Extended Work Shifts and Neurobehavioral Performance in Resident-Physicians

Shadab A. Rahman, PhD, MPH,[a,b] Jason P. Sullivan, BS,[a] Laura K. Barger, PhD,[a,b] Melissa A. St. Hilaire, PhD,[a,b] Conor S. O'Brien, BA,[a] Katie L. Stone, PhD,[c] Andrew J.K. Phillips, PhD,[a,b,d] Elizabeth B. Klerman, MD, PhD,[a,b] Salim Qadri, BS,[a] Kenneth P. Wright, Jr, PhD,[e] Ann C. Halbower, MD,[f] Jeffrey L. Segar, MD,[g] John K. McGuire, MD,[h] Michael V. Vitiello, PhD,[i] Horacio O. de la Iglesia, PhD,[j] Sue E. Poynter, MD, MEd,[k] Pearl L. Yu, MD,[l] Amy L. Sanderson, MD,[m] Phyllis C. Zee, MD, PhD,[o] Christopher P. Landrigan, MD, MPH,[a,b,n] Charles A. Czeisler, MD, PhD,[a,b] Steven W. Lockley, PhD,[a,b] THE ROSTERS STUDY GROUP

abstract

OBJECTIVES: Extended-duration work rosters (EDWRs) with shifts of 24+ hours impair performance compared with rapid cycling work rosters (RCWRs) that limit shifts to 16 hours in postgraduate year (PGY) 1 resident-physicians. We examined the impact of a RCWR on PGY 2 and PGY 3 resident-physicians.

METHODS: Data from 294 resident-physicians were analyzed from a multicenter clinical trial of 6 US PICUs. Resident-physicians worked 4-week EDWRs with shifts of 24+ hours every third or fourth shift, or an RCWR in which most shifts were ≤16 consecutive hours. Participants completed a daily sleep and work log and the 10-minute Psychomotor Vigilance Task and Karolinska Sleepiness Scale 2 to 5 times per shift approximately once per week as operational demands allowed.

RESULTS: Overall, the mean (± SE) number of attentional failures was significantly higher (P = .01) on the EDWR (6.8 ± 1.0) compared with RCWR (2.9 ± 0.7). Reaction time and subjective alertness were also significantly higher, by ∼18% and ∼9%, respectively (both P <.0001). These differences were sustained across the 4-week rotation. Moreover, attentional failures were associated with resident-physician–related serious medical errors (SMEs) (P =.04). Although a higher rate of SMEs was observed under the RCWR, after adjusting for workload, RCWR had a protective effect on the rate of SMEs (rate ratio 0.48 [95% confidence interval: 0.30–0.77]).

CONCLUSIONS: Performance impairment due to EDWR is improved by limiting shift duration. These data and their correlation with SME rates highlight the impairment of neurobehavioral performance due to extended-duration shifts and have important implications for patient safety.

[a]Division of Sleep and Circadian Disorders, Departments of Medicine and Neurology, Brigham and Women's Hospital, Boston, Massachusetts; [b]Division of Sleep Medicine, Harvard Medical School, Harvard University, Boston, Massachusetts; [c]California Pacific Medical Center Research Institute, San Francisco, California; [d]Monash Institute of Cognitive and Clinical Neurosciences, School of Psychological Sciences, Monash University, Clayton, Victoria, Australia; [e]Sleep and Chronobiology Laboratory, Department of Integrative Physiology, University of Colorado Boulder, Boulder, Colorado; [f]Department of Pediatrics, Children's Hospital Colorado and School of Medicine, University of Colorado, Aurora, Colorado; [g]University of Iowa Stead Family Children's Hospital, Iowa City, Iowa; [h]Seattle Children's Hospital, Seattle, Washington; [i]Departments of Psychiatry and Behavioral Sciences, and [j]Biology, University of Washington, Seattle, Washington; [k]Department of Pediatrics, Cincinnati Children's Hospital Medical Center and University of Cincinnati, Cincinnati, Ohio; [l]University of Virginia Children's Hospital, Charlottesville, Virginia; [m]Division of Critical Care Medicine, Department of Anesthesiology, Critical Care and Pain Medicine and [n]Division of General Pediatrics, Department of Medicine, Boston Children's Hospital, Boston, Massachusetts; and [o]Center for Circadian and Sleep Medicine, Department of Neurology, Feinberg School of Medicine, Northwestern University, Chicago, Illinois

WHAT'S KNOWN ON THIS SUBJECT: Postgraduate year 1 resident-physicians on extended-duration work rosters with shifts ≥24 hours can have impaired vigilance and increased attentional failures compared with those working on rapidly cycling work rosters that limit shifts to 16 consecutive hours.

WHAT THIS STUDY ADDS: In a multicenter clinical trial across 6 US PICUs, we found that, compared with working on extended-duration work roster shifts, resident-physicians (postgraduate year 2 and higher) had significantly lower rates of attentional failure and sleepiness on rapidly cycling work roster shifts.

The impact of extended-duration work rosters (EDWRs), commonly referred to as "on-call" shifts, on physician performance remains controversial. Interventions designed to reduce the frequency of or eliminate 24+ hour EDWRs have had mixed results. A number of studies have revealed that eliminating EDWRs improves patient safety and resident-physician education despite increases in hand-offs and reduced continuity of care (for review see Levine et al[1]). More recently, however, evaluations of policies designed to eliminate 16-hour shifts for postgraduate year (PGY) 1 resident physicians[2] have revealed no benefits in surgical morbidity and mortality (FIRST [Flexibility in Duty Hour Requirements for Surgical Trainees] Trial)[3] or medical patient mortality (iCOMPARE [Individualized Comparative Effectiveness of Models Optimizing Patient Safety and Resident Education])[4] associated with the limit. The Accreditation Council for Graduate Medical Education (ACGME) restored 24- to 28-hour EDWR for PGY 1 resident-physicians in 2017.[5]

Remaining awake for ≥16 hours causes substantial impairment in a range of performance measures.[6] Similarly, sleep deficiency due to chronic partial sleep loss also rapidly impairs performance.[7] Sleeping while on call, a practice generally recommended for physicians, does not alleviate performance degradation[8] and also risks introducing substantial impairment due to sleep inertia.[9] Finally, working 24+ hours requires physicians to work at an adverse circadian phase when performance impairment and subjective sleepiness is highest, an effect that is exacerbated when combined with extended time awake.[1,6]

Given the difficulty of evaluating the impact of alertness management solutions on medical error rates, putative intermediate outcomes have been used, including medical simulators[7,9,10] and simple neurobehavioral performance tests.[11–13] The Psychomotor Vigilance Task (PVT), or reaction time test variants thereof, requires reacting to repeated, randomly appearing visual or auditory stimuli as quickly as possible. In the medical setting, reaction time and attentional failures are sensitive to the effects of a single night of acute sleep deprivation while on call,[13–15] as well as the effects of chronic sleep loss associated with repeated extended-duration work shifts.[16,17] Although the PVT has been used to assess occupational performance impairment in other safety-sensitive professions, the validity of the PVT for predicting risks to medical performance is not documented. The aims of the current study were to assess the impact of a work schedule intervention designed to minimize extended-duration work shifts on PVT performance and to examine the relationship between PVT performance and medical error rates.

METHODS

Design Overview

Full details of the study design are reported elsewhere.[18] Briefly, the Randomized Order Safety Trial Evaluating Resident-Physician Schedules (ROSTERS) study was a multicenter cluster-randomized crossover clinical trial designed to evaluate the effectiveness of eliminating resident-physicians' traditional shifts of 24 hours or longer in 6 US PICUs. The units were studied under both the traditional EDWR, with regular 24- to 28-hour shifts or to a rapid cycle work roster (RCWR) intervention schedule that attempted to limit resident-physicians' scheduled continuous work to 16 hours maximum. The order of the schedule condition was randomly assigned within 3 pairs of units.[18] Each condition had a 4-month wash-in followed by 8 months of data collection.

Participants

All PGY 2 and PGY 3 resident-physicians working in the PICU over the study period were invited to participate and could volunteer for observation only (ie, direct observation by a physician while working in the unit) or full participation (ie, observation plus individual sleep, work, and performance data collection). The study did not include PGY 1 resident-physicians. Individual rotations in the PICU lasted ∼4 weeks, and resident-physicians could complete multiple PICU rotations during the study. Volunteers provided written consent and were offered an incentive (eg, iPad or cash equivalent) for participation. The institutional review board at each academic medical center as well as at Sutter Health (Data Coordinating Center [DCC]) and the Partners Human Research Committee (Clinical Coordinating Center [CCC]) approved the study. A Certificate of Confidentiality was obtained from the National Institutes of Health to protect the privacy of research participants.

A total of 312 resident-physicians volunteered for the individual-level data collection, as shown in Figure 1. Of these, a total of 358 data sets with both PVT and sleep diary data were analyzed from 294 unique individuals (173 women, age 29.4 ± 2.4 (SD) years; range 25–42 years, 2 unknown). 239 resident-physicians contributed data from 1 rotation in the analysis (n = 109 while working a traditional EWDR, n = 130 while working an intervention RCWR), 49 contributed 2 rotations (in order; n = 8 EDWR and RCWR; n = 10 RCWR and EDWR; n = 16 EDWR and EDWR; n = 15 RCWR and RCWR), 3 resident-physicians contributed 3 rotations (n = 2 RCWR, RCWR, and EDWR; n =
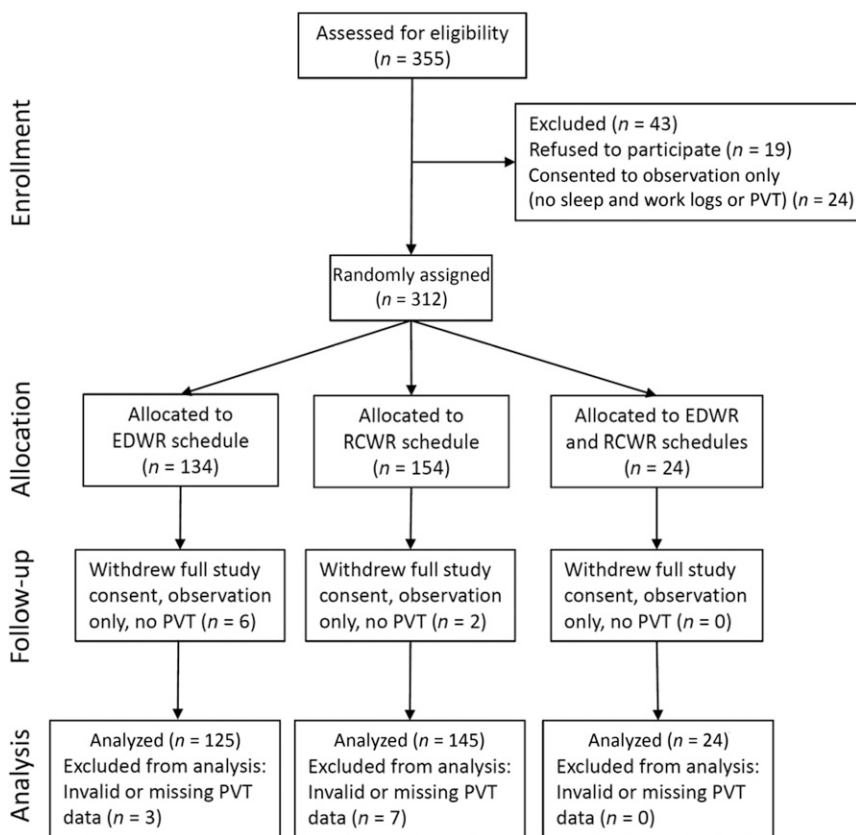
FIGURE 1
Consolidated Standards of Reporting Trials diagram.

1 RCWR, EDWR, and EDWR) and 3 contributed 4 rotations (n = 3 RCWR, RCWR, EDWR, and EDWR).

Shift Schedules

Full details of the shift schedules are provided elsewhere.[18] The traditional schedules required resident-physicians to work an EDWR of 24 to 28 scheduled consecutive hours every third or fourth shift (starting every fourth or fifth shift) for 4 weeks. During the EDWR, resident-physicians at 5 of 6 sites worked a 4-day rotation with 2 ~12-hour day shifts followed by an on-call shift beginning in the morning and ending 24 to 28 hours later. One site worked a 5-day rotation with 2 ~12-hour day shifts followed by a day off and then an on-call shift beginning at ~11 AM and ending ~24 hours later (Supplemental Fig 7).

During the intervention RCWR, resident-physicians at 5 of 6 sites worked a 4-day rotation with 2 ~11 to 15-hour day shifts followed by a 16-hour overnight shift that started in the evening and ended the next morning. The remaining site worked a 5-day rotation with 2 ~13-hour day shifts followed by a day off and then a 16-hour overnight shift that started at ~6 PM (Supplemental Fig 7).

Measurements and Outcomes

Objective Neurobehavioral Performance

Neurobehavioral performance was assessed by using a 10-minute PVT, a standard measure of sustained visual attention. Resident-physicians were asked to complete the PVT approximately every 5 hours during a shift, including the beginning and end of the shift, once per week, as operational demands allowed

(Supplemental Fig 7). Mean number of attentional failures (responses ≥500 milliseconds), mean reaction time, and mean of the slowest 10% of responses were used as measures of objective performance.

A total of 358 resident-physician rotations were analyzed (traditional n = 169, intervention n = 189), including a total of 4414 tests (traditional n = 2364, intervention n = 2050; mean ± SD tests per resident-physician rotation n = 14.0 ± 5.1 [range: 2–23] for the EDWR group and 10.9 ± 3.2 [range: 2–17] tests for the RCWR group). Each PVT test was assigned a time-of-day bin.[11] For the EDWR, the bins were the following: bin 1 (day 1): 6:00 AM to 2:59 PM; bin 2 (evening): 3:00 to 10:59 PM; bin 3 (night): 11:00 PM to 5:59 AM; bin 4 (day 2): 6:00 AM to 2:59 PM. For the RCWR, the bins had the same times, but data were binned in relation to the long day-call (tests between 6:00 AM and 10:59 PM [bins 1–2]) or long night-call (tests between 3:00 PM on day 1 and 2:59 PM on day 2 [bins 2–4]). Note, bin 1 spanned 5:00 AM to 2:59 PM for 1 site because of an earlier shift start time. Each resident did not always contribute to each bin; when 2 tests occurred in the same bin, they were averaged for that individual in subsequent analysis. Unequal contributions were accounted for in the statistical analysis as described in detail in the Supplemental Information; briefly, all 4 regression models were adjusted for the log number of tests taken by each individual.

Subjective Sleepiness, Sleep, and Work

Resident-physicians completed the Karolinska Sleepiness Scale (KSS) before each PVT. KSS ≥8 was defined as severe sleepiness on the basis of previous work associating these levels consistently to accident risk.[19] Resident-physicians also completed a daily electronic sleep and work log.[20]

Medical Errors and Resident-Physician Workload

Full details of the medical errors assessment are available elsewhere.[21] Briefly, an intensive 4-pronged approach was used to gather data on suspected events: (1) direct observation of resident-physicians; (2) daily review of unit charts by trained research nurses; (3) collection of voluntary reports from unit staff; and (4) collection of formal hospital incident reports. The retrospective review included every chart of every patient cared for by residents, and the study units were reviewed 5 days per week (Monday review included a review of the weekend). Two people from a team of independent, trained physician-reviewers, blinded to study condition, reviewed each event and defined whether it was an adverse event and caused harm; a potential adverse event, also known as "near miss"; error with little or no potential for harm; or excluded. Harmful events were further classified by the level of harm and preventability. In the current analysis, only directly traceable resident-physician–related preventable adverse events, classified as resident-physician–related serious medical errors (SMEs), were used. The ratio of ICU patients per resident-physician (IPRP) was calculated as a proxy measure of resident-physician workload.[18,21] See Supplemental Information for additional information.

Statistical Analysis

The main a priori performance hypothesis (major secondary end point of the overall trial) was that resident-physicians' risk of objective performance attentional failures, as assessed through PVT lapses, would be lower on the RCWR as compared with the EDWR. Exploratory analyses were performed on PVT mean reaction time and mean of the slowest 10% of responses and KSS. All 3 were hypothesized to be higher during the

EDWR. Additional exploratory analyses were performed to compare the main performance outcome (attentional failures on the PVT) with medical error rates. Neurobehavioral performance and error rates were calculated for each resident and correlated at the resident-physician level. See Supplemental Information for additional information.

RESULTS

Effect of Study Condition

There was a significant main effect of study condition for all outcomes, with significantly fewer attentional failures (mean $\pm$ SE, 2.9 $\pm$ 0.7 vs 6.8 $\pm$ 1.0; P = .01), faster mean reaction time (283.7 $\pm$ 6.3 vs 336.2 $\pm$ 7.2; P < .0001), slowest 10% reaction time (509.9 $\pm$ 24.6 vs 683.8 $\pm$ 31.8; P < .0001), and lower sleepiness ratings (4.4 $\pm$ 0.1 vs 4.8 $\pm$ 0.1; P < .0001) on the RCWR as compared with the EDWR, respectively. The distribution and regression-estimated mean ($\pm$

SE) for each outcome is revealed in Fig 2. Furthermore, there was a 43% decrease in the odds of having a 10-min PVT session with at least 2 attentional failures[22] on the EDWR compared with the RCWR (odds ratio: 0.57 [95% confidence interval (CI): 0.50–0.64]; P < .05) and a 29% reduction in the odds of reporting severe sleepiness (KSS $\geq$8; odds ratio: 0.71 [95% CI: 0.62–0.78]; P < .0001).

Effect of Time on Shift

After evaluating the overall effect of the intervention on the outcomes, we further evaluated the effect of the intervention on each of the outcomes across the day and night shifts. During the day shift (bins 1–2), there was a significant difference (P < .0001) in attentional failures between conditions, but no difference in mean or slowest 10% reaction times (Fig 3). There was, however, a significant interaction between time on shift and study condition for the
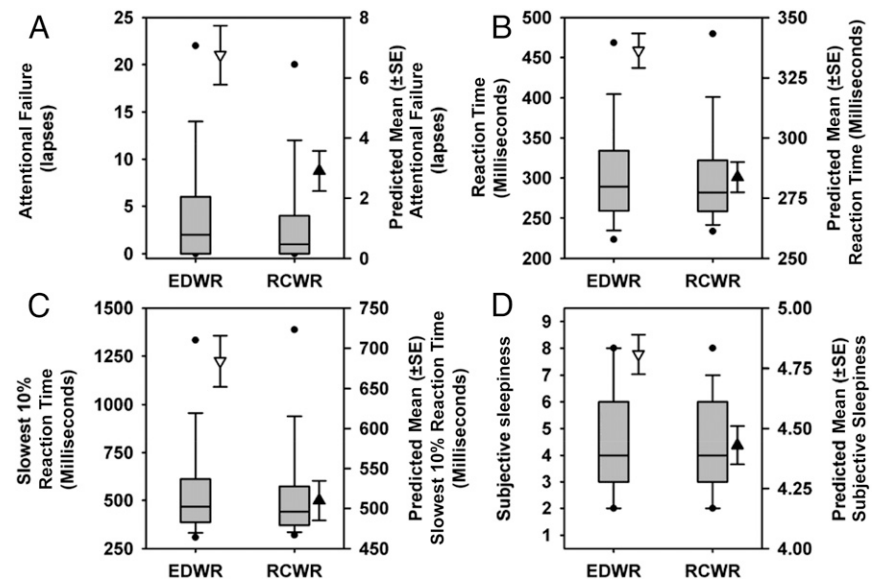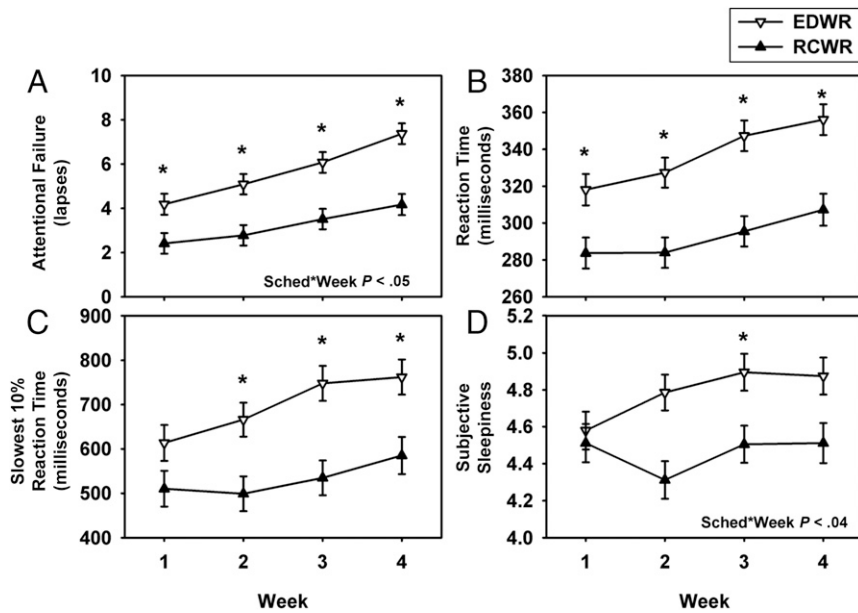

FIGURE 2
Objective performance and subjective sleepiness ratings during the traditional EDWR and intervention RCWR schedules. The box and whisker plots depict the median (solid horizontal line), 25th and 75th percentiles (box limits), 10th and 90th percentiles (whiskers), and fifth and 95th percentiles (black circle) along with the predicted values ($\pm$ SE) of each outcome estimated from the regression analyses shown next to each box and whisker plot under the EDWR (white triangles) and RCWR (black triangles) schedules for attentional failures (A) (lapses, reaction time >500 milliseconds), reaction time (B), slowest 10% reaction time (C), and subjective sleepiness ratings (D).
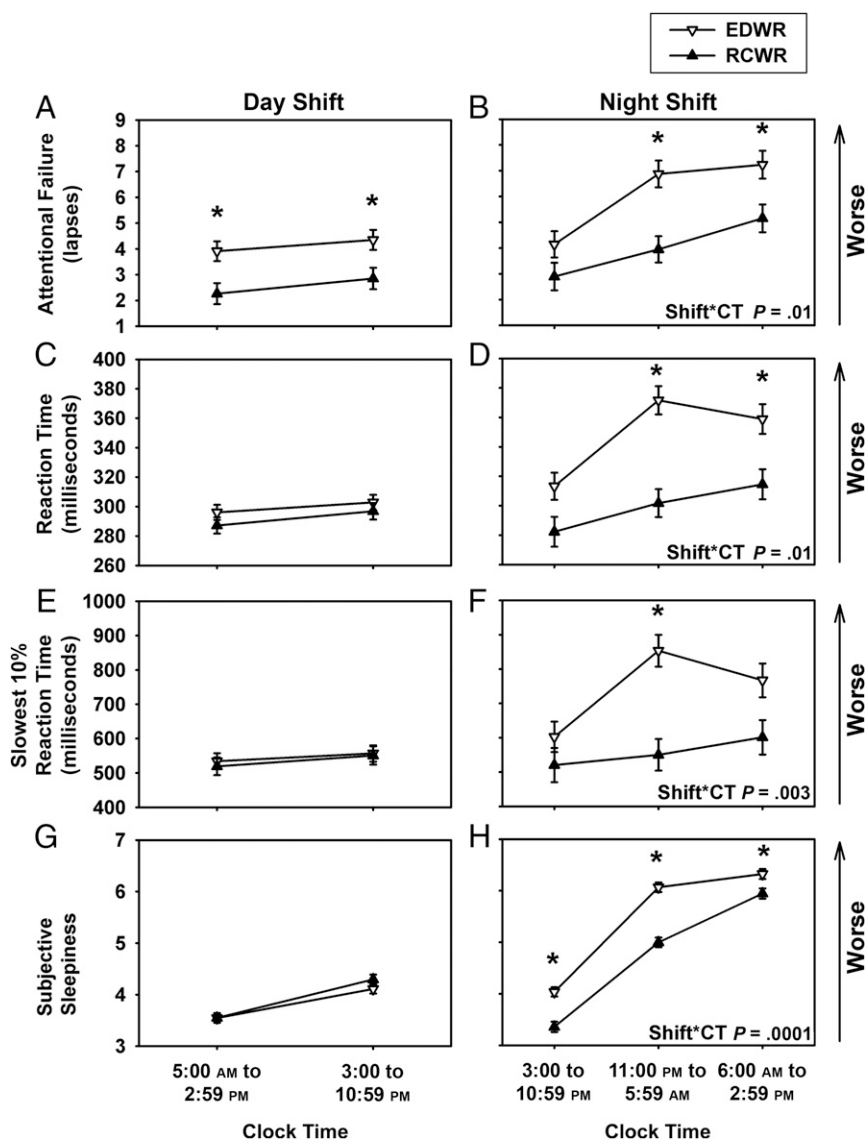
FIGURE 3
Objective performance measures and subjective sleepiness ratings during the traditional EDWR and intervention RCWR schedules during the day and night. Least-square means (± SE) are estimated from the mixed-effects model under the EDWR (white triangles) and RCWR (black triangles) schedules during the day (top row) and night (bottom row) for attentional failures (A and B) (lapses, reaction time >500 milliseconds), reaction time(C and D), slowest 10% reaction time (E and F), and subjective sleepiness ratings (G and H) plotted across the day. Note that under the day shift interval of 5:00 AM to 2:59 PM, Cincinnati started at ∼5:00 AM, and all others started at ∼6:00 AM. *P < .05, significant differences between conditions at each time point with Tukey-Kramer adjusted post hoc analysis. Shift*CT, interaction between the fixed effects shift and clock time.

nightshift (bins 2–4, 3:00–2:59 PM), with post hoc analyses revealing that all measures of performance were significantly worse overnight (bin 3, 11:00 PM to 5:59 AM) on the EDWR compared with the RCWR (all $P \leq$ .01; Fig 3). Sleepiness during the nightshift was also higher on the

EDWR, including overnight (bin 3, 11:00 PM to 5:59 AM) (Fig 3).

Effect of Successive Shifts

We further assessed the impact of cumulative shifts across the 4 weeks of the resident rotations and compared between the EDWR and

RCWR conditions. As revealed in Fig 4, there was an increase in attentional failures with increasing weeks for both schedules ($P <$ .0001) with post hoc analyses revealing a significant worsening of attentional failures during the third and fourth weeks of each schedule (both $P <$ .05). Attentional failures were consistently lower on the RCWR than on the EDWR ($P <$ .0001), however, and the worsening of attentional failures across the 4 weeks was significantly reduced on the RCWR compared with the EDWR ($P <$ .05). Other measures of performance and sleepiness also deteriorated with study week ($P <$ .05; Fig 4).

Moreover, we assessed whether the effect of the scheduling intervention was modulated by time on shift, when on the night shift or day shift, and across the 4-week rotation. When examined by time of day and week of rotation, both conditions revealed a gradual increase in attentional failures with increasing duration on shift, primarily during the night hours (11:00 PM to 05:59 AM) in the third and fourth weeks of the rotation (Fig 5). Similar results were observed for subjective sleepiness (Supplemental Fig 8).

Attentional Failures and Medical Error Rates

The average rate of resident-physician–related SMEs increased by∼1% with each additional PVT attentional failure (rate ratio [RR] 1.01 [95% CI: 1.00–1.02]; *P* =.04, Supplemental Fig 9) after adjusting for site, schedule and order of randomization. Although neurobehavioral performance improved, which, on the basis of the observed inverse association between neurobehavioral performance (attentional failures) and resident-physician–related SMEs, would indicate a reduction in the rate of medical errors under the intervention condition, we instead found that the average rate of resident-

FIGURE 4
Objective performance measures and subjective sleepiness ratings during the traditional EDWR and intervention RCWR schedules across the 4-week study. Least-square means (± SE) are estimated from the mixed-effects model under the EDWR (white triangles) and RCWR (black triangles) schedules for attentional failures (A) (lapses, reaction time >500 milliseconds), reaction time (B), slowest 10% reaction time (C), and subjective sleepiness ratings (D) plotted across the 4-week rotation. *P < .05, significant differences between conditions at each time point with Tukey-Kramer adjusted post hoc analysis. Sched*Week, interaction between the fixed effects schedule and week.

physician–related SMEs per day in the study was 70% higher under the RCWR compared with the EDWR (RR 1.70 [95% CI: 1.48–1.96]; P <.0001). There were, however, a number of potential confounders that could account for the discrepancy between the neurobehavioral performance data and rates of SME across schedules. Notably, resident-physician workload increased by 22% to 30% across the 6 sites under the RCWR compared with the EDWR.[18]

Given that increase in workload, we investigated the relationship between resident-physician workload and the rate of SMEs. As revealed in Fig 6, we found that the rate of SMEs increased significantly with workload in this group. Therefore, we analyzed the impact of the intervention on the rates of resident-physician–related SMEs while adjusting for workload (IPRP). After adjusting across sites for average IPRP during the EDWR and RCWR intervals, the rate of resident-

physician–related SMEs was significantly lower for the RCWR than the EDWR (RR 0.48 [95% CI: 0.30–0.77], P = .002) in this cohort.

DISCUSSION

We found that eliminating 24- to 28-hour EDWRs and scheduling 16-hour maximum work shifts improved neurobehavioral performance and reduced sleepiness in resident-physicians over a 4-week PICU rotation, particularly overnight. Repeated exposure to EDWR resulted in cumulative performance decrements such that by the fourth week of the rotation, performance and attentional failures at the start of shift were worse than performance at the end of shift in the first week. Although performance also deteriorated over the 4-week RCWR rotation, the severity was much reduced compared with the EDWR. The rate of resident-physician–related SME was correlated

to the average number of attentional failures, suggesting that this relatively simple test is operationally relevant in assessing physician performance and the likelihood of committing a preventable adverse event. Contrary to the observed improvement in neurobehavioral performance, however, the overall rate of resident-physician–related SME increased under the RCWR compared with the EDWR. After adjusting for workload, which increased between 22% and 30% across sites under the RCWR, the rates of SMEs on the RCWR decreased in this cohort by >50%.

Improved sleep during the RCWR may account in part for the improvements observed in neurobehavioral performance and reduction in risk of SMEs when adjusted for workload. As we reported elsewhere, under the EDWR, resident-physicians slept significantly less (mean ± SD, 47.5 ± 4.5 hours per week) as compared with the RCWR (49.1 ± 4.9 hours per week).[23] Importantly, the distribution of sleep loss relative to work was also altered: During the EDWR, 10% of work hours were preceded by ≤2 hours of sleep in the preceding 24 hours, as compared with only 4% during the RCWR. Furthermore, the percentage of work hours preceded by ≥8 hours of sleep in the last day was higher during the RCWR (38%) than the EDWR (23%).[23]

The direct demonstration of a significant correlation between attentional failures as measured by the PVT and medical errors is important because the PVT is easier to measure and has been used extensively to evaluate the impact of work schedules on performance. The correlation observed between this laboratory measure and the risk of medical errors is consistent with a previous meta-analysis revealing that 24 hours of sleep loss had a similar impact on laboratory measures of performance in non-physicians as it did on clinical
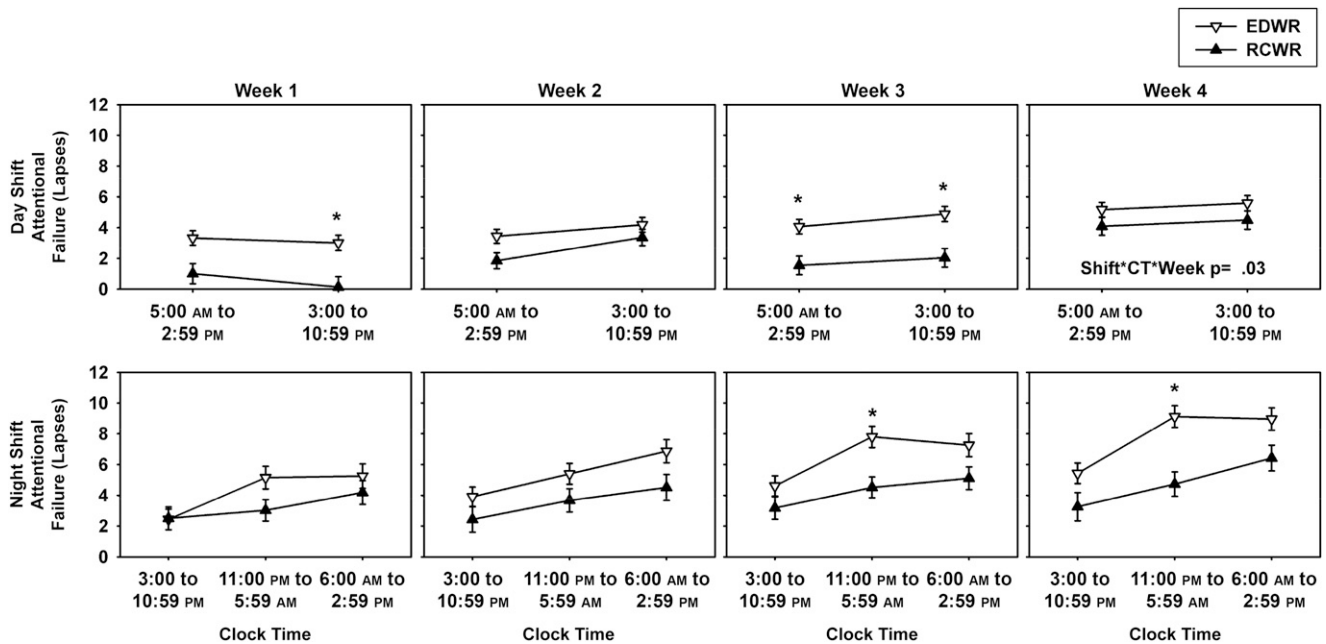
FIGURE 5
Objective performance during the traditional EDWR and intervention RCWR schedules during the day and night across the 4-week study. Least-square means ($\pm$ SE) are estimated from the mixed-effects model under the EDWR (white triangles) and RCWR (black triangles) schedules during the day (top row) and night (bottom row) for attentional failures (lapses, reaction time >500 milliseconds) plotted across the 4 weeks. Note that under the day shift interval of 5:00 AM to 2:59 PM, Cincinnati started at ~5:00 AM, and all others started at ~6:00 AM. *$P < .05$, significant differences between conditions at each time point with Tukey-Kramer adjusted post hoc analysis. Shift*CT*Week, interaction between the fixed effects shift, clock time, and week.

performance among physicians.[24] Our finding that PVT attentional failures and medical error rates are correlated suggests that, at least at a group level, PVT performance may be a useful workplace metric to predict medical performance.

In previous studies, researchers have examined psychomotor performance in medical residents using the 10-minute PVT reported herein and have found a similar degree of impairment due to EDWR.[11] The level of impairment in resident-physicians post call is equivalent to that observed when rested but with a blood alcohol concentration of 0.04% to 0.05%.[17] Similar performance impairments have been observed post call by using a shorter 3-minute test,[14] and, in a negative noninferiority analysis of PVT performance in PGY 1 resident-physicians, there was also evidence of poorer performance post call in programs with EDWRs, as compared with those which complied with a 16-

hour duty limit.[25] Collectively, these and the current data are consistent in revealing that the PVT can detect impairments in resident-physician vigilance when working EDWR and that abolition of 24-hour shifts reduces this impairment. In addition, we note that the average KSS scores in this trial were elevated at the start of shift by the fourth week compared with the first week (Fig 5), suggesting that the KSS combined with the PVT may be a useful indicator of impairment under such an extreme challenge.[16] Such "fitness-for-duty" testing may therefore be able to prevent avoidable sleepiness-related medical errors from reaching the patient.

In our study, we address the shortcomings of previous work[20,26] by studying several hundred resident-physicians at multiple sites and including more experienced PGY 2 and PGY 3 resident-physicians. There remain limitations, however. First, the RCWR schedule was not uniformly

implemented and, in some sites, resulted in work hours that may not be considered an improvement over the EDWR. Second, workload at baseline was highly variable between sites, with resident-physicians at some PICUs caring for as many as ~10 patients per resident-physician, which could affect overall sleepiness and medical performance, although the intervention still resulted in a clear improvement in indicators of sustained attention. Although we analyzed the impact of resident-physician workload on rates of SMEs using IPRP estimates as a proxy measure of workload, and our results suggest that variability in resident-physician workload may have impacted the effectiveness of our scheduling intervention across sites, we did not set out to explicitly test the effects of workload on our intervention. The significant interaction between site and average workload (IPRP) makes it difficult to disentangle conclusively the main
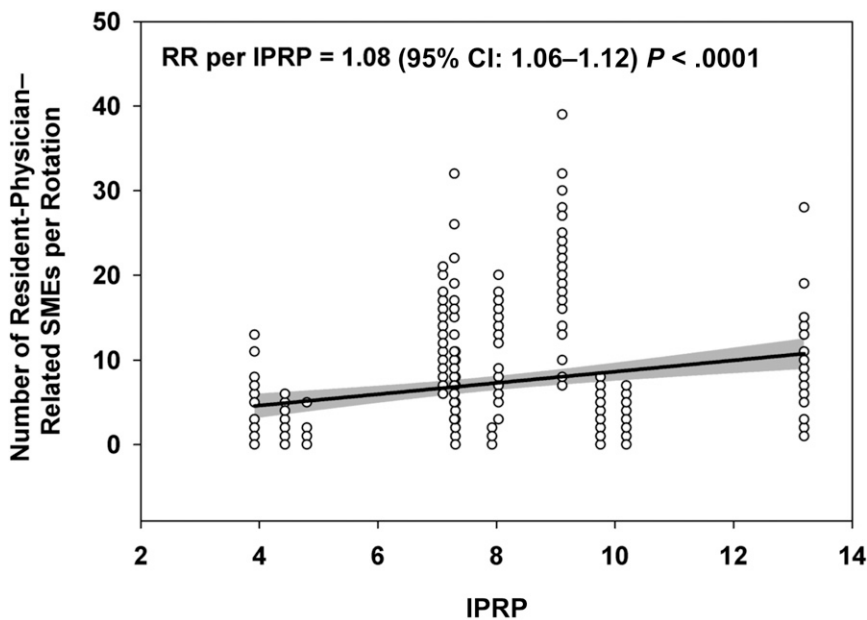
FIGURE 6

Association between resident-physician–related SMEs and workload. Resident-physician–related SMEs increased with resident-physician workload (IPRP). The best fit regression line is shown with 95% CI as a gray shaded band.

independent effects of the intervention and workload on medical error rates. Furthermore, IPRP does not take into account other caregivers who may have impacted overall workload at each site, and our analysis was restricted to resident-physician–related errors and not overall unit-wide errors. Future studies with more exact measures of workload are needed to confirm these exploratory findings and should explicitly test the interaction between workload and schedule design. Third, individual resident-physicians did not contribute equally to the data distribution, a problem inherent in real-world studies. We addressed this by using linear mixed models and generalized estimating equations that are robust to missing data and unbalanced designs[27,28] as well as including offsets to further account for the unequal number of observations for each participant. Despite these limitations, the impact of the RCWR intervention schedule on neurobehavioral and medical performance, together with the role of workload on medical error rates,

reveals that both physiologic and external variables need to be carefully considered when designing and implementing work schedules.

In 2011, on the basis of the 2009 recommendations of the National Academy of Medicine, the ACGME mandated a limit of 16 hours of continuous work for PGY 1 resident-physicians (although 88 hour work weeks, averaged over 4 weeks, were still permitted in more senior resident-physicians, and many programs conducting research were exempted from the regulations).[1,20,26] The 2011 ACGME work hour restrictions, however, were unfunded, contrary to the National Academy of Medicine recommendations, which specifically warned against implementation without additional staffing to prevent an increase in workload and workload-related risk. Other recommendations (for example, ensuring adequate time off between shifts) were not included in the ACGME regulations, allowing rosters that could induce severe chronic sleep deficiency.[29] In several large

studies (but not all[30]), it has since been reported that the limitations had little clinical benefit,[3,4] but these did not account for workload, patient census, or the way in which the schedules were implemented. In 2017, the limitations on PGY 1 duty hours were reversed, again permitting 28 hours of continuous duty for all residents.[5]

In our study, we indicate that this policy reversal is premature. Both extended-duration work shifts and increased workload compose a serious threat to patient safety, and both of these factors should be considered in the design of work rosters for resident-physicians. Moreover, the deterioration in cognitive performance associated with EDWRs is of particular concern for graduate medical education, given the essential role of sleep for learning and memory.[31] We found that the substantial impairment in objective performance and subjective sleepiness inherent in an EDWR is significantly improved by implementation of a RCWR that limits shift duration, including a significant reduction in attentional failures. These data, and their correlation with serious medical error rates, highlight the impairment of neurobehavioral performance that occurs when resident-physicians work extended-duration shifts and have important implications for both resident-physician and patient safety.

ABBREVIATIONS

ACGME: Accreditation Council for Graduate Medical Education
CCC: clinical coordinating center
CI: confidence interval
DCC: data coordinating center
EDWR: extended-duration work roster
IPRP: ICU patients per resident-physician
KSS: Karolinska Sleepiness Scale
PGY: postgraduate year
PVT: Psychomotor Vigilance Task
RCWR: rapid cycling work roster
ROSTERS: Randomized Order Safety Trial Evaluating Resident-Physician Schedules
RR: rate ratio
SME: serious medical error

Data can be obtained from the National Heart, Lung, and Blood Institute Biologic Specimen and Data Repository Information Coordinating Center at https://biolincc.nhlbi.nih.gov/home/. These are cleaned, analysis-ready data sets containing data collected during the trial. Data are at the individual level and deidentified. They include study protocol, annotated data collection forms, data dictionary, and calculated variable documentation. Data availability will begin no later than 3 years after the end of the trial or 2 years after the main findings have been published, whichever comes first. The study ended March 2017, and data were finalized (adjudicated) October 2017. There is no end date.

Dr Rahman conducted data analysis, drafted the initial manuscript, and reviewed and revised the manuscript; Mr Sullivan coordinated data collection and conducted data analysis; Dr Barger coordinated and supervised data collection and analysis; Drs St. Hilaire, Phillips, and Klerman conducted data analysis; Mr O'Brien and Mr Qadri coordinated data collection; Dr Stone conducted data coordinating center oversight and coordinated data collection and analysis; Drs Wright, Halbower, Segar, McGuire, Vitiello, de la Iglesia, Poynter, Yu, Sanderson, and Zee were site principal investigators and coordinated and supervised data collection; Drs Landrigan and Czeisler were study principal investigators, conceived and designed the study, and coordinated and supervised data collection and analysis; Dr Lockley coordinated and supervised data collection and analysis, drafted the initial manuscript, and reviewed and revised the manuscript; and all authors provided a critical revision of the intellectual content and approved the final manuscript as submitted. The corresponding author (Dr Rahman) had full access to all the data in the study and had the final responsibility for submission of the manuscript; he affirms that the manuscript is an honest, accurate, and transparent account of the study being reported, that no important aspects of the study have been omitted, and that any discrepancies from the study as originally planned have been explained.

This trial has been registered at www.clinicaltrials.gov (identifier NCT02134847).

Address correspondence to Shadab A. Rahman, PhD, MPH, Division of Sleep and Circadian Disorders, Departments of Medicine and Neurology, Brigham and Women's Hospital, 221 Longwood Ave, Boston, MA 02115. E-mail: sarahman@rics.bwh.harvard.edu

REFERENCES

1. Levine AC, Adusumilli J, Landrigan CP. Effects of reducing or eliminating resident work shifts over 16 hours: a systematic review. *Sleep*. 2010;33(8):1043–1053

2. The Accreditation Council for Graduate Medical Education. History of duty hours. Available at: https://www.acgme.org/What-We-Do/Accreditation/Clinical-Experience-and-Education-formerly-Duty-Hours/History-of-Duty-Hours. Accessed November 30, 2020

3. 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(8):713–727

4. Silber JH, Bellini LM, Shea JA, et al.; iCOMPARE Research Group. Patient safety outcomes under flexible and standard resident duty-hour rules. *N Engl J Med*. 2019;380(10):905–914

5. Burchiel KJ, Zetterman RK, Ludmerer KM, et al. The 2017 ACGME common work hour standards: promoting physician learning and professional development in a safe, humane environment. *J Grad Med Educ*. 2017;9(6):692–696

6. Killgore WDS. Effects of sleep deprivation on cognition. *Prog Brain Res*. 2010;185:105–129

7. Goel N. Neurobehavioral effects and biomarkers of sleep loss in healthy adults. *Curr Neurol Neurosci Rep*. 2017;17(11):89

8. St Hilaire MA, Anderson C, Anwar J, et al.; Harvard Work Hours Health and Safety Group. Brief (<4 hr) sleep episodes are insufficient for restoring performance in first-year resident physicians working overnight extended-duration work shifts. *Sleep (Basel)*. 2019;42(5):zsz041

9. Jewett ME, Wyatt JK, Ritz-De Cecco A, Khalsa SB, Dijk DJ, Czeisler CA. Time course of sleep inertia dissipation in human performance and alertness. *J Sleep Res*. 1999;8(1):1–8

10. Lockley SW, Landrigan CP, Barger LK, Czeisler CA; Harvard Work Hours Health and Safety Group. When policy meets physiology: the challenge of reducing resident work hours. *Clin Orthop Relat Res*. 2006;449(449):116–127

11. Anderson C, Sullivan JP, Flynn-Evans EE, Cade BE, Czeisler CA, Lockley SW. Deterioration of neurobehavioral performance in resident physicians during repeated exposure to extended duration work shifts. *Sleep*. 2012;35(8):1137–1146

12. Gordon JA, Alexander EK, Lockley SW, et al.; Harvard Work Hours, Health, and Safety Group (Boston, Massachusetts). Does simulator-based clinical performance correlate with actual hospital behavior? The effect of extended work hours on patient care provided by medical interns. *Acad Med*. 2010;85(10):1583–1588

13. Grantcharov TP, Bardram L, Funch-Jensen P, Rosenberg J. Laparoscopic performance after one night on call in a surgical department: prospective study. *BMJ*. 2001;323(7323):1222–1223

14. Basner M, Dinges DF, Shea JA, et al. Sleep and alertness in medical interns and residents: an observational study on the role of extended shifts. *Sleep*. 2017;40(4):zsx027

15. Persico N, Maltese F, Ferrigno C, et al. Influence of shift duration on cognitive

performance of emergency physicians: a prospective cross-sectional study. *Ann Emerg Med.* 2018;72(2):171–180

16. Anderson C, Ftouni S, Ronda JM, Rajaratnam SMW, Czeisler CA, Lockley SW. Self-reported drowsiness and safety outcomes while driving after an extended duration work shift in trainee physicians. *Sleep.* 2018;41(2):zsx195

17. Arnedt JT, Owens J, Crouch M, Stahl J, Carskadon MA. Neurobehavioral performance of residents after heavy night call vs after alcohol ingestion. *JAMA.* 2005;294(9):1025–1033

18. Blackwell T, Kriesel DR, Vittinghoff E, et al.; ROSTERS Study Group. Design and recruitment of the randomized order safety trial evaluating resident-physician schedules (ROSTERS) study. *Contemp Clin Trials.* 2019;80:22–33

19. Akerstedt T, Anund A, Axelsson J, Kecklund G. Subjective sleepiness is a sensitive indicator of insufficient sleep and impaired waking function. *J Sleep Res.* 2014;23(3):240–252

20. Lockley SW, Cronin JW, Evans EE, et al.; Harvard Work Hours, Health and Safety Group. Effect of reducing interns' weekly work hours on sleep and attentional failures. *N Engl J Med.* 2004; 351(18):1829–1837

21. 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(26):2514–2523

22. St Hilaire MA, Kristal BS, Rahman SA, et al. Using a single daytime performance test to identify most individuals at high-risk for performance impairment during extended wake. *Sci Rep.* 2019;9(1):16681

23. Barger LK, Sullivan JP, Blackwell T, et al.; ROSTERS Study Group. 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

24. Philibert I. Sleep loss and performance in residents and nonphysicians: a meta-analytic examination. *Sleep.* 2005; 28(11):1392–1402

25. Basner M, Asch DA, Shea JA, et al.; iCOMPARE Research Group. Sleep and alertness in a duty-hour flexibility trial in internal medicine. *N Engl J Med.* 2019;380(10):915–923

26. 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(18):1838–1848

27. Fitzmaurice GM, Ravichandran C. A primer in longitudinal data analysis. *Circulation.* 2008;118(19):2005–2010

28. Patel CB. Letter by Patel regarding article, "A primer in longitudinal data analysis". *Circulation.* 2009;120(4):e25

29. Cohen IG, Czeisler CA, Landrigan CP. Making residency work hour rules work. *J Law Med Ethics.* 2013;41(1): 310–314

30. Landrigan CP, Czeisler CA. Patient safety under flexible and standard duty-hour rules. *N Engl J Med.* 2019;380(24): 2379–2380

31. Stickgold R, Walker MP. Sleep-dependent memory triage: evolving generalization through selective processing. *Nat Neurosci.* 2013;16(2): 139–145