National improvements in resident physician-reported patient safety after limiting first-year resident physicians’ extended duration work shifts: a pooled analysis of prospective cohort studies

Matthew D Weaver, Christopher P Landrigan, Jason P Sullivan, Conor S O’Brien, Salim Qadri, Natalie Viyaran, Charles A Czeisler, Laura K Barger

ABSTRACT

Background The Accreditation Council for Graduate Medical Education (ACGME) enacted a policy in 2011 that restricted first-year resident physicians in the USA to work no more than 16 consecutive hours. This was rescinded in 2017.

Methods We conducted a nationwide prospective cohort study of resident physicians for 5 academic years (2002–2007) before and for 3 academic years (2014–2017) after implementation of the 16 hours 2011 ACGME work-hour limit. Our analyses compare trends in resident physician-reported medical errors between the two cohorts to evaluate the impact of this policy change.

Results 14,796 residents provided data describing 78,101 months of direct patient care. After adjustment for potential confounders, the work-hour policy was associated with a 32% reduced risk of resident physician-reported significant medical errors (rate ratio (RR) 0.68; 95% CI 0.64 to 0.72), a 34% reduced risk of reported preventable adverse events (RR 0.66; 95% CI 0.59 to 0.74) and a 63% reduced risk of reported medical errors resulting in patient death (RR 0.37; 95% CI 0.28 to 0.49).

Conclusions These findings have broad relevance for those who work in and receive care from academic hospitals in the USA. The decision to lift this work hour policy in 2017 may expose patients to preventable harm.

INTRODUCTION

Resident physicians’ work hours have been a subject of controversy in the USA for more than two decades.

From 2003 to 2011, the Accreditation Council for Graduate Medical Education (ACGME) limited residents in their first postgraduate year to a maximum of 30 consecutive work hours, including 6 hours for continuity of care and educational activities (30 hours 2003 ACGME work-hour limit). Subsequent evaluations found that shifts of 24 or more hours were associated with increased odds of fatigue-related medical errors and preventable adverse events (PAEs).
percutaneous injuries and motor vehicle crashes. A randomised controlled trial found that limiting first-year resident physicians to 16 consecutive work hours significantly improved resident alertness and patient safety. Altogether, a body of evidence accumulated suggesting that reducing or eliminating shifts longer than 16 hours did not negatively impact resident education and likely improved patient safety and resident quality of life. Subsequently, the Institute of Medicine of the National Academies (IOM) reviewed the available evidence and concluded that it was unsafe for any resident physician to provide clinical care for >16 consecutive hours without sleep. In response, the ACGME issued new work-hour regulations on 1 July 2011, limiting first-year resident physicians to a maximum of 16 consecutive work hours and emphasising a commitment to patient safety and mitigation of fatigue-related risks (16 hours 2011 ACGME work-hour limit). The response within the medical community to the 16 hours 2011 ACGME work-hour limit was mixed. Many stakeholders expected the changes to diminish the educational experience. The increased frequency of patient handoffs raised concerns, as physician-to-physician handoffs have historically been non-standardised and prone to error. In addition, the work-hour limitations were not accompanied by an increased number of residency slots, leading to work compression and a shift in some responsibilities to other clinical providers, as well as concerns about resident physician understaffing. Several studies of the 16 hours 2011 ACGME work-hour limit found that it had no impact on hospital-level mortality or mortality following surgical procedures. In light of these studies and opposition to the work-hour limit from within the medical community, the ACGME lifted the 16-hour limit as of 1 July 2017, again allowing first-year resident physicians to be scheduled for 24 hours of continuous work, plus up to 4 hours for care transitions (28 hours 2017 ACGME work-hour limit).

A limitation of several studies that informed this latest policy change is that the primary end points lacked sensitivity (eg, hospital-wide mortality) or were only tangentially related to the care provided by first-year resident physicians (eg, evaluating the impact of the 16 hours work-hour limit for first-year surgical residents using the outcome of fatal or serious complications after surgical procedures within institutions where first-year surgical residents have a minimal role in surgical procedures). The impact of the 16 hours 2011 ACGME work-hour limit on patient safety outcomes directly reported by resident physicians has not been studied. Therefore, we undertook an evaluation to determine whether resident physician-reported patient safety outcomes changed following implementation of the 16 hours 2011 ACGME work-hour limit.

METHODS

We carried out nationwide prospective cohort studies of resident physicians for 5 academic years (2002–2007) before and for 3 academic years (2014–2017) after implementation of the 16 hours 2011 ACGME work-hour limit. Data were not available for transition years (2008–2013). During the transition period resident physician work-hour changes were being made nationwide in response to the IOM report (findings first made public in 2008), and a 2010 ACGME report reaffirmed many of the IOM conclusions and established a target date of 1 July 2011 for implementation of the new policies, with exceptions for programmes that needed one or more additional years to make the transition. Our analyses compare trends in resident physician-reported medical errors between the two cohorts. A Certificate of Confidentiality was issued by the Centers for Disease Control and Prevention. All variables used in this analysis were collected using the same wording and frequency in both cohorts. The methods have been described in detail in prior publications.

Recruitment and data collection

In the initial cohort, all US medical school graduates and all additional individuals who matched to a US residency programme received an invitation to participate in the study. In the second cohort, all medical school graduates who completed an application through the Electronic Residency Application Service were emailed the same invitation to participate.

Respondents were asked to provide electronic informed consent. In the first month of residency, individualised links were sent via email to complete the baseline survey. The baseline survey collected demographic information. Subsequent monthly reports collected work-hour information, including the frequency of extended duration shifts, hours engaged in patient care and total hours of work. Participants also self-reported the primary outcome: significant medical errors in the past month (“Do you believe that you made a significant medical error, whether or not an adverse patient outcome occurred”), whether or not they believed fatigue contributed to the error, the number of errors and the patient outcome resulting from the errors. Resident physician-reported errors resulting in patient harm were considered PAEs. We also queried how many times participants nodded off or fell asleep (attentional failures) during specific patient-care activities (during surgery or while talking to, or examining, patients).

Statistical analysis

The primary analysis tested the association between the 16 hours 2011 ACGME work-hour limit and the resident physician-reported patient safety outcomes of interest. Secondary analyses tested the association between weekly work hours, extended duration
shifts and resident physician-reported patient safety outcomes independent of cohort. We excluded months where participants reported ≥14 work-free days, or where work-hour information was missing or exceeded 168 hours of work per week. We further excluded months where no direct patient care was reported (eg, research rotation).

Demographics were compared between groups using Wilcoxon rank-sum test and Pearson’s χ² test. The sum of hours spent physically awake in the hospital, classes or workplace (excluding occasional employment outside the residency (moonlighting)), plus the number of hours asleep in the hospital were considered weekly work hours. Monthly reports of work, sleep and stress were compared using generalised linear mixed models which accounted for clustering of respondents and the dependence between repeated measures.

We calculated the incidence of each resident-reported patient safety outcome before and after the 16 hours 2011 ACGME work-hour limit. Incidence rate ratios (IRR) were constructed, and the significance of the IRR was tested using likelihood ratio tests in log-linear models. Similar results were obtained using Pearson’s and deviance-based scaled Poisson models. Rare outcomes were dichotomised. We used generalised linear mixed models with a binomial distribution and log-link function to estimate the risk of each outcome. All models adjusted for hours spent in patient care that month. We identified additional confounders a priori based on relevance to the research question and biological plausibility. We imputed age with the median when missing using the missing indicator method.22 Fully adjusted models controlled for age, gender, specialty programme, university-based versus community hospital, an imputation indicator variable and the hours of patient care for that month. Secondary analyses used mixed effects models with data from both cohorts to examine extended duration shifts or weekly work hours as independent variables of interest. These pooled analyses controlled for cohort using a conditional likelihood approach. Sensitivity analyses stratified by cohort and surgical/non-surgical specialty. SAS (V9.4, Cary, North Carolina, USA) was used for statistical analysis. All tests were two-sided and p<0.05 was considered statistically significant.

RESULTS
A total of 21862 residents who matched to a US residency programme during the study interval consented to participate (including 9% of all US medical residents from 2002 to 2007 and 18% of US medical residents from 2014 to 2017). Approximately 75% of those who consented provided data and were included in the study cohort (cooperation rate). Six thousand two hundred eleven first-year resident physicians completed 45261 monthly reports between 2002 and 2007 and 9778 first-year resident physicians completed 61205 monthly reports between 2014 and 2017 (106466 total monthly reports). After applying exclusion criteria, 80266 monthly reports from 15276 first-year resident physicians were available21 and 14796 individual residents reporting on 78101 months with direct patient care were included in the current analyses (online supplemental figure 1).

Characteristics of the study sample
Resident physicians who contributed data after 2011 were less likely to report female gender and affiliation with a university-based programme (table 1).21 The composition of specialties was largely similar. Weekly hours of work decreased following the 16 hours 2011 ACGME work-hour limit from 71 hours (SD 18) to 62 hours (SD 17), p<0.001, while hours engaged in patient care remained the same (50 hours (SD 20) vs 49 hours (SD 19), p=0.40). Participants reported fewer extended duration shifts per month (4.0 (SD 3.4) vs 0.2 (SD 1.2), p<0.001).

Incidence and adjusted risk of resident physician-reported adverse safety outcomes following the 16 hours 2011 ACGME work-hour limit
We observed an absolute excess of 0.611 (95% CI 0.305 to 0.913) resident physician-reported serious medical errors per-person-year, 0.086 (95% CI 0.043 to 0.129) PAEs per-person-year, and 0.042 (95% CI 0.020 to 0.064) errors resulting in patient death per-person-year prior to the 16 hours 2011 ACGME work-hour limit (relative to after the work-hour limit). The incidence for each adverse resident physician-reported patient safety outcome studied was significantly reduced following the 16 hours 2011 ACGME work-hour limit, both in unadjusted analyses (table 2) and after adjustment for potential confounders (figure 1). The risk of at least one significant medical error reported in a month was reduced 32% (rate ratio (RR) 0.68; 95%CI 0.64 to 0.72, fully adjusted model). A similar decrease was noted in the rate of reported PAEs (RR 0.66; 95%CI 0.59 to 0.74), while the risk of first-year resident physicians reporting a fatal PAE was reduced >60% (RR 0.37; 95%CI 0.28 to 0.49). The rate of patient-care-related attentional failures was reduced >40% (IRR 0.58; 95%CI 0.54 to 0.62).

Work hours, extended duration shifts and resident physician-reported patient safety outcomes
Participants in both cohorts reported exceeding 80 hours of work per week (averaged across 4 weeks) and working extended duration shifts (EDS). Independent of cohort, working >80 hours per week and working EDS were each independently associated with an increased risk of resident physician-reported significant medical error (table 3). Compared with working no EDS, working even one EDS in a month increased the risk of resident physician-reported adverse outcomes. Similarly, compared with working up to 60 hours per week, increased risks of resident physician-reported adverse outcomes were seen for
those residents working >60–70 hours per week, those working >70–80 hours per week and those working >80 hours per week in a dose-dependent fashion (table 4). Stratified analyses are presented in online supplemental tables 1 and 2.

**DISCUSSION**
Following implementation of the 2011 ACGME work-hour restrictions, we found that first-year resident physicians reported making fewer significant medical errors, errors resulting in patient harm and errors
resulting in patient death. Despite fears that the policy would limit hands-on training, reported hours spent in patient care remained unchanged. Extrapolating these findings to the first-year resident physician population in the USA would suggest that prior to the implementation of the 2011 reforms, first-year resident physicians perceived that they committed an additional 22,000 errors, 3100 of which led to PAE, and 1500 of which resulted in death, on a yearly basis compared to after the reform.

Our results are in contrast with other recent evaluations of the ACGME’s implementation of the 16-hour shift limit in 2011. In the FIRST (Flexibility in duty hour Requirements for Surgical Trainees) trial, first-year resident physicians at surgical programmes assigned to the flexible work-hour group were allowed to exceed 16 consecutive hours of work and required time off between shifts was reduced. The mortality rate following general surgery was not inferior at teaching hospitals allowed to exceed the 16 hours 2011 ACGME work-hour limit compared with hospitals assigned to follow those limits. However, in those hospitals, the first-year surgical residents have minimal roles in the surgical outcomes that were evaluated (surgical mortality and major complications, the study’s primary end point), as first-year surgical residents spend only a small fraction of their time in the operating room (estimates suggest 1 hour per day), and their roles there are constrained by design (with involvement in 1% of complex cases). Moreover, no information was collected on the actual hours worked by the surgical residents in the FIRST trial, or whether the institutions that were allowed to schedule their surgical residents for >16 hours per shift actually did so, given the potential exposure to malpractice liability due to the IOM report and concerns regarding the ethics of the trial.

Another recent effort, the iCOMPARE (Individualized Comparative Effectiveness of Models Optimizing Patient Safety and Resident Education) trial, compared programmes randomised to flexible work hours (where extended duration shifts were permitted) with programmes that maintained the 2011 ACGME work-hour restrictions, and similarly found that flexible programmes were not inferior to standard programmes with respect to 30-day hospital-level mortality (primary outcome) or patient safety indicators. Like the FIRST trial, the iCOMPARE trial had a primary end point that lacked sensitivity and a direct connection to the care provided by first-year resident physicians. In addition, the prespecified non-inferiority threshold allowed for non-inferiority to be declared as long as hospital-level 30-day mortality for any cause did not increase in the flexible group by >1%. Whether this prespecified threshold is consistent with an acceptably small degree of harm is questionable, and clinically significant decrements in patient safety may have gone unrecognised as a result. In any case, neither of these studies suggested limiting work hours

### Table 3

<table>
<thead>
<tr>
<th>N cases</th>
<th>N person months</th>
<th>WWH</th>
<th>EDS</th>
<th>Combined categorical WWH and EDS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>≤80</td>
<td>&gt;80</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Attentual failures | 26 576 | 78 101 | Ref | 2.15 (2.04 to 2.26) | * Ref | 2.17 (2.03 to 2.32) | * Ref | 2.06 (1.92 to 2.21) | 1.95 (1.79 to 2.13) | 3.69 (3.42 to 3.98) | *
| Medical failures | 25 166 | 78 101 | Ref | 2.07 (1.97 to 2.18) | * Ref | 1.83 (1.71 to 1.95) | * Ref | 1.86 (1.74 to 2.00) | 2.18 (2.00 to 2.37) | 3.07 (2.85 to 3.32) | *
| PAE | 1673 | 78 101 | Ref | 2.07 (1.82 to 2.35) | * Ref | 1.62 (1.38 to 1.89) | * Ref | 1.63 (1.36 to 1.94) | 2.37 (1.94 to 2.88) | 2.67 (2.22 to 3.19) | *
| Fatal PAE | 274 | 78 101 | Ref | 3.2 (1.97 to 1.78) | Ref | 1.88 (1.25 to 2.84) | † Ref | 2.05 (1.34 to 3.14) | 1.77 (0.99 to 3.14) | 2.01 (1.28 to 3.15) | †

All models adjusted for age, gender, specialty, hours in patient care, community versus university programme and cohort. Created by the authors.

*P<0.001.
1P=0.002.
is harmful to patient safety. Therefore, given the associated improvements in physician safety that we have previously reported, and the improvement in resident physician-reported patient safety that we report here, taken together, these studies indicate that the 2011 ACGME policy limiting first-year resident physician work hours was associated with an improvement in both physician and patient safety.

In addition to these two trials, prior retrospective studies using administrative data similarly concluded that the 16 hours 2011 ACGME work-hour limit did not appreciably impact hospital-level mortality or process of care measures, but like FIRST and iCOMPARE, these studies relied on process and outcome measures that were relatively insensitive and lacked a direct connection to the work of first-year resident physicians. In contrast to previous efforts, our study design enabled collection of patient-safety and work-hour data directly from the residents who provided care. We found that resident physician-reported significant medical errors, PAEs and errors resulting in fatal patient outcomes were reported at substantially increased rates in months when extended duration shifts were worked. The findings from this study, which focused on the impact of resident physician work-hour limits on patient safety using metrics that were directly related to care provided by the resident physicians, are consistent with extensive evidence compiled by the ACGME indicating that sleep deficiency is associated with degradation of performance among physicians performing clinical tasks.

We also found that when resident physicians averaged 70–80 hours of work per week over a month, which is within the current standard of an average of 80 work hours per week averaged over a month, they had >2.5 times the odds of reporting a significant medical error, a medical error that injured a patient and a medical error that resulted in the death of a patient, as compared with weeks in which the resident physicians worked no more than 60 hours per week on average. In the USA, resident physicians are often scheduled to work more weekly work hours than equivalents in other countries. We found that even after controlling for the increased hours in patient care that could be expected to result from increased total work hours, increasing total work hours was associated with increased risk of reported significant medical errors, PAEs and errors resulting in patient death. Our data suggest that reducing weekly work hours would improve patient safety. The increased risk of long work weeks and extended duration shifts should be made more transparent to patients and providers.

Our study has several limitations. Our study population included 9% of all first-year US medical residents from 2002 to 2007 and 18% of all first-year US medical residents from 2014 to 2017. We examined demographic characteristics for evidence of non-response bias. While the demographic characteristics of participants in this sample are similar to the demographic characteristics of resident physicians nationally (online supplemental table 3), those who chose to participate in the study may differ from non-participants in ways that are not captured by our analysis. As it was observational, we are unable to directly attribute the findings to the 2011 ACGME work-hour restrictions. The estimated associations could be affected by residual confounding or other forms of bias. We controlled for multiple potential confounders, but residual confounding may persist as a result of factors we did not collect or imprecise measurement of the confounders that were collected.

Significant medical errors were collected by self-report. This approach to the collection of medical errors could be subject to social desirability bias, recall bias or erroneous self-observation. Self-report is an established methodology that is commonly used. Prior work has shown that self-reported errors are corroborated in the medical record >80% of the time and produce similar or lower estimates of medical error compared with retrospective chart review. Alternative detection methods, such as trigger tools and direct observation have been shown to detect higher rates of medical errors compared with self-report. By using self-report, our estimates for medical errors are likely to be conservative, which would bias our results towards the null. Although we expect that resident physician-reported errors represent only a fraction of all errors that occur throughout the healthcare system, we do not have reason to believe that the propensity to report errors would have changed over time in a

### Table 4

<table>
<thead>
<tr>
<th>Weekly work hours (n=person months)</th>
<th>≤60 (n=27989)</th>
<th>&gt;60 and ≤70 (n=17175)</th>
<th>&gt;70 and ≤80 (n=19221)</th>
<th>&gt;80 (n=13716)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attentional failures</td>
<td>Ref</td>
<td>1.71 (1.60 to 1.81)</td>
<td>2.25 (2.12 to 2.39)</td>
<td>3.50 (3.28 to 3.73)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Medical errors</td>
<td>Ref</td>
<td>1.82 (1.71 to 1.94)</td>
<td>2.53 (2.38 to 2.68)</td>
<td>3.58 (3.36 to 3.82)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>PAE</td>
<td>Ref</td>
<td>2.13 (1.79 to 2.53)</td>
<td>2.66 (2.26 to 3.14)</td>
<td>3.89 (3.27 to 4.63)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Fatal PAE</td>
<td>Ref</td>
<td>1.75 (1.16 to 2.64)</td>
<td>2.69 (1.84 to 3.94)</td>
<td>2.39 (1.59 to 3.59)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

All models are adjusted for age, gender, specialty, community versus university programme, hours in patient care and cohort. Created by the authors.
manner that introduced bias. Clinician perceptions of medical errors can also have a pervasive impact on physician well-being, and have been associated with increased odds of physician burnout and depression.39

While some reductions in the rate of medical errors may be attributable to secular trends, most studies have not found temporald reductions in the overall rate of harm attributable to medical care over the time period under study that could explain our results.40–43 Evaluations prior to the reform showed no reduction in the rate of medical error over time, although one recently published effort reported a declining trend in patient safety indicators from 2000 to 2013.44 Most estimates following the reform suggest the problem may be larger than previously reported, and remains highly prevalent.31 43 44 Our findings have important policy implications for >100,000 resident physicians in the USA. The very substantial reductions in harmful resident physician-reported medical errors over a relatively short period of time are striking, and together with the previously reported improvements in occupational safety among the resident physicians,21 suggest a causal link to the policy implemented at that point in time for this subset of clinicians and the improvements in both patient and physician safety.

CONCLUSIONS
In a large nationwide prospective cohort study, we found that implementation of a policy that eliminated shifts exceeding 16 consecutive hours for first-year resident physicians was associated with a substantial reduction in harmful and fatal medical errors reported by the resident physicians themselves. Extended duration shifts were associated with significant and harmful medical errors even after the implementation of the policy. Surveillance efforts are needed now that national regulations allow shifts of up to 28 consecutive hours for first-year resident physicians.

Author affiliations
1Brigham and Women’s Hospital, Division of Sleep and Circadian Disorders, Departments of Medicine and Neurology, Boston, Massachusetts, USA
2Division of Sleep Medicine, Harvard Medical School, Boston, Massachusetts, USA
3Division of General Pediatrics, Department of Pediatrics, Boston Children’s Hospital, Boston, Massachusetts, USA
4Departments of Pediatrics and Medicine, and Division of Sleep Medicine, Harvard Medical School, Boston, Massachusetts, USA

Acknowledgements We thank the resident physicians who took time from their busy work schedules to participate in this study; to the Association of American Medical Colleges, especially Marie Caulfield and Brianna Gunter, for their assistance with recruitment; to Steven W. Lockley for his assistance with the study design and data interpretation; to research assistants Michael Shrieve and Justin Bue, and students Kelly Racicot, Lars Johnsen, Erin Rothchild, Autumn Gertz, Niamh Mulrooney, Matthew Stewart, Jacob Stroud, Gregory Pavlov, Krystal Lipomi, Arlene Perez, and Julian Smith-Sparks.

Contributors CPL, CAC and LKB designed the parent study and developed the data collection tools. SQ managed data collection and data security. JPS, CSO and NV provided daily study oversight. MDW conducted the analysis and wrote the initial draft. All authors provided substantive contributions and critical revisions to the final product. MDW is the guarantor, accepting full responsibility for the work and the conduct of the study, had access to the data, and controlled the decision to publish.

Funding This study was funded by National Institute for Occupational Safety and Health (R01OH010300, R01OH011773, R01OH07567) and the National Heart, Lung, and Blood Institute (F32HL134249, R56HL151637, T32HL.007901, U01HL11478).

Competing interests MDW reports grant support from the CDC, NIOSH, NHLBI and the Brigham Research Institute. He also reports institutional support and/or consulting fees from Delta Airlines, the National Sleep Foundation, Puget Sound Pilots, and the University of Pittsburgh. CPL has served as a paid consultant to the Midwest Lighting Institute to help study the effect of blue light on health care provider performance and safety. CPL 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, CPL 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. CAC reported receiving grants from BWH from CDC, NHLBI, National NIA, NIOSH and DOD during the conduct of the study; being a paid consultant to or speaker for Genesco, Institute of Digital Media and Child Development, Klarman Family Foundation, M. Davis and Co, Physician’s Seal, Puget Sound Pilots, Samsung Group, State of Washington Board of Pilotage Commissioners, Tencen Holdings, Teva Pharma Australia, and Vanda Pharmaceuticals, in which Dr Czeisler holds an equity interest; receiving travel support from Aspen Brain Institute, Bloomage International Investment Group, UK Biotechnology and Biological Sciences Research Council, Bouley Botanical, Dr Stanley Ho Medical Development Foundation, Illuminating Engineering Society, National Safety Council, Tencen Holdings, and The Wonderful Co; receiving institutional research and/or education support from Cephalon, Mary Ann and Stanley Snider via Combined Jewish Philanthropies, Harmony Biosciences, Jazz Pharmaceuticals PLC, Johnson and Johnson, Neurocare, Peter Brown and Margaret Hamburg, Philips Respironics, Regeneron Pharmaceuticals, Regional Home Care, Teva Pharmaceuticals Industries, Sanofi A., Optum, ResMed, San Francisco Bar Pilots, Schneider National, Serta, Simmons Betting, Sysco, Vanda Pharmaceuticals; being or having been an expert witness in legal cases, including those involving Advanced Power Technologies; Aegis Chemical Solutions; Amtrak; Casper Sleep; and J Energy Services; Complete General Construction; Dallas Police Association; Enterprise Rent-A-Car; Steel Warehouse Co; FedEx; Greyhound Lines; Palomar Health District; PAR Electrical, Product, and Logistics Services; Packet Emergency Medical Services; South Carolina Central Railroad Co; Union Pacific Railroad; UPS; and Vanda Pharmaceuticals; serving as the incumbent of an endowed professorship provided to Harvard University by Cephalon; and receiving royalties from McGraw Hill and Philips Respironics for the Actiwatch-2 and Actiwatch Spectrum devices. Dr Czeisler’s interests were reviewed and are managed by the Brigham and Women’s Hospital and Mass General Brigham in accordance with their conflict-of-interest policies. LKB was supported by grants from NIOSH, NIA and NHLBI, and has consulted with and/or received institutional support from Boston Children’s Hospital, the AAA Foundation, Delta Airlines, Puget Sound Pilots, the University of Pittsburgh, and the CDC.

Patient consent for publication Not applicable.

Ethics approval This study was approved by Partners Human Research Committee (Protocol #: 2002P000202). Participants gave informed consent to participate in the study before taking part.

Weaver MD, et al. BMJ Qual Saf 2022;0:1–9. doi:10.1136/bmjqs-2021-014375
Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID ID Matthew D Weaver http://orcid.org/0000-0003-3578-336X

REFERENCES
41 Makary MA, Daniel M. Medical error—the third leading cause of death in the US. *BMJ* 2016;353:i2139.
Supplemental Table of Contents

Page 1. Table of contents
Page 2: Flow diagram detailing participation in the study.
Page 3: The adjusted association between extended duration shifts and adverse safety outcomes stratified by cohort and surgical vs. non-surgical specialty.
Page 4: Representativeness of the study sample
Supplemental Figure 1. Participation in the study and exclusion criteria.

- 2002-2007
  - 93,984 PGY1s
  - Matched to residency
  - 9% participation rate
  - 2002-2007
  - 8,570 PGY1s
  - Consented to participate
  - 73% cooperation rate

- Exclusions
  - Months with vacation or ≥14 free days

- 2002-2007
  - 6,211 PGY1s
  - 45,261 Monthly Reports

- 2014-2017
  - 73,764 PGY1s
  - Matched to residency
  - 18% participation rate
  - 2014-2017
  - 13,292 PGY1s
  - Consented to participate
  - 74% cooperation rate

- 2014-2017
  - 9,778 PGY1s
  - 61,205 Monthly Reports

- Exclusions
  - Months with missing weekly work hour information, >168 weekly work hours, or <12 hours of work for all 4 weeks

- 2014-2017
  - 9,671 PGY1s
  - 50,666 Monthly Reports

- 2014-2017
  - 9,601 PGY1s
  - 49,495 Monthly Reports

- Months where ≥16 extended duration shifts were reported

- 2014-2017
  - 9,596 PGY1s
  - 49,418 Monthly Reports

- Months where direct patient care was reported

- 2014-2017
  - 9,311 PGY1s
  - 47,946 Monthly Reports

- 2014-2017
  - 14,796 PGY1s
  - 78,101 Monthly Reports
### Supplemental Table 1 The adjusted association between extended duration shifts and resident-physician-reported adverse safety outcomes stratified by cohort.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>≥1</td>
<td>None</td>
<td>≥1</td>
</tr>
<tr>
<td>Attentional Failures</td>
<td>Ref 2.17 (2.03-2.32) ***</td>
<td>Ref 2.50 (2.28-2.74) ***</td>
<td>Ref 1.61 (1.43-1.81) ***</td>
</tr>
<tr>
<td>Medical Errors</td>
<td>Ref 1.83 (1.71-1.95) ***</td>
<td>Ref 2.00 (1.84-2.18) ***</td>
<td>Ref 1.55 (1.38-1.74) ***</td>
</tr>
<tr>
<td>PAE</td>
<td>Ref 1.62 (1.38-1.89) ***</td>
<td>Ref 1.81 (1.45-2.26) ***</td>
<td>Ref 1.37 (1.03-1.83) *</td>
</tr>
<tr>
<td>Fatal PAE</td>
<td>Ref 1.88 (1.25-2.84) *</td>
<td>Ref 1.67 (1.09-2.56) *</td>
<td>Failed to converge</td>
</tr>
</tbody>
</table>

All models adjusted for age, gender, specialty, hours in patient care, and community vs. university program. P-values are obtained from likelihood ratio tests.  
*** p-value<0.001  
** p-value=0.001.  
* p-value>0.001 & <0.05.

### Supplemental Table 2. The adjusted association between extended duration shifts and resident-physician-reported adverse safety outcomes stratified by surgical vs. non-surgical specialties.

<table>
<thead>
<tr>
<th></th>
<th>Extended Duration Shifts (overall)</th>
<th>Extended Duration Shifts (Non-Surgical)</th>
<th>Extended Duration Shifts (Surgical)</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>≥1</td>
<td>None</td>
<td>≥1</td>
</tr>
<tr>
<td>Attentional Failures</td>
<td>Ref 2.17 (2.03-2.32) ***</td>
<td>Ref 2.37 (2.20-2.55) ***</td>
<td>Ref 1.33 (1.14-1.55) ***</td>
</tr>
<tr>
<td>Medical Errors</td>
<td>Ref 1.83 (1.71-1.95) ***</td>
<td>Ref 1.92 (1.79-2.06) ***</td>
<td>Ref 1.34 (1.13-1.60) **</td>
</tr>
<tr>
<td>PAE</td>
<td>Ref 1.62 (1.38-1.89) ***</td>
<td>Ref 1.77 (1.47-2.13) ***</td>
<td>Failed to converge</td>
</tr>
<tr>
<td>Fatal PAE</td>
<td>Ref 1.88 (1.25-2.84) *</td>
<td>Ref 2.00 (1.19-3.32) **</td>
<td>Failed to converge</td>
</tr>
</tbody>
</table>

All models adjusted for age, gender, hours in patient care, and community vs. university program. P-values are obtained from likelihood ratio tests.  
*** p-value<0.001  
** p-value=0.001.  
* p-value>0.001 & <0.05.
Supplemental Tables 3A and 3B. The demographics of the study sample compared to the characteristics of first-year residents nationally during the study interval.

<table>
<thead>
<tr>
<th>A</th>
<th>Study Sample</th>
<th>ACGME National Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2014-2017 (n=9,778)</td>
<td>2014-2015 (n=28,220)</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td>28.7±3.2</td>
<td>30.6</td>
</tr>
<tr>
<td><strong>Female Gender</strong></td>
<td>48%</td>
<td>44%</td>
</tr>
<tr>
<td><strong>Specialty</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internal Medicine</td>
<td>25%</td>
<td>34%</td>
</tr>
<tr>
<td>Family Medicine</td>
<td>12%</td>
<td>13%</td>
</tr>
<tr>
<td>Pediatrics</td>
<td>11%</td>
<td>11%</td>
</tr>
<tr>
<td>General Surgery and surgical specialties</td>
<td>10%</td>
<td>13%</td>
</tr>
<tr>
<td>Emergency Medicine</td>
<td>7%</td>
<td>7%</td>
</tr>
<tr>
<td>Obstetrics/Gynecology</td>
<td>5%</td>
<td>5%</td>
</tr>
<tr>
<td>Psychiatry</td>
<td>4%</td>
<td>5%</td>
</tr>
<tr>
<td>Anesthesiology</td>
<td>4%</td>
<td>3%</td>
</tr>
<tr>
<td>Other (including combined)</td>
<td>15%</td>
<td>9%</td>
</tr>
</tbody>
</table>

*Percentage of residents in each specialty.

<table>
<thead>
<tr>
<th>B</th>
<th>Study Sample</th>
<th>ACGME National Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2002-2007 (n=6,211)</td>
<td>2007-2008 (n=36,012)</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td>28.9±3.9</td>
<td>Not Reported</td>
</tr>
<tr>
<td><strong>Female Gender</strong></td>
<td>56%</td>
<td>41%</td>
</tr>
<tr>
<td><strong>Specialty</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internal Medicine</td>
<td>24%</td>
<td>31%</td>
</tr>
<tr>
<td>Family Medicine</td>
<td>11%</td>
<td>10%</td>
</tr>
<tr>
<td>Pediatrics</td>
<td>14%</td>
<td>10%</td>
</tr>
<tr>
<td>General Surgery and surgical specialties</td>
<td>11%</td>
<td>9%</td>
</tr>
<tr>
<td>Emergency Medicine</td>
<td>7%</td>
<td>4%</td>
</tr>
<tr>
<td>Obstetrics/Gynecology</td>
<td>6%</td>
<td>3%</td>
</tr>
<tr>
<td>Psychiatry</td>
<td>4%</td>
<td>5%</td>
</tr>
<tr>
<td>Anesthesiology</td>
<td>1%</td>
<td>4%</td>
</tr>
<tr>
<td>Other (including combined)</td>
<td>20%</td>
<td>22%</td>
</tr>
</tbody>
</table>

*Percentage of residents in each specialty.

Footnote: National PGY-1 specific demographic information is not available for years prior to 2007. Reporting practices may explain some differences in the demographic characteristics of our study sample compared to residents nationally. The lower prevalence of internal medicine residents in our study may be explained by the reporting of planned specialty rather than preliminary or transitional year.