ORIGINAL ARTICLE

Effect of Reducing Interns' Work Hours on Serious Medical Errors in Intensive Care Units

Christopher P. Landrigan, M.D., M.P.H., Jeffrey M. Rothschild, M.D., M.P.H., John W. Cronin, M.D., Rainu Kaushal, M.D., M.P.H., Elisabeth Burdick, M.S., Joel T. Katz, M.D., Craig M. Lilly, M.D., Peter H. Stone, M.D., Steven W. Lockley, Ph.D., David W. Bates, M.D., and Charles A. Czeisler, Ph.D., M.D., for the Harvard Work Hours, Health and Safety Group

From the Divisions of Sleep Medicine (C.P.L., J.W.C., S.W.L., C.A.C.), General Internal Medicine (J.M.R., R.K., E.B., D.W.B.), Infectious Disease (J.T.K.), Pulmonary and Critical Care Medicine (J.W.C., C.M.L.), and Cardiology (P.H.S.) and the Internal Medicine Residency Program (J.T.K.), Department of Medicine, Brigham and Women's Hospital; the Division of Sleep Medicine, Harvard Medical School (C.P.L., J.W.C., S.W.L., C.A.C.); and the Division of General Pediatrics, Department of Medicine, Children's Hospital Boston and Harvard Medical School (C.P.L.) — all in Boston. Address reprint requests to Dr. Landrigan at the Division of Sleep Medicine, Department of Medicine, Brigham and Women's Hospital, 221 Longwood Ave., Boston, MA 02115, or at clandrigan@rics.bwh.harvard.edu.

N Engl J Med 2004;351:1838-48.
Copyright © 2004 Massachusetts Medical Society.

ABSTRACT

BACKGROUND
Although sleep deprivation has been shown to impair neurobehavioral performance, few studies have measured its effects on medical errors.

METHODS
We conducted a prospective, randomized study comparing the rates of serious medical errors made by interns while they were working according to a traditional schedule with extended (24 hours or more) work shifts every other shift (an "every third night" call schedule) and while they were working according to an intervention schedule that eliminated extended work shifts and reduced the number of hours worked per week. Incidents were identified by means of a multidisciplinary, four-pronged approach that included direct, continuous observation. Two physicians who were unaware of the interns' schedule assignments independently rated each incident.

RESULTS
During a total of 2203 patient-days involving 634 admissions, interns made 35.9 percent more serious medical errors during the traditional schedule than during the intervention schedule (136.0 vs. 100.1 per 1000 patient-days, $P<0.001$), including 56.6 percent more nonintercepted serious errors ($P<0.001$). The total rate of serious errors on the critical care units was 22.0 percent higher during the traditional schedule than during the intervention schedule (193.2 vs. 158.4 per 1000 patient-days, $P<0.001$). Interns made 20.8 percent more serious medication errors during the traditional schedule than during the intervention schedule (99.7 vs. 82.5 per 1000 patient-days, $P=0.03$). Interns also made 5.6 times as many serious diagnostic errors during the traditional schedule as during the intervention schedule (18.6 vs. 3.3 per 1000 patient-days, $P<0.001$).

CONCLUSIONS
Interns made substantially more serious medical errors when they worked frequent shifts of 24 hours or more than when they worked shorter shifts. Eliminating extended work shifts and reducing the number of hours interns work per week can reduce serious medical errors in the intensive care unit.

Downloaded from www.nejm.org on January 17, 2008 . For personal use only. No other uses without permission.
Copyright © 2004 Massachusetts Medical Society. All rights reserved.

In a pioneering study published in the *Journal* 33 years ago, Friedman and colleagues reported that interns made almost twice as many errors reading electrocardiograms after an extended (24 hours or more) work shift than after a night of sleep.[1] More recent studies have similarly found that surgical residents made up to twice the number of technical errors in the performance of simulated laparoscopic surgical skills after working overnight than after a night of sleep.[2,3] Although many prior studies have been methodologically limited by the use of nonvalidated self-reports on the timing of sleep and inadequate accounting for circadian phase and chronic sleep loss, as reviewed elsewhere,[4-6] the literature as a whole suggests that sleep deprivation causes substantial decrements in physicians' performance of discrete neurocognitive and simulated clinical tasks.[4-8] The clinical importance of sleep curtailment has remained unclear, however,[4-6] owing to a lack of studies conducted in clinical care environments[4,9] and the possibility that scheduling interventions designed to mitigate sleep deprivation may simultaneously introduce discontinuities in care.[10,11]

Within hospitals, of all trainees, interns (postgraduate year 1) typically work the greatest number of hours per week.[12,13] The extended (24 hours or more) work shifts and long workweeks of interns may make them especially prone to fatigue-induced errors. In a survey of house officers, 41 percent reported fatigue as a cause of their most serious mistake. Most of these events occurred while they were interns, and 31 percent reportedly resulted in fatalities.[14]

To understand the effects of interns' sleep deprivation on serious medical errors, we conducted a comprehensive comparison of errors while interns followed a traditional work schedule and errors while they followed an intervention work schedule that was designed to reduce sleep deprivation. Our goals were to compare the rates of serious errors directly involving interns on the two schedules, since interns were the focus of our scheduling intervention, and to compare the overall rates of serious medical errors in order to track the effects of interns' schedules on the system as a whole.

METHODS

The Intern Sleep and Patient Safety Study was conducted as part of the Harvard Work Hours, Health and Safety Study from July 2002 to June 2003 in the medical intensive care unit (MICU) and coronary care unit (CCU) of Brigham and Women's Hospital, a large academic hospital in Boston, after approval by the institutional review board. The MICU and CCU were selected for study because they are the rotations of this internal-medicine training program with the longest work hours and because medical errors have been detected at higher rates in critical care settings than in other settings.[15,16] Both units have 10 adult critical care beds. Data were not collected on patients admitted for fewer than four hours, patients undergoing elective allergy desensitization, or the rare patients who boarded on the units but who were not cared for by the MICU or CCU team.

DESIGN OF INTERVENTION TRIAL

In collaboration with the leadership of the residency program and unit directors, we designed an intervention work schedule for interns that eliminated extended (24 hours or more) work shifts and reduced the number of scheduled hours of work to 63 per week (Fig. 1). The traditional MICU housestaff team consisted of three interns and three third-year residents, whereas the CCU team consisted of three interns and two second-year residents. Each intern and resident on these teams worked overnight in the hospital every third night. A resident from another hospital service assumed patient care responsibilities in the CCU on nights when neither of the daytime CCU residents was working. Under this rotation, interns' scheduled workweeks averaged 77 to 81 hours, depending on the clinic assignment, with up to 34 continuous hours of scheduled work when clinic occurred after they were on call (Fig. 1A).

During the intervention schedule, interns' work hours and overnight work schedules were changed. Interns' traditional extended work shifts were divided in two: a "day-call" intern worked the first half of a traditional call (from 7 a.m. to 10 p.m.); a "night-call" intern worked the second half (from 9 p.m. to 1 p.m. the following day). To effect this schedule, four interns shared patient care responsibilities during the rotation. The maximal scheduled hours of work were 60 to 63 per week, with consecutive hours of work limited to approximately 16 hours (Fig. 1B). The intervention did not alter the schedules or staffing of second- or third-year residents or other clinical personnel.

Our goal was to improve interns' opportunities

Downloaded from www.nejm.org on January 17, 2008 . For personal use only. No other uses without permission.
Copyright © 2004 Massachusetts Medical Society. All rights reserved.

to sleep while minimizing errors due to handoffs of patient care and cross-coverage.[10] To minimize cross-coverage errors, we developed a sign-out template for interns to use in all critical care rotations (both intervention and traditional schedules)[17] and incorporated an hour of overlap in the evenings for interns on the intervention schedule to sign out formally (see Figure A of the Supplementary Appendix, available with the full text of this article at www.nejm.org) under the supervision of the senior resident.

After providing written informed consent, interns were randomly assigned to work either the intervention schedule in the CCU and the traditional schedule in the MICU or the converse; these rotations were distributed throughout the year. Data collected during a pilot intervention schedule involving four interns that was discontinued after the first ICU rotation were not included. As detailed in the article by Lockley et al. in this issue of the *Journal*,[18] although actual work hours often exceeded those scheduled during both the tradi-

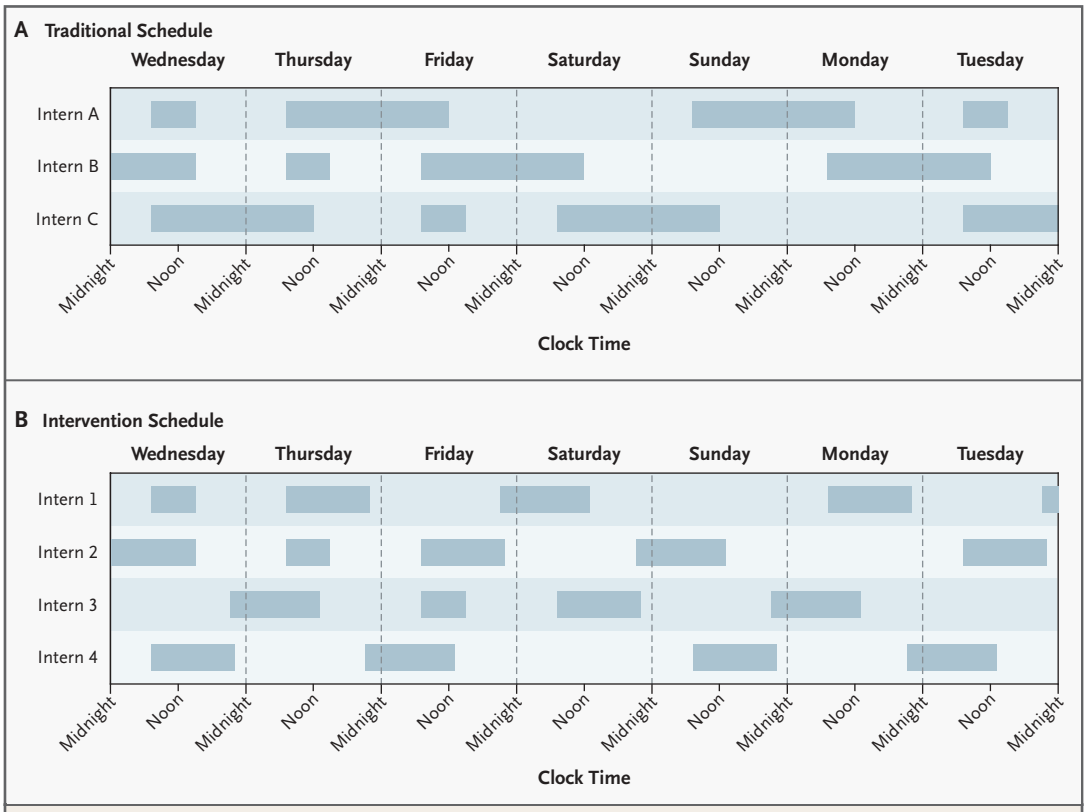

Figure 1. Representative Work Hours during a Single Week for the Whole Team of Interns during the Traditional Schedule (Panel A) and the Intervention Schedule (Panel B).

Scheduled work hours are indicated by the bars. Panel A shows the traditional rotation in which a team of three interns provided continuous coverage on a repeated three-day schedule, consisting of a daytime "swing" shift on day 1 (7 a.m. to 3 p.m.) (e.g., Wednesday for Intern A), followed by an extended on-call shift from 7 a.m. on day 2 to noon on day 3 (e.g., Thursday through Friday for Intern A). Interns had the day off when a swing shift was to occur on a Saturday, Sunday, or Monday (e.g., Saturday for Intern A). Interns staffed weekly ambulatory clinics when they coincided with a swing shift or the latter half of an extended on-call shift. Panel B shows the intervention rotation in which a team of four interns provided continuous coverage on a repeated four-day schedule. Day 1 is the standard swing shift (e.g., Wednesday for Intern 1); day 2 is "day call" from 7 a.m. to 10 p.m. (representing the first half of the traditional call) (e.g., Thursday for Intern 1); days 3 and 4 are "night call" from 9 p.m. on day 3 to 1 p.m. on day 4 (representing the second half of a traditional call) (e.g., Friday through Saturday for Intern 1). There is a one-hour scheduled overlap between the outgoing day-call intern and the incoming night-call intern (e.g., Wednesday from 9 p.m. to 10 p.m. for Intern 4 and Intern 3, respectively); this overlap was often extended as clinically required. Interns had the day off when a swing shift was to occur on a Saturday, Sunday, or Monday (e.g., Sunday for Intern 1). Interns only attended clinics when they coincided with the swing day.

Downloaded from www.nejm.org on January 17, 2008 . For personal use only. No other uses without permission.
Copyright © 2004 Massachusetts Medical Society. All rights reserved.

tional and intervention schedules, the intervention successfully eliminated shifts of 24 hours or more, reduced the number of hours worked by interns by nearly 20 per week, increased the average daily duration of sleep by nearly an hour, and reduced attentional failures.

DATA COLLECTION AND CLASSIFICATION

To measure patients' safety during the two schedules, we developed an intensive system of data collection and evaluation that expanded on methods previously used in the study of medication errors[16,19] and also included continuous observation of interns by physicians. In this study, we focused on procedural and diagnostic errors in addition to medication errors. The definitions used to classify incidents are provided in Table 1.

A team of two nurse chart reviewers and six physician observers collected data, supplemented by voluntary reports from clinical staff and a computerized event-detection monitor. Direct observation was the principal means of detecting serious errors in which interns were directly involved; physician observers followed study interns continuously, day and night in the hospital. In the afternoons after work rounds, when more than one intern was working simultaneously, only one intern was observed at a time owing to staffing limitations. Residents and other personnel on the units were not directly observed. Data collection for personnel other than interns was less comprehensive and relied on chart review, voluntary reports, and computerized event-detection monitoring. Other methods of data collection, though less comprehensive, were designed to identify all serious medical errors — both those in which interns were involved and those in which they were not involved. Before beginning data collection, all staff received intensive training in the consistent, objective collection of data using standardized forms. Because it was not possible to blind data collectors to the study schedule, determinations of the preventability and classification of events were not made by the primary data collectors. Instead, each suspected error or adverse event identified was independently rated by two physician investigators who were unaware of the identity of those involved or whether the incident occurred during the traditional or intervention schedule.

In the vast majority of cases, the serious errors identified by observers were promptly addressed by medical staff with no need for action on the part of the observers. Nonintercepted serious errors were generally detected by observers when they were discussed by clinical staff. In the handful of cases in which observers identified possible errors in the making with substantial potential to cause harm, they immediately alerted clinical staff to prevent harm to the patient.

Blinded reviewers categorized each incident as an adverse event, nonintercepted serious error, intercepted serious error, or error with little potential for harm (a category that was excluded from the

Table 1. Definitions Used in the Study.

Term	Definition
Medical error	Any error in the delivery of medical care, whether harmful or trivial
Serious medical error	A medical error that causes harm or has substantial potential to cause harm, including preventable adverse events, nonintercepted serious errors, and intercepted serious errors, but not including errors with little or no potential for harm or unpreventable adverse events
Intercepted serious error	A serious medical error that is intercepted before reaching the patient
Nonintercepted serious error	A serious medical error that is not intercepted and therefore reaches the patient but causes no clinically detectable harm
Adverse event	Any injury due to medical management
Nonpreventable adverse event	Unavoidable injury resulting from appropriate medical care
Preventable adverse event	Injury due to a nonintercepted serious error in medical management
Serious medication error	A serious medical error related to the ordering or administration of pharmaceutical agents, blood products, or intravenous fluids
Serious procedural error	A serious medical error related to the performance of an invasive procedure, such as placement of a central venous or arterial catheter
Serious diagnostic error	A serious medical error related to history taking, the performance of a physical examination, or the ordering or interpretation of a diagnostic test

Downloaded from www.nejm.org on January 17, 2008 . For personal use only. No other uses without permission.
Copyright © 2004 Massachusetts Medical Society. All rights reserved.

analysis) and rated the preventability of adverse events using a Likert scale (was prevented, was definitely preventable, was probably preventable, was probably not preventable, or was definitely not preventable); the preventability scale was dichotomized to include only "preventable events" and "nonpreventable events" before analysis. Events deemed more likely to be due to patients' underlying illness than to medical therapy were excluded. Disagreements were resolved by discussion; the interrater reliability was calculated before such discussion by means of the kappa statistic, as described below.

STATISTICAL ANALYSIS

Patients' characteristics and the mean daily census of the units during the intervention and traditional rotations were compared by means of Fisher's exact test; Wilcoxon's nonparametric test for dichotomous, nonnormally distributed continuous variables; and a t-test for normally distributed continuous variables. All statistical tests were two-tailed. The rates of diagnostic tests and procedures per patient-day were compared between the two schedules, and the distribution was assumed to be binomial. We compared the rates of medication orders per patient-day between the two schedules, assuming a Poisson distribution, since the presence of rates of more than one order per patient-day precluded the use of the binomial distribution.

We compared the rates of intern-associated serious medical errors per patient-day (for all interns combined) and of total serious medical errors per patient-day between the intervention and traditional schedules, assuming a binomial distribution. The rates of all serious medical errors include all intern-associated serious errors (those detected by direct observation and other methods) plus non–intern-associated errors (identified by chart review, staff reports, and the computerized monitor). We also compared the rates of type-specific errors (medication, procedural, and diagnostic) per patient-day, assuming a binomial distribution. For all tests, two-tailed P values of less than 0.05 were considered to indicate statistical significance.

The study was powered to determine differences in rates of serious medical errors. Analyses of the rates of adverse events were performed, but the results were considered exploratory since we had only 11 percent power to detect a 25 percent difference in intern-associated preventable adverse events. By contrast, the study was designed to have 80 percent power to detect a 16 percent difference in the rate of serious errors between groups.

We evaluated the reliability of the primary data-collection process by conducting dual direct observation for a total of 10 patient-days; there was 82 percent agreement between independent observers with respect to the occurrence of a serious medical error. At the review stage conducted by the blinded investigators, we performed comprehensive reliability testing of all incidents rated using the kappa statistic. For reviewers' judgments about whether an incident was an adverse event, an intercepted serious error, a nonintercepted serious error, or an excluded event, the κ was 0.90; the κ was 0.80 for the preventability of adverse events.

RESULTS

PATIENT POPULATION

The study involved 2203 patient-days (1294 during the traditional schedule and 909 during the intervention schedule), representing 634 admissions to the units (385 during the traditional schedule and 249 during the intervention schedule) and 5888 hours of direct observation of interns. The patients' characteristics and the units' characteristics were very similar during the traditional and intervention schedules (Table 2). The number of days included in the traditional schedule exceeded that of the intervention schedule primarily because four interns were required for the intervention schedule as compared with only three for the traditional schedule. Since all interns rotated through both schedules, more traditional than intervention rotations were required to allow each intern to spend three weeks on each schedule. The patients' length of stay and mortality rate did not differ significantly between the two schedules.

SERIOUS MEDICAL ERRORS BY INTERNS

Interns made 35.9 percent more serious medical errors during the traditional schedule than during the intervention schedule (136.0 vs. 100.1 per 1000 patient-days, P<0.001) (Table 3). Interns made 27.8 percent more serious errors that were intercepted during the traditional schedule than during the intervention schedule (70.3 vs. 55.0 per 1000 patient-days, P=0.02) and 56.6 percent more nonintercepted serious errors that reached the patients (44.8 vs. 28.6 per 1000 patient-days, P<0.001). The rates of preventable adverse events did not differ significantly between the two schedules.

Downloaded from www.nejm.org on January 17, 2008 . For personal use only. No other uses without permission.
Copyright © 2004 Massachusetts Medical Society. All rights reserved.

ALL SERIOUS MEDICAL ERRORS AND ADVERSE EVENTS

The rate of all serious medical errors was 22.0 percent higher during the traditional schedule than during the intervention schedule (193.2 vs. 158.4 per 1000 patient-days, P<0.001) (Table 3). Intercepted serious errors occurred 37.2 percent more frequently during the traditional schedule than during the intervention schedule (95.1 vs. 69.3 per 1000 patient-days, P<0.001). The overall rates of nonintercepted serious errors did not differ significantly between the two schedules, nor did the rates of preventable adverse events. There was no significant difference in the rates of total adverse events (preventable plus nonpreventable) between the traditional and intervention schedules (85.0 vs. 93.5 per 1000 patient-days, P=0.31). Secondary analysis of the rates of serious medical errors in which interns were not involved revealed no significant differences between the traditional schedule and the intervention schedule (40.2 vs. 38.5 per 1000 patient-days, P=0.69).

TYPES OF SERIOUS MEDICAL ERRORS

Interns made 20.8 percent more serious medication errors during the traditional schedule than during the intervention schedule (99.7 vs. 82.5 per 1000 patient-days, P=0.03). Interns made 5.6 times as many serious diagnostic errors during the traditional schedule as during the intervention schedule (18.6 vs. 3.3 per 1000 patient-days, P<0.001). The rates of serious procedural errors among interns did not differ significantly between the two schedules (Table 3).

Analysis of the types of all errors (errors made by interns plus errors in which interns were not involved) showed similar patterns (Table 3). Serious medication errors occurred 17.1 percent more frequently during the traditional schedule than during the intervention schedule (135.2 vs. 115.5 per 1000 patient-days, P=0.03). The rates of serious procedural errors did not differ significantly between the two schedules. Serious diagnostic errors were nearly twice as common during the traditional schedule as during the intervention schedule (21.6 vs. 11.0 per 1000 patient-days, P<0.001).

Examples of each type of serious medical error and adverse event observed in the study are provided in Table 4. Subcategories of medication and nonmedication errors are available in Table A of the Supplementary Appendix.

Table 2. Characteristics of the Patients and the System.*

Characteristic	Traditional Schedule	Intervention Schedule
Patients		
No. of patients	354	227
No. of unit admissions	385	249
No. of patient-days	1294	909
Mean age — yr	64.9±0.8	63.2±1.10
Male sex — no./total no. of unit admissions (%)	214/385 (55.6)	126/249 (50.6)
Charlson comorbidity index†	4.0±0.2	4.1±0.2
APACHE II score‡	17.7±0.5	17.9±0.7
Length of unit stay — days		
Median	2.9	3.0
Interquartile range	5.1	5.7
No. who died in unit/total no. of unit admissions — %	49/385 (12.7)	36/249 (14.5)
CCU and MICU		
Daily censuses	9.2±0.1	9.4±0.1
Interns		
No. of medication orders/patient-day	8.2	7.8
No. of procedures/patient-day§	0.28	0.33¶
No. of interpretations of diagnostic tests/patient-day‖	0.28	0.29

* Plus–minus values are means ±SE.
† Scores for the Charlson comorbidity index can range from 0, indicating no serious coexisting conditions, to 6, indicating the presence of metastatic cancer or infection with the human immunodeficiency virus.
‡ Acute Physiology and Chronic Health Evaluation (APACHE) scores can range from 0 to 71, with higher scores indicating an increased likelihood of death.
§ Procedures performed by interns included placement (or rethreading) of central venous catheters, placement of arterial catheters, drawing of arterial blood, intubation, thoracentesis, placement of nasogastric and orogastric tubes, lumbar puncture, and removal of central catheters or tubes.
¶ P<0.001 for the comparison with the traditional schedule.
‖ Interpretations of diagnostic tests by interns included interpretation of chest radiographs, other radiographs, electrocardiograms, and arterial blood gas values.

DISCUSSION

Interns made 36 percent more serious medical errors during a traditional work schedule than during an intervention schedule that eliminated extended work shifts. These included significantly more serious medication errors and 5.6 times as many serious diagnostic errors. As a consequence, the overall rates of serious medical errors were significantly higher during the traditional schedule than during the intervention schedule. Fortunately, most

Downloaded from www.nejm.org on January 17, 2008 . For personal use only. No other uses without permission.
Copyright © 2004 Massachusetts Medical Society. All rights reserved.

Table 3. Incidence of Serious Medical Errors.

Variable	Traditional Schedule	Intervention Schedule	P Value
	no. of errors (rate/1000 patient-days)		
Serious medical errors made by interns			
Serious medical errors	176 (136.0)	91 (100.1)	<0.001
Preventable adverse events	27 (20.9)	15 (16.5)	0.21
Intercepted serious errors	91 (70.3)	50 (55.0)	0.02
Nonintercepted serious errors	58 (44.8)	26 (28.6)	<0.001
Types of serious medical errors made by interns			
Medication	129 (99.7)	75 (82.5)	0.03
Procedural	11 (8.5)	6 (6.6)	0.34
Diagnostic	24 (18.6)	3 (3.3)	<0.001
Other	12 (9.3)	7 (7.7)	0.47
All serious medical errors, unit-wide			
Serious medical errors	250 (193.2)	144 (158.4)	<0.001
Preventable adverse events	50 (38.6)	35 (38.5)	0.91
Intercepted serious errors	123 (95.1)	63 (69.3)	<0.001
Nonintercepted serious errors	77 (59.5)	46 (50.6)	0.14
Types of serious medical errors, unit-wide			
Medication	175 (135.2)	105 (115.5)	0.03
Procedural	18 (13.9)	11 (12.1)	0.48
Diagnostic	28 (21.6)	10 (11.0)	<0.001
Other	29 (22.4)	18 (19.8)	0.45

serious medical errors were either intercepted or did not result in clinically detectable harm to the patient. Although the study was not designed to have sufficient statistical power to detect a difference in preventable adverse events, the incidence of intern-associated preventable adverse events was 27 percent higher during the traditional schedule than during the intervention schedule, a difference that was not statistically significant (20.9 vs. 16.5 per 1000 patient-days, P=0.21). The overall rates of preventable adverse events (intern-associated and non–intern-associated) were not significantly different during the traditional and intervention schedules (38.6 and 38.5 per 1000 patient-days, respectively; P=0.91), although our intervention and observations were focused on the interns. This study was not designed or powered to assess comprehensively the effect of the intervention on adverse event rates in the units as a whole. Therefore, it remains to be determined whether the decrease in the rate of serious medical errors by interns will translate into a reduction in the rate of adverse events.

The prospective, randomized nature of this study allowed for a rigorous evaluation of the effects on patients' safety of an intervention designed to improve interns' sleep and thus decrease medical errors. Prior studies using before-and-after cohort designs to assess the effects of scheduling interventions have provided limited and conflicting data. A before-and-after analysis of a scheduling intervention in one hospital that reduced residents' work hours and decreased cross-coverage of unfamiliar patients by senior residents found that the efficiency of care increased and the rates of errors among residents decreased.[20] In contrast, an unblinded, retrospective study of a New York State regulation that decreased the number of hours worked by house staff but increased cross-coverage found that the efficiency of care declined and rates of medical complications increased.[11] Each was limited by a before-and-after design, which precluded the exclusion of secular trends, increasing experience of house staff, cohort effects, or other external confounders as possible explanations for the changes. Because of concurrent changes in work hours, cross-coverage, and other aspects of care in these studies, it was not possible to identify the elements that may have been responsible for the findings.

The overall incidence of serious errors and adverse events we detected is similar to that reported in other studies of patients' safety in the ICU. For example, Giraud et al.[21] and Rubins and Moskowitz[22] documented the occurrence of 13 to 40 preventable adverse events per 1000 patient-days. The Harvard Medical Practice Study[23] reported lower rates but used a less comprehensive method of data collection and a more restrictive definition of harm, since it sought to detect injuries due to negligence. Donchin et al. reported a higher rate of 1.7 errors per patient-day but included errors with little potential for harm.[15] The rates detected by Donchin et al. may also be higher because they focused on errors in the unit as a whole, whereas we directly observed only interns. Moreover, during daytime hours, when two or more interns were working simultaneously in different parts of the units, our staffing limitations allowed us to observe only one intern at a time. Consequently, the true rate of serious errors in the units as a whole may have been higher.

The article by Lockley et al.[18] demonstrates that eliminating extended work shifts and reducing the number of hours worked by interns led to signifi-

Downloaded from www.nejm.org on January 17, 2008 . For personal use only. No other uses without permission.
Copyright © 2004 Massachusetts Medical Society. All rights reserved.

Table 4. Examples of Serious Medical Errors and Nonpreventable Adverse Events.

Category and Type	Description
Intercepted serious error	
Procedural	As intern is preparing to perform a thoracentesis on the left side of the patient's chest, the senior resident enters the room and informs the intern that the pleural effusion is on right side of the patient's chest.
Diagnostic	Several days after a patient with a history of flash pulmonary edema is admitted for congestive heart failure, intern reports that patient is in clinically stable condition, having miscalculated that 24-hour input and output volumes are well matched (positive by 20 ml). The nurse is concerned that patient seemed overloaded with fluid and in mild respiratory distress and requests a reevaluation. A recalculation by the senior resident reveals an error by a factor of 100: the patient's input and output volume has, in fact, been positive by 2000 ml for the prior 24 hours. Furosemide is promptly administered and the patient's symptoms improve.
Medication	Intern orders an intravenous vasopressin drip at rate of 0.2 U/min (overdose by a factor of 10). Nurse intercepts the order, and the rate is changed to 0.02 U/min.
Nonintercepted serious error	
Procedural	Patient with defibrillator implanted on left side urgently needs central access for inotropic support. Intern inserts a central venous catheter in the left subclavian vein. Not recognizing that the vein contains the wire from the defibrillator, the intern is having repeated difficulty advancing the introducer. In the middle of the placement, the cardiology fellow enters and asks the intern to abort the procedure immediately. The catheter is removed before it can interfere with or dislodge the defibrillator wire.
Diagnostic	A middle-aged patient with a complete heart block is admitted to the CCU. The intern fails to examine the patient's back. The following day, the patient is noted to have a well-developed erythema migrans rash on the back, consistent with the presence of Lyme disease, which is later confirmed by serologic testing. Initiation of Lyme therapy is delayed.
Medication	Intern orders an antibiotic for a patient with a listed allergy to the medication. One dose is given before the error is detected, but the patient does not have an allergic reaction.
Preventable adverse event	
Procedural	A right-sided tension pneumothorax develops after a technical error during placement of a subclavian venous catheter leads to pleural-space puncture.
Diagnostic	The attending physician devised a plan to transfuse a patient for a hematocrit of <30. Despite these instructions, the intern fails to check laboratory results for 36 hours. When the laboratory results are finally checked, hematocrit is found to have been 26 in the interim. The patient has tachycardia for a protracted time as a consequence.
Medication	Bradycardia and hypotension develop owing to an inadvertent overdose of a benzodiazepine.
Nonpreventable adverse event	
Procedural	Transfusion is required for severe bleeding resulting from placement of a medically indicated nasogastric tube in a patient with coagulopathy. There is no error in placement or technique.
Medication	A rash related to nafcillin develops in a patient with no known drug allergies.

cant improvements in interns' sleep and reductions in attentional failures. Although causality cannot be established, it was our a priori hypothesis that increases in sleep resulting from the elimination of extended work shifts and reduction of work hours would lead to a decrease in serious medical errors.[24] There were no significant differences between the two schedules in the patients' severity of illness or other individual or systemic variables that could independently account for the observed differences in the rates of medical errors. Our randomized study design greatly diminished the likelihood of hidden confounding owing to secular trends, seasonal effects, learning over the course of the year, or other external factors unrelated to our study.

Before we initiated the intervention schedule, concern was expressed that decreasing the number of hours interns worked might diminish their

Downloaded from www.nejm.org on January 17, 2008 . For personal use only. No other uses without permission.
Copyright © 2004 Massachusetts Medical Society. All rights reserved.

role in the units, thereby shifting the burden of order writing and procedures and, hence, the risk of errors to more senior staff. Our results did not bear out this concern: the number of medications ordered and tests interpreted by interns per patient-day did not differ significantly between the two schedules, and interns performed significantly more procedures per patient-day during the intervention schedule. Moreover, the error rates among senior residents and other staff members were not increased during the intervention schedule. Thus, the substantially lower rates of errors by interns during the intervention schedule cannot be due to shifting of errors to more senior staff.

The Institute of Medicine's report "To Err Is Human"[25] was notably silent regarding the issue of sleep deprivation, largely because data directly linking sleep deprivation and medical error have been lacking. Our study helps to fill this knowledge gap and provides data suggesting that the sleep deprivation associated with the traditional extended shifts of 24 hours or more worked by interns may contribute to the high risk of medical errors in critical care units.

It is important to emphasize that not all interventions that reduce interns' work hours will increase interns' sleep[26] or improve patients' safety. Schedule design is a critical factor in determining the extent to which around-the-clock work schedules disrupt wake–sleep cycles, even when the number of weekly work hours remains the same.[27] Furthermore, any systemic intervention that reduces work hours necessarily increases either providers' workload (i.e., the number of patients covered by a provider at any time) or the number of handoffs in care between medical personnel on shorter work shifts. Either can lead to increased rates of errors and adverse events.[10] "Night-float" systems, which use residents on night shifts to allow physicians working extended work shifts protected time for sleep, have their own set of risks. Night-float residents often know patients less well than do other team members (particularly if multiple residents share responsibilities as night floats over the course of a week, or if night floats are responsible for an increased number of patients), and may themselves be sleep-deprived and error-prone.[28] For these reasons, we ultimately decided not to implement a night-float system as a means of reducing interns' work hours, as originally planned.[24] Our data support the hypothesis that elimination of extended work shifts in a system that minimizes cross-coverage can improve patients' safety. These gains might not be realized in systems that use extensive cross-coverage.

Although our intervention decreased the rate of serious errors overall, our efforts to optimize the sign-out process were only partially successful. The computerized template was never fully adopted, and the effectiveness of the planned evening sign-out was frequently suboptimal. Although some groups of interns worked successfully as teams and effectively signed out every evening, even in the absence of formal training in team management, others did not. In the latter case, the night-call intern was often unaware of historical details regarding patients admitted by the day-call intern and sometimes performed poorly when describing these patients on morning rounds. This led to a widespread impression that communication on the intervention schedule was problematic, making the improvements in patients' safety we observed all the more remarkable. We suggest that future scheduling interventions address this issue by adding formal evening rounds for the entire team. Such improvements, coupled with the elimination of extended work shifts, could further improve patients' safety.

Our study has several limitations. The intervention schedule improved work hours but still involved shifts that were long enough to induce a number of attentional failures that was greater than would be expected among fully rested people.[18] We studied two ICUs in a single hospital, and our results may not be generalizable to other settings. In addition, although our study was very large as compared with prior observational safety studies,[15] the study was not powered to detect differences in the rates of preventable adverse events. Larger-scale, multicenter trials are needed to investigate this aspect.

Another important limitation was our inability to blind the medical observers to the schedule of the interns, an issue commonly encountered in investigations of systemic interventions to maximize patients' safety. We addressed this in two ways: first, we instructed observers — none of whom were study investigators — in the importance of consistent, objective detection of serious errors, regardless of study schedule. Second, all initial observations were also reviewed by two independent investigators who were blinded to the study's conditions and who classified incidents with extremely high reliability. Nonetheless, we cannot exclude the possibility that some bias may have resulted

Downloaded from www.nejm.org on January 17, 2008 . For personal use only. No other uses without permission.
Copyright © 2004 Massachusetts Medical Society. All rights reserved.

from the inability to blind the primary detection process, though our reliability data suggest that this bias was probably minimal.

Notably, our data on the high incidence of intercepted serious errors in ICU settings indicate that the ability of other personnel to act as interns' safety net — nurses, pharmacists, and senior medical staff — is very important in preventing injury to patients as a result of interns' errors. Therefore, future studies should seek to improve and measure objectively the sleep and performance of all clinical unit personnel, since team performance may critically affect patients' safety.[29] Having interns on a different schedule than supervising residents may have introduced discontinuities in education and interfered with the traditional resident–intern mentorship bond. We would recommend that future studies investigate the effects of eliminating the extended work shifts of interns and senior residents, both to avoid this problem and because it is unlikely that interns are uniquely susceptible to the adverse effects of sleep deprivation.

Prior interventions that have proved successful in reducing serious medical errors in ICU settings have included the use of computerized provider order entry (CPOE)[30] and on-site monitoring of orders by clinical pharmacists.[31] The higher intern-associated rate of serious medical errors during the traditional schedule, even in the presence of CPOE, clinical pharmacists, unrestricted use of caffeine by interns,[32] and a perceived increase in the risk of handoff errors, indicates the extent of impairment associated with extended work shifts. This observation corroborates the prior experimental finding that a single night of continuous sleep deprivation causes decrements in performance similar to those induced by a blood alcohol level of 0.10 percent.[33]

By reducing consecutive and weekly work hours, our scheduling intervention attempted to address both acute sleep deprivation and chronic partial sleep deprivation. By reducing interns' sleep deprivation and, hence, depth of subsequent sleep, we also indirectly addressed the adverse effects of sleep inertia (i.e., an increased tendency to err on awakening) on performance, since such impairment is a function of sleep depth.[34] The schedule was also designed to attenuate the circadian performance nadir by taking advantage of the blunting of this nadir that occurs when the homeostatic sleep drive is lower.[35,36] By providing interns with the opportunity to sleep in the afternoon before working overnight, the schedule thereby muted the effect of circadian misalignment on performance. Medical or surgical simulators could help isolate the effects of these interacting factors, since the relative importance of these variables remains unclear. Strategic use of a novel regimen of caffeine[32] or ambient light of specific intensity and wavelengths[37,38] may further mitigate the deterioration in performance resulting from circadian misalignment.

In conclusion, the rates of serious medical errors in two ICUs were lowered by eliminating extended work shifts and reducing the number of hours interns worked each week. Our results may have important implications for health policy, since more than 100,000 physicians are currently in training in the United States.[39] Most of these residents are regularly scheduled to work 30-hour shifts, since extended work shifts and long workweeks continue to be permitted, even under the scheduling reforms instituted last year by the Accreditation Council for Graduate Medical Education. Further modifications of these standards, particularly with respect to the duration of work shifts, may be needed to improve patients' safety in teaching hospitals nationwide.

Supported by a grant (RO1 HS12032) from the Agency for Healthcare Research and Quality (AHRQ), affording data-confidentiality protection by federal statute (Public Health Service Act 42 U.S.C.); by a grant (RO1 OH07567) from the National Institute for Occupational Safety and Health, which provided a Certificate of Confidentiality for data protection; by the Department of Medicine, Brigham and Women's Hospital; by the Division of Sleep Medicine, Harvard Medical School; by the Brigham and Women's Hospital; and by a General Clinical Research Center grant (M01 RR02635) from the National Center for Research Resources. Dr. Landrigan is the recipient of an AHRQ career development award (K08 HS13333); Dr. Cronin is the recipient of an AHRQ National Research Service Award (F32 HS14130) and a National Heart, Lung, and Blood Institute fellowship in the program of training in Sleep, Circadian, and Respiratory Neurobiology at Brigham and Women's Hospital (T32 HL079010); Dr. Lockley is the recipient of a traveling fellowship from the Wellcome Trust, United Kingdom (060018/B/99/Z); and Dr. Czeisler is supported in part by the National Space Biomedical Research Institute through NASA (NCC 9-58).

We are indebted to our data-collection team, without whom this project could not have occurred; to Patricia Aboagye-Kumi, M.D., Megan Callahan, M.D., Vilma L. Castenada, M.D., Rajneesh S. Hazarika, M.B., B.S., Kristina Martell-Waldrop, R.N., Tamara Sikharulidze, M.D., Noma Rehman, M.D., and Victor Tsveybel, R.N., for their hard work and dedication; to the house staff, attending physicians, nurses, and clinical pharmacists of the CCU and MICU for their ongoing support; to Victor J. Dzau, M.D., Anthony D. Whittemore, M.D., Jeffrey Otten, Matthew Van Vranken, Gary L. Gottlieb, M.D., M.B.A., and Joseph Martin, M.D., Ph.D., for their support and leadership in fostering this study; and to Laura K. Barger, Ph.D., M.P.H., Erin E. Evans, B.S., R.P.S.G.T., Heather L. Gornik, M.D., and DeWitt C. Baldwin, Jr., M.D., for assistance in designing and coordinating the many facets of this effort.

Downloaded from www.nejm.org on January 17, 2008 . For personal use only. No other uses without permission.
Copyright © 2004 Massachusetts Medical Society. All rights reserved.

REFERENCES

1. Friedman RC, Bigger JT, Kornfeld DS. The intern and sleep loss. N Engl J Med 1971; 285:201-3.
2. Grantcharov TP, Bardram L, Funch-Jensen P, Rosenberg J. Laparoscopic performance after one night on call in a surgical department: prospective study. BMJ 2001; 323:1222-3.
3. Eastridge BJ, Hamilton EC, O'Keefe GE, et al. Effect of sleep deprivation on the performance of simulated laparoscopic surgical skill. Am J Surg 2003;186:169-74.
4. Gaba DM, Howard SK. Fatigue among clinicians and the safety of patients. N Engl J Med 2002;347:1249-55.
5. Veasey S, Rosen R, Barzansky B, Rosen I, Owens J. Sleep loss and fatigue in residency training: a reappraisal. JAMA 2002;288: 1116-24.
6. Weinger MB, Ancoli-Israel S. Sleep deprivation and clinical performance. JAMA 2002;287:955-7.
7. Samkoff JS, Jacques CHM. A review of studies concerning effects of sleep deprivation and fatigue on residents' performance. Acad Med 1991;66:687-93.
8. Pilcher JJ, Huffcutt AI. Effects of sleep deprivation on performance: a meta-analysis. Sleep 1996;19:318-26.
9. Buysse DJ, Barzansky B, Dinges D, et al. Sleep, fatigue, and medical training: setting an agenda for optimal learning and patient care. Sleep 2003;26:218-25.
10. 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; 121:866-72.
11. 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: 374-8.
12. Schwartz RJ, Dubrow TJ, Rosso RF, Williams RA, Butler JA, Wilson SE. Guidelines for surgical residents' working hours: intent vs reality. Arch Surg 1992;127:778-82.
13. Baldwin DC Jr, Daugherty SR, Tsai R, Scotti MJ Jr. A national survey of residents' self-reported work hours: thinking beyond specialty. Acad Med 2003;78:1154-63.
14. Wu AW, Folkman S, McPhee SJ, Lo B. Do house officers learn from their mistakes? JAMA 1991;265:2089-94.
15. Donchin Y, Gopher D, Olin M, et al. A look into the nature and causes of human errors in the intensive care unit. Crit Care Med 1995;23:294-300.
16. Kaushal R, Bates DW, Landrigan C, et al. Medication errors and adverse drug events in pediatric inpatients. JAMA 2001;285:2114-20.
17. Petersen LA, Orav EJ, Teich JM, O'Neil AC, Brennan TA. Using a computerized sign-out program to improve continuity of inpatient care and prevent adverse events. Jt Comm J Qual Improv 1998;24:77-87.
18. Lockley SW, Cronin JW, Evans EE, et al. Effect of reducing interns' weekly work hours on sleep and attentional failures. N Engl J Med 2004;351:1829-37.
19. Bates DW, Boyle DL, Vander Vliet MB, Schneider J, Leape LL. Relationship between medication errors and adverse drug events. J Gen Intern Med 1995;10:199-205.
20. Gottlieb DJ, Parenti CM, Peterson CA, Lofgren RP. Effect of a change in house staff work schedule on resource utilization and patient care. Arch Intern Med 1991;151: 2065-70.
21. Giraud T, Dhainaut JF, Vaxelaire JF, et al. Iatrogenic complications in adult intensive care units: a prospective two-center study. Crit Care Med 1993;21:40-51.
22. Rubins HB, Moskowitz MA. Complications of care in a medical intensive care unit. J Gen Intern Med 1990;5:104-9.
23. Brennan T, Leape LL, Laird NM, et al. Incidence of adverse events and negligence in hospitalized patients: results of the Harvard Medical Practice Study I. N Engl J Med 1991; 324:370-6.
24. Lamberg L. Long hours, little sleep: bad medicine for physicians-in-training? JAMA 2002;287:303-6.
25. Kohn LT, Corrigan JM, Donaldson MS, eds. To err is human: building a safer health system. Washington, D.C.: National Academy Press, 1999.
26. Richardson GS, Wyatt JK, Sullivan JP, et al. Objective assessment of sleep and alertness in medical house staff and the impact of protected time for sleep. Sleep 1996;19: 718-26.
27. Czeisler CA, Moore-Ede MC, Coleman RH. Rotating shift work schedules that disrupt sleep are improved by applying circadian principles. Science 1982;217:460-3.
28. Cavallo A, Ris MD, Succop P. The night float paradigm to decrease sleep deprivation: good solution or a new problem? Ergonomics 2003;46:653-63.
29. Leape LL, Berwick DM, Bates DW. What practices will most improve safety? Evidence-based medicine meets patient safety. JAMA 2002;288:501-7.
30. Bates DW, Leape LL, Cullen DJ, et al. Effect of computerized physician order entry and a team intervention on prevention of serious medication errors. JAMA 1998;280: 1311-6.
31. Leape LL, Cullen DJ, Clapp MD, et al. Pharmacist participation on physician rounds and adverse drug events in the intensive care unit. JAMA 1999;282:267-70. [Erratum, JAMA 2000;283:1293.]
32. Wyatt JK, Cajochen C, Ritz-De Cecco A, Czeisler CA, Dijk DJ. Low-dose repeated caffeine administration for circadian-phase-dependent performance degradation during extended wakefulness. Sleep 2004;27:374-81.
33. Dawson D, Reid K. Fatigue, alcohol and performance impairment. Nature 1997;388: 235.
34. Dinges DF. Are you awake? Cognitive performance and reverie during the hypnopompic state. In: Bootzin R, Kihlstrom J, Schacter DL, eds. Sleep and cognition. Washington, D.C.: American Psychological Association, 1990:159-75.
35. Jewett ME, Kronauer RE. Interactive mathematical models of subjective alertness and cognitive throughput in humans. J Biol Rhythms 1999;14:588-97.
36. Wyatt JK, Ritz-De Cecco A, Czeisler CA, Dijk DJ. Circadian temperature and melatonin rhythms, sleep, and neurobehavioral function in humans living on a 20-h day. Am J Physiol 1999;277:R1152-R1163.
37. Czeisler CA, Johnson MP, Duffy JF, Brown EN, Ronda JM, Kronauer RE. Exposure to bright light and darkness to treat physiologic maladaptation to night work. N Engl J Med 1990;322:1253-9.
38. Lockley SW, Brainard GC, Czeisler CA. High sensitivity of the human circadian melatonin rhythm to resetting by short wavelength light. J Clin Endocrinol Metab 2003; 88:4502-5.
39. Accreditation Council for Graduate Medical Education. Number of programs and filled positions by specialty for the current academic year (Ending June 30th, 2005). (Accessed October 4, 2004, at http://www.acgme.org/adspublic/reports/specialty_prognum.asp.)

Copyright © 2004 Massachusetts Medical Society.

Downloaded from www.nejm.org on January 17, 2008 . For personal use only. No other uses without permission.
Copyright © 2004 Massachusetts Medical Society. All rights reserved.