Differences in Day and Night Shift Clinical Performance in Anesthesiology

Caroline G. L. Cao, Tufts University, Medford, Massachusetts, **Matthew B. Weinger, Jason Slagle,** and **Chuan Zhou,** Vanderbilt University Medical Center, Nashville, Tennessee, and **Jennie Ou, Shakha Gillin, Bryant Sheh,** and **William Mazzei,** University of California, San Diego, La Jolla, California

Objective: This study examined whether anesthesia residents (physicians in training) performed clinical duties in the operating room differently during the day versus at night. **Background:** Fatigue from sleep deprivation and working through the night is common for physicians, particularly during residency training. **Methods:** Using a repeated-measures design, we studied 13 pairs of day-night matched anesthesia cases. Dependent measures included task times, workload ratings, response to an alarm light latency task, and mood. **Results:** Residents spent significantly less time on manual tasks and more time on monitoring tasks during the maintenance phase at night than during the day. Residents reported more negative mood at night than during the day, both pre- and postoperation. However, time of day had no effect on the mood change between pre- and postoperation. Workload ratings and the response time to an alarm light latency task were not significantly different between night and day cases. **Conclusions:** Because night shift residents had been awake and working for more than 16 hr, the observed differences in task performance and mood may be attributed to fatigue. The changes in task distribution during night shift work may represent compensatory strategies to maintain patient care quality while keeping perceived workload at a manageable level. **Applications:** Fatigue effects during night shifts should be considered when designing work-rest schedules for clinicians. This matched-case control scheme can also be applied to study other phenomena associated with patient safety in the actual clinical environment.

INTRODUCTION

The ongoing debate regarding work hour restrictions for resident physicians reflects a growing concern about the effect of sleep deprivation and fatigue on physicians' ability to provide quality care (Lamberg, 2002; Weinger & Ancoli-Israel, 2002). Before the policy change instituted by the Accreditation Council on Graduate Medical Education to restrict resident work hours to 80 hr per week, residents routinely worked 100 hr or more, with shifts of more than 30 consecutive hr every 3rd or 4th day (Lamberg, 2002).

Residents working such schedules experience acute and chronic sleep deprivation and circadian rhythm disturbances (Gaba & Howard, 2002; Howard, Gaba, Rosekind, & Zarcone, 2002; Howard, Rosekind, Katz, & Berry, 2002; Lamberg, 2002; Veasey, Rosen, Baransky, Rosen, & Owens, 2002; Weinger & Ancoli-Israel, 2002). Fatigue (i.e., the cognitive, behavioral, and physiologic outcomes of sleep loss and circadian disruption) ranks among the first complaints of medical staff (Daugherty, Baldwin, & Rowley, 1998) and may be a major contributor to medical errors (Gaba & Howard, 2002; Gaba, Howard, & Jump, 1994).

A great deal of research has demonstrated that fatigue related to sleep loss may cause a variety of deficits, including lack of innovation and creativity, increased distractibility, an inability to deal with unexpected events or to deviate from previous problem-solving strategies, unreliable temporal

Address correspondence to Matthew B. Weinger, Center for Perioperative Research in Quality, Vanderbilt University Medical Center, 1121 21st Ave. S., 732 Medical Arts Building, Nashville, TN 37212; matt.weinger@vanderbilt.edu. ***HUMAN FACTORS,* Vol. 50, No. 2, April 2008, pp. 276–290. DOI 10.1518/001872008X288303.**

memory, impaired language skills, and negative mood (Bonnet, 2000; Harrison & Horne, 2000). Furthermore, sleep deprivation impairs cognitive performance and hand-eye coordination (Dawson & Reid, 1997), altering at the same time the behavioral status of the individual toward depression, aggressiveness, anxiety, anger, and decreased vigilance (Ford & Wantz, 1984). Studies have also shown that 24 hr without sleep is equivalent to a blood alcohol level of 0.10% on a hand-eye coordination task (Dawson & Reid, 1997).

Although early studies suggested that because of high motivation to avoid critical errors, fatigued residents did not suffer a decrement in cognitive task performance (Spurgeon & Harrington, 1989), more recent work has found that motor skill performance can be impaired (Eastridge et al., 2003; Grantcharov, Bardram, Funch-Jensen, & Rosenberg, 2001; Taffinder, McManus, Gul, Russell, & Darzi, 1998; Wesnes et al., 1997). Using a laparoscopic surgery simulator to measure motor dexterity, Taffinder et al. (1998) found that skill errors and task completion times increased after a sleepless overnight duty shift. Therefore, for residents who are typically less experienced in responding to unusual situations, the probability of an adverse outcome may increase when they are asked to perform while fatigued in high-risk, complex, and dynamic task environments, such as the operating room (OR; Howard, Rosekind, et al., 2002; Howard et al., 2003; Weinger & Ancoli-Israel, 2002).

A recent randomized, controlled trial of interns working shifts of 24 continuous hr or longer documented more serious medical errors and more attentional failures (measured by polysomnography) when compared with those working shifts of 16 hr or less (Landrigan et al., 2004; Lockley et al., 2004). Subsequently, the same team conducted a national longitudinal Web-based survey of almost 3,000 interns, who reported an approximately sevenfold higher incidence of fatigue-related medical errors or preventable adverse events (and significantly more attentional failures) during months when they worked more than five extended duty shifts than during months with no extended duty shifts (Barger et al., 2006).

However, other studies have questioned whether the mandated resident work hour restrictions have had the predicted beneficial effects on the quality of patient care (Poulose et al., 2005) or the rate of physician burnout (Gelfand et al., 2004). Poulose et al. (2005) found that 3 years after the implementation of resident work hour restrictions, the incidence of surgical adverse events had not improved in New York's teaching hospitals. In fact, the rates of accidental puncture/laceration and of postoperative thrombosis events were actually higher than before the policy change. The lack of evidence to support the benefits of work hour limitations may be attributable to the short period since implementation, the methodology used to assess the effects of fatigue (Fletcher et al., 2005), or inadequate processes to ensure continuity of care as the resulting shorter or less frequent resident duty shifts increase the number of care transitions (i.e., handoffs) between providers.

Veasey et al. (2002) reviewed 10 studies addressing the effects of sleep loss on cognition and task performance in health care providers, and they indicated that few investigators have successfully resolved the challenges inherent in studying how fatigue affects physicians' actual performance and decision making during patient care–related tasks. The majority of studies examining the effects of sleep deprivation on physicians' functioning have been conducted in laboratory and nonmedical settings, rather than during actual patient care, limiting their ability to generalize results to the clinical setting. Further, many of the earlier studies (Leung & Becker, 1992; Samkoff & Jacques, 1991) had serious methodological flaws, such as inadequate controls, lack of randomization, insensitive outcome measures, and inadequate consideration of effects over time (Weinger & Ancoli-Israel, 2002; Weinger & Englund, 1990).

In the past decade, more studies have employed rigorous methodologies and realistic outcomes measures (e.g., performance on simulated cases; Grantcharov et al., 2001; Howard et al., 2003; Leonard, Fanning, Attwood, & Buckley, 1998; Nelson, Dell'Angela, Jellish, Brown, & Skaredoff, 1995; Smith-Coggins, Rosekind, Buccino, Dinges, & Moser, 1997; Taffinder et al., 1998; Wesnes et al., 1997). Recent studies by Landrigan et al. (2004) and Lockley et al. (2004) of medical interns are notable, but such complex randomized, controlled trials are very resource intensive and may not generalize to more experienced clinicians or to other clinical disciplines.

The goal of the present study was to examine whether anesthesia residents perform their clinical duties in the OR differently during regular day shifts and extended night shifts. According to the literature, there are two major sources of nighttime sleepiness or sense of fatigue: circadian rhythm disruption and sleep loss (Akerstedt, 1991; Lavie, 1991). Subjective alertness and performance efficiency are determined by the homeostatic process (i.e., how long the participant has been awake) and circadian rhythms (e.g., regular fluctuation of arousal level with time of day; Carrier & Monk, 1999). Furthermore, performance efficiency depends on the task demands and strategies adopted by the participant, which in turn may change because of fatigue.

In this study, we postulated that residents would be more fatigued during the night shifts that occur as part of their extended on-call duties. Regularly scheduled on-call duty begins in the morning and typically lasts 24 hr. While residents are on call, sleep is rare and, when possible, usually fragmented. Therefore, in the midst of a night shift, anesthesia residents have generally been awake and working for 16 hr or more. Furthermore, given the nature of anesthesia care, which primarily involves patient monitoring and vigilance of sometimes subtle changes in patient status, anesthesiologists may be more susceptible to the effects of fatigue than other perioperative clinicians such as surgeons and nurses, whose primary clinical duties involve psychomotor tasks.

This study was also designed to address some of the intrinsic limitations of real-world clinical performance studies by including case matching in daytime and nighttime work shifts for the same participant. This study also employed an innovative use of multiple techniques to elicit converging evidence on physician task distribution and workload. Clinicians' workload was obtained using both self-report and externally observed ratings. The reliability of observers' task analysis and workload ratings was assessed through video review by a third trained and well-rested observer, who was blinded to the experimental conditions.

It was hypothesized that fatigue would impair the perceptual and cognitive abilities of anesthesia residents as evidenced by altered task patterns, increased workload, and decreased vigilance (as measured by increased response time to an alarm latency task) during night-shift work, compared with their daytime performance after a night of normal sleep. To test the hypothesis, a prospective study was designed to observe anesthesia residents working on matched night and day shift cases.

METHODS

Participants

Initially, 14 anesthesia residents (11 men and 3 women ranging in age from 27 to 31 years) at the University of California, San Diego, participated in the study. There were 9 first-year, 1 second-year, and 4 third-year anesthesia residents. One of the residents took part in two different trials, and another resident slept before his night case. To ensure maximal data quality, data from the rested resident and from one of the two trials from the single resident were excluded from analyses (the less well matched trial of the two was dropped), resulting in a total of 13 trials from 13 different residents. The average length of anesthesia training for the residents in the night trials (always studied first) was 10.8 months (range 1.5–30 months), whereas in the day trials it was 12.7 months (range 2.0–31 months; see Table 1). Before participating in the study, residents were thoroughly familiarized with the procedures and the data to be collected, and provided written informed consent according to a research protocol approved by the Institutional Review Board.

Experimental Conditions

Each resident took part in one trial consisting of two paired cases: one at night during their on-call (night) shift and one during their regular (noncall) day shift. Residents' regular day shift began at 06:30 and generally ended before 18:00. Residents were usually on call every fourth or fifth night. The on-call shifts typically began at 10:00 (but sometimes at 06:30, depending on work demands) and were completed at 07:00 the following morning. The first case observation in the pair occurred during the night (between 22:00 and 06:00), and the second occurred during the day (between 09:00 and 17:00).

Because nighttime cases were much less frequent than daytime cases, and the likelihood of unusual or emergency cases was greater at night, night cases were selected first. The resident was then assigned to a comparable daytime case within a 1-month period. This arrangement was necessary to accommodate both the residents' and ORs' scheduling constraints. Therefore, the second case in the pair was prospectively matched, as closely as possible, for type, length, and difficulty of surgery; expected type of anesthesia to be provided; and the patients' age and health status as rated

TABLE 1: Case and Participant Matching Data

Case No.	Months of Training	Day/ Night	Anesthesia Start Time	Patient ASA Status	Case Duration (min)	Surgical Procedure
			Observer 1			
1[a]	2	Night	05:00	1	100	Knee arthroscopy
	2	Day	11:40	2	65	Cystoscopy
2	25	Night	22:45	2	285	ORIF right ankle
	26	Day	12:30	2	195	ORIF left ankle
3	25	Night	00:55	1E[c]	110	Laparoscopy/salpingectomy
	29	Day	08:45	2	65	Hysteroscopy/dilation and curettage
4	2	Night	03:00	2E[c]	125	Appendectomy
	2	Day	07:30	2	115	Removal nasal mass
5	26	Night	22:00	2E[c]	85	Repair elbow fracture
	26	Day	12:05	2	330	Repair ankle fracture
6	30	Night	23:20	4E[c]	340	Hemicolectomy
	31	Day	11:35	3	160	Exploratory laparotomy
7	3	Night	02:15	2	180	Debridement of elbow
	5	Day	07:20	1	100	Knee arthroscopy
8	19	Night	22:40	1	170	Repair hip fracture
	20	Day	12:55	2	215	Repair hip fracture
			Observer 2 and Videotaped Cases			
9[b]	2	Night	00:30	2E[c]	145	Debridement hand and knee
	2	Day	12:15	1	225	Shoulder arthroscopy
10[b]	2	Night	01:35	2	115	Axillary-femoral bypass graft
	2	Day	07:20	2	85	Laparoscopic cholecystectomy
11	2	Night	02:15	1	68	Dilation and curettage
	2	Day	07:15	3	230	Carotid endarterectomy
12	2	Night	00:00	1E[c]	195	Repair distal femur fracture
	2	Day	16:40	2	245	Repair tibial plateau fracture
13	5	Night	23:45	2	180	Repair hip fracture
	6	Day	07:15	2	175	Total hip arthroplasty
14	5	Night	03:50	1E[c]	150	Laparoscopic appendectomy
	6	Day	12:18	2	132	Laparoscopic cholecystectomy
15	5	Night	22:00	2	420	Spinal fusion
	5	Day	07:25	3	310	Spinal fusion

Note. ASA = American Society of Anesthesiologists Physical Status Scale. ORIF = open reduction, internal fixation.
[a]Resident rested before the start of the night case; Case 1 was excluded from analysis. [b]Resident was observed for two case pairs; Case 9 was excluded from analysis. [c]E indicates cases designated as true emergencies by the involved anesthesia providers. Other nighttime cases were considered "urgent" (i.e., couldn't wait until the next day following the scheduled list of cases to start).

according to the American Society of Anesthesiologists (ASA) Physical Status Scale (Owens, Veasey, & Rosen, 2001), but not for the number of hours into the shift the case was scheduled to start. This design approached the best-controlled conditions possible in this real-world environment.

Procedures and Dependent Measures

Demographics. Before each case, residents completed a questionnaire providing demographic information (age, gender, etc.), sleep and activity status, and drug and caffeine consumption over the previous 24 hr. The sleep and activity data collected included number of hours of sleep participants received over the previous night and number of hours they were awake prior to the start of each study case.

Behavioral task analysis. During each case, a trained observer sat in the OR with a laptop computer and performed an intraoperative behavioral task analysis, beginning when the patient entered the OR and ending when the patient left the room. Data collection was temporarily suspended if the participant left the OR for a break during the case. The behavioral task analysis utilized 37 task categories that had been previously defined and validated (Slagle, Weinger, Dinh, Wertheim, & Williams, 2002; Weinger, Herndon, & Gaba, 1997;

Weinger et al., 1994). These task categories were displayed to the observer on a single screen by custom software (see Figure 1). The start time of each observed task was recorded with the click of a radio button. Simultaneous tasks were recorded by toggling between task categories based on the proportion of time spent on each task.

All data were automatically organized and tabulated. For each task observed, the following measures were recorded: total time spent on task, percentage of case time spent on task, the number of times a task occurred, and the average task duration or dwell time (i.e., the time a task was performed before a different task was initiated). All measures were calculated separately for each phase (induction, maintenance, and emergence) of the anesthetic case. "End of induction" was defined as the time when the patient's trachea had been intubated and the endotracheal tube had been secured or when the anesthesia resident told the surgeons that they could begin operating, whichever occurred first. "Beginning of emergence" was defined as occurring when the anesthesia provider shut off all anesthetic agents and began delivering 100% oxygen. The maintenance phase of anesthesia care occurred between the end of induction and the beginning of emergence (Weinger et al., 1994, 1997).

To establish the validity of the intraoperative observers' task analysis data, during 14 of the cases (7 case pairs) the residents' clinical activities were also videotaped for off-line analysis by a second trained observer, who was blinded to the night versus daytime case status.

Vigilance and workload. To measure response time to an alarm latency task, the observer was also prompted by the computer, at random 10- to 15-min intervals throughout the case, to illuminate a small, bright red "alarm" light, which was positioned adjacent to the physiological monitors at the center of the anesthesia workspace (Weinger et al., 1994, 1997). Residents were instructed to indicate to the observer, either verbally or by hand signal, their detection of the illuminated light. The response time (RT) between light illumination and its detection by the participant was recorded to provide a measure of the response latency to a new (secondary) task demand. This type of probe detection task is well supported in the literature (Parasuraman, 1985) to be an indirect measure of perceptual or cognitive workload (Eysenck & Eysenck, 1979).

Independently, the computer prompted the observer, at random 10- to 15-min intervals, to rate and record the resident's clinical workload using a previously validated 15-point scale (Borg, 1977, 1998; Weinger et al., 1994, 1997) that ranges from 6 (e.g., completely sedentary participant) to 20 (e.g., during a full-blown OR resuscitation) and, only then, to obtain the resident's self-rating to

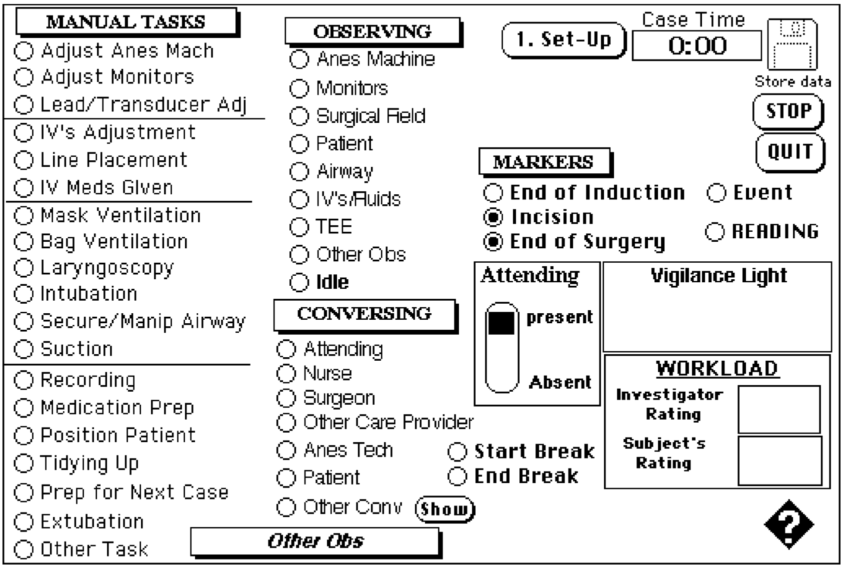

Figure 1. Screen snapshot of custom data collection software.

avoid bias. This univariate rating scale integrates multiple workload constructs, including physical effort, mental effort, and psychological stress. Participants were explicitly instructed during the prestudy briefing that the typical workload rating at the time of a routine oral intubation was about 12. A copy of the scale with anchor values was observable by the participant during each study.

Workload density was also determined using a technique that permits real-time measurement of workload by weighing the contribution of each task performed with a workload factor score (Vredenburgh, Weinger, Williams, Kalsher, & Macario, 2000). The workload factor score is a statistically determined numerical value, derived from the prior ratings of anesthesia providers, of the physical effort, mental effort, and psychological stress associated with each particular task relative to all other anesthesia tasks (Vredenburgh et al., 2000). Workload and alarm latency measures were calculated for each phase of the anesthetic and for the entire case.

Mood. Before and after each case, residents completed a mood survey instrument (modified Profile of Mood States, or POMS-Brief; McNair, Lorr, & Droppleman, 1992). The mood questionnaire asked residents to rate 13 "feeling" words on a Likert scale from 1 (*not at all*) to 10 (*very much*). Two of these words were positive mood descriptors (i.e., "alert," "relaxed"), and the remaining 11 words were negative mood descriptors (e.g., "tense," "tired," "nervous"). A negative mood score was calculated by subtracting the sum of the 2 positive mood items from the sum of the 11 negative mood items. Therefore, the upper bound (most negative mood) was a score of 108 and the lower bound was –9 (most positive mood).

Data Analysis

Data were analyzed using the R statistical package (R Development Core Team, 2006). Summary statistics such as means, medians, and standard deviations were calculated for each variable. Univariate analyses of association between the dependent variables (task allocations, mood scores, alarm latency, and workload, which included self-reported workload, observer-scored workload, and calculated workload density) and the primary independent variable (day/night) were performed with paired Wilcoxon's signed-rank sum test and Fisher's exact test for continuous and categorical variables, respectively.

Given the small number of participants and the scheduling constraints, we could not have randomized the residents to different cases. Even though the pair of cases performed by each resident was carefully matched, several important covariates were unbalanced between the day and night cases. For example, the amount of anesthesia training for each resident in the night and day cases was unbalanced because night cases were always observed first. These potential confounders were addressed, to the extent possible, within linear mixed effect regression models. However, because of the relatively small sample size, only the most important potential confounders could be included in the models while retaining statistical validity.

On the basis of scientific evidence, clinical knowledge, and statistical considerations, the variables that were included in the multivariable models were time of day of the case (i.e., day or night), case duration, months of residency training, and number of hours the participant had slept the previous night. Hours awake before the case was highly correlated with time of day. Because both are surrogate measures of sleeplessness, including them both in the model would introduce collinearity and reduce the power of corresponding tests. We opted to use time of day because it addressed our primary research question, is statistically simple to interpret, and would make it easier to design mitigation strategies if it proved to be a significant factor.

Patient's ASA status was highly correlated with case duration. Because the focus of the study was on residents' clinical behavior, we deemed the case characteristics as more relevant than patient characteristics. Hours slept the previous night (before the start of the shift) was included in the models to allow us to estimate the magnitude of sleeplessness and fatigue.

Linear mixed effect regression models were used to evaluate the relationship between residents' task performance and time of day (i.e., day vs. night shift) while adjusting for months of resident training, case duration, and the number of hours participants had slept the previous night. For the behavioral task analysis, we focused on the percentage of time spent on specific task categories because they were standardized, regardless of the duration of each phase or of the entire case.

Because of the paired design of the study, observations within the same resident tend to be highly correlated. To account for such correlation, a random intercept was introduced in the linear model, whereas the five predictors were treated as fixed effects. Thus, the total variability was decomposed into within-pair variability and between-pair variability, allowing for better standard error estimates of coefficients and increased power of corresponding tests. The estimation of coefficients was based on maximum likelihood, and the effect that time of day had on resident task performance was evaluated using *t* tests (Pinheiro & Bates, 2000). Similarly, we applied the linear mixed effect regression models to compare alarm light response time, mood changes, and workload ratings between the day and night shift cases. A *p* value less than .05 was considered statistically significant.

Interrater reliability assessment. One trained observer collected data on the first eight pairs of cases; a second observer, similarly trained and experienced, collected data on the subsequent seven pairs of cases (all of which were videotaped and reviewed by a third trained observer). To allow us to test interrater reliability of the task analysis and workload assessment measures, the third observer studied six of the seven cases from videotape. One pair of cases (Case Pair 12; see Table 1) could not be analyzed off line because one of its videotapes was unusable. Spearman's rho rank correlation coefficients (Myles et al., 1999; Myles, Weitkamp, Jones, Melick, & Hensen, 2000; Turner-Stokes et al., 1998) were calculated across task categories for each phase (i.e., induction, maintenance, and emergence) and over the entire case for each rater. The reliability of the observers' workload ratings was similarly assessed.

RESULTS

Table 1 provides a summary of the case and participant matching data, including the experience of the residents, case time and duration, patient ASA status, and type of surgical procedure performed. The night prior to daytime cases, residents averaged 6.6 ± 1.1 (range 4.5–8.5) hr of sleep and had been awake an average of 4.6 ± 2.7 (range 1.8–9.7) hr prior to case start. In contrast, prior to night cases, they had been awake for a significantly longer period (an average of 18.2 ± 2.4 hr, range 0.2–22.8; $p < .001$ compared with day cases). The night prior to being on call (i.e., the night before the night cases), residents averaged 7.1 ± 1.7 (range 3–9) hr of sleep. Seven of the 13 cases were emergency cases, all occurring during the night trials. Because the emergency designation was entirely confounded by night, these data provided no information on any emergency versus time of day interaction.

The average start time of the day cases was midmorning (10:19 a.m. ± 10 min) and of the night cases was after midnight (00:48 a.m. ± 19 min). The day and night cases were well matched. There was no statistically significant difference between the day (187 ± 93 min) and night (187 ± 104 min) case durations. No significant critical events occurred during any of the study cases.

An attending anesthesiologist was present in the OR for at least some period during all daytime case but was absent in 5 of the 15 night cases ($\chi^2 = 6.00$; $p = .01$). However, the attending physicians spent a similar amount of time overall with residents in the OR (33 ± 8 min in day and 26 ± 8 min in night cases), and one was always immediately available should the resident require assistance. The residents took significantly more breaks during day (1.3 ± 1.1) than night cases (0.5 ± 0.6; $p < .05$). In addition, total break time per case was significantly longer during daytime (27 ± 25 min) compared with nighttime (9 ± 12 min; $p < .05$) cases. Whenever a resident left the OR on a break, another anesthesia provider substituted, but no data were collected during these breaks.

Table 2 summarizes the results of the multivariate regression analyses.

Behavioral task analysis. On average, surgical cases were 144 ± 74 min in duration, with the shortest case lasting 44 min and the longest 318 min. Each case was divided into the three phases of anesthesia care: induction, maintenance, and emergence. The average durations of the three phases were 20 ± 9 min, 109 ± 64 min, and 21 ± 14 min, respectively. There were no significant differences in the duration of any phase of care between day and night cases.

Overall, the amount of time spent on many clinical tasks was not significantly different between the day and night conditions, except for two key task categories: manual tasks and observation tasks (see Tables 2 and 3). During nighttime cases, residents spent significantly less time, on a total case percentage basis, on manual tasks ($p = .01$) but significantly more time on observing tasks ($p = .02$), when adjusted for similar resident experience, case duration, and hours slept the previous

TABLE 2: Abbreviated Summary of Results of Regression Analyses for Differences Between Night and Day Cases

Dependent Measure	Phase or Task	Significant Difference[a]	*ns*[a]
Total task time			
All phases:	Observing	Night > day	
Induction			*ns*
Maintenance:	Observing	Night > day	
Emergence			*ns*
% Task time			
All phases:	Manual	Night < day	
	Observing	Night > day	
Induction			*ns*
Maintenance:	Manual	Night < day	
	Observing	Night > day	
Emergence			*ns*
Negative mood scale			
Precase score		Night > day	
Postcase score		Night > day	
Precase–postcase score			*ns*
Workload measures			
Self-reported workload			
All phases			*ns*
Induction			*ns*
Maintenance			*ns*
Emergence			*ns*
Observer-assessed workload			*ns*
All phases			*ns*
Induction			*ns*
Maintenance			*ns*
Emergence			*ns*
Workload density			*ns*
Vigilance latency (alarm light response time)			
All phases			*ns*
Induction			*ns*
Maintenance			*ns*
Emergence			*ns*

Note. All models adjusted for participant months of experience, number of hours participant slept during the previous night, and case duration.

[a]$p < .05$.

night. These differences were most notable during the maintenance phase (the longest of the three phases), when the residents spent more time observing their physiological monitors, the patient's airway, and intravenous fluids.

During the entire case, observing tasks occurred more often and lasted longer for the residents at night, even after controlling for the case duration. There were more observation tasks (i.e., individual occurrences of observation) at night during emergence ($p < .001$); also during that period, these observation tasks lasted an average of 11 s longer than they did during the daytime. The duration and frequency of other task categories did not differ significantly between night and daytime.

As would be expected, residents spent more time on manual tasks relative to other tasks during the induction phase. In the maintenance phase, observing tasks consumed the largest proportion of the residents' time. Nevertheless, participants spent more time on observing tasks and less time on manual tasks at night than during daytime during maintenance. In the emergence phase, residents again spent more time on manual tasks, with no differences in task distribution between day and night.

TABLE 3: Summary of Task Performance Between Day and Night Cases (% Time)

	Total Case		Induction		Maintenance		Emergence	
Tasks	Day	Night	Day	Night	Day	Night	Day	Night
Manual	30.6 ± 13.4[a]	24.88 ± 6.7[a]	57.2 ± 13.2[b]	55.9 ± 11.3[b]	22.3 ± 13.2[a]	16.8 ± 6.8[a]	41.7 ± 12.9[b]	37.7 ± 11.2[b]
Observing	27.3 ± 9.3[a]	36.49 ± 10.7[a]	13.7 ± 4.1	12.4 ± 3.9	30.5 ± 13.2[a]	40.5 ± 12.2[a]	27.2 ± 16.5	34.8 ± 13.0
Conversing	11.4 ± 6.5	8.9 ± 3.7	13.2 ± 7.3	13.9 ± 9.0	10.7 ± 7.7	7.8 ± 4.4	13.1 ± 5.6	10.5 ± 6.8
Recording	12.4 ± 4.9	11.2 ± 4.7	0.3 ± 0.7	0.7 ± 1.6	17.8 ± 8.9	15.2 ± 5.8	3.3 ± 4.2	3.0 ± 3.5
Other	18.3 ± 12.5	18.5 ± 5.9	15.6 ± 10.4	17.1 ± 8.7	18.7 ± 15.4	19.8 ± 7.8	14.7 ± 10.9	13.9 ± 8.1

Note. All data are presented as mean ± *SD;* the reported significance levels are based on multiple regression models with months in training, case duration, and hours slept the night prior to the shift as covariates.
[a]$p < .05$, day versus night shift. [b]$p < .05$, compared with other task groups, independent of day/night allocation.

Vigilance and workload. After resident experience, case duration, and sleep hours were controlled for, residents' response latency to the alarm light (a secondary task) at night was not significantly different from that during the day (see Table 4). Similarly, in contrast to our hypothesis, there were no significant differences between night and day cases in the OR observers' workload ratings, in the residents' self-reported workload assessments, or in calculated workload density (see Table 4). In the subset of cases subjected to blinded video review, the observer's workload ratings of the residents during maintenance were significantly lower in the night cases (7.9 ± 0.8) than in the day cases (8.6 ± 1.0; $p < .001$). This contrasted with the ratings by the real-time observer of the same cases (night 9.2 ± 1.3 vs. day 9.1 ± 1.4; $p = 0.5$).

Mood. Both the pre- and postcase mood scores were significantly more negative at night than during the day (see Table 5). There was a general trend for providers' mood to be slightly more positive at the end of the case. Interestingly, time of day was not significantly associated with the changes in mood scores after the cases. In other words, regardless of the time of day, residents experienced similar amounts of mood relief after the case, even though they may have started the case at a more negative mood at night.

Interrater reliability. A direct comparison of interrater reliability of the task data between the live (in OR) and blinded video reviewers revealed acceptable reliability across all cases, with a Spearman's rho correlation coefficient of .78 ± .12.

DISCUSSION

In comparing day and night cases within residents, significant differences in task distribution and negative mood were evident, although clinical workload and alarm response latency did not show significant differences between day and night. The failure to fully support the original hypothesis can be interpreted in several ways. As we will discuss, we argue that the changes in task distribution and mood were attributable to sleeplessness and fatigue. The reason for similar perceived workload ratings and alarm response latency between the study conditions could be explained by participants' compensatory strategies during night work (i.e., load shedding to maintain low workload and attention to primary task cues), methodological issues (e.g., the measures were insufficiently sensitive to fatigue effects), or the possibility that generally young, motivated professionals are much less affected by extended night shift work than has been previously asserted.

Anesthesia residents allocated more attention to primary monitoring tasks during the maintenance phase when working at night. Anesthesiology is similar in some ways to a pilot's tasks in flying, in which takeoff and landing are active, whereas maintaining flight is a lower workload period involving primarily monitoring system status. The maintenance phase of anesthesia is dominated by monitoring of the patient's physiological status, whereas the induction and emergence phases involve more time-critical manual tasks (e.g., inserting and extracting a tube from a patient's airway). Thus, more visual sampling of monitors can be observed in the maintenance phase. The increase in visual sampling at night in this study suggests a compensatory strategy to maintain a high level of performance at a time when fatigue is known to affect vigilance.

Past research has shown that fatigue has an adverse impact on the performance of tasks involving memory, visual encoding, and monitoring, in which novel responses or interventions are required, as well as in other tasks that require constant attention or vigilance (May & Kline, 1987). Workers performing complex monitoring tasks spend more time watching their primary instruments when workload increases (Harris, Tole, Stephens, & Ephrath, 1982).

One explanation for our findings may be that as task load increases (presumably because of diminished attentional capacity and lack of sleep), the resident under stress (fatigue) strategically allocates more time to collecting and verifying information prior to making a decision (Hancock & Dirkin, 1983; Harris et al., 1982). Employing such a sampling strategy may reduce the burden placed upon memory by fatigue. The reallocation of attentional resources serves to reduce the increased demands upon short-term memory during fatigued states that result from extended work hours and sleeplessness.

These results also corroborate previous research findings (Howard et al., 2003; Lamberg, 2002; Owens et al., 2001; Weinger & Ancoli-Israel, 2002) that offer new insights into how health care providers may alter their cognitive and task strategies in an attempt to maintain performance under

TABLE 4: Residents' Vigilance Latency, Workload Density, and Workload Ratings During Day Versus Night Cases

	Vigilance Latency (s)		Workload Density		Observer-Scored Workload (6–10 Scale)		Participant Self-Rated Workload (6–10 Scale)	
Tasks	Day	Night	Day	Night	Day	Night	Day	Night
Induction	106.1 ± 115.8	135.9 ± 142.6	1.5 ± 0.1	1.5 ± 0.1	12.1 ± 2.3	12.4 ± 1.6	12.1 ± 2.9	12.6 ± 2.5
Maintenance	36.5 ± 24.0	30.8 ± 20.7	1.1 ± 0.2	1.1 ± 0.1	9.1 ± 1.4	9.2 ± 1.3	9.8 ± 2.0	9.9 ± 1.8
Emergence	32.4 ± 31.0	41.6 ± 50.3	1.4 ± 0.1	1.2 ± 0.4	10.7 ± 2.0	11.0 ± 2.6	12.2 ± 2.9	11.6 ± 3.4
Total case	42.8 ± 21.1	41.0 ± 27.4	1.2 ± 0.2	1.2 ± 0.1	9.6 ± 1.4	9.7 ± 1.3	10.3 ± 2.2	10.4 ± 1.9

Note: n = 13 paired cases. All data are presented as mean ± *SD*. There is no statistical significant difference in all measures between day and night based on multiple regression models.

TABLE 5: Participants' Mood Before and After Day and Night Cases

	Day	Night
Mood score (precase)[a]	6.2 ± 10.1	23.5 ± 19.1
Mood score (postcase)[a]	5.4 ± 8.0	22.4 ± 15.7
Mood change (postcase–precase)[b]	–0.8 ± 7.4	–1.1 ± 9.3

Note: $n = 13$ paired cases.

[a] $p < .05$. There is significant difference between day and night, after adjusting for months in training, case duration, and hours slept prior to study shift. [b] Negative value indicates more positive mood postcase compared with precase. Nonsignificant difference between day and night.

stress during different clinical situations. As more attention was being allocated to monitoring, less attention was allocated to manual tasks (and perhaps other less essential tasks) during the night cases. Such compensatory strategies may have consequences for patient safety, for example, if unexpected and unusual events were to occur.

Inconsistent with our initial conceptual framework, the response to the alarm light was not slower at night than during the day, nor was perceived workload higher at night than during the day. The task of responding to an alarm light, a secondary task (as compared with the primary task of patient monitoring), may not have been sufficiently meaningful in the clinical environment. The alarm light was an artificial task that required minimal attentional resources and did not seem to affect primary task performance even when the resident was fatigued, resulting in a floor effect. Alternatively, response to the alarm light, which was embedded in the central monitoring array, was confounded at night by the participants' greater attention to their primary monitors. This hypothesis could be tested, for example, by examining the response to peripherally located stimuli. Allowing the participants to respond to the light's illumination either verbally or using hand signals may have introduced variability into the data and obscured possible between-group differences (Type II error).

The lack of between-group differences in workload scores could be explained by the participants' motivation to provide consistently high-quality patient care. Spurgeon and Harrington (1989) suggested that motivation can overcome the effect of fatigue in physicians when maintaining clinical performance is critical. However, our study did not address motivation specifically, and the mood scores at night might be viewed as countervailing evidence. An alternative explanation may be that the observed changes in task distribution at night represent the participant's deliberate effort to maintain workload (and primary task performance) through the shedding of secondary tasks (Harris et al., 1982).

The finding of a greater negative mood at night than during the day is consistent with previous research that showed greater hostility (Hart, Buchsbaum, Wade, Hamer, & Kwentus, 1987), depression (Friedmann, Bigger, & Kornfeld, 1971; Friedmann, Kornfeld, & Bigger, 1973), and irritability (Friedmann et al., 1971, 1973) in sleep-deprived residents. The mood change from case start to end was minimal, suggesting that mood is affected more by the time of day than by the clinical case itself.

Sleep deprivation is known to be associated with more negative mood (Bonnet, 2000; Harrison & Horne, 2000), which can be harmful to both the patient and the health care provider (Mittal, 1998; Sicard, 2001). For example, negative mood may result in lack of compassion for patients (Leung & Becker, 1992), and clinicians who experience chronic stress or dysphoria may also have an increased risk of psychological problems, family problems, substance abuse, or suicide (Samkoff & Jacques, 1991). More research is needed to assess the role of mood, age, and clinical experience in the effects of fatigue on clinical performance.

The results of the present study suggest that changes in clinical task performance at night may be attributable to sleeplessness and fatigue and that at least some fatigue and sleep-deprivation effects, such as negative mood, previously demonstrated in laboratory and nonmedical naturalistic research, do generalize to clinicians providing actual patient care. Other fatigue effects, such as perceived workload and vigilance, may be modified by contextual factors during real work. This is an important consideration for the interpretation of fatigue effects in naturalistic research. This study also demonstrates methods to study physicians' clinical behavior in dynamic patient care settings.

Limitations

Given the fundamental differences in surgical work performed during the day and at night, as well as the constraints of clinical scheduling, a prospective randomized, controlled trial was not feasible. Nighttime cases are generally unplanned, whereas day cases are usually planned. Seven of the 15 observed nighttime cases, but no daytime cases, were classified as true emergencies. Because the nature of a case to be performed at night is unknown in advance, we addressed this confound to the extent possible by controlling for the type and anticipated duration of the surgical procedure, intended anesthesia care, and the patient's preoperative status through prospective case matching in the day case. Nonetheless, there is still a possibility that residual differences in case or patient factors could have affected the results.

Another limitation of this study was the time delay between the two study conditions. In our original design this delay was intended to be 2 weeks or less; however, because of practical constraints in identifying well-matched day cases and then scheduling the appropriate residents, the delay was often longer. With delays of a month or more, it is possible that the residents could have acquired enough additional clinical experience to confound the results. That is, the observed differences may have been attributable to experience and not to fatigue effects. However, we addressed this confounder by adjusting for clinician experience within our statistical models, utilizing the number of months of training for both day and night cases.

We were unable to control for variation in the residents' acute and chronic sleep debt. It is possible, given our relatively small sample size, that individual variation in sleep habits (e.g., the possibility of substantial chronic sleep deprivation in some residents performing during the day; Howard, Gaba, et al., 2002) or the shift work schedule in the week prior to the study (Tepas, Paley, & Popkin, 1997) could have obscured or reduced real differences between day and night cases. Most residents reported having had more sleep the night prior to working nights than the night prior to participating in the daytime case.

As is true with all applied research, the residents were aware that they were being studied (i.e., Hawthorne effect) and may have compensated for their acute sleep loss by expending extra effort at night to be more active, vigilant, or diligent. This too may have decreased differences observed between day and nighttime work. Finally, we were not able to control for the amount of direct involvement by attending physicians or for the number or length of participant breaks during study cases. These variables were not the primary contrasts of interest in this study and were not included as covariates in the regression models. Future studies would be required to address these possible confounds.

Potential Applications

Unlike in other high-risk industries (e.g., commercial transportation), at least in the United States, there are no mandatory restrictions on practicing physicians' work hours (Gaba & Howard, 2002). The Accreditation Council for Graduate Medical Education and some state legislatures have begun to apply restrictions to resident work hours, but the success and impact of these efforts is still uncertain (Poulose et al., 2005).

Recent evidence suggests that reducing and restructuring the shift schedules of medical interns can decrease serious medical errors (Landrigan et al., 2004). However, placing even stricter limits on residents' work schedules is a hotly debated issue involving questions of economics, quality of patient care, and quality of clinical training. Research on the effect of fatigue on human performance in other domains suggests that the time of day in which a task is actually conducted is as important as, if not more important than, the amount of sleep prior to completing the task (Hartley, Arnold, Penna, Corry, & Feyer, 1997; Smiley, 1998). The present study contributes new data and methods to the growing literature on this important public policy issue.

This study has potential applications in restructuring resident work hours to avoid undue disruption to circadian rhythm and sleep loss, such as work-rest schedules. For example, an 80-hr work week could consist of evenly distributed on-duty periods separated by adequate periods of rest. However, determination of the minimum duration of rest periods may be domain specific and thus requires further research. Nevertheless, adjustments to work schedules to better accommodate circadian rhythm disruptions and sleep loss effects are probably needed. This work also has methodological applications in patient safety research. The matched-case control scheme for investigating effects of fatigue during actual patient care can be adapted to study other phenomena in the clinical environment when a randomized, controlled study is not feasible.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the contributions of Drs. David Kobus, Ron Thomas, Amy Bronstone, and Sonia Jain. Review of the manuscript by Drs. Sonia Ancoli-Israel and Robert Kaplan and long-time support and encouragement by Drs. David Gaba and Steve Howard were greatly appreciated. This study would not have been possible without the indulgence and cooperation of the participating anesthesia residents, their faculty supervisors, and the dedicated operating room nurses. This work was supported by grants to Dr. Weinger from the Department of Veterans Affairs (HSRD IIR HR20-066) and the Agency for Healthcare Research and Quality (R01-HS11375 and P20-HS11521). The writing of this paper was also made possible with support from the Tufts Mellon Grant to Dr. Cao and endowment funds from Vanderbilt University for Dr. Weinger.

REFERENCES

Akerstedt, T. (1991). Sleepiness at work: Effects of irregular work hours. In T. H. Monk (Ed.), *Sleep, sleepiness and performance* (pp. 129–152). New York: Wiley.

Barger, L. K., Ayas, N. T., Cade, B. E., Cronin, J. W., Rosner, B., Speizer, F. E., et al. (2006). Impact of extended-duration shifts on medical errors, adverse events, and attentional failures. *Public Library of Science Medicine, 3,* e487.

Bonnet, M. H. (2000). Sleep deprivation. In M. H. Kryger, T. Roth, & W. C. Dement (Eds.), *Principles and practice of sleep medicine* (3rd ed., pp. 53–71). Philadelphia: W. B. Saunders.

Borg, G. (1977). Simple rating methods for estimation of perceived exertion. In G. Borg (Ed.), *Physical work and effort* (Vol. 28, pp. 39–47). Oxford, UK: Permagon Press.

Borg, G. (1998). *Borg's perceived exertion and pain scales.* Stockholm: Human Kinetics.

Carrier, J., & Monk, H. M. (1999). Effects of sleep and circadian rhythms on performance. In F. W. Turek & P. C. Zee (Eds.), *Regulation of sleep and circadian rhythms* (pp. 527–556). New York: Marcel Dekker.

Daugherty, S. R., Baldwin, D. C., Jr., & Rowley, B. D. (1998). Learning, satisfaction, and mistreatment during medical internship: A national survey of working conditions. *Journal of the American Medical Association, 279,* 1194–1199.

Dawson, D., & Reid, K. (1997). Fatigue, alcohol and performance impairment. *Nature, 388,* 235.

Eastridge, B. J., Hamilton, E. C., O'Keefe, G. E., Rege, R. V., Valentine, R. J., Jones, D. J., et al. (2003). Effect of sleep deprivation on the performance of simulated laparoscopic surgical skill. *American Journal of Surgery, 186,* 169–174.

Eysenck, M. W., & Eysenck, M. C. (1979). Processing depth, elaboration of encoding, memory stores, and expending processing capacity. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 472–484.

Fletcher, K. E., Underwood, W., Davis, S. Q., Mangrulkar, R. S., McMahon, L. F., & Saint, S. (2005). Effects of work hour reduction on residents' lives: A systematic review. *Journal of the American Medical Association, 294,* 1088–1100.

Ford, C. V., & Wantz, D. K. (1984). The internship year: A study of sleep, mood states, and psychophysiological parameters. *Southern Medical Journal, 77,* 1435–1442.

Friedmann, R. C., Bigger, J. T., & Kornfeld, D. S. (1971). The intern and sleep loss. *New England Journal of Medicine, 285,* 201–203.

Friedmann, R. C., Kornfeld, D. S., & Bigger, J. T. (1973). Psychological problems associated with sleep deprivation in interns. *Journal of Medical Education, 48*, 436–441.

Gaba, D. M., & Howard, S. K. (2002). Patient safety: Fatigue among clinicians and the safety of patients. *New England Journal of Medicine, 347,* 1249–1255.

Gaba, D. M., Howard, S. K., & Jump, B. (1994). Production pressure in the work environment: California anesthesiologists' attitudes and experiences. *Anesthesiology, 81,* 488–500.

Gelfand, D. V., Podnos, Y. D., Carmichael, J. C., Saltzman, D. J., Wilson, S. E., & Williams, R. A. (2004). Effect of the 80-hour workweek on resident burnout. *Archives of Surgery, 139,* 933–940.

Grantcharov, T. P., Bardram, L., Funch-Jensen, P., & Rosenberg, J. (2001). Laparoscopic performance after one night on call in a surgical department: Prospective study. *British Medical Journal, 323,* 1222–1223.

Hancock, P. A., & Dirkin, G. R. (1983). Stressor induced attentional narrowing: Implications for design and operation of person-machine systems. In *Proceedings of the Human Factors Association of Canada 16th Annual Meeting* (pp. 19–21). Mississauga, Ontario, Canada: Human Factors Association of Canada.

Harris, R. L., Tole, J. R., Stephens, A. T., & Ephrath, A. R. (1982). Visual scanning behavior and pilot workload. *Aviation, Space, and Environmental Medicine, 53,* 1067–1072.

Harrison, Y., & Horne, J. A. (2000). The impact of sleep deprivation on decision making: A review. *Journal of Experimental Psychology: Applied, 6,* 236–249.

Hart, R. P., Buchsbaum, D. G., Wade, J. B., Hamer, R. M., & Kwentus, J. A. (1987). Effect of sleep deprivation on first-year residents' response times, memory, and mood. *Journal of Medical Education, 62,* 940–942.

Hartley, L. R., Arnold, P. K., Penna, F., Corry, A., & Feyer, A. M. (1997). *Comprehensive review of fatigue research* (Rep. No. 116). Perth, Australia: Murdoch University, Institute for Research in Safety and Transport.

Howard, S. K., Gaba, D. M., Rosekind, M. R., & Zarcone, V. P. (2002). Excessive daytime sleepiness in resident physicians: Risks, intervention, and implication. *Academic Medicine, 77,* 1019–1025.

Howard, S. K., Gaba, D. M., Smith, B., Weinger, M. B., Herndon, C., Keshavacharya, S., et al. (2003). Simulation study of rested versus sleep deprived anesthesiologists. *Anesthesiology, 98,* 1345–1355.

Howard, S. K., Rosekind, M. R., Katz, J. D., & Berry, A. J. (2002). Fatigue in anesthesia. *Anesthesiology, 97,* 1281–1294.

Lamberg, L. (2002). Long hours, little sleep: Bad medicine for physicians-in-training? *Journal of the American Medical Association, 287,* 303–307.

Landrigan, C. P., Rothschild, J. M., Cronin, J. W., Kaushal, R., Burdick, E., Katz, J. T., et al. (2004). Effect of reducing interns' work hours on serious medical errors in intensive care units. *New England Journal of Medicine, 351,* 1838–1848.

Lavie, P. (1991). The 24-hour sleep propensity function (SPF): Practical and theoretical implications. In T. H. Monk (Ed.), *Sleep, sleepiness and performance* (pp. 65–93). New York: Wiley.

Leonard, C., Fanning, N., Attwood, J., & Buckley, M. (1998). The effect of fatigue, sleep deprivation and onerous working hours on the physical and mental wellbeing of pre-registration house officers. *Irish Journal of Medical Sciences, 167,* 22–25.

Leung, L., & Becker, C. E. (1992). Sleep deprivation and house staff performance. *Journal of Occupational Medicine, 34,* 1153–1160.

Lockley, S. W., Cronin, J. W., Evans, E. E., Cade, B. E., Lee, C. J., Landrigan, C. P., et al. (2004). Effect of reducing interns' weekly work hours on sleep and attentional failures. *New England Journal of Medicine, 351,* 1829–1837.

May, J., & Kline, P. (1987). Measuring the effects upon cognitive abilities of sleep loss during continuous operations. *British Journal of Psychology, 78,* 443–455.

McNair, D. M., Lorr, M., & Droppleman, L. F. (1992). *EdITS manual for the Profile of Mood States*. San Diego, CA: Educational & Industrial Testing Service.

Mittal, V. (1998). The impact of positive and negative affect and issue framing on issue interpretation and risk taking. *Organizational Behavior and Human Decision Processes, 76,* 298–324.

Myles, P. S., Hunt, J. O., Nightingale, C. E., Fletcher, H., Beh, T., Tanil, D., et al. (1999). Development and psychometric testing of a quality of recovery score after general anesthesia and surgery in adults. *Anesthesia and Analgesia, 88,* 83–90.

Myles, P. S., Weitkamp, B., Jones, K., Melick, J., & Hensen, S. (2000). Validity and reliability of a postoperative quality of recovery score: The QoR-40. *British Journal of Anaesthesia, 84,* 11–15.

Nelson, C. S., Dell'Angela, K., Jellish, W. S., Brown, I. E., & Skaredoff, M. (1995). Residents' performance before and after night call as evaluated by an indicator of creative thought. *Journal of the American Osteopathic Association, 95,* 600–603.

Owens, J. A., Veasey, S. C., & Rosen, R. C. (2001). Physician, heal thyself: Sleep, fatigue, and medical education. *Sleep, 24,* 493–495.

Parasuraman, R. (1985). Sustained attention: A multifactorial approach. In M. I. Posner & O. Marin (Eds.), *Attention and performance XI: Mechanism of attention (International Symposium on Attention and Performance)* (pp. 493–511). Mahwah, NJ: Erlbaum.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-plus*. New York: Springer.

Poulose, B. K., Wayne, A. R., Patrick, G. A., Needleman, J., Buerhaus, P. I., Griffin, M. R., et al. (2005). Resident work hour limits and patient safety. *Annals of Surgery, 241*, 847–860.

R Development Core Team. (2006). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available from http://www.R-project.org

Samkoff, J. S., & Jacques, C. H. M. (1991). A review of studies concerning effects of sleep deprivation and fatigue on residents' performance. *Academic Medicine, 66,* 687–693.

Sicard, B. (2001). Risk propensity assessment in military special operations. *Military Medicine*, *166,* 871–874.

Slagle, J., Weinger, M. B., Dinh, M.-T. T., Wertheim, V. V., & Williams, K. (2002). Intra- and inter-rater reliability of a clinical task analysis methodology. *Anesthesiology, 96,* 1129–1139.

Smiley, A. (1998). Fatigue management: Lessons from research. In L. Hartley (Ed.), *Managing fatigue in transportation* (pp. 1–23). Oxford, UK: Elsevier Science.

Smith-Coggins, R., Rosekind, M. R., Buccino, K. R., Dinges, D. F., & Moser, R. P. (1997). Rotating shiftwork schedules: Can we enhance physician adaptation to night shifts? *Academic Emergency Medicine, 4,* 951–961.

Spurgeon, A., & Harrington, J. M. (1989). Work performance and health of junior hospital doctors – A review of the literature. *Work and Stress, 3,* 117–128.

Taffinder, N. J., McManus, I. C., Gul, Y., Russell, R. C., & Darzi, A. (1998). Effect of sleep deprivation on surgeons' dexterity on laparoscopy simulator. *Lancet, 352,* 1191.

Tepas, D. I., Paley, M. J., & Popkin, S. M. (1997). Work schedules and sustained performance. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (2nd ed., pp. 1021–1058). New York: Wiley.

Turner-Stokes, L., Tonge, P., Nyein, K., Hunter, M., Nielson, S., & Robinson, I. (1998). The Northwick Park Dependency Score (NPDS): A measure of nursing dependency in rehabilitation. *Clinical Rehabilitation, 12,* 304–318.

Veasey, S. C., Rosen, R. C., Baransky, B., Rosen, I., & Owens, J. (2002). Sleep loss and fatigue in residency training: A reappraisal. *Journal of the American Medical Association, 288,* 1116–1124.

Vredenburgh, A. G., Weinger, M. B., Williams, K. J., Kalsher, M. J., & Macario, A. (2000). Developing a technique to measure anesthesiologists' real-time workload. In *Proceedings of the XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 4.241–4.244). Santa Monica, CA: Human Factors and Ergonomics Society.

Weinger, M. B., & Ancoli-Israel, S. (2002). Sleep deprivation and clinical performance. *Journal of the American Medical Association, 287,* 955–957.

Weinger, M. B., & Englund, C. E. (1990). Ergonomic and human factors affecting anesthetic vigilance and monitoring performance in the operating room environment. *Anesthesiology, 73,* 995–1021.

Weinger, M. B., Herndon, O. W., & Gaba, D. M. (1997). The effect of electronic record keeping and transesophageal echocardiography on task distribution, workload, and vigilance during cardiac anesthesia. *Anesthesiology, 87,* 144–155.

Weinger, M. B., Herndon, O. W., Paulus, M. P., Gaba, D., Zornow, M. H., & Dallen, L. D. (1994). Objective task analysis and workload assessment of anesthesia providers. *Anesthesiology, 80,* 77–92.

Wesnes, K., Walker, M., Walker, L., White, L., Warren, R., & Eremin, O. (1997). Cognitive performance and mood after a weekend on call in a surgical unit. *British Journal of Surgery, 84,* 493–495.

Caroline G. L. Cao is the director of the Human Factors Program in the School of Engineering, an assistant professor of mechanical engineering and of biomedical engineering, and the founding director of the Ergonomics in Remote Environments Laboratory at Tufts University, Medford, Massachusetts. She received her Ph.D. in mechanical and industrial engineering from the University of Toronto in 2002.

Matthew B. Weinger holds the Norman Ty Smith Chair in Patient Safety and Medical Simulation and is a professor of anesthesiology, biomedical informatics, and medical education at the Vanderbilt University School of Medicine. He also serves as the director of the Vanderbilt Center for Perioperative Research in Quality, Nashville, Tennessee. He received his M.D. from the University of California, San Diego, in 1982.

Jason Slagle is an assistant professor in the Department of Anesthesiology at Vanderbilt University and is a faculty member in the Vanderbilt Center for Perioperative Research in Quality, Nashville, Tennessee. He received his Ph.D. in industrial and organizational psychology from Alliant University, San Diego, California, in 2004.

Chuan Zhou is an assistant professor of biostatistics at Vanderbilt University, Nashville, Tennessee. He received his Ph.D. in biostatistics from the Department of Biostatistics at the University of Washington, Seattle, Washington, in 2003.

Jennie Ou is a pediatrician at Sharp Rees-Stealy Medical Group in San Diego, California. She received her M.D. from the University of California, San Diego, in 2007.

Shakha Gillin (née Vora) is in private practice at El Camino Pediatrics in Encinitas, California, and is an adjunct assistant professor of pediatrics at the University of California, San Diego, where she received her M.D. in 1998.

Bryant Sheh is a resident at the Olive View Medical Center in Sylmar, California. He received his M.D. from the University of California, San Diego, in 2007.

William Mazzei is a clinical professor of anesthesiology at the University of California, San Diego, and the director of perioperative services at the University of California San Diego Medical Center. He received his M.D. from Harvard Medical School, Boston, Massachusetts, in 1981.

Date received: April 7, 2006
Date accepted: January 18, 2008