

REVIEW ARTICLE

# Systematic Review of the Impact of Physician Work Schedules on Patient Safety with Meta-Analyses of Mortality Risk

Matthew D. Weaver, PhD, MPH; Jason P. Sullivan; Christopher P. Landrigan, MD, MPH; Laura K. Barger, PhD

Resident physician work hour limits continue to be controversial. Numerous trials have come to conflicting conclusions about the impact on patient safety of eliminating extended duration work shifts. We conducted meta-analyses to evaluate the impact of work hour policies and work schedules on patient safety. After identifying 8,362 potentially relevant studies and reviewing 688 full-text articles, 132 studies were retained and graded on quality of evidence. Of these, 68 studies provided enough information for consideration in meta-analyses. We found that patient safety improved following implementation of the Accreditation Council for Graduate Medical Education's 2003 and 2011 resident physicians work hour guidelines. Limiting all resident physicians to 80-hour work weeks and 28-hour shifts in 2003 was associated with an 11% reduction in mortality ( $p < 0.001$ ). Limited shift durations and shorter work weeks were also associated with improved patient safety in clinical trials and observational studies not specifically tied to policy changes. Given the preponderance of evidence showing that patient and physician safety is negatively affected by long work hours, efforts to improve physician schedules should be prioritized. Policies that enable extended-duration shifts and long work weeks should be reexamined. Further research should expand beyond resident physicians to additional study populations, including attending physicians and other health care workers.

Approximately 21 million people in the United States are employed in the health care and social assistance sector, and more than 5 million work in shifts to provide 24/7 care, support, and oversight.<sup>1,2</sup> Extended ( $\geq 24$  hours), overnight, and rotating shifts, along with long weekly work hours, are particularly common in health care, exposing health care workers to circadian misalignment, extended wakefulness, and sleep deficiency, all known to impair performance, health, and safety.<sup>3</sup>

Resident physicians are required to endure some of the most intensive of all work schedules. These include frequent extended-duration work shifts, which have been a hallmark of medical education training and the subject of controversy for decades.<sup>4</sup> Policy has vacillated over the years as new evidence has emerged. In 2003 the Accreditation Council for Graduate Medical Education (ACGME) implemented guidelines that limited resident physicians to 80 hours per week, but it continued to sanction up to nine extended-duration shifts (up to 30 consecutive hours) per month.<sup>5</sup> Following a year-long review by the Institute of Medicine in 2008 that found resident physician extended-duration work shifts posed hazards to health and safety,<sup>6</sup> the

ACGME issued new work hour guidelines in 2011, limiting postgraduate year 1 resident physicians (PGY1s; interns) to work shifts of 16 consecutive hours or less. Work shifts of up to 28 consecutive hours (24 hours of work plus 4 hours for transitions of care and education) continued to be allowed for more senior residents.<sup>7</sup> In 2017, after publication of two clinical trials showing that flexible work hours that included extended-duration shifts were associated with noninferior hospital-level patient safety outcomes, the ACGME reversed course and lifted the 16-hour restriction on first-year resident work hours.<sup>8</sup>

Recent policy changes that relaxed work hour limits were informed by a subset of studies relevant to this policy issue. However, the body of knowledge that informs the design of safe and healthy work schedules is much larger than these few influential studies and multifaceted. Studies regarding the effects of resident physician work hours on patient safety, as well as resident safety, have evaluated a range of metrics and have come to different conclusions. Although randomized clinical trials provide the highest-quality evidence, to better quantify and synthesize all the literature regarding the role of work schedule characteristics on patient safety, we conducted a systematic review followed by a series of meta-analyses to pool similar studies. Our primary research question was, "Are resident physician work hour policies or physician work schedule characteristics associated with adverse effects on patient or physician safety?" We hypothesized that policies and schedules that limited work

**Table 1. US Preventive Services Task Force Quality Ratings.**

Level	Description
I	Evidence obtained from at least one properly designed randomized controlled trial
II-1	Evidence obtained from well-designed controlled trials without randomization
II-2	Evidence obtained from well-designed cohort or case-control analytic studies, preferably from more than one center or research group
II-3	Evidence obtained from multiple time series with or without the intervention; dramatic results in uncontrolled trials might also be regarded as this type of evidence.
III	Opinions of respected authorities, based on clinical experience, descriptive studies, or reports of expert committees

hours would either improve or have no negative effect on patient or physician safety.

## METHODS

Potentially relevant studies were identified by searching the PubMed and Embase databases. Inclusion criteria were as follows: English language (published prior to June 26, 2019); study population included physicians; outcomes studied included one or more patient safety outcome (for example, medical error, preventable adverse event, complications, mortality) or one or more physician safety outcome (for example, motor vehicle crash, occupational or percutaneous injury); and individual-level scheduling variables for resident physicians were collected (for example, shift length, weekly work hours). Our specific search criteria are listed in Appendix 1 (available in online article).

One investigator independently screened titles and abstracts of search results and reviewed references to identify additional potentially relevant studies through citation chaining. Two investigators then worked independently to review the full-text articles of all retained studies and extracted key information. If there was any ambiguity in whether a study was appropriate, it was discussed in a group of at least three investigators until there was unanimous agreement.

Observational studies, intervention studies (including randomized clinical trials), and other systematic reviews were retained. We excluded nonoriginal research (such as nonsystematic review articles) and non-peer-reviewed papers (for example, abstracts, editorials). We excluded staffing studies; that is, studies that focused on types of physician coverage or unit staffing rather than on the schedule of individual physicians. We considered studies evaluating the weekend effect (effects of weekends/holidays vs. midweek admissions, and admissions after hours vs. daytime admissions) to be staffing exclusions. Although reaction time and alertness are associated with performance and patient safety, we excluded studies that used neurobehavioral performance tests that did not directly simulate a medical procedure or driving task. As a result, we included studies that had clinically relevant outcomes, such as surgical simulator studies, but excluded studies that used the psychomotor vigilance task to assess alertness. Similarly, we excluded evaluations of perceptions, feelings, or predictions

about patient care without direct collection of patient care outcomes. Studies without any quantitative analyses were excluded, as were reviews and opinion pieces. Finally, we excluded evaluations of short-term work intensity, drug trials of wake-promoting and sleep-promoting pharmaceuticals, and studies that assessed modeling of schedules. Some studies were excluded for more than one reason.

## Quality of Evidence

Two investigators assessed the quality of each study, including risk of bias, using the US Preventive Services Task Force quality ratings, which provide a standardized metric of study design and implementation.<sup>9</sup> The size of the cohorts, the number of centers involved in the research, long-term outcomes, and statistical tests and covariates used in statistical adjustments were considered when assigning the rating. All studies were classified as Level I, II-1, II-2, II-3, or III, and any discrepancies were resolved through discussion. Table 1 provides description of each classification level. For randomized clinical trials, we used the Cochrane Collaboration's Risk of Bias tool to assess biases and limitations associated with experimental study designs.<sup>10</sup> Each item was graded as + (low risk of bias), – (high risk of bias), or ? (unclear risk of bias). The assessment of the overall study risk of bias was graded as + (low risk of bias) when the majority of items ( $\geq 4$ ) were rated as low risk of bias ratings or as – (high risk of bias) when  $\leq 3$  items were rated as low risk of bias. We used the GRADE tool for evaluating biases and limitations associated with observational studies.<sup>11</sup> In our discussion of the observational studies, we evaluated the four domains specified in the GRADE guidelines (participant selection, measurement of exposure and outcome, control for confounding, and completeness of follow-up).

## Meta-Analysis

Multiple studies had similar outcome measures that could be pooled for meta-analysis. For a study to be included in a meta-analysis, the outcome and study design needed to be sufficiently similar and the effect estimate needed to be reported in a way that enabled translation (if necessary) to a risk ratio (RR). Information from each study was extracted to calculate the RR and 95% confidence intervals (CIs). We used the generic inverse variance and implemented fixed effects methods in RevMan 5.3 (Copenhagen, The Nordic Cochrane Centre, The Cochrane Col-

laboration, 2014) to calculate the weighted average effect of the 2003 and 2011 policy changes on mortality and the weighted average effect of shorter shift durations on mortality in clinical trials. Data from clinical trials of varying shift durations that examined patient safety outcomes were pooled using random effects in the meta-analysis to account for the varied distribution of true effect sizes that likely follow the varied outcomes and study populations. The Z statistic was used to test the significance of the overall effect. The I<sup>2</sup> statistic was evaluated for evidence of heterogeneity. Sensitivity analyses implemented random effects rather than fixed effects models and vice versa—the conclusions were the same for all comparisons. Funnel plots were examined for evidence of publication bias. Forest plots were generated to illustrate the individual point estimates and the weighted average effect. Studies applying noninferiority approaches and/or a CI other than 95% CI were recalculated to 95% CI in RevMan and the associated figures.

## RESULTS

The search strategy yielded 7,963 studies, and 399 additional studies were found through reviewing the references listed in 24 systematic literature reviews identified in the search process. After removal of irrelevant studies and duplicates, 688 full-text articles were reviewed. An additional 556 articles were excluded, leaving 132 studies that met inclusion criteria for review. Ultimately, 68 studies provided enough information and were retained for consideration in a meta-analysis<sup>12–79</sup> (Figure 1). Most studies excluded from the meta-analyses either were not sufficiently similar in design or outcome to be pooled or did not include sufficient precision in outcome reporting to facilitate calculations (for example, counts, rates, RRs).

Study design, key results, and quality ratings of each of the 68 included studies are presented in Appendix 2.

### Quality Ratings

Four randomized controlled trials were rated at the highest level (I) and 2 nonrandomized trials were rated as II-1. The majority of the studies were rated at Level II-2 ( $n = 46$ ), indicating well-designed cohort or case-control analytic studies, generally with more than 1,000 subjects or more than one site. Twelve studies were rated as II-3, and 4 studies were purely descriptive (III). The risk of bias for the 4 randomized controlled trials is shown in Appendix 3. The overall risk of bias for randomized controlled trials was considered low for 2 trials and high for 2 trials. The quality of the nonrandomized trials and observational studies was generally rated as low. Common reasons for low quality ratings included lack of prospective data collection, limited generalizability, issues with exposure and/or outcome ascertainment, and limited adjustment for potentially confounding factors.

We completed four meta-analyses. Studies with multiple comparison groups and/or multiple outcomes were entered in the meta-analysis for each relevant outcome or comparator. To accommodate meta-analyses in the four randomized clinical trials, patient safety and mortality were conducted separately. Outcomes and methodologies in the simulator studies were too disparate for inclusion in a meta-analysis. We instead evaluated the direction of effect for each study. Because so few studies examined physician safety, we were unable to perform some of the preplanned meta-analyses.

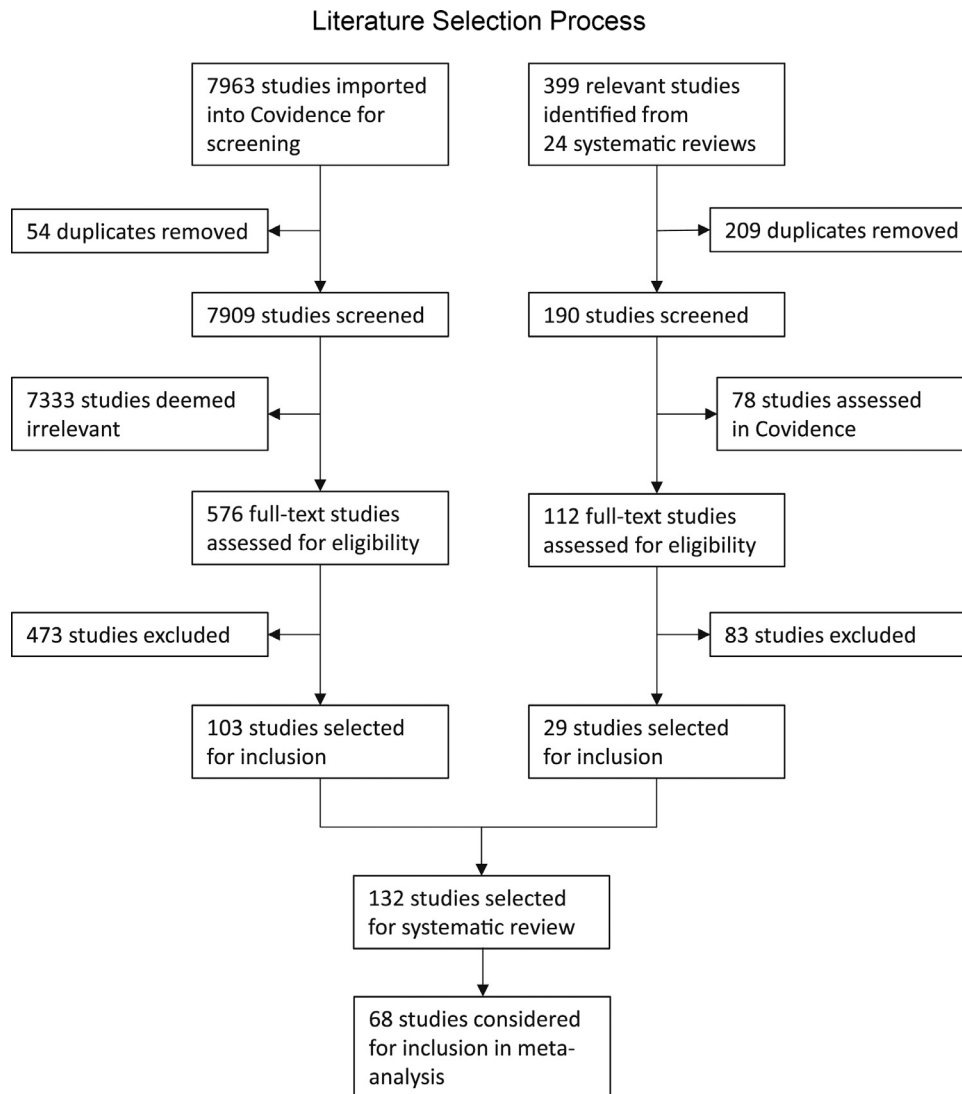
### Before/After Study Designs

The vast majority of studies with patient safety outcomes focused on the work hours of resident physicians. Further, most studies specifically examined changes in outcomes following the 2003 and 2011 changes in ACGME work hour policies. We conducted two meta-analyses of these before/after study designs: mortality before/after the 2003 ACGME policy change and mortality before/after the 2011 ACGME policy change.

### Before/After the 2003 ACGME Guidelines

Many studies evaluated patient safety before and after the ACGME work hour guidelines that were implemented in July 2003 using large databases such as the Nationwide Inpatient Sample (NIS), Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) Neonatal Research Network, and Improved Methods of Patient Information Access of Core Clinical Tasks (IMPACT). The breadth of the studies was wide. For example, the NIS was used to study patient safety in relationship to diagnosis related to stroke,<sup>12</sup> brain tumor and cerebrovascular disease,<sup>16</sup> neurosurgery,<sup>35,49</sup> otolaryngology,<sup>27</sup> thyroid or parathyroid procedures,<sup>66</sup> a wide range of post-operative complications.<sup>31</sup> Many of the studies used nonteaching hospitals or services as comparison groups,<sup>12,25,27,35,36,53,58,64,66,72–75</sup> (as do other studies in the literature<sup>80–82</sup>). Most of the 68 studies evaluated medical records for several years before and after the ACGME work hour changes. The length of the studies ranged up to nine years. All provided metrics of morbidity and mortality.

Thirty-three comparisons were included in the meta-analysis comparing mortality before and after the 2003 ACGME rule changes (Figure 2). The pooled effect across studies was a highly significant 11% reduction in mortality risk (weighted RR 0.89, 95% CI 0.88–0.90,  $p < 0.001$ ) after implementation of the new work rules. Only 1 study showed patient safety to be worse following the 2003 ACGME work hour reductions.<sup>35</sup> Heterogeneity of outcomes precluded pooling the 39 available comparisons estimating the risk of morbidity following the 2003 ACGME work hour rules (Figure 3). Eleven of the 39 found that the work hour rules had a protective effect, 9 reported harm, and the remainder found no difference.



**Figure 1:** This flow chart illustrates the literature selection process.

### Before/After the 2011 ACGME Work Rules

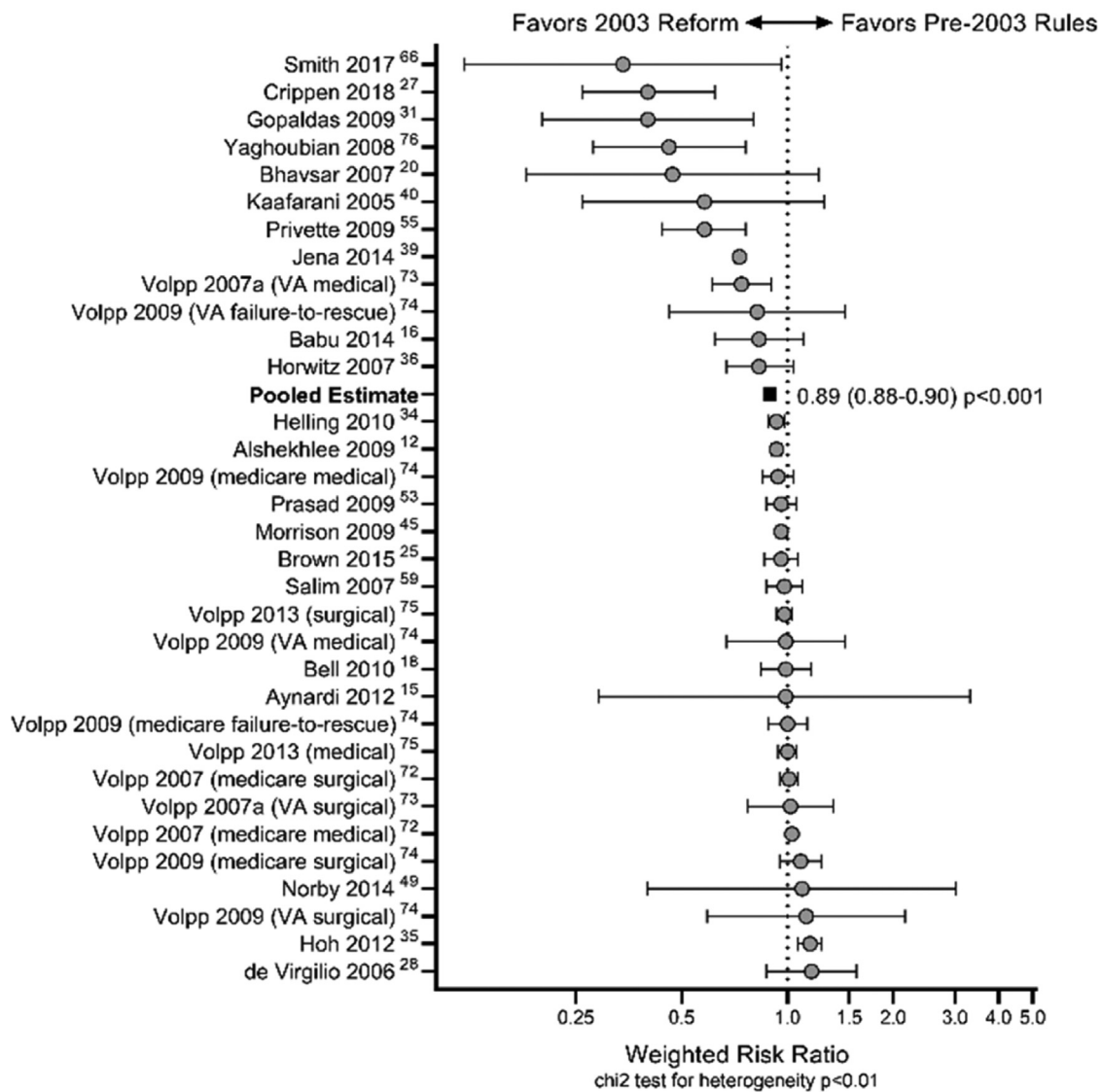
Before/after study designs were also used to study the 2011 ACGME work hour changes, but far fewer studies have been published on this change. Fifteen comparisons were included in a meta-analysis of patient mortality (Figure 4). None showed a significant difference in patient mortality before and after the 2011 ACGME work hour rule changes (weighted RR 1.00, 95% CI 0.97–1.02,  $p = 0.84$ ). Seven studies conducted 11 eligible comparisons of patient morbidity pre- vs. post-2011 ACGME work hour changes. Heterogeneity of the outcomes prevented pooling estimates for meta-analysis. Two studies found that the 2011 ACGME work rule changes were associated with reductions in patient morbidity, while the remainder found no difference (Figure 5).

### Randomized Clinical Trials of Shift Length

Many studies investigated shift length, using both clinical and observational study designs. Four randomized clinical

trials were considered for meta-analyses comparing longer and shorter shift lengths. Landrigan and colleagues compared shifts of 24 hours or more to shifts of no more than 16 scheduled consecutive hours,<sup>43,44</sup> while Parshuram and colleagues added a 12-hour condition.<sup>50</sup> Silber and colleagues and Bilimoria and colleagues compared the ACGME 2011 work hours (first-year resident physicians restricted to shifts of 16 or fewer consecutive hours) to a flexible schedule of unrestricted continuous hours (Figure 6).<sup>21,65</sup> Two of the four clinical trials examined mortality as an outcome.<sup>21,65</sup> Pooling data from the two trials, the weighted RR for mortality is estimated to be 0.97 (95% CI 0.95–1.00,  $p = 0.06$ ), with a protective point estimate for work hour restrictions that did not meet our criteria for statistical significance (Appendix 4, Supplemental Figure 3). After estimates from the three clinical trials that examined medical errors or patient safety indicators were pooled,<sup>43,50,65</sup> shorter shift lengths were associated with improved patient safety (weighted RR 0.74, 95% CI 0.60–0.91,  $p = 0.01$ ) (Appendix 4, Supplemental Figure 4). None of the trials found that shorter shift

## Risk Ratios, Patient Mortality Before and After 2003 ACGME Work Rules Introduced



**Figure 2:** Shown here are the weighted risk ratios of patient mortality before and after the 2003 Accreditation Council for Graduate Medical Education (ACGME) work rules were introduced. VA, US Department of Veterans Affairs.

lengths were associated with harm. McCoy et al. conducted a nonrandomized trial of 24-hour shifts (plus 6 hours of call) to an alternative model that limited shifts to no more than 16 consecutive hours.<sup>44</sup> This study was not pooled with the other trials due to the nonrandomized design. There was no difference between the groups for patient safety outcomes.

### Observational Studies of Shift Length

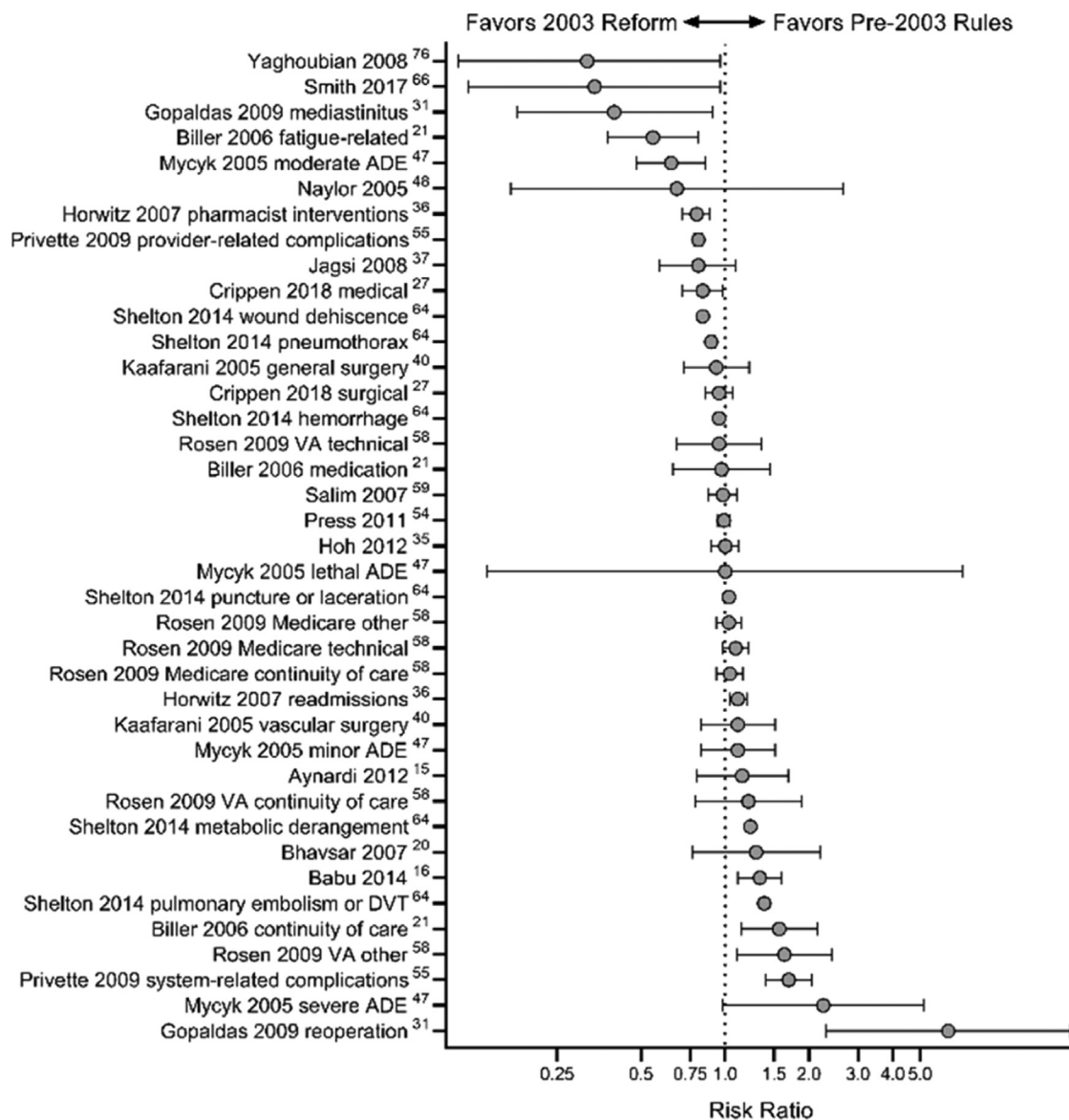
Six observational studies conducted eight comparisons of longer shifts, including exceeding 10,<sup>42</sup> 12,<sup>63</sup> 16,<sup>77,78</sup> and 24 hours or more,<sup>17</sup> whereas Anderson and colleagues compared a 16-hour limit to a flexible schedule of unrestricted continuous hours.<sup>13</sup> Two studies (three distinct comparisons) found that longer shifts were associated with harm.

No difference was found for another two studies reporting on three distinct comparisons. The remaining two studies found that longer shifts were associated with lower rates of patient morbidity (Figure 7).

### Before/After Overnight Shifts

Sixteen studies reported 17 total comparisons comparing performance on simulated tasks following a night of work vs. a night of not working. Eleven of the studies showed worse performance post-work or following an on-call shift, and 5 studies' results were neutral (Figure 8). The 1 study that showed better performance following work or an on-call shift did not use a randomized order, instead comparing the tasks sequentially, allowing for the possibility of a learning effect to explain the results.<sup>61</sup>

Risk Ratios of Patient Morbidity Before and After 2003 ACGME Work Rules Introduced



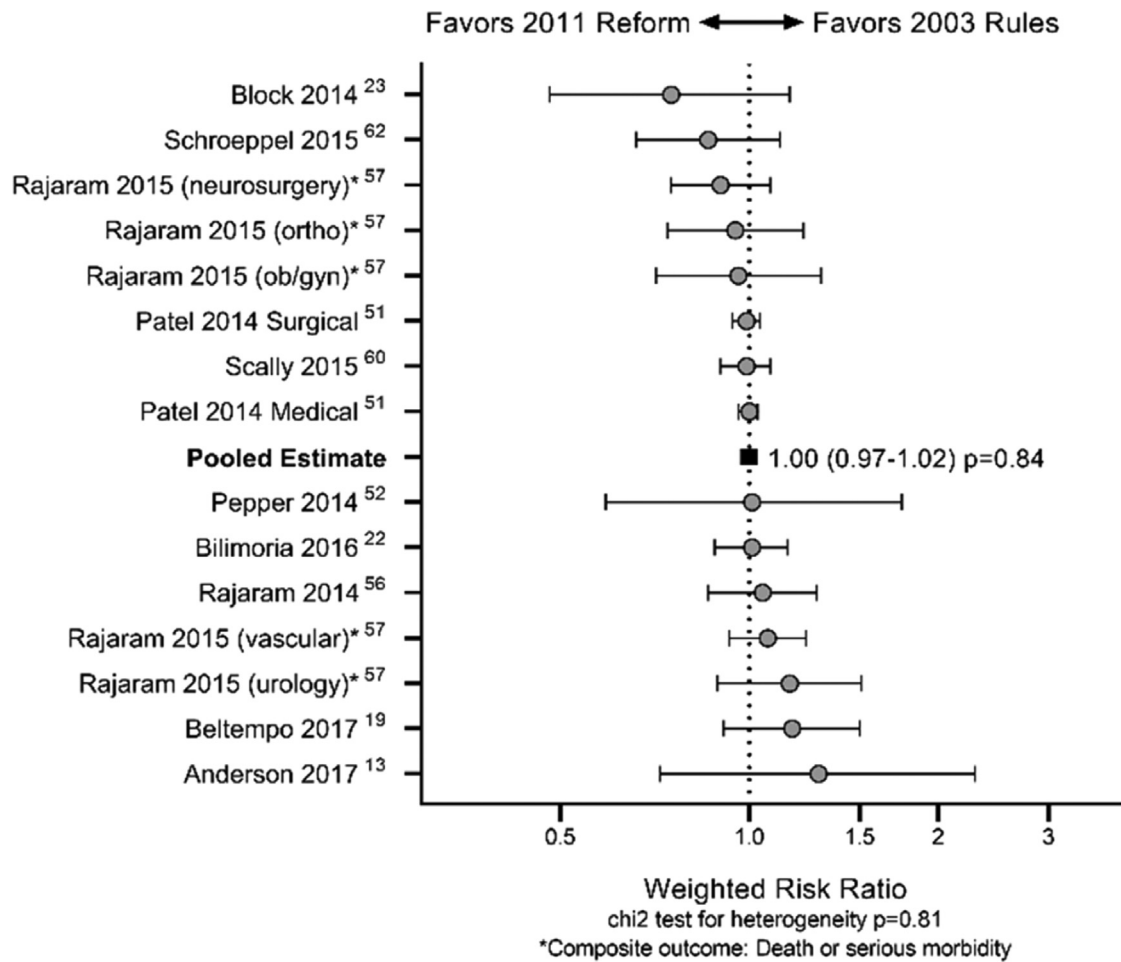
**Figure 3:** Shown here are the weighted risk ratios of patient morbidity before and after the 2003 Accreditation Council for Graduate Medical Education (ACGME) work rules were introduced. ADE, adverse drug event; VA, US Department of Veterans Affairs; DVT, deep vein thrombosis.

**DISCUSSION**

The overall body of evidence suggests that scheduling resident physicians to work hours that are less than current regulations permit is associated with lower mortality and improved patient safety. Randomized clinical trials comparing shifts of 16 or fewer consecutive hours to those of 24 or more hours estimate a 3% reduction in mortality ( $p = 0.06$ ) and a 26% reduction in patient safety ( $p = 0.01$ ) with shorter shifts. Limited shift durations and shorter work weeks in observational and quasi-experimental design studies also suggest improved patient safety. Although these nonrandomized designs represent a lower quality of

evidence that is insufficient to demonstrate causality, and many studies were inconclusive, the majority suggest that shorter shifts were associated with patient safety benefit or that work hour regulations were associated with improvements in patient safety. Observational studies comparing shifts of 16 or fewer consecutive hours to shifts of 24 or more hours more commonly reported patient harm with longer shifts. In simulation studies using a before/after design, 94.0% showed neutral or worsened performance of resident physicians following a work or on-call shift. These results are consistent with meta-analyses from other industries showing increased risk for injuries and accidents with

## Risk Ratios of Patient Mortality Before and After 2011 ACGME Work Rules Introduced



**Figure 4:** Shown here are the weighted risk ratios of patient mortality before and after 2011 Accreditation Council for Graduate Medical Education (ACGME) work rules were introduced.

increasing shift duration.<sup>83,84</sup> Smaller previous reviews of resident work hours have not been definitive.<sup>85-87</sup>

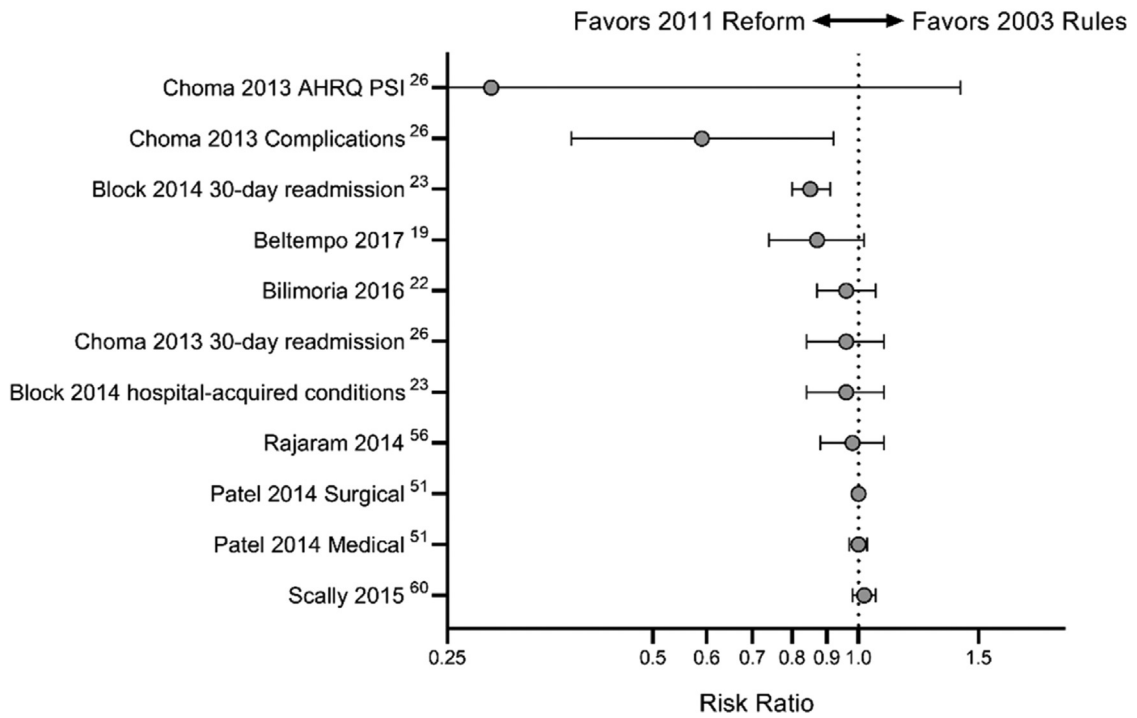
Meta-analyses of studies using a before/after design suggest that patient safety improved following implementation of the ACGME 2003 and 2011 work hour guidelines for resident physicians. Limiting all resident physicians to 80-hour work weeks and 28-hour shifts in 2003 was associated with an 11% reduction in mortality. Changes in hospital-wide patient safety outcomes were less likely following the 2011 ACGME reforms, as they restricted the work hours of only first-year resident physicians (~25% of all resident physicians).

Not all studies reviewed reported sufficient details of work schedules such as the shift length, frequency, and timing. These are all critical independent predictors of performance, and they may interact in complex ways. Multiple physiological variables that are known to influence performance are tied to scheduling factors—time awake is directly correlated with shift duration; nightly sleep duration is correlated with time off between shifts; and bio-

logical time of day is influenced by shift timing and the rate/direction of rotation. Sleep debt, or an accumulation of insufficient sleep, degrades performance<sup>88</sup> and was rarely assessed in the reviewed studies. Furthermore, chronic sleep deprivation potentiates the performance degradation resulting from acute sleep deprivation and circadian misalignment.<sup>89</sup> Future studies should be designed to measure and evaluate the interaction of each of these components of sleep and scheduling factors known to affect performance.

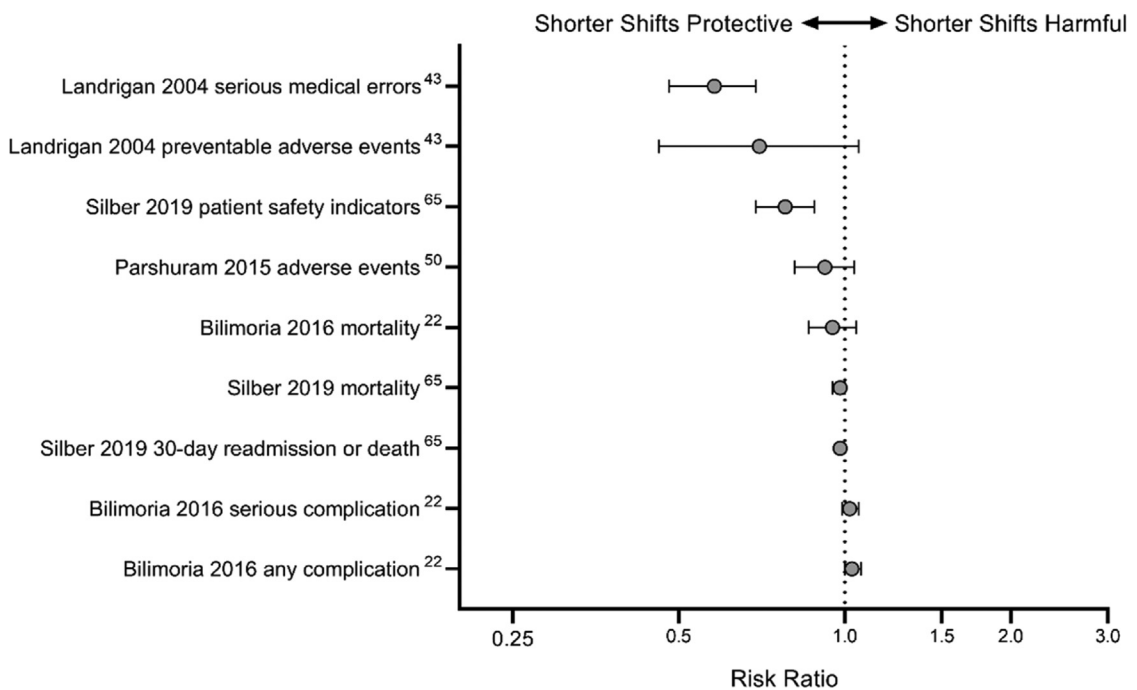
In all meta-analyses, it is important to recognize the quality of the studies included. Although they are subjective, US Preventive Services Task Force quality ratings provide a means of categorizing varied unique studies into five fixed categories. Only four studies met the gold standard of a randomized clinical trial, and none were double-blinded, as there was no way for subjects to be unaware of their work schedules. Even in laboratory studies using clinical simulators, double blinding is not possible because physician participants know their schedule (for example, whether they worked the night prior to simulation). The

Risk Ratios, Patient Morbidity Before and After 2011 ACGME Work Rules Introduced



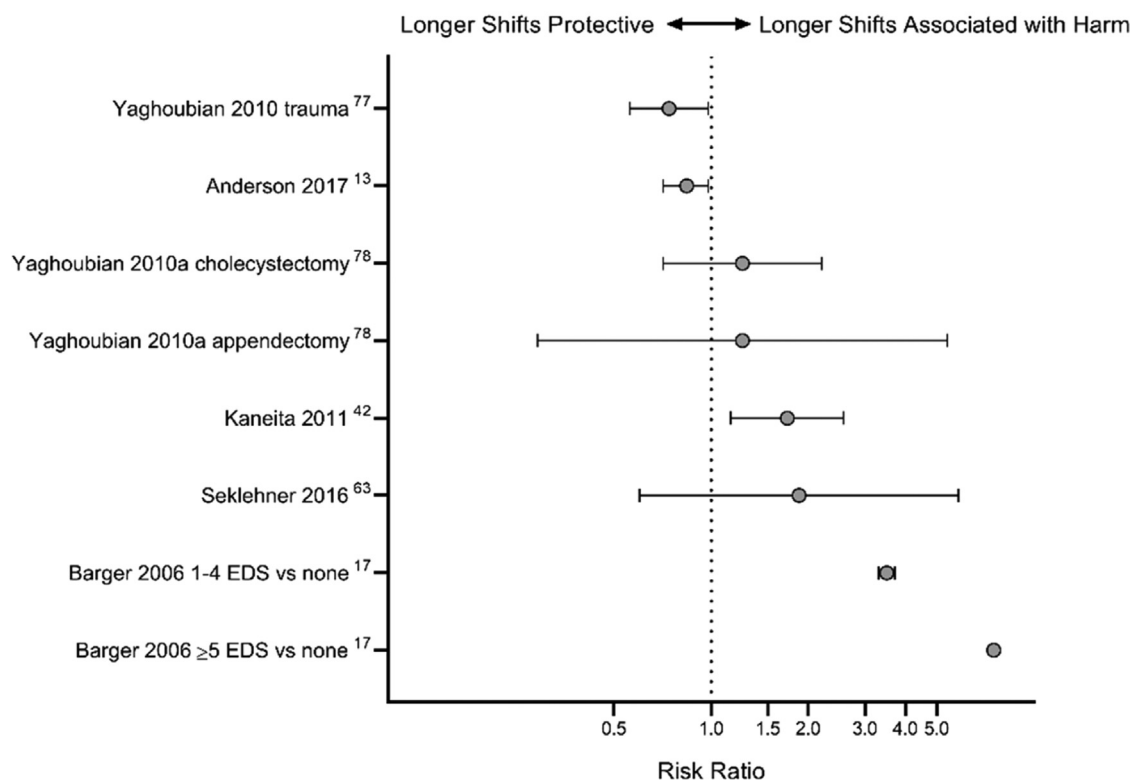
**Figure 5:** Shown here are the risk ratios of patient morbidity before and after the 2011 Accreditation Council for Graduate Medical Education (ACGME) work rules were introduced. AHRQ, Agency for Healthcare Research and Quality; PSI, Patient Safety Indicator.

Mortality and/or Morbidity of Four Randomized Clinical Trials



**Figure 6:** Shown here are risk ratios for mortality and/or morbidity of four randomized clinical trials of shorter and longer work shifts of resident physicians.

## Morbidity Associated with Resident Physicians' Shifts



**Figure 7:** Shown here are the risk ratios of morbidity associated with resident physicians working shorter and longer shifts in six observational studies.

quality of evidence in this project overall, with 67.6% of evidence obtained from well-designed cohort or case-control analytic studies (Level II-2), is moderate. Further assessment of biases revealed low quality ratings for observational studies. However, in the meta-analysis limited to randomized clinical trials, 50.0% rated low risk of bias and also demonstrated that patient safety outcomes were superior for shorter shifts compared to longer shifts, strengthening this finding.

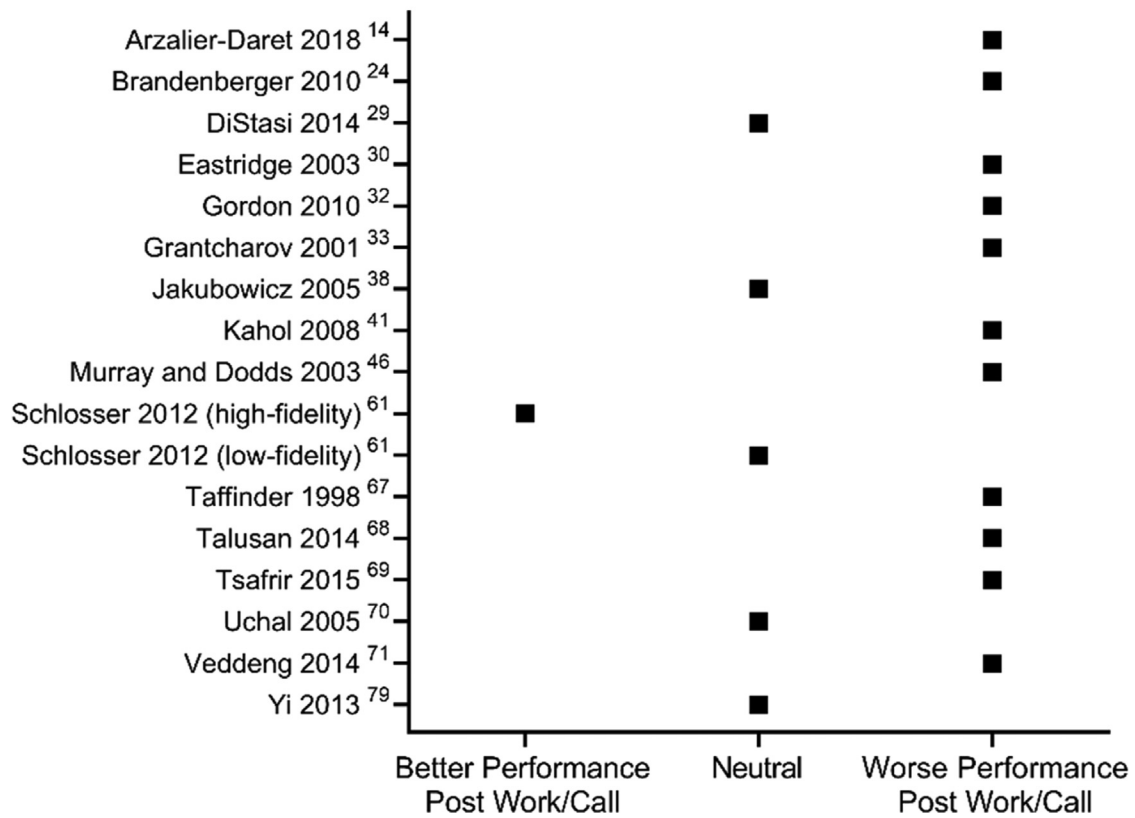
### Limitations

When evaluating the results, biases and other factors should be considered in addition to the formal assessment of study quality. For example, although the FIRST<sup>21</sup> and iCOMPARE<sup>65</sup> trials were given a quality rating of I because they were randomized controlled trials, these reports had other limitations, such as outcomes (hospital-level mortality) that were very distal from the intervention, that are not indicated by this rating system. Many other studies also evaluated hospitalwide mortality following the 2011 ACGME work hour changes<sup>19,23,26,51,52,56,57,60,62</sup> when only first-year resident physicians were bound by the new shift limits introduced at that time. Similarly, studies using nonresident services or nonteaching or less teaching-intensive hospitals as controls to look at the effect of a scheduling intervention at a systemic level have substantial limitations, as changes in

resident work schedules represent only a small part of very complex systems. Indeed, choosing an appropriate metric to measure the impact of a change in schedule in a way that accurately and comprehensively reflects patient care is difficult.

In addition, the high quality rating given to the randomized controlled trials included in this review did not reflect the controversial choice of statistical analysis in those trials. Noninferiority clinical trials seek to determine if the experimental arm is not worse than the control arm by an “acceptably small amount, with a given degree of confidence.”<sup>90</sup>(p. 1357) The prespecified noninferiority margin of 1 percentage point in 30-day mortality used in the iCOMPARE trial conservatively translates to a minimum of 39,270 additional deaths per year in the United States.<sup>91,92</sup> Whether nearly 40,000 deaths is an “acceptably small” or a strikingly high number of deaths is debatable, as some of our most critical public health interventions have smaller effects. For example, the National Highway Traffic Safety Administration estimates seat belts saved 15,000 lives in 2016.<sup>93</sup> Therefore, the conclusions of these studies need to be considered carefully. Meta-analyses executed in this review enabled the pooling of data from cohort studies and all clinical trials, including those that originally reported their results using noninferiority tests. A recalculation of outcomes from the iCOMPARE trial, originally reported

Physician Performance on Clinical Simulators



**Figure 8:** Shown here is a comparison of resident physician performance on clinical simulators before and after an on-call shift.

as noninferior with respect to mortality, 30-day readmission or death, and patient safety indicators, shows that programs randomized to maintain the 16-hour ACGME limit actually had significantly fewer patient safety events compared to programs randomized to more flexible work hours that allowed for extended-duration shifts (using two-sided tests with  $\alpha = 0.05$ ). Examining the clinical trials as a whole, some trials were neutral and others showed a protective effect with assignment to shorter shift durations. None indicated harm to morbidity or mortality with shift restrictions, and after pooling data from all clinical trials, shorter shifts were associated with a 26% reduction in the risk of medical errors or adverse events.

The many before/after studies included in our review also have limitations. For example, changes in Medicare reimbursement in 2010 may have confounded the finding that 30-day readmission rates fell in some studies.<sup>23,51</sup> Any changes in standard clinical practices (at a site or systemic level, depending on the study) that coincided with the implementation of new resident work hour policies could also confound the before/after study designs. Further, not all study designs adequately accounted for circadian variations on alertness and performance and thus could present misleading results. For example, some studies compared per-

formance measured near the end of a day shift (evening) to performance in the morning following a night shift<sup>70,94</sup> or before and after a 12-hour shift.<sup>79</sup> Any differences or lack of differences in performance may not be completely ascribed to the schedule, as there are circadian differences in performance at different times of day.<sup>95</sup>

Although studies were similar enough to use meta-analytical techniques to evaluate patient safety, heterogeneity of exposure variables (such as length of shift), follow-up intervals, and outcome variables (for example, medical errors, adverse events) limits interpretation and generalization. Other important outcomes, such as quality of residency training, medical knowledge, and the skills of resident physicians following their training were not addressed in this study. The exclusion of non-English publications further limited the study. Moreover, because gray literature was not included in this project, publication bias cannot be discounted. However, funnel plots for each of the meta-analyses do not indicate evidence of publication bias. Meta-analysis procedures assign weights to each individual study based on the sample size and the precision of effect estimates. In situations in which very large studies are pooled with other small studies, the effect estimates from the larger studies contribute proportionally

more to the pooled RR estimates. We provide output in Appendix 4, Supplemental Figures 1 to 4, to review the weight assigned to each study when interpreting the results.

Several factors in addition to schedules and work hours that affect patient safety were not evaluated in this literature review. Physician workload is one factor that requires additional attention. Hanna and colleagues found that the number of discrepancies in radiologic examinations was increased when study volume increased.<sup>96</sup> A more recent multicenter clinical trial, published outside of the timeline of this review, also suggested that increasing workload is associated with medical errors and can confound analyses of work schedules.<sup>97</sup> Handoffs between work shifts and the supervision of resident physicians can also have an important impact on patient safety. The interaction of all these factors with work schedules requires additional research.

Additional research into the schedules of attending physicians is greatly needed. Given that multiple work hour policy changes have been instituted for resident physicians, particularly those in their first year, over the past two decades, most of the studies conducted have involved resident physicians. The schedules of attending physicians in the United States are not regulated, and little information is available about them or how they might have changed due to changes in resident work hours. The few studies that have evaluated attending physician schedules<sup>98,99</sup> showed longer hours were associated with higher risks to patient safety, similar to the results of this meta-analysis.

## CONCLUSION

The collective literature suggests that resident physician work schedules have an important impact on patient safety. Given that the ACGME work hour reductions of 2003 and 2011 have been associated with significant improvements in physician safety and health,<sup>100</sup> as well as subtler but still important improvements in patient safety, efforts should continue to safely do away with resident physicians' traditional long work weeks and extended shifts. For those programs seeking to perpetuate them, the burden of proof should be to show that longer hours offer benefits greater than their demonstrated risks in their particular settings. Regulators and researchers should view any such assertions critically, given the large volume of research now demonstrating the hazards of long work hours for resident physicians.

Long work weeks and extended shifts are hazardous, but not all interventions to reduce them are equivalent. The literature demonstrates that the manner in which schedules or policies are implemented can have an important bearing on their effectiveness. Future policy efforts aimed at improving patient safety must be crafted in a sufficiently specific manner to effectively achieve their intended aims. Sufficient re-

sources must be provided to ensure that efforts to improve work hour policy do not result in a worsening of workload or other structural factors in hospitals. Finally, future research should assess the associations among scheduling characteristics and physician sleep, quality of life, long-term health, well-being, and burnout. Further research should expand beyond resident physicians to additional study populations, including attending physicians and other health care workers.

**Funding.** Funding for this project was provided by Liberty Mutual Research Institute for Safety. Drs. Weaver and Barger and Mr. Sullivan were supported in part by the National Institute for Occupational Safety and Health (NIOSH R01 OH011773-01).

**Acknowledgments.** The authors would like to thank Dr. George Brogmus and Dr. Jesse Jacobs for their review of this work.

**Conflict of Interest.** Dr. Weaver reports institutional support from the US Centers for Disease Control and Prevention, National Institutes of Occupational Safety and Health, National Heart, Lung and Blood Institute, Delta Airlines, and the Puget Sound Pilots during the conduct of the study; as well as consulting fees from the Fred Hutchinson Cancer Center and the University of Pittsburgh for assisting with research projects on sleep and circadian rhythms. Mr. Sullivan has nothing to declare. Dr. Landrigan has consulted with and holds equity in the I-PASS Institute, which seeks to train institutions in best handoff practices and aid in their implementation. He has consulted with the Missouri Hospital Association / Executive Speaker's Bureau regarding I-PASS. In addition, Dr. Landrigan has received monetary awards, honoraria, and travel reimbursement from multiple academic and professional organizations for teaching and consulting on sleep deprivation, physician performance, handoffs, and safety, and has served as an expert witness in cases regarding patient safety and sleep deprivation. Dr. Barger reports institutional support from the US Centers for Disease Control and Prevention, National Institutes of Occupational Safety and Health, National Institute of Aging, Delta Airlines, and the Puget Sound Pilots during the conduct of the study; as well as consulting fees from the Boston Children's Hospital and honorariums from University of Helsinki, the AAA Foundation, the University of Arizona and the University of British Columbia.

## SUPPLEMENTARY MATERIALS

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.jcjq.2023.06.014](https://doi.org/10.1016/j.jcjq.2023.06.014).

**Matthew D. Weaver, PhD, MPH**, is Associate Epidemiologist, Division of Sleep and Circadian Disorders, Departments of Medicine and Neurology, Brigham and Women's Hospital, Boston, Instructor in Medicine and Division of Sleep Medicine, Harvard Medical School, Boston. **Jason P. Sullivan** is Senior Project Manager, Division of Sleep and Circadian Disorders, Departments of Medicine and Neurology, Brigham and Women's Hospital. **Christopher P. Landrigan, MD, MPH**, is Chief, Division of General Pediatrics, Boston Children's Hospital, and Director, Sleep and Patient Safety Program, Division of Sleep and Circadian Disorders, Departments of Medicine and Neurology, Brigham and Women's Hospital. **Laura K. Barger, PhD**, is Associate Physiologist, Division of Sleep and Circadian Disorders, Departments of Medicine and Neurology, Brigham and Women's Hospital, and Assistant Professor, Division of Sleep Medicine, Harvard Medical School. Please address correspondence to Laura K. Barger, [lbarger@hms.harvard.edu](mailto:lbarger@hms.harvard.edu).

## REFERENCES

1. US Bureau of Labor Statistics. Industries at a glance: Healthcare and Social Assistance: NAICS 62. (Updated: Jul 14, 2023.) Accessed Jul 16, 2023. <https://www.bls.gov/iag/tgs/iag62.htm>.

2. Alterman T, et al. Prevalence rates of work organization characteristics among workers in the U.S.: data from the 2010 National Health Interview Survey. *Am J Ind Med.* 2013;56:647–659.
3. Barger LK, et al. Neurobehavioral, health, and safety consequences associated with shift work in safety-sensitive professions. *Curr Neurol Neurosci Rep.* 2009;9:155–164.
4. Steinbrook R. The debate over residents' work hours. *N Engl J Med.* 2002 Oct 17;347:1296–1302.
5. Accreditation Council for Graduate Medical Education (ACGME). Common Program Requirements—Duty Hour Requirements. Chicago: ACGME, 2003.
6. Institute of Medicine. Resident Duty Hours: Enhancing Sleep, Supervision, and Safety, Washington, DC: National Academies Press, 2008. Accessed Jul 16, 2023 <https://www.ncbi.nlm.nih.gov/books/NBK214948/>.
7. Accreditation Council for Graduate Medical Education (ACGME). Common Program Requirements. Chicago: ACGME, 2011.
8. Accreditation Council for Graduate Medical Education (ACGME). Common Program Requirements. Chicago: ACGME, 2017.
9. Harris RP, et al. Current methods of the US Preventive Services Task Force: a review of the process. *Am J Prev Med.* 2001;20(3 Suppl):21–35.
10. Cochrane Collaboration Cochrane Handbook for Systematic Reviews of Interventions, version 5.1.0, 2023. (Updated: Mar 2011.) Accessed Jul 16 <https://handbook-5-1.cochrane.org/>.
11. Guyatt GH, et al. GRADE guidelines: 4. Rating the quality of evidence—study limitations (risk of bias). *J Clin Epidemiol.* 2011;64:407–415.
12. Alshekhlee A, et al. The impact of Accreditation Council for Graduate Medical Education duty hours, the July phenomenon, and hospital teaching status on stroke outcomes. *J Stroke Cerebrovasc Dis.* 2009;18:232–238.
13. Anderson JE, et al. Restrictions on surgical resident shift length does not impact type of medical errors. *J Surg Res.* 2017 May 15;212:8–14.
14. Arzallier-Daret S, et al. Effect of sleep deprivation after a night shift duty on simulated crisis management by residents in anaesthesia. A randomised crossover study. *Anaesth Crit Care Pain Med.* 2018;37:161–166.
15. Aynardi M, et al. Effect of work-hour restrictions and resident turnover in orthopedic trauma. *Orthopedics.* 2012;35:e1649–e1654.
16. Babu R, et al. Worse outcomes for patients undergoing brain tumor and cerebrovascular procedures following the ACGME resident duty-hour restrictions. 2014;121:262–276.
17. Barger LK, et al. Impact of extended-duration shifts on medical errors, adverse events, and attentional failures. *PLoS Med.* 2006;3:e487.
18. Bell EF, et al. Impact of timing of birth and resident duty-hour restrictions on outcomes for small preterm infants. *Pediatrics.* 2010;126:222–231.
19. Beltempo M, et al. Association of resident duty hour reform and neonatal outcomes of very preterm infants. *Am J Perinatol.* 2017;34:1396–1404.
20. Bhavsar J, et al. Impact of duty hours restrictions on quality of care and clinical outcomes. *Am J Med.* 2007;120:968–974.
21. Biller CK, et al. The 80-hour work guidelines and resident survey perceptions of quality. *J Surg Res.* 2006;135:275–281.
22. Bilimoria KY, et al. National cluster-randomized trial of duty-hour flexibility in surgical training. *N Engl J Med.* 2016 Feb 25;374:713–727.
23. Block L, et al. Inpatient safety outcomes following the 2011 residency work-hour reform. *J Hosp Med.* 2014;9:347–352.
24. Brandenberger J, et al. Effects of duty hours and time of day on surgery resident proficiency. *Am J Surg.* 2010;200:814–819.
25. Brown SES, Ratcliffe SJ, Halpern SD. Assessing the utility of ICU readmissions as a quality metric: an analysis of changes mediated by residency work-hour reforms. *Chest.* 2015;147:626–636.
26. Choma NN, et al. Effect of the ACGME 16-hour rule on efficiency and quality of care: duty hours 2.0. *JAMA Intern Med.* 2013 May 13;173:819–821.
27. Crippen MM, et al. The impact of duty-hour restrictions on complication rates following major head and neck procedures. *Laryngoscope.* 2018;128:2804–2810.
28. de Virgilio C, et al. The 80-hour resident workweek does not adversely affect patient outcomes or resident education. *Curr Surg.* 2006;63:435–440.
29. Di Stasi LL, et al. Saccadic eye movement metrics reflect surgical residents' fatigue. *Ann Surg.* 2014;259:824–829.
30. Eastridge BJ, et al. Effect of sleep deprivation on the performance of simulated laparoscopic surgical skill. *Am J Surg.* 2003;186:169–174.
31. Gopaldas RR, et al. The impact of resident work-hour restrictions on outcomes of cardiac operations. *J Surg Res.* 2009;157:268–274.
32. Gordon JA, et al. 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;85:1583–1588.
33. Grantcharov TP, et al. Laparoscopic performance after one night on call in a surgical department: prospective study. *BMJ.* 2001 Nov 24;323:1222–1223.
34. Helling TS, et al. The effect of resident duty hour restriction on trauma center outcomes in teaching hospitals in the state of Pennsylvania. *J Trauma.* 2010;69:607–612.
35. Hoh BL, et al. Higher complications and no improvement in mortality in the ACGME resident duty-hour restriction era: An analysis of more than 107,000 neurosurgical trauma patients in the Nationwide Inpatient Sample database. *Neurosurgery.* 2012;70:1369–1381.
36. Horwitz LI, et al. Changes in outcomes for internal medicine inpatients after work-hour regulations. *Ann Intern Med.* 2007 Jul 17;147:97–103.
37. Jagi R, et al. The Accreditation Council for Graduate Medical Education's limits on residents' work hours and patient safety: a study of resident experiences and perceptions before and after hours reductions. *Arch Intern Med.* 2008 Mar 10;168:493–500.
38. Jakubowicz DM, et al. Effects of a twenty-four hour call period on resident performance during simulated endoscopic sinus surgery in an Accreditation Council for Graduate Medical Education-compliant training program. *Laryngoscope.* 2005;115:143–146.
39. Jena AB, Schoemaker L, Bhattacharya J. Exposing physicians to reduced residency work hours did not adversely affect patient outcomes after residency. *Health Aff (Millwood).* 2014;33:1832–1840.
40. Kaafarani HM, et al. Does resident hours reduction have an impact on surgical outcomes? *J Surg Res.* 2005 Jun 15;126:167–171.

41. Kahol K, et al. Effect of fatigue on psychomotor and cognitive skills. *Am J Surg.* 2008;195:195–204.
42. Kaneita Y, Ohida T. Association of current work and sleep situations with excessive daytime sleepiness and medical incidents among Japanese physicians. *J Clin Sleep Med.* 2011 Oct 15;7:512–522.
43. Landrigan CP, et al. Effect of reducing interns' work hours on serious medical errors in intensive care units. *N Engl J Med.* 2004 Oct 28;351:1838–1848.
44. McCoy CP, et al. Effect of 16-hour duty periods on patient care and resident education. *Mayo Clin Proc.* 2011;86:192–196.
45. Morrison CA, Wyatt MM, Carrick MM. Impact of the 80-hour work week on mortality and morbidity in trauma patients: an analysis of the National Trauma Data Bank. *J Surg Res.* 2009 Jun 1;154:157–162.
46. Murray D, Dodds C. The effect of sleep disruption on performance of anaesthetists—a pilot study. *Anaesthesia.* 2003;58:520–525.
47. Mycyk MB, et al. Hospitalwide adverse drug events before and after limiting weekly work hours of medical residents to 80. *Am J Health Syst Pharm.* 2005 Aug 1;62:1592–1595.
48. Naylor RA, Rege RV, Valentine RJ. Do resident duty hour restrictions reduce technical complications of emergency laparoscopic cholecystectomy? *J Am Coll Surg.* 2005;201:724–731.
49. Norby K, et al. The effect of duty hour regulations on outcomes of neurological surgery in training hospitals in the United States: duty hour regulations and patient outcomes. *J Neurosurg.* 2014;121:247–261.
50. Parshuram CS, et al. Patient safety, resident well-being and continuity of care with different resident duty schedules in the intensive care unit: a randomized trial. *CMAJ.* 2015 Mar 17;187:321–329.
51. Patel MS, et al. Association of the 2011 ACGME resident duty hour reforms with mortality and readmissions among hospitalized Medicare patients. *JAMA.* 2014 Dec 10;312:2364–2373.
52. Pepper DJ, Schweinfurth M, Herrin VE. The effect of new duty hours on resident academic performance and adult resuscitation outcomes. *Am J Med.* 2014;127:337–342.
53. Prasad M, et al. Effect of work-hours regulations on intensive care unit mortality in United States teaching hospitals. *Crit Care Med.* 2009;37:2564–2569.
54. Press MJ, et al. The impact of resident duty hour reform on hospital readmission rates among Medicare beneficiaries. *J Gen Intern Med.* 2011;26:405–411.
55. Privette AR, et al. Implementation of resident work hour restrictions is associated with a reduction in mortality and provider-related complications on the surgical service: a concurrent analysis of 14,610 patients. *Ann Surg.* 2009;250:316–321.
56. Rajaram R, et al. Association of the 2011 ACGME resident duty hour reform with general surgery patient outcomes and with resident examination performance. *JAMA.* 2014 Dec 10;312:2374–2384.
57. Rajaram R, et al. Association of the 2011 ACGME resident duty hour reform with postoperative patient outcomes in surgical specialties. *J Am Coll Surg.* 2015;221:748–757.
58. Rosen AK, et al. Effects of resident duty hour reform on surgical and procedural Patient Safety Indicators among hospitalized Veterans Health Administration and Medicare patients. *Med Care.* 2009;47:723–731.
59. Salim A, et al. Impact of the 80-hour workweek on patient care at a Level I trauma center. *Arch Surg.* 2007;142:708–712.
60. Scally CP, et al. Early impact of the 2011 ACGME duty hour regulations on surgical outcomes. *Surgery.* 2015;158:1453–1461.
61. Schlosser K, et al. Call-associated acute fatigue in surgical residents—subjective perception or objective fact? A cross-sectional observational study to examine the influence of fatigue on surgical performance. *World J Surg.* 2012;36:2276–2287.
62. Schroepel TJ, et al. How to further decrease the efficiency of care at a Level I trauma center: implement the amended resident work hours. *Am Surg.* 2015;81:698–703.
63. Seklehner S, et al. Impact of hours worked by a urologist prior to performing ureteroscopy on its safety and efficacy. *Scand J Urol.* 2016;50:56–60.
64. Shelton J, et al. Patient safety in the era of the 80-hour workweek. *J Surg Educ.* 2014;71:551–559.
65. Silber JH, et al. Patient safety outcomes under flexible and standard resident duty-hour rules. *N Engl J Med.* 2019 Mar 7;380:905–914.
66. Smith A, et al. Association of otolaryngology resident duty hour restrictions with procedure-specific outcomes in head and neck endocrine surgery. *JAMA Otolaryngol Head Neck Surg.* 2017 Jun 1;143:549–554.
67. Taffinder NJ, et al. Effect of sleep deprivation on surgeons' dexterity on laparoscopy simulator. *Lancet.* 1998 Oct 10;352:1191.
68. Talusan PG, et al. Effects of fatigue on driving safety: a comparison of brake reaction times in night float and postcall physicians in training. *J Grad Med Educ.* 2014;6:653–657.
69. Tsafrir Z, et al. Effects of fatigue on residents' performance in laparoscopy. 2015;221:564–570.e3.
70. Uchal M, et al. The impact of sleep deprivation on product quality and procedure effectiveness in a laparoscopic physical simulator: a randomized controlled trial. *Am J Surg.* 2005;189:753–757.
71. Veddeng A, et al. Impact of night shifts on laparoscopic skills and cognitive function among gynecologists. *Acta Obstet Gynecol Scand.* 2014;93:1255–1261.
72. Volpp KG, et al. Mortality among hospitalized Medicare beneficiaries in the first 2 years following ACGME resident duty hour reform. *JAMA.* 2007 Sep 5;298:975–983.
73. Volpp KG, et al. Mortality among patients in VA hospitals in the first 2 years following ACGME resident duty hour reform. *JAMA.* 2007 Sep 5;298:984–992.
74. Volpp KG, et al. Did duty hour reform lead to better outcomes among the highest risk patients? *J Gen Intern Med.* 2009;24:1149–1155.
75. Volpp KG, et al. Teaching hospital five-year mortality trends in the wake of duty hour reforms. *J Gen Intern Med.* 2013;28:1048–1055.
76. Yaghoubian A, et al. Decreased bile duct injury rate during laparoscopic cholecystectomy in the era of the 80-hour resident workweek. *Arch Surg.* 2008;143:847–851.
77. Yaghoubian A, et al. Trauma surgery performed by "sleep deprived" residents: are outcomes affected? *J Surg Educ.* 2010;67:449–451.
78. Yaghoubian A, et al. Acute care surgery performed by sleep deprived residents: are outcomes affected? *J Surg Res.* 2010;163:192–196.

79. Yi WS, Hafiz S, Sava JA. Effects of night-float and 24-h call on resident psychomotor performance. *J Surg Res.* 2013;184:49–53.
80. Browne JA, et al. Resident duty-hour reform associated with increased morbidity following hip fracture. *J Bone Joint Surg Am.* 2009;91:2079–2085.
81. Churnin I, Michalek J, Seifi A. Association of resident duty hour restrictions on mortality of nervous system disease and disorder. *J Grad Med Educ.* 2016;8:576–580.
82. Smith A, et al. Duty hour restrictions and surgical complications for head and neck key indicator procedures. *Laryngoscope.* 2017;127:1797–1803.
83. Fischer D, et al. Updating the “Risk Index”: a systematic review and meta-analysis of occupational injuries and work schedule characteristics. *Chronobiol Int.* 2017;34:1423–1438.
84. Folkard S, Lombardi DA. Modeling the impact of the components of long work hours on injuries and “accidents. *Am J Ind Med.* 2006;49:953–963.
85. Bolster L, Rourke L. The effect of restricting residents’ duty hours on patient safety, resident well-being, and resident education: an updated systematic review. *J Grad Med Educ.* 2015;7:349–363.
86. Philibert I. What is known: examining the empirical literature in resident work hours using 30 influential articles. *J Grad Med Educ.* 2016;8:795–805.
87. Levine AC, Adusumilli J, Landrigan CP. Effects of reducing or eliminating resident work shifts over 16 hours: a systematic review. *Sleep.* 2010;33:1043–1053.
88. Lowe CJ, Safati A, Hall PA. The neurocognitive consequences of sleep restriction: a meta-analytic review. *Neurosci Biobehav Rev.* 2017;80:586–604.
89. Cohen DA, et al. Uncovering residual effects of chronic sleep loss on human performance. *Sci Transl Med.* 2010 Jan 13;2:14ra3.
90. Mauri L, Sr D’Agostino RB. Challenges in the design and interpretation of noninferiority trials. *N Engl J Med.* 2017 Oct 5;377:1357–1367.
91. Agency for Healthcare Research and Quality. HCUP Statistical Brief 246. Overview of U.S. Hospital Stays in 2016: Variation by Geographic Region. Freeman WJ, Weiss AJ, Heslin KG. Dec 2018. Accessed Jul 16, 2023 <https://www.hcup-us.ahrq.gov/reports/statbriefs/sb246-Geographic-Variation-Hospital-Stays.pdf>.
92. Shea JA, Silber JH, Desai SV the iCOMPARE Research Group, et al. Development of the individualised Comparative Effectiveness of Models Optimizing Patient Safety and Resident Education (iCOMPARE) trial: a protocol summary of a national cluster-randomised trial of resident duty hour policies in internal medicine. *BMJ Open.* 2018;8:e021711. doi:10.1136/bmjopen-2018-021711.
93. National Highway Traffic Safety Administration. Seat Belts Save Lives. Accessed Jul 16, 2023. <https://www.nhtsa.gov/seat-belts/seat-belts-save-lives>.
94. Hanna TN, et al. The effects of fatigue from overnight shifts on radiology search patterns and diagnostic performance. *J Am Coll Radiol.* 2018;15:1709–1716.
95. Goel N, et al. Circadian rhythms, sleep deprivation, and human performance. *Prog Mol Biol Transl Sci.* 2013;119:155–190.
96. Hanna TN, et al. Effect of shift, schedule, and volume on interpretive accuracy: a retrospective analysis of 2.9 million radiologic examinations. *Radiology.* 2018;287:205–212.
97. Landrigan CP, et al. Effect on patient safety of a resident physician schedule without 24-hour shifts. *N Engl J Med.* 2020 Jun 25;382:2514–2523.
98. Barber EL, Eisenberg DL, Grobman WA. Type of attending obstetrician call schedule and changes in labor management and outcome. *Obstet Gynecol.* 2011;118:1371–1376.
99. Rothschild JM, et al. Risks of complications by attending physicians after performing nighttime procedures. *JAMA.* 2009 Oct 14;302:1565–1572.
100. Weaver MD, et al. The association between resident physician work-hour regulations and physician safety and health. *Am J Med.* 2020;133:e343–e354.