Saccadic Eye Movement Metrics Reflect Surgical Residents' Fatigue

Leandro L. Di Stasi, PhD,†‡ Michael B. McCamy, PhD,* Stephen L. Macknik, PhD,§* James A. Mankin, MD,¶ Nicole Hooft, MD,¶ Andrés Catena, PhD,|| and Susana Martinez-Conde, PhD**

Objective: Little is known about the effects of surgical residents' fatigue on patient safety. We monitored surgical residents' fatigue levels during their call day using (1) eye movement metrics, (2) objective measures of laparoscopic surgical performance, and (3) subjective reports based on standardized questionnaires.

Background: Prior attempts to investigate the effects of fatigue on surgical performance have suffered from methodological limitations, including inconsistent definitions and lack of objective measures of fatigue, and nonstandardized measures of surgical performance. Recent research has shown that fatigue can affect the characteristics of saccadic (fast ballistic) eye movements in nonsurgical scenarios. Here we asked whether fatigue induced by time-on-duty (∼24 hours) might affect saccadic metrics in surgical residents. Because saccadic velocity is not under voluntary control, a fatigue index based on saccadic velocity has the potential to provide an accurate and unbiased measure of the resident's fatigue level.

Methods: We measured the eye movements of members of the general surgery resident team at St. Joseph's Hospital and Medical Center (Phoenix, AZ) (6 males and 6 females), using a head-mounted video eye tracker (similar configuration to a surgical headlight), during the performance of 3 tasks: 2 simulated laparoscopic surgery tasks (peg transfer and precision cutting) and a guided saccade task, before and after their call day. Residents rated their perceived fatigue level every 3 hours throughout their 24-hour shift, using a standardized scale.

Results: Time-on-duty decreased saccadic velocity and increased subjective fatigue but did not affect laparoscopic performance. These results support the hypothesis that saccadic indices reflect graded changes in fatigue. They also indicate that fatigue due to prolonged time-on-duty does not result necessarily in medical error, highlighting the complicated relationship among continuity of care, patient safety, and fatigued providers.

Conclusions: Our data show, for the first time, that saccadic velocity is a reliable indicator of the subjective fatigue of health care professionals during prolonged time-on-duty. These findings have potential impacts for the development of neuroergonomic tools to detect fatigue among health professionals and in the specifications of future guidelines regarding residents' duty hours.

Keywords: eye metrics, fatigue assessment, neuroergonomics, patient safety, saccades, shift work, surgical skills assessment, time-on-task

(*Ann Surg* 2013;00:1–6)

The use of extended work shifts and overtime has escalated at hospitals, especially among residents.[1–3] Little is known about the effects of long work hours on patients' safety,[4,5] but contrary to common assumptions,[6,7] recent evidence suggests that fatigue (ie, the mental and physical exhaustion generated by time-on-duty) does not always impair surgical performance.[8–11] Studies focused on residents[12] report conflicting results, both in terms of behavioral[13] and physiological measures.[14] These contradictions arise largely from the use of nonstandardized measures of surgical performance and from inconsistent definitions and lack of objective measures of fatigue.[2]

Traditionally, risk management experts have used subjective tests and questionnaires as the primary tool for evaluating the work demands imposed on physicians.[15,16] Such questionnaires are relatively easy to administer and interpret but have methodological caveats, including the following: (1) the usual off-line administration does not allow for continuous evaluation of physicians' performance; (2) test sensitivity is generally insufficient to signal small variations in work demand; and (3) personal and motivational factors (eg, social acceptance) can bias test scores.[17,18] Thus, the sensitivity and objectivity of work demand assessment in ecologically valid surgical scenarios remains a major challenge in the health care field.[19–21]

Objective eye movement metrics have been used to differentiate novice from expert surgeons, to quantify and to study surgeons' scanning behavior,[22–24] to assess concentration during surgical performance,[25,26] and as a measure of laparoscopic surgical skills.[27] Fatigue resulting from manipulations of the sleep/wake cycle can alter eye movements[28–30] in nonsurgical scenarios, in particular, the saccadic magnitude/peak velocity relationship.[21,31,32] No research has investigated the effect of fatigue on the eye movements of health care professionals, however. Here, we set out to determine the effects of time-on-duty on the saccadic eye movements of surgical residents. Because saccadic velocity is not under voluntary control,[33] a fatigue index based on saccadic velocity has the potential to provide an accurate and unbiased measure of a resident's fatigue level.

We measured the eye movements of general surgery residents before and after their call day while they performed 3 standardized tasks: 1 guided saccade task[21] and 2 surgical simulation tasks defined by the Fundamentals of Laparoscopic Surgery (FLS) manual skills test[34,35] (peg transfer and precision cutting). The surgical simulation tasks had different difficulty levels[36–37]: low (peg transfer) and high (precision cutting).

Time-on-duty decreased saccadic velocity and increased subjective fatigue, but it did not affect laparoscopic performance in either

From the *Department of Neurobiology, Barrow Neurological Institute, Phoenix, AZ; †Cognitive Ergonomics Group, Mind, Brain, and Behavior Research Center (CIMCYC), University of Granada, Granada, Spain; ‡Joint Center University of Granada-Spanish Army Training and Doctrine Command, Spain; §Department of Neurosurgery, Barrow Neurological Institute, Phoenix, AZ ¶St. Joseph's Hospital and Medical Center, Phoenix, AZ; and ||Learning, Emotion and Decision Group, Mind, Brain, and Behavior Research Center (CIMCYC), University of Granada, Granada, Spain.

Supported by the Barrow Neurological Foundation (to S.L.M. and S.M.-C.), the St. Joseph's Foundation (to J. M.), the National Science Foundation (awards 0852636 and 1153786 to S.M.-C), the Spanish Ministry of Economy and Finance (project PSI2012-39292 to A. C.), and the MEC-Fulbright Postdoctoral Fellowship program (grant PS-2010-0667 to L.L.D.S.).

Disclosure: The authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest, or nonfinancial interest in the subject matter or materials discussed in this manuscript. The authors declare no conflicts of interest.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.annalsofsurgery.com).

Reprints: Susana Martinez-Conde, PhD, Department of Neurobiology, Barrow Neurological Institute, 350 W Thomas Rd, Phoenix, AZ 85013. E-mail: smart@neuralcorrelate.com.

ISSN: 0003-4932/13/00000-0001
DOI: 10.1097/SLA.0000000000000260

surgical simulation task (despite performance data and subjective reports indicating higher difficulty in the precision cutting than in the peg transfer task). Within the limitations of our study, these results support the hypothesis that saccadic indices reflect graded changes in subjective fatigue, and they, moreover, indicate that resident fatigue does not result necessarily in medical error. These findings highlight the complicated relationship between continuity of care, patient safety, and fatigued providers[10] and have potential impacts for the development of neuroergonomic tools to detect fatigue among health care professionals.

METHODS

Experimental Design

The experiment included 2 levels of fatigue determined by time-on-duty: we tested subjects precall and postcall, with each experimental session being approximately 45 minutes in duration. Each session included 2 simulated laparoscopic exercises from the FLS manual skills test[34] (peg transfer task and precision cutting task), corresponding to 2 levels of task difficulty (low and high),[36–37] and 2 guided saccade tasks (see "Apparatus and Tasks" section). Potential practice/learning effects on the laparoscopic tasks were controlled by a Latin square design across both experimental sessions (half of the residents started with the postcall condition and the other half with the precall condition) and across the same session (half of the residents performed the precision cutting task before the peg transfer task and the other half did the opposite). Thus, the experimental design minimized the possible effects of confounding factors, including learning or series effects, and task-switching costs (ie, the costs associated with going from a complex task to an easy one). We will use the terms time-on-duty and fatigue interchangeably in this article.

Participants

Twelve general surgery residents (6 men and 6 women), constituting the entire trauma resident team for the 2011/2012 academic year at St. Joseph's Hospital and Medical Center (SJHMC, Phoenix, AZ), volunteered to participate in the study (mean age: 30 years; SD = 2.2 years). Participants included 4 junior, 3 middle, and 5 senior residents and received postgraduate training for an average of approximately 3 years (minimum: 2 years; maximum: 5 years). One resident dropped out at the beginning of the experiment (data not obtained). All participants had normal or corrected to normal vision. Subjects were nonsmokers and right-hand dominant. All subjects were naive to the aim of the experiment and participated on a 24-hour call day. They ran 2 experimental sessions: 1 precall (ie, rested) and 1 postcall (ie, fatigued). Their shift length ranged from approximately 13 to 18 hours (with a working time range of 61–79 hours per week). They reported an average 5 hours of sleep (range: 3–7 hours) before the precall session and 2 hours of sleep (range: 0–4 hours) before the postcall session. Before the experiment, residents filled in the Epworth Sleepiness Scale[38] and the Groningen Sleep Quality Scale[39] for screening purposes. The Epworth Sleepiness Scale scores indicated significant levels of excessive daytime sleepiness (mean score = 12.8, SD = 2.1) in half of the participants; the other half had normal levels (mean score = 5.7; SD = 3.1). Most residents scored lower than 3 in the Groningen Sleep Quality Scale (mean score = 2.27; SD = 1.95), however, indicating an optimal quality of sleep the night before their call day. No participants were eliminated on the basis of their Epworth Sleepiness Scale or Groningen Sleep Quality Scale scores. Subjects received $40 per experimental session. Written informed consent was obtained from each participant. The study was carried out under the guidelines of the SJHMC's Institutional Review Board and in conformity with the declaration of Helsinki.[40]

Apparatus and Tasks

Participants performed a guided saccade task[21] at the beginning and end of each experimental session. Subjects were instructed to follow a fixation spot on a gray computer screen. The fixation spot was composed of 2 concentric circles; the outer one was black with a 1 degree of visual angle (degree) radius and the inner one was white with a 0.25-degree radius. Residents made saccades starting from 4 randomly selected locations (each of the 4 corners of a square centered on the middle of the monitor with 22.5-degree side length) of 7 randomly selected sizes (measured from the starting location; 7.5 degrees, 10 degrees, 12.5 degrees, 15 degrees, 17.5 degrees, 20 degrees, or 22.5 degrees) and in 3 randomly selected directions (vertical, horizontal, or diagonal). There were thus $84 = 4 \times 7 \times 3$ possible guided saccades. The guided saccade sequences were exactly the same in the 2 experimental sessions, thus the saccade magnitude distributions were equal in both sessions. This task lasted 8 minutes for a total of 16 minutes of recorded data for both the precall and postcall sessions. After completing the first guided saccade task, participants performed 2 simulated laparoscopic tasks from the FLS (Society of American Gastrointestinal Endoscopic Surgery)[34] manual skills test, using the FLS training toolbox (Fig. 1). In the peg transfer task, participants used Maryland graspers to pick up 6 rings with their left hand, 1 at a time, from the left side of the pegboard, transferred them to a pair of graspers in their right hand, and placed them on the right side of the pegboard. After transferring all 6 rings from left to right, they reversed the process, transferring the rings from right to left hand and right to left side of the pegboard. The task always started with the left hand and the maximum time allowed was 300 seconds. Each transfer failure due to dropping the peg outside the field of view added a penalty of 120 seconds to the total execution time. In the precision cutting task, participants cut out the biggest circle from a 10×10 cm piece of gauze with 2 premarked concentric circles (5-cm and 6.5-cm diameter), suspended between 4 alligator clips. Residents used 1 pair of Maryland graspers and 1 pair of endoscopic scissors to perform this task. The maximum time limit was 300 seconds. Not finishing

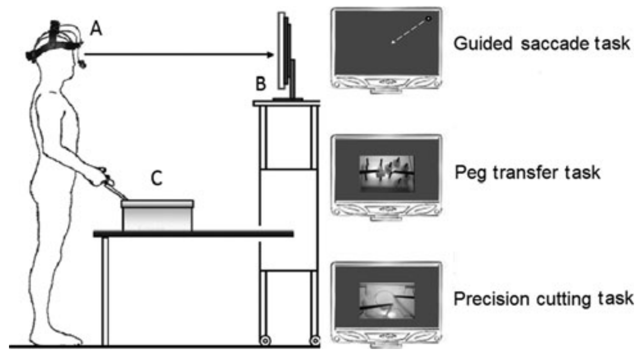

FIGURE 1. Experimental setup. Participants used the FLS box simulator for the laparoscopic tasks. The position of the simulator box was fixed (C). A USB Web camera (iREZ K2r USB; GlobalMedia Group, LLC; 640 × 480 pixel resolution) fixed to the inside of the FLS box displayed the laparoscopic task area on the center of a computer monitor (28.2 degrees × 21.1 degrees display area) facing the subject. The operating table's height was fixed at 87 cm, but participants could stand on a variable step if necessary. The monitor was mounted on a video cart and positioned about 170 cm above the floor level and about 57 cm from the participants' eyes. The EyeLink II helmet-mounted eye tracker (A) measured the subjects' eye movements. Adapted with permission from Rogers et al.[24]

the cutting added a penalty of 120 seconds to the total execution time. All experimental tasks were performed in a darkened and quiet room.

Procedure

Each precall session took place between 5:30 AM and 9:30 AM and each postcall session took place between 6:00 AM and 11:30 AM. The 2 sessions took place within an approximately 24-hour period. Previous to the beginning of the first experimental session, we recorded the age, sex, hand dominance, and years of laparoscopic-operating experience for each resident. Then, each participant watched a standardized video on how to use the FLS box. Setup of the eye-tracking system followed the training session. An instruction screen indicating the type of task to be performed preceded each experimental task, and the eye tracker was calibrated at the beginning of every task. Residents did not rest between tasks, except to answer the questionnaires described later. We used custom code and the Psychophysics Toolbox to create and display the visual stimuli.[41]

Questionnaires to Assess Fatigue Induced by Time-on-Duty

We asked participants to fill in 2 questionnaires at the beginning of each experimental session to evaluate the effectiveness of the fatigue-inducing manipulation: the Stanford Sleepiness Scale (SSS)[42] and an adapted version of the Borg rating of perceived exertion.[43] The SSS provided a global measure of sleepiness, and the Borg rating of perceived exertion indicated the level of fatigue. Residents continued to fill in the Borg scale every 3 hours from the start of their call day (8 measuring times) to control for potential diurnal fluctuations of fatigue. We used the NASA-Task Load Index (NASA-TLX) questionnaire[44] as an indicator of the degree of task complexity that participants experienced while performing the laparoscopic tasks, only during the first experimental session, irrespective of experimental condition (ie, pre/postcall session). Thus, half of the subjects filled the NASA-TLX during the postcall session and the other half during the precall session. The NASA-TLX is a multidimensional rating scale with 6 bipolar dimensions: mental demand, physical demand, temporal demand, own performance, effort, and frustration. The first 3 dimensions (mental demand, physical demand, temporal demand) reflect task-related factors such as task complexity.[44,45] Residents filled in the Self-Assessment-Manikin scale[46] (valence and arousal scales) to assess their emotional state at the beginning of each experimental session.

Eye Movement Recordings and Analyses

We sampled eye movements binocularly at 500 Hz using the Eyelink II helmet-mounted eye tracking system (SR Research, Ontario, Canada), with a resolution of 0.01-degree root mean square. See the recent work by Richstone et al[27] for a detailed description of the Eyelink II system. We identified and removed blink periods as portions of the Eyelink II raw data where pupil information was missing. We also removed portions of data where very fast decreases and increases in pupil area occurred (>50 units/sample; such periods are probably semi-blinks where the pupil is never fully occluded).[48,49] We added 200 ms before and after each blink/semi-blink to eliminate the initial and final parts where the pupil was still partially occluded.[48,49] Saccades were identified with a modified version of the algorithm developed by Engbert and Kliegl[47,50–53] with $\lambda = 6$ (threshold used for saccade detection) and a minimum saccadic duration of 6 ms. To reduce the amount of potential noise, we considered only binocular saccades, that is, saccades with a minimum overlap of one data sample in both eyes.[50-54] Additionally, we imposed a minimum intersaccadic interval of 20 ms so that potential overshoot corrections might not be categorized as new saccades.[55].

Saccadic Slope Analysis

We analyzed the slope of the saccadic peak velocity-magnitude relationship[31,50–60] to investigate the potential effects of time-on-duty on saccadic dynamics.[61] We assumed a power-law relationship—rather than a linear one—between saccadic magnitude and peak velocity, because the average r^2 was higher for the linear fits of the log-transformed data (0.93) than for the linear fits of the raw data (0.84). Thus, to obtain the slope for each peak velocity-magnitude relationship, we performed robust linear regressions (using the robust fit function in MATLAB) on the log-transformed data for each subject. That is, we did a robust linear regression on the following: $\ln(\text{peak velocity}) = m \ln(\text{magnitude}) + \text{b}$, which assumes the power-law, peak $velocity = e^b(\text{magnitude})^m$ Here, b is the y-intercept and m is the slope.

The different experimental tasks induced varying saccadic dynamics (see Supplemental Digital Content Table S2, available at http://links.lww.com/SLA/A479). To evaluate saccades with equivalent parameter (magnitude, peak and mean velocity, duration) distributions and to avoid confounding factors from potentially varying visual strategies used in each task, we compared saccades across the precall and postcall versions of the same task. That is, we compared saccades made during the precall-guided saccade task with saccades made during the postcall-guided saccade task, saccades during the precall peg transfer task with saccades during the postcall peg transfer task, and finally saccades during the precall peg precision cutting task with saccades during the postcall peg precision cutting task.

Performance Data

We used the time to complete the tasks (execution time) as an indicator of the effect of fatigue on laparoscopic procedures. Being aware that "a fast surgeon is not necessarily a good surgeon,"[62] we also assessed the residents' performance during the laparoscopic tasks. One of SJHMC's expert attendants (author J.A.M.) evaluated the recordings using a modified version of the rating scale for operative performance.[63] A single-blind process was adopted to avoid potential ranking bias.

Statistical Analyses

To analyze the effect of time-on-duty, we calculated separate single-factor mixed linear models (1 for each dependent variable: peak velocity-magnitude relationship slope, Borg, SSS, Self-Assessment-Manikin scales) with the 2 measuring times (precall vs postcall) as the within-subjects factor. To study the effect of time-on-duty on laparoscopic performance (ie, execution time and expert evaluation data) we also considered the variable task (2 task difficulty levels: peg transfer [low] and precision cutting [high]). We included the laparoscopic procedure training level and the number of years of postgraduate training as covariates in all calculated models. Finally, we conducted a single-factor repeated measures analysis of variance for the NASA-TLX data with the task as the within-subjects factor. We conducted power analyses (Graphpad Statmate version 2.0) of all results lacking statistical significance to ensure their validity, and that the subject pool size was sufficient.[64]

RESULTS

We determined the effect of time-on-duty on the saccadic eye movements, surgical performance, and subjective fatigue of surgical residents. Residents performed 3 tasks during 2 experimental sessions before and after their 24-hour call day: 2 standardized simulated laparoscopic tasks (peg transfer and precision cutting) and a guided saccade task (twice, at the beginning and at the end of each experimental session).

Effect of Time-on-Duty on Saccadic Eye Movements and Surgical Performance

Saccadic peak velocity-magnitude relationship slopes decreased with increased time-on-duty during the guided saccade task ($F_{1,10} = 5.5$, $P = 0.04$; mean precall = 0.634, SD = 0.037; mean postcall = 0.619, SD = 0.029), indicating decreased saccadic velocity with increased fatigue in surgical residents (Fig. 2 and Supplemental Digital Content Fig. S1, available at http://links.lww.com/SLA/A479; for additional analyses on the effects of time-on-duty on saccadic parameters during task performance, see Supplemental Digital Content Table S2, available at http://links.lww.com/SLA/A479). These data are consistent with previous reports of the effects of fatigue on saccadic metrics,[17,19] now applied, for the first time, to the health care field.

Time-on-duty did not affect the saccadic peak velocity-magnitude relationship slopes during the 2 laparoscopic tasks, however (all F values < 1) (for further details on the effects of time-on-duty and task complexity on saccadic parameters during task performance, see Supplemental Digital Content Table S2, available at http://links.lww.com/SLA/A479). Participants may have compensated for the effects of fatigue during laparoscopic performance by increasing their arousal.[61] This possibility is consistent with previous research showing that residents can overcome the results of fatigue during brief and familiar psychomotor tasks when they are aware that their performance will be evaluated.[12,65] In agreement with this hypothesis, peak velocity-magnitude slopes were higher during the laparoscopic tasks (mean values: 0.68, SD: 0.04) than during the guided saccade task (mean values: 0.63, SD: 0.03); for more details, see Supplemental Digital Content Table S2, available at http://links.lww.com/SLA/A479.

Laparoscopic procedure training level and years of postgraduate training were not significantly related to the change (or lack thereof) in saccadic peak velocity-magnitude relationship slope (all F values < 1), during either the guided saccade or the laparoscopic tasks.

Time-on-duty did not affect surgical performance: execution times and expert evaluations of task performance were stable across the 2 experimental sessions (see Table 1 and Supplemental Digital Content Table S1, available at http://links.lww.com/SLA/A479; all F values < 1). Power analyses revealed a power level in excess of 80% for all measures, indicating that our subject pool size was sufficiently large to support the conclusion that time-on-duty did not diminish surgical performance significantly.[64] Laparoscopic procedure training level and the number of years in postgraduate training did not significantly affect performance (all F values < 1).

Higher levels of unhappiness in sleep-deprived residents are known to be correlated to improved performance in short psychomotor tasks.[12] Accordingly, participants in the current study were less happy and less activated (lower valence and lower arousal on the Self-Assessment-Manikin scales) in the postcall session than in the precall session ($F_{1,10} = 7.6$, $P = 0.02$ and $F_{1,10} = 3.2$, $P = 0.10$; see Table 2).

Perceived task complexity was higher for the precision cutting than for the peg transfer task ($F_{1,10} = 11.7$, $P = 0.006$; NASA-TLX questionnaire: mean precision cutting = 52.65, SD = 14.44; mean peg

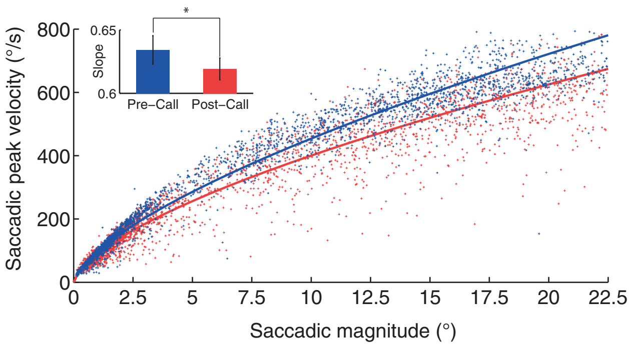

FIGURE 2. Effect of time-on-duty on the saccadic peak velocity-magnitude relationship. Peak velocity-magnitude relationships for 1 subject during the precall (blue) and postcall (red) sessions. Each dot represents 1 saccade during the guided saccade task. The curves are the power-law fits of the data from each experimental session [precall fit: peak velocity $= e^{4.58}(\text{magnitude})^{0.67}$ vs postcall fit: peak velocity $= e^{4.51}(\text{magnitude})^{0.64}$]. Insets: average slopes across all subjects for each experimental session. Error bars represent the SEM across subjects (n = 11). *indicates statistically significant differences between the precall and postcall session, $P < 0.05$.

TABLE 2. Effect of Time-on-Duty on Subjective Ratings

	Precall	Postcall
SSS*	1.73 (0.79)	3.45 (1.29)
Borg*	7.36 (1.57)	13.91 (3.59)
SAM *valence**	7.45 (1.81)	6.00 (1.61)
SAM *arousal*	4.09 (2.21)	3.00 (1.00)

SSS values range between 0 and 7. Borg's perceived exertion scale scores range between 6 and 20. For both scales, higher scores indicate more subjective tiredness. The Self-Assessment-Manikin uses a 9-point scale to rate valence and arousal. Values range between 1 and 9, with higher scores indicating higher valence/arousal. The mean and standard deviation of the SSS, BORG, and SAM scales were calculated from all residents (n = 11).

*Statistically significant differences across the precall and postcall sessions, all P values < 0.02. Nonparametric statistics (ie, Wilcoxon matched pairs tests) produced equivalent results.

SAM indicates Self-Assessment-Manikin; SSS, Stanford Sleepiness Scale.

TABLE 1. Effect of Time-on-Duty on Surgical Performance

	Peg Transfer (Precall)	Peg Transfer (Postcall)	Precision Cutting (Precall)	Precision Cutting (Postcall)
Execution time*	136.45 (52.19)	127.91 (45.63)	266.82 (89.49)	250.55 (97.24)
Surgical performance (expert evaluation)	2.6 (0.6)	2.4 (0.5)	2.2 (0.6)	2.3 (0.6)

Execution times are reported in seconds. Expert evaluation scores range between 1 and 5, with values closer to 1 indicating better performance. Measurement time (precall or postcall) did not affect execution time or surgical performance and there was no interaction between variables (all F values < 1).

*Statistically significant differences between tasks.

transfer = 33.62, SD = 19.83; values range between 0 and 100, with higher values indicating higher task complexity). Correspondingly, execution time was longer in the precision cutting than in the peg transfer task ($F_{1,10} = 32.3$, $P = 0.0002$; see Table 1).[36,37] Laparoscopic procedure training level influenced the resident's execution times in both tasks ($F_{1,8} = 5.6$, $P = 0.045$), that is, execution times were inversely related to laparoscopic procedure training level. Details on the saccadic main sequences and other saccadic parameters during laparoscopic performance are included in Supplemental Digital Content Table S2, available at http://links.lww.com/SLA/A479.

Effect of Time-on-Duty on Perceived Fatigue

Subjective fatigue (as expressed by the Borg scale) increased linearly across the 24 hours of shift ($F_{1,9} = 15.9$, $P = 0.003$—answers obtained from 10 residents out of 12). Accordingly, residents felt more tired during the postcall than the precall session ($F_{1,10} = 39.9$, $P = 0.00009$ and $F_{1,10} = 23.1$, $P = 0.00071$, respectively; SSS and Borg scales; see Table 2). Within the limitations imposed by subjective reports,[18,32] these results indicate that the time-on-duty manipulation was successful, that is, residents experienced higher levels of fatigue with time-on-duty. Laparoscopic procedure training level and years of postgraduate training were not significantly related to subjective fatigue (all F values < 1).

DISCUSSION

We examined the effect of time-on-duty on the saccadic metrics, laparoscopic performance, and subjective fatigue of surgical residents. Our results show that time-on-duty increases subjective fatigue and modulates the saccadic peak velocity-magnitude relationship (saccadic peak velocity decreased with fatigue) of surgical residents during standardized oculomotor testing, but it does not affect surgical performance or the related eye movement dynamics significantly. No previous studies have examined the effects of time-on-duty on the eye movement metrics of health professionals.

Understanding the factors that affect eye movement metrics is crucial to determining their cognitive and perceptual consequences.[66] Our study provides concrete evidence that time-on-duty affects the eye movement metrics of surgical residents. Time-on-duty modulated the saccadic peak velocity-magnitude relationship in a manner consistent with previous observations in laboratory and non–health care scenarios.[21,30] A physiologically plausible explanation for the effects of fatigue on eye movements has been provided recently.[21,32,61] Changes in attentional processing (ie, due to fatigue) can affect the strength of excitatory connections from the frontal cortex to the brainstem reticular formation,[67] thus modifying the characteristics of the peak velocity-magnitude relationship. It follows that fatigue may affect saccadic velocity via the inhibitory connections from the sleep-regulating centers and the superior colliculus to the reticular formation and cerebellum. The present experiment was conducted in conditions of partial sleep deprivation; thus variations in saccadic velocity may be related to the activation of the brain's sleep centers. Consistent with this idea, there was an increase in subjective sleepiness after the approximately 24-hour call day, associated to a reduced level of arousal. Thus, the present results support the hypothesis of a sleep/wake cycle influence on fatigue-induced decreases in saccadic velocity.[53]

Our finding that time-on-duty did not impact surgical performance or the related eye movement dynamics is consistent with the hypothesis that fatigued residents are not a main source of medical error[10,68] and highlights the complicated relationship between continuity of care, patient safety, economical factors, and fatigued medical providers.[69]

We must note that residents may compensate for high levels of fatigue during short familiar psychomotor tasks when they know that they are being tested, however,[12,65] maybe by increasing their arousal.[61] Because residents were familiar with the standardized laparoscopic tasks presented, their performance may have been more resistant to the effects of extended time-on-duty than that in alternative unfamiliar tasks or in tasks dependent on analytical, rather than visuomotor, abilities. Future research should investigate fatigued residents in novel and/or longer duration tasks that require sustained vigilance (eg, using a dual-task paradigm[70] or working memory tasks)[71,72] and/or are interrupted by unexpected emergency situations.

CONCLUSIONS

Saccadic eye movements are a reliable indicator of subjective fatigue in surgical residents after a 24-hour call day. These findings have potential impacts for the development of neuroergonomic tools to monitor fatigue in health care professionals and in the specifications of future guidelines and fit-for-duty standards regarding resident work hours.

ACKNOWLEDGMENTS

The authors thank Behrooz Kousari and Peter Wettenstein for technical assistance. They also thank Dr. Jorge Otero-Millan and Hector Rieiro for their helpful comments.

Drs. Leandro L. Di Stasi and Michael B. McCamy contributed equally to this work.

REFERENCES

1. Asch DA, Parker RM. The Libby Zion case. *N Engl J Med.* 1988;318:771–775.
2. Gaba DM, Howard SK. Fatigue among clinicians and the safety of patients. *N Engl J Med.* 2002;347:1249–1255.
3. Mastin DF, Siddalingaiah HS, Singh A, et al. Excessive daytime sleepiness, sleep hygiene, and work hours among medical residents in India. *J Trop Psychol.* 2012;2:e4.
4. Rogers A, Hwang W-T, Scott LD, et al. The working hours of hospital staff nurses and patient safety. *Health Aff.* 2004;23:202–212.
5. Sturm L, Dawson D, Vaughan R, et al. Effects of fatigue on surgeon performance and surgical outcomes: a systematic review. *ANZ J Surg.* 2011;81:502–509.
6. Waqar S, Park J, Kersey TL, et al. Assessment of fatigue in intraocular surgery: analysis using a virtual reality simulator. *Graefes Arch Clin Exp Ophthalmol.* 2011;249:77–81.
7. Hull L, Arora S, Aggarwal R, et al. The impact of nontechnical skills on technical performance in surgery: a systematic review. *J Am Coll Surg.* 2012;214:214–230.
8. Schieman C, MacLean AR, Buie WD, et al. Does surgeon fatigue influence outcomes after anterior resection for rectal cancer? *Am J Surg.* 2008;195:684–688.
9. Britt LD, Sachdeva AK, Healy GB, et al. Resident duty hours in surgery for ensuring patient safety, providing optimum resident education and training, and promoting resident well-being: a response from the American College of Surgeons to the Report of the Institute of Medicine, 'Resident Duty Hours: Enhancing Sleep, Supervision, and Safety'. *Surgery*. 2009;146:398–409.
10. Mitchell CD, Mooty CR, Dunn EL, et al. Resident fatigue: is there a patient safety issue? *Am J. Surg.* 2009;198:811–816.
11. Kramer M. Sleep loss in resident physicians: the cause of medical errors? *Front Neurol.* 2010;1:128.
12. Samkoff JS, Jacques CH. A review of studies concerning effects of sleep deprivation and fatigue on residents' performance. *Acad Med.* 1991;66:687–693.
13. Davies EC, Henderson S, Balcer LJ, et al. Residency training: the King-Devick test and sleep deprivation study in pre- and post-call neurology residents. *Neurology*. 2012;78:e103–e106.
14. Schlosser K, Maschuw K, Kupietz E, et al. Call-associated acute fatigue in surgical residents-subjective perception or objective fact? A cross-sectional observational study to examine the influence of fatigue on surgical performance. *World J Surg*. 2012;36:2276–2287.
15. Bertram DA, Hershey CO, Opila DA, et al. A measure of physician mental work load in internal medicine ambulatory care clinics. *Med Care*. 1990;28:458–467.
16. Young MS, Stanton NA. Mental workload: theory, measurement, and application. *Int Encyclopedia Ergon Hum Factors*. 2001;1:507–509.

17. Di Stasi LL, Antolí A, Gea M, et al. A neuroergonomic approach to evaluating mental workload in hypermedia interactions. *Int J Ind Ergon*. 2011;41:298–304.
18. Siegenthaler E, Bochud Y, Bergamin P, et al. Reading on LCD vs e-Ink displays: effects on fatigue and visual strain. *Ophthalmic Physiol Opt.* 2012;32:367–374.
19. Dawson D, Searle AK, Paterson JL. Look before you (s)leep: evaluating the use of fatigue detection technologies within a fatigue risk management system for the road transport industry [published online ahead of print June 21, 2013]. *Sleep Med Rev*. doi:10.1016/j.smrv.2013.03.003.
20. Lim J, Ebstein R, Tse C-Y, et al. Dopaminergic polymorphisms associated with time-on-task declines and fatigue in the Psychomotor Vigilance Test. *PLoS ONE*. 2012;7:e33767.
21. Di Stasi LL, McCamy MB, Catena A, et al. Microsaccade and drift dynamics reflect mental fatigue. *Eur J Neurosci*. 2013;38:2389–2398.
22. Nicolaou M, James A, Darzi A, et al. A study of saccade transition for attention segregation and task strategy in laparoscopic surgery. In: Barillot C, Haynor DR, Hellier P, et al., eds. *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2004*. Vol. 3217. Berlin, Germany: Springer Berlin/Heidelberg; 2004:97–104.
23. Atkins MS, Tien G, Khan R, et al. What do surgeons see: capturing and synchronizing eye gaze for surgery applications. *Surg Innov*. 2013;20:241–248.
24. Rogers ML, Heath WB, Uy CC, et al. Effect of visual displays and locations on laparoscopic surgical training task. *Appl Ergon.* 2012;43:762–767.
25. Zheng B, Cassera MA, Martinec DV, et al. Measuring mental workload during the performance of advanced laparoscopic tasks. *Surg Endosc*. 2010;24:45–50.
26. Berguer R, Smith WD, Chung YH. Performing laparoscopic surgery is significantly more stressful for the surgeon than open surgery. *Surg Endosc*. 2001;15:1204–1207.
27. Richstone L, Schwartz MJ, Seideman C, et al. Eye metrics as an objective assessment of surgical skill. *Ann Surg.* 2010;252:177–182.
28. Di Stasi LL, Renner R, Catena A, et al. Towards a driver fatigue test based on the saccadic main sequence: a partial validation by subjective report data. *Transportation Res Part C: Emerg Technol*. 2012;21:122–133.
29. Hirvonen K, Puttonen S, Gould K, et al. Improving the saccade peak velocity measurement for detecting fatigue. *J Neurosci Methods*. 2010;187:199–206.
30. Ahlstrom C, Nyström M, Holmqvist K, et al. Fit-for-duty test for estimation of drivers' sleepiness level: eye movements improve the sleep/wake predictor. *Transportation Res Part C: Emerg Technol.* 2013;26:20–32.
31. Bahill AT, Stark L. Overlapping saccades and glissades are produced by fatigue in the saccadic eye movement system. *Exp Neurol.* 1975;48:95–106.
32. Di Stasi LL, Antolí A, Cañas JJ. Main sequence: an index for detecting mental workload variation in complex tasks. *Appl Ergon*. 2011;42:807–813.
33. Leigh R, Zee D. *The Neurology of Eye Movements: Text and CD-ROM: Text and CD-ROM*. Oxford, UK: Oxford University Press; 1999.
34. Edelman DA, Mattos MA, Bouwman DL. FLS skill retention (learning) in first year surgery residents. *J Surg Res.* 2010;163:24–28.
35. Peters JH, Fried GM, Swanstrom LL, et al. Development and validation of a comprehensive program of education and assessment of the basic fundamentals of laparoscopic surgery. *Surgery*. 2004;135:21–27.
36. Ritter EM, Scott DJ. Design of a proficiency-based skills training curriculum for the fundamentals of laparoscopic surgery. *Surg Innov.* 2007;14:107–112.
37. Panait L, Akkary E, Bell RL, et al. The role of haptic feedback in laparoscopic simulation training. *J Surg Res.* 2009;156:312–316.
38. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep*. 1991;14:540–545.
39. Meijman TF, De Vries-Griever AHG, De Vries GM, et al. *The Evaluation of the Groningen Sleep Quality Scale*. Groningen, The Netherlands: Heijmans Bulletin; HB 88-13-EX; 1988.
40. World Medical Association. WMA declaration of Helsinki—ethical principles for medical research involving human subjects. http://www.wma.net/en/30publications/10policies/b3/. Updated 2008. Accessed August 21, 2012.
41. Kleiner M, Brainard D, Pelli D, et al. What's new in Psychtoolbox-3 [no. ECVP Abstract Supplement]. *Perception*. 2007;36:14. doi: 10.1068/v070821.
42. Hoddes E, Zarcone V, Smythe H, et al. Quantification of sleepiness: a new approach. *Psychophysiology*. 1973;10:431–436.
43. Borg G. *Borg's Perceived Exertion and Pain Scales*. Champaign, IL: Human Kinetics Publishers; 1998.
44. Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. *Hum Ment Workload*. 1988;1:139–183.
45. Hart SG. NASA-task load index (NASA-TLX); 20 years later. *Proc Hum Fact Ergon Soc Ann Meet.* 2006;50:904–908.
46. Bradley MM, Lang PJ. Measuring emotion: the Self-Assessment Manikin and the Semantic Differential. *J Behav Ther Exp Psychiatry*. 1994;25:49–59.
47. Engbert R, Kliegl R. Microsaccades uncover the orientation of covert attention. *Vision Res.* 2003;43:1035–1045.
48. McCamy MB, Otero-Millan J, Macknik SL, et al. Microsaccadic efficacy and contribution to foveal and peripheral vision. *J Neurosci.* 2012;32:9194–9204.
49. Troncoso XG, Macknik SL, Martinez-Conde S. Microsaccades counteract perceptual filling-in. *J Vis.* 2008;8:1–9.
50. McCamy MB, Collins N, Otero-Millan J, et al. Simultaneous recordings of ocular microtremor and microsaccades with a piezoelectric sensor and a video-oculography system. *Peer J.* 2013;1:e14.
51. Engbert R. Microsaccades: a microcosm for research on oculomotor control, attention, and visual perception. *Prog Brain Res.* 2006;154:177–192.
52. Laubrock J, Engbert R, Kliegl R. Microsaccade dynamics during covert attention. *Vision Res.* 2005;45:721–730.
53. McCamy MB, Najafian Jazi A, Otero-Millan J, et al. The effects of fixation target size and luminance on microsaccades and square-wave jerks. *Peer J.* 2013;1:e9.
54. Engbert R, Kliegl R. Microsaccades keep the eyes' balance during fixation. *Psychol Sci*. 2004;15:431–436.
55. Møller F, Laursen M, Tygesen J, et al. Binocular quantification and characterization of microsaccades. *Graefes Arch Clin Exp Ophthalmol.* 2002;240:765–770.
56. Bahill AT, Clark MR, Stark L. The main sequence, a tool for studying human eye movements. *Math Biosci*. 1975;24:191–204.
57. Robinson DA. The mechanics of human saccadic eye movement. *J Physiol.* 1964;174:245–264.
58. Collewijn H, van der Mark F, Jansen TC. Precise recording of human eye movements. *Vision Res*. 1975;15:447–450.
59. Becker W, Fuchs AF. Further properties of the human saccadic system: eye movements and correction saccades with and without visual fixation points. *Vis Neurosci.* 1969;9:1247–1258.
60. Fuchs AF, Kaneko CRS, Scudder CA. Brainstem control of saccadic eye movements. *Annu Rev Neurosci.* 1985;8:307–337.
61. Di Stasi LL, Catena A, Cañas JJ, et al. Saccadic velocity as an arousal index in naturalistic tasks. *Neurosci Biobehav Rev.* 2013;37:968–975.
62. Darzi A, Smith S, Taffinder N. Assessing operative skill. *BMJ*. 1999;318:887–888.
63. Grantcharov TP, Kristiansen VB, Bendix J, et al. Randomized clinical trial of virtual reality simulation for laparoscopic skills training. *Br J Surg*. 2004;91:146–150.
64. Cohen J. A power primer. *Psychol Bull*. 1992;112:155–159.
65. Hockey GR. Compensatory control in the regulation of human performance under stress and high workload: a cognitive-energetical framework. *Biol Psychol.* 1997;45:73–93.
66. McCamy MB, Macknik SL, Martinez-Conde S. Natural eye movements and vision. In: Werner JS, Chalupa LM, eds. *The New Visual Neurosciences. A Bradford Book.* Cambridge, MA: MIT Press; 2014:849–863.
67. Munoz DP, Everling S. Look away: the anti-saccade task and the voluntary control of eye movement. *Nat Rev Neurosci.* 2004;5:218–228.
68. Fletcher KE, Davis SQ, Underwood W, et al. Systematic review: effects of resident work hours on patient safety. *Ann Intern Med.* 2004;141:851.
69. Veasey S, Rosen R, Barzansky B, et al. Sleep loss and fatigue in residency training: a reappraisal. *JAMA*. 2002;288:1116–1124.
70. Benedetto S, Pedrotti M, Bridgeman B. Microsaccades and exploratory saccades in a naturalistic environment. *J Eye Movement Res.* 2011;4:1–10.
71. Gaunt J, Bridgeman B. Microsaccades and visual-spatial working memory. *J Eye Movement Res.* 2012;5:1–16.
72. Azuma M, Ikeda T, Minamoto T, et al. High working memory performers have efficient eye movement control systems under Reading Span Test. *J Eye Movement Res.* 2012;5:1–10.

