

# UC Berkeley

## Theses

### Title

I Can Do This With My Eyes Closed: Overnight Call and the Quality of the Patient Encounter

### Permalink

<https://escholarship.org/uc/item/0rt6t99p>

### Author

Sussman, Jeremy B

### Publication Date

2003-04-01

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives License, available at <https://creativecommons.org/licenses/by-nc-nd/4.0/>

I Can Do This With My Eyes Closed:  
Overnight Call and the Quality of the Patient Encounter

by

Jeremy Broder Sussman

B.A. (Amherst College) 1998

A thesis submitted in partial satisfaction of the

requirements for the degree of

Master of Science

in

Health and Medical Sciences

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:

Professor Jeffrey Burack, Chair

Professor Sharad Jain

Professor Nicholas Jewell

Spring 2003



---

## Acknowledgements

This project was the work of many people, to only thank a few.

- Sharad Jain, whose patience and thoughtfulness often carried me when the burdens of the work were weighing me down.
- John Peabody, whose work developing the vignettes made the thesis possible, and whose work on my project consistently assured its quality.
- Dan Bertenthal and Brett Lewis, who provided technical help on a moments notice more times than I ever thought I would need.
- Jeff Burack and Nicholas Jewell, who were available when I needed them to be.
- Will Huen and Divy Ravindranath, whose constant availability for academic and non-academic advice was constantly utilized.
- To my family and friends who were supportive and thoughtful at times that I often was not.

# Table of Contents

<b>Preface</b>	<b>1</b>
<b>Overnight call, resident fatigue, and medical care: a review of the issues</b>	<b>3</b>
Introduction	3
Methods	6
Results	7
The Effect of Call and Resident Fatigue on Mood	7
Real-world studies	8
Performance on Medical Boards and Board-type Questions	10
Assessments of Medical Skills	12
Discussion	17
Limitations of the Research	18
Possible Future Studies	22
Conclusion	25
<b>Clinical Vignettes to Examine the Effect of Fatigue on the Quality of Medical Care</b>	<b>26</b>
Introduction	27
Methods	28
Participants	28
Interventions	28
Measurement	29
Objectives	30
Outcomes	30
Sample Size	30
Blinding	30
Statistical Methods	31
Results	33
Baseline Information	33
Outcomes	33
Comment	37
<b>Works Cited</b>	<b>40</b>

## List of tables

Table 1: Summary of studies on fatigue and board-type tests	12
Table 2: Summary of studies on fatigue and medical skills	14
Table 3: Sampled scoring categories for uncomplicated diabetes mellitus	32
Table 4: Effect of call status on performance of individual domains of medical care	35
Table 5: Effect of call status on performance of individual medical conditions	36

## List of Figures

Figure 1: clinical scenarios portrayed by standardized patients	31
Figure 2: Effect of call status on vignette performance. Error bars are SE. 1-sided paired t-test $p = 0.28$	35

# **I CAN DO THIS WITH MY EYES CLOSED: OVERNIGHT CALL AND THE QUALITY OF THE PATIENT ENCOUNTER**

## **INTRODUCTION**

Medical residents are neither independent professionals nor dependent students. They have finished four years of medical school, but are usually not allowed to independently practice medicine. They perform work that is nearly identical to the doctors who oversee them, but are often paid less than minimum wage. While all residents have chosen their path and all had been given many other options, once they choose their residency program they have almost no control over their lifestyles or their schedules.

One aspect of this lack of control is the long hours that residents are required to work. One hundred hour weeks are common, as are shifts that last 36 consecutive hours. Shifts are usually based around overnight call—the long days where a resident takes control over new patients in the morning and does not leave until after those patients have stabilized sometime the next day. This assures that a patient will have one doctor for the entire length of their acute illness. One doctor who knows their case and their problems.

There is concern, however, that residents long hours inhibit their ability to provide effective medical care. The public and the medical community have both been concerned that the expense of these long hours may actually be to the patient's detriment. Perhaps the long hours, and fatigue that go with them, may harm the doctors ability to provide effective patient care.

Research on the health effects of the hours that medical residents work matter has been inconsistent in quality and scope. The studies have often relied on psychometric

tests of questionable validity to medical practice and different practice patterns between residency programs leaves a tremendous uncertainty about how different fields of medicine are affected differently by fatigue.

In the first half of this project, I will review the literature on the effect of fatigue and overnight call by residents on the medical care that they provide. I will briefly examine the effect of fatigue on the mood of residents and then examine more thoroughly the studies that have looked at the effect of fatigue on performance on board exams and board-type questions, the effect of fatigue in “real-world” studies, and the effect of fatigue in studies using medical simulators and medical skills. I will examine the flaws in the research and propose a set of studies that could help more-clearly demonstrate the effect that fatigue, overnight call, and long working hours have on medical care.

In the second half of the project I will present the results of my own study, which was designed to answer some of these questions. Using a crossover design, I looked at residents after a normal full work-day and after a post-call work-day and, using a series of computerized clinical vignettes, attempted to determine if their ability to provide effective medical care was compromised by overnight call.



# Overnight call, resident fatigue, and medical care: a review of the issues

---

## INTRODUCTION

It's a topic that will not go away. The current system of training for medical interns and residents was introduced to America in 1889 by none other than Drs. Osler, Halsted, and Kelly.<sup>1</sup> It seems like criticism of it has been around for nearly as long.

Medical residents are both students and professionals. They have already received their diploma and are often legally licensed to practice medicine, but must complete the training to specialize. They are paid an average of \$40,000 per year and often work 100 hour weeks<sup>2</sup> to practice medicine, with little formal educational training and little academic support.

Concerns over the implications of residents' long hours, stressful environment, laborious non-medical work (often referred to as scut), and few organized educational activities has been a standard argument in the medical and non-medical community alike. The first research on the effects of residency on patients and the residents themselves was carried out in the early 1970's and academic research and calls to reform the system have occurred regularly since then.<sup>3, 4</sup>

The issue became a major national discussion beginning in 1984. This is when Libby Zion, daughter of lawyer and journalist Sidney Zion, died in New York Hospital of mysterious causes. Mr. Zion did not blame the individual doctors for his daughter's death, he blamed the entire system of medical education. In 1986, a grand jury agreed with him. In 1987, the results of the Bell Commission were published, making New York State the only state in the country to limit the number of hours that medical

residents could work. Along with other limitations, these regulations limited residents' work hours to no more than 80 per week.<sup>4,5</sup>

The issue moved in and out of the national consciousness until the last 3 years. In 1999, the Institute of Medicine released its hugely influential report on medical errors, *To Err is Human*.<sup>6</sup> While resident fatigue did not play a large role in the study, the public and interested professionals quickly made the association and began lobbying many organizations to limit resident working hours, which finally occurred in 2002, when the Accreditation Council on Graduate Medical Education<sup>7</sup> established a set of guidelines that will be required for training program accreditation beginning in June, 2003. The regulations limit resident working hours to 320 hours per four weeks, four days off per four weeks, permit call no more frequently than every third night, place a 24-hour limit on on-call duty (plus 6 extra hours for transfer of care), and a mandate a 10-hour minimum rest between duty periods.<sup>6</sup> This change occurred after a long campaign by the American Medical Students Association followed by an endorsement (some would say capitulation) by the American Medical Association.<sup>8</sup>

There is a fairly large amount of research on the effects of fatigue on medical care, but it is, generally of poor quality. The studies usually look at the effects of fatigue on residents' mood, cognitive functioning, physical functioning, or health. Most do this by using psychometric evaluations and survey instruments, while a few others utilize board-type exam questions and various types of medical simulators.

I am going to analyze in-depth all studies which examine the effect of fatigue and overnight call on work-related activities and board-type examination questions. I am going to do a less-thorough analysis of the effects of call on mood and the physical

effects on residents due clear results in these studies. I will also look at a number of factors which could influence the outcomes of such studies, including limitations of the studies, relevant biases, the role of chronic fatigue, and the role stress could play on the relationship between call, acute fatigue, and patient care. I will include a brief analysis of strengths and weaknesses of the state of the research and possible future directions. I am not going to examine the large and inconsistent literature that measures the effect of fatigue using psychometric tests because of psychometric tests unclear relationship with medical practice.

---

## Methods

Articles were included in the review if they examined the effect of call status or fatigue using work-related activities or board-type questions. Articles examining the effect of call status and fatigue on the mood and physical status of physicians were also included. Articles that used self-report of objective criteria (ie, reports where residents said they made an error due to fatigue) were not included. Self-report of emotional status was only used for prospective studies. Studies examining different types of call status (ie shift length, night float systems) were not reviewed. Articles were obtained by internet searches on Medline, citations from previous review articles on related subjects<sup>9-15</sup>, citations in the articles themselves, and citations of the articles themselves using the Scientific Citation Index.

---

## RESULTS

### *THE EFFECT OF CALL AND RESIDENT FATIGUE ON MOOD*

Virtually every study on acute resident fatigue has asked about the mood of the resident and has consistently found fatigue to have a negative effect on mood. The evidence on chronic fatigue implies that residents are immediately upset by the new conditions, but acclimate to them slightly with time. Social problems are common among residents, including divorce. However, studies of mood have suffered from infrequent use of control groups and over-reliance on retrospective designs, allowing legitimate questions about bias and uncertain implications of results. This section of the review is meant to be representative of the strongest literature in the field and does not intend to be a complete examination.

Studies of the effects of acute fatigue on mood have shown strong effects due to call in spite of what are often very small sample sizes. Lingenfelter et al<sup>16</sup> showed that 40 German residents had worse moods after a night on-call than a night off-duty. Leonard<sup>17</sup>, Engel<sup>18</sup> (with only seven participants), Browne<sup>19</sup>, Lewis<sup>20</sup> and others all found worsened mood after call than before it.

Studies of the effect of long-term fatigue on mood and personality have shown slightly more complicated results. One methodologically-strong study,<sup>21</sup> examined the effects of long-term fatigue on mood and personality by giving Profile of Mood States and Interpersonal Reactivity Instruments to 61 people throughout their internship year. The study found that over the course of the first five months the interns had increased scores on the depression-dejection, anger-hostility, fatigue-inertia scales, and personal

distress scales, along with decreased vigor-activity and IRI empathic concern measures. These changes “persisted throughout the internship period.” Some point out that this decrement in mood is compared to the already high-stress environment of medical school.<sup>9</sup>

Other studies also examined the effect of residency on fatigue. Reuben<sup>22</sup> found that emotional state worsened at the beginning of internship, but declined after the first 6 months. Gordon found consistent decrements of scores on the Profile of Mood States instrument.<sup>23</sup> Firth-Cozens<sup>24</sup> found that problems were stronger in women, people with poor diets, and high levels of self-criticism had higher rates of depression in residency. Hours were not associated with negative mood. Baldwin,<sup>25</sup> found that depression was more related to stress from job-related factors (number of emergency admissions, number of deaths on the ward) than to hours worked. Ford<sup>26</sup>, on the other hand, found that anger progressively increased during the internship year.

### *REAL-WORLD STUDIES*

While there is a small literature on the effects of changes in call structure on patient care, such as changes in shift-length and the institution of a night-float system, I could only find two studies examining the effects of resident hours on real-world physician actions.

One study by Haynes et al<sup>27</sup>, was a retrospective chart review of post-surgical complications. They found no change in complications for surgeries performed by post-call surgeons vs surgeons on their first day of work. This study has many methodological flaws, however. It is unclear what role others (attendings, nurses, etc) played in limiting

the effects of fatigued residents' care, it did not measure near misses, and neglects the possibility of a selection bias—that more complicated surgeries will be given to residents who are less fatigued and have more time left in their shift.

The other real-world study of fatigue and medical care examined the clinical effects of the 1989 regulation that limited resident hours for New York State residents.<sup>28</sup> The study used administrative data and chart review to examine a single general medical service of the 1989 New York State regulation limiting the hours that residents can work. After a fairly thorough demonstration of unchanged case-mix, the study found that *more* patient suffered at least one medical complications and (35% vs 22%) and a diagnostic test delay of at least 24 hours (17% vs 2%). They interpreted this as possible due to worsened continuity of care or decreased physician-staff hours. While an excellent study, it has numerous limitations, mostly because a non-reductionist study such as this makes it impossible to determine which aspect of the change caused this decreased care. The effects of “physician hand-offs,” which shorter call nights will necessarily create, are understudied but may be large.<sup>29</sup> Similarly, the proper arrangement of physicians and physician extenders to replace the residents is extremely difficult to determine. Research focus could go towards limiting the effects of these flaws, potentially ameliorating the negative effects of shorter resident hours. It is ultimately a fine line between this study of the effects of call-related fatigue and the effect of any other change in call structure that a hospital undergoes. While I am not reviewing the complete effects of call-structure changes, the results are not consistent or consistently negative. This is ultimately a study with  $n=1$ .

## *PERFORMANCE ON MEDICAL BOARDS AND BOARD-TYPE QUESTIONS*

A number of studies have examined the effects of call using medical questions similar to those on board exams, or scores on board exams themselves. I will divide these studies in three parts: scores on actual board exams, scores on board-type exams, and tests of ability to learn while fatigued. There is, of course, a large difference in incentive to perform well between actual board exams and studies that use board-type questions.

The results of the effects of fatigue on actual board exams is mixed. Stone et al<sup>30</sup> found no effect comparing prior night's call schedule to scores on 424 American Board of Surgery In-Training Examinations after adjusting for post-graduate year and training track. Jacques et al<sup>31</sup> found a strong association of prior night's sleep to score on the American Board of Family Practice in-training examination for 353 residents, by linear regression. Notice that in this study, fatigue is by participant choice, not call status. Another study found no effect of the NYS regulations changes on scores for 33 students from the year before the regulations went into effect and 21 from the year after on the Council on Resident Education in Obstetrics and Gynecology (CREOG) exam.<sup>32</sup>

Studies using board-type questions have also shown mixed effects. Lewis, Blagrove, and Ebdon<sup>20</sup> gave 36 junior medical staff-persons from South Wales 10 questions from medical students final exams and "published questions in the style of both parts of the membership examinations for the Royal College of Physicians" and found a relationship with sleep. The same group found no change in scores in 23 residents based on who had under 3.3 hours of continuous sleep and who had over 3.3 hours.<sup>33</sup> Storer et



al<sup>34</sup> gave 45 pediatrics residents, half of whom were randomly given a 24 hour call and the other half 34 hour call 30 questions “like those on the pediatrics board certification examination” before and after call and found no effects for either group. The 34-hour time gap is problematic, however because there is evidence that circadian effects throughout the day can have a large effect.

The only study that I found of the effect of fatigue on learning was performed by Browne et al,<sup>19</sup> who gave 35 medical students and 21 residents a half-hour each to read 6 journal articles over 6 days, half of which were post-call and then tested them on recall 1 week and 3 months later. They found no effect of call status on the ability to learn medical material.

**Table 1. Summary of studies on fatigue and board-type tests**

Author	Year	Length of Tests	N	Design	Results
Browne et al <sup>19</sup>	1994	30 min reading/day	35 students, 21 residents	Read 1 article per day, 3 per week, tested on it 7 days after reading.	Factual recall of articles no worse for ones read post-call
Jacques et al <sup>31</sup>	1990	4 hours	353	Linear regression of scores on Family Practice board exam vs sleep	"Loss of one night's sleep resulted in changes in test scores similar to the change that occurred in test scores between residents in the first and third year of training"
Blagrove et al <sup>33</sup>	2000	46 questions	23	43 questions of medical knowledge to interns	Same % correct, same confidence in answers
Lewis et al <sup>20</sup>	2002	45-90 minutes	11 house officers, 15 senior house officers	Board-type questions	Rested had more correct answers, but more inappropriate confidence in their answers
Stone et al <sup>30</sup>	2000	Board exam	424	Surgical in-training exam	70 post-call residents received similar scores to 324 non-post-call
Storer et al <sup>34</sup>	1989	unclear	45	Board-type questions to pediatrics residents	No effect

### *ASSESSMENTS OF MEDICAL SKILLS*

A number of studies have examined the effects of fatigue on skills that are necessary to medicine. The first, and one of the best, was published in 1971 by Friedman et al.<sup>35</sup> In this study, one of few to not use a crossover design, 14 participants (7 post-call, 7 not) were told to read an electrocardiographic strip with common arrhythmias for 20 minutes and bracket every arrhythmic event. In spite of the small sample size and lack of a crossover design, they found that fatigued residents performed significantly

worse (1 standard deviation in performance) on the test than non-fatigued. Bertram<sup>36</sup> found with two common diagnoses the notation in the medical record included less comprehensive physical and history findings for more hours worked and more patients seen. Christensen<sup>37</sup> found no effect of fatigue on 14 anesthesiology residents who read 25 artificial chest x-rays with 0-4 artificial primary nodules using a crossover design. Their study, however, was susceptible to circadian effects—the post-work testing occurred immediately after a 15 hour work day, the other group took the test before a work day. Denisco<sup>38</sup> et al gave 17 anesthesiology residents a simulated anesthesiology experience to test vigilance and found roughly a 0.8 standard deviation decline in ability to recognize problems. A decade earlier, in a fairly rudimentary experiment, Beatty et al<sup>39</sup> did not find a significant effect on a similar study of fifty-minute-long vigilance among six anesthesiology residents.

**Table 2. Summary of studies on fatigue and medical skills**

Author	Year	N	Design	Results
Beatty et al <sup>39</sup>	1977	6	Simulated anesthesia monitoring for 50 minutes	Minimal decrease in vigilance
Storer et al <sup>34</sup>	1989	45	Intubation and vein cannulation of sedated cats, catheterization of human umbilical artery	Catheterization took longer, other skills did not
Bertram <sup>36</sup>	1988	???	Retrospective chart review of 2 common diagnoses	Less comprehensive physical and history for those who had worked longer house
Christensen et al <sup>37</sup>	1977	14	25 x-rays with simulated nodules	No change with fatigue
Denisco et al <sup>38</sup>	1987	17	Simulated anesthesia monitoring	0.9 SD decline
Engel et al <sup>18</sup>	1987	7	Clinical exam on standardized patient actors in office setting, with predetermined criteria	No change with fatigue
Friedman et al <sup>35</sup>	1971	14	20 minute EKG strip reading test	1.5 SD increase in errors for fatigued
Grantcharov et al <sup>40</sup>	2001	14	Laparoscopic surgery simulator	Significant decline
Reznick et al <sup>41</sup>	1987	33	Reading lab charts, simulated wound suturing	No change with fatigue
Taffinder et al <sup>42</sup>	1998	6	Laparoscopic surgery simulator	"Surgeons awake all night made 20% more errors and took 14% longer to complete the tasks."
Poulton et al <sup>43</sup>	1978	30	Laboratory forms reading test	Reliable deterioration only with 8 hour sleep debt

Reznick and Folse<sup>41</sup> gave a variety of tests to 21 surgical residents, including ability to scan laboratory results with 18 error values over 48 pages of reports and the ability to suture a stimulated wound. The tests were given between 8 am and 6 pm (no indication if it was different between the groups) and twelve eligible participants (36% declined to participate). They found no effects.

Poulton et al<sup>43</sup> also had fatigued physicians read laboratory forms and found that it required a sleep debt of almost 8 hours before there was a noticeable decline in

performance. The poor organization of the paper makes it difficult to understand the exact design.

Engel<sup>18</sup> presented 7 interns with 4 standardized patient actors, complete with artificial charts and medical reports. Each intern saw 2 standardized patients on post-call clinic days, the other two on non-post-call clinic days. Evaluation was by a checklist of “specific items considered essential to an adequate evaluation.” The study found no effect, although the power and reliability of the measures was not determined. It was also not mentioned if the doctors were able to “guess” who the standardized patients were. This is a very intriguing study, but the importance of its null effect is impossible to determine without a great deal more knowledge about the details of the trial.

Storer et al<sup>34</sup>, along with the board-type questions mentioned above, had residents perform vein cannulation and endotracheal intubation on sedated cats, and umbilical artery catheterization on sections from recent deliveries. They found no difference in intubation or vein cannulation skills, but umbilical artery catheterization took longer

Grantcharov et al<sup>40</sup> used a laparoscopic surgical simulator on 14 Danish residents in gastroenterological surgery. They found significant effects with a few uncommon potential biases: severe circadian effects (post-call tests were at 9:30 am, when I can barely tie my shoelaces, and non-call tests were during working hours) and limited experience for all residents, who were performing a simulation of a surgery over half of them had never performed in reality.

Taffinder et al<sup>42</sup> gave 6 British surgical residents multiple performances on laparoscopic simulators, including before and after undisturbed nights of sleep, sham sleep (disturbed at midnight, 3 am, and 6 am) and no sleep. Surgeons who were awake

all night made more errors and took longer to complete the tasks than those who had a full night's sleep. The results were insignificant as stress was considered as a covariate. This is an interesting indication that stress may be a more important cause of decreased quality of care than fatigue, but it should be emphasized that stress is not an appropriate confounder of the relationship between fatigue and decreased laparoscopic skill because it is clearly in the causal pathway.

---

## Discussion

The overall inconsistency of the results is surprising. A meta-analysis of fatigue on humans' abilities that used primarily psychometric measures determined that partial sleep deprivation led to a 1.33 standard deviation decrease on individual's physical abilities and a 2.88 standard deviation effect on intellectual abilities.<sup>44</sup> These studies that I examined were not nearly so consistent. Of the six studies that looked at the effect of fatigue on board-type questions, only 2 had significant effects. Of the two that used actual board scores, the one with a stronger design<sup>30</sup> found an effect, the other study was biased by the fact that the fatigue was by the student's choice, not the effect of call status. The only study that looked at the effect of fatigue on resident's ability to learn found no effect.<sup>19</sup>

The studies of the effect of fatigue on medical simulators was similarly inconsistent. Of the four studies that examined physical ability,<sup>34, 40-42</sup>, all but one<sup>41</sup> found an effect. Of the 8 studies that examined mental issues or vigilance, only half found any effect.

However, these studies all have severe flaws and limitations. In an attempt to clarify what is left to learn and propose future directions for research, the remainder of the paper will be devoted to examining these limitations and proposing future research that could help more fully answer the question of the effect of fatigue on medical care.

### *Limitations of the Research*

The residency is a strenuous time in a physician's life. While adults, medical residents are not yet independent in their field, but are often asked to act as if they are. While knowledgeable about medicine, the resident often spends much of his or her time performing work that is not specifically medical in nature. Residents are given large expectations with limited time and limited support. Therefore, any study that examines resident fatigue can easily get lost in the many other factors complicating a resident's life.

Furthermore, the role of a resident is uncertain and varied. The amount of work performed by residents varies dramatically between specialties, locations, and individuals. A fifth-year surgical resident at an over-burdened public hospital has a very different job than a radiology intern in a small private hospital. Similarly, the amount of stress associated with the job varies dramatically. It is virtually impossible to isolate these different factors in single studies. These are some of the reasons why there is such a long history of small, methodologically-flawed studies.

Sample sizes in these studies were often very small, amount of sleep received was often unclear, and outcome measures were of questionable relevance to medical practice. Effect sizes were usually provided in nearly indecipherable terms (ie number of seconds to perform a task on a laparoscopic simulator test of unknown validity). Major limitations and relevant factors of the study designs were regularly not reported. Among these, the length of the tests (vital since vigilance seems to decline faster than cognitive processing), if participants received payment, what percentage of eligible participants were involved in the study, and whether the studies were controlled for time of day were rarely reported.



As I see it, the primary limitations in the majority of the research are effort bias (the possibility of an effect due to differential effort, not differential capacity), failure to account for stress as differed from fatigue, failure to account for local differences between individual residency programs and hospitals, and participation bias. I will examine each of these in more detail.

### *EFFORT BIAS*

The biggest potential bias in most of these studies is due to different levels of effort between fatigued and non-fatigued participants. In medicine, cutting corners can be a very dangerous thing. If tired residents are capable of effective performance, but do not expend their full effort, the result could be just as bad as if they were incapable of effective actions. Most studies ignore this problem. Those that accounted for it used different techniques and found unclear results.

Some studies attempted to account for effort bias by utilizing incentives for performance. The most well-known recent study,<sup>45</sup> provided a payment (\$50) for participation, an additional award (worth \$200) for the best performance), and a “negative incentive ... by informing the subjects that all scores subsequently would be publicized to their peers” and found null results.

Christensen<sup>37</sup> gave \$100 to the resident with the best performance and received negative results. Friedman’s<sup>35</sup> study also provided \$50 and \$25 to the groups with the fewest errors *when sleep-deprived*. This study did find positive effects in spite of the incentive.

Two other studies used residents' performances on real board-qualifying exams, an unmistakable self-incentive. One of these<sup>30</sup> found that the scores of 424 surgical residents taking the American Board of Surgery In-Training Examination did not differ by prior nights' call status once postgraduate year level and training track were adjusted for. The other<sup>31</sup> used a linear regression of the amount that 353 family practice residents slept the night before the American Board of Family Practice in-training examination. They found that "loss of one night's sleep resulted in changes in test scores that were approximately equivalent in magnitude to the change that occurred in test scores between residents in the first and third year of training." There is a potential bias created from the fact that call was not necessarily the reason people were sleeping less: perhaps many people received less sleep because they were studying, had personal problems, were ill, or had other problems that would have resulted in lower test scores.

The problem with all of these methods is that while money, board scores, and public shaming are incentives, they are *different* incentives than the needs of a sick patient. A board qualifying exam forces a person to concentrate in a way that a comatose patient may not, but it cannot create the immediate physical stimulus that a patient with a myocardial infarction will instill in any resident. Perhaps greater, perhaps lesser, but different. As long as this study flaw continues, it is impossible to know if the residents would have expended more or less effort had their been a patient, rather than an exam.

I could only find two studies that did not risk an effort bias. Engel et al did obscure the reason for study<sup>18</sup> and scored the performance of residents in the practice of medicine using standardized patient actors whom the residents thought were real patients. He does not clarify if the blinding was successful. While limited by a small sample size

and a non-thorough study design (too little training for patient actors, unclear scoring techniques, etc) I believe in a more thorough examination this would be an excellent methodology. Bertram's<sup>36</sup> retrospective chart review also looked at real actions with real patients and doctors who were not being observed at the time, but the end-point of a less thorough chart is probably of more interest to lawyers than physicians, because medical charts have been proven to be poor indicators of patient care<sup>46</sup> and the lack of case-mix adjustment calls the results into question.

### *Stress*

Few studies have effectively isolated the effect of resident stress from fatigue. The number of hours residents work has been shrinking, but improved technology, shorter lengths of stay, and increased severity of illness have all made hospital life faster and more stressful. Furthermore, as more women enter medicine, men take on more substantial roles in their family lives, and more nontraditional students enter medicine, the implication of working long hours changes. Many researchers consider stress the primary sin of residency, not fatigue.<sup>47</sup> If this is the case, lowering hours will have an effect, but other interventions may be more appropriate such as supplying allied healthcare help, nursing, day-care, easier access to attendings. Grantcharov<sup>40</sup> did find that the effect of fatigue disappeared when stress was taken into account as a covariate.

### *LOCAL DIFFERENCES*

Very few studies have examined the effect of specific, local differences on causing the fatigue and stress of medical residencies. Do residents of smaller hospitals

have much less stress? Do less busy hospitals create less stress? Do students of different residency programs feel the effect of fatigue differently? What about residents in different areas of the country?

Similarly, how do individual differences between residents affect their response to call status? Tanz et al<sup>48</sup> found that some residents had reputations for having harder call nights. In spite of having the same number of admissions, patients, and deaths, these residents received less sleep and perceived more stress than other residents. This is almost certainly due to how efficiently the resident works and perceives the effects of the call schedule on their life. A large interpersonal variation could explain the effects of many of these studies, however, which almost all have small sample sizes.

### *PARTICIPATION BIAS*

I've had troubles getting people excited about staying an extra hour after work to sit in front of a computer, so I have troubles believing other people haven't, also. In spite of this problem, very few studies have reported participation rate, those that did had extremely good participation rates (except for some survey studies, eg Reznick<sup>41</sup>), and even fewer said how they obtained such excitement. The possibility of a severe participation bias in the literature is very real and impossible to determine.

### POSSIBLE FUTURE STUDIES

#### *Hand-offs*

As mentioned above, the fear of the effects of shortening resident working hours on patient continuity seems quite reasonable.<sup>29</sup> One aspect of this is to perform further research on ways to limit the ill effects of continuity of care. There is very little research

on ways to improve physician hand-offs to minimize their ill effects. Possible studies could involve requiring incoming physicians to fill out check-lists on all of their patients or see the effect of improving the organization of the medical record.

### *CHRONIC FATIGUE*

None of the studies that I have seen examining medical performance have effectively looked at the effect of chronic fatigue on quality of care. The studies of mood and affect all found that changes were not only acute (eg post-call) but also long-term. Most residents will say that the first call is tiring, but not dangerous. It's only over time that the long hours affect people.

Most of the studies of the effects of fatigue in surgeons compared a morning post-call to a morning after a non-call night for residents who have every-other-night call. This is not a rested state. In fact, one study<sup>49</sup> found that after a single night of call scores on the Epworth Sleepiness scale and fatigue on the Profile of Mood States (POMS) scale fell for 3 consecutive days, while POMS confusion and vigor scores fell for two, making any test of acute fatigue during every-other-night call seem completely unreasonable. Another study using a laboratory test of fatigue found that even in their "rested" state, anesthesia residents were as tired as patients with narcolepsy.<sup>50</sup>

I don't believe any study of studies reported how long into their residency the study occurred, making examining how chronically fatigued the intern is impossible.

The effect of chronic fatigue on patient care could be studied relatively easily. Most residents get four weeks of vacation a year. A study could use one of the designs

seen above and, using the same matched crossover design, have them perform the study before and after a vacation.

### *Patient Care*

While it is clear that fatigued residents are unhappier and more anxious than non-fatigued ones, what has not been studied is how residents pass these moods off onto patients and the medical students who they are training. The nature of call makes this very difficult to study—since new patients are admitted at the beginning of call and discharged at the end of it, there is no control for how residents treat a patient when they first meet him or her when fatigued. The closest study that I've found to examining this issue is the previously-mentioned study by Bellini, which showed that physicians performed poorer on a test of empathy six months into their internship than before it.<sup>21</sup>

### *The Current Regulations Change*

The Accreditation Council on Graduate Medical Education's recent decree that no residents may work more than 320 hours per four weeks, in conjunction with the medical community's new interest in patient safety and medical errors and rapidly-improving clinical information systems, provides an excellent opportunity to study the effects of changes in resident call schedules. A few obvious studies include repeating Laine's<sup>28</sup> study under other specialties and locations. Since different hospitals and departments will all come up with different ways to address the issue, many different comparative studies are possible.

---

## Conclusion

In many ways, not only has the topic of residency hours not gone away-- the discussion itself has hardly advanced. Thirty years of studies has found surprisingly few breakthroughs about the effects of fatigue on medical practice and researchers (unfortunately, myself included) have learned surprisingly few lessons from prior studies with which to improve their work.

Some overall trends do appear:

- fatigue has very negative effects on physician mood and affect
- until better studies are designed that recognize variation between people, hospitals and specialties, without major biases and with strong outcomes measures, the exact extent of the damage to care by fatigued residents will not be known
- effort can overcome fatigue for short time periods
- the interaction of fatigue and stress is vital, but understanding is limited

While it is clear that policy decisions will have to be made on limited knowledge, it is equally clear that we will never know the exact effects of our actions until we fill in the gaps in our knowledge.

# Clinical Vignettes to Examine the Effect of Fatigue on Quality of Care

---

## INTRODUCTION

The exact effects that overnight call and long work hours for medical residents has on the quality of the medical care they provide has been a source of concern for over thirty years<sup>35</sup>. This concern took a major shift in 2002, when the Accreditation Council for Graduate Publication published controversial new guidelines on resident hours, limiting the amount of time that residents may work to 80 per week and limiting the length of call to 24 hours consecutively<sup>51-54</sup>.

In spite of the many studies and several review articles, questions still remain about the effect of fatigue on medical care and the effect that the new work hours regulations will have on patient care.<sup>14, 15, 55</sup> Most studies have used either psychometric measures with unclear relevance to medical practice or simulators that are relevant only to a specific specialty<sup>37, 38, 40, 42</sup> and results have been so inconsistent that even structured review articles disagree on the overall trends<sup>9-11, 13, 14, 44, 56</sup>. The primary conclusion is that the effect of fatigue is strongest on an individual's mood, less strong on his or her vigilance and physical dexterity, with mental abilities being the quality least susceptible to acute fatigue<sup>44</sup>.

One piece of evidence that is lacking in the research are examinations of the patient encounters that constitute the daily practice of the resident physician. The goal of this study was to look at the effect of a single night of call on the process of the medical encounter. A secondary goal was to see if fatigue had different effects on the component



parts of the medical encounter: history-taking, physical exam, diagnostic tests, diagnosis, and plan.

Computerized clinical vignettes or case simulations are a relatively inexpensive method of measuring a doctor's performance in the patient encounter<sup>46, 57, 58</sup>. These vignettes control for case mix and utilize an explicit scoring system based on national practice guidelines. Physicians' performances on the vignettes have been well-validated as very similar to a doctor's performance with standardized patient actors. They also enable a patient encounter to be isolated into its individual domains of care.

---

## METHODS

The study used a crossover design measuring performance on computerized clinical vignettes comparing residents after a full day of work (baseline) and a full day of work after a night on call (post-call).

### PARTICIPANTS

The study was conducted at an internal medicine unit at a teaching hospital in California. All second and third year residents who rotated through the inpatient ward service between November 2002 and January 2003 were eligible to participate. Each participant was provided a \$50 gift certificate in remuneration.

19 physicians were invited to participate, of which 18 agreed and one more was lost to follow-up. There were 5 participants in November, six in December, 6 in January, and 1 who began in November and ended in December. Eight second-year and nine third-year residents participated. There were 9 women and 8 men in the study. Participants ranged in age from 27 to 35, with six 30 or over at the time of participation. 10 participants performed the baseline sitting first, 7 performed the post-call sitting first.

### INTERVENTIONS

Every participant completed the clinical vignettes at the end of a full work-day—once after a night on call, the other time after a regular work-day. Participants were assigned to perform the post-call or baseline sitting first by call status on the day of the researcher's availability, with an attempt made to match first performances with equal

numbers of baseline and post-call participants. The cases that they saw were randomized using a random number generator.

## MEASUREMENT

We assessed quality of care using computerized clinical vignettes. Performance on these vignettes has been validated using standardized patients, and strongly approximates the quality of care in actual patient situations<sup>58</sup>. There were eight vignettes, each describing hypothetical patients with one of four common outpatient medical conditions: diabetes mellitus (DM), chronic obstructive pulmonary disease (COPD), vascular disease, and depression. Two clinical scenarios (cases) were developed for each condition, 1 simple and 1 complex. The complex cases were differentiated by a comorbid condition, the need for additional diagnostic tests, or indications for more involved treatment (Figure 1). Participants were assigned 1 case of each of the 4 medical conditions at each sitting. Stability of the results benefited from every participant having seen four vignettes of four medical conditions at each sitting.

Explicit scoring criteria for each case were developed based upon published national practice guidelines, such as the American Diabetes Association guidelines for diabetes mellitus. The scoring techniques were also reviewed by academic and local caregivers. A representative scoring sample of the uncomplicated case of diabetes mellitus is included (BOX 2).

## OBJECTIVES

We hypothesized that residents in internal medicine would perform worse on clinical vignettes after a full post-call work-day compared to a baseline condition of after a full, non-post-call work day.

## OUTCOMES

The primary outcome measure was the change in average score of four clinical vignettes between the post-call and baseline (non-post-call) setting. Secondary outcomes included performance on each of five clinical domains of medical care (history, physical exam, laboratory tests, diagnosis, and treatment and plan), scores on individual medical conditions, and scores of complex cases as opposed to simpler cases. We also assessed self-reported fatigue and self-reported mood on 5-point Likert scales and self-reported hours of sleep the two nights prior to performing the vignettes.

## SAMPLE SIZE

Sample size was calculated for a power of 0.8 and a correlation between non-call and post-call of 0.7 to find a 0.5 SD change in quality of care, to an intended sample size of 38.

## BLINDING

The scorer was blinded to call status and participant identification. Analysis was blind to participant identification. Participants were aware of the study design.

## STATISTICAL METHODS

All comparisons between post-call and non-call participants used paired t-tests. The primary and secondary outcome measures used one-tailed tests, all background analysis (comparing amount of sleep, reported fatigue, etc) were two-tailed. All other analyses are labeled in the text.

### Figure 1: clinical scenarios portrayed by standardized patients

- COPD with a mild exacerbation and history of hypertension
- COPD with an exacerbation associated with productive sputum, slight fever, and past history of hypertension
- Type 2 diabetes with limited preventive care in the past and untreated hypercholesterolemia
- Poorly controlled type 2 diabetes with early renal damage
- Congestive heart failure secondary to long standing hypertension and non-compliance with treatment
- New onset amaurosis fugax in patient with multiple risk factors
- Depression in an older patient with no other major clinical illness
- Depression complicated by substance abuse

**Table 3: Sampled scoring categories for uncomplicated diabetes mellitus**

<b>Domain</b>	<b>Necessary</b>	<b>Unnecessary/ inappropriate</b>
History: ask patient about	Duration of DM	
	Current medications	
	Associated symptoms	
	Complications: neurologic and peripheral organs	
	Comorbidities: other illnesses, hypertension, cholesterol	
	Risk Factors	
Physical Examination	Blood Pressure	
	Evaluate for signs of CHF	
	Foot sensory and vascular exam	
	DRE (by age)	
Tests Ordered	Chemistry 7	Ultrasound
	CBC	CT or MRI
	Hepatic panel	ETT
	Glucose	Catheterization
	Lipid panel	X-rays
	HgbA1c	
	Urine protein microalb	
Diagnosis	Diabetes Mellitus	
	Hypercholesterolemia	
Treatment		
Prescription	Diabetes medications	
	Cholesterol medications	
Management	Smoking cessation	Cardiology referrals
	Follow-up appointment	Other diagnosis
	Diet	Other treatment
	Exercise	
	Ophthalmology referral	
	Podiatry	
	Diabetic Nurse Educator	

---

## RESULTS

### BASELINE INFORMATION

Post-call residents reported significantly less sleep the previous night than non-call (3.8 $\pm$ 1.6 hrs SD vs 7.3 $\pm$ 0.37 hrs SD,  $p < 0.001$ ) and more subjective fatigue 2.6 $\pm$ 0.87 vs 3.5 $\pm$ 1  $p < 0.01$  on a 5-pt scale). They expressed the same subjective mood (3.94 $\pm$ 0.23 SD vs 3.71  $\pm$  0.21,  $p = 0.22$  on a 5-pt scale).

There did not appear to be any learning effect – residents' performance on the vignettes did not differ between their first and second sitting ( $p = 0.18$  two-tailed t-test with equal variance). The residents also did not have different performance between the first two cases of the day than the second two (66.7  $\pm$  7 for first 2 vignettes of a day vs 68.9  $\pm$  7 for second 2; 2-sided  $p = 0.18$ ). There were also no circadian effects— the time when a participant began taking the vignettes did not have an effect upon their performance ( $p = 0.99$ , linear regression). There was no difference in score between the uncomplicated cases and the complex ones ( $p = 0.8$ , unpaired two-tailed t-test with equal variance). There was no difference in scores between level of training (70.9  $\pm$  6.7 for PGY2, 65.0  $\pm$  2.2 for PGY3,  $p = 0.051$  t-test with equal variance). There was no effect of gender on scores ( $p = 0.57$ , unpaired t-test with equal variance).

### OUTCOMES

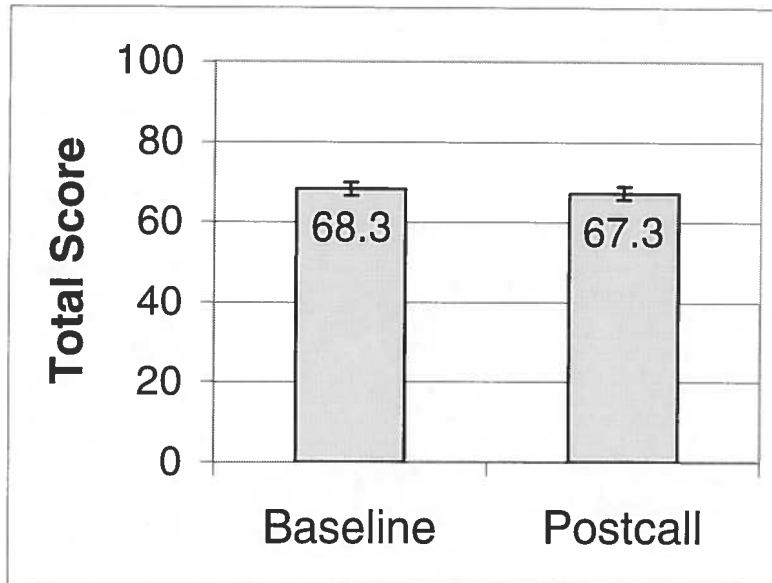
Call status did not predict overall performance on the vignettes (FIGURE 2) In sub-analyses, we found no effect after stratifying performance by 5 discrete skill domains of history, physical exam, test request, diagnosis, or treatment (TABLE 4). We also found no effect of call status on the performance in any specific condition (TABLE 5).

Call status did not affect performance on the vignettes in complex or uncomplicated ones ( $p=0.38$  and  $p=0.73$ , two-sided, unpaired analysis).

Post-hoc power calculations with the observed correlation between observations of 0.58 and a final sample size of 17 found a 0.89 power to detect a 1 SD change (7% score) and a 0.28 power to detect a 0.5 SD change (3.6%) in score.



**Figure 2: Effect of call status on vignette performance. Error bars are SE. 1-sided paired t-test  $p = 0.28$**



**Table 4: Effect of call status on performance of individual domains of medical care\***

	Baseline	Post-call	P Values
History	76.0 (9.4)	76.9 (8.8)	0.3
Physical Exam	57.9 (9.2)	58.6 (8.7)	0.38
Laboratory Tests	68.5 (19.8)	65.9 (15.8)	0.32
Diagnosis	60.4 (10.9)	59.5 (11.3)	0.37
Treatment	64.9 (11.0)	61.3 (11.6)	0.14

\* Data are presented as % (SD). P values are 1-sided. N=17

**Table 5: Effect of call status on performance of individual medical conditions\***

	<b>Baseline</b>	<b>Post-call</b>	<b>P Values</b>
COPD	67.0 (6.4)	66.7 (7.9)	0.43
Diabetes	66.6 (11.1)	66.8 (11.7)	0.47
Vascular disease	73.4 (10.2)	71.2 (7.3)	0.18
Depression	66.1 (13.5)	64.6 (9.2)	0.35

\* Data are presented as % (SD). P values are 1-sided. N=17

---

## COMMENT

This study may be one of the first to examine how fatigue affects the individual domains of medical practice and one of few studies to examine the effects of fatigue using methods that have been validated to approximate actual medical care. The study design limited interpersonal variation by using a crossover design and limited circadian effects by performing baseline and post-call vignettes at the end of their working day. The validity of the design was reinforced by the correlation between pre-call and post-call scores of  $r=0.57$ . The lack of circadian effects or learning effects were demonstrated by the scores being consistent by time of day, by first or second performance of the vignettes, and the stability of scores throughout an individual testing day. This also ensures that there are no biases due to having an unequal number of participants who took the vignettes post-call first rather than non-call.

The null results might be attributed to a combination of two factors: first, if a single night of call has any effect on the patient encounter, it appears to not be large and, second, that the study was underpowered. While our original power calculations had predicted a  $\frac{1}{2}$  SD change in performance, resource limitations and a somewhat smaller-than-expected correlation between performances by individuals led to a power that would only observe a 1 SD change. This would be a large deficit for an average difference of 3.5 hours of sleep, but one that has been found in other studies.

Participants' knowledge of the nature of the study could have biased their effort. Similarly, the decreased incentive for fatigued physicians to perform well on a computer program as opposed to seeing a living patient could also undermine the results of the study. The strongly null results, however, imply that this did not occur.

The primary limitation to generalizability was the differences in sleep that residents obtain in different academic hospitals. This was designed as a study of the acute effects of overnight call, not fatigue per se. Hours of sleep and difficulty of performance were not standardized, so differences between hospitals and programs could limit generalizability. There is no reason to believe fatigue would affect these residents dramatically differently than any others and most major medical specialties utilize the skills examined here.

Overall, the literature of the effect of acute fatigue on medical care utilizes widely varying techniques leading to fairly widely-varying results. The lack of an effect in this study is further indication that mental abilities such as those needed in the daily practice of medicine may be fairly resistant to the effects of acute fatigue.

---

## WORKS CITED

1. Wentz DK, Ford CV. A brief history of the internship. *Jama*. Dec 28 1984;252(24):3390-3394.
2. Daugherty S, Baldwin D, Jr., Rowley B. Learning, satisfaction, and mistreatment during medical internship: a national survey of working conditions. *JAMA*. April 15, 1998 1998;279:1194-1199.
3. McCall TB. The impact of long working hours on resident physicians. *N Engl J Med*. Mar 24 1988;318(12):775-778.
4. Wallack MK, Chao L. Resident work hours: the evolution of a revolution. *Arch Surg*. 2001;136(12):1426-1431; discussion 1432.
5. Asch DA, Parker RM. The Libby Zion case. One step forward or two steps backward? *N Engl J Med*. 1988;318(12):771-775.
6. Kohn LTC, Janet M.; Donaldson, Molla S., ed. *To Err Is Human: Building a Safer Health System*. Washington, DC: National Academy Press; 1999. Medicine 10, ed.
7. ACGME. *Report of the ACGME Work Group on Resident Duty Hours*: Accreditation Council for Graduate Medical Education; June 11, 2002 2002.
8. Tanner L. AMA endorses 80-hour work week for medical residents. *Washington Post*. June 20, 2002, 2002.
9. Asken MJ, Raham DC. Resident performance and sleep deprivation: a review. *J Med Educ*. 1983;58(5):382-388.
10. Samkoff JS, Jacques CH. A review of studies concerning effects of sleep deprivation and fatigue on residents' performance. *Acad Med*. Nov 1991;66(11):687-693.
11. Leung L, Becker CE. Sleep deprivation and house staff performance. Update 1984-1991. *J Occup Med*. 1992;34(12):1153-1160.
12. Jha AK, Duncan BW, Bates DW. Fatigue, Sleepiness, and Medical Errors. In: Shojania KG, Duncan BW, McDonald KM, Wachter RM, eds. *Making Health Care Safer: A Critical Analysis of Patient Safety Practices*. Washington, DC: Agency for Healthcare Research and Quality; 2001:519-533.
13. Owens JA. Sleep loss and fatigue in medical training. *Curr Opin Pulm Med*. 2001;7(6):411-418.
14. Veasey S, Rosen R, Barzansky B, Rosen I, Owens J. Sleep loss and fatigue in residency training: a reappraisal. *Jama*. Sep 4 2002;288(9):1116-1124.
15. Steinbrook R. The debate over residents' work hours. *N Engl J Med*. Oct 17 2003;347(16):1296-1302.
16. Lingenfelter T, Kaschel R, Weber A, Zaiser-Kaschel H, Jakober B, Kuper J. Young hospital doctors after night duty: their task-specific cognitive status and emotional condition. *Med Educ*. 1994;28(6):566-572.

17. Leonard C, Fanning N, Attwood J, Buckley M. The effect of fatigue, sleep deprivation and onerous working hours on the physical and mental wellbeing of pre-registration house officers. *Ir J Med Sci.* Jan-Mar 1998;167(1):22-25.
18. Engel W, Seime R, Powell V, D'Alessandri R. Clinical performance of interns after being on call. *South Med J.* 1987;80(6):761-763.
19. Browne BJ, Van Susteren T, Onsager DR, Simpson D, Salaymeh B, Condon RE. Influence of sleep deprivation on learning among surgical house staff and medical students. *Surgery.* May 1994;115(5):604-610.
20. Lewis KE, Blagrove M, Ebden P. Sleep deprivation and junior doctors' performance and confidence. *Postgrad Med J.* Feb 2002;78(916):85-87.
21. Bellini LM, Baime M, Shea JA. Variation of mood and empathy during internship. *Jama.* Jun 19 2002;287(23):3143-3146.
22. Reuben DB. Depressive symptoms in medical house officers. Effects of level of training and work rotation. *Arch Intern Med.* Feb 1985;145(2):286-288.
23. Gordon GH, Hubbell FA, Wyle FA, Charter RA. Stress during internship: a prospective study of mood states. *J Gen Intern Med.* July/August 1986;1(4):228-231.
24. Firth-Cozens J. Emotional distress in junior house officers. *Br Med J (Clin Res Ed).* Aug 29 1987;295(6597):533-536.
25. Baldwin PJ, Dodd M, Wrate RM. Young doctors' health--II. Health and health behaviour. *Soc Sci Med.* 1997;45(1):41-44.
26. Ford CV, Wentz DK. The internship year: a study of sleep, mood states, and psychophysiologic parameters. *South Med J.* Nov 1984;77(11):1435-1442.
27. Haynes DF, Schwedler M, Dyslin DC, Rice JC, Kerstein MD. Are postoperative complications related to resident sleep deprivation? *South Med J.* 1995;88(3):283-289.
28. Laine C, Goldman L, Soukup JR, Hayes JG. The impact of a regulation restricting medical house staff working hours on the quality of patient care. *Jama.* 1993;269(3):374-378.
29. Petersen LA, Brennan TA, O'Neil AC, Cook EF, Lee TH. Does housestaff discontinuity of care increase the risk for preventable adverse events? *Ann Intern Med.* 1994;121(11):866-872.
30. Stone MD, Doyle J, Bosch RJ, Bothe A, Jr., Steele G, Jr. Effect of resident call status on ABSITE performance. American Board of Surgery In-Training Examination. *Surgery.* 2000;128(3):465-471.
31. Jacques CH, Lynch JC, Samkoff JS. The effects of sleep loss on cognitive performance of resident physicians. *J Fam Pract.* Feb 1990;30(2):223-229.
32. Kelly A, Marks F, Westhoff C, Rosen M. The effect of the New York State restrictions on resident work hours. *Obstet Gynecol.* 1991;78(3 Pt 1):468-473.

33. Blagrove M, Lewis K, P. E. Sleep loss and the relationship between confidence and accuracy of hospital interns' answers to medical questions. *Sleep*. 2000;23(S2):A246.
34. Storer JS, Floyd HH, Gill WL, Giusti CW, Ginsberg H. Effects of sleep deprivation on cognitive ability and skills of pediatrics residents. *Acad Med*. Jan 1989;64(1):29-32.
35. Friedman RC, Bigger JT, Kornfeld DS. The intern and sleep loss. *N Engl J Med*. 1971;285(4):201-203.
36. Bertram DA. Characteristics of shifts and second-year resident performance in an emergency department. *N Y State J Med*. Jan 1988;88(1):10-14.
37. Christensen EE, Dietz GW, Murry RC, Moore JG. The effect of fatigue on resident performance. *Radiology*. Oct 1977;125(1):103-105.
38. Denisco RA, Drummond JN, Gravenstein JS. The effect of fatigue on the performance of a simulated anesthetic monitoring task. *J Clin Monit*. 1987;3(1):22-24.
39. Beatty J, Ahern S, K. K. Sleep deprivation and the vigilance of anesthesiologists during simulated surgery. In: Mackie R, ed. *Vigilance: theory, operational performance and physiological correlates*. New York: Plenum; 1977:511-527.
40. 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(7323):1222-1223.
41. Reznick RK, Folse JR. Effect of sleep deprivation on the performance of surgical residents. *Am J Surg*. Nov 1987;154(5):520-525.
42. Taffinder NJ, McManus IC, Gul Y, Russell RC, Darzi A. Effect of sleep deprivation on surgeons' dexterity on laparoscopy simulator. *Lancet*. 1998;352(9135):1191.
43. Poulton EC, Hunt GM, Carpenter A, Edwards RS. The performance of junior hospital doctors following reduced sleep and long hours of work. *Ergonomics*. Apr 1978;21(4):279-295.
44. Pilcher JJ, Huffcutt AI. Effects of sleep deprivation on performance: a meta-analysis. *Sleep*. May 1996;19(4):318-326.
45. Deaconson TF, O'Hair DP, Levy MF, Lee MB, Schueneman AL, Codon RE. Sleep deprivation and resident performance. *Jama*. 1988;260(12):1721-1727.
46. Luck J, Peabody JW, Dresselhaus TR, Lee M, Glassman P. How well does chart abstraction measure quality? A prospective comparison of standardized patients with the medical record. *Am J Med*. 2000;108(8):642-649.
47. Luck C. Reducing stress among junior doctors. *Bmj*. 2000;321(7268):S2-S7268.
48. Tanz RR, Charrow J. Black clouds. Work load, sleep, and resident reputation. *Am J Dis Child*. May 1993;147(5):579-584.
49. Rose MW, Ware JC, Kolm P, Risser MR. Residual effects of call on sleep and mood in medical residents. *Sleep*. 2000;23(S2):A253.

50. Howard SK, Gaba DM, Rosekind MR, Zarcone VP. The risks and implications of excessive daytime sleepiness in resident physicians. *Acad Med.* Oct 2002;77(10):1019-1025.
51. Reece JS, McKenzie JS. New ACGME requirements can be met. *Fam Med.* Oct 2001;33(9):651.
52. Srinivas SK, Rising JP. The ACGME is unable to address residents' overwork. *Acad Med.* May 2001;76(5):397-398.
53. ACGME clamps down on hours for residents. *Hosp Peer Rev.* Aug 2002;27(8):111-112.
54. Romano M. Hours of doctors-in-training: who's counting? ACGME's restrictions limiting residents' workhours denounced as inadequate. *Mod Healthc.* Aug 19 2002;32(33):18-19.
55. Weinstein DF. Duty Hours for Resident Physicians - Tough Choices for Teaching Hospitals. *N Engl J Med.* Oct 17 2002;347(16):1275-1278.
56. Weinger MB, Ancoli-Israel S. Sleep deprivation and clinical performance. *Jama.* Feb 27 2002;287(8):955-957.
57. Dresselhaus TR, Peabody JW, Lee M, Wang MM, Luck J. Measuring compliance with preventive care guidelines: standardized patients, clinical vignettes, and the medical record. *J Gen Intern Med.* 2000;15(11):782-788.
58. Peabody JW, Luck J, Glassman P, Dresselhaus TR, Lee M. Comparison of vignettes, standardized patients, and chart abstraction: a prospective validation study of 3 methods for measuring quality. *Jama.* 2000;283(13):1715-1722.