Assessment of surgeon fatigue by surgical simulators

Khaled Tuwairqi1
Jessica H Selter2
Shameema Sikder3
1College of Medicine, University of Utah, Salt Lake City, UT, 2Johns Hopkins School of Medicine, 3Wilmer Eye Institute, Johns Hopkins University, Baltimore, MD, USA

Correspondence: Shameema Sikder
Wilmer Eye Institute, 7315 Wisconsin Ave, Suite 610-W, Bethesda, MD 20910, USA
Tel +1 240 482 1100
Fax +1 240 482 1105
Email ssikder1@jhmi.edu

Background: The impact of fatigue on surgical performance and its implications for patient care is a growing concern. While investigators have employed a number of different tools to measure the effect of fatigue on surgical performance, the use of the surgical simulator has been increasingly implemented for this purpose. The goal of this paper is to review the published literature to achieve a better understanding of evaluation of fatigue on performance as studied with surgical simulators.

Methods: A PubMed and Cochrane search was conducted using the search terms “simulator”, “surgery”, and “fatigue”. In total, 50 papers were evaluated, and 20 studies were selected after application of exclusion criteria. Articles were excluded if they did not use the simulator to assess the impact of fatigue on surgeon performance. Systematic reviews and case reports were also excluded.

Results: Surgeon fatigue led to a consistent decline in cognitive function in six studies. Technical skills were evaluated in 18 studies, and a detrimental impact was reported in nine studies, while the remaining nine studies showed either no change or positive results with regard to surgical skills after experience of fatigue. Two pharmacological intervention studies reversed the detrimental impact of fatigue on cognitive function, but no change or a worsening effect was recognized for technical skills.

Conclusion: Simulators are increasingly being used to evaluate the impact of fatigue on the surgeon’s performance. With regard to the impact of fatigue in this regard, studies have demonstrated a consistent decline in cognitive function and mixed outcomes for technical skills. Larger studies that relate the simulator’s results to real surgical outcomes could help residency programs and policy makers in structuring more efficient training.

Keywords: surgical education, fatigue, simulator, sleep deprivation, duty hours

Introduction
Patient safety is a growing concern in contemporary clinical medicine. Provider fatigue is among the potential risk factors that adversely affect patient care through decreasing resident performance and attitude as well as increasing the rates of medical error.1,2 These findings have supported increased regulation of duty hours for physicians in training.

In 2003, the American Council of Graduate Medical Education (ACGME) stated new regulations for duty hours, limiting the trainee to a maximum of 80 hours of work per week averaged over a 4-week period.3 In 2011, an updated statement restricted the intern’s continuous duty to a maximum of 16 hours in duration. The introduction of these regulations was intended to promote patient safety and resident well-being.
However, since the introduction of these regulations, there have been concerns over the economic, educational, and patient safety implications of duty hours. Nuckols et al have estimated the annual labor cost due to the duty hour restrictions to be $1.6 billion of substitute labor.4 While many were optimistic that these restrictions would be associated with better patient outcomes and fewer medical errors, recent studies have shown conflicting results. In 2007, two studies by Volpp et al showed mixed results related to patient outcomes after the ACGME reform.5,6 The first study showed no significant improvement or worsening in mortality rate among Medicare patients after duty restrictions were instituted.7 On the other hand, the second study showed a significant improvement in the mortality rate for four medical conditions in Veterans Affairs hospitals.6 Both studies and a meta-analysis published in 2012 showed no significant change in outcomes for surgical patients after implementation of work hour restrictions.3–7 Similar variation in the literature has been found with regard to the impact of the ACGME regulations on the education and training of residents. Some studies have shown no effect of the regulations on the educational activities of interns, time spent in the operating room or clinic, number of patients on service, or knowledge-based multiple choice questions assessment.8,9 However, the studies have shown significant concerns about residents losing a central role in patient care, as well as limitations on the educational and training experience of residents.10–12

Duty hour restrictions may help reduce the level of sleep deprivation resulting in trainee fatigue. However, many factors, such as physical and emotional stress, can also contribute to fatigue.13 Therefore, the impact of surgeon fatigue on patient outcomes remains a significant issue. Researchers have recently started to utilize simulators as a more controlled environment in which to examine directly the effect of fatigue on surgeon performance without jeopardizing patient safety.

Surgical simulators have been acquired by many residency programs for training and educational purposes. In the surgical field, simulators have proven to be valid tools in differentiating the user's level of experience, helping in skill acquisition, and providing an excellent environment for evaluation of performance without potential harm to the patient.14–16 The number of studies that use simulators to measure the effect of fatigue on surgical performance is growing. The purpose of this article is to review how researchers have used simulators to assess surgeon fatigue, the impact of fatigue on simulated performance, and the utility of simulators in surgical training.

Methodology
A PubMed, EMBASE, and Cochrane search was conducted using the following keywords: “surgeon” AND “fatigue” AND “simulators”. Two reviewers independently examined each article in full (n=50 articles), evaluated the individual article’s relevance, and recorded the main findings of the study. The two reviewers then determined if the article should be included or excluded based on the article’s relevance to the topic. A score of zero or one was given to each paper; one, if relevant to the study, zero if not. Papers that received a total score of two were included, and papers with a score of one were given to a third reviewer as a tie-breaker. Inclusion criteria included papers that provided results about the effect of fatigue on the participant’s performance using a surgical simulator. Case reports and studies that did not use simulators for the purpose of assessing fatigue and performance were excluded. Studies published until October 2014 were included. Gray literature was excluded. A total of 20 studies are reviewed in this paper. A summary of these studies is shown in Table 1.

Results
Simulator studies demonstrating a decline in surgical performance with fatigue
Of the 20 studies reviewed, nine demonstrated at least some aspect of fatigue-related decline in performance on surgical simulators. However, the study design and simulator system used varied significantly across the nine studies. Two studies examined the impact of fatigue using the Minimally Invasive Surgical Training Virtual Reality (MIST-VR) system, which allows investigators to more closely simulate laparoscopic surgical skills. Eastridge et al evaluated 35 surgical residents on MIST-VR pre-call/on-call (resting state) and immediately post-call when they were acutely sleep-deprived.17 A questionnaire was given before each simulation to examine levels of fatigue. Time taken to perform tasks, number of errors, and economy of motion were recorded using the MIST-VR system. It was found that technical errors in the performance of simulated laparoscopic surgery skills was increased in the post-call period (P<0.001). Leff et al examined the effect of multiple night shifts on surgical performance of residents on the MIST-VR.18 In this study, 21 residents were trained to achieve a preset level of proficiency on the most complex of the MIST-VR tasks. After reaching proficiency, the residents...
were assessed on these tasks after 7 consecutive night shifts to examine if technical skill performance would deteriorate. The residents completed the Epworth Sleepiness Scale and a questionnaire after each night shift to examine fatigue. It was found that the residents’ surgical skills deteriorated the most after the first night shift. The residents took significantly longer to complete the tasks compared with baseline \( (P=0.002) \) and made more technical errors \( (P=0.025) \). However, improved performance on the tasks was observed after subsequent night shifts and performance came back to baseline levels.

Four of the studies were designed to examine the effect of fatigue on both cognitive and psychomotor skills of surgeons. These studies used a visuohaptic simulation as a variation of the ProMIS system designed to assess more complex cognitive skills as well as the standard psychomotor

<table>
<thead>
<tr>
<th>Reference</th>
<th>Participants and simulator</th>
<th>Parameters evaluated</th>
<th>Drug intervention</th>
<th>Significant outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eastridge et al(^7)</td>
<td>35 surgical residents, MIST-VR simulator</td>
<td>Psychomotor skills</td>
<td>No</td>
<td>Technical errors increased</td>
</tr>
<tr>
<td>Leff et al(^8)</td>
<td>21 residents, MIST-VR simulator</td>
<td>Psychomotor skills</td>
<td>No</td>
<td>More technical errors and deterioration on first night</td>
</tr>
<tr>
<td>Kahol et al(^9)</td>
<td>37 surgical residents, visuohaptic simulator</td>
<td>Psychomotor and cognitive skills</td>
<td>No</td>
<td>Deterioration of cognitive functions</td>
</tr>
<tr>
<td>Kahol et al(^10)</td>
<td>7 surgical residents, visuohaptic simulator</td>
<td>Psychomotor skills</td>
<td>No</td>
<td>Decreases in surgical proficiency and increased errors</td>
</tr>
<tr>
<td>Brandenberger et al(^11)</td>
<td>14 surgical residents, visuohaptic simulator</td>
<td>Psychomotor and cognitive skills</td>
<td>No</td>
<td>More cognitive errors and declines in performance of night float residents</td>
</tr>
<tr>
<td>Gerdes et al(^12)</td>
<td>Nine attending surgeons and five residents, visuohaptic simulator</td>
<td>Psychomotor skills and cognitive skills</td>
<td>No</td>
<td>All had overall performance decline; attendings had less decline in cognitive errors</td>
</tr>
<tr>
<td>Daruwalla et al(^13)</td>
<td>114 participants, LapSim (FLS) simulator</td>
<td>Psychomotor skills</td>
<td>No</td>
<td>Decrease in overall score per day elapsed, performed worse on certain laparoscopic skills</td>
</tr>
<tr>
<td>Naughton et al(^14)</td>
<td>20 residents, endovascular skills simulator</td>
<td>Psychomotor skills</td>
<td>No</td>
<td>It took one day longer for night shift residents to reach plateau in proficiency</td>
</tr>
<tr>
<td>Yamany et al(^15)</td>
<td>13 residents, Da Vinci Skills simulator</td>
<td>Psychomotor skills</td>
<td>No</td>
<td>Time to complete tasks was significantly increased when residents were fatigued</td>
</tr>
<tr>
<td>Tomasko et al(^16)</td>
<td>31 medical students, MIST-VR and Rapid-fire simulators</td>
<td>Psychomotor skills</td>
<td>No</td>
<td>No significant impact of fatigue on completing old and new tasks</td>
</tr>
<tr>
<td>Uchal et al(^17)</td>
<td>64 surgeons, MIST-VR simulator</td>
<td>Psychomotor skills</td>
<td>No</td>
<td>No significant impact of fatigue on performance</td>
</tr>
<tr>
<td>Waqar et al(^18)</td>
<td>Seven ophthalmology surgeons, EyeSi simulator</td>
<td>Psychomotor skills</td>
<td>No</td>
<td>Improvements in total score and total time taken to complete the simulator task</td>
</tr>
<tr>
<td>Erie et al(^19)</td>
<td>Nine ophthalmology residents, EyeSi simulator</td>
<td>Psychomotor skills</td>
<td>No</td>
<td>No significant impact of fatigue on performance</td>
</tr>
<tr>
<td>Yi et al(^20)</td>
<td>Nine surgery residents, Simbionix simulator</td>
<td>Psychomotor skills</td>
<td>No</td>
<td>No significant impact of fatigue on performance</td>
</tr>
<tr>
<td>Jakubowicz et al(^21)</td>
<td>Eight surgery residents, ESS simulator</td>
<td>Psychomotor skills</td>
<td>No</td>
<td>No significant impact of fatigue on performance</td>
</tr>
<tr>
<td>Schlosser et al(^22)</td>
<td>38 surgical residents, LapSim simulator</td>
<td>Psychomotor skills</td>
<td>No</td>
<td>Residents had better performance post call on more complex virtual reality tasks</td>
</tr>
<tr>
<td>Olasky et al(^23)</td>
<td>22 surgical residents and attendings, FLS simulator</td>
<td>Psychomotor skills</td>
<td>No</td>
<td>Sleep hours and fatigue did not affect performance</td>
</tr>
<tr>
<td>Di Stasi et al(^24)</td>
<td>12 surgical residents, FLS simulator</td>
<td>Psychomotor skills</td>
<td>No</td>
<td>No significant impact of fatigue on performance</td>
</tr>
<tr>
<td>Sugden et al(^25)</td>
<td>39 residents, MIST-VR simulator</td>
<td>Psychomotor and cognitive skills</td>
<td>Yes (modafinil)</td>
<td>Modafinil did improve cognitive performance, no improvement in psychomotor skills</td>
</tr>
<tr>
<td>Aggarwal et al(^26)</td>
<td>18 medical students, MIST-VR simulator</td>
<td>Psychomotor and cognitive skills</td>
<td>Yes (caffeine and taurine)</td>
<td>Caffeine and taurine did improve cognitive performance, more errors on technical skills in comparison with rested state</td>
</tr>
</tbody>
</table>

skills usually assessed in simulation. The simulator assessed gesture proficiency, smoothness of hand movement, smoothness of tool movement, time taken to complete the task, and cognitive errors. Although all these studies used the ProMIS system, they differed in study design and focus. Kahol et al used visuoaptic simulation to examine more extensively the cognitive errors that can occur as a result of fatigue in residents after call. In their study, 37 surgical residents were assessed before call and after call. Post call, the residents made more cognitive errors and showed decreased proficiency in areas of memory, attention, and intermodal coordination (P<0.01). Another study by Kahol et al examined seven surgical residents who performed this visuoaptic simulation before and after call. The residents also wore electroencephalography caps while performing the tasks to examine the immediate impact of workload, distraction, and attention on performance. The authors found significant decreases in surgical proficiency after call (P<0.015) along with an increased number of errors, and electroencephalography showed significant decreases in attention (P<0.015) associated with the errors. Brandenberger et al examined night float residents and compared their performance with that of day shift residents using visuoaptic simulation both before and after shifts. Fatigue and sleep deprivation measures were assessed via questionnaire. Both night and day shift residents showed decreased proficiencies (both psychomotor and cognitive) after call (P<0.05), but night float residents showed greater declines in overall performance (P<0.05). Post call, cognitive errors were higher in the night shift residents compared with the day shift residents (P<0.01). Gerdes et al examined if there was a difference in effect of fatigue on surgical proficiency in residents compared with surgical attendings. In this study, nine attending trauma surgeons and five trauma surgery residents were assessed on the simulator before and after call. A questionnaire was used to measure fatigue levels at the time of assessment. The study found that both attending surgeons and surgical residents had a decline in overall performance post call, but attendings showed less of a decline in cognitive errors compared with residents (P<0.05). Both attending surgeons and surgical residents showed an equal decline in psychomotor skills. 

Daruwalla et al used an FLS (Fundamentals of Laparoscopic Surgery) or LapSim simulator to compare the effect of fatigue on surgical performance between a study cohort and historical non-fatigued controls. In this study, 114 participants ranging from medical students to experienced surgeons (73 study participants and 41 historical controls) were taught to reach a preset level of proficiency on the FLS or LapSim simulator. The participants were then retested on the simulator tasks approximately one month later. The study cohort was retested directly after a 10-hour work shift, while the historical controls were tested at a time of their own choosing. Fatigue was assessed using the Epworth Sleepiness Scale. The fatigued participants exhibited a decrease in overall score per day elapsed compared with the historical non-fatigued participants (P=0.02). Fatigued participants performed worse on certain laparoscopic skills, including peg transfer, extracorporeal knot tying, and clip application. 

Naughton et al compared ten residents working night shifts with 10-day shift residents on an endovascular skills simulator to determine if the night shift workers could obtain a similar proficiency in skill as the day shift workers. The residents completed a validated renal artery stent module on an endovascular simulator over a series of seven sequential shifts. It was found that the night shift workers did obtain proficiency on the module similar to the daytime workers, but took one day longer to reach plateau than day shift training.

Yamany et al utilized the da Vinci Skills simulator to assess surgical residents on a suturing task before and after a 2-hour call. Residents who were post-call took significantly longer (P<0.05) to complete the simulator task.

**Simulator studies demonstrating no impact or improvement in surgical performance with fatigue**

Of the 20 studies evaluated, nine demonstrated that fatigue had either no impact or led to improved performance on surgical simulators. However, there was a wide variety of types of simulators and study designs used to evaluate fatigue. Two of the nine studies utilized the MIST-VR simulator to assess the impact of fatigue on performance. Tomasko et al trained 31 medical students to proficiency on the MIST-VR simulator. The students were then randomized to a sleep deprivation group or a control group. On the next day, the students’ performance was assessed on a novel task. Fatigue was assessed using the Epworth Sleepiness Scale. The study found that the sleep deprivation group was able to complete both the old task and the new task just as well as controls despite an increase in sleepiness, but used an increase in mental workload to do so. Uchal et al performed a randomized controlled trial to examine the effect of sleep deprivation and fatigue on the technical performance of surgeons. Surgeons post a 24-hour duty call were compared with surgeons post an 8-hour work day on a MIST-VR simulator task that involved suturing a perforated ulcer on a foam stomach. Fatigue was assessed using the Epworth Sleepiness Scale. The study showed no
significant difference between the two surgeon groups with regard to product quality or procedural effectiveness on the simulator task.

Two of the nine studies specifically used the EyeSi simulator to examine the impact of fatigue on performance of ophthalmic surgery. In one study, Waqar et al assessed the performance of seven ophthalmic surgeons on the EyeSi simulator before and after they completed their scheduled theater lists. The study found that the surgeons had significant improvements in total score ($P=0.028$) and total time taken to complete the simulator task ($P=0.033$) after completing their theater lists. No deleterious effect of fatigue on surgical performance was noted on the EyeSi simulator.

Erie et al used the EyeSi simulator to measure the effect of sleep deprivation on performance of surgical skills in nine ophthalmology residents. The residents were assessed using a forceps and anti-tremor module and fatigue was assessed using the Epworth Sleepiness Scale. Although sleepiness did increase significantly ($P<0.001$), there were no statistically significant differences in technical performance on the simulator modules after call as compared with before call.

One study utilized the Simbionix Lap Mentor stimulator, which is used to teach and assess skills in basic surgical procedures. In the study, residents on night float were compared with residents on 24-hour call to assess the impact of fatigue on procedural skills using the simulator. Neither night float nor 24-hour call residents showed decreases in psychomotor performance after call.

Jakubowicz et al utilized the Endoscopic Sinus Surgery (ESS) simulator instead of the MIST-VR as a more complex simulator to assess surgical residents on their ability to learn a complete procedure as opposed to a more simplified task. The study used the ESS to examine the impact of fatigue on the performance of surgical residents. Eight general surgery residents were trained on the ESS, and then tested before and after on-call duties while complying with ACGME working hour mandates. The residents completed questionnaires to assess their level of fatigue. The study found no statistically significant differences in the number of errors, time to task completion, or overall performance of the residents before call versus after call.

One study by Schlosser et al used the LapSim virtual reality simulator to examine the technical performance of 38 surgical residents before and after a 24-hour call. The study also examined physiological measures, such as cortisol and pupillary activity, to determine if physiological correlates of fatigue correlated with subjective experiences of fatigue by the surgeons. The Stanford Sleepiness Scale was used to assess fatigue. The study found that there was no difference post call in physiological measures of fatigue or in performance on lower-complexity simulator tasks. However, residents actually had better performance post call on the more complex virtual reality tasks ($P<0.044$).

Two studies utilized the FLS trainer in their assessment of surgeon fatigue. Olasky et al assessed both surgical residents and attendings using peg transfer tasks on the FLS simulator and compared their performance with their hours of sleep and self-reported fatigue levels. Performance on the simulator was found not to be affected by the number of hours of sleep. Di Stasi et al used the FLS trainer but also aimed to determine if saccadic eye movement metrics would reflect fatigue in surgical residents. To perform the assessment, 12 surgical residents wore a head-mounted video eye tracking device while performing simulated laparoscopic surgery tasks throughout their shifts. The study found that a longer time on duty led to decreased saccadic velocity ($P=0.04$), but no difference in simulated laparoscopic surgery performance.

**Simulator studies assessing impact of treatments for fatigue on surgical performance**

Two of the 20 studies evaluated the use of MIST-VR simulators to examine pharmacological treatments for fatigue affecting surgical performance. Sugden et al performed a randomized controlled trial of modafinil versus placebo to examine the impact of modafinil on resident performance when sleep-deprived. Modafinil is a prescribed drug (200 mg) that has been used in the treatment of shiftwork sleep disorder and has been shown to ameliorate the cognitive effects of fatigue in healthy sleep-deprived individuals. In the study, 39 residents received either placebo or modafinil after a night of sleep deprivation and were then assessed using both neuropsychological tests and the MIST-VR system. While modafinil did improve performance on cognitive tasks undertaken during the neuropsychological tests, there was no difference between groups on MIST-VR psychomotor performance. Aggarwal et al performed a single-blind crossover study to measure the outcomes of using pharmacologic stimulants (caffeine 150 mg and taurine 2 g) in 18 sleep-deprived medical students. Taurine is an over-the-counter drug that is often found in combination with caffeine in the energy drink market. The participants were evaluated by neurocognitive tests and psychomotor skills performance on the MIST-VR simulator. Participants taking placebo took longer ($P=0.016$), made more errors ($P=0.021$), and were less economical with movement ($P=0.016$) compared
with baseline. However, those who took caffeine and taurine had similar reaction times on neurocognitive tasks compared with baseline. Those taking caffeine still made significantly more errors compared with baseline ($P=0.046$).

**Discussion**

We have reviewed the simulator-based studies that assessed the effect of fatigue on surgeon performance. However, there are multiple limitations that restrict the ability to compare the outcomes of these studies. These limitations include: examining multiple surgical specialties that vary from study to study; different kinds of simulators, with each device having its own tasks and scoring systems; variation in complexity from task to task; and variation in the experience level of study participants.

Consistent declines in cognitive function were seen in the included studies. Six of the reviewed studies showed a detrimental impact of fatigue on cognitive function using visuohaptic and MIST-VR simulators. A reversed action of fatigue that can lead to improvement in task performance has been described in the literature, including coping strategies allowing the operator to focus their attention on the task to be performed and ignore distractions. Environmental factors, such as motivation and working conditions, could have a similar impact.

Evaluation of the impact of fatigue on technical performance yielded mixed results. While nine of the included studies showed a detrimental impact of fatigue on technical performance, nine further studies had findings that showed no impact of fatigue. A reversed action of fatigue that can lead to improvement in task performance has been described in the literature, so it is interesting to note that the same phenomenon was noted in one of the studies. An exploration of the circumstances that produced this finding, and studying its sustainability may enhance clinicians’ daily practice.

The variable effect of fatigue on psychomotor skills could be explained by the nature of the procedures that were performed. For example, in the field of ophthalmology, two studies showed no impact of fatigue on surgeon performance. Both studies utilized tasks that are part of the cataract surgery training modules. A consideration of such results may help modify training and duty hour structure according to the specialty and task complexity.

The night float system has been implemented to comply with the recent AGME duty hour reforms. Recent studies reported a negative impact on patient safety related to this system. Three of the reviewed studies showed a similar detrimental impact of fatigue on performance and cognitive errors secondary to this system, with the deterioration being most marked on the first night and a delay in skill acquisition by the trainee. However, a study by Yi et al showed no significant impact on psychomotor performance related to this system.

Understanding of the relationship between doctor fatigue and patient outcomes is key to regulating and organizing health care services. While most studies have focused on sleep deprivation, it is important to consider all the factors that may contribute to fatigue, such as emotional and physical distress, or suffering from a medical condition. It is also important to emphasize that simulator performance is intended to be an objective surrogate to correlate with real-practice outcomes. Therefore, studies that are able to correlate simulator outcomes with actual patient outcomes are needed. For example, Gordon et al found that medical internship with extended shifts had a consistently negative impact on simulated case performance and hospital outcomes data. Findings such as these help provide greater validity to simulator studies. However, in the surgical field, the reviewed studies fell short of providing such a clear correlation.

**Future opportunities**

The growing number of simulator devices in the surgical field should encourage developers to create a standardized scoring tool and task complexity classification. Additionally, studying the impact of fatigue according to specialty with inclusion of different factors that may lead to fatigue could help tailor the regulations in a more specific way.

In order to validate the simulator results, further studies must correlate simulation with real practice data. Further, in the future, the study of fatigue in the surgical team and its impact on coordination between team members should be explored. Researchers are also encouraged to study the safety and outcomes of reversing declines in cognitive function with pharmacological interventions. Eventually, larger multicenter studies will be required to give a better understanding of the relationship between fatigue, simulation, and patient outcomes.

**Conclusion**

Simulators have provided a safe environment to assess the impact of fatigue on the surgeon performance. A consistent...
Decline in cognitive function related to fatigue has been recognized, which could be reversed by pharmacological intervention. A variable impact of fatigue on technical performance has been established. Further simulator development, larger studies in each surgical field, and correlation of results with clinical outcomes should give better understanding and enable planning for training programs and by regulatory organizations.

Disclosure
The authors report no conflicts of interest in this work.

References

