Sleep and Motor Performance in On-call Internal Medicine Residents

Akash D. Saxena, MD, FRCPC; Charles F.P. George, MD, FRCPC

University of Western Ontario, Division of Respirology, London, Ontario, Canada

Study Objectives: To compare vigilance and performance among internal medicine residents doing in-house call versus residents not doing in-house call.
Design: Prospective study of resident cohorts with repeated testing.
Setting: University Teaching Hospital
Participants: Internal medicine residents doing in-house call and residents not doing in-house call (pathology, endocrinology) (controls).
Measurements And Results: Subjective sleepiness scores (daily Stanford Sleepiness Scale and Epworth Sleepiness Scale at start and end of the test period), actigraphy, and daily sleep logs as well as regular psychomotor vigilance testing using a Palm® version (Walter Reed Army Institute of Research) of the Psychomotor Vigilance Test (PVT®). Subjects were enrolled for a period of 28 to 32 days, which included 4 to 6 on-call nights for the internal medicine residents. Controls took call from home. Participants were compensated for their time.
Results: Twenty residents were evaluated, 13 internal medicine and 7 controls. Overall median reaction time was slower in the internal medicine residents (264.7 ± 102.9 vs 239.2 ± 26.1 milliseconds; $P < .001$). Internal medicine residents showed no difference in reaction time postcall versus other periods (269.9 ± 131.2 vs 263.6 ± 95.6; $P = .65$). Actigraphic sleep time was shorter during on-call than noncall nights and in internal medicine residents as compared with controls (287.48 ± 143.8 vs 453.49 ± 178.5 and 476.08 ± 71.9 minutes; $P < .001$). Internal medicine residents had significantly greater major and minor reaction-time lapses compared with controls (1.26 ± 3.4 vs 0.53 ± 1.1 & 2.4 ± 7.4 vs 0.45 ± 1.0; $P < .001$). They reported increased sleepiness on postcall days compared with the start of their call (Stanford Sleepiness Scale: 3.26 ± 1.2 vs 2.22 ± 0.8; $P < .001$) but had scores similar to those of controls by their next call (2.22 ± 0.8 vs 2.07 ± 0.8; $P = .13$).
Conclusions: Internal medicine residents have impaired reaction time and reduced vigilance compared with controls. Despite subjective improvements in sleepiness postcall, there was no change in their objective performance across the study period, suggesting no recovery. Internal medicine residents did not get extra sleep on postcall nights in an attempt to recover their lost sleep time. Implications for residents' well-being and patient care remain unclear.
Keywords: Sleepiness, performance, on-call, internal medicine, residents
Citation: Saxena A; George CFP. Sleep and motor performance in on-call internal medicine residents. *SLEEP* 2005;28(11): 1386-1391.

INTRODUCTION

COGNITIVE AND MOTOR FUNCTION ARE IMPAIRED WITH ACUTE AND CHRONIC SLEEP DEPRIVATION.[1-6] MEDICAL RESIDENTS ARE SUBJECTED TO REGULAR sleep restriction due to their on-call responsibilities.[7] Recent reports indicate that interns doing frequent emergency room night shifts of 12 hours show significant deterioration in visual memory and psychomotor vigilance from the beginning to the end of their duty.[8] Subjective assessment by residents using the Epworth Sleepiness Scale (ESS) suggests levels of fatigue equivalent to that found in patients with serious sleep disorders such as sleep apnea and narcolepsy.[9] In driving-simulation experiments, postcall pediatric residents have response times equivalent to those produced in normal subjects with a blood alcohol level of 0.05.[10] Current regulations in Ontario, Canada, permit in-house on-call duties averaging up to 1 in 4 nights (average 4-6 calls per month). Residents may work continuously for up to 28 hours without rest or sleep. There are limited objective data to indicate whether or not residents have recovered by their next night shift.[11] Moreover, the long-term health effects on them and the effects on patient care are not well known.

This study was designed to test the following 3 hypotheses: (1) internal medicine residents have overall impaired performance in objective testing, as compared with residents who do not take in-house call; (2) internal medicine residents show an acute deterioration in their vigilance and motor performance across a single on-call duty; and 3) internal medicine residents will demonstrate objective improvement (recovery) between call duties.

METHODS

We recruited internal medicine residents performing in-house call (postgraduate years 1-3) and control residents from services that are traditionally quiet during nighttime hours and who do not take in-house call (pathology, endocrinology, nuclear medicine). A 45-minute PowerPoint presentation was given to all potential subjects to outline the study. All participants were aware of this study's intention to obtain objective and subjective measures of sleepiness. Subjects came from 2 tertiary care teaching hospitals within the same city and healthcare network. Due to division of services within our hospital network, internal medicine residents were enrolled at 1 site, while the control residents came from the other campus. The study ran from September 2003 through to March 2004.

Inclusion criteria were as follows: all medical residents doing in-house medicine call and residents not taking in-house call and who are from services with a reasonable expectation of getting at least 5 hours of continuous sleep even when on call (control group). Exclusion criteria were failing to provide informed consent, age under 18 or over 35 years as of their enrollment

Disclosure Statement
This was not an industry supported study. Drs. George and Saxena have indicated no financial conflicts of interest.

Submitted for publication September 2004
Accepted for publication July 2005
Address correspondence to: Charles F.P. George, MD, FRCPC, London Health Sciences Centre—South Street Hospital, 375 South Street, London, ON Canada N6A 4G5; Tel: (519) 667-6860 ; Fax: (519) 667-6552; E-mail: cgeorge@uwo.ca

Table 1—Demographics for the Resident Groups

Variable	Medicine n=13	Controls n=7	P Value
Age, y*	27.0 ± 1.1	29.9 ± 2.9	.001
Sex† (Men:Women)	7:6	3:4	NS
Race (Caucasian:Other)	5:8	3:4	NS
Handedness (Right:Left)	11:2	6:1	NS

*Data are presented as mean ± SD.
†Data are presented as number.

start date, residents on service other than internal medicine doing in-house call, those with a history of a sleep disorder or serious comorbid illness (ie, diabetes, epilepsy), current prescription sedative or stimulant use, medicine residents not doing at least 3 in-house calls during a 30-day period, taking holidays (>3 days) during the study period, pregnancy, those who moonlight, control subjects who had recently (within the previous 10 days) done in-house call, participants in another study, or those failing to complete at least 80% of the required tasks.

Subjects were studied for between 28 and 32 days. Each participant gave informed consent, completed a demographic survey, and met both inclusion and exclusion criteria. As a measure of overall sleepiness, they completed the ESS at study entry and exit.[12] Participants wore an activity monitor throughout their study period, which was used to estimate sleep (Actiwatch®, Minimitter, Bend, Ore). They also completed a weekly sleep log, a daily 5-minute Palm® version of the psychomotor vigilance test (PVT), and the Stanford Sleepiness Scale (SSS) at repeated intervals.[13,14] Both the PVT and SSS measures were completed between 7 and 10 AM Eastern Standard Time. To get a sense of acute changes in sleepiness, medicine residents were required to complete the SSS scale both on call and postcall, while the control group completed this scale twice per week (either every Monday and Thursday or Tuesday and Friday). Figure 1 outlines the protocol. Monthly on-call schedules for medicine residents were obtained from the department of medicine. The study results were analyzed on a per-protocol basis. Eighty percent completion of each task was required for analysis inclusion. For analysis of the actigraphy data, if a subject failed to wear the activity monitor for 1 night, the entire time of no activity could either be considered completely asleep or as missing data. Data were analyzed both ways.

Several outcome measures were determined: median and 10% slowest and fastest reaction times for controls and medicine residents from their on-call, postcall and noncall days. Also, we compared (square root transformed) lapses[15] between the 2 groups (medicine/controls). For this study, minor lapses were defined as any reaction time > 500 milliseconds. Major lapses were an adjustable variable defined as the number of reaction times greater than 2 times the median reaction time for that specific trial. We reasoned that this measurement of lapses would be more sensitive to changes in vigilance within an individual on a given day. Lapses exceeding 30 seconds were also measured, but the number of instances in which this occurred among and between the groups was extremely rare. Scores on the SSS throughout the study and ESS at study start and end were also compared between the resident groups. Finally, sleep times recorded through self-reported sleep logs and an objective activity monitor were compared for agreement. Statistical analysis included unpaired student *t* tests (assuming equal variances, Bonferroni correction for multiple

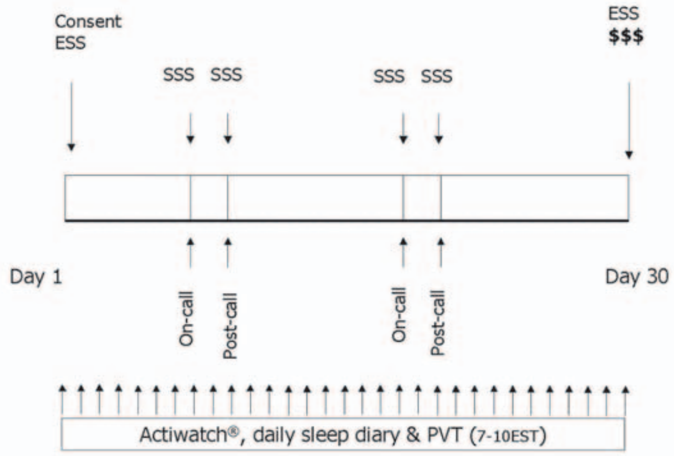

Figure 1—Outline of testing protocol. ESS refers to Epworth Sleepiness Scale; SSS, Stanford Sleepiness Scale; PVT, Psychomotor Vigilance Task; $$$, compensation at the end of the study. Actiwatch®, sleep diary, and PVT performed daily. SSS recorded both before and after on-call periods, and ESS performed at start and end of the study.

comparisons), 1-way analysis of variance, and simple regression analysis.

This project met the approval of the Health Sciences Research Ethics Board at the University of Western Ontario. Subjects were free to withdraw at anytime, and all participants were compensated. The amount of compensation was based on the degree of compliance with the study procedures. Completion of all tasks resulted in payment of $100 (Canadian). All participants were contacted each week to encourage compliance and reminded about the study requirements. Subjects were also discouraged from communicating with each other about reaction times. A bonus of $100 was awarded to the resident with the overall fastest mean reaction time on the PVT, independent of their group (medicine or control).

RESULTS

There were 56 potential internal medicine residents. However, only 35 worked at the main site for enrollment during the study. All were aware of the study via lectures and announcements, but only 26 were individually approached. Four declined to be studied. One subject was over the maximum age requirement. Three other subjects planned to take holidays during their possible enrollment period. Finally, 1 medicine resident did not meet the minimum call per month requirement.

In all, 18 internal medicine residents were enrolled. Unfortunately, only 13 met the final requirements for analysis. Of the 5 excluded, 4 failed to meet the protocol requirements. The reasons for exclusion were multifactorial. Subjects were unable to complete the PVT during the specific time periods (7-10 AM), complete the task on at least 80% of their study day's or consistently wear the activity monitor. One individual had a technical failure with the activity monitor (water damage). The demographics of those excluded did not differ from the other medical residents. Ten control subjects from the departments of endocrinology, pathology and nuclear medicine were asked to participate. Two of these subjects were subsequently declined enrollment due to exceeding the maximum age limit, and the third was moonlighting (n = 7; 3 endocrinology, 3 pathology, 1 nuclear medicine).

Table 2—Psychomotor Vigilance Test Data for Control and Medicine Residents

Group	Median RT (Overall)	Slowest 10% RT	Fastest 10% RT
Medicine	264 ± 103	582 ± 539	191 ± 12
Control	239 ± 26*	374 ± 111**	179 ± 14**
Medicine on call		573 ± 537	192 ± 13
Medicine post call		611 ± 611	191 ± 12
Medicine regular		584 ± 473	191 ± 12

RT refers to reaction time in milliseconds. Results are mean (or median) ± SD.
*P < .001 compared with medicine (overall).
**P < .001 compared with all medicine conditions.

A total of 13 internal medicine and 7 control residents met criteria for analysis. Their demographics are presented in Table 1. The only statistically significant difference between groups was age, with control residents being slightly older (29.9 ± 2.9 years vs 27.0 ± 1.1 years, P < .001). However, the age difference was small (3 years) and unlikely to be clinically significant in terms of performance-outcome data. There was a significant difference in overall median reaction time (264 ± 103 milliseconds vs 239 ± 26 milliseconds, P < .001), with medical residents being slower. Surprisingly, internal medicine subjects showed no difference in reaction time after call compared with all other times (270 ± 103 milliseconds vs 264 ± 96 milliseconds; P = .65). The mean of the fastest and slowest 10% of responses were significantly different between groups but did not differ within the medicine residents regardless of condition (Table 2). Also, medical residents had significantly more major (2.5 ± 2.2 vs 1.53 ± 0.92) and minor reaction-time lapses compared with controls (2.9 ± 3.2 vs 1.5 ± 0.91; both P < .001; Table 1). False starts, defined as initiating a response when no stimulus was presented on the Palm® PVT, was less likely to occur among medicine residents (2.2 ± 1.1 vs 3.2 ± 1.5; P < .001). Figure 2 demonstrates the differences in lapses between subjects from each group. Control residents had few, if any, lapses, while medicine residents had several—often up to 4 seconds.

Figure 3 illustrates the difference in activity at night between a medicine and control subject. The medicine resident, as expected, showed a more erratic sleep schedule due to call duties. The control subject had quite regular sleep hours.

There was a significant discrepancy between subjective total sleep time (diary) and objective estimate of sleep time (actigraphy), particularly for the medical residents. As seen in Figure 4, medical residents consistently underestimated or overestimated their actual sleep time, and this was different than that of controls (r = 0.46 vs 0.55; P < .05 Figure 4). Average time of on-call sleep (in minutes) measured by Actiwatch monitors, as expected, differed from those on noncall nights and of control subjects (287.5 ± 143.8 vs 453.5 ± 178.5 or vs 476.1 ± 71.9; P < .001). However, medicine residents did not show any difference in the total amount of sleep on postcall nights compared with either other noncall nights or the sleep of controls (Figure 5).

Based upon the SSS scores, medical residents did report more sleepiness after call (2.22 ± 0.8 vs 3.26 ± 1.2; P < .001), but the scores were equivalent to those of control subjects by their next on-call duty (2.22 ± 0.8 vs 2.07 ± 0.8, P = .13). ESS scores for the

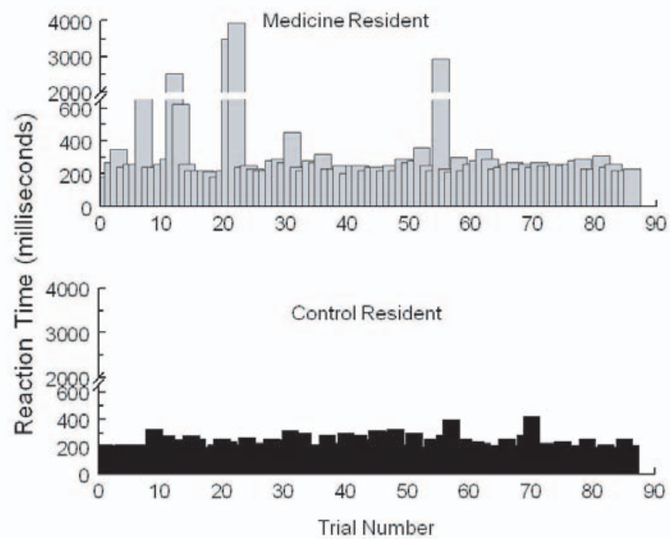

Figure 2—Reaction time and lapses on the Psychomotor Vigilance Test (PVT). For illustration, a single PVT trial for a medicine resident is contrasted with the PVT for a control resident. Only the medicine resident has prolonged lapses.

medicine residents trended higher during their study period (6.9 ± 3.2 vs 8.0 ± 3.4; P = .07), while there was no change for the controls (8.0 ± 3.4 vs 7.7 ± 3.5; P = .44; Figure 6).

DISCUSSION

This study differs in design from previously reported literature in 3 ways. First and most important, to our knowledge, this is the only published study to have a prospective control cohort. Second, while previous studies have used in-laboratory or complex tasks to measure vigilance, we used a simple-to-use Palm® version of the PVT, allowing measurements in the field. Finally, we measured sleep time through both objective estimates (actigraphy) and subjective (sleep-log diary) methods. Previous reports have used only sleep-log data or an activity monitor. While not the gold standard for measuring sleep (c.f. polysomnography), actigraphy provides reliable objective estimates of sleep, especially in subjects without sleep apnea or insomnia.[16-18] Given the longitudinal design with 28 to 32 days of data collection for each subject, we felt this was the most practical means of data collection.

As expected, medical residents demonstrated impaired vigilance. But, more concerning, is their lack of improvement or "recovery" between call duties and misperception about their chronic level of impairment. We suggest that this lack of change in post-call reaction times compared with other periods for internal medicine residents is due to chronic fatigue (sleep deprivation). Our study suggests that current guidelines allowing 4 to 6 on-call duties per month may not allow sufficient time for motor performance recovery. Recently, in the United States, similar on-call guidelines have been enacted as those in Ontario. However, our study suggests that, even with the current schedule, residents experience significant chronic fatigue.

Also of importance is the wider variances seen among medical residents in most of their measurements, both objective and subjective. Two possible explanations exist for why medical residents had a trend toward higher ESS scores. First, as sleepiness was one of the stated outcome measures, it might be argued that, through the course of the study, medicine residents

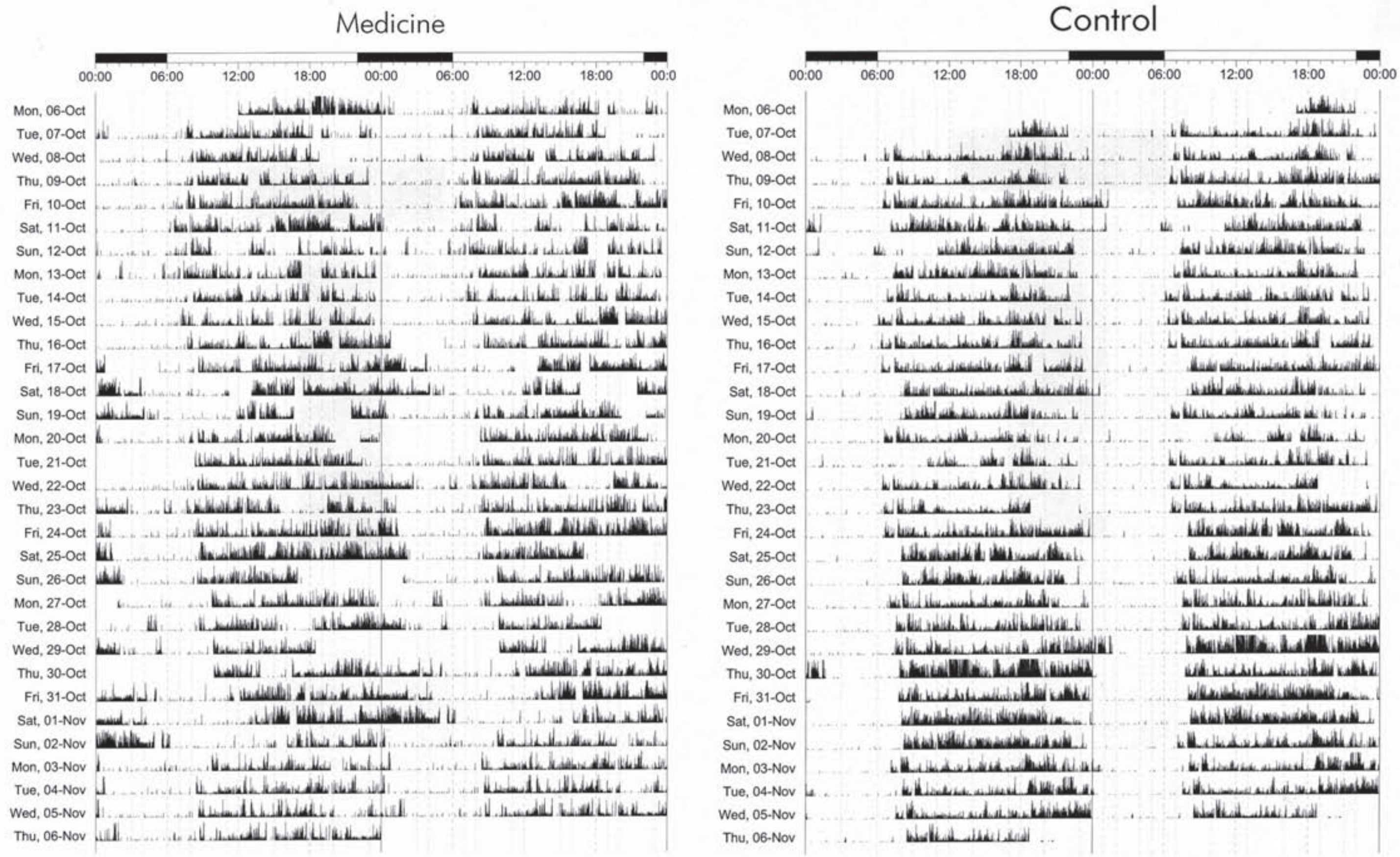

Figure 3—Monthly actigraphy for medicine resident (left) and control resident (right). The control resident has a very regular activity (sleep-wake) pattern over the month, whereas that of the medical resident is more erratic.

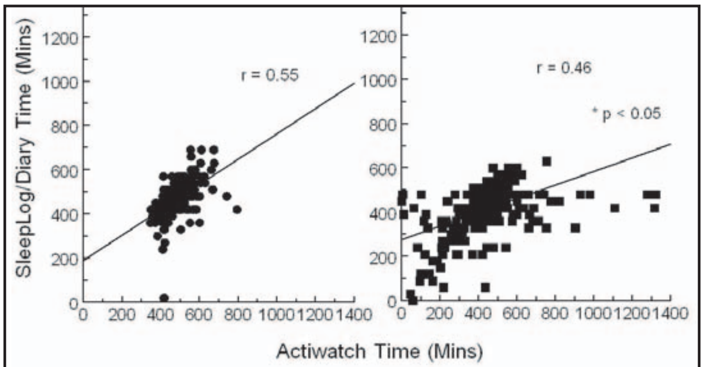

Figure 4—Relationship between Actiwatch® (objective) and sleep log (subjective) sleep time for medicine and control residents. Agreement between sleep measures is greater for control residents. *P < .05 for difference between groups.

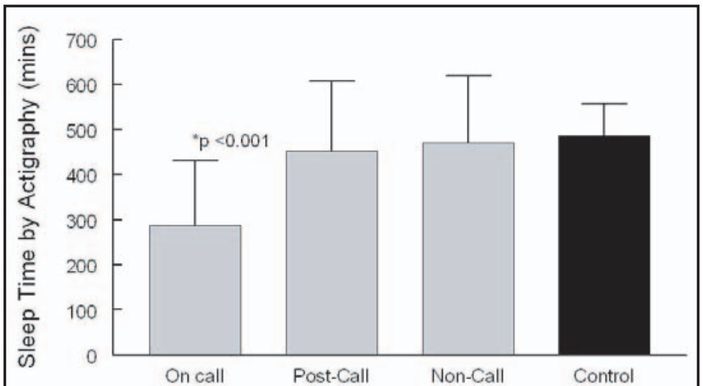

Figure 5—Total sleep time, as measured by actigraphy, for medicine residents (grey bars) and control residents (Black bar). Values are mean ± SD. Only on-call sleep time is shorter.

might be biased to be more aware of these variables. However, control subjects had the same study cues and did not report any change. The second, and we believe more likely explanation, is that general internal medicine residents develop a chronic sleep debt that does not recover during the interval between on-call duties despite a frequency of only 4 to 6 nights per month. Although we did not measure or survey off-duty activities, we believe that residents either cannot or do not avail themselves of the opportunity to recover sleep during their off-duty and noncall days. This is supported by the lack of increase in sleep time inferred from actigraphy.

We acknowledge that there may be the potential for selection bias, and, while our control group may not be ideal, we believe it was the best available for comparison. Compared with internal medical residents, residents with regular working hours may differ in some factors, including personality. Certainly career selection is a personal choice. These differences could explain some of our differences. However, practically speaking, we did not control for personality. Still, some control residents, such as those in endocrinology, will have had to complete 3 years of internal medicine training prior to subspecialty training. And, conversely, some of the internal medicine residents could well be destined to be become endocrinology residents. Thus, we suggest that, while not measured, it may be just as likely to find these "control" traits in either group, thus limiting the differences between groups. Table 1 indicates a significant but small age difference between the 2 groups (controls being older by 3 years). This is likely a clinically insignificant difference, and, while reaction times on

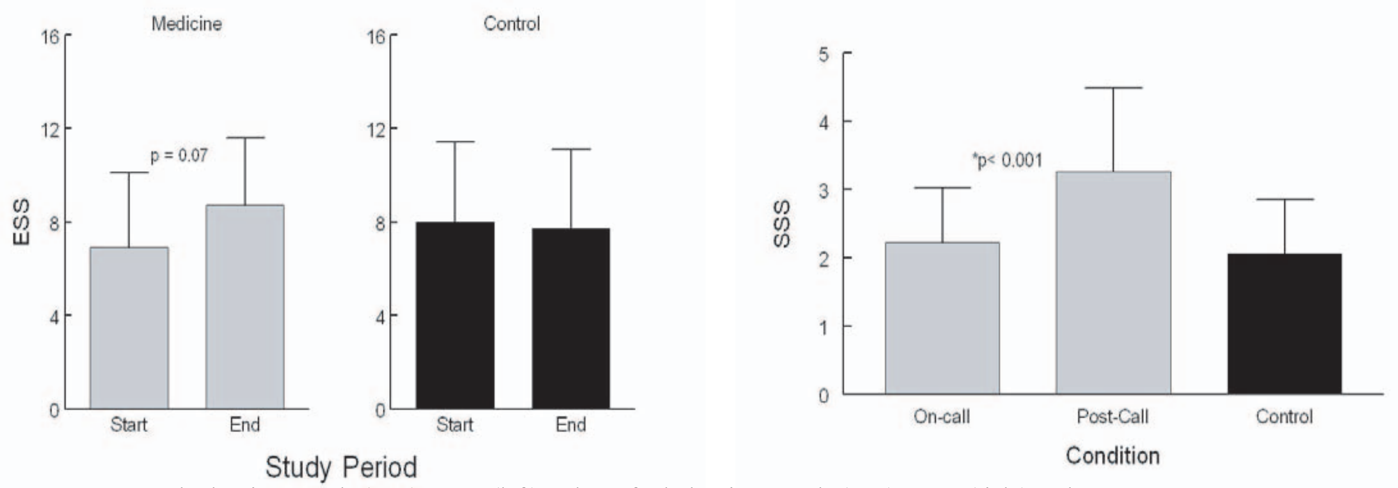

Figure 6—Epworth Sleepiness Scale (ESS) scores (left) and Stanford Sleepiness Scale (SSS) scores (right). Values are mean ± SD

vigilance tasks slow with age, it is unlikely that this small age difference in this age group would account for the differences in PVT. In fact, if age is an important confounder, then we would expect the older control residents would have worse results, which would bias against the observed differences in reaction time.

Medical residents show on average of a 40-milliseconds prolongation in their mean reaction time. This difference is primarily related to brief prolonged lapses as opposed to a diffuse slowing. The clinical significance of this finding is not clear. False starts were also reduced among medical residents. Typically, false starts increase with increasing sleepiness, so we should expect to see these increased in the medical residents. We cannot readily explain this finding.

Current Ontario regulations, in place for more than 5 years, permit 1-in-4 in-house call but require that residents be relieved of duty not later than noon the day after a night on call (≤ 28-hour shift). Despite these regulations, medicine residents do not seem to avail themselves of the opportunity to sleep (as evidenced by the lack of increased actigraphic sleep times) and, by extension, to improve their (reaction-time) performance. The United States has recently adopted similar call restrictions, but it is as yet unknown how these will influence resident behavior and performance in the long term. There are recent data showing that short-term changes in call schedule can reduce medical errors in the intensive care unit,[19] but whether that change can be maintained in other settings and in the long term remains unanswered.

Implications for patient care are unclear. Likely, in an emergency situation, judgment would not be impaired, but "routine" tasks (ie, medication reorders) might be missed, which could later lead to more serious consequences. This study did not address medical error due to sleep deprivation, but the persistent impairment in vigilance between call duties and misperception about their state among internal medicine residents is concerning.

Practical clinical recommendations from this study include (1) educating medicine residents about their persistent impairment compared against some other specialties requiring one to actively focus especially on "routine" tasks and (2) encouraging residents to increase their sleep time on postcall nights. The major limitation to this study was the sample size (N = 20). However, we feel that the rigorous nature of our design and previous reports on this topic raise significant concerns about resident well-being and patient care with an on-call frequency of up to 6 times per month.

In conclusion, we have shown that internal medicine residents have overall both impaired reaction time and vigilance compared against residents who do not take in-house call. There was no change in their objective performance across the study period, suggesting persistent impairment. Medicine residents did report a relative increase in their level of sleepiness after call. However, of great concern is that they rated their level of alertness comparable to controls by the start of their next on-call duty despite objective data to the contrary. Medicine residents misperceive their level of vigilance. Finally, medical residents did not get extra sleep on postcall nights in an attempt to recover. Reasons for lack of recovery sleep were not addressed in this study but likely relate to personal and family commitments. Implications for resident well-being and patient care remain unclear. Studies directly addressing sleep impairment and medical mistakes are needed.

REFERENCES

1. Wilkinson RT, Edwards RS, Haines E. Performance following a night of reduced sleep. Psychonom Sci 1966;5:471-2.
2. Webb WB, Agnew HW. The effects of a chronic limitation of sleep length. Psychophysiology 1974;11:265-74.
3. 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.
4. Jewett ME, Dijk DJ, Kronauer RE, Dinges DF. Dose-response relationship between sleep duration and human psychomotor vigilance and subjective alertness. Sleep 1999;22:171-9.
5. 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;26:117-26.
6. 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.
7. Deaconson TF, O'Hair DP, Levy MF, et al. Sleep deprivation and resident performance. JAMA 1988;260: 1721-5.
8. Rollinson DC, Rathlev NK, Moss M, et al. Effects of consecutive night shifts on neuropsychological performance of interns in the ER. Ann Emerg Med 2003;41:400-6.
9. Sleep, Alertness, and Fatigue in Residency (S.A.F.E.R.) education Module. American Academy of Sleep Medicine, 2002. Westchester IL.
10. Owens JA, Arnedt J, Crouch M, Stahl J. Sleep loss and fatigue in pediatric residents: self-report of sleep patterns and impact on performance. Sleep 2004;27: A149.

11. 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.
12. Johns MW A new method for measuring daytime sleepiness: the Epworth sleepiness scale. Sleep 1991;14:540-5.
13. Hoddes E, Zarcone V, Smythe H, Phillips R, Dement WCl. Quantification of sleepiness: a new approach (Stanford sleepiness scale). Psychophysiology 1973;10:431-6.
14. Thorne DR, Johnson DE, Redmond DP, Sing HC, Belenky G. The Walter Reed Palm-held Psychomotor Vigilance test. Sleep 2003;26: A182.
15. Dinges DF, Powell JW. Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. Behav Res Methods Instrum Comput 1985;17:625-55.
16. Cole RJ, Kripke DF, Gruen W, Mullaney DJ, Gillin JC. Automatic sleep/wake identification from wrist activity. Sleep 1992;15:461-9.
17. Jean-Louis G, von Gizycki H, Zizi F, et al. Determination of sleep and wakefulness with the actigraph data analysis software (ADAS). Sleep 1996;19:739-43.
18. Blood ML, Sack RL, Percy DC, Pen JC. A comparison of sleep detection by wrist actigraphy, behavioral response, and polysomnography. Sleep 1997;20:388-95.
19. 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;351:1838-48.