

Workload assessment of surgeons: correlation between NASA TLX and blinks

Bin Zheng · Xianta Jiang · Geoffrey Tien ·
Adam Meneghetti · O. Neely M. Panton ·
M. Stella Atkins

Received: 26 January 2012 / Accepted: 20 March 2012 / Published online: 24 April 2012
© Springer Science+Business Media, LLC 2012

Abstract

Background Blinks are known as an indicator of visual attention and mental stress. In this study, surgeons' mental workload was evaluated utilizing a paper assessment instrument (National Aeronautics and Space Administration Task Load Index, NASA TLX) and by examining their eye blinks. Correlation between these two assessments was reported.

Methods Surgeons' eye motions were video-recorded using a head-mounted eye-tracker while the surgeons performed a laparoscopic procedure on a virtual reality trainer. Blink frequency and duration were computed using computer vision technology. The level of workload experienced during the procedure was reported by surgeons using the NASA TLX.

Results A total of 42 valid videos were recorded from 23 surgeons. After blinks were computed, videos were divided into two groups based on the blink frequency: infrequent group (≤ 6 blinks/min) and frequent group (more than 6

blinks/min). Surgical performance (measured by task time and trajectories of tool tips) was not significantly different between these two groups, but NASA TLX scores were significantly different. Surgeons who blinked infrequently reported a higher level of frustration (46 vs. 34, $P = 0.047$) and higher overall level of workload (57 vs. 47, $P = 0.045$) than those who blinked more frequently. The correlation coefficients (Pearson test) between NASA TLX and the blink frequency and duration were -0.17 and 0.446 .

Conclusion Reduction of blink frequency and shorter blink duration matched the increasing level of mental workload reported by surgeons. The value of using eye-tracking technology for assessment of surgeon mental workload was shown.

Keywords Endoscopy training · Simulation · Mental workloads · Eye-tracking · Surgical education · Patient safety

Presented at the SAGES 2012 Annual Meeting, March 7–March 10, 2012, San Diego, CA.

B. Zheng (✉)
Department of Surgery, 3-002F Li Ka Shing Centre for Health
Research Innovation, University of Alberta, 8440-112 St. NW,
Edmonton, AB, Canada
e-mail: bin.zheng@ualberta.ca

X. Jiang · G. Tien · M. S. Atkins
Computing Science, Simon Fraser University, Burnaby, BC,
Canada

A. Meneghetti · O. N. M. Panton
Department of Surgery, University of British Columbia,
Vancouver, BC, Canada

Surgeons face multiple challenges while performing a laparoscopic procedure [1–3]. To achieve a designed surgical goal using long-shafted surgical instruments, surgeons need to coordinate visual input with hand motions in order to control these surgical tools. The eye–hand coordination in laparoscopic surgery is more complicated than in performing an open procedure. Consequently, surgeons reported higher levels of physical and mental stresses at the end of performing a laparoscopic than an open procedure [4, 5].

Various methods have been used for assessing the mental workload of surgeons to develop strategies for managing intraoperative stresses. Some used subjective survey forms, such as NASA-TLX, to collect feedback from surgeons at the end of an operation [6–9]. Surgeons

can give a general description of the stresses they experienced during the procedure; however, it is difficult to recall the causes of the stresses. This poses difficulties when providing constructive feedback for future improvement. In addition, surgeons' answers to the survey questions can be affected by their working memory and ability in judging the task difficulty [7].

Other investigators used an information-processing model to examine surgeons' mental workload in laparoscopic surgery [10, 11]. Surgeons, as human beings, have limited capability to process multiple sources of information. When the primary surgical task requires a high level of attention, the surgeon will be left with a small amount of mental resources with which to manage a secondary task presented to his/her peripheral vision. As a result, the deterioration of the performance of the secondary task often correlates to the increasing difficulty of the primary task [10, 11]. This gives us a chance to label the stressful surgical steps and develop measures to deal with difficulties. However, the secondary task can also be recognized as a distraction to the primary task, and it is irrational for us to use it in the real operating room (OR) due to concerns about patient safety [7].

Given the drawbacks of the above two methods, several authors have turned toward using the physical response of surgeons for assessing their mental workload [5]. While surgeons are performing a surgical task, sensors placed on the chest, forehead, and ear can measure heart rate as well as respiratory and skin conductance. These physiological responses, controlled by the sympathetic nervous system, are associated with the change in the mental stress experienced by the operator [7].

To further explore the method of using physical signs for mental assessment of surgeons, we decided to include eye motions as a psychophysiological sign. Among the metrics of eye motions, blinks are an important indicator of mental activity [12–14]. Involuntary eye blinks in humans are controlled by a central mechanism that can be affected by the mental state. Changes in blink frequency and duration are correlated with the moment of mental fatigue, lapsed attention, and stress overload. This has been documented in various studies outside the context of surgery, including checking the drowsiness of individuals [15] and evaluating the mental workload of drivers [16, 17] and pilots [18].

In this study we tested the feasibility of using blink frequency and duration as a way to assess surgeons' mental workload and correlated the surgeons' eye blinks with their score on the NASA Task Load Index (NASA TLX), a known assessment instrument, based on the surgeons' comments on six domains of task requirements. We hypothesized that surgeons who display physical signs of stress (less frequent blinks and shorter blink duration)

during a laparoscopic procedure will report a high level of mental workload, assessed at the end of the procedure with the NASA TLX.

Methods

Environment and apparatus

The study was conducted in a simulated OR located in the Department of Surgery at the University of British Columbia. In the centre of this simulated OR, we set up a PC-based virtual simulator (SurgicalSim VR, METI Inc., Sarasota, FL, USA) which can create a virtual environment for training several intracorporeal procedures. Surgeons in this project were required to remove the gallbladder from attached liver tissue (laparoscopic cholecystectomy). For this procedure, the SurgicalSim displayed a series of images of human gallbladder and adjacent organs, similar to settings in our previous paper [19]. Images of the surgical site refreshed at 36 Hz in response to the surgeon's manipulation on the tissues using a pair of laparoscopic graspers and a monopolarity. The SurgicalSim did not provide haptic feedback to the surgeon, but visual feedback was close to what the surgeon receives during a real procedure. The system can record a surgeon's performance during each session and provides immediate feedback once the task is completed.

When surgeons were called to the lab, each was required to perform two simulated cases of laparoscopic cholecystectomy whilst wearing a pair of goggles which were attached to an eye-tracker (PT-Mini, Locarna System Inc., Victoria, BC, Canada). The goggles include two small cameras, one aimed toward the wearer's eye to capture the eye motions and the other pointing forward to capture the scene observed by the wearer. The video recorded by the eye camera was used for processing blink data for the study.

Participants

General surgery residents, minimally invasive surgery (MIS) fellows, and attending surgeons with various levels of experience were recruited from the Department of Surgery at the University of British Columbia for this study.

Procedure and measures

Consent was obtained from the participants before entering the study. Before starting the procedure, the surgeons needed to watch nine black dots printed on a paper poster to complete the calibration process with the Locarna eye-tracker (Fig. 1).

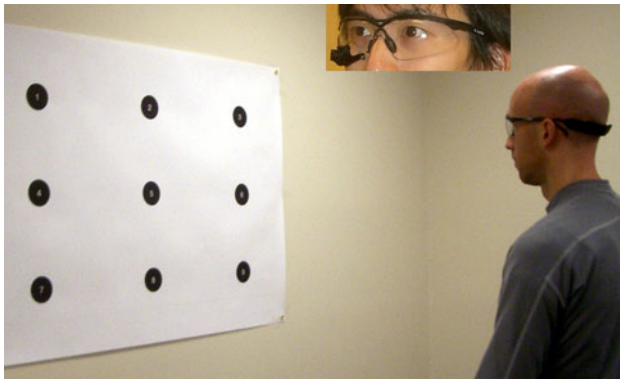


Fig. 1 Calibration process is performed on a subject before starting the trial after wearing a special goggle for tracking eye motions

During the task, the SurgicalSim VR recorded the surgical performance along with the following measures: (1) total procedure time (in seconds), (2) instrument tip trajectory (in centimeters), and (3) number of errors during the procedure, which includes misplacement of electrocautery onto the nontarget tissue, and the number of excessive traction movements on tissues. To achieve the best performance, surgeons needed to perform the procedure in a short amount of time and with caution.

While performing the simulated tasks, the surgeons' eye-motions were recorded by the head-mounted eyetracker (Locarna PT-Mini) at 30 Hz. In the postoperative phase, a custom-designed algorithm was applied to the eye-motion videos to compute the number of blinks, as we previously described [20]. The algorithm was developed on the basis of computer vision techniques. Briefly, the pupil and the eyelid were identified by the computer in each video frame. The positions of the eyelid and the pupil were compared between any two consecutive frames. The frame where the eyelid starts to drop down was annotated as the moment of a blink start. The moment that the pupil size returns to full size indicated the blink end. The blink duration was calculated as the number of video frames between blink start and end and then converting that number and reporting the duration in milliseconds. To avoid bias in eye measures, participants were not informed of blink measures during the data collection phase.

Once the simulated task was completed, the NASA TLX form was used to evaluate the workload experienced by the

surgeon. NASA-TLX requires participants to rate their perceived levels of mental, physical, and time demands associated with a task on a scale of 20 points, as well as their effort, performance, and frustration during that task. NASA-TLX has been used extensively in a variety of projects for assessing the mental workload experienced while performing both open and laparoscopic operations [7, 10, 21].

Data analyses

Descriptive analysis was conducted on the blink data. Median of blink frequency was used to divide surgeons into infrequent and frequent blinking groups. Task performance (task times, distance of tool trajectories, and number of errors) and the NASA TLX measures were compared between these two groups. Surgeons' blink frequency and duration were correlated to the outcomes of the NASA TLX reports using a Pearson test on correlation.

Results

A total of 46 surgical trials were recorded from 23 surgeons. Eye motions were not recorded in 4 trials due to the equipment failure. The descriptive analysis of blinks based on 42 valid trials is reported in Table 1. The median blink frequency (6 blinks/min) was used to divide data into two groups—those with low blink frequency ($n = 21$) and those with high blink frequency ($n = 21$). The correlation coefficients (Pearson test) between NASA TLX and the blink frequency and duration were -0.17 ($P = 0.282$) and 0.446 ($P = 0.776$), respectively. Surgeons who blinked less frequently (≤ 6 blinks/min) made a higher level of effort on performance (60 vs. 42, $P = 0.005$) and reported more frustration (46 vs. 34, $P = 0.047$) and overall level of workload (57 vs. 47, $P = 0.045$) compared with those who blinked more frequently (>6 blinks/min; see Table 2). Interestingly, surgical task performance was not associated with the blink frequency of the surgeons. Although the low-blink surgeons were more satisfied with their performance (60 vs. 43, $P = 0.005$), task time ($P = 0.439$), length of tool tip trajectory ($P = 0.925$), and number of surgical errors made in the procedure ($P = 0.671$) were not significantly different between the two blink groups.

Table 1 Descriptive statistics

	Mean \pm SD	Range	Median
Age (years) ($n = 23$ participants)	34.8 \pm 9.3	25–57	31
Years performing laparoscopic procedures ($n = 23$)	3.0 \pm 3.6	0–19	2
Blinks frequency ($n = 42$ videos, blinks/min)	6.4 \pm 5.8	0.2–25.8	6.0
Blink duration ($n = 42$ videos, ms)	181.2 \pm 71.7	100–477.8	162.7

Table 2 Group comparison

	Low blink frequency ($n = 21$) (mean \pm SD)	High blink frequency ($n = 21$) (mean \pm SD)	<i>P</i> value
NASA-TLX			
Mental demands	56.7 \pm 21.8	49.8 \pm 22.3	0.316
Physical demands	46.2 \pm 22.3	40.2 \pm 18.5	0.353
Time demands	50.2 \pm 19.5	45.5 \pm 21.1	0.452
Performance	60.0 \pm 18.1	42.9 \pm 19.1	0.005
Efforts	55.2 \pm 19.8	51.2 \pm 18.3	0.496
Frustration	46.2 \pm 20.5	33.8 \pm 18.6	0.047
Sum workload	56.8 \pm 15.0	47.3 \pm 14.6	0.045
Task performance			
Task time (s)	196 \pm 58.9	198.3 \pm 64.9	0.925
Tip trajectory (cm)	259.8 \pm 115.2	279.2 \pm 172.4	0.671
Errors (n)	14.3 \pm 11.9	10.0 \pm 6.1	0.144

Discussion

Our research hypothesis was supported by the results of our study: surgeons who made fewer and shorter blinks reported higher mental workloads through a self-reported instrument.

A surgeon's workload in the operating room can be affected by multiple factors, including hours of work and hours of sleep prior to the operation, as well as the complexity of surgical procedure [7, 11, 22]. Previous studies have found that sleep deprivation of surgeons leads to a higher level of mental workload [23], and more complex surgical procedures such as endoscopic procedures will increase the mental workload compared to routine laparoscopic procedures [24]. Our long-term goal in this line of research is to further our knowledge on what elevates surgeons' workloads and how they should cope with this factor for a safe performance. To achieve this goal, we have taken the first step, i.e., to develop a reliable tool for assessing surgeons' mental workload.

The novelty of this study is that we employed computer vision techniques for capturing eye blinks with the eye camera of a head-mounted eye-tracker. The high-quality image of the eye enabled us to compute the number of blinks and report blink durations. As we reported elsewhere, the accuracy of this algorithm can be as high as 98 % when compared to the detection by human operators using video [20].

We noticed that the correlation between NASA-TLX and blink duration and frequency is not significant (both $P > 0.05$) due to some outliers. This is quite disappointing for us. Our interpretation is based on findings from two blinks groups. We understand that separating data into two groups and comparing the means between them may limit the generalizability of our findings. Nevertheless, the

tendency was still presented for all data and was confirmed by the group difference.

There is a strong advantage in using the blink to assess surgeons' mental workload. This technology provides the opportunity to continuously monitor the change in the mental status of the surgeon during a surgical procedure. Explicitly, we can plot the change in blinks over the course of an operation, label the stressful moments, and subsequently identify which surgical tasks were associated with the stress. In this study, we did not plot the data in this way because the procedure was too short, only about 3 min on average. In addition, the task in this virtual training simulation was relatively simple, requiring only the dissecting movement from beginning to end. There were no steps in the operation that increased task difficulty for the surgeons. Although we have not had the chance to compare blinks over different surgical tasks, we will include this analysis in our next project where complex surgical procedures with multiple surgical tasks will be performed.

The second limitation of this study was that we did not record surgeons' blinks before they performed the task. Baseline blink data would help us understand individual differences and would provide the basis for making adjustments on blink changes with differing experimental conditions. We will include baseline blink data in our future study to improve the reliability of the data.

Conclusion

Eye-tracking technology can be a valuable tool for assessing the mental workload of surgeons in the operating room. As we can record the eye motion of surgeons in addition to intraoperative video, we expect to develop a way to annotate difficult steps in the procedure and

visualize the surgeon's mental state during the surgical procedure.

Acknowledgments We thank the Canadian Natural Sciences and Engineering Research Council (NSERC) and the Royal College of Physicians and Surgeons in Canada (RCPSC) Medical Education Research Grant for funding this project.

Disclosures Dr. Neely Pantou receives funds from Ethicon Endo-Surgery for fellow education and is on the scientific advisory board of Ethicon Endo-Surgery. Dr. Bin Zheng, Xianta Jiang, Geoffrey Tien, Dr. Adam Meneghetti, and Dr. Stella Atkins have no conflicts of interest or financial ties to disclose.

References

1. Tendick F, Jennings RW, Tharp G, Stark L (1993) Sensing and manipulation problems in endoscopic surgery: experiment, analysis, and observation. *Presence* 2:66–81
2. Keehner MM, Tendick F, Meng MV et al (2004) Spatial ability, experience, and skill in laparoscopic surgery. *Am J Surg* 188: 71–75
3. Cuschieri A (1995) Visual displays and visual perception in minimal access surgery. *Semin Laparosc Surg* 13:209–214
4. Berguer R, Chen J, Smith WD (2003) A comparison of the physical effort required for laparoscopic and open surgical techniques. *Arch Surg* 138:967–970
5. Berguer R, Smith WD, Chung YH (2001) Performing laparoscopic surgery is significantly more stressful for the surgeon than open surgery. *Surg Endosc* 15:1204–1207
6. Klein MI, Warm JS, Riley MA, Matthews G, Gaitonde K, Donovan JF (2008) Perceptual distortions produce multidimensional stress profiles in novice users of an endoscopic surgery simulator. *Hum Factors* 50:291–300
7. Carswell CM, Clarke D, Seales WB (2005) Assessing mental workload during laparoscopic surgery. *Surg Innov* 12:80–90
8. Cao CG (2007) Guiding navigation in colonoscopy. *Surg Endosc* 21:480–484
9. O'Connor A, Schwaartzberg SD, Cao CG (2008) How much feedback is necessary for learning to suture? *Surg Endosc* 22: 1614–1619
10. Stefanidis D, Haluck R, Pham T et al (2007) Construct and face validity and task workload for laparoscopic camera navigation: virtual reality versus videotrainer systems at the SAGES Learning Center. *Surg Endosc* 21:1158–1164
11. Zheng B, Cassera MA, Martinec DV, Spaun GO, Swanstrom LL (2010) Measuring mental workload during the performance of advanced laparoscopic tasks. *Surg Endosc* 24:45–50
12. Ryu K, Myung R (2005) Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic. *Int J Ind Ergon* 35:991–1009
13. Andreassi JL (2000) *Psychophysiology: human behavior and physiological response*, 4th edn. Lawrence Erlbaum Associates, London
14. Marshall SP (2007) Identifying cognitive state from eye metrics. *Aviat Space Environ Med* 78:B165–B175
15. Caffier PP, Erdmann U, Ullsperger P (2003) Experimental evaluation of eye-blink parameters as a drowsiness measure. *Eur J Appl Physiol* 89:319
16. Benedetto S, Pedrotti M, Minin L, Baccino T, Re A, Montanari R (2011) Driver workload and eye blink duration. *Transp Res F* 14:199–208
17. Smith P, Shah M, Lobo NDV (2000) Monitoring Head/Eye Motion for Driver Alertness with One Camera. In: *Proceedings of the 15th international conference on pattern recognition (ICPR'00)*, Barcelona, Spain, 3–7 September 2000
18. Veltman JA, Gaillard AWK (1998) Physiological workload reactions to increasing levels of task difficulty. *Ergonomics* 41: 656–669
19. Zheng B, Tien G, Atkins SM et al (2011) Surgeon's vigilance in the operating room. *Am J Surg* 201:673–677
20. Jiang X, Tien G, Huang D, Zheng B, Atkins SM (2012) Capturing eye blinks from video-based eye-tracking. *Behav Res Methods* (accepted). Paper ID: BR-Org-12-012
21. Berguer R, Smith W (2006) An ergonomic comparison of robotic and laparoscopic technique: the influence of surgeon experience and task complexity. *J Surg Res* 134:87–92
22. Howard SK, Gaba DM (2004) Trainee fatigue: are new limits on work hours enough? *CMAJ* 170:975–976
23. Dexter F, Epstein RH (2006) Holiday and weekend operating room on-call staffing requirements. *Anesth Analg* 103:1494–1498
24. Zheng B, Rieder E, Cassera MA et al (2012) Quantifying mental workloads of surgeons performing natural orifice transluminal endoscopic surgery (NOTES) procedures. *Surg Endosc*. doi: [10.1007/s00464-011-2038-x](https://doi.org/10.1007/s00464-011-2038-x)