Journal of Pediatric Surgery 51 (2016) 598–602

Contents lists available at ScienceDirect

Journal of Pediatric Surgery

journal homepage: www.elsevier.com/locate/jpedsurg

Cumulative sum: An individualized proficiency metric for laparoscopic fundamentals ☆

Yinin Hu [a,1], Harry L. Warren [a,1], Timothy L. McMurry [b], Robyn N. Goodrich [a], Joanna Choi [a], Adela Mahmutovic [a], Helen Kim [a], Sara K. Rasmussen [a,]*

[a] Division of Pediatric Surgery, Department of Surgery, University of Virginia School of Medicine, PO Box 800679, Charlottesville, VA, 22908, USA
[b] Department of Public Health Sciences, University of Virginia School of Medicine, PO Box 800679, Charlottesville, VA, 22908, USA

ARTICLE INFO

Article history:
Received 24 June 2015
Received in revised form 25 September 2015
Accepted 12 October 2015

Key words:
Fundamentals of laparoscopic surgery
Surgical education
Adaptive training
Medical student education
Simulation
Cumulative sum

ABSTRACT

Background: A reliable metric of technical proficiency is indispensable to the training of fellows and residents. The purpose of this study was to determine whether cumulative sum (Cusum) has predictive validity in laparoscopic training. We hypothesized that Cusum would be a better predictor of technical ability in fundamentals of laparoscopic surgery (FLS) than traditional practice volume metrics.
Methods: Twenty medical students were recruited to practice three FLS tasks: peg transfer, circle cut, and intracorporeal knot tie. Up to 7 hours of self-directed practice was allotted to each participant. Practice attempts were scored by standard FLS criteria and monitored via Cusum. Each participant's terminal Cusum performance was analyzed retrospectively. Posttests were conducted by faculty blinded to practice performance.
Results: Eighteen participants completed the study (90%). Median adjusted posttest scores were 102.3, 84.1, and 78.6 for peg transfer, circle cut, and knot tie, respectively. For the knot tie task, participants who exceeded the Cusum decision interval during their final practice attempts performed significantly higher on posttesting (81.2 vs 71.5, $p = 0.015$). Knot tie terminal Cusum score was positively associated with posttest performance after adjusting for practice volume ($p = 0.031$). Total practice volume and practice time were not significantly associated with posttest performance for any FLS task.
Conclusion: Cusum score is a more valid representation of FLS proficiency than practice volume or practice time. Incorporating Cusum in a clinical setting may promote more efficient allocation of time resources and operative volume.

Today's surgical trainees are educated within an environment of diminishing financial and time resources [1]. Residents and fellows are required to fill often competing roles as learner, clinician, teacher, mentor, and researcher. Fulfilling these roles in the setting of work hour restrictions is a persistent challenge [2]. Compounding this challenge is the need to deliver hands-on operative experiences to residents and fellows while maintaining healthcare quality and patient safety. Pediatric surgery experienced a rapid expansion of training programs during the first decade of the 21st century. With national index case volumes stagnant, this growth introduced variability in the operative experiences of graduating fellows [3,4]. As a result, a number of fellowship programs have since reduced enrollment in an attempt to bolster the operative experiences of trainees.

In addition to responsible management of trainee recruitment, a thoughtful reevaluation of the practice volume paradigm is indicated. When it comes to surgical training, volume-based metrics such as practice time or number of cases are frequently adopted because of their simplicity and objectivity [5]. However, universal volume metrics fail to account for learning rate variability. Ideally, a trainee should advance from simple to complex cases based on demonstrated proficiency in an adaptive manner. Training adequacy should not rely on universal volume criteria, but rather on a longitudinal demonstration of proficient performance. Cumulative sum (Cusum) is a promising metric that is competency based and independent of practice volume. Originally designed as a dichotomous monitoring tool for quality control purposes [6], Cusum has recently seen increasing application in medical simulation [7,8]. In the clinical setting, Cusum has been used to track proficiency in procedures ranging from bronchoscopy to epidural placement [9–12]. Cusum differentiates slow from fast learners, and can allow training protocols to conform to individualized learning rates [13].

Before Cusum can be adopted in the clinical setting, its utility as an adaptive training tool requires validation in a learning environment.

☆ Funding Source: Funding support was provided by the University of Virginia Academy of Distinguished Educators, and by NIH T32 CA163177 (to YH).

* Corresponding author at: PO Box 800709, Charlottesville, VA 22908-0679, USA. Tel.: +1 434 982 2796; fax: +1 434 243 0036.
E-mail addresses: YH9B@hscmail.mcc.virginia.edu (Y. Hu), hlw3b@virginia.edu (H.L. Warren), tmcmurry@virginia.edu (T.L. McMurry), rng3he@virginia.edu (R.N. Goodrich), ic6ad@virginia.edu (J. Choi), am6bd@virginia.edu (A. Mahmutovic), hk3vh@virginia.edu (H. Kim), skr3f@virginia.edu, SKR3F@hscmail.mcc.virginia.edu (S.K. Rasmussen).
[1] Authors Yinin Hu and Harry L. Warren contributed equally to this manuscript.

http://dx.doi.org/10.1016/j.jpedsurg.2015.10.049
0022-3468/

Surgical simulation is suitable for this purpose, and there is perhaps no more extensively tested and validated simulation platform than the fundamentals of laparoscopic surgery curriculum (FLS). Comprised of low-fidelity tasks performed on a laparoscopic simulator, FLS boasts excellent interrater reliability and correlates with intraoperative ability [14]. Training in FLS is typically conducted independently by surgery residents, with little feedback on learning progress until the testing phase. As a result, many residents erroneously allocate valuable time to overtrain on simple FLS tasks. It stands to reason that an accurate predictor of training adequacy would streamline the learning process.

The purpose of this study was to determine if Cusum analysis of self-directed FLS training is a valid predictor of technical proficiency. We hypothesized that Cusum would be a stronger predictor of performance than traditional metrics such as practice time and practice volume. Through this study, we hope to provide proof of concept that Cusum could be useful for monitoring training progress in the operative setting.

1. Materials and methods

1.1. Participants

Twenty second-year medical students without prior clinical experience volunteered for practice in three laparoscopic simulation tasks based on the McGill Inanimate System for the Training and Evaluation of Laparoscopic Skills (MISTELS): peg transfer, circle cut, and intracorporeal knot tie [15]. Participants were chosen based on schedule availability and in the order of response. A recruitment questionnaire determined that none of the participants had prior exposure to laparoscopic surgery or the FLS curriculum; thus, formal pretesting for baseline proficiency was not performed.

1.2. Simulation protocol

Enrolled students initially underwent a 3-hour group orientation session during which instrument handling and proper laparoscopic techniques were introduced. Subsequently, participants each underwent repetitive practice for up to 7 combined hours, distributed more than two months. During practice sessions, participants dictated the practice order and volume of the three FLS tasks. This participant-driven training protocol was designed to emulate laparoscopic simulation training during residency, in which a limited amount of time is self-allocated for independent practice outside of the operating room. All sessions were proctored one on one by trained assistant instructors (TAI). Each practice attempt at a given task was scored using previously reported grading criteria [15,16]. Raw scores were normalized based on calculations reported by Fraser and colleagues [17]. Feedback was permitted in between attempts. However, to maximize repetitive practice, demonstrations by TAIs were discouraged. Participants were assigned to TAIs on a rotating schedule, such that exposure to each TAI was equivalent across participants. After a maximum of 7 hours of practice, participants underwent posttesting by a surgical faculty member experienced in laparoscopic surgery who was blinded to participants' practice performance. Posttests were comprised of three attempts at each FLS task; scores across the three attempts were averaged and normalized.

1.3. Cumulative sum

The normalized score for each practice attempt for each task was first compared against "competent" score thresholds: 82.5 for peg transfer, 69.5 for circle cut, and 76 for intracorporeal knot tie. The protocol for creating Cusum learning curves using continuous data was based on methods as described by Montgomery [18]. The guiding principle of this method is to compare the standardized score of each consecutive task attempt (x) against a reference standard – in this case, the task-specific competent score thresholds – in order to assess whether a participant is performing below, on par with, or above this standard (μ_0).

We first proposed that scoring more than 5 points above or below the reference standard would be considered a meaningful difference. A reference value (K) was then set at half this value (K = 2.5). For every practice attempt performed, the participant's positive and negative Cusum scores (C^+ and C^-, respectively) would then accumulate based on deviations from μ_0 greater than this reference value. In this way, for each practice attempt, a participant's task-specific Cusum scores would accumulate based on the following equations:

$$C^+{}_1 = \max\left[0,\ x_1 - (\mu_0 + K) + C^+{}_0\right]$$

$$C^-{}_1 = \max[0, (\mu_0 - K) - x_1 + C^-{}_0]$$

For example, a score of 90 on the peg transfer task would cause the corresponding participant's C^+ score to increase by 90 − (82.5 + 2.5) = 5 points, while the C^- score would remain unchanged. In order to simplify comparative and associative statistical analyses between Cusum performance and posttest score, we assigned a single overall Cusum score (C_i) to be the difference between $C^+{}_i$ and $C^-{}_i$. When a participant's task-specific overall Cusum score is plotted throughout the course of practice, a learning curve is generated in which negative deflections indicate subpar performance and positive deflections indicate superior performance (Fig. 1A).

We separately assessed Cusum performance over each participant's final few practice attempts in order to represent learning status at the end of practice. The length of this terminal performance snapshot was determined by calculating the average run length (ARL), which is a function of K, the aforementioned meaningful difference in performance (in this case 5 points), the participant's standard deviation in performance, and the decision interval (H) beyond which a participant is considered to be performing out of control — either exceeding or deficient compared to the reference standard. For this study, we calculated each participant's ARL based on Siegmund's approximation [19], setting the decision interval to be four times the standard deviation based on preexisting recommendations [18]. In this way, we captured learning performance by calculating each participant's task-specific Cusum score during his or her terminal ARL (Fig. 1B). If a participant's ARL Cusum score surpassed the decision interval (C > H), we considered the participant to be significantly exceeding the reference standard. A participant whose ARL Cusum score was lower than the negative decision interval (C ≤ H) was considered to still be in the learning phase of practice. Finally, if a participant's Cusum score was within the decision interval (C ≤ |H|), then his or her performance was considered to be on par with the reference standard.

As a secondary analysis, we fit natural cubic splines to each subject's terminal ARL Cusum performance. We then used the fitted spline to estimate the derivative of the Cusum score at the final repetition. This derivative represents the rate of Cusum change relative to the reference standard when the subject finishes their training. A derivative of zero indicates stable performance equivalent to the reference standard; a positive derivative indicates superior performance, and a negative derivative indicates subpar performance.

1.4. Analysis

Participants who significantly exceeded the reference standard on ARL Cusum score (C > H) were compared against participants performing equivalent to or below the reference standard (C ≤ H) using Student's t test because of normal data distribution. The relationships between four independent variables – total number of practice repetitions, total practice time, ARL Cusum score, and ARL derivative – and posttest performance were assessed using Spearman correlation coefficients. To assess the impact of ARL Cusum while controlling for