Effect of the 2011 vs 2003 Duty Hour Regulation–Compliant Models on Sleep Duration, Trainee Education, and Continuity of Patient Care Among Internal Medicine House Staff

A Randomized Trial

Sanjay V. Desai, MD; Leonard Feldman, MD; Lorrel Brown, MD; Rebecca Dezube, MD; Hsin-Chieh Yeh, PhD; Naresh Punjabi, MD, PhD; Kia Afshar, MD; Michael R. Grunwald, MD; Colleen Harrington, MD; Rakhi Naik, MD; Joseph Cofrancesco Jr, MD, MPH

Importance: On July 1, 2011, the Accreditation Council for Graduate Medical Education implemented further restrictions of its 2003 regulations on duty hours and supervision. It remains unclear if the 2003 regulations improved trainee well-being or patient safety.

Objective: To determine the effects of the 2011 Accreditation Council for Graduate Medical Education duty hour regulations compared with the 2003 regulations concerning sleep duration, trainee education, continuity of patient care, and perceived quality of care among internal medicine trainees.

Design and Setting: Crossover study design in an academic research setting.

Participants: Medical house staff.

Intervention: General medical teams were randomly assigned using a sealed-envelope draw to an experimental model or a control model.

Main Outcome Measures: We randomly assigned 4 medical house staff teams (43 interns) using a 3-month crossover design to a 2003-compliant model of every fourth night overnight call (control) with 30-hour duty limits or to one of two 2011-compliant models of every fifth night overnight call (Q5) or a night float schedule (NF), both with 16-hour duty limits. We measured sleep duration using actigraphy and used admission volumes, educational opportunities, the number of handoffs, and satisfaction surveys to assess trainee education, continuity of patient care, and perceived quality of care.

Results: The study included 560 control, 420 Q5, and 140 NF days that interns worked and 834 hospital admissions. Compared with controls, interns on NF slept longer during the on call period (mean, 5.1 vs 8.3 hours; P = .003), and interns on Q5 slept longer during the post-call period (mean, 7.5 vs 10.2 hours; P = .05). However, both the Q5 and NF models increased handoffs, decreased availability for teaching conferences, and reduced intern presence during daytime work hours. Residents and nurses in both experimental models perceived reduced quality of care, so much so with NF that it was terminated early.

Conclusions and Relevance: Compared with a 2003-compliant model, two 2011 duty hour regulation–compliant models were associated with increased sleep duration during the on-call period and with deteriorations in educational opportunities, continuity of patient care, and perceived quality of care.

Raduate medical education training programs must balance 3 key priorities, namely, training excellence, resident well-being, and safe, effective patient care. Duty hours, which had no limits when the first modern medical residency was established in 1889 by William Osler, have become an important variable in this balance. A 1971 study that found exhausted interns tended to misinterpret electrocardiograms prompted discussion but no action on duty hours. Subsequently, the well-publicized death of Libby Zion prompted the first state-level regulation of duty hours in 1989 in New York. The Accreditation Council for Graduate Medical Education (ACGME) imposed
the first national regulation of duty hours in 2003, with a July 1, 2011, revision. The 2011 rules mandate rest periods between duty periods, increased supervision for junior trainees, and a 16-hour limit on continuous duty hours for postgraduate year 1 (PGY-1) trainees (interns).

The 2003 ACGME limits on duty hours were intended to improve resident well-being and patient safety, but studies have not consistently demonstrated improvements in either. Therefore, we conducted a controlled experiment to determine the effects of the 2011 ACGME duty hour regulations compared with the 2003 regulations concerning sleep duration, trainee education, continuity of patient care, and perceived quality of care using an experienced group of trainees in internal medicine. We randomly assigned 4 medical house staff teams (43 interns) using a 3-month crossover design to a 2003-compliant model of every fourth night overnight call (control) with 30-hour duty limits or to one of two 2011-compliant models of every fifth night overnight call (Q5) or a night float schedule (NF), both with 16-hour duty limits.

### METHODS

#### STUDY POPULATION

Data from all the patients admitted to house staff general internal medicine services at The Johns Hopkins Hospital, Baltimore, Maryland, during two 4-week periods (January 27, 2011, to February 23, 2011, and March 24, 2011, to April 20, 2011) were included; patients admitted before or discharged after the study period were excluded. Postgraduate years 1 through 3 trainees and ward nurses on these services during the study periods were also included. The Osler Medicine Training Program consists of 4 general medical teams (firms), each composed of the following physician members: 1 attending-level chief resident (assistant chief of service), 2 PGY-3 trainees (senior residents), and 4 PGY-1 trainees (interns). After stratification by sex, program track, and medical school, interns are randomly assigned to firms for their entire residency. The study was approved by the institutional review board, and all the study participants provided informed consent.

#### INTERVENTIONS AND RANDOMIZATION

Two “control” firms operated within the ACGME 2003 duty hour regulations, with the team composition as already described in the “Study Population” subsection. Control interns took overnight call every fourth night, beginning at 9 PM and concluding no later than 6 PM the next day, with a maximal continuous duty of 30 hours. Two experimental models were designed to comply with the ACGME 2011 duty hours regulations (the schedules are given in the eAppendix; http://www.jama(internalmed.com).

The first experimental model with an every fifth night on-call schedule (Q5) consisted of an intern on overnight call every fifth night beginning at 9 PM and concluding no later than 1 PM the next day, for a maximal continuous duty of 16 hours. The second experimental model was a night float system (NF), which used day and night shifts with an intern working for approximately 6 consecutive nights, each with maximal continuous duty of 14 hours and with day shifts the remainder of the study period. Both models included continuous PGY-2 or PGY-3 supervision. Reduced duty hours and increased supervision required each experimental firm to include 1 additional intern and 1 additional resident compared with the control firm. All firms were bound by the other 2003 duty hour regulations. Interns were the primary providers for all patients at all times.

The firms were randomly assigned using a sealed-envelope draw to an experimental model or a control model. We planned a crossover design with two 4-week blocks separated by a 4-week washout period. During block 1, there was Q5 firm, 1 NF firm, and 2 control firms. After a 4-week washout period during which all firms operated under control conditions, we planned to cross over each experimental firm to the opposite experimental model, while each control firm would continue as a control. Although participants on the teams could not be blinded to their model, the data handling and analysis were done by those blinded to the team assignments.

The primary outcome of this study was the effect of the 2011 duty hour and supervision regulations on an internal medicine training program compared with the 2003 regulations on PGY-1 on-call period sleep duration. Sleep duration was measured using wristwatch actigraphy (Activwatch Spectrum; Respironics), a valid and convenient alternative to polysomnography. Total sleep duration for every 24-hour period that the actigraph was worn was determined by software (Activware 5; Respironics) using a computerized algorithm. Automated results were then reviewed for accuracy and edited by a trained sleep technician and one of us (N.P.). Interns wore wristwatch actigraphs 24 hours a day during the study periods.

Secondary outcomes for the study were operations, trainee education, continuity of patient care, sleep duration outside of the on-call period, and satisfaction of interns and nurses across domains of education and patient care. Operations outcomes included length of stay, 30-day readmissions, and the number of discharges before 11 AM, an institutional objective. Educational outcomes were assessed using trainee surveys, intern admission volumes, daytime presence in the hospital, and availability to attend the daily weekday noon conferences. Continuity of patient care was assessed by calculating the number of handoffs and the number of different interns for 1 patient during a 3-day length of stay. Satisfaction was assessed by trainee and nurse surveys (the surveys are given in the eAppendix). All surveys were administered at the midpoint and end of each 4-week study period. All survey questions used a Likert-type scale ranging from 1 to 5 (least favorable outcome to most favorable outcome).

### STATISTICAL ANALYSIS

We used the analysis of variance F test, Wilcoxon rank sum test, or Pearson χ2 test to compare the difference and to assess the statistical significance across models. Because duty hours for the on-call period varied in the 3 models, sleep was compared in multiple ways (Figure 1). Poor-quality actigraphic sleep data were excluded; for analyses based on 48-hour periods, only data from complete pairs were used (eAppendix). Generalized estimating equations were applied to estimate the difference in sleep time across firms after considering the repeated sleep measures clustered by individual interns. Tests of significance were 2-tailed, with an α level of .05. Data were analyzed using statistical software (SAS, version 9.2; SAS Institute, Inc; and STATA/SE, version 11; StataCorp LP). We estimated the minimal detectable difference based on a sample size of 5 interns per group, with 8 individual sleep measures per intern during 2 months in the on-call schedule. With 80% power, an SD of 3, and the intracluster correlation of 0.1, the minimal detectable difference would be 3.2 hours using a 2-sided t test with a significance level of .03.

Interim analysis of trainee and nurse survey data, along with feedback, indicated that NF call received lower satisfaction scores compared with Q5 call. Specifically, trainees were less satis-
fied with the quality of care on NF (mean, 2.70; 95% CI, 2.39-3.01) compared with Q5 (mean, 3.19; 95% CI, 2.82-3.36) (P < .001). A trend toward lower satisfaction for NF compared with Q5 was also observed for trainee education, outpatient experience, and team membership. In addition, nurses’ satisfaction with quality of care was lower in NF (mean, 3.18; 95% CI, 2.97-3.38) than in Q5 (mean, 3.24; CI, 3.08-3.41) (P = .02) and demonstrated a trend toward lower satisfaction for communication and patient safety. As a result, NF was not continued, and the second 4-week period included 2 Q5 firms and 2 control firms.

**RESULTS**

The study period included 4 control study periods, 3 Q5 periods, and 1 NF period, corresponding to 43 interns, 26 PGY-2 and PGY-3 residents, and 834 discrete hospital admissions. The analysis comprised 560 control, 420 Q5, and 140 NF days that interns worked. Patients’ data on severity of illness were similar in all 3 models (eAppendix). Response rates from trainee surveys were 73% for control firms, 77% for Q5, and 81% for NF.

Sleep data were analyzed for 274 control, 273 Q5, and 63 NF days and are given in Figure 2. Poor quality precluded analysis of 20% of sleep data (eAppendix). On average, interns in the control model slept 3 hours less across 48-hour on-call periods (mean, 12.9; 95% CI, 11.0-14.8 hours) compared with interns in the Q5 model (mean, 16.1; 95% CI, 14.7-17.5 hours; P = .26) and the NF model (mean, 15.9; 95% CI, 13.0-18.9 hours; P = .39). Although these differences were not statistically significant, the variance in sleep duration was significantly reduced in the Q5 model (SD, 4.7 hours) and the NF model (SD, 2.8 hours) (P < .001 for both) compared with the control model (SD, 6.9 hours). Control interns slept significantly less than NF interns during the on-call period (mean, 5.1; 95% CI, 4.1-6.1 vs 8.3; 95% CI, 7.2-9.4 hours; P = .003) and slept less than Q5 interns during the post-call period (mean, 7.5; 95% CI, 6.2-8.7 vs 10.2; 95% CI, 9.2-11.1 hours; P = .05). In both the on-call and postcall days, the variance in sleep duration was significantly reduced in the experimental models. Assessing any day during the study independent of role, no differences in sleep duration were observed between the control model (mean, 7.6; 95% CI, 6.9-8.1 hours) and the Q5 model (mean, 7.7; 95% CI, 7.3-8.1 hours; P = .90) and the NF model (mean, 8.2; 95% CI, 7.7-8.7 hours; P = .98).

Educational opportunities were decreased in both experimental models compared with the control model. For example, interns had fewer admission experiences in the experimental models. Specifically, interns admitted a higher proportion of the firm’s patients each month on the control model (79%) compared with the Q5 model (61%) or the NF model (64%) (P < .001 across groups). On average, more patients per month were admitted by each control intern (mean, 24.8) compared with each Q5 intern (mean, 16.5) and NF intern (mean, 17.4), and more patients were primarily cared for by each control intern (mean, 31.5) compared with each Q5 intern (mean, 27.0) and NF intern (mean, 27.2). Opportunities to attend a daily noon conference were reduced by 25% in both experimental models. Last, control interns worked a mean of 39 hours per week between 8 AM and 6 PM (standard work hours), 30% more than Q5 interns and 13% more than NF interns. As such, traditional educational activities occurring during standard work hours, including attending and teaching rounds, were less available to interns in experimental models.

The minimal number of handoffs between interns increased from 3 to as high as 9, a 130% to 200% increase, in the experimental models compared with the control model. The minimal number of different interns caring for a patient during a 3-day stay increased from 3 to as high as 5, a 33% to 67% increase, in the experimental models compared with the control model (Table). Trainee satisfaction was higher in the control model compared with either experimental model across several domains (Table), including perceived quality of care and team membership, assessed using a collection of questions focused on teamwork and perceptions of being a central member of the team. Nurses perceived that the highest quality of care was provided to patients in the control model. Nurses also reported lower satisfaction with communication and patient safety in the NF model compared with the control model.

Intern duty hour violations for maximal permitted continuous hours, while occurring more frequently in the experimental models, occurred in all models, with few excess hours. Violations of the 30-hour rule occurred in 4% of the work periods, for a mean of 1.5 hours, in the control arm. Violations of the 16-hour rule occurred in 36% of the work periods, for a mean of 1.8 hours, in the
The results of this experimental study suggest that implementing the 2011 ACGME duty hour regulations may present challenges and could have unintended consequences. While the regulations produced increased sleep duration during the on-call period, they also decreased continuity of patient care, intern and nurse perceptions of quality of care, and educational opportunities from teaching and patient care.

These findings are consistent with prior research about duty hour regulations. A study of 220 pediatric residents demonstrated no change in the amount of burnout, hours slept, depression, motor vehicle crashes, or resident educational outcomes related to the 2003 regulations. Similarly, 3 neurology training programs piloting the 2011 regulations found no improvement in education, sleepiness, study hours, or sleep duration, but found adverse effects on continuity of care, transitions of care, faculty satisfaction, and trainees' knowledge of their patients. Faculty satisfaction suffered also from the 2003 regulations, with a shift in effort from teaching, research, and academics to direct clinical care.

Studies of the 2003 regulations have not consistently demonstrated improved patient outcomes. A review of more than 14 million admissions demonstrated no change in the rate of patient safety events after the 2003 regulations were implemented. Studies conducted in intensive care units (ICUs) show mixed results. For ex-
ample, interns with shorter duty hours in the medical and coronary ICU at a single center made fewer serious medical errors than interns with longer duty hours. However, an analysis of more than 200,000 ICU patients in the United States between 2001 and 2005 revealed no reduction in ICU or hospital mortality attributable to the 2003 regulations. Likewise, large studies of Medicare, Veterans Affairs, and other high-risk patient populations showed no reduction in mortality following implementation of the 2003 regulations.

Sleep duration increased within the on-call period in both experimental models. In addition, there was much less variability in interns’ sleep duration in either experimental model compared with the control model. This may be a result of personal preferences and behaviors, with some individuals prioritizing sleep over other tasks both inside and outside of the hospital. Notably, only a mean of 3 of 14 hours (21%) newly gained from work periods was used for sleep. The decreased variance and increased duration of sleep in the interns with reduced duty hours suggests that, if a group of interns is given the opportunity to sleep by leaving the hospital sooner, most will use at least part of the increased time outside of the hospital to sleep more. However, the clinical significance of the sleep changes observed in the experimental models remains unclear because intern satisfaction with education or quality of care did not improve. This may be because sleep duration on days outside of the on-call period did not differ. In fact, the mean duration of sleep an intern had on any day was the same regardless of the model and was similar to the mean daily sleep of young adults in the United States. Sleep and sleep architecture represent a complex science, and the influence on performance from total hours, weekly means, circadian alignment and the number of interruptions all need greater study.

Compared with controls, trainees on the experimental models admitted fewer patients and followed them up for shorter continuous periods. While the optimal balance between admission volumes (or workload) and educational experience is unclear, concerns have been raised about the competency achievable with less in-hospital experience during any fixed duration of training. In addition, opportunities for trainee education were reduced with restricted shifts, many of which occur solely during evening hours, precluding participation in traditional core educational components of medicine residency programs, such as noontime conference and morning rounds.

This disruption in education can reduce the effectiveness of training programs’ current provision of formal and informal curricula. Our models preserved 2-hour morning bedside rounds led by our assistant chief of service, a cornerstone of our educational curriculum. However, interns on our experimental models were on the wards less during standard work hours and had fewer opportunities to work with our faculty, consultants, and other health care professionals who are present more often during these hours. Programs have expanded curricula to include evening teaching by attending physicians, but there will be inherent limitations in the content delivered during these hours because of faculty availability and patient convenience. Finally, faculty often face increased clinical duties to compensate for the work previously done by residents, reducing their time to teach. This, coupled with pressures to “get the resident out on time,” will further reduce teaching opportunities even during daytime hours. These changes in trainee education...
may have adverse short-term and long-term effects on clinical competence and ultimately on patient care and safety.

Handoffs, a known risk factor for medical errors, increased 130% to 200% in the experimental models. Increased supervision and training in handoffs may mitigate some of the threat. Our results suggest an urgent need to study, standardize, teach, and improve this critical component of care.

Furthermore, the lower satisfaction of nurses with quality of care in both experimental models was particularly striking. In general, this is because they may hold a longer-term view of care and because they are less affected by the implications that duty hour limits have on physicians.

Our study has several limitations. First, the study focuses on internal medicine training at one institution, although we believe the issues forced by duty hour rules affect most types of training programs at most academic medical centers. Second, by intentionally conducting the study with experienced interns to minimize patient risk, we may have underestimated the adverse consequences of the new models. Similarly, although we studied the models during 8 weeks, the number of patient encounters may not have been sufficient to identify small differences in operation outcomes. Third, we could not exclude the possibility that some of the dissatisfaction and perceptions of interns might be a result of unfamiliarity with or prejudice against the new models, as well as reluctance to systematic change. Despite these limitations, such an experiment cannot be performed again because pre-2011 conditions are no longer permissible.

The main implication of this study is that the 2011 ACGME duty hour regulations may have unintended adverse consequences. There is a complex relationship between the many variables that influence patient safety and residency education. Furthermore, training programs differ substantially in history, culture, and the preferences and attitudes of the trainees whom they attract. Therefore, the ACGME might consider allowing programs to experiment with alternate models under close scientific and regulatory scrutiny. Such experimentation could generate successful innovations that leverage differences in local conditions, resources, expertise, and culture to improve quality of care and training nationally.

Accepted for Publication: November 5, 2012.
Published Online: March 25, 2013. doi:10.1001/jamainternmed.2013.2973

Correspondence: Sanjay V. Desai, MD, 1830 E Monument St, Room 9029, Baltimore, MD 21205 (sanjaydesai@jhmi.edu).

Author Contributions: Dr Desai had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Study concept and design: Desai, Feldman, Brown, Dezube, Punjabi, Afshar, Grunwald, Harrington, Naik, and Cofrancesco. Acquisition of data: Desai and Cofrancesco. Analysis and interpretation of data: Desai, Feldman, Brown, Dezube, Yeh, Punjabi, Afshar, Grunwald, Harrington, Naik, and Cofrancesco. Drafting of the manuscript: Desai, Feldman, and Cofrancesco. Statistical analysis: Desai, Yeh, Punjabi, and Cofrancesco.

Conflict of Interest Disclosures: None reported.


Additional Contributions: Myron L. Weisfeld, MD, provided leadership and support, Fredrick L. Brancati, MD, offered a thorough review and invaluable guidance, the Institute for Clinical and Translational Research gave actigraph support, and the Clinician-Educator Mentoring and Scholarship Program contributed administrative support. We thank all the residents, nurses, and patients who participated in the study.

REFERENCES

23. American College of Surgeons Division of Education. Position of the American Coll...
Resident Workload—Let’s Treat the Disease, Not Just the Symptom

Work compression is doing the same amount of work in fewer hours. The term is often used to describe an effect of the restriction of residents’ work hours by the Accreditation Council for Graduate Medical Education (ACGME). But before work hour limitations were implemented in 2003, residents were already experiencing work compression. From 1990 to 2010, annual admissions to major teaching hospitals increased by 46% (Katherine Brandenburg, Association of American Medical Colleges, written communication, December 11, 2012), while first-year residency positions, limited by restrictions in Graduate Medical Education funding, grew only 13%. During the same period, length of stay fell by almost one-third, and intensity of care per admission greatly increased. In short, by the time ACGME restrictions were implemented, residents were already doing much more, in less time and for more and sicker patients, than were previous generations of house staff.

Work hour limitations, by restricting the length of hospital shifts and spacing them, smooth out the peaks and valleys of work and ensure more regular sleep periods. But they do not change the total amount of work by residents in our teaching hospitals; they simply redistribute it.

Two serious problems remain. First, the current work hour limitations, which became more stringent in 2011, are too inflexible, particularly in the absolute restrictions they impose on the length of shifts (interns are restricted to 16-hour shifts, and other house staff are restricted to 24-hour shifts). This leads to overly complicated scheduling, which greatly increases handoffs, a known cause of errors. It also can prevent residents from taking advantage of educational opportunities, including seeing patients through critical times in their hospital course.

The second problem is more fundamental but receives less attention. Limiting work hours without commensurately decreasing workload exacerbates the already extreme work compression for residents. Some teaching hospitals have shifted some patient care to nonresident services or have hired more residents, but most have relied heavily on simply redistributing workload among existing residents, and wresting some additional patient care out of a mostly reluctant faculty. Residents still perform most of the work but are now racing the clock. In large part, long work hours before the work hour limitations reflected the choices residents made to handle their workload rather than specific scheduling requirements. In focusing on work hours rather than workload, the ACGME—under intense pressure from many groups, both within and outside medicine—treated the symptom, not the disease.

Why this conspicuous lack of attention to workload, despite the logic of dealing with it directly? One reason is that many teaching hospitals are heavily dependent on the work provided by residents. Shifting patients to nonresident services is more expensive than implementing work hour limitations, at least in direct costs. In fact, estimates of the costs of implementing work hour limitations assumed that not all of the work corresponding to lost hours would be transferred to other providers. Work compression makes changes in work hours affordable. The critical wager is that the gains from more alert residents will offset the losses from work compression and from decreased flexibility in scheduling and continuity of care.

Has the wager worked? Findings following the implementation of the 2003 work hour limitations, which established the 80-hour workweek, were inconclusive in terms of patient safety and educational outcomes. But residents and faculty consistently reported concerns about the effects on quality of care and education. Recently, studies have begun to evaluate the 2011 work hour limitations.

Desai et al., elsewhere in this issue of the journal, examined effects of the 2011 work hour limitations on sleep, education, and continuity and quality of care at The Johns Hopkins Hospital, Baltimore, Maryland. Residents were randomized to a schedule compliant with the 2003 work hour regulations or to 1 of 2 schedules compliant with the 2011 regulations. The authors found an increase in sleep during the on-call period with the 2011-compliant schedules and found more consistency in the duration of sleep. However, no difference was observed in average sleep per day. Handoffs more than doubled, and time for education decreased. Of most concern, residents and nurses reported a steep drop in the quality of care, so much so that one of the two 2011 schedules was abandoned before completion of the study.

The findings are consistent with those of other studies that have compared 2003 and 2011 schedules; these studies found that residents believed the quality of care and their education suffered. When the ACGME sur-