ORIGINAL ARTICLE

Effect on Patient Safety of a Resident Physician Schedule without 24-Hour Shifts

C.P. Landrigan, S.A. Rahman, J.P. Sullivan, E. Vittinghoff, L.K. Barger, A.L. Sanderson, K.P. Wright, Jr., C.S. O'Brien, S. Qadri, M.A. St. Hilaire, A.C. Halbower, J.L. Segar, J.K. McGuire, M.V. Vitiello, H.O. de la Iglesia, S.E. Poynter, P.L. Yu, P.C. Zee, S.W. Lockley, K.L. Stone, and C.A. Czeisler, for the ROSTERS Study Group*

ABSTRACT

BACKGROUND

The effects on patient safety of eliminating extended-duration work shifts for resident physicians remain controversial.

METHODS

We conducted a multicenter, cluster-randomized, crossover trial comparing two schedules for pediatric resident physicians during their intensive care unit (ICU) rotations: extended-duration work schedules that included shifts of 24 hours or more (control schedules) and schedules that eliminated extended shifts and cycled resident physicians through day and night shifts of 16 hours or less (intervention schedules). The primary outcome was serious medical errors made by resident physicians, assessed by intensive surveillance, including direct observation and chart review.

RESULTS

The characteristics of ICU patients during the two work schedules were similar, but resident physician workload, described as the mean (±SD) number of ICU patients per resident physician, was higher during the intervention schedules than during the control schedules (8.8±2.8 vs. 6.7±2.2). Resident physicians made more serious errors during the intervention schedules than during the control schedules (97.1 vs. 79.0 per 1000 patient-days; relative risk, 1.53; 95% confidence interval [CI], 1.37 to 1.72; P<0.001). The number of serious errors unitwide were likewise higher during the intervention schedules (181.3 vs. 131.5 per 1000 patient-days; relative risk, 1.56; 95% CI, 1.43 to 1.71). There was wide variability among sites, however; errors were lower during intervention schedules than during control schedules at one site, rates were similar during the two schedules at two sites, and rates were higher during intervention schedules than during control schedules at three sites. In a secondary analysis that was adjusted for the number of patients per resident physician as a potential confounder, intervention schedules were no longer associated with an increase in errors.

CONCLUSIONS

Contrary to our hypothesis, resident physicians who were randomly assigned to schedules that eliminated extended shifts made more serious errors than resident physicians assigned to schedules with extended shifts, although the effect varied by site. The number of ICU patients cared for by each resident physician was higher during schedules that eliminated extended shifts. (Funded by the National Heart, Lung, and Blood Institute; ROSTERS ClinicalTrials.gov number, NCT02134847.)

The authors' full names, academic degrees, and affiliations are listed in the Appendix. Address reprint requests to Dr. Landrigan at the Division of General Pediatrics, Department of Pediatrics, Boston Children's Hospital, 300 Longwood Ave., Boston, MA 02115, or at christopher.landrigan@childrens.harvard.edu.

*A list of the members of the ROSTERS Study Group is provided in the Supplementary Appendix, available at NEJM.org.

This article was updated on June 25, 2020, at NEJM.org.

N Engl J Med 2020;382:2514-23.
DOI: 10.1056/NEJMoa1900669

Since publication of a study in 1971 showing that sleep-deprived resident physicians made more errors in reading electrocardiograms,[1] a robust literature has accumulated indicating that sleep deprivation adversely affects the alertness and performance of resident physicians.[2-13] In a previous randomized, controlled trial, we found that resident physicians who worked according to a schedule that included frequent shifts of 24 or more consecutive hours (extended-duration work schedule) made 36% more serious medical errors than when they worked a schedule that cycled them through day and night shifts limited to no more than 16 consecutive hours.[11,13]

In recent years, policy regarding resident physician work hours has shifted. In 2008, the National Academy of Medicine recommended that resident physicians work no more than 16 consecutive hours without sleep.[14] In 2011, the Accreditation Council for Graduate Medical Education (ACGME) partially acted on this recommendation, prohibiting shifts exceeding 16 consecutive hours for first-year residents.[15] In 2017, the ACGME reversed its policy[16] and again began allowing shifts of 24 to 28 consecutive hours for all resident physicians after the FIRST (Flexibility in Duty Hour Requirements for Surgical Trainees) trial showed that no changes in the incidence of death or serious surgical complications were associated with shift limits among first-year surgical residents, although most of them spend a minority of their time in the operating room.[17] More recently, the iCOMPARE (Individualized Comparative Effectiveness of Models Optimizing Patient Safety and Resident Education) trial also showed no change in mortality among medical patients when shift limits were implemented,[18] although we believe that the power of the iCOMPARE trial was suboptimal.[19]

Questions remain as to why the duration of shifts for resident physicians appears to be a major driver of patient safety in some studies and inconsequential in others. Possibly, differing approaches to eliminating extended shifts (e.g., having resident physicians cycle through day and night shifts vs. having them work six consecutive night shifts) have differing effectiveness in promoting resident physician performance.[20] Alternatively, poorly managed transitions between shifts (known as handoffs) in some settings could undermine the potential benefits of reducing sleep deprivation in residents.[21-23] A third possibility is that reduced staffing levels[24] could counterbalance any benefit to patient safety of reduced work hours in some settings,[25] since, contrary to National Academy of Medicine recommendations,[14] the ACGME 2011 work-hour limits were not accompanied by firm workload limits or funding to support increased staffing.

To address these knowledge gaps, we conducted a multicenter, cluster-randomized, crossover trial of the effects on patient safety of implementing a rapidly cycling work roster that eliminated extended shifts. We concurrently captured data on resident physician work schedules, sleep, workload, and other systemic factors.[26,27]

METHODS

TRIAL DESIGN

The Randomized Order Safety Trial Evaluating Resident-Physician Schedules (ROSTERS) was a multicenter, cluster-randomized, crossover trial conducted from July 1, 2013, to March 5, 2017, in six pediatric intensive care units (ICUs) across the United States. Trial investigators obtained a certificate of confidentiality from the National Institutes of Health to protect the privacy of the participants, and institutional review board approval was granted. Detailed methods for the trial have been described previously.[28] We studied pediatric ICUs because medical errors occur at high rates in critical care settings, and the pediatric ICUs we included were staffed by resident physicians who were second-year and above and thus not subject to the ACGME's changing policies for first-year residents.[29,30]

To be considered for the trial, each participating pediatric ICU was required to have resident physicians who were following a schedule that included extended work shifts at baseline. The frequency of extended shifts varied across sites from every third shift (which required staying overnight in the hospital every fourth night) to every fourth shift (which required staying overnight in the hospital every fifth night); between extended shifts, resident physicians worked shorter day shifts and had occasional days off. This baseline schedule at each site served as the control schedule for our trial. Each ICU had an established handoff process in place at baseline (Table S1 in the Supplementary Appendix, avail-

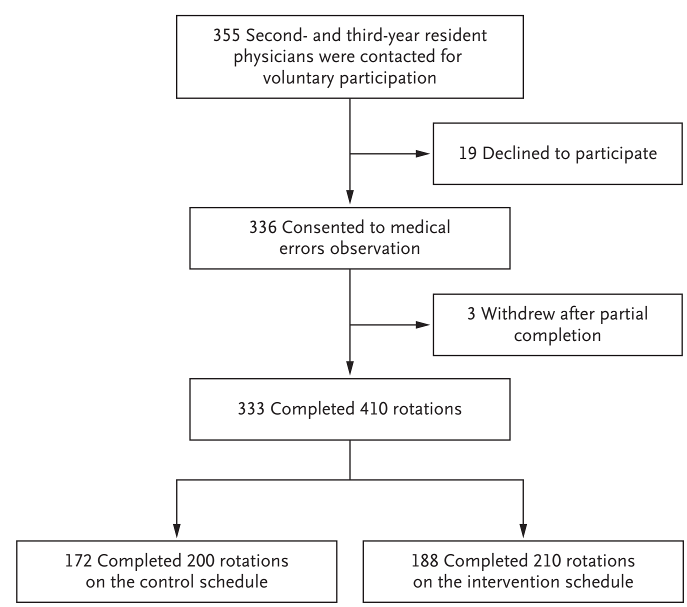


Figure 1. Participants and Rotations.

The control schedules included shifts of 24 hours or more; the intervention schedules eliminated extended shifts and cycled residents through day and night shifts of 16 hours or less. Of resident physicians who completed more than 1 rotation, 58 completed 2 rotations, 5 completed 3 rotations, and 3 completed 4 rotations. The 333 resident physicians who completed 410 rotations include 27 participants who rotated through at least one control and one intervention cycle.

able with the full text of this article at NEJM.org). All patients (except the subgroup of patients cared for during the day primarily by resident physicians working extended shifts) had their care handed off to resident physicians working extended shifts in the evening.

The trial was completed over several years, with each site beginning a 2-year trial at a different time. Sites were paired on the basis of the date they began the trial; one site from each pair was randomly assigned to start with the extended-shift schedule (control schedule), and the other site started with the schedule that eliminated extended shifts (intervention schedule). Each site had a 4-month wash-in interval before data collection began during which resident physicians followed the schedule that was about to be tested. Eight months of data were then collected on this schedule. This interval was followed by another 4-month wash-in interval during which sites crossed over to the other schedule. Then 8 months of data were collected on this second schedule. This design allowed each site to serve as its own control, matched by time of year.

INTERVENTION SCHEDULE DESIGN

During the intervention schedule, resident physicians typically worked a night shift followed by approximately 24 hours off duty, and then two or three consecutive day shifts (depending on the site); this pattern was repeated over the course of a month-long rotation, with occasional additional days off. Specific details about the schedule for each site have been reported previously (Table S2).[28] Our objective for the intervention schedule was to eliminate extended-duration (≥24 hours) work shifts and increase the amount of sleep for residents. Owing to substantial site-level differences in unit organization at baseline, sites made individual determinations about how best to organize staffing to accommodate the intervention.[28] All patients had their care handed off to night-shift resident physicians in the evening.

STUDY OVERSIGHT

Written informed consent was obtained from resident physicians for the collection of identifiable information (Fig. 1). Families of patients were informed that the trial was being conducted, and the institutional review boards waived informed consent for the collection of patient safety data. Data were reviewed on a regular basis by a data and safety monitoring board. A subgroup of resident physicians also gave written informed consent to provide data on their sleep, work hours, neurobehavioral performance (e.g., on the basis of psychomotor vigilance tasks), and subjective sleepiness, as reported previously.[26-28] The authors vouch for the accuracy and completeness of the data and for the fidelity of trial to the protocol, available at NEJM.org.

COLLECTION OF DATA AND CATEGORIZATION OF SERIOUS MEDICAL ERRORS

We used an intensive data collection and adjudication method to capture and classify adverse events and medical errors.[28] This method was used in our earlier trial of resident physician work schedules and patient safety[11] and was adapted from a well-tested approach used in multiple studies.[31-34] Categories of errors and adverse events are described in Table 1.

At each hospital, a team of chart reviewers (nurses) and observers (physicians) who were centrally trained through a series of webinars collected data, which were supplemented by voluntary reports from clinical staff. The team of physician observers followed participating resident physicians around the clock during each schedule, gathering information on any suspected serious errors. Concurrently, research nurses performed chart reviews (generally 5 days a week, with Monday reviews including charts from the weekend) and gathered reports of incidents of suspected serious errors from clinical staff. Incidents were classified as being attributable to resident physicians or to other staff.

Data were collected on electronic forms and securely transferred to the trial data coordinating center. Subsequently, data on all suspected incidents were sent to trained physician reviewers who were unaware of site and schedule and who independently classified each suspected incident (Table 1). Two physicians independently rated each suspected incident, classifying it as an adverse event, near miss, error with little or no potential for harm, or excluded event. Adverse events were further classified according to preventability with the use of a 4-point Likert scale; events were subsequently dichotomized to preventable or nonpreventable incidents. Disagreements were resolved by discussion; pre-discussion interrater reliability was good (weighted kappa score, 0.52 to 0.67).

Table 1. Classification of Errors and Adverse Events.

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 (i.e., the sum of preventable adverse events plus near misses). Errors with little or no potential for harm are not serious errors, nor are nonpreventable adverse events.
Adverse event	Any injury due to medical management
Nonpreventable adverse event	Injury caused by medical care, without any apparent error
Preventable adverse event	Injury caused by an error in medical management
Near miss	An error in care that has substantial potential to cause harm but does not, either because it is intercepted or because it unexpectedly causes no apparent harm despite reaching the patient
Error with little or no potential for harm	An error in care delivery that is unlikely to injure a patient
Exclusion	An incident detected on initial surveillance that is determined on review to be neither an error nor an adverse event

PATIENTS PER RESIDENT PHYSICIAN

We obtained work rosters for each site. Average hourly resident physician staffing for a 24-hour interval was derived from these rosters, from which an average estimate of daily staffing by resident physicians at each site and for each schedule was determined. The number of ICU patients per resident physician for each site–schedule combination was calculated as the average of the estimates of daily patient census at each site per schedule divided by the average number of resident physicians present daily at each site per schedule.

STATISTICAL ANALYSIS

The unit of analysis for our primary analysis was the rate of serious medical errors (preventable adverse events and near misses) made by resident physicians per admission. In accordance with the prespecified statistical analysis plan, we compared the rates of serious medical errors during one schedule with those during the other schedule using log-link Poisson models, with patient admission to the pediatric ICU as the unit of analysis; with site, period of randomization, and schedule as fixed effects; with robust standard errors to account for potential overdispersion; and with the log of adjusted patient-days at risk as an offset. All patients in the participating units were included in the analysis; there were no dropouts. Adjusted patient-days at risk were estimated, with exclusion of shifts that were not observed, although estimates that did not exclude missed shifts (in sensitivity analyses) were essentially unchanged. Rates are presented as numbers of medical errors per 1000 adjusted patient-days at risk during the two schedules. Secondary outcomes included rates of unitwide serious medical errors. Overall rates as well as site-specific rates are reported. A two-tailed P value (with $P<0.05$ considered to indicate statistical significance) is reported for the primary outcome in the primary analyses. There was no prespecified plan to account for multiple comparisons; for all analyses other than the primary analyses, point estimates and 95% con-

Table 2. Patient and ICU Characteristics.*

Characteristic	Control Schedule	Intervention Schedule
Patients — no.	3,267	3,310
ICU admissions — no.	3,508	3,591
Patient-days — no.	18,749	20,072
Age — yr	7.3±6.7	7.1±6.6
Male sex — no./total no. (%)	1853/3508 (52.8)	1943/3591 (54.1)
Median length of stay (IQR) — days	2 (2–5)	2 (2–5)
Median chronic condition indicator (IQR)†	2 (1–4)	2 (1–4)
ICU patients per resident physician — no.‡	6.7±2.2	8.8±2.8

* Plus–minus values are means ±SD. The control schedule included shifts of 24 hours or more. The intervention schedule eliminated extended shifts and cycled resident physicians through day and night shifts of 16 hours or less. ICU denotes intensive care unit, and IQR interquartile range.

† The chronic condition indicator is a marker of a patient's coexisting conditions, derived from administrative billing codes. Higher numbers indicate the presence of more coded chronic conditions.[35]

‡ The number of ICU patients per resident physician is calculated as the average census of patients at each site during each schedule divided by the average number of resident physicians present at each site during each schedule.

fidence intervals are reported without P values. Confidence intervals have not been adjusted for multiple comparisons, and inferences drawn from them may not be reproducible.

We conducted secondary analyses comparing the rates of medical errors during the two schedules in which we adjusted for the number of patients per resident physician as a potential confounder, because the number of patients per resident physician was unbalanced between the trial groups. In these analyses, resident physician rotation was used as the unit of analysis, with the log of the length of resident physician rotation as an offset, since this analysis accounted for varying lengths of individual residents' rotations when we adjusted for workload as a potential confounder. To assess the effects of these potential confounders, we used log-link Poisson regression with robust standard errors. The model included linear and quadratic terms for number of patients per resident physician and for site, schedule, and period. We also assessed variation in the number of patients per resident physician by site and schedule and conducted post hoc analyses to further explore site-related and workload-related effects (Fig. S1).

RESULTS

CHARACTERISTICS OF SHIFTS

In total, 38,821 patient-days (18,749 in the control schedule with extended shifts and 20,072 in the intervention schedule with extended shifts eliminated) were studied, representing 7099 admissions (3508 and 3591, respectively). Resident physicians consented to be directly observed for patient safety data during 413 of 432 rotations (a total of 72,102 hours of observation).

Patient characteristics varied among hospitals but were generally similar between the two schedules (Table 2; site-specific data are shown in Table S3). Unit characteristics differed between schedules; specifically, the mean (±SD) number of patients per resident physician was higher during the intervention schedules than during the control schedules (8.8±2.8 vs. 6.7±2.2) (Table 2).

As reported previously,[27] residents' mean weekly work hours were lower during the intervention schedule than during the control schedule (61.9±4.8 hours vs. 68.4±7.4 hours), and mean weekly sleep hours were greater (52.9±6.0 hours vs. 49.1±5.8 hours). The percentage of 24-hour intervals with fewer than 4 hours of sleep was 25% in the control group and 9% in the intervention group.

SERIOUS MEDICAL ERRORS

Resident physicians made significantly more serious medical errors during the intervention schedules (without extended shifts) than during the control schedules (with extended shifts) (1723 vs. 1268; unadjusted rates, 97.1 vs. 79.0 per 1000 patient-days at risk; adjusted relative risk, 1.53 [95% confidence interval {CI}, 1.37 to 1.72]; P<0.001) (Fig. 2). There were wide discrepancies in the effect of the intervention across sites (Fig. 2). At three sites, resident physicians made more serious errors during the intervention schedule than during the control schedule (adjusted relative risk, 1.51, 2.38, and 5.90); at two sites, there was no difference; and at one site, resident physicians made fewer serious errors during the intervention schedule (adjusted relative risk, 0.24).

INCIDENCE OF ERRORS UNITWIDE

The unitwide incidence of serious errors (including those that involved resident physicians and those that did not) was higher during the intervention schedule than during the control schedule (unadjusted rates, 181.3 vs. 131.5 per 1000 patient-days at risk; adjusted relative risk, 1.56 [95% CI, 1.43 to 1.71]) (Fig. 2). There was wide variability in the incidence of serious errors at the site level (Fig. 2).

RELATIONSHIP BETWEEN WORKLOAD AND PATIENT SAFETY

Wide site-level variability existed in the number of patients per resident physician at baseline, and the degree of change in the number of patients per resident physician with implementation of the intervention schedule also varied among sites. A secondary analysis with resident physician rotation as the unit of analysis and with adjustment for the number of patients per resident physician as a continuous variable showed that the relative risk of a serious error during the intervention schedule as compared with the control schedule was 0.54 (95% CI, 0.35 to 0.85). However, when the number of patients per resident physician was included as a categorical variable, in quartiles and thirds, the relative risk estimate was 0.74 and 1.32, respectively, in the statistical model, which suggests instability of the model. However, in these secondary analyses, there was a substantial interaction between schedule and workload variables, making interpretation of results difficult. In additional post hoc analyses, we observed that at the three sites with the highest number of patients per resident physician at baseline (i.e., with the control schedule), the incidence of medical errors worsened when intervention schedules were implemented; conversely, at the site with the lowest number of patients per resident physician at baseline, the incidence of medical errors declined when the intervention schedule was implemented (Fig. S1A). Rates of serious errors made by resident physicians increased with increasing numbers of patients per resident physician (Fig. S1B).

DISCUSSION

Contrary to our hypothesis, introduction of a schedule that eliminated extended shifts for resident physicians in six pediatric ICUs was associated with a significant increase in the rates of serious medical errors. There was substantial site-level variability in the effect of the intervention, however, with three sites having higher incidents of serious medical errors with the schedule that eliminated extended shifts (the intervention schedule) than with the extended-shift schedule (control schedule), one site having fewer serious medical errors with the intervention schedule, and two others having no significant difference in the incidence of serious medical errors between the two schedules. These data were not explained by differences in the demographics or complexity of illness of the patients. However, we noted that hospitals with the highest resident physician workloads had the most negative results with the intervention. Secondary analyses suggested that the results might have been confounded by concurrent increases in workload with the intervention, although this finding should be viewed as exploratory.

Our trial builds on a growing literature evaluating the effects of eliminating extended shifts. Our previous randomized trial[11-13] showed a benefit of eliminating extended shifts, as did a systematic review.[36] The more recent FIRST and iCOMPARE trials, by contrast, showed no benefit.[17,18] The FIRST trial, involving surgical programs, did not standardize the manner in which hospitals implemented schedule changes, which made the effects of any particular approach to scheduling unknown. In addition, programmatic data on resident physician workload, patient census, and other variables were not gathered. The iCOMPARE trial, in which internal medicine programs were randomly assigned to allow or prohibit extended shifts, likewise did not specify an approach to eliminating extended shifts.[18]

Our current trial adds to this literature in

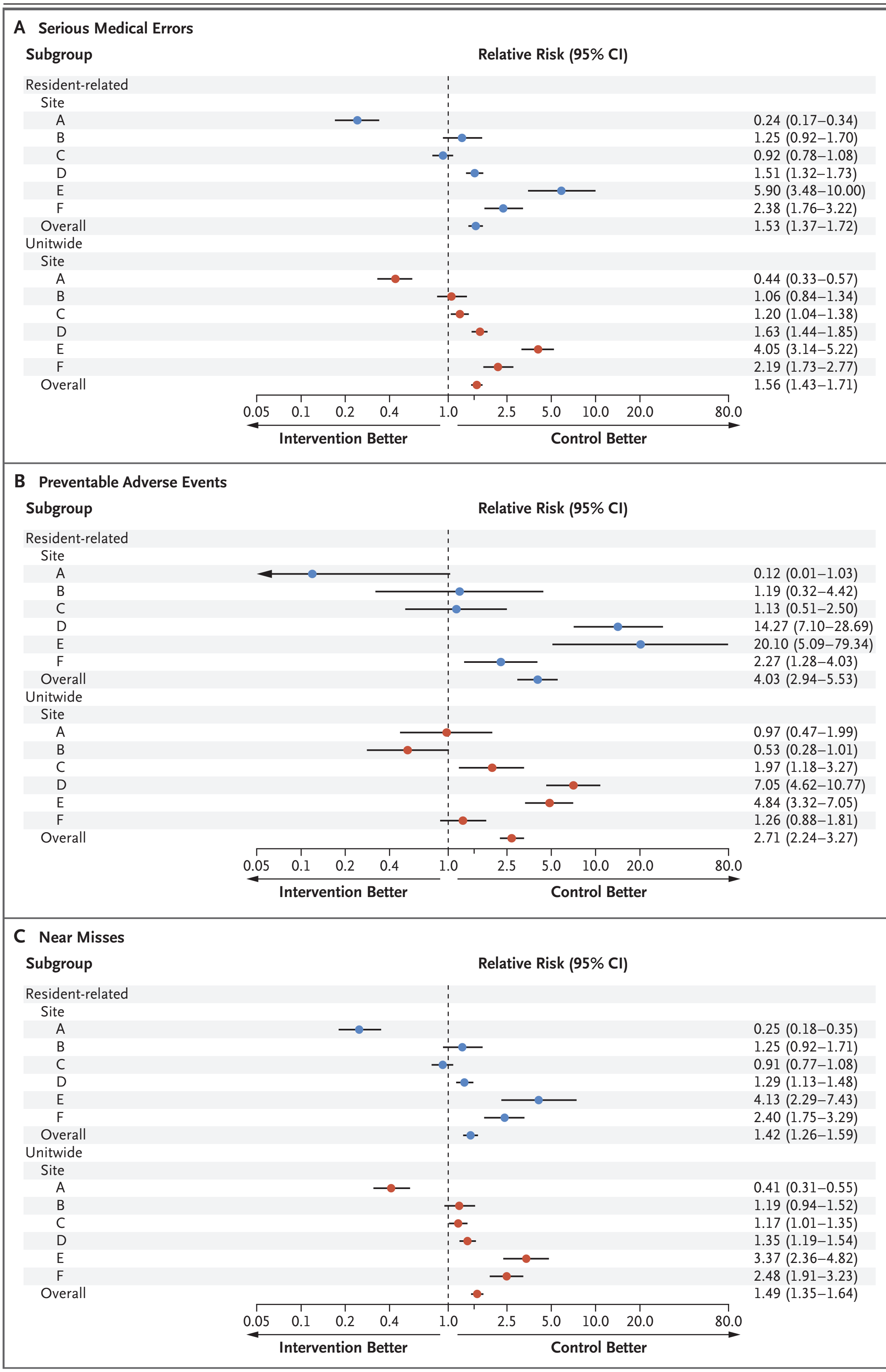

A Serious Medical Errors
Subgroup
Relative Risk (95% CI)
Resident-related
Site
A 0.24 (0.17–0.34)
B 1.25 (0.92–1.70)
C 0.92 (0.78–1.08)
D 1.51 (1.32–1.73)
E 5.90 (3.48–10.00)
F 2.38 (1.76–3.22)
Overall 1.53 (1.37–1.72)
Unitwide
Site
A 0.44 (0.33–0.57)
B 1.06 (0.84–1.34)
C 1.20 (1.04–1.38)
D 1.63 (1.44–1.85)
E 4.05 (3.14–5.22)
F 2.19 (1.73–2.77)
Overall 1.56 (1.43–1.71)
0.05 0.1 0.2 0.4 1.0 2.5 5.0 10.0 20.0 80.0
Intervention Better
Control Better
B Preventable Adverse Events
Subgroup
Relative Risk (95% CI)
Resident-related
Site
A 0.12 (0.01–1.03)
B 1.19 (0.32–4.42)
C 1.13 (0.51–2.50)
D 14.27 (7.10–28.69)
E 20.10 (5.09–79.34)
F 2.27 (1.28–4.03)
Overall 4.03 (2.94–5.53)
Unitwide
Site
A 0.97 (0.47–1.99)
B 0.53 (0.28–1.01)
C 1.97 (1.18–3.27)
D 7.05 (4.62–10.77)
E 4.84 (3.32–7.05)
F 1.26 (0.88–1.81)
Overall 2.71 (2.24–3.27)
0.05 0.1 0.2 0.4 1.0 2.5 5.0 10.0 20.0 80.0
Intervention Better
Control Better
C Near Misses
Subgroup
Relative Risk (95% CI)
Resident-related
Site
A 0.25 (0.18–0.35)
B 1.25 (0.92–1.71)
C 0.91 (0.77–1.08)
D 1.29 (1.13–1.48)
E 4.13 (2.29–7.43)
F 2.40 (1.75–3.29)
Overall 1.42 (1.26–1.59)
Unitwide
Site
A 0.41 (0.31–0.55)
B 1.19 (0.94–1.52)
C 1.17 (1.01–1.35)
D 1.35 (1.19–1.54)
E 3.37 (2.36–4.82)
F 2.48 (1.91–3.23)
Overall 1.49 (1.35–1.64)
0.05 0.1 0.2 0.4 1.0 2.5 5.0 10.0 20.0 80.0
Intervention Better
Control Better

Figure 2 (facing page). Serious Errors, Adverse Events, and Near Misses by Site and Schedule.
The control schedules included shifts of 24 hours or more; the intervention schedules eliminated extended shifts and cycled resident physicians through day and night shifts of 16 hours or less. The relative risk is for the intervention schedule as compared with the control schedule. Panel A shows the relative risk of serious medical errors, both resident-physician–related (primary outcome) and unitwide, Panel B the relative risk of preventable adverse events, and Panel C the relative risk of near misses.

several respects. We focused on a particular approach to the intervention scheduling, which cycled resident physicians through day and night shifts. Patient safety worsened under this schedule. However, we concurrently collected detailed data that allowed us to explore possible reasons for this.

We found that our intervention led to a decrease in weekly work hours and an increase in residents' hours of sleep.[27] In addition, as reported elsewhere, we observed an improvement in residents' neurobehavioral performance,[26] and poorer neurobehavioral performance has been correlated with a higher risk of serious medical errors.[26] Since sleep and neurobehavioral performance improved on the intervention schedule as expected, it appears unlikely that the worsening in patient safety was due to worsening fatigue on this schedule.

A possible explanation for the deterioration in patient safety despite improvements in sleep and neurobehavioral performance is the increase in handoff frequency across sites. The number of patients whose care was handed off each evening increased at all six sites during the intervention schedule. However, only three sites had worse patient safety outcomes with the intervention schedule than with the extended-shift schedule, and one had substantially better safety outcomes with the intervention, which suggests that the increase in handoffs overall was unlikely to account for our results. Moreover, in our previous trial, safety improved after extended shifts were eliminated, despite increased handoffs.[11] It is possible that handoff processes at some sites might have protected against degradations in safety more effectively than the processes at other sites, but no obvious trends were apparent to support this possibility (Table S1).

Increases in resident physician workload that occurred as programs eliminated 24-hour shifts could account for our findings. There is evidence that when ICU physicians care for more than seven patients per day, patient safety may deteriorate.[37] In our previous trial, in which a schedule eliminating extended work shifts (intervention schedule) was shown to be beneficial, an additional resident physician was added to the roster in trial units during months with the intervention schedule (i.e., four resident physicians were in the units during the intervention schedule vs. three during the control schedule), in order to keep the daily workload for resident physicians constant as each resident's average work hours decreased.[11-13] By contrast, in the current trial, resident physician workload increased overall when the intervention schedule was introduced. In secondary analyses that controlled for the increase in workload, we did not observe increases in errors during the intervention schedule. However, we did not set out to explicitly test the effects of workload on our intervention.

This trial has several limitations. First, although our methods for collecting data on medical errors are well established, measuring and classifying medical errors is an imperfect science. Our primary data collectors were aware of the residents' schedules. We provided all primary data collectors standardized training to minimize bias and variability in data collection. In addition, all final incident classification was made at a second stage by two independent physicians who were unaware of site and schedule and who classified with good reliability. Despite these measures, some variability in data collection may have occurred across sites, but we believe that this is unlikely to account for our main findings.

Second, although our results suggest that variability in workload may have influenced the intervention, other site-level factors (e.g., unmeasured differences in handoff processes and attending physicians' supervision or performance) may have influenced these findings. Our workload findings should be viewed as exploratory and tested further in future research, although they raise the possibility that the debate currently playing out in some states regarding health care provider–patient ratios may be germane to physicians as well as to nurses.[38]

Finally, we studied the effects on patient safety of a specific work schedule in pediatric ICUs. Although our findings may be relevant to other settings, particularly other ICUs, generalizability is uncertain. We found that local systems of care and variation in implementation had a substantial effect on the effectiveness of the intervention schedule.

In this multicenter trial, incidents of harmful medical errors by resident physicians were higher during an intervention schedule that eliminated extended work shifts than during a schedule that included shifts of 24 hours or more. However, the intervention schedule also increased residents' workload. Residents' sleep and neurobehavioral performance improved with the intervention,[26,27] as we expected. A decade ago, the National Academy of Medicine[14] recommended that resident physician work-hour reduction should not occur without an investment of resources to support adequate staffing and infrastructure. Excessive work hours degrade patient safety, but so too do excessive workloads and poor handoffs. The results of our trial suggest that future interventions to address the persistent patient safety problems in academic health centers must address and rigorously evaluate all these challenges concurrently.

A data sharing statement provided by the authors is available with the full text of this article at NEJM.org.

Supported by grants (U01-HL-111478 and U01-HL-111691) from the National Heart, Lung, and Blood Institute. Drs. Barger, Lockley, and Czeisler were supported in part by a grant (R01-OH-010300) from the National Institute of Occupational Safety and Health.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

APPENDIX

The authors' full names and academic degrees are as follows: Christopher P. Landrigan, M.D., M.P.H., Shadab A. Rahman, Ph.D., M.P.H, Jason P. Sullivan, B.S., Eric Vittinghoff, Ph.D., Laura K. Barger, Ph.D., Amy L. Sanderson, M.D., Kenneth P. Wright, Jr., Ph.D., Conor S. O'Brien, B.A., Salim Qadri, B.S., Melissa A. St. Hilaire, Ph.D., Ann C. Halbower, M.D., Jeffrey L. Segar, M.D., John K. McGuire, M.D., Michael V. Vitiello, Ph.D., Horacio O. de la Iglesia, Ph.D., Sue E. Poynter, M.D., M.Ed., Pearl L. Yu, M.D., Phyllis C. Zee, M.D., Ph.D., Steven W. Lockley, Ph.D., Katie L. Stone, Ph.D., and Charles A. Czeisler, M.D., Ph.D.

The authors' affiliations are as follows: the Division of Sleep and Circadian Disorders, Departments of Medicine and Neurology, Brigham and Women's Hospital (C.P.L., S.A.R., J.P.S., L.K.B., C.S.O., S.Q., M.A.S.H., S.W.L., C.A.C.), the Division of Sleep Medicine, Harvard Medical School (C.P.L., S.A.R., L.K.B., M.A.S.H., S.W.L., C.A.C.), and the Division of General Pediatrics, Department of Pediatrics (C.P.L.), and the Division of Critical Care Medicine, Department of Anesthesiology, Critical Care, and Pain Medicine (A.L.S.), Boston Children's Hospital — all in Boston; the University of California, San Francisco (E.V., K.L.S.), and California Pacific Medical Center Research Institute (K.L.S.), San Francisco; the Sleep and Chronobiology Laboratory, Department of Integrative Physiology, University of Colorado Boulder, Boulder (K.P.W.), and Children's Hospital Colorado, University of Colorado School of Medicine, Aurora (A.C.H.); the University of Iowa Stead Family Children's Hospital, Iowa City (J.L.S.); Seattle Children's Hospital (J.K.M.) and the University of Washington (M.V.V., H.O.I.), Seattle; Cincinnati Children's Hospital Medical Center, University of Cincinnati, Cincinnati (S.E.P.); University of Virginia Children's Hospital, Charlottesville (P.L.Y.); and the Department of Neurology and Center for Circadian and Sleep Medicine, Northwestern University, Feinberg School of Medicine, Chicago (P.C.Z.).

REFERENCES

1. Friedman RC, Bigger JT, Kornfeld DS. The intern and sleep loss. N Engl J Med 1971;285:201-3.
2. Philibert I. Sleep loss and performance in residents and nonphysicians: a meta-analytic examination. Sleep 2005;28:1392-402.
3. 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.
4. 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.
5. Gaba DM, Howard SK. Patient safety: fatigue among clinicians and the safety of patients. N Engl J Med 2002;347:1249-55.
6. Veasey S, Rosen R, Barzansky B, Rosen I, Owens J. Sleep loss and fatigue in residency training: a reappraisal. JAMA 2002;288:1116-24.
7. Weinger MB, Ancoli-Israel S. Sleep deprivation and clinical performance. JAMA 2002;287:955-7.
8. Samkoff JS, Jacques CH. A review of studies concerning effects of sleep deprivation and fatigue on residents' performance. Acad Med 1991;66:687-93.
9. Pilcher JJ, Huffcutt AI. Effects of sleep deprivation on performance: a meta-analysis. Sleep 1996;19:318-26.
10. 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.
11. Landrigan CP, Rothschild JM, Cronin JW, et al. Effect of reducing interns' work hours on serious medical errors in intensive care units. N Engl J Med 2004;351:1838-48.
12. Barger LK, Cade BE, Ayas NT, et al. Extended work shifts and the risk of motor vehicle crashes among interns. N Engl J Med 2005;352:125-34.
13. 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.
14. Institute of Medicine. Resident duty hours: enhancing sleep, supervision, and safety. Washington, DC: National Academies Press, 2009.
15. Accreditation Council for Graduate Medical Education. Common program requirements — effective: July 1, 2011 (https://www.acgme.org/Portals/0/PDFs/Common_Program_Requirements_07012011[2].pdf).
16. Accreditation Council for Graduate Medical Education. Common program requirements — effective: July 1, 2017 (https://www.acgme.org/Portals/0/PFAssets/ProgramRequirements/CPRs_2017-07-01.pdf).
17. Bilimoria KY, Chung JW, Hedges LV, et al. National cluster-randomized trial of duty-hour flexibility in surgical training. N Engl J Med 2016;374:713-27.
18. Silber JH, Bellini LM, Shea JA, et al.

Patient safety outcomes under flexible and standard resident duty-hour rules. N Engl J Med 2019;380:905-14.
19. Landrigan CP, Czeisler CA. Patient safety under flexible and standard duty-hour rules. N Engl J Med 2019;380:2379-80.
20. Folkard S, Lombardi DA. Modeling the impact of the components of long work hours on injuries and "accidents." Am J Ind Med 2006;49:953-63.
21. 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.
22. Cooper JB, Long CD, Newbower RS, Philip JH. Critical incidents associated with intraoperative exchanges of anesthesia personnel. Anesthesiology 1982;56:456-61.
23. 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.
24. Chua KP, Gordon MB, Sectish T, Landrigan CP. Effects of a night-team system on resident sleep and work hours. Pediatrics 2011;128:1142-7.
25. Needleman J, Buerhaus P, Pankratz VS, Leibson CL, Stevens SR, Harris M. Nurse staffing and inpatient hospital mortality. N Engl J Med 2011;364:1037-45.
26. Rahman SA, Sullivan JP, Barger LK, et al. Effect on neurobehavioral performance of eliminating extended duration work shifts and implementing a rapidly cycling work roster in the Randomized Order Safety Trial Evaluating Resident-physician Schedules (ROSTERS). Presented at SLEEP 2019, June 8–12, 2019. abstract.
27. Barger LK, Sullivan JP, Blackwell T, et al. Effects on resident work hours, sleep duration, and work experience in a Randomized Order Safety Trial Evaluating Resident-physician Schedules (ROSTERS). Sleep 2019;42(8):zsz110.
28. Blackwell T, Kriesel DR, Vittinghoff E, et al. Design and recruitment of the Randomized Order Safety Trial Evaluating Resident-physician Schedules (ROSTERS) study. Contemp Clin Trials 2019;80:22-33.
29. Rothschild JM, Landrigan CP, Cronin JW, et al. The Critical Care Safety Study: the incidence and nature of adverse events and serious medical errors in intensive care. Crit Care Med 2005;33:1694-700.
30. Kaushal R, Bates DW, Franz C, Soukup JR, Rothschild JM. Costs of adverse events in intensive care units. Crit Care Med 2007;35:2479-83.
31. Bates DW, Boyle DL, Vander Vliet MB, Schneider J, Leape L. Relationship between medication errors and adverse drug events. J Gen Intern Med 1995;10:199-205.
32. 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.
33. Starmer AJ, Spector ND, Srivastava R, et al. Changes in medical errors after implementation of a handoff program. N Engl J Med 2014;371:1803-12.
34. Khan A, Coffey M, Litterer KP, et al. Families as partners in hospital error and adverse event surveillance. JAMA Pediatr 2017;171:372-81.
35. Berry JG, Ash AS, Cohen E, Hasan F, Feudtner C, Hall M. Contributions of children with multiple chronic conditions to pediatric hospitalizations in the United States: a retrospective cohort analysis. Hosp Pediatr 2017;7:365-72.
36. Levine AC, Adusumilli J, Landrigan CP. Effects of reducing or eliminating resident work shifts over 16 hours: a systematic review. Sleep 2010;33:1043-53.
37. Gershengorn HB, Harrison DA, Garland A, Wilcox ME, Rowan KM, Wunsch H. Association of intensive care unit patient-to-intensivist ratios with hospital mortality. JAMA Intern Med 2017;177:388-96.
38. Aiken LH, Sloane DM, Cimiotti JP, et al. Implications of the California nurse staffing mandate for other states. Health Serv Res 2010;45:904-21.

Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Landrigan CP, Rahman SA, Sullivan JP, et al. Effect on patient safety of a resident physician schedule without 24-hour shifts. N Engl J Med 2020;382:2514-23. DOI: 10.1056/NEJMoa1900669

Table of Contents

ROSTERS STUDY GROUP

Clinical Coordinating Center: Laura K. Barger, PhD; Charles A. Czeisler, PhD, MD; Melissa A. St. Hilaire, PhD; Elizabeth B. Klerman, MD, PhD; Christopher P. Landrigan, MD, MPH*; Steven W. Lockley, PhD; Conor S. O'Brien, BA; Andrew J.K. Phillips, PhD; Salim Qadri, BS; Shadab A. Rahman, PhD, MPH; Jason P. Sullivan, BS; and Natalie C. Viyaran, BS.

Data Coordinating Center: Terri Blackwell, MA; Dana R. Kriesel, MPH, MS; and Katie L. Stone, PhD**.**

Colorado: Angela S. Czaja, MD; Ann C. Halbower, MD; Adam Rosenberg, MD; and Kenneth P. Wright Jr, PhD.

Iowa: Gretchen Cress, RN, MPH; Gwen E. Erkonen, MD, MEd; and Jeffrey L. Segar, MD.

Massachusetts: Lindsey B. Armstrong, MD; Ben D. Albert, MD; Erin A. Bressler, MD; Dennis Daniel, MD; Christopher P. Landrigan, MD, MPH*; Bradley S. Podd, MD, PhD; Amy L. Sanderson, MD; Theodore C. Sectish, MD; Patrick A. Upchurch, MD; and Traci A. Wolbrink, MD, MPH.

Ohio: Sue E. Poynter, MD, MEd

Virginia: Jeannean Carver, MD and Pearl L. Yu, MD.

Washington: Maneesh Batra, MD, MPH; Reid W.D. Farris, MD, MS; Horacio O. de la Iglesia, PhD; John K. McGuire, MD; and Michael V. Vitiello, PhD.

Other: Phyllis C. Zee, MD, PhD

*Dr. Christopher Landrigan fulfilled two roles: ROSTERS Study Multiple Principal Investigator (with Dr. Charles Czeisler) of the Clinical Coordinating Center and Site Principal Investigator at Boston Children's Hospital.

ACKNOWLEDGMENTS

We thank the Data Safety and Monitoring Board members for their oversight: Donald L. Bliwise, PhD; Barry Markovitz, MD, MPH; Eva Petkova, PhD; Wasima N. Rida, PhD; and Ramesh C. Sachdeva, MD, PhD, DBA.

We thank the National Heart, Lung, and Blood Institute for their support: Carol J. Blaisdell, MD; Peyvand Ghofrani, MDE, CCRA; Lora A. Reineck, MD, MS; Robert A. Smith, PhD, FCCM; Michael Twery, PhD; Gail G. Weinmann, MD; and Colin O. Wu, PhD.

Thank you also to the resident physicians, attending physicians, nurses, and clinical pharmacists of the participating Pediatric Intensive Care Units for their ongoing support.

Thank you to the following ROSTERS team members:

Clinical Coordinating Center: Justin D. Buie, BS; and Joshua T. Stephens, BS.

Data Coordinating Center: Lynn Harvey, BS and Vicki Li, BS.

Colorado: Bradley Brainard, MBA, MHA, PMP; Tristan Bakke Dear, MS, MD; Tondeleyo Gonzalez, MA, BSN, RN; Jonathan D. Haywood, MD, MPH; Heather Hoch, MD, MSCS; Brian M. Jackson, MD, MA; Ayoub Lahlou, MD, MBA; Kathryn J. Yucha, MSN, BSN; Karen Meyer, RN, BSN, CPN; Tolulope Oyewumi, MD, MPH; Kimberly Ralston, MPH, BSN, RN; Nabeel Sawaged, MD; Beth E. Smith, MSN, RN, WHNP-BC; and Vanitha K. Varre, MBBS, MPH, MPS.

Iowa: Safa Abukhalil, MD; Ihab Ahmed, MBBS, MPH; Safa Abdelwahid Mohamed Ahmed, MBBS; Shilpa C. Balikai, DO; Maria Ana C. Canaya-Voskov, MD; Janice M. Jeter, RN, CCRC; Sameer Kamath, MBBS; Crystal Tuley, RN; Jessica G. Moreland, MD; Vani C. Movva, MBBS; Geoffrey Ounda Obel, MPH, MD; Angie Platt, BSN, RN; Thomas D. Scholz, MD; Ruthann Schrock, BSN, RN; Amy Stier, MD, MME; Alexandra Paige Davis Volk, MD; and Jin Zhou, RN, MNHP.

Massachusetts: Oluwafunmilola Alabi, MD; Joseph Asemota, MD, MPH; Abimbola Chris-Olaiya, MD, MPH; Virginia Leon, BSN, MEd, CCRN; Alexandra Male, MPH; Siyu Ma, MS; Adeolu O. Oladunjoye, MBChB, MPH; Olubunmi O. Oladunjoye, MBBS, MPH; Saki Onda, MD, MPH; Kimberly Ralston, MPH, BSN, RN; Bhavya Atul Shah, MBBS; Lisa Tse, MPH; and Sandra Wooldridge, BSN, RN.

Ohio: Juanita Dudley, BSN, RN; Tatiana Elson, BS; Narinderpal Kaur, MD; Samuel Lee, MD, MBA; Najima Mwase, MD; Narissa Puran, MD, MS, MPH; Ndidi Unaka, MD, MEd; Andrew M. Warner, MD; Robin Widing, RN, MSN, CCRP; and Hector R. Wong, MD.

Virginia: Indu Aggarwal, MD**;** Fatimah Begom, MBBS**;** Kateryna Bilanovych, MD, ND, PhD, BCMAS**;** Ashley C. Eason, MD; Nasir Farhat, MD; Nicole M. Frank, PA-C; Robin L. Kelly, BSN, RN; Evan B. Kudron, MD; Abigail V.W. Kumral, MD; Brock D. Libby, MD; Jules Mukunde Katotola, MD; Trevor Pollock, MD; Justin Rizer, MD; Isaac A. Shields, MD; Terrell D. Smith, BS; Carolyn Spilman, MS, RN, CPNP; Albert T. Tang, MD; Linlin Wang, MD, PhD; Weonpo Yarl, MD; Hong Zhu, RN, CCRC; and Jenna V. Zschaebitz, BS.

Washington: Mouammar M. Abouagila, MBBS; Ibrahim K. Abukenda, MBChB, MPH; Canan Akture, MD, CCRC; Jennifer Jane Gile, BS; Carol Mendivil, MD; Anas L. Najjar, MS; Gowri Rajendran, MBBS; Shahar Robinzon, MD, MSc; Erin M. Sullivan, MPH; and Nastassya West, BS.

Other: David Gozal, MD, MBA; Leila Kheirandish-Gozal, MD, MSC; and Sharon M. Unti, MD.

SUPPLEMENTAL TABLE 1. Evening Handoff Procedures by Site

Site	Systems or problem-based approach?	Consistent use of any organizing framework?	Supervised by fellow or attending?	Any change with intervention schedule?
A	systems-based	no	yes	no
B	systems-based	I-PASS*	yes	no
C	systems-based	no	yes	yes, PM handoff staggered to occur in two parts to accommodate fellow and resident schedules
D	systems-based	no	yes	no
E	systems-based	I-PASS*	yes	no
F	systems-based	I-PASS*	yes	no

**Illness severity; Patient summary: Action List; Situational awareness and contingency planning; Synthesis by receiver*

SUPPLEMENTAL TABLE 2. Summary of Shifts on Control Schedule vs. Intervention

	Control Schedule*	**Intervention Schedule†**
Day shift	6-7am through 3-7pm (8-14 hours)	6-7am through 5-9pm (11-14 hours)
Night shift	n/a	6-8pm through 8am-12pm (16 hours)
Extended shift	6-11am through 8-10am (24-28 hours)	n/a

*on the Control Schedule, 2-3 day shifts typically preceded an extended shift, with occasional days off built into the schedule

†on the Intervention Schedule, 2-3 day shifts typically preceded a night shift, with occasional days off built into the schedule

SUPPLEMENTAL TABLE 3. Patient Population and Unit Characteristics by Site and Schedule

	Site A		Site B		Site C	
Characteristic (site difference comparison)	Control	Intervention	Control	Intervention	Control	Intervention
Number of patients	545	754	382	353	586	537
Number of unit admissions	547	754	395	386	627	579
Number of patient-days	2674	3480	2195	2096	3451	3323
age, yr, mean ± SD	7.1 ± 6.0	6.5 ± 6.3	9.2 ± 7.4	8.8 ± 7.1	7.1 ± 6.4	7.1 ± 6.3
Male sex, n (%)	305 (55.8)	425 (56.4)	216 (54.7)	215 (55.7)	316 (50.4)	297 (51.3)
Length of unit stay, days, median (IQR)	2 (1, 5)	3 (2, 5)	3 (2, 5)	2 (2, 4)	3 (2, 5)	3 (2, 5)
Median Chronic Condition Indicator (IQR)	2 (1, 3)	2 (1, 3)	3 (2, 5)	3 (2, 5)	3 (1, 5)	3 (2, 5)
ICU patients per resident physician†	3.9	4.8	4.4	7.3	7.3	8.0

	Site D		Site E		Site F	
Characteristic (site difference comparison)	Control	Intervention	Control	Intervention	Control	Intervention
Number of patients	432	392	617	601	705	673
Number of unit admissions	487	421	684	661	768	790
Number of patient-days	2884	2776	3659	3892	3886	4505
age, yr, mean ± SD	5.9 ± 6.0	5.6 ± 5.9	7.3 ± 6.7	6.5 ± 6.3	7.6 ± 7.3	8.1 ± 7.1
Male sex, n (%)	244 (50.1)	239 (56.8)	373 (54.5)	362 (54.8)	399 (51.9)	405 (51.3)
Length of unit stay, days, median (IQR)	2 (1, 5)	2 (1, 5)	2 (2, 4)	3 (2, 4)	3 (2, 4)	2 (2, 4)
Median Chronic Condition Indicator (IQR)	2 (1, 3)	2 (1, 3)	2 (1, 4)	2 (1, 4)	2 (1, 3)	2 (1, 3)
ICU patients per resident physician†	7.1	9.1	7.9	10.2	9.8	13.2

†ICU patients per resident-physician (IPRP) calculated as average census at each site, per schedule, over average number of resident-physicians present daily at each site, per schedule

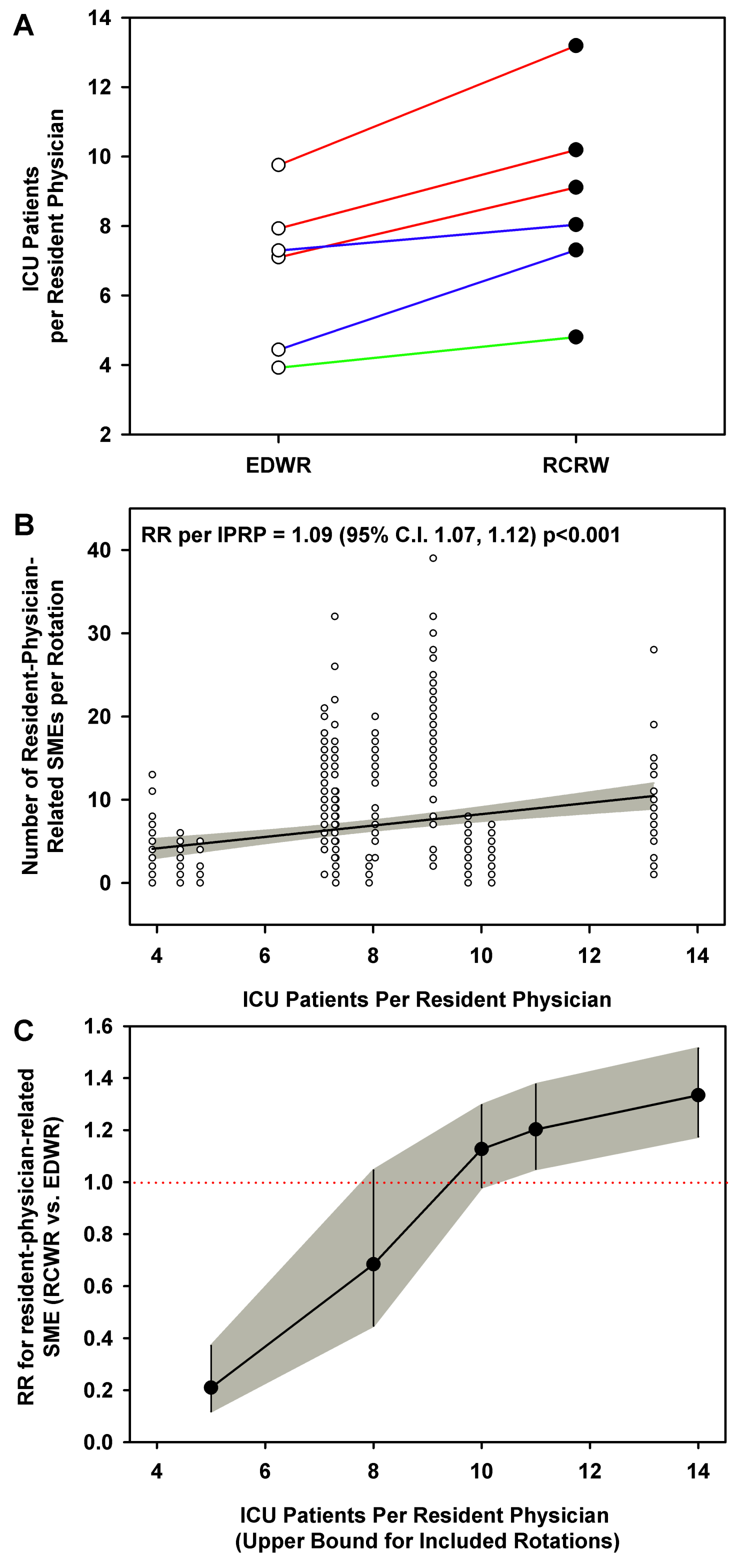
P-values from a chi-square test or Wilcoxon rank-sum test for within site comparisons.

P-values from a chi-square test for homogeneity or a Kruskal Wallis test for site difference comparisons.

SUPPLEMENTAL TABLE 4. Number of Serious Medical Errors, by Site

	Overall		Site A		Site B		Site C		Site D		Site E		Site F	
	control	intervention	control	intervention	control	intervention	control	intervention	control	intervention	control	intervention	control	intervention
Resident-physician related	1268	1723	129	42	93	117	490	435	419	678	18	111	119	340
Unit-wide	2112	3217	304	179	200	213	673	781	600	1046	74	313	261	685

SUPPLEMENTAL FIGURE 1. Effect of workload on serious medical error rates



(A) Number of ICU patients per resident-physician (IPRP), a measure of resident-physician workload, under the extended duration work roster (EDWR; control schedule) vs. the rapid cycle work roster (RCWR; intervention schedule). Sites in red experienced a worsening in rates of resident-physician-related serious medical errors (SMEs) with implementation of the RCWR schedule; sites in blue experienced no significant change; the site in green experienced an improvement. (B) We used a Poisson model with robust standard errors to estimate the unadjusted dependence of the number of resident-physician SMEs on IPRP. Site- and schedule-level average resident-physician workload was correlated with resident-physician-related serious medical errors (SME). (C) On the basis of an initial Poisson model showing modification of the effect of schedule by IPRP, we assessed its effects allowing for dependence in IPRP, again using log-link Poisson models, but with resident-rotation as the unit of analysis instead of admissions to the unit as the unit of analysis, and with site and schedule as fixed effects, robust standard errors, and the log of the duration of resident rotation as an offset. Separate Poisson models were run, restricting each model to rotations with IPRP at discrete thresholds from 5 to 14 to estimate the rate ratio of resident-physician-related SMEs under RCWR and EDWR at each of these thresholds. The rate ratio estimates from these separate Poisson models showed that the effectiveness of the RCWR on the rate of resident-physician-related SMEs across sites depended on IPRP. In these exploratory analyses, the rate ratio of resident-physician-associated SMEs on the RCWR vs. EDWR was significantly <1.0 in analyses including rotations below the IPRP inflection point [RR 0.21 (95% CI: 0.12 – 0.37)], but detrimental [RR 1.46 (95% CI: 1.27 – 1.67)] when IPRP was above the inflection point.

Covariate-adjusted rate ratio estimates of resident-physician-related SMEs are shown with corresponding 95% confidence intervals.

Protocol

This trial protocol has been provided by the authors to give readers additional information about their work.

Protocol for: Landrigan CP, Rahman SA, Sullivan JP, et al. Effect on patient safety of a resident physician schedule without 24-hour shifts. N Engl J Med 2020;382:2514-23. DOI: 10.1056/NEJMoa1900669

This supplement contains the following items for the ROSTERS study:

Contents:

ROSTERS
Randomized Order Safety Trial Evaluating Resident Schedules

A Multi-Center Trial of Limiting PGY 2&3 Resident Work Hours on ICU Patient Safety

Principal Investigators:
Charles A. Czeisler, PhD, MD
Christopher P. Landrigan, MD, MPH
Katie L. Stone, PhD

Supported by:
National Heart Lung and Blood Institute

Version 1.0

March 5, 2013

Table of Contents

PROTOCOL SUMMARY

Objective:
To conduct a multi-center randomized crossover trial in six pediatric ICUs staffed by PGY2 and PGY3 residents to compare the effectiveness and safety of a sleep and circadian science-based (SCS) intervention schedule with a traditional schedule that includes frequent shifts of 24 hours or longer.

Study Population:
Second and third year (PGY2 and PGY3) residents

Study Design:
Multi-center randomized crossover trial

Interventions:
Sleep and circadian science-based intervention schedule versus a traditional schedule

Primary outcomes:
Resident-related preventable adverse events and near misses

Secondary Outcomes:
ICU-wide preventable adverse events and near misses
Resident neurobehavioral performance and predicted driving safety

Study Duration:
12 months (4 month wash-in period followed by 8 months of data collection)

1.1 STUDY AIMS

1.1.1 Aim 1:

To test the hypothesis that PGY2&3 residents working on an *SCS intervention schedule* will make significantly fewer harmful medical errors (preventable adverse events) and other serious medical errors (near misses) while caring for ICU patients than residents working on a *traditional schedule*; *(primary endpoints: resident-related preventable adverse events and near misses)*

1.1.2 Aim 2:

To test the hypothesis that rates of harmful medical errors (preventable adverse events) and other serious medical errors (near misses) *throughout the ICU* (i.e., those involving and those not involving residents) will be lower in ICUs when PGY2&3 residents work on an *SCS intervention schedule* than when residents work on a *traditional schedule; (major secondary endpoints: ICU-wide preventable adverse events and near misses)*

1.1.3 Aim 3:

To test the hypothesis that resident physicians' risk of neurobehavioral performance failures and motor vehicle crashes – as assessed through simple visual reaction time tasks [Johns Drowsiness Score (JDS) and Psychomotor Vigilance Task (PVT) lapses] – will be lower on the *SCS intervention schedule* than on the *traditional schedule. (major secondary endpoints: resident neurobehavioral performance and predicted driving safety)*

1.2 BACKGROUND AND RATIONALE

Sleep deficiency and circadian disruption degrade human alertness and performance both in laboratory and occupational settings. Over the past decade, a series of studies have found that first-year residents (interns; PGY1s) working recurrent extended duration shifts (≥24 hours) make more serious medical errors than do those working shifts of ≤16 consecutive hours; moreover, PGY1s working extended duration shifts suffer more injuries on the job, and have an increased risk of motor vehicle crashes (MVCs) on the drive home from work.(1-8) In 2009, after a year-long comprehensive study, the Institute of Medicine (IOM) concluded that while it remained unclear whether resident sleep deprivation led to patient harm, "the scientific evidence base establishes that human performance begins to deteriorate after 16 hours of wakefulness."(9;10) They consequently called for the elimination of all resident physician shifts without sleep over 16 consecutive hours.

In response, beginning in July 2011, the Accreditation Council for Graduate Medical Education (ACGME) limited interns to no more than 16 consecutive hours of work; second year (PGY2) and higher residents, however, will continue to work for up to 28 consecutive hours.(11) In choosing not to more substantively limit the hours of PGY2 and higher residents

– who represent approximately 80% of all physicians-in-training – the ACGME indicated that in its view, insufficient data existed to support policy change for more senior trainees.

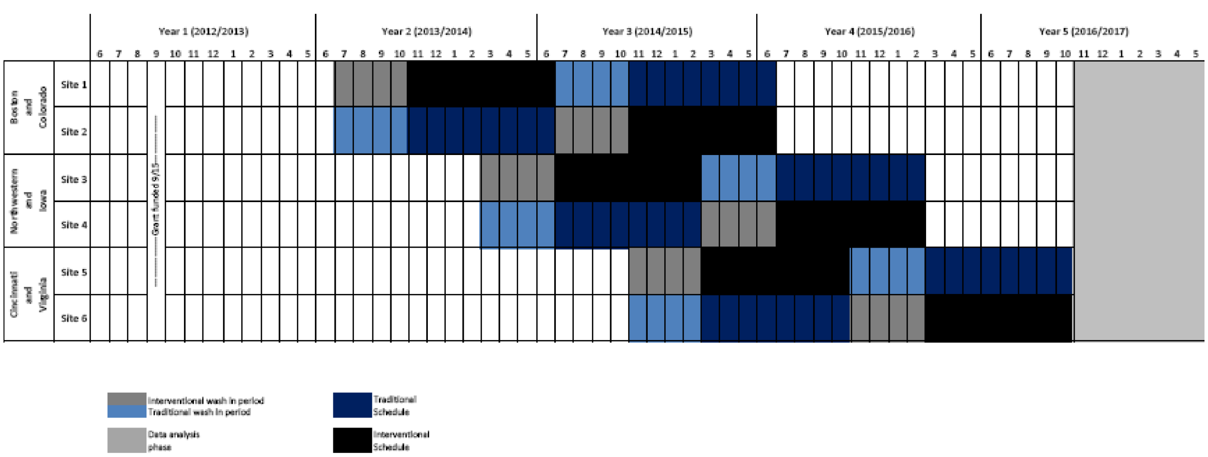
In this study, we seek to conclusively address two knowledge gaps: 1) the lack of data on the relationship between PGY2 and higher sleep deprivation and patient safety; and 2) the lack of data on the relationship between resident sleep deprivation and preventable patient injuries.

1.3 STUDY DESIGN AND INTERVENTIONS

The study is a clustered randomized controlled trial to evaluate the effectiveness of eliminating residents' traditional 24-hour shifts in ICUs. The trial will take place in 6 academic medical centers nationwide, in three waves of two centers each.(Figure 1) One of each pair of units will initially be randomly assigned to the traditional schedule (i.e., in which overnight shifts of 24-28 hours continue to occur every 4 nights), or to the SCS intervention schedule (i.e., in which residents are limited to 16 consecutive work hours; sleep before night duty is promoted; and time off is arranged to allow recovery from sleep debt.)

Our planned timeline will be as follows: after a 2 month planning and startup period, during which staff will be recruited and hired, the intervention will be implemented in one randomly selected PICU in study wave 1; the traditional schedule will continue in the other PICU. There will be a 4-month wash-in period, after which time 8 months of data collection will take place in intervention (in black, Figure 1) and traditional (in blue) ICUs. After 8 months of data collection, the units will cross over (traditional to intervention schedule; intervention to traditional). Following another 4-month wash-in period, there will be 8 additional months of data collection. Study sites in waves two and three will follow the same pattern. Of note, each intervention period will be scheduled to occur entirely within a single academic year, and will be preceded or followed (12 months apart) by a control period that will also occur entirely within an academic year.

Figure 1. Timeline



1.4 SELECTION AND ENROLLMENT OF PARTICIPANTS

All second and third year residents at the six clinical centers will be invited to enroll in the study. The only exclusion criteria is that the resident is at least 21 years of age. Four to eight residents will be recruited per month at each clinical center for total of approximately 50 residents per site and approximately 300 subjects study-wide.

1.4.1 Recruitment

Sites will submit a waiver of informed consent for the collection of patient data, but will obtain residents' written informed consent to observe them, and collect resident-specific data. Working with program directors at each hospital, the Principal Investigators will make presentations each year to all residents to describe our study and request volunteer participants. Residents are free not to participate in the study; if they choose not to participate, they will not be followed by observers and no other data will be collected from them, but the unit schedule will proceed on the *traditional* or *SCS intervention schedule* as planned.

1.4.2 Randomization

The intervention will be implemented in one randomly assigned study site in each of the three waves and the second site will remain on the traditional schedule. After 8 months of data collection, the units will cross over (traditional to intervention schedule; intervention to traditional). The Data Coordinating Center (DCC) will be responsible for randomly assigning sites to their initial study arm.

1.4.3 Preparation of subjects

Before the start of each study, volunteers will receive a detailed explanation of the procedures involved in the study. They will also attend an educational seminar prior to the implementation of the intervention schedule designed to provide an overview of sleep and circadian science, and to convey the importance of complying with the protocol by attempting to sleep prior to night shifts. They will be asked to complete a baseline survey, which will include the Sleep Disorders and Berlin Sleep Questionnaires.

1.5 STUDY PROCEDURES

The study will take place in pediatric intensive care units (PICUs). In general, subjects will continue to carry out their normal activities and responsibilities when working either the intervention or traditional schedule. Throughout the PICU rotations, subjects will complete a daily sleep diary; will wear actiwatches, described further below, to validate the results of self-reported sleep; and will wear Optalert glasses on the commute to and from work. They will also periodically complete psychomotor vigilance tests (PVTs) to monitor their vigilance. At the completion of their rotations, all subjects will complete an End of Rotation survey. For every experimental intervention, there will be written protocols and checklists used to insure uniformity in the execution of standard procedures.

1.5.1 Detection of Errors and Adverse Events

We will use a very intensive, four-pronged data collection approach to comprehensively measure rates of all errors and adverse events.

1.5.1.1 Continuous Observation

A team of five physician research associates will conduct direct observation of resident subjects working in the units, 24 hours per day, 7 days per week. The observers will share this responsibility, working in eight-hour shifts. All suspected adverse events and errors will be documented on tablet-based data forms, and transmitted to the research nurse, who will gather follow-up data on them while conducting his or her daily chart reviews, as described below. The observer will also record and classify all medical activities in which the study subject engages, including but not limited to performance of procedures, test and medication ordering, and test interpretation. Suspected adverse events and incidents will be identified and classified. Follow-up of all suspected adverse events and errors detected by the observers will be performed by the nurse data extractors, who will collect additional information.

1.5.1.2 Voluntary and solicited reports

Forms will be made available and prominently posted in the ICUs to facilitate voluntary reporting of possible errors and events by nurses and other clinical staff. Chart reviewers will also request reports from staff of errors and adverse events 5 days per week. Any reported error or event will be pursued by the nurse data extractors, who will collect additional information.

1.5.1.3 Collection of formal incident reports

In each hospital, formal incident reports will be collected if permitted; if any institutions will not allow access to these data, we will request that duplicate study reports be filed by clinical staff on our study units when they complete formal reports. In addition, in any hospitals with computerized adverse event detection systems, the computerized AE monitors will also be interrogated for study-unit events. Nurse data extractors will collect additional information on each incident identified.

1.5.1.4 Chart surveillance

The nurse data extractors will serve as the focal point for data collection and organization, and will follow up and review all data collected by observers, reported by staff, and detected via incident reporting systems. In addition to coordinating collection from other sources, reviewers will examine all orders and charts 5 days per week; a focused version of the Institute for Healthcare Improvement (IHI) Global Trigger Tool (consisting of the intensive care module, cares module, and medication module triggers)(12) will be used to increase the sensitivity of adverse event detection. Reviews on Monday will include a review of the weekend. Data collected for each incident will include a description and classification of the event, patient information, services and personnel involved, and additional work resulting from the event. Medication incidents will further include name, dose, route and category of the drug involved.

1.5.2 Classification by severity, attribution, and preventability

Physician observers and research nurses will identify suspected errors and adverse events. Two independent physician reviewers will subsequently classify events as errors, potential adverse events (near misses), or adverse events. All events will be rated on severity using the modified NCC-MERP scale.(12;13) Preventability will be rated using a four point Likert scale. Disagreements will be resolved by discussion. Events for which consensus cannot be reached will be re-rated by a third reviewer. Pre-discussion inter-rater reliability will be evaluated with the Kappa statistic.

1.5.3 Identification of patient risk factors

Clinical and demographic data for all patients admitted to study units will be collected by the observers from patient records and institutional administrative databases during lulls in unit activity. Severity of illness will be assessed using a standardized illness severity score, such as PRISM.(14) Morbidity and disability data will be collected until discharge for all patients with an adverse event, as will data about the complexity of conditions, interventions, and drug regimens, including number and types of drugs and interventions.

1.5.4 Measuring Sleep and Fatigue

In addition to data on patient safety, we will collect data on residents' sleep and work hours using the methods described below.

1.5.4.1 Sleep and work logs

A diary of sleep and wake times will be maintained by the research volunteers. A post-sleep questionnaire will be completed immediately following wake time from all sleep episodes and will provide information on subjective evaluation of sleep onset, duration, consolidation, quality, and wakefulness during sleep, as well as daily work hours. Our sleep and work logs have previously been validated; hours of sleep and work reported using this methodology have a high correlation with polysomnographically-validated total sleep time (r=0.94) as well as 3rd-party documented work hours (r=0.98).(2)

1.5.4.2 Ambulatory physiologic monitoring

To further validate reported sleep times in this study, wrist activity and ambient light levels will be monitored for the entire duration of residents' rotations with a solid-state, portable data collection device (Motionlogger BASIC; Ambulatory Monitoring, Inc., Ardsley, NY). The Motionlogger recorder is a small wrist worn device that measures activity and ambient light exposure; it is waterproof and powered by a 3V, 150 mAmp-hr Lithium Manganese battery that has a lifetime of 60 days. Data are preserved if the battery expires.

1.5.5 Measuring Resident Vigilance and Driving Safety

1.5.5.1 Optalert

Figure 2. Optalert glasses

Optalert is an innovative technology that uses infrared oculography to monitor the alertness of an individual continuously. The Optalert glasses contain a small light emitting diode (LED) positioned below and in front of the eye, attached to a frame designed to hold clear or prescription lenses. **(Figure 2)** Brief pulses of invisible infrared (IR) light (each lasting

70 μs, wavelength 935 nm) are directed up in a 30 degree cone of light centered on the lower edge of the upper eyelid and repeated at a frequency of 500 Hz. The total IR light reflected back from the eye and eyelid is detected by a phototransistor in the frame beside the LED.

A microprocessor, housed in the arm of the glasses, controls the timing and other characteristics of the IR pulses, and digitizes the analogue output from the sensors. The power supply and the serial output from the glasses via a light cable connected to a personal digital assistant (PDA). This PDA provides a variety of different analyses of the recorded signals including reflecting position and the velocity of movements. The velocity of movements are calculated 500 times per second as the change in position (uncalibrated units) per 50 ms. The durations of the separate components of blinks and of other eye and eyelid movements, and their AVRs, are measured by software that included period-amplitude analysis of both the position and velocity signals and classification of all periods and amplitudes per minute.

These analyses automatically generate an alertness score (Johns Drowsiness Scale, JDS), a composite measure of alertness based on many variables characterizing blinks, including the ratio of the amplitude to the maximum velocity (AVR) of eyelid movements for closing and reopening of the eyelids, as well as the duration of eyelids closing, of remaining closed, and of reopening during blinks. The weighting for each variable in the JDS was derived from multiple regression analysis, comparing results of Optalert recordings before and after sleep deprivation. The JDS is calculated each minute on a 0-10 scale. Normal values are 0-4, and do not require adjustment for individual subjects.

1.5.5.2 PVT.

In the proposed study, we will also have resident physician subjects complete Psychomotor Vigilance Testing – an established metric of vigilance that is sensitive to sleep deprivation and circadian misalignment (15) – during one shift per week, every five hours. Completing the PVT requires 10 minutes and provides data on vigilance that will be used to derive an independent measure of neurobehavioral performance while subjects are on the job.

1.5.6 Additional Measures

1.5.6.1 Self-reported attentional failures, motor vehicle crashes (MVCs), and percutaneous injuries

Through end-of-rotation surveys, we will also collect data on residents' self-reported attentional failures, MVCs, near-miss MVCs, and percutaneous injuries using the instruments we previously developed for use in our national cohort study. Any reported MVCs and percutaneous injuries will be validated by the collection of objective data (e.g. police reports, repair bills, etc. for MVCs), as was done in our cohort study.(1;4) When MVCs occur on the commute to and from work, we will also have detailed Optalert data available. We believe that collecting MVC and percutaneous data is important given their implications for resident safety, but with 300 anticipated subjects in the study, and the relative infrequency of these occupational injuries,(1;4) we will have power to detect an effect of the intervention on these outcomes only if they are reduced more than two-fold. As such, analysis of these outcomes will be considered exploratory.

1.5.6.2 Educational Measures

We will also, through the daily diaries and the end-of-rotation survey, collect preliminary data on residents' educational experiences. Residents will provide daily estimates on their logs of the amount of time spent in didactic education and the amount of time spent reading for knowledge acquisition. On their end-of-rotation surveys, they will report their impressions of their educational experience.

1.5.6.3 Collection of Salivary Samples for Subsequent Genetic Analyses

In light of the emerging science exploring the genetic predictors of susceptibility to sleep loss and circadian misalignment, salivary samples will be collected to measure genetic modifiers of the *SCS intervention's* effects. Unfortunately, the limited number of subjects (~300) being studied in this proposed trial precludes conducting genome-wide association studies to determine what genes may convey an increased risk of fatigue-related error, and as yet, no candidate gene has been independently verified to convey altered vulnerability to the performance-impairment associated with sleep deprivation. However, given that this study will gather unprecedented data on sleep, performance, and safety, and the likelihood that one or more candidate genes will be verified in the near future, we will collect samples and analyze the DNA of all participants who agree to participate in a future genetic evaluation. These specimens will then be available at a later time to evaluate whether candidate genes verified to affect vulnerability to sleep loss predict the probability of Optalert lapses or of making an error that leads to an AE.

1.6 SAFETY ASSESSMENTS

The study will be looking for the occurrence of adverse events that occur in ICU care; some are preventable, and some are not. Data on adverse events will be systematically collected and reviewed by the DSMB after each 8 month data collection period. Rates of adverse events will be compared and the DSMB will subsequently make periodic recommendations on whether to continue, modify, or terminate the study. From prior studies, we have found that the cause of detected adverse events in hospitals cannot in most cases be reliably assigned on a case by case basis (e.g., was acquisition of a catheter-related bloodstream infection due to a resident's work schedule? A nurse error? Another cause?), so we anticipate that there will not be discrete, individual adverse events that would modify an assessment of the safety of the intervention or traditional work schedule's risks; rather, we expect that through epidemiologic methods, we will obtain useful data on the safety of the intervention vs. traditional work schedules, and act as needed. However, if there are discrete events that clinical staff or study investigators believe are attributable to either the intervention of traditional work schedule, these events will be brought to the attention of the DSMB for review and action as needed.

1.7 STATISTICAL CONSIDERATIONS

1.7.1 Outcomes and Statistical Analyses

We will compare resident-related (Specific Aim 1) and total (Specific Aim 2) rates of harmful medical errors (preventable adverse events) on the two schedules. As a secondary measure, rates of non-harmful serious medical errors (i.e., near misses) will also be compared. In an intention-to-treat analysis, rates will be compared by schedule using Poisson models with compound symmetric working correlation and robust standard errors to account for both over-dispersion and clustering by clinical center (16), and the number of patient-days at risk in each center and period included as so-called offset. The models will control for period effects and will be used to assess treatment-period interaction, a standard check with crossover trials(17).

In addition, actigraphy and sleep diaries will be used to ascertain the minimum and mean number of hours of sleep obtained per night, total sleep time, wake after sleep onset (WASO; minutes), and sleep efficiency. We will also compare performance on the Johns Drowsiness Score (JDS), based on results captured by Optalert, and the Psychomotor Vigilance Test (PVT). Specifically, we will compare the time-weighted average JDS score for each shift; the proportion of the shift minutes spent with a JDS score above 4.5, the cutoff for impairment; and the numbers of lapses in vigilance, defined as a reaction time >500 ms, on the PVT. These Intention-to-treat comparisons by will be made using linear mixed models for repeated measures, with nested random effects to account for clustering within clinical center and residents[12]. Outcomes will be normalized as necessary; generalized linear mixed models appropriate for other outcome distributions, and/or bootstrap standard errors will be used if adequate normalization cannot be achieved. In preliminary analyses, we will assess the comparability of the residents in the intervention and control periods in terms of age, gender, post-graduate year (PGY2 vs PGY3), as well as other potential confounders of the intervention. If imbalances are found, we will conduct sensitivity analyses in which we flexibly adjust for the potential confounders.

1.7.2 Sample Size and Minimum Detectable Effects

Based on the Intern Sleep and Patient Safety Study, during which 2,203 patient-days were accrued under both schedules over a total of 8.5 months at a single center with 2 ICUs totaling 20 beds, we estimate that approximately 46,080 patient-days will be accrued across the six centers in this study, each contributing an observation period of 8 months on each schedule, with an average census of 16 patients, or 80% of an average 20 beds. Assuming that the baseline error rates per-patient day on the standard schedule are the same as those observed in the Intern Sleep and Patient Safety Study, the new sample will provide 90% power to detect relative rate reductions (MD-RRRs) on the sleep and circadian science-based (SCS) intervention schedule as shown in Figure 3. Estimates are shown first assuming a scale factor of 1.0, corresponding to a Poisson distribution, while conservative estimates assuming an over-dispersed Poisson distribution with a scale factor of 2.0, are shown in parentheses. For overall serious medical errors, we will have 90% power to detect reductions in both

Serious medical errors	Resident-related rate per 1000 patient-days	MD-RRR with Scale Factor of 1.0 (2.0)	Unit rate per 1000 patient-days	MD-RRR with Scale Factor of 1.0 (2.0)
overall	136.0	8.0% (11.2%)	193.2	6.7% (9.4%)
intercepted	70.3	11.0% (15.5%)	95.1	9.5% (13.3%)
non-intercepted	44.8	13.8% (19.2%)	59.5	11.9% (16.7%)
Preventable AEs	**20.9**	**19.8% (27.5%)**	**38.6**	**14.8% (20.6%)**

Figure 3, ***power table.*** Minimum detectable relative rate reductions (MD-RRRs) detectable with 90 percent power in two-sided tests (with alpha of 5%) for serious medical errors, by sub-type of error, using Poisson models. The primary outcomes of interest, resident-related and total preventable AEs, are bolded.

resident-related and unit error rates of 7-11%; MD-RRRs for intercepted and non-intercepted medical errors will be between 9.5 and 19%. Even for the least frequent outcome but that of greatest interest, resident-related preventable AEs, we will have 90% power to detect a relative reduction of 19.8% if the distribution is Poisson, and 27.5% under our conservative assumptions about over-dispersion. The actual reductions seen in our preliminary study were larger than these thresholds in each sub-category of resident-related serious errors, including resident-related preventable adverse events. Thus, we expect to be adequately powered to draw definitive conclusions regarding the effectiveness of the *SCS schedule* intervention. In assessing interaction/period interactions, MD-RRRs will be fourfold larger than the main intervention effects. However, because residents will participate in at most one period, the potential for carryover effects is much smaller than in most crossover studies.

In addition, for sleep and work hours as well as JDS and PVT results, we estimate the expected sample of 300 residents will provide 90% power in 2-sided tests with alpha of 5% to detect standardized between-group differences of 0.30 to 0.37 standard deviations of the outcome, depending on the design effect due to clustering within centers and residents; calculations are shown in Figure 4 for minimum detectable effects (MDEs) for sleep and work hours, based on standard deviations observed for sleep and work hours observed in the ICU pilot study conducted at Brigham and Women's Hospital. Because this pilot study did not collect data on other sleep outcomes (e.g. WASO, sleep efficiency) or the performance outcomes (JDS and PVT), the standardized units are also shown in Figure 4. These calculations show that the study will be powered to detect small-to-moderate intervention effects on these outcomes, even if the design effect is as large as 2.

Outcome	MDE
Sleep hours (weekly)	1.75 / 1.92 / 2.21
Work hours (weekly)	1.39 / 1.53 / 1.76
Standardized units	0.30 / 0.32 / 0.37

Figure 4, *minimum effects on resident outcomes* detectable with 90% power in two-sided 5% tests, with design effects of 1.25/1.5/2.0.

1.8 DATA COLLECTION AND QUALITY ASSURANCE

1.8.1 Data Acquisition and Transmission

The DCC will develop a front-end component for data entry and transmission. Data would be entered at remote sites on secure mobile computing devices and transmitted a minimum of once a day via secure transfer. Data received will be automatically imported into SQL tables and integrated into our standard data system. Data will be available for review/updating via the standard study website within 24-48 business hours. The DCC will create and maintain a comprehensive programmer's guide for all the data systems it provides.

1.8.2 Data Collection and Editing

Capturing data electronically will reduce the risk of missing data. Furthermore, edits checks may be built into the instruments to flag out of range or inconsistent data. Further custom edit checks will be written and run on data received at the DCC. Each hour, the cumulative study data is subjected to these custom edits to ensure completeness, consistency and validity of the data. The results of the error-checking procedures are posted to the study website, which the study staff will check daily to both confirm that the DCC has successfully received all of the transmitted forms and to address errors that have been detected by the edit system. The Project Manager and other Clinical Coordinating Center (CCC) staff will be responsible for checking the daily data edits and working in conjunction with the sites to address them. This data entry/data editing process is monitored by a number of standard monitoring reports that will be available on the website. An audit trail is maintained which includes the date of change, time of change, description of data change, and the study ID of the person making the change.

The DCC will also serve as the Reading Center for the actigraphy, Optalert and PVT datasets. DCC staff will perform an initial review of the files upon receipt and report any issues to the clinical site staff; later DCC staff will clean and merge the device data with the form data.

The DCC will prepare comprehensive ‘clean’ datasets for analysis, providing the analysis-ready files to the CCC and others as necessary, for statistical analysis purposes. The DCC will also ensure that any entity receiving the dataset will have a Data Use Agreement on file, signed by the appropriate institutional authority. The DCC will also be responsible for preparing a public release dataset to be housed within NHLBI’s BioLINCC (or a comparable public data repository) no later than 3 years after the end of the trial or 2 years after the main paper has been published, whichever comes first.

1.8.3 Data Management and Study Progress

The study website which includes a number of data management features and reports to enable clinical sites and study investigators to monitor data collection and study progress. Data Management features include a data inventory, lists of missing forms and queries by participant ID, and tools to resolve queries and update data via the website. Reports can be custom-designed to meet study needs and could include:

- Enrollment in aggregate and by site
- Demographics in aggregate and by site
- Early Discontinuation in aggregate and by site

1.8.4 Staff Training

Site staff (physician observers, nurse data extractors and study coordinators) will be trained intensively in the detection of medical errors and adverse events using an established training protocol. Inter-rater reliability will be verified during the study. Staff will also be instructed on procedures for entering data into electronic data collection forms and transmitting these data to the DCC. Lastly, staff will be trained in the operation of the actigraph and Optalert devices, as well as the PVT tests.

1.9 PARTICIPANT RIGHTS AND CONFIDENTIALITY

1.9.1 Institutional Review Board (IRB) Review

This protocol and informed consent documents, and any subsequent modifications, will be reviewed and approved by the IRB or ethics committee responsible for oversight of each of the clinical sites, as well as the clinical and data coordinating centers.

1.9.2 Informed Consent Forms

Informed Consent documents will be maintained at individual sites and coordinators must ensure that they are using an up-to-date, IRB approved version of the consent form. A signed

Informed Consent form will be obtained from each participant. The Informed Consent will describe the purpose of the study, the procedures to be followed, and the risks and benefits of participation and must comply with the requirements of the study site's Institutional Review Board. A copy of the Informed Consent will be given to each participant and provision of the copy will be documented in the study participant's record. The DCC will ensure that all clinical site informed consents are consistent and include required elements.

1.9.3 Participant Confidentiality

1.9.3.1 Clinical Sites

In order to protect confidentiality of the research participants as well as the patients they treat, both will be assigned a unique study identification number, which will be used to refer to them on all study forms. The list of codes linking resident-subjects' and patients' names and study IDs will be kept by each site PI on a separate password protected computer in a locked office at that site. Each site PI will be responsible for ensuring the security of these data. These identifying data will not be transmitted outside of the originating institution. Risk to patient confidentiality will likewise be minimized. All information will be de-identified at the point of data analysis.

1.9.3.2 Data Coordinating Center

The DCC follows standard operating procedures (SOPs) for computer system security to ensure the confidentiality and validity of study data. The SOPs are designed to prevent unauthorized access and limit authorized access to our computer systems and are in compliance with established standards for Information Technology Security. Our network is privately maintained, hardware fire-walled and none of the workstations or database servers can be directly addressed from outside the Local Area Network. Study website and database access requires a network domain account with appropriate account-specific permission on the database. All requests for new accounts and access to the database must be documented by a System Access Request Form signed by the project director.

1.10 COMMITTEES

1.10.1 Steering Committee

The Steering Committee will be responsible for all decisions concerning the scientific and technical conduct of the study. It will appoint the analysis and publications committee and writing groups, ensuring that information from the study is disseminated in the scientific literature and at scientific meetings. The committee will be chaired by Dr. Czeisler and will include the lead site investigator from the clinical and data coordinating center and each of the clinical sites. The Steering Committee will have one in-person meeting on an annual basis, and will meet by teleconference on at least a bi-monthly basis. The Steering Committee consists of the following members:

Charles A. Czeisler, PhD, MD	Chair, PI, CCC, Brigham and Women's Hospital
Christoper P. Landrigan, MD, MPH	PI, CCC, Brigham and Women's Hospital
Katie L. Stone, PhD	PI, DCC, California Pacific Medical Center
Susan Redline, MD, MPH	Chair, Sleep Research Network

Ken Wright, PhD	PI, University of Colorado
Gwen Erkonen, MD	PI, University of Iowa
Pearl Yu, MD	PI, University of Virginia
Phyllis Zee, MD, PhD	PI, Northwestern University
Sue Poynter, MD	PI, University of Cincinnati
Michael Twery, PhD	Project Scientist, NHLBI

1.10.2 Executive Committee

An Executive Subgroup of the Steering Committee will be responsible for decisions that require attention between Steering Committee Meetings and for major financial, administrative and operational decisions. The Executive Committee will meet via teleconference at least monthly and will consist of the following members:

Charles A. Czeisler, PhD, MD	PI, CCC, Brigham and Women's Hospital
Christoper P. Landrigan, MD, MPH	PI, CCC, Brigham and Women's Hospital
Katie L. Stone, PhD	PI, DCC, California Pacific Medical Center
Michael Twery, PhD	Project Scientist, NHLBI

1.10.3 Principal Planning Group

A planning group comprised of the Principal Investigators and project staff of both the CCC and DCC will be responsible for the development of study documents, including the forms, protocol and operations manual, as well as the day-to-day management of the trial. The Principal Planning Group will meet via teleconference at least monthly and will consist of the following members:

Charles A. Czeisler, PhD, MD	PI, CCC, Brigham and Women's Hospital
Christoper P. Landrigan, MD, MPH	PI, CCC, Brigham and Women's Hospital
Joshua Stephens	CCC, Brigham and Women's Hospital
Katie L. Stone, PhD	PI, DCC, California Pacific Medical Center
Dana R. Kriesel, MPH	PD, DCC, California Pacific Medical Center

1.10.4 Working Groups

Members of the Steering Committee will form working groups to manage and oversee specific aspects of the trial. These include: (1) Data Management and Data Quality; (2) Patient Safety and PICU Quality Outcomes; (3) Sleep, Performance and Health Outcomes; and (4) Education Outcomes. Working groups will meet on an as needed basis.

1.10.5 Data and Safety Monitoring Board

To monitor the study for the possibility that the intervention schedule has an adverse effect on safety, a multidisciplinary Data and Safety Monitoring Board (DSMB) comprised of financially disinterested members will be appointed by NHLBI, with guidance from DCC. The DSMB will monitor the quality and integrity of the data emerging from the proposed study; assess the adequacy of recruitment, compliance, follow-up and other aspects of study execution; and review interim analyses of the major outcomes of the trial to identify safety issues and make periodic recommendations on whether to continue, modify, or terminate the study. The DSMB will be an advisory board to NHLBI and the Steering Committee. The DSMB will work under a Charter that will be developed by the DCC, working in tandem with

NHLBI. The DCC will support the DSMB by preparing interim reports for review at DSMB meetings; and by organizing scheduled and ad hoc meetings of the DSMB as needed. Please see the DSMB Charter for more details.

1.11 CONFLICT OF INTEREST

The DCC will be responsible for surveying the key personnel at each study site on an annual basis regarding any conflicts of interest that may have arisen. The Executive Committee will review the results of these annual surveys and determine if any action is necessary.

1.12 REFERENCES

1. Barger LK, Cade BE, Ayas NT, Cronin JW, Rosner B, Speizer FE, Czeisler CA. Extended work shifts and the risk of motor vehicle crashes among interns. N Engl J Med 2005;352:125-34

2. Lockley SW, Cronin JW, Evans EE, Cade BE, Lee CJ, Landrigan CP, Rothschild JM, Katz JT, Lilly CM, Stone PH, et al. Effect of reducing interns' weekly work hours on sleep and attentional failures. N Engl J Med 2004 Oct 28;351(18):1829-37

3. Landrigan CP, Rothschild JM, Cronin JW, Kaushal R, Burdick E, Katz JT, Lilly CM, Stone PH, Lockley SW, Bates DW, et al. Effect of reducing interns' work hours on serious medical errors in intensive care units. N Engl J Med 2004 Oct 28;351(18):1838-48

4. Ayas NT, Barger LK, Cade BE, Hashimoto DM, Rosner B, Cronin JW, Speizer FE, Czeisler CA. Extended work duration and the risk of self-reported percutaneous injuries in interns. JAMA 2006 Sep 6;296(9):1055-62

5. Grantcharov TP, Bardram L, Funch-Jensen P, Rosenberg J. Laparoscopic performance after one night on call in a surgical department: prospective study. B.M.J. 2001 Nov 24;323:1222-3

6. Barger LK, Ayas NT, Cade BE, Cronin JW, Rosner B, Speizer FE, Czeisler CA. Impact of extended-duration shifts on medical errors, adverse events, and attentional failures. PLoS Med 2006 Dec;3(12):e487

7. Arnedt JT, Owens J, Crouch M, Stahl J, Carskadon MA. Neurobehavioral performance of residents after heavy night call vs after alcohol ingestion. JAMA. 2005 Sep 7;294(9):1025-33

8. Philibert I. Sleep loss and performance in residents and nonphysicians: a meta-analytic examination. Sleep 2005 Nov 1;28(11):1392-402

9. Institute of Medicine. Resident Duty Hours: Enhancing Sleep, Supervision, and Safety. Washington, D.C.: The National Academies Press; 2009. 1-322 p.

10. Iglehart JK. Revisiting duty-hour limits--IOM recommendations for patient safety and resident education. N Engl J Med 2008 Dec 18;359(25):2633-5

11. ACGME. Common Program Requirements. 2010 Sep 26.

12. Griffin FA, Resar RK. Global Trigger Tool for Measuring Adverse Events: IHI Innovation Series white paper. Cambridge, MA: Institute for Healthcare Improvement; 2007.

13. National Coordinating Council for Medication Error Reporting and Prevention (NCC MERP). Index for Categorizing Errors. 2009.

14. Pollack MM, Ruttimann UE, Getson PR. Pediatric risk of mortality (PRISM) score. Crit Care Med 1988 Nov;16(11):1110-6

15. Wyatt JK, Dijk DJ, Ronda JM, Jewett ME, Powell JW, Dinges DF, Czeisler CA. Interaction of circadian and sleep/wake homeostatic-processes modulate psychomotor vigilance test (PVT) performance. Sleep Res 26, 759. 1997.

16. Zeger SL. A regression model for time series of counts. *Biometrika.* 1988;75(4):621-629.

17. Armitage P, Hills M. The two-period crossover trial. *The Statistician.* 1982;31(2): 119-131.

ROSTERS
Randomized Order Safety Trial Evaluating Resident Schedules

A Multi-Center Trial of Limiting PGY 2&3 Resident Work Hours on ICU Patient Safety

Principal Investigators:
Charles A. Czeisler, PhD, MD
Christopher P. Landrigan, MD, MPH
Katie L. Stone, PhD

Supported by:
National Heart Lung and Blood Institute

Version 1.3

June 12, 2015

Table of Contents

PROTOCOL SUMMARY

Objective:
To conduct a multi-center randomized crossover trial in six pediatric ICUs staffed by PGY2 and PGY3 residents to compare the effectiveness and safety of a sleep and circadian science-based (SCS) intervention schedule with a traditional schedule that includes frequent shifts of 24 hours or longer.

Study Population:
Second and third year (PGY2 and PGY3) residents

Study Design:
Multi-center randomized crossover trial

Interventions:
Sleep and circadian science-based intervention schedule versus a traditional schedule

Primary outcomes:
Resident-related preventable adverse events and near misses

Secondary Outcomes:
ICU-wide preventable adverse events and near misses
Resident neurobehavioral performance and predicted driving safety

Study Duration:
12 months (4 month wash-in period followed by 8 months of data collection)

1.1 STUDY AIMS

1.1.1 Aim 1:

To test the hypothesis that PGY2&3 residents working on an *SCS intervention schedule* will make significantly fewer harmful medical errors (preventable adverse events) and other serious medical errors (near misses) while caring for ICU patients than residents working on a *traditional schedule*; *(primary endpoints: resident-related preventable adverse events and near misses)*

1.1.2 Aim 2:

To test the hypothesis that rates of harmful medical errors (preventable adverse events) and other serious medical errors (near misses) *throughout the ICU* (i.e., those involving and those not involving residents) will be lower in ICUs when PGY2&3 residents work on an *SCS intervention schedule* than when residents work on a *traditional schedule; (major secondary endpoints: ICU-wide preventable adverse events and near misses)*

1.1.3 Aim 3:

To test the hypothesis that resident physicians' risk of neurobehavioral performance failures and motor vehicle crashes – as assessed through drive diaries, End of Rotation Survey and simple visual reaction time tasks [Psychomotor Vigilance Task (PVT) lapses] – will be lower on the *SCS intervention schedule* than on the *traditional schedule. (major secondary endpoints: resident neurobehavioral performance and predicted driving safety)*

1.2 BACKGROUND AND RATIONALE

Sleep deficiency and circadian disruption degrade human alertness and performance both in laboratory and occupational settings. Over the past decade, a series of studies have found that first-year residents (interns; PGY1s) working recurrent extended duration shifts ($\geq$24 hours) make more serious medical errors than do those working shifts of $\leq$16 consecutive hours; moreover, PGY1s working extended duration shifts suffer more injuries on the job, and have an increased risk of motor vehicle crashes (MVCs) on the drive home from work.(1-8) In 2009, after a year-long comprehensive study, the Institute of Medicine (IOM) concluded that while it remained unclear whether resident sleep deprivation led to patient harm, "the scientific evidence base establishes that human performance begins to deteriorate after 16 hours of wakefulness."(9;10) They consequently called for the elimination of all resident physician shifts without sleep over 16 consecutive hours.

In response, beginning in July 2011, the Accreditation Council for Graduate Medical Education (ACGME) limited interns to no more than 16 consecutive hours of work; second year (PGY2) and higher residents, however, will continue to work for up to 28 consecutive hours.(11) In choosing not to more substantively limit the hours of PGY2 and higher residents – who represent approximately 80% of all physicians-in-training – the ACGME indicated that in its view, insufficient data existed to support policy change for more senior trainees.

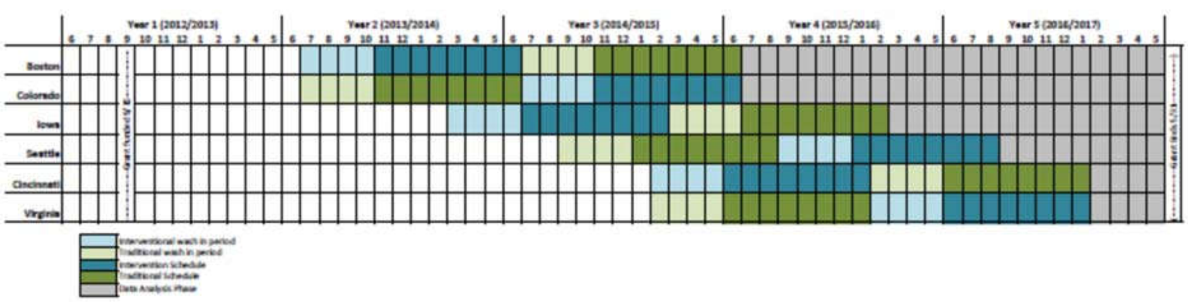
In this study, we seek to conclusively address two knowledge gaps: 1) the lack of data on the relationship between PGY2 and higher sleep deprivation and patient safety; and 2) the lack of data on the relationship between resident sleep deprivation and preventable patient injuries.

1.3 STUDY DESIGN AND INTERVENTIONS

The study is a clustered randomized controlled trial to evaluate the effectiveness of eliminating residents' traditional 24-hour shifts in ICUs. The trial will take place in 6 academic medical centers nationwide, in three waves of two centers each.(Figure 1) One of each pair of units will initially be randomly assigned to the traditional schedule (i.e., in which overnight shifts of 24-28 hours continue to occur every 4 nights), or to the SCS intervention schedule (i.e., in which residents are limited to 16 consecutive work hours; sleep before night duty is promoted; and time off is arranged to allow recovery from sleep debt.)

Our planned timeline will be as follows: after a 2 month planning and startup period, during which staff will be recruited and hired, the intervention will be implemented in one randomly selected PICU in study wave 1; the traditional schedule will continue in the other PICU. There will be a 4-month wash-in period, after which time 8 months of data collection will take place in intervention (in black, Figure 1) and traditional (in blue) ICUs. After 8 months of data collection, the units will cross over (traditional to intervention schedule; intervention to traditional). Following another 4-month wash-in period, there will be 8 additional months of data collection. Study sites in waves two and three will follow the same pattern. Of note, each intervention period will be scheduled to occur entirely within a single academic year, and will be preceded or followed (12 months apart) by a control period that will also occur entirely within an academic year.

Figure 1. Timeline



1.4 SELECTION AND ENROLLMENT OF PARTICIPANTS

All second and third year residents at the six clinical centers will be invited to enroll in the study. The only exclusion criteria is that the resident is at least 21 years of age. Four to eight residents will be recruited per month at each clinical center for total of approximately 50 residents per site and approximately 300 subjects study-wide.

1.4.1 Recruitment

Sites will submit a waiver of informed consent for the collection of patient data, but will obtain residents' written informed consent to observe them, and collect resident-specific data. Working with program directors at each hospital, the Principal Investigators will make presentations each year to all residents to describe our study and request volunteer participants. Residents are free not to participate in the study; if they choose not to participate, they will not be followed by observers and no other data will be collected from them, but the unit schedule will proceed on the *traditional* or *SCS intervention schedule* as planned.

1.4.2 Randomization

The intervention will be implemented in one randomly assigned study site in each of the three waves and the second site will remain on the traditional schedule. After 8 months of data collection, the units will cross over (traditional to intervention schedule; intervention to traditional). The Data Coordinating Center (DCC) will be responsible for randomly assigning sites to their initial study arm.

1.4.3 Preparation of subjects

Before the start of each study, volunteers will receive a detailed explanation of the procedures involved in the study. They will also attend an educational seminar prior to the implementation of the intervention schedule designed to provide an overview of sleep and circadian science, and to convey the importance of complying with the protocol by attempting to sleep prior to night shifts. They will be asked to complete a baseline survey, which will include the Sleep Disorders and Berlin Sleep Questionnaires.

1.5 STUDY PROCEDURES

The study will take place in pediatric intensive care units (PICUs). In general, subjects will continue to carry out their normal activities and responsibilities when working either the intervention or traditional schedule. Throughout the PICU rotations, subjects will complete a daily sleep diary and wear actiwatches, described further below, to validate the results of self-reported sleep.. They will also periodically complete psychomotor vigilance tests (PVTs) to monitor their vigilance. At the completion of their rotations, all subjects will complete an End of Rotation survey. For every experimental intervention, there will be written protocols and checklists used to insure uniformity in the execution of standard procedures.

1.5.1 Detection of Errors and Adverse Events

We will use a very intensive, four-pronged data collection approach to comprehensively measure rates of <u>all</u> errors and adverse events.

1.5.1.1 Continuous Observation

A team of five physician research associates will conduct direct observation of resident subjects working in the units, 24 hours per day, 7 days per week. The observers will share this responsibility, working in eight-hour shifts. All suspected adverse events and errors will be documented on tablet-based data forms, and transmitted to the research nurse, who will gather follow-up data on them while conducting his or her daily chart reviews, as described below. The observer will also record and classify all medical activities in which the study subject engages, including but not limited to performance of procedures, test and medication ordering, and test interpretation. Suspected adverse events and incidents will be identified and classified. Follow-up of all suspected adverse events and errors detected by the observers will be performed by the nurse data extractors, who will collect additional information.

1.5.1.2 Voluntary and solicited reports

Forms will be made available and prominently posted in the ICUs to facilitate voluntary reporting of possible errors and events by nurses and other clinical staff. Chart reviewers will also request reports from staff of errors and adverse events 5 days per week. Any reported error or event will be pursued by the nurse data extractors, who will collect additional information.

1.5.1.3 Collection of formal incident reports

In each hospital, formal incident reports will be collected if permitted; if any institutions will not allow access to these data, we will request that duplicate study reports be filed by clinical staff on our study units when they complete formal reports. In addition, in any hospitals with computerized adverse event detection systems, the computerized AE monitors will also be interrogated for study-unit events. Nurse data extractors will collect additional information on each incident identified.

1.5.1.4 Chart surveillance

The nurse data extractors will serve as the focal point for data collection and organization, and will follow up and review all data collected by observers, reported by staff, and detected via incident reporting systems. In addition to coordinating collection from other sources, reviewers will examine all orders and charts 5 days per week; a focused version of the Institute for Healthcare Improvement (IHI) Global Trigger Tool (consisting of the intensive care module, cares module, and medication module triggers)(12) will be used to increase the sensitivity of adverse event detection. Reviews on Monday will include a review of the weekend. Data collected for each incident will include a description and classification of the event, patient information, services and personnel involved, and additional work resulting from the event. Medication incidents will further include name, dose, route and category of the drug involved.

1.5.2 Classification by severity, attribution, and preventability

Physician observers and research nurses will identify suspected errors and adverse events. Two independent physician reviewers will subsequently classify events as errors, potential adverse events (near misses), or adverse events. All events will be rated on severity using the modified NCC-MERP scale.(12;13) Preventability will be rated using a four point Likert scale. Disagreements will be resolved by discussion. Events for which consensus cannot be reached will be re-rated by a third reviewer. Pre-discussion inter-rater reliability will be evaluated with the Kappa statistic.

1.5.3 Identification of patient risk factors

Clinical and demographic data for all patients admitted to study units will be collected by the research nurses from patient records and institutional administrative databases during lulls in unit activity. Severity of illness will be assessed using ICD-9 codes.

1.5.4 Measuring Sleep and Fatigue

In addition to data on patient safety, we will collect data on residents' sleep and work hours using the methods described below.

1.5.4.1 Sleep and work logs

A diary of sleep and wake times will be maintained by the research volunteers. A post-sleep questionnaire will be completed immediately following wake time from all sleep episodes and will provide information on subjective evaluation of sleep onset, duration, consolidation, quality, and wakefulness during sleep, as well as daily work hours. Our sleep and work logs have previously been validated; hours of sleep and work reported using this methodology have a high correlation with polysomnographically-validated total sleep time (r=0.94) as well as 3rd-party documented work hours (r=0.98).(2)

1.5.4.2 Ambulatory physiologic monitoring

To further validate reported sleep times in this study, wrist activity and ambient light levels will be monitored for the entire duration of residents' rotations with a solid-state, portable data collection device (Motionlogger BASIC; Ambulatory Monitoring, Inc., Ardsley, NY). The Motionlogger recorder is a small wrist worn device that measures activity and ambient light exposure; it is waterproof and powered by a 3V, 150 mAmp-hr Lithium Manganese battery that has a lifetime of 60 days. Data are preserved if the battery expires.

1.5.5 Measuring Resident Vigilance

1.5.5.1 PVT.

In the proposed study, we will also have resident physician subjects complete Psychomotor Vigilance Testing – an established metric of vigilance that is sensitive to sleep deprivation and circadian misalignment (14) – during one shift per week, every five hours. Completing the PVT requires 10 minutes and provides data on vigilance that will be used to derive an independent measure of neurobehavioral performance while subjects are on the job. Subjects will also be asked to complete the Karolinska Sleepiness Scale (KSS), a self-report scale that measures alertness, before and after each test.

1.5.6 Additional Measures

1.5.6.1 Self-reported attentional failures, motor vehicle crashes (MVCs), and percutaneous injuries

Through end-of-rotation surveys, we will also collect data on residents' self-reported attentional failures, MVCs, near-miss MVCs, and percutaneous injuries using the instruments we previously developed for use in our national cohort study. Any reported MVCs and percutaneous injuries will be validated by the collection of objective data (e.g. police reports, repair bills, etc. for MVCs), as was done in our cohort study.(1;4) When MVCs occur on the commute to and from work, we will also have drive diary data available. We believe that collecting MVC and percutaneous data is important given their implications for resident safety, but with 300 anticipated subjects in the study, and the relative infrequency of these occupational injuries,(1;4) we will have power to detect an effect of the intervention on these outcomes only if they are reduced more than two-fold. As such, analysis of these outcomes will be considered exploratory.

1.5.6.2 Educational Measures

As part of the study, participants will have the option of enrolling in an online educational platform entitled OPENPediatrics. OPENPediatrics is a free, open access, peer-reviewed, digital learning platform that utilizes the latest in innovative technology to provide robust continuing medical education. OPENPediatrics is sponsored through Boston Children's Hospital and IBM and located at Boston Children's Hospital (Website: http://openpediatrics.org). Through a feature entitled the Learning Pathways, participants will have access to the ROSTERS curriculum, or a similar ICU learning curriculum approved by each participant's hospital. This curriculum is comprised of 18 lessons by educational experts. Each lesson begins with a pre-test followed by a didactic or procedural demonstration video and concludes with a post-test. This unique format will allow for asynchronous learning, so that participants can complete their educational lessons outside the hospital and apply their knowledge during on-duty hours. OPENPediatrics' robust analytics will allow research staff to track each participant's submissions as they progress through the curriculum. Pre-test and post-test scores as well as duration in each module will allow researchers to note the educational gains made throughout the curriculum, and compare each participant's results in the control and Sleep and Circadian Science-based (SCS) intervention arms. A vigorous set of security measures will allow the analytic data from the application to be securely passed from the application to the data collection program, Cognos. As standard practice for the OPENPediatrics data storage procedures this data will be stored on IBM's secure servers, and IBM's analytic department will control access to this data. All residents in the unit will have access to the platform so no delineation between non subject residents and subject residents can be made by anyone except each site coordinator who will maintain the only identified list of subjects for the site. Resident data specific to each site will be sent from OPENPediatrics to each site coordinator; the site coordinator will then delete all data associated with non-subject residents. These measures are being taken so only site coordinators have knowledge of which data sets are associated with resident subjects. The team will work with the site PI to de-identify the data by removing the user's name/email. The data will only be stored on password-protected computers. Data will only be reported as de-identified, aggregated data. On their end-of-rotation surveys, they will report their impressions of their educational experience.

1.5.6.3 Collection of Salivary Samples for Subsequent Genetic Analyses

In light of the emerging science exploring the genetic predictors of susceptibility to sleep loss and circadian misalignment, salivary samples will be collected to measure genetic modifiers of the *SCS intervention's* effects. Unfortunately, the limited number of subjects (~300) being studied in this proposed trial precludes conducting genome-wide association studies to determine what genes may convey an increased risk of fatigue-related error, and as yet, no candidate gene has been independently verified to convey altered vulnerability to the performance-impairment associated with sleep deprivation. However, given that this study will gather unprecedented data on sleep, performance, and safety, and the likelihood that one or more candidate genes will be verified in the near future, we will collect samples and analyze the DNA of all participants who agree to participate in a future genetic evaluation. These specimens will then be available at a later time to evaluate whether candidate genes verified to affect vulnerability to sleep loss or of making an error that leads to an AE.

1.6 SAFETY ASSESSMENTS

1.6.1 Patient Safety

The study will be looking for the occurrence of adverse events that occur in ICU care; some are preventable, and some are not. Data on adverse events will be systematically collected and reviewed by the DSMB after each 8 month data collection period. Rates of adverse events will be compared and the DSMB will subsequently make periodic recommendations on whether to continue, modify, or terminate the study. From prior studies, we have found that the cause of detected adverse events in hospitals cannot in most cases be reliably assigned on a case by case basis (e.g., was acquisition of a catheter-related bloodstream infection due to a resident's work schedule? A nurse error? Another cause?), so we anticipate that there will not be discrete, individual adverse events that would modify an assessment of the safety of the intervention or traditional work schedule's risks; rather, we expect that through epidemiologic methods, we will obtain useful data on the safety of the intervention vs. traditional work schedules, and act as needed. However, if there are discrete events that clinical staff or study investigators believe are attributable to either the intervention of traditional work schedule, these events will be brought to the attention of the DSMB for review and action as needed.

1.6.2 Resident Safety

As part of this study, we will ask resident subjects questions about depression, including questions about suicidal thoughts and plans, motor vehicle incidents, drowsy driving, including falling asleep at the wheel, and occupational exposures. The Data Coordinating Center for the study reviews survey responses, and if responses indicate suicidality, 3 or more occupational exposure reports, and/or motor vehicle crashes in which the damage was >$1000 and/or drowsiness was reported as a factor, the study principle investigator will be asked to follow up with the resident subject. Near misses will not be followed up on. Additionally, initial reports of falling asleep at the wheel will trigger a follow up with by each site's research study coordinator; 3 or more responses of falling asleep while driving will trigger a follow up by each site's principle investigator. Specifically in regards to questions concerning suicide, if resident subjects respond to the question "Over the past two weeks, how often have you thought about or wanted to commit suicide?" with "Some of the time," "Most of the time," or "All of the time," regardless of whether they respond "Yes" or "No" to "Do you have a plan?", the following message will appear on the form: "Your response to the previous questions causes us to be concerned for your welfare. All reports of suicide ideation will be reported to your site PI, who will then initiate your institution's protocols for timely intervention and action, in compliance with their organizational rules and standards." These criteria were discussed and decided on by the Clinical Coordinating Center, Steering Committee, and voted on and approved by the Data Safety Monitoring Board for this protocol.

1.7 STATISTICAL CONSIDERATIONS

1.7.1 Outcomes and Statistical Analyses

We will compare resident-related (Specific Aim 1) and total (Specific Aim 2) rates of harmful medical errors (preventable adverse events) on the two schedules. As a secondary measure, rates of non-harmful serious medical errors (i.e., near misses) will also be compared. In an intention-to-treat analysis, rates will be compared by schedule using Poisson models with compound symmetric working correlation and robust standard errors to account for both over-dispersion and clustering by clinical center (15), and the number of patient-days at risk in each center and period included as so-called offset. The models will control for period effects and will be used to assess treatment-period interaction, a standard check with crossover trials(16).

In addition, actigraphy and sleep diaries will be used to ascertain the minimum and mean number of hours of sleep obtained per night, total sleep time, wake after sleep onset (WASO; minutes), and sleep efficiency. We will also compare performance based on results captured by the Psychomotor Vigilance Test (PVT). Specifically, we will compare the the numbers of lapses in vigilance, defined as a reaction time >500 ms, on the PVT. This Intention-to-treat comparison will be made using linear mixed models for repeated measures, with nested random effects to account for clustering within clinical center and residents[12]. Outcomes will be normalized as necessary; generalized linear mixed models appropriate for other outcome distributions, and/or bootstrap standard errors will be used if adequate normalization cannot be achieved. In preliminary analyses, we will assess the comparability of the residents in the intervention and control periods in terms of age, gender, post-graduate year (PGY2 vs PGY3), as well as other potential confounders of the intervention. If imbalances are found, we will conduct sensitivity analyses in which we flexibly adjust for the potential confounders.

1.7.2 Sample Size and Minimum Detectable Effects

Based on the Intern Sleep and Patient Safety Study, during which 2,203 patient-days were accrued under both schedules over a total of 8.5 months at a single center with 2 ICUs totaling 20 beds, we estimate that approximately 46,080 patient-days will be accrued across the six centers in this study, each contributing an observation period of 8 months on each schedule, with an average census of 16 patients, or 80% of an average 20 beds. Assuming that the baseline error rates per-patient day on the standard schedule are the same as those observed

Serious medical errors	Resident-related rate per 1000 patient-days	MD-RRR with Scale Factor of 1.0 (2.0)	Unit rate per 1000 patient-days	MD-RRR with Scale Factor of 1.0 (2.0)
overall	136.0	8.0% (11.2%)	193.2	6.7% (9.4%)
intercepted	70.3	11.0% (15.5%)	95.1	9.5% (13.3%)
non-intercepted	44.8	13.8% (19.2%)	59.5	11.9% (16.7%)
Preventable AEs	**20.9**	**19.8% (27.5%)**	**38.6**	**14.8% (20.6%)**

Figure 3, ***power table.*** Minimum detectable relative rate reductions (MD-RRRs) detectable with 90 percent power in two-sided tests (with alpha of 5%) for serious medical errors, by sub-type of error, using Poisson models. The primary outcomes of interest, resident-related and total preventable AEs, are bolded.

in the Intern Sleep and Patient Safety Study, the new sample will provide 90% power to detect relative rate reductions (MD-RRRs) on the sleep and circadian science-based (SCS) intervention schedule as shown in Figure 3. Estimates are shown first assuming a scale factor of 1.0, corresponding to a Poisson distribution, while conservative estimates assuming an over-dispersed Poisson distribution with a scale factor of 2.0, are shown in parentheses. For overall serious medical errors, we will have 90% power to detect reductions in both resident-related and unit error rates of 7-11%; MD-RRRs for intercepted and non-intercepted medical errors will be between 9.5 and 19%. Even for the least frequent outcome but that of greatest interest, resident-related preventable AEs, we will have 90% power to detect a relative reduction of 19.8% if the distribution is Poisson, and 27.5% under our conservative assumptions about over-dispersion. The actual reductions seen in our preliminary study were larger than these thresholds in each sub-category of resident-related serious errors, including resident-related preventable adverse events. Thus, we expect to be adequately powered to draw definitive conclusions regarding the effectiveness of the *SCS schedule* intervention. In assessing interaction/period interactions, MD-RRRs will be fourfold larger than the main intervention effects. However, because residents will participate in at most one period, the potential for carryover effects is much smaller than in most crossover studies.

In addition, for sleep and work hours as well as PVT results, we estimate the expected sample of 300 residents will provide 90% power in 2-sided tests with alpha of 5% to detect standardized between-group differences of 0.30 to 0.37 standard deviations of the outcome, depending on the design effect due to clustering within centers and residents; calculations are shown in Figure 4 for minimum detectable effects (MDEs) for sleep and work hours, based on standard deviations observed for sleep and work hours observed in the ICU pilot study conducted at Brigham and Women's Hospital. Because this pilot study did not collect data on other sleep outcomes (e.g. WASO, sleep efficiency) or the performance outcomes (PVT), the standardized units are also shown in Figure 4. These calculations show that the study will be powered to detect small-to-moderate intervention effects on these outcomes, even if the design effect is as large as 2.

Outcome	MDE
Sleep hours (weekly)	1.75 / 1.92 / 2.21
Work hours (weekly)	1.39 / 1.53 / 1.76
Standardized units	0.30 / 0.32 / 0.37

Figure 4, *minimum effects on resident outcomes* detectable with 90% power in two-sided 5% tests, with design effects of 1.25/1.5/2.0.

1.8 DATA COLLECTION AND QUALITY ASSURANCE

1.8.1 Data Acquisition and Transmission

The DCC will develop a front-end component for data entry and transmission. Data would be entered at remote sites on secure mobile computing devices and transmitted a minimum of once a day via secure transfer. Data received will be automatically imported into SQL tables and integrated into our standard data system. Data will be available for review/updating via the standard study website within 24-48 business hours. The DCC will create and maintain a comprehensive programmer's guide for all the data systems it provides.

1.8.2 Data Collection and Editing

Capturing data electronically will reduce the risk of missing data. Furthermore, edits checks may be built into the instruments to flag out of range or inconsistent data. Further custom edit checks will be written and run on data received at the DCC. Each hour, the cumulative study data is subjected to these custom edits to ensure completeness, consistency and validity of the data. The results of the error-checking procedures are posted to the study website, which the study staff will check daily to both confirm that the DCC has successfully received all of the transmitted forms and to address errors that have been detected by the edit system. The Project Manager and other Clinical Coordinating Center (CCC) staff will be responsible for checking the daily data edits and working in conjunction with the sites to address them. This data entry/data editing process is monitored by a number of standard monitoring reports that will be available on the website. An audit trail is maintained which includes the date of change, time of change, description of data change, and the study ID of the person making the change.

The DCC will also serve as the Reading Center for the actigraphy and PVT datasets. DCC staff will perform an initial review of the files upon receipt and report any issues to the clinical site staff; later DCC staff will clean and merge the device data with the form data.

The DCC will prepare comprehensive ‘clean’ datasets for analysis, providing the analysis-ready files to the CCC and others as necessary, for statistical analysis purposes. The DCC will also ensure that any entity receiving the dataset will have a Data Use Agreement on file, signed by the appropriate institutional authority. The DCC will also be responsible for preparing a public release dataset to be housed within NHLBI’s BioLINCC (or a comparable public data repository) no later than 3 years after the end of the trial or 2 years after the main paper has been published, whichever comes first.

1.8.3 Data Management and Study Progress

The study website which includes a number of data management features and reports to enable clinical sites and study investigators to monitor data collection and study progress. Data Management features include a data inventory, lists of missing forms and queries by participant ID, and tools to resolve queries and update data via the website. Reports can be custom-designed to meet study needs and could include:

- Enrollment in aggregate and by site
- Demographics in aggregate and by site
- Early Discontinuation in aggregate and by site

1.8.4 Staff Training

To ensure inter-rater reliability, the physician observers and research nurses will undergo training in the detection of medical errors and adverse events prior to observing residents. The training will include evaluation of test cases designed to calibrate the observers’ assessments. At least twice during each 8 month period of data collection, for each wave of sites, the DCC will examine the rates of events by observer at each of the sites to determine if there is significant variability. Re-training of physician observers will occur as needed to ensure consistent reporting of suspected medical errors across staff and sites. Documentation of the analyses of event rates as well as the staff training will be maintained by the DCC. Staff will also be instructed on procedures for entering data into electronic data collection forms and transmitting these data to the DCC. Lastly, staff will be trained in the operation of the actigraphs as well as administration of the PVT tests.

1.9 PARTICIPANT RIGHTS AND CONFIDENTIALITY

1.9.1 Institutional Review Board (IRB) Review

This protocol and informed consent documents, and any subsequent modifications, will be reviewed and approved by the IRB or ethics committee responsible for oversight of each of the clinical sites, as well as the clinical and data coordinating centers.

1.9.2 Informed Consent Forms

Informed Consent documents will be maintained at individual sites and coordinators must ensure that they are using an up-to-date, IRB approved version of the consent form. A signed Informed Consent form will be obtained from each participant. The Informed Consent will describe the purpose of the study, the procedures to be followed, and the risks and benefits of participation and must comply with the requirements of the study site's Institutional Review Board. A copy of the Informed Consent will be given to each participant and provision of the copy will be documented in the study participant's record. The DCC will ensure that all clinical site informed consents are consistent and include required elements.

1.9.3 Participant Confidentiality

1.9.3.1 Clinical Sites

In order to protect confidentiality of the research participants as well as the patients they treat, both will be assigned a unique study identification number, which will be used to refer to them on all study forms. The list of codes linking resident-subjects' and patients' names and study IDs will be kept by each site PI on a separate password protected computer in a locked office at that site. Each site PI will be responsible for ensuring the security of these data. These identifying data will not be transmitted outside of the originating institution. Risk to patient confidentiality will likewise be minimized. All information will be de-identified at the point of data analysis.

1.9.3.2 Data Coordinating Center

The DCC follows standard operating procedures (SOPs) for computer system security to ensure the confidentiality and validity of study data. The SOPs are designed to prevent unauthorized access and limit authorized access to our computer systems and are in compliance with established standards for Information Technology Security. Our network is privately maintained, hardware fire-walled and none of the workstations or database servers can be directly addressed from outside the Local Area Network. Study website and database access requires a network domain account with appropriate account-specific permission on the database. All requests for new accounts and access to the database must be documented by a System Access Request Form signed by the project director.

1.10 COMMITTEES

1.10.1 Steering Committee

The Steering Committee will be responsible for all decisions concerning the scientific and technical conduct of the study. It will appoint the analysis and publications committee and writing groups, ensuring that information from the study is disseminated in the scientific literature and at scientific meetings. The committee will be chaired by Dr. Czeisler and will include the lead site investigator from the clinical and data coordinating center and each of the clinical sites. The Steering Committee will have one in-person meeting on an annual basis, and will meet by teleconference on at least a bi-monthly basis. The Steering Committee consists of the following members:

Charles A. Czeisler, PhD, MD	Chair, PI, CCC, Brigham and Women's Hospital
Christoper P. Landrigan, MD, MPH	PI, CCC, Brigham and Women's Hospital
Katie L. Stone, PhD	PI, DCC, California Pacific Medical Center
Susan Redline, MD, MPH	Chair, Sleep Research Network
Ken Wright, PhD	PI, University of Colorado
Jeffrey Segar, MD	PI, University of Iowa
John McGuire, MD	PI, Seattle Children's Hospital
Pearl Yu, MD	PI, University of Virginia
Sue Poynter, MD	PI, University of Cincinnati
Phyllis Zee, MD, PhD	PI,
Robert Smith, PhD	Project Scientist, NHLBI

1.10.2 Executive Committee

An Executive Subgroup of the Steering Committee will be responsible for decisions that require attention between Steering Committee Meetings and for major financial, administrative and operational decisions. The Executive Committee will meet via teleconference at least monthly and will consist of the following members:

Charles A. Czeisler, PhD, MD	PI, CCC, Brigham and Women's Hospital
Christoper P. Landrigan, MD, MPH	PI, CCC, Brigham and Women's Hospital
Katie L. Stone, PhD	PI, DCC, California Pacific Medical Center
Robert Smith, PhD	Project Scientist, NHLBI

1.10.3 Principal Planning Group

A planning group comprised of the Principal Investigators and project staff of both the CCC and DCC will be responsible for the development of study documents, including the forms, protocol and operations manual, as well as the day-to-day management of the trial. The Principal Planning Group will meet via teleconference at least monthly and will consist of the following members:

Charles A. Czeisler, PhD, MD	PI, CCC, Brigham and Women's Hospital
Christoper P. Landrigan, MD, MPH	PI, CCC, Brigham and Women's Hospital
Conor O'Brien	CCC, Brigham and Women's Hospital
Katie L. Stone, PhD	PI, DCC, California Pacific Medical Center
Dana R. Kriesel, MPH	PD, DCC, California Pacific Medical Center

1.10.4 Working Groups

Members of the Steering Committee will form working groups to manage and oversee specific aspects of the trial. These include: (1) Data Management and Data Quality; (2) Patient Safety and PICU Quality Outcomes; (3) Sleep, Performance and Health Outcomes; and (4) Education Outcomes. Working groups will meet on an as needed basis.

1.10.5 Data and Safety Monitoring Board

To monitor the study for the possibility that the intervention schedule has an adverse effect on safety, a multidisciplinary Data and Safety Monitoring Board (DSMB) comprised of financially disinterested members was appointed by NHLBI, with guidance from DCC. The board is comprised of five members with expertise in pediatric intensive care, pediatric critical care, sleep medicine, clinical trial design and biostatistics, bioethics, healthcare quality and outcomes research. The DSMB will monitor the quality and integrity of the data emerging from the proposed study; assess the adequacy of recruitment, compliance, follow-up and other aspects of study execution; and review interim analyses of the major outcomes of the trial to identify safety issues and make periodic recommendations on whether to continue, modify, or terminate the study. The DSMB will be an advisory board to NHLBI and the Steering Committee. The DSMB will work under a Charter that will be developed by the DCC, working in tandem with NHLBI. The DCC will support the DSMB by preparing interim reports for review at DSMB meetings; and by organizing scheduled and ad hoc meetings of the DSMB as needed. Please see the DSMB Charter for more details.

1.11 CONFLICT OF INTEREST

The DCC will be responsible for surveying the key personnel at each study site on an annual basis regarding any conflicts of interest that may have arisen. The Executive Committee will review the results of these annual surveys and determine if any action is necessary.

1.12 REFERENCES

1. Barger LK, Cade BE, Ayas NT, Cronin JW, Rosner B, Speizer FE, Czeisler CA. Extended work shifts and the risk of motor vehicle crashes among interns. N Engl J Med 2005;352:125-34

2. Lockley SW, Cronin JW, Evans EE, Cade BE, Lee CJ, Landrigan CP, Rothschild JM, Katz JT, Lilly CM, Stone PH, et al. Effect of reducing interns' weekly work hours on sleep and attentional failures. N Engl J Med 2004 Oct 28;351(18):1829-37

3. Landrigan CP, Rothschild JM, Cronin JW, Kaushal R, Burdick E, Katz JT, Lilly CM, Stone PH, Lockley SW, Bates DW, et al. Effect of reducing interns' work hours on serious medical errors in intensive care units. N Engl J Med 2004 Oct 28;351(18):1838-48

4. Ayas NT, Barger LK, Cade BE, Hashimoto DM, Rosner B, Cronin JW, Speizer FE, Czeisler CA. Extended work duration and the risk of self-reported percutaneous injuries in interns. JAMA 2006 Sep 6;296(9):1055-62

5. Grantcharov TP, Bardram L, Funch-Jensen P, Rosenberg J. Laparoscopic performance after one night on call in a surgical department: prospective study. B.M.J. 2001 Nov 24;323:1222-3

6. Barger LK, Ayas NT, Cade BE, Cronin JW, Rosner B, Speizer FE, Czeisler CA. Impact of extended-duration shifts on medical errors, adverse events, and attentional failures. PLoS Med 2006 Dec;3(12):e487

7. Arnedt JT, Owens J, Crouch M, Stahl J, Carskadon MA. Neurobehavioral performance of residents after heavy night call vs after alcohol ingestion. JAMA. 2005 Sep 7;294(9):1025-33

8. Philibert I. Sleep loss and performance in residents and nonphysicians: a meta-analytic examination. Sleep 2005 Nov 1;28(11):1392-402

9. Institute of Medicine. Resident Duty Hours: Enhancing Sleep, Supervision, and Safety. Washington, D.C.: The National Academies Press; 2009. 1-322 p.

10. Iglehart JK. Revisiting duty-hour limits--IOM recommendations for patient safety and resident education. N Engl J Med 2008 Dec 18;359(25):2633-5

11. ACGME. Common Program Requirements. 2010 Sep 26.

12. Griffin FA, Resar RK. Global Trigger Tool for Measuring Adverse Events: IHI Innovation Series white paper. Cambridge, MA: Institute for Healthcare Improvement; 2007.

13. National Coordinating Council for Medication Error Reporting and Prevention (NCC MERP). Index for Categorizing Errors. 2009.

14. Wyatt JK, Dijk DJ, Ronda JM, Jewett ME, Powell JW, Dinges DF, Czeisler CA. Interaction of circadian and sleep/wake homeostatic-processes modulate psychomotor vigilance test (PVT) performance. Sleep Res 26, 759. 1997.

15. Zeger SL. A regression model for time series of counts. *Biometrika.* 1988;75(4):621-629.

16. Armitage P, Hills M. The two-period crossover trial. *The Statistician.* 1982;31(2): 119-131.

Modifications to the original ROSTERS protocol:

Section 1.1.3 Aim 3: Assessment of motor vehicle crashes was changed from Johns Drowsiness Score to drive diaries and data gathered on the End of Rotation Survey.

Section 1.5, Study Procedures. Data collection with Optalert was deleted.

Section 1.5.3, Identification of patient risk factors. Severity of illness was changed from assessment by PRISM scores to use of ICD-9 codes.

Section 1.5.5.1, Optalert. This section was deleted because the Optalert was not used to assess driving safety.

Section 2.5.5.2, PVT. The measurement of alertness via the Karolinska Sleepiness Scale (KSS), before and after each PVT test, was added.

Section 1.5.6.2, Educational measures. The optional enrollment of residents in OpenPediatrics was added.

Section 1.6.2, Resident safety. This section was added at the request of the Data and Safety Monitoring Board (DSMB).

Section 1.7, Statistical considerations. Information regarding the analysis of Optalert data was removed.

Section 1.8.4, Staff training. This additional information was added regarding the reliability of the detection of medical errors as follows: To ensure inter-rater reliability, the physician observers and research nurses will undergo training in the detection of medical errors and adverse events prior to observing residents. The training will include evaluation of test cases designed to calibrate the observers' assessments. At least twice during each 8 month period of data collection, for each wave of sites, the DCC will examine the rates of events by observer at each of the sites to determine if there is significant variability. Re-training of physician observers will occur as needed to ensure consistent reporting of suspected medical errors across staff and sites. Documentation of the analyses of event rates as well as the staff training will be maintained by the DCC.

Section 1.10.1, Steering Committee. Changes in membership were noted.

Section 1.10.2, Executive Committee. Changes in membership were noted.

Section 1.10.3, Principal Planning Group. Changes in membership were noted.

Section 1.10.5, Data and Safety Monitoring Board. This information was added: . The board is comprised of five members with expertise in pediatric intensive care, pediatric critical care, sleep medicine, clinical trial design and biostatistics, bioethics, healthcare quality and outcomes research.

ROSTERS Statistical Analysis Plan

Definition of medical error outcomes. The primary outcome of the study will be resident-related harmful medical errors (preventable adverse events) and other medical errors (near misses), as determined by the adjudication process specified in the protocol. In brief, all potential medical errors and adverse events reported by the study sites will be reviewed by two site investigators, blinded to the site of origin. Physician reviewers will classify events as: (1) adverse event / harm; (2) intercepted near misses with potential for harm; (3) non-intercepted near miss with potential for harm; (4) error with little or no potential for harm; and (5) exclusion. If an event is classified as "adverse event / harm", the reviewers will then classify the level of harm and indicate to what extent the incident was preventable. Discordant reviews will be resolved by teleconference or assigned to a 3rd reviewer, as needed. Total ICU-wide preventable adverse events and near misses will be secondary outcomes.

Definition of secondary sleep outcomes. Actigraphy data and sleep diaries will be used to ascertain the minimum and mean 24-hour sleep times. If there is sufficient data collected, we will also measure performance on the Johns Drowsiness Score (JDS), based on results captured by Optalert. Specifically, we will calculate the time-weighted average JDS score for each commute; the proportion of the shift minutes spent with a JDS score above 4.5, the cutoff for impairment. Additional outcomes include variables derived from the Psychomotor Vigilance Test (PVT), including the numbers of lapses in vigilance, defined as a reaction time >500 ms, on the PVT.

Definition of safety outcomes. Safety outcomes consist of motor vehicle accidents (MVAs) and near misses, drowsy driving, needle sticks and other body fluid exposures, suicide ideation and resident SAEs. MVAs, near misses, and drowsy driving are captured both in the End of Rotation Survey (completed by all enrolled residents) and in the Drive Diary (completed by residents who drive to and from the hospital). Information about needle sticks, body fluid exposures, depression, and suicidal ideation are captured in the End of Rotation Survey as well. Resident SAEs are reported as they occur.

Preliminary analyses. Preliminary analyses include ongoing efforts to check and clean the data. The electronic data collection forms are designed to minimize missing data and skip pattern errors, by requiring responses to questions and only displaying sub-questions, based on responses to parent questions. A series of edit checks are run daily to detect data inconsistencies (eg, inconsistent dates). Furthermore, additional checks are run on the entire data set to check for inconsistencies or errors across forms. The Resident Sleep/Work and Drive Diary data is reviewed a few times per week for errors and missing data. Diary findings are communicated to the sites, who confirm any diary changes or additions with Residents.

Objectively measured sleep data will be collected using an actigraph (Motionlogger-L model; Ambulatory Monitoring Inc, Ardsley, NY). Actigraph files will be processed at the DCC using manufacturer-specific software (Action-W2 software, version 2.7.2288). Participating residents will complete sleep diaries for the time period they wore the actigraph, which includes time into and time out of bed and times the actigraph was removed. This information will be used in editing the actigraphy data files to set intervals for when the participant was in bed trying to sleep, and to delete time when the actigraph was removed. Inter-scorer reliability for editing the actigraphy data files has been previously found to be high in our group (intra-class coefficient = 0.95). (Blackwell 2005) Sleep-scoring algorithms available in the software will be used to determine sleep from wake times. (Cole 1992, Girardin 2001) These algorithms calculate a moving average, which takes into account the activity levels immediately prior to and after the current minute, to determine if the time point should be coded as sleep or wake. The summary measure of total sleep time across a 24-hr period will reflect data averaged over all 24-hr periods the device was worn in order to obtain a more representative characterization of usual sleep patterns during the rotation.

Once we have needed data from the site, we will determine whether the observers report similar rates of suspected events within centers, by schedule and period. We will also examine rates of discordance between adjudicators periodically to identify and address potential problems.

Assessment of balance between schedules. To assess balance of the treatment and control samples, we will compare resident characteristics by schedule, stratified by clinical center, assessing within-center and overall differences, as well as center by schedule interaction. These checks will use linear, logistic, and other generalized linear models as appropriate to the distribution of each of the baseline covariates. Baseline variables to be considered will include resident age, gender, race, post-graduate year (PGY2 vs PGY3), body mass index, lifestyle factors (including smoking, alcohol use and caffeine intake), physical activity and medical history (including sleep apnea).

Analysis of medical error outcomes. Rates per 1000 patient-days of the medical error outcomes will be compared by schedule using log-link Poisson models, with clinical center, period, and schedule as fixed effects, robust standard errors to account for potential over-dispersion, and the log of the number of patient-days at risk as an offset. Outcome events and total patient days will be summed over each study period for each clinical site, omitting missed observer shifts, based on reports of the missed shifts and corresponding patient censuses from the sites. We will also assess treatment-period interaction, a standard check with crossover trials.

Analysis of secondary outcomes. Comparisons of actigraphic outcomes will be made using regression models for summary outcomes constructed for each resident over their entire period, again controlling for clinical center and period. These outcomes will be normalized as necessary; generalized linear models appropriate for other outcome distributions, and/or bootstrap standard errors will be used if adequate normalization cannot be achieved.

Analyses of safety outcomes. Numbers of resident needle sticks, fluid exposures, MVAs, MVA near misses, MVAs with injuries, as well as numbers of each reported as related to fatigue or drowsy driving, will be compared by schedule using the methods proposed above for harmful medical errors. Offsets will be calculated as the numbers of resident days determined using work and drive diaries. Because MVA injuries are expected to too uncommon for meaningful statistical comparison, they will be presented using individual narrative descriptions.

Mediation of the effects schedule on medical errors by sleep measures. We will assess mediation of the effects of schedule on resident-related harmful medical errors by the effects of schedule on 24-hour sleep times and PVT scores. We will use simulation methods (Imai 2010) as implemented in the *mediation* package in R [R Foundation for Statistical Computing, Vienna].

Sensitivity analyses. Two sensitivity analyses will be performed.

- If imbalances are found between schedules in our preliminary assessment, described above, we will conduct sensitivity analyses, both for primary and secondary outcomes, in which we flexibly adjust for the potential confounders that are associated at $P<0.1$ with schedule either overall or within at least one center. These will include analyses adjusting for the number of residents with sleep disordered breathing, including sleep apnea; additional sensitivity analysis of the secondary outcomes will exclude these residents.
- If the numbers of missed observer shifts are imbalanced by schedule, we will conduct sensitivity analyses using multiple imputation of resident and total medical errors during unobserved shifts under both standard missing at random (MAR) and also plausible missing not at random (MNAR) assumptions. Specific MNAR scenarios will include doubling of the error rates during unobserved shifts in both schedules, first simultaneously, then one at a time.

Stopping Guidelines

<u>Stopping for benefit and harm.</u> The DSMB Guidelines for recommending continuation, modification, or stopping the study will explicitly include consideration of both statistical and non-statistical issues and include the possibilities of early stopping for benefit, stopping early for harm, stopping for futility, and modification of protocol.

<u>Primary outcome for assessing benefit and harm.</u> We propose to use the numbers of resident-related harmful medical errors as the primary outcome for both benefit and harm. Additional considerations for stopping for harm could include serious adverse events among residents, including needle sticks, fluid exposures, and MVAs.

<u>Rules for stopping.</u> We propose to use Lan-DeMets methods (Lan 1983) to define symmetric stopping boundaries; the DSMB may decide to use a more liberal rule for stopping for harm or to allocate type-I error differentially to efficacy and harm. We will use an alpha spending function resulting in boundaries similar to O'Brien-Fleming, making it difficult to stop early but inducing minimal shrinkage in the alpha for the end-of-trial analysis. We propose two interim analyses of treatment effects on the primary benefit/harm endpoint. The first analysis will occur when Sites 1 and 2 will have completed both schedules; and the second when Sites 1-4 will have completed both schedules. To aid in decision making, partial data will be provided for sites 3 and 4 at the first look, as well as for sites 5 and 6 at the second, but not included in the interim analysis; the rationale for these exclusions is that between-site differences would confound schedule if the sites contributing outcomes for only one schedule were included. With an overall type-I error rate of 5% allocated equally to benefit and harm, critical values of the Z-statistics at the first, second, and final analyses will be 2.67, 2.32, and 2.03 respectively, corresponding to 2-sided p-values of 0.008, 0.021, and 0.043. The DSMB will consider stopping the study early for efficacy or safety if the absolute Z-statistic at the first and second interim analysis exceeds the boundary values of 2.67 and 2.32 respectively.

<u>Stopping for futility.</u> The decision to stop for futility will be informed using simulations of conditional power, given the current data, to detect a 10% reduction in overall resident-related medical errors, and a 25% reduction in resident-related preventable AEs (Lan 1982; Halperin 1982). These analyses entail no inflation of the overall type-I error rate, and will be conducted in conjunction with the two interim analyses of benefit/harm. The DSMB will consider stopping the study for futility if estimated conditional power to detect the specified treatment effects falls below 50% for both outcomes. These decisions are complex and the guidelines may be further developed by the DSMB and NHLBI, with the assistance of the DCC.

Interim Analyses

An unblinded statistical analyst at the DCC will analyze the major endpoints at the time points specified above, and reports will be prepared for presentation to the DSMB. Except for designated NHLBI and DCC staff, all investigators and staff will remain blinded to interim results. Interim reports will be derived from the database as it exists on pre-specified dates, and full copies of the database at the time of each interim analysis will be archived. The interim analysis tables will be the same as the tables used in the open and closed sessions of the DSMB meetings, but results will be stratified by sites and include P-values. Please see a list of the tables below:

Interim Analysis Tables
Study Progress
Table 1: Summary of Residents Declined Participation
Table 2: Enrollment Status by Schedule, Wave and Site
Figure 1: Cumulative Enrollment by Month, by Site/Schedule
Table 3: Patient Days by Schedule, Wave and Site

Descriptives

Primary Outcome

Secondary Outcomes

References:

Blackwell T, Ancoli-Israel S, Gehrman PR, Schneider JL, Pedula KL, Stone KL. Actigraphy scoring reliability in the study of osteoporotic fractures. Sleep 2005;28:1599-605.

Cole RJ, Kripke DF, Gruen W, Mullaney DJ, Gillin JC. Automatic sleep/wake identification from wrist actigraphy. Sleep 1992;15:461- 9.

Girardin JL, Kripke DF, Mason WJ, Elliot JA, Youngstedt SD. Sleep estimation from wrist movement quantified by different actigraphic modalities. J Neurosci Methods 2001;105:185-91.

Imai K, Keele L, Yamamoto T. Identification, inference, and sensitivity analysis for causal mediation effects. Statistical Science, 2010:25(1):51-71.

Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. *Biometrika.* 1983;70:649-653

Lan KKG, Simon R, Halperin M. Stochastically curtailed tests in long-term clinical trials. *Comm Statist, Series C (Sequential Analysis).* 1982;1:207-219.

Halperin M, Lan KKG, Ware JH, Johnson WJ, DeMets DL. An aid to data monitoring in long- term clinical trials. *Controlled Clin Trials.* 1982;3:311-323.

Modifications to the original ROSTERS statistical analysis plan:

Additional analysis were requested by the DSMB as follows:

- Requested at April 2016 meeting. Show medical errors results broken down by severity of error.
- Requested at November 2016 meeting. Adjust models with medical errors outcomes by patient risk score.