

Simulation Study of Rested Versus Sleep-deprived Anesthesiologists

Steven K. Howard, M.D.,* David M. Gaba, M.D.,† Brian E. Smith, M.D.,‡ Matthew B. Weinger, M.D.,§ Christopher Herndon, B.A.,|| Shanthala Keshavacharya, M.D.,# Mark R. Rosekind, Ph.D.**

Background: Sleep deprivation causes physiologic and subjective sleepiness. Studies of fatigue effects on anesthesiologist performance have given equivocal results. The authors used a realistic simulation environment to study the effects of sleep deprivation on psychomotor and clinical performance, subjective and objective sleepiness, and mood.

Methods: Twelve anesthesia residents performed a 4-h anesthetic on a simulated patient the morning after two conditions of prior sleep: sleep-extended (EXT), in which subjects were allowed to arrive at work at 10:00 AM for 4 consecutive days, thus allowing an increase in nocturnal sleep time, and total sleep deprivation (DEP), in which subjects were awake at least 25 h. Psychomotor testing was performed at specified periods throughout the night in the DEP condition and at matched times during the simulation session in both conditions. Three types of vigilance probes were presented to subjects at random intervals as well as two clinical events. Task analysis and scoring of alertness were performed retrospectively from videotape.

Results: In the EXT condition, subjects increased their sleep by more than 2 h from baseline ($P = 0.0001$). Psychomotor tests revealed progressive impairment of alertness, mood, and performance in the DEP condition over the course of the night and when compared with EXT during the experimental day. DEP subjects showed longer response latency to vigilance probes, although this was statistically significant for only one probe

type. Task analysis showed no difference between conditions except that subjects "slept" more in the DEP condition. There was no significant difference in the cases' clinical management between sleep conditions. Subjects in the DEP condition had lower alertness scores ($P = 0.02$), and subjects in the EXT condition showed little video evidence of sleepiness.

Conclusions: Psychomotor performance and mood were impaired while subjective sleepiness and sleepy behaviors increased during simulated patient care in the DEP condition. Clinical performance between conditions was similar.

ANESTHESIOLOGISTS often work extended duty shifts that result in acute and chronic sleep loss and circadian disruption.¹ A workweek of greater than 60 h is common, and residents sometimes work more than 80 h in a week.²⁻⁴ Although anesthesiology residents are precluded by guidelines of the Accreditation Council for Graduate Medical Education from administering anesthesia the day after in-house call, they are not prohibited from working more than 24 h at a time in other settings, including the intensive care unit.* There are no regulations or guidelines concerning duty and rest periods for experienced anesthesiologists or nurse anesthetists, either. Sleep deprivation is not only an issue for the immediate postcall period. We recently reported that anesthesia residents who are not postcall have a level of daytime sleepiness similar to that seen in patients with narcolepsy or sleep apnea.¹ These effects were reversed by 2 h additional sleep for 4 consecutive days.

The effects of chronic sleep deprivation on performance have been studied in physicians, but the results of such studies are mixed.⁵⁻⁹ In the laboratory, a single night of sleep loss can produce measurable performance decrements on psychomotor tasks.^{10,11} With sleep loss, subjects exhibit a progressive decrease in reaction time, increased response time variability, and a propensity to fall asleep (epochs of sleep lasting for as little as a few seconds are called "microsleeps").^{12,13} Sleep-deprived workers fail to appropriately allocate attention, set task priorities, or sample for sources of potentially faulty information.¹⁴⁻¹⁶ Whether clinical performance is similarly degraded has not been demonstrated.

We conducted a study of the performance of anesthesiology residents during routine long simulated surgical cases under different conditions of prior sleep. We compared subjects after 25-30 h of total sleep deprivation and the same individuals after 2 h of additional sleep for

This article is featured in "This Month in Anesthesiology." Please see this issue of ANESTHESIOLOGY, page 5A.

Additional material related to this article can be found on the Anesthesiology Web site. Go to the following address, click on Enhancements Index, and then scroll down to find the appropriate article and link. <http://www.anesthesiology.org>

* Associate Director, Patient Safety Center of Inquiry, VA Palo Alto Health Care System. Associate Professor of Anesthesia, Department of Anesthesia, Stanford University School of Medicine. † Director, Patient Safety Center of Inquiry, VA Palo Alto Health Care System. Professor of Anesthesia, Department of Anesthesia, Stanford University School of Medicine. ‡ Fellow, Department of Anesthesia, Stanford University School of Medicine. § Director, San Diego Center for Patient Safety. Director, Anesthesia Ergonomics Research Laboratory, VA San Diego Healthcare System. Professor of Anesthesiology, Department of Anesthesiology, University of California, San Diego School of Medicine. || Research Assistant, Anesthesia Ergonomics Research Laboratory, VA San Diego Healthcare System. # Research Associate, VA Palo Alto Health Care System. ** President and Chief Scientist, Alertness Solutions, Inc., Cupertino, California.

Received from the Anesthesia Service, VA Palo Alto Health Care System, Palo Alto, California; the Department of Anesthesia, Stanford University School of Medicine, Palo Alto, California; the Anesthesiology and HSRD Services, VA San Diego Medical Center, La Jolla, California; and the Department of Anesthesiology, University of California, San Diego School of Medicine, La Jolla, California. Submitted for publication October 14, 2002. Accepted for publication January 14, 2003. Supported by the Patient Safety Center of Inquiry Program, Department of Veterans Affairs, Washington, DC (Drs. Gaba and Howard); the Office of Technology Licensing, Stanford University, Stanford, California (Drs. Gaba and Howard); the Anesthesia Patient Safety Foundation, Pittsburgh, Pennsylvania; and the Agency for Healthcare Research and Quality, Rockville, Maryland (Dr. Weinger).

Address reprint requests to Dr. Howard: Anesthesia Service, 112A, 3801 Miranda Avenue, Palo Alto, California 94304. Address electronic mail to: showard@stanford.edu. Individual article reprints may be purchased through the Journal Web site, www.anesthesiology.org.

* Duty hours and conditions of work: Accreditation Council for Graduate Medical Education program requirements for anesthesiology. Available at: <http://www.acgme.org>. Accessed October 9, 2002

4 consecutive days. We used patient simulation rather than real cases because (1) there is no risk to real patients if errors are made; (2) the same cases can be presented reproducibly to each subject; (3) the subjects can be monitored extensively; (4) the key independent variables (e.g., sleep deprivation *vs.* sleep extension) can be manipulated more easily; and (5) performance probes and abnormal clinical events can be presented at predetermined times under controlled conditions. We hypothesized that the patterns and adequacy of performance (psychomotor and clinical) during a long anesthetic would be different for residents who were sleep-deprived relative to the patterns and adequacy of performance seen when they were well rested. We also hypothesized that acute sleep deprivation would result in an increased propensity to fall asleep even when conducting simulated patient care. The amount of total sleep deprivation in this study replicates that faced by anesthesiologists when they provide patient care the morning after a long on-call period, a practice that still occurs for experienced clinicians.⁴

Materials and Methods

The joint Institutional Review Board of Stanford University and VA Palo Alto Health Care System approved the study protocol. Twelve anesthesia residents participated in the study after providing written informed consent. Each had prior experience working in the simulator facility during one or more full-day sessions of Anesthesia Crisis Resource Management training.^{17,18}

Prestudy Period

In the week preceding each simulation session, subjects completed the Sleep Disorders Questionnaire¹⁹ and the Owl and Lark Questionnaire.²⁰ The Sleep Disorders Questionnaire is used to test for the presence of clinical sleep disorders, whereas the Owl and Lark Questionnaire evaluates the subjects' preference for daytime or nighttime work. To measure their sleep-wake cycles, subjects wore a wrist activity monitor²¹ (AMA-32; Ambulatory Monitoring, Inc., Ardsley, NY) continuously and completed a computerized sleep log once daily for 1 week before the experiment. Action3 software (Ambulatory Monitoring, Inc.) was used to score actigraph sleep and wake periods.

Sleep Conditions

In the sleep-deprived condition (DEP), subjects were kept awake for at least 25 h before the simulated case. They performed a regular day's work in the operating room (OR) the previous day and then began a pseudo-call night accompanied by an investigator who ensured that they did not sleep. When possible, they assisted a real on-call team; otherwise, they did other activities

(e.g., read, played cards). During the pseudo-call night, subjects completed a 15-min psychomotor test battery every 2 h from 22:00 through 06:00 the morning of the experiment. In the sleep-extended condition (EXT), subjects were instructed to maximize their sleep for 4 consecutive nights before being studied in the simulator. Subjects were relieved of clinical duties until 10:00 AM each morning of sleep extension.

Subjects were instructed to refrain from drinking caffeinated beverages for 24 h before each simulation session. To minimize training effects, the order of the sleep conditions was randomized.

Simulated Laparoscopic Surgery Sessions

Subjects conducted anesthesia for two similar 4-h simulated laparoscopic surgery cases on separate days, once under each condition of prior sleep. For each subject, the two simulation sessions were conducted no more than 30 days apart.

The Simulator and Simulation Environment. The MedSim/Eagle Patient Simulator (Binghamton, NY) used in this study consisted of a computer control system and a patient mannequin that generated physiologic signals and allowed for realistic airway management. The simulator generated responses appropriate for a patient under anesthesia using mathematical models of physiologic systems and of anesthesia drug pharmacokinetics and pharmacodynamics. Subjects could query the investigator (sequestered in a control room) for information regarding skin color, diaphoresis, or other clinical attributes not supported by the simulator. The mannequin had a speaker behind its head; its voice was played by one of the investigators, making it also a "standardized patient."^{22,23} This allowed the subject to clarify with the "patient" issues from the preoperative assessment.

The simulated OR was equipped with an operating table, a surgical light, a Modulus II Plus anesthesia machine, and an AS3 physiologic monitor with a Capnomac Ultima respiratory gas analyzer (all Datex-Ohmeda products, Madison, WI). A picture of the monitoring array can be found in the Web Enhancement. A standard fully stocked anesthesia supply cart was provided. A complete patient chart was present, as were blank anesthesia records. Subjects were instructed to conduct all clinical activities as they would during real patient care.

For this experiment, the auditory alarms on the physiologic monitors were preset to institutional default values, and subjects were instructed not to alter the alarm limits. The simulator was run primarily in its autonomous mode. Response to drug administration or clinical maneuvers occurred due to the deterministic predictions of the simulator's internal pharmacologic and physiologic models. Prescribed events (embedded vigilance probes and abnormal clinical events) were initiated at predetermined but randomly distributed intervals through the case. Manual control of simulator functions

was exercised as little as possible and followed preestablished written protocols.

Conduct of Simulated Cases. Prior to each simulated case, subjects were given 20 min to set up their anesthesia equipment. Subjects then left the OR to review a written preoperative evaluation of the patient, after which they reentered the OR and were allowed to interview the simulated patient. They then began their conduct of anesthesia.

During simulated laparoscopy, the room lights were off but the anesthesia work area was illuminated with an overhead surgical light. An investigator played the role of surgeon by using laparoscopic instruments to mimic the surgical actions displayed on a videotape of real laparoscopic surgery. A retired OR nurse acted as the circulating nurse. Thus, subjects could interact normally with the surgeon and nurse, who were instructed to be pleasant and responsive, but not to engage the subjects in conversation. Classical music played softly in the background during all cases.

Subjects did not know how long the case would last. They were provided with a 30-min lunch break in normal office illumination at a predefined time point, approximately halfway through each case. They were relieved by another anesthesiologist using a typical relief protocol. Data collection was suspended during the lunch break. When subjects returned from lunch, they were briefed on the status of the patient and data collection resumed. After 4 h of case time, the subjects were given an afternoon break, and the simulation was stopped.

Simulated Cases. Two simulated patients of similar apparent clinical difficulty were developed (case A and case B). Both simulated patients had medical conditions consistent with an American Society of Anesthesiologists physical status of III and required laparoscopic surgery (Nissen fundoplication or lysis of peritoneal adhesions). Half of the subjects (selected randomly) performed case A under the DEP condition and case B under the EXT condition. The opposite allocation was used for the other subjects.

Experimental Day and Performance Measures

In each of the two sleep conditions, subjects arrived at the laboratory at approximately 08:00. The course of the experimental day is shown in figure 1. During each simulated case, the subjects' psychomotor and clinical performance was measured using a variety of techniques.

Psychomotor Test Battery. A battery of psychomotor tests was administered on three occasions during each experiment: once prior to initiation of the experiment, the second time prior to eating lunch (mid-experiment), and a final time on completion of the simulation. The methods for the psychomotor test battery are described in the Appendix.

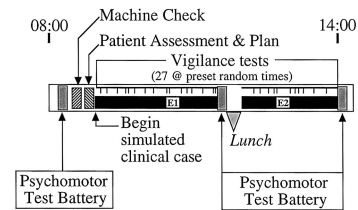


Fig. 1. Schematic diagram of the experimental day for both conditions of prior sleep. In the sleep-deprived condition, subjects spent the previous night in the hospital without sleep. *E1* and *E2* are abnormal clinical events that were inserted during the simulated case (see text for times).

Check of Anesthesia Equipment with Known Faults. In case A, the laryngoscope had dead batteries, and the ventilator hose was not connected to the carbon dioxide absorber assembly. In case B, airway suction tubing was missing, and the isoflurane vaporizer was left on at 2%. The subject's machine checkout was timed and scored in real-time by three observers and again from the videotape by a single observer blinded to subject condition. In addition to detection of the faults, the checkouts were scored for the completion of 11 tasks. After the subjects completed the equipment check and left the OR, the investigators restored all equipment to full working order.

Clinical Management of Preoperative Conditions. In case A, the simulated patient had severe gastroesophageal reflux disease due to a hiatal hernia and was also known to be allergic to tetracycline. In case B, the patient had a history of severe gastroesophageal reflux disease and was known to be allergic to cephalosporin antibiotics. Both patients had a normal (Mallampati class I) airway examination. We expected subjects to perform a rapid sequence induction in these patients who were at risk of pulmonary aspiration of gastric contents. In each case, the surgeon requested an intravenous antibiotic in the class to which the patient had a known allergy. We expected subjects to request an alternate class of antibiotic.

Vigilance Probes. Three types of task probes were presented to subjects during each simulated case to assess vigilance and/or spare capacity.²⁴⁻²⁷ Subjects were instructed to respond verbally as soon as any probe was detected. The probe types were (1) illumination of a red light placed next to the physiologic monitor; (2) sudden change of the normal arterial waveform to a flat line reading 0 mmHg; and (3) a "fully embedded" probe, whereby either the blood pressure or heart rate ramped up or down (as predetermined by the protocol) at a constant rate (typically over a 10- to 30-s time period), eventually to cross a reporting threshold that was explained to subjects prior to each case (heart rate < 70 or > 100 beats/min; mean arterial pressure < 70 or > 100 mmHg). There were nine probes of each type during each case (total of 27). These occurred at preset but randomly distributed elapsed case times, set separately

for case A *versus* case B (so that subjects could not anticipate a specific probe by its elapsed time in the second simulation session). The response time was defined as the time between the onset of the probe (or threshold crossing for fully embedded probes) and the subject's response. If the subject did not respond after 3 min, the probe was terminated and the maximum response time (180 s) was recorded.

These three probe types were chosen because they have different characteristics. The red light is sudden in onset and unequivocal in occurrence but is not part of the clinical data stream. This probe has been used extensively in studies of vigilance or spare capacity in real cases.²⁴⁻²⁷ The arterial pressure waveform change is sudden in onset and part of the clinical data stream but represents a counterfactual occurrence (the patient's true pulse and pressure were unaffected). The fully embedded probes are real data streams behaving in a plausible fashion, but they are not sudden in onset because the progressive changes predict the upcoming threshold crossing.

Abnormal Clinical Events. Two clinical events were inserted into each case; they were designed to be detected by routine clinical observations and to provoke evaluation and treatment. In each case, one event was pulmonary and one was cardiac. The events were *bronchospasm* (wheezing on auscultation, a decreased pulmonary compliance to yield a 75% increase in peak inspiratory pressure, and increased shunt fraction to yield a decrease in oxygen saturation from 99% to 95%); *atrial fibrillation* (irregularly irregular rhythm, heart rate of 100 beats/min, and a decrease in peripheral resistance to yield a 20% decrease in blood pressure); *myocardial ischemia* (2-mm ST depression and reduction of myocardial contractility to yield a 15% reduction in systolic blood pressure); and *atelectasis* (increased shunt fraction to yield oxygen saturation decrease from 99% to 92% and decreased pulmonary compliance to yield a 25% increase in peak inspiratory pressure). Case A had bronchospasm at 95 min elapsed time and atrial fibrillation at 220 min elapsed time. Case B had myocardial ischemia at 100 min elapsed time and atelectasis at 210 min elapsed time.

Events resolved without sequelae when either the subject initiated any one of several acceptable predetermined corrective actions or a preset maximum duration elapsed (10 min except for myocardial ischemia, which terminated after 15 min to allow time to initiate treatment such as nitroglycerin infusion). Observers blinded to sleep condition reviewed the videotapes to record response times to detection (evidenced by verbalization or by action) and to treatment.

Audio-Video Data Collection. All simulation activities were captured on time-coded videotape for retrospective review. The three views were (1) wide-angle view of the anesthetizing location shot toward the anes-

Table 1. Scoring of Alertness

Behavior	Score
Awake, working	6
Eyes open, staring into space	5
Slow, rolling eye movements, slow eyelid blinking	4
Eyes closed, "head nodding" (loss of neck muscle tone allowing head to fall), eyes open when head nods	3
Eyes closed, head nodding, eyes do <i>not</i> open when head nods	2
Eyes closed, no movement, sleeping	1

thesia machine and monitor (camera 1), (2) a view of the subject's face by a camera placed above the anesthesia machine (camera 2), and (3) the display output of the physiologic monitor.

Task Analysis. A single trained observer (C. N. H.) blinded to the subjects' sleep condition performed a task analysis of each case from the videotapes, primarily using the camera 1 view. The task analysis technique has been well validated and described in detail in the investigators' previous publications.^{24,25,28,29} In brief, the activities of subjects were resolved into 37 specific predefined task categories using custom software on a Macintosh computer.^{24,25} For this task analysis, a category of "sleeping" was added to the categories used in previous studies and was defined as "the subjects' eyes are closed *and* there is no detectable body movement." A total of 96 h of videotape was analyzed in this manner.

Assessment of Behavioral Alertness. Since the task analysis showed that some subjects were sleeping during the simulated cases, the videotapes of camera view 2 (subjects' faces) were also rated by a different observer (S. K.), blinded as to sleep condition, for signs of behavioral alertness using a six-point ordinal scale that included behaviors of profound sleepiness intermediate between clearly awake and completely asleep (table 1). These ratings were conducted with 1-s resolution using MacShapa video annotating software (CSERIAC, Wright-Patterson AFB, OH) from the beginning of the simulated anesthetic until the end of the study case (excluding the lunch break). The rater was trained on the alertness scale using data from the first subject's cases (8 h of rating). The subsequent analysis was performed on the data from the remaining 11 subjects (22 cases; 88 h total).

Postsimulation Questionnaire. The subjects completed a postsimulation questionnaire after each session. The questions on this one-page survey used five-point Likert scales to assess the simulated cases' perceived realism, clinical difficulty, and similarity to each other.

Data Analysis and Statistical Techniques

Data analysis and statistical analyses were conducted using a variety of software, including Microsoft Excel 98 (Microsoft, Redmond, WA), Statview 4.1 and SuperANOVA (both from Abacus Concepts, Berkeley, CA), and STATISTICA MAC 4.1 (Oklahoma City, OK). In general,

the experiment comprised a nested repeated-measures design in which subjects were their own controls for two sleep conditions and various measures were repeated throughout the simulated on-call night and the simulation sessions. Where possible, comparisons between equivalent time points in the simulation sessions were analyzed using nested repeated-measures analysis of variance using SuperANOVA, with significance levels corrected for sphericity by Greenhouse-Geisser epsilon. Ordinal data and proportions were analyzed nonparametrically (Wilcoxon signed-rank test for paired data and Mann-Whitney U for nonpaired data). Vigilance probe response times deviated markedly from a normal distribution and were also analyzed nonparametrically. Nominal data (e.g., detection of clinical events) were tested using chi-square. Aggregate data in tables and graphs are shown as mean \pm SD unless otherwise specified. Statistical significance was considered at $P < 0.05$.

Task Analysis and Workload Assessment. Task data from each case were processed and collated using custom software as described in previous publications.^{24,25,29} The total time and percent of each phase spent on each task category, the duration of individual occurrences of each task ("dwell time"), and the frequency of occurrence of individual tasks were calculated for each case. Task analysis data were analyzed using a two-way analysis of variance with tasks performed analyzed as within-subjects' variables and with sleep condition as a between-subjects variable. Significant main effects were assessed using Newman-Keuls *a posteriori* tests. The blinded observer also scored the workload of the anesthesia providers at random 7- to 12-min intervals during the simulated cases under both sleep conditions. Psychological workload was scored using a standardized scale ranging from 6 (e.g., completely sedentary) to 20 (e.g., in the middle of a full-blown cardiac arrest resuscitation).^{24,25} Workload data were analyzed using Mann-Whitney U tests.

Assessment of Behavioral Alertness. For each case, the varying alertness scores constituted a time series (see fig. 2 for one subject); thus, a time-weighted average of the alertness score was computed to aggregate the score over an entire experimental session. The fraction of case time spent by a subject at each alertness score was also determined. Because alertness scores of 3 or less represented unequivocal sleepy behavior (table 1), the fraction of time spent at this level was determined. To verify this behavioral assessment technique, the original rater rescored a randomly chosen subset of videotapes (a total of 10 h of case data) on a different day and in a different random order to assess *intraobserver* variability. *Interobserver* variability was assessed by comparing ratings on the same subset of tapes from the original rater (mean of first and second ratings) with that of a second rater (S. H.). For both *intraobserver* and *interobserver* variability, a correlation coefficient using linear regression was

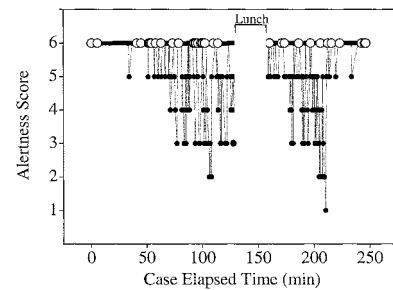


Fig. 2. Behavioral alertness score as a function of case elapsed time for one subject (sleep-deprived condition = ●; sleep-extended condition = ○). For this subject, no sleepy behaviors were noted in the sleep-extended condition, whereas there was significant cycling through sleepy behaviors in the sleep-deprived condition.

computed between the ratings for weighted average alertness score and fraction of case with alertness score of 3 or below.

Results

The subjects (six female and six male) had an average age of 31.8 ± 3.1 yr and 18 ± 11 months of clinical anesthesia experience. All subjects scored within normal limits on the Sleep Disorders Questionnaire and had no evidence of clinical sleep disorders. Subjects showed no preference for morning or evening work on the Owl and Lark Questionnaire.

Sleep Data

For the 3 nights preceding the night before the simulation session, all subjects should have slept normally, with extra sleep in the EXT condition. In fact, they had sleep times of 7.25 ± 0.9 and 9.23 ± 0.85 h for the DEP and EXT conditions, respectively ($F = 33.9$, $P = 0.0001$). During the period of sleep extension, all subjects were able to increase their total sleep time (average increase of sleep time was 127 ± 59 min; range, 23–202 min) primarily by delaying their awakening. When the night immediately before the simulation session (total sleep deprivation for the DEP condition) was included along with the other 3 nights, the subjects in the DEP condition averaged 5.44 ± 0.68 h total sleep time over 4 nights and 9.1 ± 0.68 h of sleep over 4 nights in the EXT condition ($F = 242.7$, $P = 0.0001$).

Psychomotor Test Battery

The psychomotor test battery results are shown in table 2. Additional Psychomotor Vigilance Task data are available in the Web Enhancement.

Stanford Sleepiness Scale. Subjects felt less sleepy in the EXT condition at all time periods ($F = 858$, $P = 0.0001$). Subjective sleepiness increased over the course of the call night, peaked at 08:00, and was consistently elevated throughout the simulation experiment.

Table 2. Data from Psychometric Test Battery

Performance Variable	Time of Day							
	22:00	24:00	02:00	04:00	06:00	08:00	12:00	14:00
SSS (DEP)	2.8 ± 0.6	3.8 ± 1.1	4.7 ± 1.2	5.9 ± 0.9	6.5 ± 0.7	6.5 ± 0.7	5.8 ± 0.9	6.4 ± 0.7
SSS (EXT)	-	-	-	-	-	1.7 ± 0.7	1.7 ± 0.8	2.4 ± 0.7
PRM correct (DEP)	3.5 ± 0.7	2.7 ± 1.3	2.9 ± 1.0	2.4 ± 1.2	1.9 ± 1.5	2.1 ± 1.1	2.1 ± 1.1	2.8 ± 0.9
PRM correct (EXT)	-	-	-	-	-	3.8 ± 0.4	3.1 ± 1.0	3.1 ± 0.7
PVT transform lapse (DEP)	1.76 ± 1.0	2.33 ± 1.8	2.87 ± 2.3	4.99 ± 3.7	6.59 ± 4.1	7.06 ± 3.3	4.78 ± 2.6	6.48 ± 2.8
PVT transform lapse (EXT)	-	-	-	-	-	1.53 ± 0.8	1.82 ± 1.1	1.84 ± 1.2
POMS—TMD (DEP)	0.50 ± 10	-	-	27 ± 22	-	35 ± 22	-	48 ± 29
POMS—TMD (EXT)	-	-	-	-	-	-5.1 ± 11	-	-3.2 ± 9.2

In some cells, data are missing because in the sleep-extended condition, subjects did not perform the test battery the night before their simulation study.

DEP = sleep-deprived; EXT = sleep-extended; POMS = Profile of Mood States; PRM = Probed Recall Memory; PVT, Psychomotor Vigilance Task (transform lapse described in Appendix); SSS = Stanford Sleepiness Scale; TMD = Total Mood Disturbance.

Profile of Mood States. The Profile of Mood States has six subscores (depression, fatigue, vigor, confusion, tension/anxiety, and anger) and an overall score for "Total Mood Disturbance." Over the course of the on-call period, all subscores increased monotonically and significantly except for depression ($F = 2.6$, $P = 0.1$). Total Mood Disturbance increased over the sleep deprivation period ($F = 22.9$, $P = 0.0001$), not only because fatigue and vigor scores changed, but also because tension, confusion, and anger increased. Total Mood Disturbance was significantly worse in the DEP condition than in the EXT condition during the simulation experiment ($F = 37.4$, $P = 0.0001$), an effect that was significantly greater at 14:00 than at 08:00 ($F = 5.9$, $P = 0.04$).

Probed Recall Memory. Performance on the Probed Recall Memory was significantly worse in the DEP condition than in the EXT condition ($F = 15.5$, $P = 0.02$), and the interaction with time of day was significant ($F = 5.1$, $P = 0.02$), with nearly all of the difference at 08:00. In the DEP condition, there was decreased performance and increased variability overnight, with the worst performance at 06:00.

Psychomotor Vigilance Task. All performance measures on the Psychomotor Vigilance Task were significantly worse in the DEP condition than in the EXT condition (range of F values = 25-95; $0.0001 < P < 0.0004$). An interaction effect between sleep condition and time of day during the simulation session was seen only for transform lapses ($F = 4.0$, $P = 0.04$; fig. 3). Overnight, Psychomotor Vigilance Task performance decreased and variability increased in the DEP condition, with the worst performance occurring at 08:00 (range of F values = 4.4-12; $0.0002 < P < 0.003$).

Performance Probes

For all vigilance probes combined ($n = 27$ for each experiment), there was a longer but nonstatistically significant mean response time during DEP (18.0 ± 29.7 s) versus EXT (14.4 ± 26.2 s). There was a high variability of response between individuals as well as within some individuals during some sessions (fig. 4). Response times

to the embedded probe were significantly longer in the DEP condition compared with the EXT condition ($P = 0.014$), but this was not true for the red light or arterial waveform probes ($P = 0.09$ and $P = 0.76$ respectively). There were 12 (1.9% of all probes) total lapses (*i.e.*, no probe detection before predefined time-out interval); seven of these occurred in the DEP condition, and five occurred in the EXT condition. Four of the total lapses occurred during induction of anesthesia (three DEP and one EXT), and two occurred during maintenance when DEP subjects were asleep.

Clinical Tasks

Machine Check. Mean scores for the machine check were 6.5 ± 1.3 out of a possible 11 and did not differ significantly by sleep condition ($F = 0.03$, $P = 0.87$). Some subjects in each condition left out significant portions of the machine check.

Equipment Faults. The two faults in the anesthesia environment were identified by the majority of the subjects in both conditions (faults detected = DEP 1.67 ± 0.49 ; EXT 1.83 ± 0.39 ; $F = 0.65$, $P = 0.43$).

Management of Induction. In 25% of cases (DEP 4/12, EXT 2/12; $P = 0.35$), a standard intravenous induc-

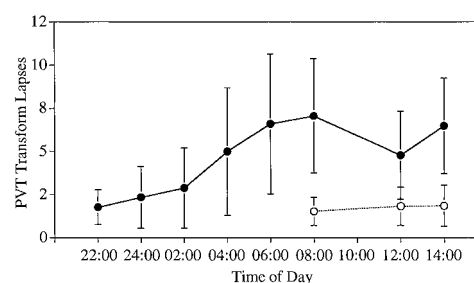


Fig. 3. Psychomotor Vigilance Task "transform lapses," $\sqrt{\text{lapses} + 1}$, as a function of condition (sleep-deprived condition = ●; sleep-extended condition = ○) and time of day. By convention, a lapse is a reaction time greater than 500 ms. Residents had significantly more lapses ($F = 25.4$; $P = 0.0004$) and greater variability of performance in the sleep-deprived condition than in the sleep-extended condition. The worst performance occurred at the 08:00 trial in the sleep-deprived condition. Error bars show SEM ($n = 12$).

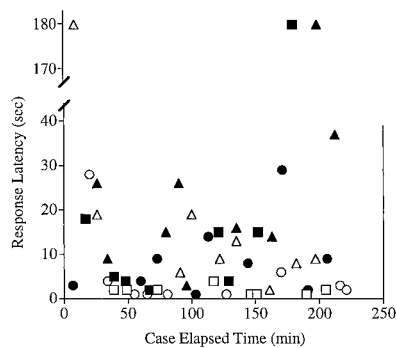


Fig. 4. Probe response time as a function of case elapsed time for one subject in both conditions (sleep-deprived condition = ●; sleep-extended condition = ○). Each symbol shape denotes one of the three types of probes. The vertical scale is broken to show extreme values. The long response latency at 15 min elapsed time occurred during anesthetic induction. Missed probes at 160 and 215 min elapsed time occurred while the subject was asleep.

tion rather than a rapid-sequence induction was performed, despite each patient's medical history that increased the risk of gastric aspiration. In 17% of the cases (DEP 3/12, EXT 1/12; $P = 0.27$), subjects administered an antibiotic drug closely related to one the patient was known to be allergic without challenging the surgeon.

Detection and Correction of Abnormal Clinical Events. Most subjects detected the clinical events and executed one of the corrective actions prior to the preset maximum time. The time to detect was greater in the DEP condition for all events except atelectasis. The time to take corrective action was similarly longer in the DEP condition for bronchospasm and atrial fibrillation, but this was not statistically significant due to the high variability observed.

Task Analysis and Psychological Workload Assessment

During both induction and maintenance, the percent time spent on individual tasks was similar between the DEP and EXT conditions. In both groups, there was appreciable intrasubject and intersubject variability in task distribution, particularly when patient care demands were low. There were no differences in average task duration or in the frequency of occurrences of each task during the entire case or during the different phases of the case. The only significant difference between the two groups was that "sleeping" was scored significantly more often in the DEP group (4.4 ± 1.9 vs. 0.2 ± 0.1 min; $P < 0.05$). In fact, 50% of DEP subjects appeared to be sleeping on at least one occasion and, in three subjects, more than 10 min of overt sleeping was recorded. The blinded observer scored the clinical workload of the anesthesia residents under the DEP condition to be 8.56 ± 0.37 (out of maximum 20; $n = 295$ samples) and under the EXT condition to be 8.62 ± 0.50 ($n = 289$; $P > 0.05$ DEP vs. EXT).

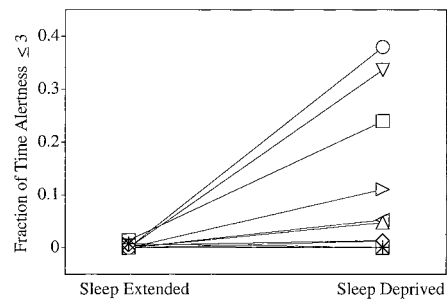


Fig. 5. Fraction of scenario time with behaviors of profound sleepiness (alertness score ≤ 3) as function of sleep condition. Each symbol represents one subject. In the sleep-deprived condition, subjects had varying amounts of sleepy behaviors—several individuals showed profound sleepiness for a substantial fraction of the scenario. In the sleep-extended condition, subjects uniformly showed virtually no such behaviors.

Behavioral Assessment of Sleepy Behaviors

The intraobserver and interobserver rating correlations were high for weighted average alertness score ($R = 0.99$ for both) and fraction of case time at a score of 3 or less ($R = 0.97$ and 0.99 , respectively), showing that the rating of alertness using this scale was reproducible. In the DEP condition, subjects' alertness scores often varied markedly over the duration of the case, as illustrated for a single representative subject in figure 2. In the EXT condition, subjects had significantly higher weighted average alertness scores (5.88 ± 0.19) than in the DEP condition (5.46 ± 0.80) ($P < 0.02$). In the EXT condition, subjects almost never exhibited a low alertness score (≤ 3) indicative of marked sleepiness (fig. 5). In the DEP condition, the amount of time spent with a low alertness score was significantly increased (*i.e.*, $8.5 \pm 14.1\%$ of case duration or 20.4 ± 33.8 min of the total 4 h case duration; $P < 0.02$ vs. EXT), and three subjects spent a substantial amount of time with low scores indicative of sleep. In fact, these three subjects exhibited profound sleepy behaviors, without awakening by head nodding or by routine noises in the environment for a notable fraction of the experiment. In the DEP condition, subjects spent an average of greater than 8 min with an alertness score of 2 or less (asleep but nodding).

Postsimulation Questionnaire

Subjects felt that the two cases were similar in difficulty (2 ± 0.9 ; where 1 = strongly agree and 5 = strongly disagree). They also agreed that the case conduct (2.1 ± 0.7), OR environment (2.3 ± 1.0), and problems encountered during the simulations (1.9 ± 0.7) were realistic.

Discussion

The three major findings of this study are:

1. Many of the clinician subjects showed sleepy behaviors when sleep-deprived, and approximately one third

fell asleep. This did not occur when they were sleep-extended. Other than the propensity to sleep, the task patterns and reported workload of subjects during simulated cases was not significantly affected by lack of sleep.

2. As a cohort, the performance of subjects on laboratory tests of psychomotor vigilance, memory, and mood showed progressive impairment during and after a night of sleep deprivation, and a significant impairment during the day after sleep deprivation *versus* when they were sleep-extended. The nadir of performance on these tests was usually around 06:00–08:00, later than the traditionally expected circadian low point of 02:00–04:00.

3. As a cohort, subjects' performance on clinically relevant tasks and probes during simulated cases showed modest, if any, impairment when sleep-deprived *versus* sleep-extended. Individuals in both sleep states made clinically relevant errors, with a trend toward more errors when sleep-deprived, but no definitive relationship to sleep deprivation could be demonstrated.

These results should not be surprising given what we know of the clinical epidemiology of patient safety. Despite the fact that sleep deprivation of interns, residents, and experienced practitioners is ubiquitous^{2,3,30,31} and residents have been shown to have substantial daytime sleepiness,¹ serious errors in patient care clearly attributable to fatigue are uncommon. In addition, a variety of factors can explain the apparent discordance between our three major findings and the relationship of these findings to clinical medicine.

Statistical Power of the Experiment was Limited

The number of subjects was limited by the logistical difficulties of conducting this study. The study had adequate power for laboratory tests of psychomotor performance, which provided a large number of repetitions and a relatively low individual variance. In calculating, *a priori*, the power for measures of clinical performance, we assumed nearly perfect performance by rested subjects on basic clinical tasks. This did not occur, as some rested subjects failed in one or more of preoperative checks, conduct of anesthesia, detecting clinical abnormalities, or responding to vigilance probes. Clinical workload (e.g., during induction) and boredom were possible factors shaping performance for both rested and sleep-deprived subjects.

This experiment should be considered a pilot study for future investigations. The results suggest that a large cohort of subjects may be required for any definitive assessment of the effects of sleep deprivation on clinical performance, given the variability in performance in the rested state and the waxing and waning level of alertness when sleep-deprived. Such studies will require substantial funding to perform.

Aggregate Results of the Cohort Do Not Capture Important Aspects of Outlier Behavior

The behavior of the cohort as an aggregate masks to some degree the behavior of individual outliers who were clearly more prone to sleep deprivation. Based on current knowledge of systems safety and accident evolution, outlier behavior may be more likely to be a factor in an accident pathway. The four subjects most prone to nodding off when sleep-deprived made some (but not all) of the errors made by all sleep-deprived subjects. There is debate in the sleep literature as to whether such outliers of sleepiness are due to a stable individual "trait" of vulnerability or are due only to random variability within individuals.³² Future studies may need to concentrate specifically on the performance of those subjects found to have decreased alertness after sleep deprivation.

Measuring Complex Clinical Performance Is More Difficult than Measuring Psychomotor Performance

Psychomotor tests provide easy and unambiguous scoring but do not represent actual work skills. Clinical tasks are more difficult to score but represent real skills. We chose to measure vigilance and the basic recognition and intervention in straightforward clinical abnormalities rather than crisis management or other complex clinical reasoning and judgment situations for two reasons. First, sleep deprivation is known to cause particular impairment in performance on vigilance tasks.^{33,34} Second, we could score recognition of and response time to vigilance probes and simple clinical events more objectively than would be possible for scoring more demanding clinical tasks involving complex clinical reasoning, leadership, teamwork, and communication. We also chose to use subtle and nonstimulating clinical situations rather than more stimulating events such as a cardiac arrest. It is possible that even when they are awake, sleep-deprived subjects could have a significant impairment in more complex clinical skills. Future studies could attempt to address this, but our ability to measure complex skills is still immature.

Sleep-deprived Individuals Cycled in and out of Reduced Alertness Frequently and Rapidly

The rapid cycling of alertness appears to be a fundamental aspect of drowsy behavior. Future studies should assess whether the periods of nodding off or apparent sleep are confirmed by simultaneous measurement of physiologic evidence of drowsiness or sleep on the electroencephalogram.

Laboratory tests of psychomotor vigilance were very sensitive to drowsy episodes, whereas the clinical vigilance probes and abnormal events were more forgiving. For example, the Psychomotor Vigilance Task requires constant, focused, and intensive attention for 10 min to stimuli that occur approximately once every 10 s. In contrast, only two clinical abnormalities occurred in

each 240-min case, and a vigilance probe occurred, on average, every 9.5 min. On several occasions, a subject who had just been asleep happened to awaken seconds before a probe was activated and was able to respond to it quickly. On two occasions, a vigilance probe was missed while the subject was clearly asleep. This outlier behavior was otherwise hidden in the aggregate statistical analysis.

A significant difference in performance on the vigilance probes was detected only for the fully embedded probe ("ramping" up or down of vital signs). The ramping behavior of this probe "telegraphed" its future threshold crossing. Subjects cycling in and out of apparent sleep either might have missed the prospective cue or might have fallen asleep before the threshold was actually attained, thereby failing to report it promptly. Future studies might use more frequent probes of clinical vigilance to better characterize the effects of brief lapses of attention.

Modest Decrease in Clinical Performance Except When Asleep

Performance when sleep-deprived probably does not degrade in a continuous fashion. When subjects in this study were awake, they maintained sufficient performance to detect most vigilance probes and to recognize and handle clinical events. When they were asleep, their performance was zero, although we captured this objectively with a performance probe only occasionally.

Unlike the Psychomotor Vigilance Task, or driving an automobile, clinical care rarely requires completely uninterrupted attention and certainly does not require reaction times of 250 ms. The complex relationship between task characteristics and the rapid shifts in level of alertness may account for some of the discordant data on sleep deprivation and clinical performance in other studies of physicians.³⁵

Compensatory Strategies were Sometimes Used to Maintain Performance

A strategy subjects in both sleep conditions sometimes used to maintain clinical performance was to focus their visual and cognitive attention primarily on the physiologic monitors and the nearby red light (see photograph in the Web Enhancement). As has been described by Weinger *et al.*³⁶ in a study of anesthesia residents performing cases in the middle of the night, this fatigue-induced focused attention strategy tends to reduce performance differences in the response to vigilance probes embedded in the central monitoring array. We did not measure attention to stimuli in other parts of the clinical environment. Future studies could assess a wider set of vigilance probes with a wider spatial distribution.

We Measured Performance in Part during a Circadian "Upswing" of Alertness

Because the simulation facility and personnel were only available during regular working hours, the simulation session was conducted in part during the morning circadian upswing of alertness. This physiologic alerting effect would tend to decrease the impairments following a night of total sleep loss, causing any differences between EXT and DEP states to be reduced. Future studies might be targeted primarily on the circadian periods of expected worst performance (such as 02:00–08:00 or 14:00–16:00).

Simulations Are Not the Same as Real Cases

Patient simulation is not the same as real patient care. Anesthesiologists might be more motivated to maintain wakefulness and performance with real patients. On the other hand, performance in the simulator in both conditions might have been better than that seen in real patient care due to the Hawthorne effect. Also, in the DEP condition, subjects only assisted the on-call teams—they were not the primary resident anesthesiologist since they had to be available to perform the psychomotor test battery every 2 h. Hence, the simulated call night might have been less fatiguing than a night of uninterrupted clinical work as the primary anesthesiologist.

Conclusions

What does it mean to have many in a cohort of sleep-deprived clinicians falling asleep and yet the cohort overall remains able to preserve vigilance and clinical performance? Theories of organizational safety predict that only a few unsafe acts actually result in injury due to multiple layers of defense in depth.³⁷ Therefore, an anesthesiologist who is asleep and perceptually unaware of the environment only rarely cause a negative patient outcome (although this has been described³⁸). Moreover, prior research demonstrates that subjects may fall asleep and yet deny that they have done so.¹ Sleeping during patient care removes a layer of protection from the system, making it vulnerable to catastrophe. Some subjects seemed particularly vulnerable to falling asleep, whereas others seemed relatively immune (fig. 5). Subjective assessment of our own sleepiness is unreliable, and there is as yet no "fitness for duty" device or blood test to tell us when we are at risk.

The cohort for our study was young and healthy residents, a group often exposed to sleep deprivation. Current rules of the Accreditation Council for Graduate Medical Education preclude anesthesia residents from performing anesthesia after a night of in-house on-call work, but there is no limit on work in other areas such as the intensive care unit. New Accreditation Council rules for all fields of medicine will go into effect in July

2003.³⁹ They will impose a 24-h limit for primary clinical duty, with a subsequent 6 h allowed for transition work and education. Although our study does not provide a “smoking gun” demonstrating clinical impairment after 24–30 h of sleep deprivation, the unequivocal impairment of alertness and laboratory psychomotor performance, combined with the trend toward more errors in the sleep-deprived condition, is consistent with the decision by the Accreditation Council for Graduate Medical Education to strengthen its rules.

However, accreditation rules for residency programs have no impact on the practices of experienced practitioners, some of whom do provide anesthesia care after being awake most of the night.⁴ There is a growing body of evidence showing greater detrimental effects of fatigue with age.^{40,41} Besides addressing limitations of our study and extending the range of clinical performance to be measured, future investigations should also include anesthesiologists from different age groups to determine whether the work practices of experienced personnel also need to be modified.

The authors thank David F. Dinges, Ph.D. (Professor of Psychology in Psychiatry, Department of Psychiatry, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania), for the use of the Probed Recall Memory test and Heidi Hwang, B.S. (VA Palo Alto Health Care System, Palo Alto, California), for technical assistance.

Appendix: Methods of Psychomotor Test Battery

Psychomotor Vigilance Task

The Psychomotor Vigilance Task is a well-validated 10-min test of simple reaction time (RT; time from observing a visual stimulus to pressing a button) that has been used extensively to evaluate sustained attention.⁴² It is known to be sensitive to sleep deprivation.^{10,34} Stimuli occur at random between 2 and 10 s after the prior response. Thus, a 10-min Psychomotor Vigilance Task run typically involves 90–100 separate RTs. A typical Psychomotor Vigilance Task RT is 250 ms, and an RT of greater than 500 ms is scored as a “lapse.” By convention, Psychomotor Vigilance Task results for each session include the following derived variables, chosen to control for disproportionate influence from long-duration lapses and to remove the proportionality between the mean and the variance for variables: median RT, mean 1/RT for the slowest 10% of RTs, mean RT for the fastest 10% of RTs, and “transformed lapses” ($\sqrt{\text{lapses}} + \sqrt{\text{lapses} + 1}$).^{33,34}

Probed Recall Memory

In the Probed Recall Memory task,⁴³ a list of four word pairs is presented for 30 s for the subject to memorize. Ten minutes later, the recall stimulus is presented, consisting of the first words of each original pair followed by a blank line. The subject has 30 s to write in the four missing words. The Probed Recall Memory score is the number of words correctly recalled.

Profile of Mood States

The Profile of Mood States is a questionnaire with six scales of mood and emotions: tension/anxiety, anger, fatigue, confusion, vigor, and depression. A summary scale, “Total Mood Disturbance,” is computed from the primary scales (low or negative values of Total Mood Distur-

bance represents positive moods; high or positive values represent negative moods).⁴⁴

Stanford Sleepiness Scale

The Stanford Sleepiness Scale is a seven-point ordinal scale of self-perceived subjective sleepiness. The Stanford Sleepiness Scale has been well validated in clinical sleep medicine and sleep research.⁴⁵

References

- Howard SK, Gaba DM, Rosekind MR, Zarcone VP: The risks and implications of excessive daytime sleepiness in resident physicians. *Acad Med* 2002; 77:1019–25
- Gravenstein JS, Cooper JB, Orkin FK: Work and rest cycles in anesthesia practice. *ANESTHESIOLOGY* 1990; 72:737–42
- Berry AJ, Hall JR: Work hours of residents in seven anesthesiology training programs. *Anesth Analg* 1993; 76:96–101
- Gaba DM, Howard SK, Jump B: Production pressure in the work environment: California anesthesiologists' attitudes and experiences. *ANESTHESIOLOGY* 1994; 81:488–500
- Asken MJ, Raham DC: Resident performance and sleep deprivation: A review. *J Med Educ* 1983; 58:382–8
- Samkoff JS, Jacques CH: A review of studies concerning effects of sleep deprivation and fatigue on residents' performance. *Acad Med* 1991; 66:687–93
- Leung L, Becker CE: Sleep deprivation and house staff performance: Update 1984–1991. *J Occup Med* 1992; 34:1153–60
- Owens JA: Sleep loss and fatigue in medical training. *Curr Opin Pulm Med* 2001; 7:411–8
- Weinger MB, Ancoli-Israel S: Sleep deprivation and clinical performance. *JAMA* 2002; 287:955–7
- Dinges DF, Pack F, Williams K, Gillen KA, Powell JW, Ott GE, Aptowicz C, Pack AI: Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4–5 hours per night. *Sleep* 1997; 20:267–77
- Dawson D, Reid K: Fatigue, alcohol and performance impairment (scientific correspondence). *Nature* 1997; 388:235
- Akerstedt T, Torsvall L, Gillberg M: Sleepiness and shift work: Field studies. *Sleep* 1982; 5(suppl 2):S95–106
- Rochrs TA, Carskadon MA, Dement WC, Roth T: Daytime sleepiness and alertness. *Principles and Practice of Sleep Medicine*, 3rd edition. Edited by Kryger MH, Roth T, Dement WC. Philadelphia, WB Saunders, 2000, pp 43–52
- Hockey GR: Changes in attention allocation in a multicomponent task under loss of sleep. *Br J Psychol* 1970; 61:473–80
- Hockey GRJ: Changes in information selection patterns in multisource monitoring as a function of induced arousal shifts. *J Exp Psychol* 1973; 101:35–42
- Krueger GP: Sustained work, fatigue, sleep loss and performance: A review of the issues. *Work Stress* 1989; 3:129–41
- Howard SK, Gaba DM, Fish KJ, Yang G, Sarnquist FH: Anesthesia crisis resource management training: Teaching anesthesiologists to handle critical incidents. *Aviat Space Environ Med* 1992; 63:763–70
- Gaba DM, Howard SK, Fish KJ, Smith BE, Sowb YA: Simulation-based training in anesthesia crisis resource management: A decade of experience. *Simulation Gaming* 2001; 32:175–93
- Douglass AB, Bornstein R, Nino-Murcia G, Keenan S, Miles L, Zarcone VP, Jr, Guilleminault C, Dement WC: The Sleep Disorders Questionnaire: I. Creation and multivariate structure of SDQ. *Sleep* 1994; 17:160–7
- Horne JA, Ostberg O: A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *Int J Chronobiol* 1976; 4:97–110
- Sadeh A, Hauri P, Kripke D, Lavie P: Role of actigraphy in the evaluation of sleep disorders. *Sleep* 1995; 18:288–302
- Stillman PL, Swanson DB, Smece S, Stillman AE, Ebert TH, Emmel VS, Caslowitz J, Greene HL, Hamolsky M, Hatem C: Assessing clinical skills of residents with standardized patients. *Ann Intern Med* 1986; 105:762–71
- McLeod PJ, Tamblin RM, Gayton D, Grad R, Snell L, Berkson L, Abrahamowicz M: Use of standardized patients to assess between-physician variations in resource utilization. *JAMA* 1997; 278:1164–8
- Weinger MB, Herndon OW, Zornow MH, Paulus MP, Gaba DM, Dallen LT: An objective methodology for task analysis and workload assessment in anesthesia providers. *ANESTHESIOLOGY* 1994; 80:77–92
- Weinger MB, Herndon OW, Gaba DM: The effect of electronic record keeping and transesophageal echocardiography on task distribution, workload, and vigilance during cardiac anesthesia. *ANESTHESIOLOGY* 1997; 87:144–55
- Loeb RG: A measure of intraoperative attention to monitor displays. *Anesth Analg* 1993; 76:337–41
- Loeb RG: Monitor surveillance and vigilance of anesthesia residents. *ANESTHESIOLOGY* 1994; 80:527–33

28. Gaba DM, Lee T: Measuring the workload of the anesthesiologist. *Anesth Analg* 1990; 71:354-61
29. Slagle J, Weinger MB, Dinh MT, Brumer VV, Williams K: Assessment of the intrarater and interrater reliability of an established clinical task analysis methodology. *ANESTHESIOLOGY* 2002; 96:1129-39
30. Dercq JP, Smets D, Somer A, Desantoine D: A survey of Belgian anesthesiologists. *Acta Anaesthesiol Belg* 1998; 49:193-204
31. Gander PH, Merry A, Millar MM, Weller J: Hours of work and fatigue-related error: A survey of New Zealand anaesthetists. *Anaesth Intensive Care* 2000; 28:178-83
32. Doran SM, Van Dongen HP, Dinges DF: Sustained attention performance during sleep deprivation: Evidence of state instability. *Arch Ital Biol* 2001; 139:253-67
33. Dinges DF, Kribbs NB: Performing while sleepy: Effects of experimentally-induced sleepiness, Sleep, Sleepiness, and Performance. Edited by Monk TH. Chichester, Wiley, 1991, pp 97-128
34. Dinges DF: Probing the limits of functional capability: The effects of sleep loss on short-duration tasks, Sleep, Arousal, and Performance. Edited by Broughton R, Ogilvie R. Boston, Birkhauser, 1992, pp 177-88
35. Veasey S, Rosen R, Barzansky B, Rosen I, Owens J: Sleep loss and fatigue in residency training: A reappraisal. *JAMA* 2002; 288:1116-24
36. Weinger MB, Vora S, Herndon CN, Howard SK, Smith BE, Mazzei WJ, Rosekind MR, Gaba DM: Evaluation of the effects of fatigue and sleepiness on clinical performance in on-call anesthesia residents during actual night time cases and in simulated cases, Proceedings of Enhancing Patient Safety and Reducing Errors in Health Care. Rancho Mirage, California, National Patient Safety Foundation, 1999, pp 306-10
37. Reason J: *Human Error*. Cambridge, United Kingdom, Cambridge University Press, 1990
38. Pankratz H: Witness: Doctor dozed. *Denver Post* September 15, 1995, pp 1A
39. Gaba DM, Howard SK: Fatigue among clinicians and the safety of patients. *N Engl J Med* 2002; 347:1249-55
40. Reilly T, Waterhouse J, Atkinson G: Aging, rhythms of physical performance, and adjustment to changes in the sleep-activity cycle. *Occup Environ Med* 1997; 54:812-6
41. Katz JD: Issues of concern for the aging anesthesiologist. *Anesth Analg* 2001; 92:1487-92
42. Dinges D, Powell J: Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Beh Res Meth Instr Comp* 1985; 17:652-5
43. Dinges D, Kribbs N, Bates B, Carlin M: A very brief probed recall memory task: Sensitivity to sleep loss. *Sleep Res* 1993; 22:330
44. McNair D, Lorr M, Droppleman L: *Profile of Mood States (manual)*. San Diego, Educational and Industrial Testing Service, 1971
45. Hoddes E, Zarcone V, Smythe H, Phillips R, Dement WC: Quantification of sleepiness: A new approach. *Psychophysiology* 1973; 10:431-6