Does Sleep Deprivation Impair Orthopaedic Surgeons’ Cognitive and Psychomotor Performance?

Michael J. O’Brien, MD, Robert V. O’Toole, MD, Mary Zadnik Newell, ScD, OTR/L, MEd, Alison D. Lydecker, MPH, Jason Nascone, MD, Marcus Sciadini, MD, Andrew Pollak, MD, Clifford Turen, MD, and W. Andrew Eglseder, MD

Investigation performed at R Adams Cowley Shock Trauma Center, University of Maryland, Department of Orthopaedics, Baltimore, Maryland

Background: Sleep deprivation may slow reaction time, cloud judgment, and impair the ability to think. Our purpose was to study the cognitive and psychomotor performances of orthopaedic trauma surgeons on the basis of the amount of sleep that they obtained.

Methods: We prospectively studied the performances of thirty-two orthopaedic trauma surgeons (residents, fellows, and attending surgeons) over two four-week periods at an urban academic trauma center. Testing sessions used handheld computers to administer validated cognitive and psychomotor function tests. We conducted a multivariate analysis to examine the independent association between test performance and multiple covariates, including the amount of sleep the night before testing.

Results: Our analysis demonstrated that orthopaedic surgeons who had slept four hours or less the night before the test had 1.43 times the odds (95% confidence interval, 1.04 to 1.95; p = 0.03) of committing at least one error on an individual test compared with orthopaedic surgeons who had slept more than four hours the previous night. The Running Memory test, which assesses sustained attention, concentration, and working memory, was most sensitive to deterioration in performance in participants who had had four hours of sleep or less; when controlling for other covariates, the test demonstrated a 72% increase in the odds of making at least one error (odds ratio, 1.72 [95% confidence interval, 1.02 to 2.90]; p = 0.04). No significant decrease in performance with sleep deprivation was shown with the other three tests.

Conclusions: Orthopaedic trauma surgeons showed deterioration in performance on a validated cognitive task when they had slept four hours or less the previous night. It is unknown how performance on this test relates to surgical performance.

Sleep deprivation can have serious effects on health, manifested as physical and mental impairment. Inadequate rest impairs the ability to think and may cloud judgment and may slow reaction time. Concern has led to calls for work-hour restrictions for resident physicians. However, to our knowledge, no previous work has examined the effects of sleep deprivation on the cognitive and psychomotor performances of orthopaedic surgeons.

Orthopaedic surgeons managing patients with high-energy fractures may be at particular risk for the harmful effects of sleep deprivation. These surgeons may spend substantial amounts of time awake at night while operating and may also operate throughout the next day after being up all night at the hospital. It is unknown how the amount of sleep deprivation typically encountered in a modern trauma center affects performance.

Disclosure: One or more of the authors received payments or services, either directly or indirectly (i.e., via his or her institution), from a third party in support of an aspect of this work. In addition, one or more of the authors, or his or her institution, has had a financial relationship, in the thirty-six months prior to submission of this work, with an entity in the biomedical arena that could be perceived to influence or have the potential to influence what is written in this work. No author has had any other relationships, or has engaged in any other activities, that could be perceived to influence or have the potential to influence what is written in this work. The complete Disclosures of Potential Conflicts of Interest submitted by authors are always provided with the online version of the article.

This article was chosen to appear electronically on September 26, 2012, in advance of publication in a regularly scheduled issue.
Our purpose was to prospectively study the cognitive and psychomotor performances of orthopaedic surgeons on the basis of the amount of sleep that they had obtained. Our hypothesis was that cognitive and psychomotor performances of orthopaedic surgeons decline when they are sleep-deprived.

**Materials and Methods**

**Participants**

After obtaining approval from our institutional review board, we prospectively tested thirty-two orthopaedic surgeons at an urban academic trauma center. Informed consent was obtained from all participants. The participants included sixteen orthopaedic residents (postgraduate years three and four), ten orthopaedic trauma fellows (postgraduate year six), and six orthopaedic trauma attending surgeons (see Appendix). All participants were blinded to the results of the study at all times. Monetary prizes were awarded to the resident and to the fellow or attending surgeon with the best overall scores as incentives to perform well on the tests. The prizes were awarded at completion of the testing periods.

**Testing Protocol**

The study was conducted during two separate testing periods. The testing periods encompassed thirty days in January and February 2008 and twenty-six days in September and October 2008. The tests were administered every morning at the start of the 7 A.M. trauma conference. The same instructions were read every day before the test was administered. We administered each individual test each testing day to all available participants to establish construct validity. Not all participants were available for each testing session. The number of participants per session ranged from five to twenty-six.

Every morning, we recorded how many hours of sleep that each participant had obtained the previous night. We recorded whether the participant was on call and whether he or she had slept at the hospital.

Individual handheld computers with validated software were used. All participants were tested at the same time in a quiet room at 7 A.M. each day. The testing protocol lasted approximately ten minutes per session.

The testing protocol was first used in January and February 2008. Because our trauma service tends to be less busy during the winter, the attending surgeons had very few sleep-deprived nights. On the basis of the initial experience, we eliminated some components of the test (see Appendix) and used the test again in September and October 2008. Because our trauma service tends to be busier during the fall, the attending surgeons were more likely to be sleep-deprived. The same six attending surgeons participated in both testing runs, but the fellows and residents were different between the two testing groups.

**Neuropsychological Assessment Software**

The cognitive function software used was developed by BrainCheckers (Behavioral Neuroscience Systems, Springfield, Missouri). The Automated Neuropsychological Assessment Metrics (ANAM) system was the source for developing the BrainCheckers battery. ANAM has been shown to be a reliable and valid tool for assessing cognitive performance and compares favorably with classic neuropsychological tests.

An ANAM battery is a collection of several specific tests that assess basic functions of cognition, such as attention, reaction time, memory, and concentration. A standard desktop personal computer is required for running ANAM. A version of ANAM called the ANAM Readiness Evaluation System (ARES) was created for the Palm Operating System (Palm, Sunnyvale, California). ARES has since evolved into BrainCheckers in validity and reliability testing, scores on the ANAM and ARES platforms correlated highly. The ANAM system has also been previously used and validated with sleep deprivation studies.

**Analysis**

All six attending surgeons participated in both testing periods. Before conducting the analysis, it was decided to include only the data from the second testing period for individual participants who participated in both testing periods because the surgeons were more likely to be sleep-deprived during that time of the year on the basis of our center’s historical referral patterns. For each participant, the first three attempts on each test were not included in the analysis because they were thought to represent the learning curve for this test.

The cognitive function battery for both testing periods included running memory, spatial processing, and simple reaction time. The Mathematical Processing and Pursuit Tracking tests were administered only during the first testing period, and the Matching to Sample test was administered only during the second testing period. The Pursuit Tracking test was eliminated from the second testing session because there was a ceiling effect, as the scores were very close to 100%. Further, this test did not have a potential for error and thus would not have been included in our final analysis, as described below. The second eliminated test was the Mathematical Processing test, which was eliminated in favor of the Matching to Sample test because the Matching to Sample test, as described below, is more likely than the Mathematical Processing test to relate to surgical performance. A brief description of each test follows.

The Running Memory test assesses sustained attention, concentration, and working memory. A single digit appears briefly on the testing screen. The task requires the participant to indicate whether the current number matches the one previously displayed. The “SAME” button is to be pressed if the numbers match and the “DIFF” button is to be pressed if the numbers are different. The number matches on 50% of the trials, and a number is not presented in more than three consecutive trials. Numbers range from zero to nine, with each digit being equally likely.

The Matching to Sample test assesses spatial processing and visuospatial working memory. It also assesses attention, short-term memory, and spatial discrimination of a visual field. The participant is presented with a preliminary screen that shows an image with a prompt to memorize that exact image (see Appendix). Next, the participant is presented with two options side by side (see Appendix). The option that matches the image on the previous screen needs to be identified and tapped.

The Spatial Processing test assesses spatial processing and visuospatial working memory. It pairs histograms, each containing four bars, are presented simultaneously on the monitor. One histogram is rotated either 90° or 270° to the other (see Appendix). The participant presses designated buttons to indicate whether the histograms are the same or different, regardless of the orientation.

The Simple Reaction Time test assesses reaction time. The task requires the participant to tap on a stimulus as fast as possible. The stimulus is an asterisk (*) that appears on the screen twenty times. The participant is instructed to tap inside a dotted box each time the stimulus is presented. This test provides a measure of pure visual reaction time and a means to differentiate between motor response speed and actual cognitive processing time.

The Mathematical Processing test assesses basic computational skills, concentration, and working memory. The task was designed to assess working memory by requiring the participant to perform basic arithmetic operations. All problems require an addition and subtraction sequence in the form of “X + Y – Z” or “X – Y + Z.” The participant indicates whether the solution to the problem is greater than or less than five. The numbers are selected at random, only digits one through nine are used, and the correct answer may be any number from one through nine except five. Test problems with answers greater than five or less than five are equally probable. Working from left to right, the same digit cannot appear twice.

The Pursuit Tracking test is a tracking and fine motor abilities test that assesses visuomotor control. It is designed to detect impairments in fine motor abilities that are evident in neuromuscular disorders such as Parkinson disease and other similar conditions that include intentional tremor, familial tremor, and side effects of medication. It requires pressing the stylus on the screen and following a moving target or bull’s-eye across the screen. The target follows a sine wave.

The cognitive function battery for both testing periods included running memory, spatial processing, and simple reaction time. The Mathematical Processing and Pursuit Tracking tests were administered only during the first testing period, and the Matching to Sample test was administered only during the second testing period. The Pursuit Tracking test was eliminated from the second testing session because there was a ceiling effect, as the scores were very close to 100%. Further, this test did not have a potential for error and thus would not have been included in our final analysis, as described below. The second eliminated test was the Mathematical Processing test, which was eliminated in favor of the Matching to Sample test because the Matching to Sample test, as described below, is more likely than the Mathematical Processing test to relate to surgical performance. A brief description of each test follows.

The Running Memory test assesses sustained attention, concentration, and working memory. A single digit appears briefly on the testing screen. The task requires the participant to indicate whether the current number matches the one previously displayed. The “SAME” button is to be pressed if the numbers match and the “DIFF” button is to be pressed if the numbers are different. The number matches on 50% of the trials, and a number is not presented in more than three consecutive trials. Numbers range from zero to nine, with each digit being equally likely.

The Matching to Sample test assesses spatial processing and visuospatial working memory. It also assesses attention, short-term memory, and spatial discrimination of a visual field. The participant is presented with a preliminary screen that shows an image with a prompt to memorize that exact image (see Appendix). Next, the participant is presented with two options side by side (see Appendix). The option that matches the image on the previous screen needs to be identified and tapped.

The Spatial Processing test assesses spatial processing and visuospatial working memory. It pairs histograms, each containing four bars, are presented simultaneously on the monitor. One histogram is rotated either 90° or 270° to the other (see Appendix). The participant presses designated buttons to indicate whether the histograms are the same or different, regardless of the orientation.

The Simple Reaction Time test assesses reaction time. The task requires the participant to tap on a stimulus as fast as possible. The stimulus is an asterisk (*) that appears on the screen twenty times. The participant is instructed to tap inside a dotted box each time the stimulus is presented. This test provides a measure of pure visual reaction time and a means to differentiate between motor response speed and actual cognitive processing time.

The Mathematical Processing test assesses basic computational skills, concentration, and working memory. The task was designed to assess working memory by requiring the participant to perform basic arithmetic operations. All problems require an addition and subtraction sequence in the form of “X + Y – Z” or “X – Y + Z.” The participant indicates whether the solution to the problem is greater than or less than five. The numbers are selected at random, only digits one through nine are used, and the correct answer may be any number from one through nine except five. Test problems with answers greater than five or less than five are equally probable. Working from left to right, the same digit cannot appear twice.

The Pursuit Tracking test is a tracking and fine motor abilities test that assesses visuomotor control. It is designed to detect impairments in fine motor abilities that are evident in neuromuscular disorders such as Parkinson disease and other similar conditions that include intentional tremor, familial tremor, and side effects of medication. It requires pressing the stylus on the screen and following a moving target or bull’s-eye across the screen. The target follows a sine wave.
For each test, the software calculated the completion time, the number of errors committed, and the throughput (the combination of accuracy and speed). The number of errors committed was dichotomized into those with no errors and those with one or more errors because of narrow distribution and substantial non-normality. Completion time, presence of errors, and throughput were examined as outcome variables. The presence of errors was used in the final analysis because we believe that the presence of errors is of the utmost clinical importance.

Of the six tests potentially available to be analyzed, two did not have the potential for error: the Simple Reaction Time and Pursuit Tracking tests. Analysis of sleep deprivation on the completion time for both tests showed no relationship to sleep as a continuous variable (p > 0.10), as described below. These two tests were eliminated from the results presented, because it was our belief that tests on which an error could be made were the most likely to be of clinical relevance.

Currently, no standardized definition of sleep deprivation based on an amount of sleep is available. Previous studies have used many different definitions for sleep deprivation. In our study, the amount of sleep obtained the previous night was examined as a continuous variable and as a dichotomous variable to indicate sleep deprivation. As a dichotomous variable, sleep deprivation was examined separately as two hours or less, three hours or less, and four hours or less of sleep. The effects were grossly similar at all three levels. However, visual inspection of scatterplots with a lowess curve (a local weighted scatterplot smoothing technique for plotting a smooth curve through data points by using regression modeling techniques) of the amount of sleep the previous night and the presence of errors on individual tests indicated that the relationship between sleep and test performance had a natural demarcation at four hours. We used four hours of sleep or less as our definition of sleep deprivation. An additional rationale against using a smaller value, such as two hours, for sleep deprivation is that this cutoff would have grouped participants with seemingly dissimilar amounts of sleep of three and seven hours into the same non-sleep-deprived group.

We therefore present our analysis that uses a cutoff of four hours of sleep or less. The sleep variable was collapsed into two categories for the final analysis: the sleep-deprived (participants getting four hours of sleep or less) and the non-sleep-deprived (those getting more than four hours of sleep).

**Statistical Methods**

The relationship between covariates and performance on cognitive and psychomotor tests was assessed with a generalized estimating equation regression with robust variance estimation to correct for correlations caused by multiple measurements within a surgeon. First, bivariate analysis was conducted to assess the association between individual covariates and performance on the tests. In instances in which a covariate had three or more categories, after the regression was performed, a Wald test was used to calculate the p value for the overall association between the categorical variable and test performance. Covariates associated with performance on the tests at the p < 0.10 level were included in the final model, except as noted. The relationship between demographic covariates that were associated with performance on the tests at the p < 0.10 level was assessed with use of the Kruskal-Wallis test on data collected at the first testing session. In the final multivariate model, with use of Wald tests, interactions were assessed between the amount of sleep the previous night and the type of test and between the amount of sleep the previous night and the type of provider. Interactions significant at the p < 0.05 level were included in the model. All statistical tests were two-sided. A value of p < 0.05 was considered to indicate significance, and a value of p < 0.10 was considered to indicate a trend. All statistical analyses were conducted with use of Stata 10 software (StataCorp LP, College Station, Texas).

**Source of Funding**

In support of this project, we received financial support in the form of educational research grants from AO North America and the Orthopaedic Research and Education Foundation. Funds were used to purchase Palm Operating System hardware, BrainCheckers software, and research coordinator support.

**TABLE I Sleep Deprivation as a Function of Job Type***

<table>
<thead>
<tr>
<th>Group</th>
<th>Amount of Sleep the Previous Night</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Four Hours or Less</td>
</tr>
<tr>
<td>Residents (n = 16)</td>
<td>61 (20)</td>
</tr>
<tr>
<td>Fellows (n = 10)</td>
<td>43 (25)</td>
</tr>
<tr>
<td>Attending surgeons (n = 6)</td>
<td>9 (11)</td>
</tr>
<tr>
<td>All (n = 32)</td>
<td>113 (20)</td>
</tr>
</tbody>
</table>

*The values are shown as the number of nights reported, with the percentage of nights by group in parentheses.

**Results**

During the testing periods, the fellows were the most likely group to be acutely sleep-deprived (25% of fellows compared with 11% of attending surgeons; p < 0.05) (Table I), followed closely by the residents (20%). Stratification of the data to periods of sleep between zero and four hours per night revealed that the surgeons had 113 such nights (20% of the data were collected under these circumstances) (Table I).

Residents obtained significantly less sleep when on call than did attending surgeons when on call (p < 0.001). The mean number of hours of sleep for all residents was 5.4 hours at baseline and 1.7 hours when on call. The mean number of hours of sleep for all fellows was 5.1 hours at baseline and 4.1 hours when on call. The mean number of hours of sleep for all attending surgeons was 6.0 hours at baseline and 4.5 hours when on call (Table II).

Bivariate analysis among all study participants indicated that the amount of sleep that the participant had obtained the previous night, the age of the participant, the type of provider, and the type of test were all associated with making an error on an individual test (p < 0.10) (Table III). Overall, orthopaedic surgeons who had slept four hours or less the previous night had 1.43 times the odds of making an error on an individual test than those who had slept more than four hours (p = 0.03). The odds of making an error on an individual test were 51% lower for attending surgeons than for residents (p = 0.06), but no significant difference was observed between fellows and residents (p = 0.66). The overall relationship between the type of provider and making an error on an individual test approached significance (p = 0.09). The odds of making an error on an individual test decreased 5% with each one-year increase in age of the participant (p = 0.02). However, age was not available for 16% of the study participants, and, considering that age is significantly associated with the type of provider, as determined by means of the Kruskal-Wallis test (p < 0.01), the type of provider was used in lieu of age in the final model. The odds of making an error on the Matching to Sample test were 39% lower than on the Mathematical Processing test (p = 0.05). The odds of making an error on the Running Memory test were 1.57 times less than for residents (p = 0.06).
the odds of making an error on Mathematical Processing test \((p = 0.04)\). No significant difference was shown in the odds of making an error on the Spatial Processing test compared with the odds of making an error on the Mathematical Processing test; the overall relationship between the type of test and making an error on an individual test was significant \((p < 0.01)\).

A model was created on the basis of the results of the bivariate analysis. As presented in Table III, those participants who had slept four hours or less the previous night had 1.43 times the odds of making an error on an individual test than those who had slept more than four hours \((p = 0.03)\). This association remained relatively unchanged after adjusting for the type of test \(\oddsratio, 1.42 \; p = 0.03\) and the type of test along with the type of provider \(\oddsratio, 1.34; \; p = 0.06\). The interaction between the type of test and sleeping four hours or less the previous night was significant \((p < 0.01)\), whereas the interaction between the type of provider and sleeping four hours or less the previous night approached significance \((p = 0.06)\). Considering the relatively small sample size, only the interaction between the type of test and sleeping four hours or less the previous night was used in the final model because the \(p\) value of that interaction was smaller.

The final model included the amount of sleep the previous night, the type of provider, the type of test, and the interaction between the type of test and the amount of sleep the previous night. The association between the amount of sleep the previous night and making an error on a specific test is presented in Table IV. Orthopaedic surgeons who had slept four hours or less the previous night had 1.83 times the odds of making one or more errors on the Running Memory test compared with those who had slept more than four hours the previous night \((p = 0.02)\). This association was present after adjusting for the type of provider \(\oddsratio, 1.72; \; p = 0.04\). Additionally, although not significant, those who had slept four hours or less the previous night were more likely to make one or more errors on the Running Memory test compared with those who had slept more than four hours the previous night \((p = 0.02)\). This association was present after adjusting for the type of provider \(\oddsratio, 1.72; \; p = 0.04\). Additionally, although not significant, those who had slept four hours or less the previous night were more likely to make one or more errors on the Running Memory test compared with those who had slept more than four hours the previous night. The direction and magnitude of these estimates remained relatively unchanged on the Matching to Sample test \(\oddsratio, 1.74; \; p = 0.10\) and the Spatial Processing test \(\oddsratio, 1.44; \; p = 0.20\) compared with those who had slept more than four hours the previous night.

The final model included the amount of sleep the previous night, the type of provider, the type of test, and the interaction between the type of test and the amount of sleep the previous night. The association between the amount of sleep the previous night and making an error on a specific test is presented in Table IV. Orthopaedic surgeons who had slept four hours or less the previous night had 1.83 times the odds of making one or more errors on the Running Memory test compared with those who had slept more than four hours the previous night \((p = 0.02)\). This association was present after adjusting for the type of provider \(\oddsratio, 1.72; \; p = 0.04\). Additionally, although not significant, those who had slept four hours or less the previous night were more likely to make one or more errors on the Running Memory test compared with those who had slept more than four hours the previous night. The direction and magnitude of these estimates remained relatively unchanged on the Matching to Sample test \(\oddsratio, 1.74; \; p = 0.10\) and the Spatial Processing test \(\oddsratio, 1.44; \; p = 0.20\) compared with those who had slept more than four hours the previous night. This relationship was not markedly changed after adjusting for the type of provider \(\oddsratio, 0.75; \; p = 0.35\).

### Table II Amount of Sleep When On Call Compared with That When Not On Call

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Amount of Sleep*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>On Call</td>
</tr>
<tr>
<td>Residents</td>
<td>1.7 (1.2 to 2.1)</td>
</tr>
<tr>
<td>Fellows</td>
<td>4.1 (3.6 to 4.6)</td>
</tr>
<tr>
<td>Attending surgeons</td>
<td>4.5 (3.6 to 5.4)</td>
</tr>
<tr>
<td>All groups combined</td>
<td>3.0 (2.6 to 3.4)</td>
</tr>
</tbody>
</table>

*The values are given as the average amount of sleep in hours, with the 95% confidence interval in parentheses.

### Table III Assessing Associations Between Covariates and Performance on Tests

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Odds Ratio*</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>0.97 (0.31 to 2.99)</td>
<td>0.96†</td>
</tr>
<tr>
<td>Age (yr)</td>
<td>0.95 (0.91 to 0.99)</td>
<td>0.02†</td>
</tr>
<tr>
<td>Post-call</td>
<td>1.23 (0.95 to 1.60)</td>
<td>0.11†</td>
</tr>
<tr>
<td>Slept in hospital the  previous night</td>
<td>1.33 (0.94 to 1.87)</td>
<td>0.11†</td>
</tr>
<tr>
<td>Slept four hours or less the previous night</td>
<td>1.43 (1.04 to 1.95)</td>
<td>0.03†</td>
</tr>
<tr>
<td>Type of provider</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resident</td>
<td>1.0</td>
<td>—†</td>
</tr>
<tr>
<td>Fellow</td>
<td>1.14 (0.63 to 2.06)</td>
<td>0.66†</td>
</tr>
<tr>
<td>Attending surgeon</td>
<td>0.49 (0.23 to 1.03)</td>
<td>0.06†</td>
</tr>
<tr>
<td>Type of test</td>
<td>&lt;0.01†</td>
<td></td>
</tr>
<tr>
<td>Mathematical Processing</td>
<td>1.0</td>
<td>—†</td>
</tr>
<tr>
<td>Matching to Sample</td>
<td>0.61 (0.38 to 1.00)</td>
<td>0.05†</td>
</tr>
<tr>
<td>Spatial Processing</td>
<td>1.26 (0.92 to 1.74)</td>
<td>0.15†</td>
</tr>
<tr>
<td>Running Memory</td>
<td>1.57 (1.02 to 2.41)</td>
<td>0.04†</td>
</tr>
</tbody>
</table>

*The values are shown as the odds of a participant making at least one error on an individual test, with the 95% confidence interval in parentheses. All odds ratios (except age, type of provider, and type of test) are the comparison of the variable name with its opposite (e.g., the odds ratio for post-call is post-call compared with not post-call). †The \(p\) value of the Wald test comparing groups.
Discussion

In this study of cognitive and psychomotor performance of orthopaedic surgeons with acute sleep deprivation, sleeping four hours or less was associated with a 43% increase in the odds of making at least one error on an individual test (p < 0.05). When examining the test-specific results, decline in performance was noted on the Matching to Sample, Spatial Processing, and Running Memory tests when participants were functioning in a sleep-deprived state. However, only the Running Memory test demonstrated a significant decline in performance. The odds of making at least one error on the Running Memory test increased by 83% when the respondents were sleep-deprived (p < 0.05). This effect persisted in a model that controlled for the type of provider (p < 0.05).

Of note, although our participants infrequently met our criteria for acute sleep deprivation (four hours of sleep or less), they may have been functioning in a state of chronic sleep deprivation. According to the National Institutes of Health, most adults need between seven and eight hours of sleep each night to be well rested and to function in a healthy state. Less sleep than that can impair concentration, reaction time, and mood. All participants in our study had less than the recommended average of seven to eight hours of sleep per night. Overall, the residents, fellows, and attending surgeons obtained a mean amount of 5.4, 5.1, and 6.0 hours of sleep each night, respectively. Previous work on chronic sleep deprivation has suggested that sleeping only five to six hours a night can lead to impairment.

If a single night of reduced sleep can impair physicians’ cognitive function, repeated nights of being on call may be even more detrimental, placing doctors in a state of chronic sleep deprivation. According to the National Institutes of Health, most adults need between seven and eight hours of sleep each night to be well rested and to function in a healthy state. Less sleep than that can impair concentration, reaction time, and mood. All participants in our study had less than the recommended average of seven to eight hours of sleep per night. Overall, the residents, fellows, and attending surgeons obtained a mean amount of 5.4, 5.1, and 6.0 hours of sleep each night, respectively. Previous work on chronic sleep deprivation has suggested that sleeping only five to six hours a night can lead to impairment.

Although no previous work on sleep deprivation in physicians has examined orthopaedic surgeons or has used our validated neuropsychological testing, multiple studies conducted during the last two decades have shown that physicians who were not surgeons demonstrated clinical performance that was adversely affected by sleep deprivation. Interns working in a medical intensive care unit had 50% more attention failures and committed 22% more serious errors while working during shifts after reduced sleep. Internal medicine house officers provided more incorrect answers on questionnaires when post-call, and sleep deprivation was associated with worsening performance, lower confidence, less energy, and mood disorders. Anesthesiology residents showed significant prolongation in response speed and reduction in accuracy after a single sleepless night. Post-call pediatric residents reported a higher incidence of vehicular accidents while driving home after a night spent at the hospital. Post-call performance has been compared with performance when intoxicated. Arnedt et al. discovered that post-call performance impairment during a heavy call rotation is comparable with impairment associated with a 0.04% to 0.05% blood alcohol content.

Few studies have evaluated surgeons’ performance when sleep-deprived, and those that have offer conflicting results. Taffinder et al. found that surgeons who had remained awake all night made 20% more errors and took 14% longer to complete tasks on a laparoscopic simulator than those who had obtained a full night’s sleep. Uchal et al. found no significant difference in outcome measures on a laparoscopic simulator when thirty-two post-call surgeons were compared with their well-rested counterparts. These studies were conducted with simulators and may not have accurately reflected actual real-time surgical procedures.

Ellman et al. were the first to evaluate clinical outcomes achieved by surgeons in a sleep-deprived state. The authors separately evaluated attending thoracic surgeons and thoracic surgical residents and found no significant differences in mortality or complication rates when cardiac operations were performed at night. They concluded that sleep deprivation does not affect morbidly or mortality associated with cardiac surgical operations and that the data did not support a need for work-hour restrictions for surgeons. However, the studies were not without limitations. Both were underpowered, retrospective reviews of surgical databases, and neither listed the amount of sleep obtained.

One of the limitations of our study was the unknown relationship between function as tested by our software and surgical performance. Tasks that assess reaction time, memory, and concentration may measure cognitive function but may not accurately mimic surgical performance. Concentrating for

<table>
<thead>
<tr>
<th>TABLE IV Association Between the Amount of Sleep the Previous Night and Making an Error on a Specific Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test</td>
</tr>
<tr>
<td>Mathematical Processing</td>
</tr>
<tr>
<td>Matching to Sample</td>
</tr>
<tr>
<td>Spatial Processing</td>
</tr>
<tr>
<td>Running Memory</td>
</tr>
</tbody>
</table>

*Odds of making at least one error on an individual test, comparing participants who had slept four hours or less the previous night with those who had slept more than four hours. †Adjusted for the type of provider.
It is interesting to note that many professions to which the public entrusts its safety have strict regulations with regard to obtaining adequate sleep because of concern for deterioration of performance with sleep deprivation. Airline pilots, bus drivers, and highway state troopers have regulations governing the amount of sleep that must be obtained before commencing a work shift. The medical field does not, although this issue is at the heart of recent resident work-hour restrictions. Airline passengers expect that pilots will not be sleep-deprived. Should patients not expect the same of surgeons?

The results of this preliminary study must be interpreted with caution, in particular because we have no data regarding the relationship between the performance on these tests and surgical performance. More data on surgeon performance in a sleep-deprived state are necessary to fully understand the relationship between these findings and clinical performance.

Only after we have obtained a better understanding of the influence of sleep deprivation on surgeon performance can we begin to make inferences with regard to the timing and scheduling of surgical procedures.

Appendix

Figures showing photographs of handheld computers demonstrating the Matching to Sample and Spatial Processing tests and tables showing the demographic characteristics of the study participants and the tests administered are available with the online version of this article as a data supplement at jbjs.org.

Note: The authors thank Drs. Aragon, Bauer, Bedford, Belin, Carter, Chomas, Chen, Cuneva, Doro, Drake, Duggan, Forward, Golden, Gordon, Guillers, Gumbo, Harvey, Hawks, Laine, Mansen, Ophlant, Osgood, Park, Quinnan, Stall, Strauss, and Wierks for participation in this study. The authors also thank senior editor and writer Dori Kelly, MA, for her expertise in the preparation of the manuscript and Jennifer Higgins, MS, for her assistance with the figures.

References


