Neurobehavioral Performance of Residents After Heavy Night Call vs After Alcohol Ingestion

J. Todd Arnedt, PhD
Judith Owens, MD, MPH
Megan Crouch, BA
Jessica Stahl, BA
Mary A. Carskadon, PhD

W ORK-RELATED SLEEP LOSS and fatigue in medical training has become a source of increasing concern.[1-4] Although some studies demonstrate post-call performance deficits,[5-8] other studies do not.[9-12] These mixed findings have been attributed to methodological limitations,[13,14] such as low power, absence of objective sleep measurement, outcome measures insensitive to sleepiness or lacking ecological validity (having little relevance to real-world demands), absence of control for circadian factors or stimulant use (caffeine), and questionable rested control groups (>4 hours of sleep).[10] The serious consequences of resident sleep loss have been demonstrated in an intervention study that found that interns obtained 5.8 hours less sleep, had 50% more attentional failures, and committed 22% more serious errors on critical care units while working a traditional schedule compared with a schedule of reduced hours.[15,16]

On self-report measures, residents express concern about occupational and interpersonal difficulties stemming from sleep loss.[1,2,17-19] Of particular concern,

Context Concern exists about the effect of extended resident work hours; however, no study has evaluated training-related performance impairments against an accepted standard of functional impairment.

Objectives To compare post-call performance during a heavy call rotation (every fourth or fifth night) to performance with a blood alcohol concentration of 0.04 to 0.05 g% (per 100 mL of blood) during a light call rotation, and to evaluate the association between self-assessed and actual performance.

Design, Setting, and Participants A prospective 2-session within-subject study of 34 pediatric residents (18 women and 16 men; mean age, 28.7 years) in an academic medical center conducted between October 2001 and August 2003, who were tested under 4 conditions: light call, light call with alcohol, heavy call, and heavy call with placebo.

Interventions Residents attended a test session during the final week of a light call rotation (non–post-call) and during the final week of a heavy call rotation (post-call). At each session, they underwent a 60-minute test battery (light and heavy call conditions), ingested either alcohol (light call with alcohol condition) or placebo (heavy call with placebo condition), and repeated the test battery. Performance self-evaluations followed each test.

Main Outcome Measures Sustained attention, vigilance, and simulated driving performance measures; and self-report sleepiness, performance, and effort measures.

Results Participants achieved the target blood alcohol concentration. Compared with light call, heavy call reaction times were 7% slower (242.5 vs 225.9 milliseconds, $P<.001$); commission errors were 40% higher (38.2% vs 27.2%, $P<.001$); and lane variability (7.0 vs 5.5 ft, $P<.001$) and speed variability (4.1 vs 2.4 mph, $P<.001$) on the driving simulator were 27% and 71% greater, respectively. Speed variability was 29% greater in heavy call with placebo than light call with alcohol (4.2 vs 3.2 mph, $P=.01$), and reaction time, lapses, omission errors, and off-roads were not different. Correlation between self-assessed and actual performance under heavy call was significant for commission errors ($r=-0.45$, $P=.01$), lane variability ($r=-0.76$, $P<.001$), and speed variability ($r=-0.71$, $P<.001$), but not for reaction time.

Conclusions Post-call performance impairment during a heavy call rotation is comparable with impairment associated with a 0.04 to 0.05 g% blood alcohol concentration during a light call rotation, as measured by sustained attention, vigilance, and simulated driving tasks. Residents' ability to judge this impairment may be limited and task-specific.

JAMA. 2005;294:1025-1033   www.jama.com

Author Affiliations are listed at the end of this article.
Corresponding Author: J. Todd Arnedt, PhD, Sleep and Chronophysiology Laboratory, Department of Psychiatry, University of Michigan, 2101 Commonwealth Blvd, Suite D, Ann Arbor, MI 48105 (tarnedt @med.umich.edu).

self-reported lifetime rates of motor vehicle near-misses and crashes among residents[20] are 2.5 and 3 times those of nonresident drivers, respectively.[21] One survey found that compared with faculty members over the previous 3 years, more pediatric house officers reported falling asleep behind the wheel (49% vs 13%) and having motor vehicle crashes (20% vs 11%) during residency.[22] A prospective study found that for every extended work shift, the monthly risk of a motor vehicle crash increased by 9.1%.[23] Despite the heightened risk, only 2 small studies have examined resident driving impairment experimentally. Both found impairment of simulated driving post-call, but neither provided objective on-call sleep duration[24,25] and 1 study used a driving task lacking face validity.[24]

One approach to measuring the magnitude of performance deficits associated with sleep loss is to compare performance following sleep loss and alcohol consumption. Dawson and Reid[26] found that impairment on a tracking task after 17 hours of wakefulness was equivalent to a blood alcohol concentration (BAC) of 0.05 g% (per 100 mL of blood) in a sample of nonresidents. These findings have been replicated with other tasks, including simulated driving, in samples of nonresident university students and truck drivers.[27-29] Alcohol serves as a useful index for comparison because it impairs performance, even at lower BACs,[30-34] and legal limits of intoxication have been established. At 0.05 g% BAC (3-4 standard drinks), alcohol increases self-confidence; decreases inhibitions; diminishes attention, judgment, and control[35]; and leads to hazardous driving.[36]

Our primary goal was to compare post-call performance during a heavy call rotation to non–post-call performance during a light call rotation with a BAC of 0.04 to 0.05 g%, using tests of sustained attention, vigilance, and simulated driving. To maximize rested and sleepy states, residents were tested during the final week of light and heavy call rotations. A prospective within-subject 2-session design was used to enhance feasibility and generalizability, reduce attrition, and experimentally control for time of testing, alcohol expectancy, beverage consumption, and test fatigue for our comparison of primary interest. A second goal was to evaluate the association between self-assessed and actual performance call-related sleep loss and alcohol ingestion.

METHODS
Participants

Participants were recruited from the Brown University Pediatrics residency program, Providence, RI. Interested residents were screened for the following exclusion criteria: aged older than 40 years; sleep disorder diagnosis; chronic medical condition; current psychiatric illness; current use of medication known to affect the sleep/wake cycle or daytime alertness, or that is a contraindication for alcohol ingestion; no or minimal prior alcohol exposure, defined as responding "never drink alcohol" or "never" to the question "How often do you have 2-4 drinks in one occasion at least once in a while"[37]; current or prior treatment for alcohol or substance abuse; and positive urine pregnancy screen.

Of the 115 residents potentially eligible for participation between October 2001 and August 2003, 43 (37%) responded to e-mail solicitations and were screened for eligibility. Six were excluded (2 were older than 40 years, 1 had a sleep disorder, 2 were taking medication, and 1 never consumed alcohol) and 2 declined. One enrolled participant completed only 1 test session. The final sample included 16 men and 18 women (mean [SD] age, 28.7 [2.7] years); 14 were interns, 15 were second-year residents, and 5 were third-year residents. Participants and nonparticipants or dropouts (n=81) did not differ by age (*P*=.74), sex (*P*=.21), or residency type (pediatric, medicine/pediatrics, triple board; *P*=.94). Participants were told that the purpose of the study was to compare performance following call-related sleep loss and following alcohol, but they were not told the primary study hypothesis. Participants provided written informed consent and received US $200 upon successful completion of each test session. The study was approved by the institutional review boards at Rhode Island Hospital and Brown University, Providence, RI, and the University of Michigan, Ann Arbor.

Study Design

Our study was a prospective 2-session within-subject study with 4 conditions: light call, light call with alcohol, heavy call, and heavy call with placebo. Light call and light call with alcohol occurred consecutively during a single session and followed a non-call night during the final week of a 4-week light call rotation. Heavy call and heavy call with placebo were completed post-call after 4 weeks of a heavy call rotation. We did not use a no-call condition as a control to maximize the generalizability of our findings by including actual rotations that are present in the training program. We did not include a light call with placebo condition because our primary comparison of interest was between the heavy call with placebo and light call with alcohol conditions, and a third testing session would likely have had a high attrition rate.

Light and Heavy Call Rotations

Light call rotations (behavioral, elective, or selective) were 4-week daytime clinic rotations averaging 44 hours per week, along with sick-call, which requires night call only if the on-call resident becomes ill. Heavy call rotations (neonatal intensive care, pediatric intensive care, or wards) averaged 90 hours per week (80 hours per week after July 2003) and mandated call every fourth or fifth night (34-36 consecutive hours per overnight call). Residents were allowed to work outside the authorized training program (moonlight) only with written permission from the residency director.

Study Protocol

One week before the first session, participants attended a 90-minute orien-

tation during which all participants practiced the tests and received a daily sleep/activity diary and wrist activity monitor (Actiwatch 64, Mini Mitter Company Inc, Bend, Ore). For 7 days before each test session, participants maintained the diary, self-selected their sleep schedules, and had activity levels continuously monitored. Sleep parameters were estimated using sleep analysis software. We adapted scoring procedures from Acebo et al.[38] Nocturnal sleep (from 2100 to 0900 hours) was analyzed separately from daytime sleep (from 0901 to 2059).

On the test day, participants were not allowed to nap, ingest caffeine after noon, or to ingest food or drink other than water within 4 hours of testing, verified by self-report before each test session.[39] Twenty-two residents were tested first during heavy call; the mean (SD) intersession interval was 106 (82) days.

Participants arrived at 1500 hours for the light call and heavy call sessions. After completing self-report sleepiness measures, participants performed a 60-minute battery that included the Psychomotor Vigilance Test (Ambulatory Monitoring Inc, Ardsley, NY),[40] the Continuous Performance Test,[41] and a simulated driving task (DriveSim, York Computer Technologies, Kingston, Ontario). The sleepiness measures were repeated every 30 minutes. After each test, participants completed self-assessments of performance and effort.

Following the light call condition testing, participants consumed alcohol (light call with alcohol condition). The alcohol dose was 0.6 g/kg for men and 0.55 g/kg for women, to produce equivalent peak BACs of 0.05 g%.[42] The alcoholic beverage consisted of a commercial brand of chilled 80-proof vodka mixed with tonic water in a 1:5 ratio and one-fourth lime. Following the heavy call condition testing, participants consumed placebo (heavy call with placebo condition). The placebo was an equal volume of chilled tonic water and one-fourth lime. The total volume was distributed among three

12-oz cups and consumed at an equal rate over 30 minutes. To enhance the appearance that participants were receiving alcohol in both conditions, the drinks were mixed in plain view, alcohol and tonic were decanted from vodka bottles, and beverages were served with fresh lime. Participants repeated the battery 20 minutes postingestion. Breath samples were analyzed before and after tests using a handheld breathalyzer (AlcoSensor IV, Intoximeters Inc, St Louis, Mo).

After each session, participants rated their certainty of having received alcohol on a 0 (certain did not receive) to 100 (certain did receive) scale. For the light call with alcohol sessions, 63.6% of the participants were completely certain they had received alcohol; for the heavy call with placebo sessions, 67.6% of the participants were completely certain that they had received placebo. Four participants in the heavy call with placebo group were at least 25% certain of receiving alcohol.

Participants in the heavy call conditions were excused and driven home by a significant other or by taxi. Participants in the light call conditions either remained in the laboratory under supervision until their BAC dropped below 0.02 g% or they signed a release and were driven home by a significant other or by taxi.

Outcome Measures

Stanford Sleepiness Scale. The Stanford Sleepiness Scale[43] is a 7-item scale that requires participants to rate their current sleepiness from 1 (feeling active and vital, wide awake) to 7 (almost in reverie, sleep onset soon, lost struggle to remain awake).

Visual Analog Scale. The visual analog scale comprised the questions, "How alert do you feel?", "How sleepy do you feel?", and "Overall, how do you feel?" Participants marked a 100-mm line, with anchors "very little/very bad" to "very much/very good." The distance from the left edge of the scale to the participant mark was the score (range, 0-100). Higher scores indicated greater levels of alertness, sleepi-

ness, and overall functioning. Scores on the alertness and overall functioning scales have been inverted for consistency.

Psychomotor Vigilance Task. The Psychomotor Vigilance Task[40] is a 10-minute visual sustained-attention test that is sensitive to sleepiness[44,45] and alcohol.[34] Participants pressed a button on the handheld unit in response to numbers scrolling on the liquid crystal display screen with a 2- to 10-second interstimulus interval. Dependent variables were median reaction time and frequency of lapse (reaction time >500 milliseconds). Higher scores indicated worse performance.

Continuous Performance Test. The Continuous Performance Test[41] is a 14-minute computer vigilance task, previously used with residents,[46,47] which requires participants to respond to any alphabetic letter except "x/X." Stimuli are displayed for 250 milliseconds with interstimulus intervals of 1, 2, and 4 seconds. Dependent variables included errors of commission (%) and omission (%). Higher scores indicated worse performance.

Simulated Driving Task. The simulated driving task is a 30-minute task, sensitive to sleepiness and alcohol,[48,49] which runs on a computer with software (DriveSim 3.00, York Computer Technologies), peripheral steering wheel, accelerator, and brake. The task presents a driver's orientation of a 2-lane highway with lane markings, speed signs, and small trees along the roadside. Other vehicles appear periodically but participants ignore them. Instructions are to stay in the center of the right lane and follow a fixed speed limit (60 miles per hour) while driving on the straight road. "Wind" periodically and randomly pushes the simulated vehicle right, left, or not at all. Traveling off the road elicits a beep and the car is automatically placed back on the road. Dependent variables were lane variability (SD of the vehicle center from the center of the right lane measured in feet), speed variability (SD of the difference in the vehicle speed from the posted speed measured in miles per

hour), and "off-roads" (number of times the vehicle left the road). Higher scores indicated worse performance.

Posttest Self-assessments. Dependent variables were performance and effort ratings. Performance was assessed on a 7-point Likert scale response to the statement "I feel my performance during this test was . . . ", anchored with 1=extremely good, 4=fair, and 7=extremely poor. Effort was assessed on a 4-point Likert scale response to the statement "The effort I had to expend to achieve this level of performance was . . . ", anchored with 1=very little effort and 4=an extreme effort.

Data Analyses

Variables that deviated significantly from normality were transformed for parametric analyses or dichotomized (lapses, omissions, off-roads) and analyzed using nonparametric McNemar tests. Data are reported as mean (SE) unless otherwise indicated, with significance level set at $P=.05$.

Continuous performance and subjective measures were analyzed with training year (interns vs second-year residents and third-year residents) by condition (light call, light call with alcohol, heavy call, or heavy call with placebo) mixed repeated measures analysis of variance. Main effects were followed by pairwise comparisons between light call and each of the other 3 experimental conditions (light call with alcohol, heavy call, and heavy call with placebo) and between light call with alcohol and heavy call with placebo. Because of our nonrandomized design, we secondarily examined order effects by separately analyzing performance in participants whose first session occurred during light call (n=12). Test-dependent variables for light call with alcohol, heavy call, and heavy call with placebo were compared with posttest self-assessments using Spearman rank order correlations. Analyses were conducted using SPSS version 12.0 for Windows (SPSS Inc, Chicago, Ill).

RESULTS

Call Schedules

Residents were on-call more frequently during the heavy than light call rotation (mean [SE], 7.3 [0.3] vs 1.4 [0.3] nights; $P<.001$). One resident reported moonlighting 6 days before light call testing; no moonlighting occurred during heavy call.

Actigraphy

Actigraphy results for the 7 days and the 24 hours before testing are shown in **TABLE 1**. For the 7 days preceding each test session, the mean nocturnal sleep period (the elapsed time between sleep onset and sleep offset as scored by the actigraphy software) was 7 hours 32 minutes during light call compared with 6 hours 17 minutes during heavy call ($P<.001$). There were similar results comparing light call and heavy call with respect to total sleep time and the cumulative sleep duration for the week.

For the 24 hours preceding each test session, there was significantly more sleep during light call than heavy call as measured by nocturnal sleep period (7 hours 24 minutes vs 3 hours 56 minutes, respectively; $P<.001$), total sleep time (6 hours 37 minutes vs 3 hours 2 minutes, respectively; $P<.001$), and the cumulative sleep duration (6 hours 48 minutes vs 3 hours 8 minutes, respectively; $P<.001$). Nocturnal sleep was also more efficient during light call than heavy call (89.4% vs 82.3%, $P=.02$), but there were no rotation differences in diurnal sleep.

Blood Alcohol Concentrations

Blood alcohol concentrations were 0.0 g% before all sessions and after light call, heavy call, and heavy call with placebo. Mean (SE) peak BACs in the light call with alcohol assessment were 0.046 g% (0.002 g%) before the Psychomotor Vigilance Test and 0.041 g%

Table 1. Actigraphically Determined Sleep Parameters for the 7 Days Before and the 24 Hours Before the Light Call and Heavy Call Test Sessions*

Actigraph Parameter	7 Days Before Testing			24 Hours Before Testing		
	Light Call (n = 30)	Heavy Call (n = 31)	P Value†	Light Call (n = 30)	Heavy Call (n = 32)	P Value†
Daily nocturnal sleep‡						
Sleep-onset time	23:52 (0:08)	00:23 (0:09)	.01	23:57 (0:12)	02:33 (0:21)	<.001
Sleep-offset time	07:24 (0:08)	06:40 (0:05)	<.001	7:21 (0:12)	06:29 (0:16)	.07
Sleep period	7:32 (0:09)	6:17 (0:10)	<.001	7:24 (0:17)	3:56 (0:19)	<.001
Total sleep time	6:38 (0:10)	5:20 (0:09)	<.001	6:37 (0:16)	3:02 (0:16)§	<.001
Sleep efficiency, %	88.2 (0.8)	85.2 (1.0)	.24	89.4 (1.0)	82.3 (2.8)	.02
Daily diurnal sleep‖						
No. of diurnal sleep episodes	0.6 (0.1)	0.8 (0.1)	.21	0.4 (0.2)	0.2 (0.1)	.77
Duration of diurnal sleep	0:13 (0:03)	0:22 (0:03)	.08	0:11 (0:05)	0:06 (0:03)	.54
Total cumulative sleep before testing	48:03 (1:23)	39:55 (1:00)	<.001	6:48 (0:17)	3:08 (0:16)§	<.001

*Results are hours:minutes mean (SE) unless otherwise specified.
†Light call vs heavy call comparisons based on paired *t* tests.
‡Nocturnal sleep defined as the major sleep bout between 2100 and 0900 hours. Sleep period is the elapsed time in hours:minutes from sleep-onset time to sleep-offset time as scored by Actiware-Sleep algorithm; it includes both sleep and wake episodes during this period. Total sleep time includes only sleep as scored by the software, within the sleep period. Sleep efficiency is the percentage of sleep period spent asleep as scored by Actiware-Sleep algorithm [(total sleep time/sleep period) × 100].
§These variables were computed on 33 residents because 1 resident had actigraphically defined 0 hour of sleep the night before testing.
‖Defined as episodes of 15 minutes to 180 minutes between 0901 and 2059 hours.

(0.002 g%) after the Psychomotor Vigilance Test; 0.041 g% (0.002 g%) before the Continuous Performance Test and 0.040 g% (0.001 g%) after the Continuous Performance Test; and 0.040 g% (0.001 g%) before the simulated driving task and 0.037 g% (0.002 g%) after the simulated driving task.

Self-reported Sleepiness

The Stanford Sleepiness Scale and visual analog scale ratings are summarized in TABLE 2. There were no main effects or interactions involving training year. The Stanford Sleepiness Scale ratings were higher in heavy call with placebo (mean, 4.5), light call with alcohol (mean, 3.3), and heavy call (mean, 4.6) vs light call (mean, 2.3; $P<.001$ for all comparisons). On the vi-

sual analog scale, the main effect of condition was significant for all subscales: alertness (light call: mean, 31.6; light call with alcohol: mean, 45.3; heavy call: mean, 68.4; heavy call with placebo: mean, 59.5; $P<.001$); sleepiness (light call: mean, 33.0; light call with alcohol: mean, 38.2; heavy call: mean, 77.9; heavy call with placebo: mean, 74.7; $P<.001$); and overall (light call: mean, 25.7; light call with alcohol: mean, 26.5; heavy call: mean, 53.5; heavy call with placebo: mean, 51.3; $P<.001$). Post hoc comparisons indicated that alertness, sleepiness, and overall ratings were higher (worse) in heavy call with placebo compared with light call with alcohol ($P<.001$) and in both heavy call and heavy call with placebo relative to light call ($P<.001$).

Actual and Self-assessed Performance

Results for actual and self-assessed performance are shown by condition in TABLE 3 and TABLE 4, respectively.

Psychomotor Vigilance Task. There were no main effects or interactions involving training year. Median reaction time for light call was 225.9 milliseconds, with reaction times 7% to 10% slower in light call with alcohol (248.4 milliseconds, $P<.001$), heavy call (242.5 milliseconds, $P=.001$), and heavy call with placebo (242.3 milliseconds, $P<.001$), and there was no difference between light call with alcohol and heavy call with placebo ($P=.19$). Lapses occurred more often in heavy call with placebo than with light call, but were not significantly differ-

Table 2. Stanford Sleepiness Scale and Visual Analog Scale Ratings in the 4 Experimental Conditions*

	Mean (SE)				P Value	
Self-report Measure	Light Call	Light Call With Alcohol	Heavy Call	Heavy Call With Placebo	Condition Effect†	Light Call With Alcohol vs Heavy Call With Placebo‡
Stanford Sleepiness Scale (n = 23)	2.3 (0.2)	3.3 (0.2)	4.6 (0.2)	4.5 (0.2)	<.001	<.001
Visual analog scale (n = 30) Alertness	31.6 (3.0)	45.3 (3.1)	68.4 (2.6)	59.5 (2.8)	<.001	<.001
Sleepiness	33.0 (3.5)	38.2 (4.1)	77.9 (2.1)	74.7 (1.8)	<.001	<.001
Overall	25.7 (2.7)	26.5 (2.6)	53.5 (3.1)	51.3 (2.7)	<.001	<.001

*Stanford Sleepiness Scale: 1-7 scale in which 1 = feeling active and vital, alert, or wide awake; 7 = almost in reverie, sleep onset soon, or lost struggle to remain awake. Visual analog scale: 0-100 scale in which higher scores indicate greater sleepiness, less alertness, and worse overall functioning.
†Based on training year (interns vs second-year residents and third-year residents) × condition (light call, light call with alcohol, heavy call, heavy call with placebo) mixed repeated measures analysis of variance. There were no significant training year × condition interactions.
‡Based on post hoc paired *t* tests.

Table 3. Performance Measures in the 4 Experimental Conditions

					P Value	
Performance Task	Light Call	Light Call With Alcohol	Heavy Call	Heavy Call With Placebo	Condition Effect*	Light Call With Alcohol vs Heavy Call With Placebo
Psychomotor Vigilance Task (n = 27) Median reaction time, mean (SE), ms	225.9 (4.0)	248.4 (7.0)	242.5 (6.1)	242.3 (5.0)	<.001	.19†
No. of lapses, median (range)	0 (0-3)	0 (0-13)	1 (0-9)	1 (0-7)	NA	.21‡
Continuous Performance Test (n = 33), % Commission errors, mean (SE)	27.2 (2.8)	46.5 (3.6)	38.2 (3.3)	40.6 (3.2)	<.001	.02†
Omission errors, median (range)	0.0 (0.0-26.8)	0.3 (0.0-3.3)	0.3 (0.0-59.3)	0.7 (0.0-33.2)	NA	.18‡
Simulated driving task (n = 34) Lane variability, mean (SE), ft	5.5 (0.2)	6.2 (0.2)	7.0 (0.4)	6.8 (0.3)	<.001	.06†
Speed variability, mean (SE), mph§	2.4 (0.3)	3.2 (0.5)	4.1 (0.5)	4.2 (0.4)	<.001	.01†
No. of off-road incidents, median (range)	0 (0-3)	1 (0-6)	1 (0-23)	1 (0-16)	NA	.55‡

Abbreviaion: NA, not applicable.
*Based on training year (interns vs second-year residents and third-year residents) × condition (light call, light call with alcohol, heavy call, heavy call with placebo) mixed repeated measures analysis of variance. There were no significant training year × condition interactions.
†Based on post hoc paired *t* tests.
‡Based on paired McNemar tests.
§Analysis of variance conducted on transformed speed variability using the transformation log (x).

Table 4. Posttest Self-assessed Performance and Effort Ratings on the Performance Measures in the 4 Experimental Conditions*

Performance Task	Light Call	Light Call With Alcohol	Heavy Call	Heavy Call With Placebo	*P* Value†
Psychomotor Vigilance Task (n = 26)					
Performance	3.0 (0.1)	3.6 (0.2)	4.1 (0.2)	4.1 (0.2)	.73
Effort	1.8 (0.2)	2.4 (0.1)	2.5 (0.1)	2.5 (0.1)	>.99
Continuous Performance Test (n = 31)					
Performance	3.5 (0.2)	4.1 (0.2)	4.4 (0.3)	4.6 (0.2)	.21
Effort	2.2 (0.1)	2.2 (0.1)	2.8 (0.1)	2.9 (0.2)	.01
Simulated driving task (n = 34)					
Performance	2.8 (0.2)	3.4 (0.2)	4.7 (0.3)	4.3 (0.2)	.007
Effort	1.8 (0.1)	2.1 (0.1)	2.8 (0.2)	2.8 (0.1)	<.001

*Performance anchors: 1 = extremely good, 2 = very good, 3 = good, 4 = fair, 5 = poor, 6 = very poor, 7 = extremely poor. Effort anchors: 1 = very little, 2 = a moderate amount; 3 = quite a lot, 4 = an extreme amount.
†Light call with alcohol vs heavy call with placebo based on pairwise McNemar tests.

ent from light call with alcohol. There were no main effects or interactions involving condition in participants whose first session occurred during light call (n=9).

More participants rated their performance as poor, very poor, or extremely poor in heavy call (26.8%, *P*=.03) and heavy call with placebo (34.5%, *P*=.008) than with light call (3.8%). Self-assessments of poor or worse performance did not differ between heavy call with placebo and light call with alcohol (26.9%) or effort ratings of quite a lot or extreme (42.3% for light call with alcohol and 46.1% for heavy call with placebo).

Posttest performance ratings were associated with reaction time for light call with alcohol ($r=-0.65, P<.001$) but not for heavy call ($r=-0.18, P=.36$) or heavy call with placebo ($r=-0.01, P=.95$).

Continuous Performance Test. There were no performance differences by training year with this test. Compared with commission errors in light call (27.2%), there were 40% to 70% more commission errors in light call with alcohol (46.5%, *P*<.001), heavy call (38.2%, *P*<.001), and heavy call with placebo (40.6%, *P*<.001), and 15% more in light call with alcohol than heavy call with placebo (*P*=.02). Omission errors occurred more often with heavy call (median, 0.3%; *P*=.01) and heavy call with placebo (median, 0.7%; *P*=.01) than with light call (median, 0%), and heavy call with placebo did not differ

from light call with alcohol (median, 0.3%; *P*=.18). For those residents who completed light call first (n=11), commission errors did not differ between light call with alcohol (48.8%) and heavy call with placebo (44.5%, *P*=.37), but in both conditions they were worse than light call (27.5% [SE, 4.3%]; *P*<.001 vs light call with alcohol, *P*=.001 vs heavy call with placebo).

Self-ratings of poor, very poor, or extremely poor performance were more common in heavy call (51.6%) than light call (12.9%, *P*=.002). There was no difference in the frequency of these performance ratings between heavy call with placebo (54.9%) and light call with alcohol (35.5%), but quite a lot and extreme effort ratings were more frequent in heavy call with placebo (64.5%) than light call with alcohol (32.3%, *P*=.01).

Performance ratings were associated with commission errors for light call with alcohol ($r=-0.61, P<.001$) and heavy call ($r=-0.45, P=.01$), but not for heavy call with placebo ($r=-0.25, P=.17$).

Simulated Driving Task. Performance was not significantly different by training year with the simulated driving task. Relative to light call lane variability (5.5 ft), lane variability was 13% to 27% higher in light call with alcohol (6.2 ft, *P*=.002), heavy call (7.0 ft, *P*<.001), and heavy call with placebo (6.8 ft, *P*<.001); light call with alcohol and heavy call with placebo did not significantly differ (*P*=.06). Speed vari-

ability was 29% greater in heavy call with placebo than light call with alcohol (4.2 vs 3.2 mph, *P*=.01) and was 34% to 75% higher in light call with alcohol (*P*=.01), heavy call with placebo (*P*<.001), and heavy call (4.1 mph, *P*<.001) compared with light call (2.4 mph). Off-roads occurred more frequently in heavy call (median, 1; *P*=.02) and heavy call with placebo (median, 1; *P*=.049) than light call (median, 0), and were not different between heavy call with placebo and light call with alcohol (median, 1). For those participants whose first session followed light call, heavy call with placebo performance was worse than light call with alcohol for lane variability and speed variability.

More than half of heavy call residents (58.8%) rated their simulated driving performance as poor or worse compared with only 5.9% of participants in the light call group (*P*<.001). These ratings were also more common in heavy call with placebo (44.1%) than light call with alcohol (11.7%, *P*=.007). Almost three quarters of participants in the heavy call with placebo group (73.6%) rated their effort as quite a lot or extreme compared with 17.6% for light call with alcohol (*P*<.001).

Performance ratings were associated with lane variability in heavy call ($r=-0.76, P<.001$) and heavy call with placebo ($r=-0.50, P=.003$), but not in light call with alcohol ($r=-0.32, P=.06$). Speed variability was associated with self-ratings in heavy call ($r=-0.71, P<.001$) and heavy call with placebo ($r=-0.51, P=.002$), but not in light call with alcohol ($r=-0.04, P=.85$).

COMMENT

Our primary findings were post-call performance decrements in attention, vigilance, and simulated driving following 4 weeks of heavy call compared with a light call rotation, similar to impairments associated with 0.04 to 0.05 g% BAC. Compared with light call, heavy call performance was characterized by slower and more variable reaction times and more commission er-

rors on validated tests of sustained attention and vigilance. Heavy call residents were also less able to maintain a consistent lane position and speed, and ran off the road more often on a simulated driving task. Compared with alcohol ingestion, heavy call simulated driving speed variability was 30% higher, and reaction time, attention lapses, omission errors, and crashes were similar. These results were independent of training year and occurred despite self-ratings of greater effort in the heavy call with placebo group on 2 of the 3 tasks.

This is the first study to our knowledge to directly compare impairment related to heavy night call with that related to alcohol ingestion, an accepted standard of functional impairment. We selected performance tasks with known sensitivity to sleep loss and alcohol. Previous studies have demonstrated that both conditions individually increase reaction time, errors of omission, and errors of commission on neurobehavioral assays,[34,50,51] and that they induce simulated driving impairments, characterized by increased variability in driving performance and a greater tendency to drive off the road.[48,49,52] Together, sleep loss and alcohol produce at least additive impairments in driving performance.[48,53] The Continuous Performance Test and Psychomotor Vigilance Test findings from our study suggest that, consistent with nonresident[45,54,55] and other resident[16,56,57] studies, sustained attention and vigilance are particularly sensitive to training-related sleep loss. The observed postcall deficits likely result not only from acute sleep loss but also from the superimposed chronic partial sleep deprivation experienced during training.[13]

These laboratory tasks have not been validated against actual medical tasks. However, post-call deterioration has been found in simulated (laparoscopy)[58,59] and actual (perioperative complications)[60] medical procedures that require the skills inherent in these assessments. The driving simulator findings are particularly provocative. In the heavy call with placebo group,

tracking and speed variability were, respectively, around 10% and 30% greater than the light call with alcohol group. These results must be interpreted cautiously because few controlled studies have compared simulated with actual driving,[61] and the strength of the relationship is likely simulator-specific. However, taken together with resident-reported increased motor vehicle crash rates,[23] it seems likely that resident driving skills are impaired post-call and contribute to increased injury risk.

Having demonstrated performance deficits, it is equally important to know whether residents recognize these deficits. We found significant associations between actual performance and self-assessed performance for the Continuous Performance Test and the simulated driving task but not the Psychomotor Vigilance Test. The associations on the 2 former tests (range, –0.45 to –0.76) are similar in magnitude to previous sleep deprivation studies,[62,63] but indicated only a limited ability of the residents to judge their impairment. The associations may have been highest on the simulated driving task because participants were better able to judge good driving rather than good reaction times. We additionally did not find systematic adaptation to chronic sleep loss effects with increasing training year despite self-reports of such adaptation.[4]

We controlled for methodological confounding variables present in previous studies[13] by requiring practice of the dependent measures, testing participants at the same time of day, objectively documenting sleep duration, including driving as a test with real-world relevance, and restricting medication, alcohol, and caffeine use. Robust call differences were found despite light call residents having a daily mean of only 6:38 and 6:37 of actigraphically defined sleep for the 7 days and the 24 hours before the test session. Greater differences might have been observed if participants maintained a consistent 8-hour sleep schedule during light call; however, our external validity is enhanced by having participants

self-select sleep schedules. Additional studies are needed that include truly rested control conditions to determine if impairment is present even on light call rotations.

Our study had several limitations. First, the small sample size meant that our main comparisons of interest had low statistical power, and we did not perform an intention-to-treat analysis. However, we successfully detected simulated driving differences between the heavy call with placebo and light call with alcohol groups and we had a relatively large sample size for studies on residents using a within-subjects design. We did not randomize or counterbalance the order of test conditions and cannot discount the possibility of order effects on our findings. However, the tests we used have relatively small practice effects[41,64,65] and all participants practiced the outcome measures before the first test session. In addition, secondary analyses, although underpowered, showed a similar pattern of results.

There may have been a self-selection bias, such that participants may have wanted to demonstrate worse impairment after heavy call than after alcohol ingestion, and our attempts to blind participants to the presence or absence of alcohol were frequently unsuccessful. However, we believe that these results are valid because we did not communicate our specific hypotheses to the participants, we found consistently worse light call with alcohol than light call performance, and effort ratings were higher in heavy call with placebo than light call with alcohol. It is unlikely that participants could have titrated their light call with alcohol performance to be systematically worse than light call but not worse than heavy call with placebo, or that they used greater effort in heavy call with placebo if the goal was to show worse heavy call with placebo than light call with alcohol impairment. Intentional poor performance in heavy call with placebo would have been achieved by exerting minimal effort on the tasks.

Although the tests selected for our study were carefully chosen surrogates for skills that we hypothesized

would be impaired by sleep loss in medical residents, we are unable to draw firm conclusions about the degree of training-related impairment associated with actual medical tasks or medical decision making. Our findings do suggest, however, that some of the constituent skills necessary to perform medical tasks are likely to be impaired post-call during a typical heavy call rotation. Finally, our results may not generalize to subspecialties other than pediatrics or to other residency programs with different light and heavy call rotation schedules.

In conclusion, our study demonstrates that resident performance impairment post-call after 4 weeks of heavy call is equivalent to or worse than the impairment observed at 0.04 to 0.05 g% BAC on tests of sustained attention, vigilance, and simulated driving. Moreover, residents' self-assessment of heavy call performance is limited and task-dependent. These findings have important clinical implications. Residents must be made aware of post-call performance impairment and the potential risk to personal and patient safety. There should be sleep loss, fatigue, and countermeasure education in residency programs. Because sleepy residents may have limited ability to recognize the degree to which they are impaired, residency programs should consider these risks when designing work schedules and develop risk management strategies for residents, such as considering alternative call schedules or providing post-call napping quarters. Additional studies should examine the impact of these operational and educational interventions on resident driving safety and on patient care and safety.

Author Affiliations: Department of Psychiatry and Human Behavior, Brown Medical School (Drs Arnedt, Owens, and Carskadon); Division of Ambulatory Pediatrics, Rhode Island Hospital (Dr Owens and Mss Crouch and Stahl); and Sleep and Chronobiology Research Laboratory, E. P. Bradley Hospital (Dr Carskadon), Providence, RI. Dr Arnedt is now with the Sleep and Chronophysiology Laboratory, Department of Psychiatry, University of Michigan, Ann Arbor.
Author Contributions: Dr Arnedt had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

REFERENCES

1. Baldwin DC Jr, Daugherty SR. Sleep deprivation and fatigue in residency training: results of a national survey of first- and second-year residents. *Sleep*. 2004; 27:217-223.
2. Baldwin DC Jr, Daugherty SR, Tsai R, Scotti MJ Jr. A national survey of residents' self-reported work hours: thinking beyond specialty. *Acad Med*. 2003;78:1154-1163.
3. Daugherty SR, Baldwin DC Jr, Rowley BD. Learning, satisfaction, and mistreatment during medical internship. *JAMA*. 1998;279:1194-1199.
4. Rosen IM, Bellini LM, Shea JA. Sleep behaviors and attitudes among internal medicine housestaff in a US university-based residency program. *Acad Med*. 2004; 79:407-416.
5. Denisco RA, Drummond JN, Gravenstein JS. The effect of fatigue on the performance of a simulated anesthetic monitoring task. *J Clin Monit*. 1987;3:22-24.
6. Jacques CH, Lynch JC, Samkoff JS. The effects of sleep loss on cognitive performance of resident physicians [see comments]. *J Fam Pract*. 1990;30:223-229.
7. Nelson CS, Dell'Angela K, Jellish WS, Brown IE, Skaredoff M. Residents' performance before and after night call as evaluated by an indicator of creative thought. *J Am Osteopath Assoc*. 1995;95:600-603.
8. Rubin R, Orris P, Lau SL, Hryhorczuk DO, Furner S, Letz R. Neurobehavioral effects of the on-call experience in housestaff physicians [see comments]. *J Occup Med*. 1991;33:13-18.
9. Deaconson TF, O'Hair DP, Levy MF, Lee MB, Schueneman AL, Codon RE. Sleep deprivation and resident performance [see comments]. *JAMA*. 1988;260: 1721-1727.
10. Bartle EJ, Sun JH, Thompson L, Light AI, McCool C, Heaton S. The effects of acute sleep deprivation during residency training. *Surgery*. 1988;104:311-316.
11. Reznick RK, Folse JR. Effect of sleep deprivation on the performance of surgical residents. *Am J Surg*. 1987;154:520-525.
12. Storer JS, Floyd HH, Gill WL, Giusti CW, Ginsberg H. Effects of sleep deprivation on cognitive ability and skills of pediatrics residents [see comments]. *Acad Med*. 1989;64:29-32.
13. Veasey S, Rosen R, Barzansky B, Rosen I,

Owens J. Sleep loss and fatigue in residency training: a reappraisal. *JAMA*. 2002;288:1116-1124.
14. Owens JA. Sleep loss and fatigue in medical training. *Curr Opin Pulm Med*. 2001;7:411-418.
15. Landrigan CP, Rothschild JM, Cronin JW, et al. Effect of reducing interns' work hours on serious medical errors in intensive care units. *N Engl J Med*. 2004;351: 1838-1848.
16. Lockley SW, Cronin JW, Evans EE, et al. Effect of reducing interns' weekly work hours on sleep and attentional failures. *N Engl J Med*. 2004;351:1829-1837.
17. Daugherty SR, Baldwin DC Jr. Sleep deprivation in senior medical students and first year residents. *Acad Med*. 1996;71(1 suppl):S93-S95.
18. Baldwin PJ, Dodd M, Wrate RW. Young doctors' health, I: how do working conditions affect attitudes, health and performance? *Soc Sci Med*. 1997;45:35-40.
19. Papp KK, Stoller EP, Sage P, et al. The effects of sleep loss and fatigue on resident–physicians: a multi-institutional, mixed-method study. *Acad Med*. 2004;79: 394-406.
20. Steele MT, Ma OJ, Watson WA, Thomas HA Jr, Muelleman RL. The occupational risk of motor vehicle collisions for emergency medicine residents. *Acad Emerg Med*. 1999;6:1050-1053.
21. McCartt AT, Ribner SA, Pack AI, Hammer MC. The scope and nature of the drowsy driving problem in New York state. *Accid Anal Prev*. 1996;28:511-517.
22. Marcus CL, Loughlin GM. Effect of sleep deprivation on driving safety in housestaff. *Sleep*. 1996;19:763-766.
23. Barger LK, Cade BE, Ayas NT, et al. Extended work shifts and the risk of motor vehicle crashes among interns. *N Engl J Med*. 2005;352:125-134.
24. Robbins J, Gottlieb F. Sleep deprivation and cognitive testing in internal medicine house staff. *West J Med*. 1990;152:82-86.
25. Risser MR, Manser TJ, Cain CL, Morewitz CL, Ware JC. Effects of call and gender on medical resident driving simulator performance [abstract]. *Sleep*. 2001; 24:A125.
26. Dawson D, Reid K. Fatigue, alcohol and performance impairment [letter]. *Nature*. 1997;388:235.
27. Lamond N, Dawson D. Quantifying the performance impairment associated with fatigue. *J Sleep Res*. 1999;8:255-262.
28. Williamson AM, Feyer A-M, Mattick RP, Friswell R, Finlay-Brown S. Developing measures of fatigue using an alcohol comparison to validate the effects of fatigue on performance. *Accid Anal Prev*. 2001;33:313-326.
29. Williamson AM, Feyer A-M. Moderate sleep deprivation produces impairments in cognitive and motor performance equivalent to legally prescribed levels of alcohol intoxication. *Occup Environ Med*. 2000; 57:649-655.
30. Liguori A, D'Agostino RB, Dworkin SI, Edwards D, Robinson JH. Alcohol effects on mood, equilibrium, and simulated driving. *Alcohol Clin Exp Res*. 1999;23:815-821.
31. Howland J, Rohsenow DJ, Cote J, Gomez B, Mangione TW, Laramie AK. Effects of low-dose alcohol exposure on simulated merchant ship piloting by maritime cadets. *Accid Anal Prev*. 2001;33:257-265.
32. Howland J, Rohsenow DJ, Cote J, Siegel M, Mangione TW. Effects of low-dose alcohol exposure on simulated ship power plant operation by maritime cadets. *Addiction*. 2000;95:719-726.
33. Vermeeren A, Riedel WJ, van Boxtel MP, Darwish M, Paty I, Patat A. Differential residual effects of Zaleplon and Zopiclone on actual driving: a comparison with a low dose of alcohol. *Sleep*. 2002;25:224-231.
34. Lamond N, Dorrian J, Burgess H, et al. Adaptation of performance during a week of simulated night work. *Ergonomics*. 2004;47:154-165.
35. Council on Scientific Affairs. Alcohol and the driver. *JAMA*. 1986;255:522-527.

36. Ogden EJ, Moskowitz H. Effects of alcohol and other drugs on driver performance. *Traffic Inj Prev*. 2004;5:185-198.

37. Cahalan D, Cisin I, Crossley H. *American Drinking Practices*. New Brunswick, NJ: Rutgers Center of Alcohol Studies; 1969.

38. Acebo C, Sadeh A, Seifer R, et al. Estimating sleep patterns with activity monitoring in children and adolescents: how many nights are necessary for reliable measures. *Sleep*. 1999;22:95-103.

39. Rohsenow DJ, Marlatt AG. The balanced placebo design: methodological considerations. *Addict Behav*. 1981;6:107-122.

40. Dinges DF, Powell JW. Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behav Res Methods Instrum Comput*. 1985;17:652-655.

41. *Conners' Continuous Performance Test* [computer program]. Toronto, Ontario: Multi-Health Systems Inc; 1995.

42. Friel PN, Logan BK, O'Malley D, Baer JS. Development of dosing guidelines for reaching selected target breath alcohol concentrations. *J Stud Alcohol*. 1999; 60:555-565.

43. Hoddes E, Dement W, Zarcone V. Quantification of sleepiness: a new approach. *Psychophysiology*. 1973;10:431-436.

44. Dinges DF, Pack F, Williams K, et al. Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4-5 hours per night. *Sleep*. 1997;20:267-277.

45. Van Dongen HP, Maislin G, Mullington JM, Dinges DF. The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. *Sleep*. 2003;26:117-126.

46. Rollinson DC, Rathlev NK, Moss M, et al. The effects of consecutive night shifts on neuropsychological performance of interns in the emergency department: a pilot study. *Ann Emerg Med*. 2003; 41:400-406.

47. Cavallo A, Ris MD, Succop P. The night float paradigm to decrease sleep deprivation: good solution or a new problem? *Ergonomics*. 2003;46:653-663.

48. Arnedt JT, Wilde GJS, Munt PW, MacLean AW. Simulated driving performance following prolonged wakefulness and alcohol consumption: separate and combined contributions to impairment. *J Sleep Res*. 2000;9:233-241.

49. Arnedt JT, Wilde GJS, Munt PW, MacLean AW. How do prolonged wakefulness and alcohol compare in the decrements they produce on a simulated driving task? *Accid Anal Prev*. 2001;33:337-344.

50. Doran SM, Van Dongen HPA, Dinges DF. Sustained attention performance during sleep deprivation: evidence of state instability. *Arch Ital Biol*. 2001; 139:253-267.

51. Casbon TS, Curtin JJ, Lang AR, Patrick CJ. Deleterious effects of alcohol intoxication: diminished cognitive control and its behavioral consequences. *J Abnorm Psychol*. 2003;112:476-487.

52. Fairclough SH, Graham R. Impairment of driving performance caused by sleep deprivation or alcohol: a comparative study. *Hum Factors*. 1999;41:118-128.

53. Roehrs T, Beare D, Zorick F, Roth T. Sleepiness and ethanol effects on simulated driving. *Alcohol Clin Exp Res*. 1994;18:154-158.

54. Pilcher JJ, Huffcutt AI. Effects of sleep deprivation on performance: a meta-analysis. *Sleep*. 1996;19: 318-326.

55. Belenky G, Wesensten NJ, Thorne DR, et al. Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: a sleep dose-response study. *J Sleep Res*. 2003;12:1-12.

56. Smith-Coggins R, Rosekind MR, Buccino KR, Dinges DF, Moser RP. Rotating shiftwork schedules: can we enhance physician adaptation to night shifts? *Acad Emerg Med*. 1997;4:951-961.

57. Bartel P, Offermeier W, Smith F, Becker P. Attention and working memory in resident anaesthetists after night duty: group and individual effects. *Occup Environ Med*. 2004;61:167-170.

58. Taffinder NJ, McManus IC, Gul Y, Russell RC, Darzi A. Effect of sleep deprivation on surgeons' dexterity on laparoscopy simulator. *Lancet*. 1998;352:1191.

59. Grantcharov TP, Bardram L, Funch-Jensen P, Rosenberg J. Laparoscopic performance after one night on call in a surgical department: prospective study. *BMJ*. 2001;323:1222-1223.

60. Haynes DF, Schwedler M, Dyslin DC, Rice JC, Kerstein MD. Are postoperative complications related to resident sleep deprivation? *South Med J*. 1995;88:283-289.

61. Lee HC, Cameron D, Lee AH. Assessing the driving performance of older adult drivers: on-road versus simulated driving. *Accid Anal Prev*. 2003;35:797-803.

62. Baranski JV, Pigeau RA. Self-monitoring cognitive performance during sleep deprivation: effects of modafinil, d-amphetamine and placebo. *J Sleep Res*. 1997;6:84-91.

63. Dorrian J, Lamond N, Dawson D. The ability to self-monitor performance when fatigued. *J Sleep Res*. 2000;9:137-144.

64. Kribbs NB, Dinges DF. Vigilance decrement and sleepiness. In: Harsh JR, Ogilvie RD, eds. *Sleep Onset: Normal and Abnormal Processes*. Washington, DC: American Psychological Association; 1994: 113-125.

65. Arnedt JT, Acebo C, Seifer R, Carskadon MA. Assessment of a simulated driving task for sleep research [abstract]. *Sleep*. 2001;24:A413.