura by three. The number of US examinations required to reach proficiency ranged from 5 to 41 depending on the student and on the type of view.

Figures 1A and 1B show plotted RLC LCs of the seven students for aorta and for subxyphoid views and illustrate heterogeneity between students. For example, albeit LCs were all crescent for aorta view (Figure 1A), student 6 needed approximately four times as many attempts to reach proficiency than student 1.

Subxyphoid view (Figure 1B) showed differences among students: two, five, and six reached proficiency with different slopes (13, 31, and 24 examinations, respectively) while the other students did not.

There was also heterogeneity between types of view performed by the same student. Figure 1C shows that student 3 had some difficulty and did not reach proficiency for kidneys and subxyphoid views while achieving proficiency with liver hilum, aorta, and inferior vena cava views.

DISCUSSION

The RLC method described in 2010 by Biau et al.[9] has never been applied to US traineeship. LCs
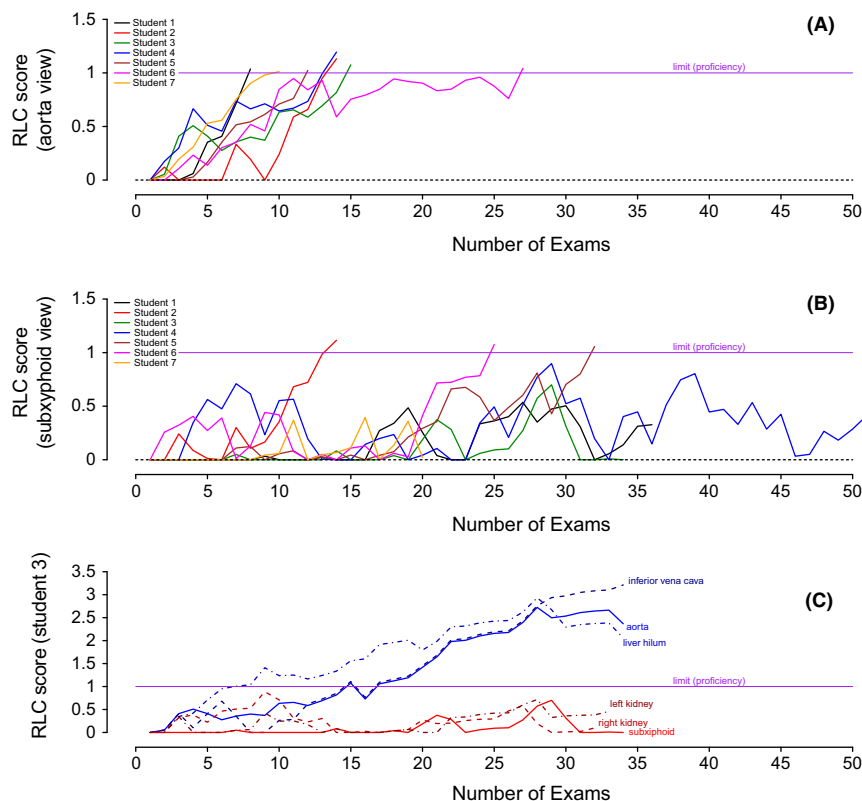
Figure 1. Learning curves with RLC of (A) all students for aorta view; (B) all students for subxyphoid view; and (C) student 3 for aorta, inferior vena cava, liver hilum, kidneys, and subxyphoid views. RLC = risk-adjusted learning curve cumulative summation. [Colour figure can be viewed at wileyonlinelibrary.com]

showed that students reached proficiency with different learning speeds. This observation strengthens our theory that rules based only on a predetermined number of examinations to reach proficiency, as recommended by guidelines,[1] are not satisfactory.

Although the use of this method may appear complex, it provides a more controlled follow-up and allows determining a more precise threshold to define proficiency. Biau et al.[11] analyzed retrospectively 532 endoscopic retrograde cholangiopancreatographies over 8 years and plot LCs with LC-CUSUM and compared it with a simple cumulative sum of failure curve. In this study, the LC-CUSUM curve indicated more precisely and sooner the moment when proficiency was reached.

The adjustment of the LCs to patient image quality is of prime interest to not penalize students in their learning process by the difficulties inherent to patients' characteristics. Steiner et al.[10] monitored a surgical procedure with CUSUM and RA-CUSUM adjusting the mortality to the patient characteristics and inherent risk of death according to the validated Parsonnet score. He showed that without adjustment, experienced surgeons had worse performances than trainee surgeons because they attended to higher-risk cases. We believe that risk adjustment based on patients' characteristics is crucial in a learning process as it may impact the outcome of the procedure.

LIMITATIONS

Our predictive model could have benefitted from a larger cohort and a validation study. Ideally, we would have directly modeled the probability of failure instead of the probability of poor image quality. The choice of the limit h, which necessitates simulation data, can be somewhat arbitrary and depends on the students' performance one whishes to validate. Consequently, the model parameters used in our department could not be generalized. Furthermore, the use of pocket-size US machines may have resulted in some difficulty linked to image acquisition and image quality. In addition, the LCs we monitored targeted the acquisition of images and not their interpretation, which is a different process and demands further expertise. However, we underline that one of the students found a peritoneal effusion in a patient who had malaria and permitted making the diagnosis of splenic rupture and immediate transfer the patient to intensive care unit.[12]

CONCLUSIONS

Our study shows some heterogeneity between student learning curves for the acquisition of abdominal emergency ultrasound views with a pocket-size device. This goes against the idea that students have to perform a predetermined number of examinations to be proficient. Even though the risk-adjusted learning-curve cumulative summation technique is complex and requires some practice, it allows personalized monitoring of learning processes with precision and strictly controlling the threshold level required to achieve proficiency. Furthermore, risk-adjusted learning-curve cumulative summation permits adjusting of the learning curves to the patient underlying risk of failure intrinsic to his or her characteristics.

References

1. American College of Emergency Physicians. Ultrasound Guidelines: Emergency, Point-Of-Care And Clinical Ultrasound Guidelines In Medicine. Ann Emerg Med 2017;69:e27–54.
2. Société Française de Médecine d'Urgence. Premier niveau de compétence pour l'échographie clinique en médecine d'urgence. Recommandations de la Société française de médecine d'urgence par consensus formalisé. Ann Fr Med Urgence 2016;6:284–95.
3. Heegeman DJ, Kieke B Jr. Learning curves, credentialing, and the need for ultrasound fellowships. Acad Emerg Med 2003;10:404–5.
4. Jang TB, Ruggeri W, Dyne P, Kaji AH. The learning curve of resident physicians using emergency ultrasonography for cholelithiasis and cholecystitis. Acad Emerg Med 2010;17:1247–52.
5. Lapostolle F, Petrovic T, Catineau J, Lenoir G, Adnet F. Training emergency physicians to perform out-of-hospital ultrasonography. Am J Emerg Med 2005;23:572.
6. Blehar DJ, Barton B, Gaspari RJ. Learning curves in emergency ultrasound education. Acad Emerg Med 2015;22:574–82.
7. Nguyen AT, Hill GB, Versteeg MP, Thomson IA, van Rij AM. Novices may be trained to screen for abdominal aortic aneurysms using ultrasound. Cardiovasc Ultrasound 2013;11:42.
8. 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 Safe Health Care 2007;16:203–7.
9. Biau DJ, Porcher R. A method for monitoring a process from an out of control to an in control state: application to the learning curve. Stat Med 2010;29:1900–9.

10. Steiner SH, Cook RJ, Farewell VT, Treasure T. Monitoring surgical performance using risk-adjusted cumulative sum charts. Biostatistics 2000;1:441–52.
11. Biau DJ, Williams SM, Schlup M, Nizard RS, Porcher R. Quantitative and individualized assessment of the learning curve using LC-CUSUM. Br J Surg 2008;95:925–9.
12. Wittwer A, Azarnoush K, Peyrony O. Malarial spontaneous splenic rupture revealed by echostethoscopy. Ann Fr Med Urgence 2016;65:337–8.

Supporting Information

The following supporting information is available in the online version of this paper available at http://onlinelibrary.wiley.com/doi/10.1002/aet2.10073/full

Data Supplement S1. Figure with the eleven ultrasound views of interest.

Data Supplement S2. R code.