



HHS Public Access

Author manuscript

N Engl J Med. Author manuscript; available in PMC 2019 September 07.

Published in final edited form as:

N Engl J Med. 2019 March 07; 380(10): 915–923. doi:10.1056/NEJMoa1810641.

Sleep and Alertness in a Duty-Hour Flexibility Trial in Internal Medicine

Mathias Basner, M.D., Ph.D., David A. Asch, M.D., M.B.A., Judy A. Shea, Ph.D., Lisa M. Bellini, M.D., Michele Carlin, B.A., Adrian J. Ecker, Susan K. Malone, Ph.D., Sanjay V. Desai, M.D., Alice L. Sternberg, Sc.M., James Tonascia, Ph.D., David M. Shade, J.D., Joel T. Katz, M.D., David W. Bates, M.D., Orit Even-Shoshan, M.S., Jeffrey H. Silber, M.D., Ph.D., Dylan S. Small, Ph.D., Kevin G. Volpp, M.D., Ph.D., Christopher G. Mott, M.S., Sara Coats, B.S., Daniel J. Mollicone, Ph.D., David F. Dinges, Ph.D., and for the iCOMPARE Research Group* University of Pennsylvania (M.B., D.A.A., J.A.S., L.M.B., M.C., A.J.E., S.K.M., D.S.S., K.G.V., D.F.D.), Corporal Michael J. Crescenz Veterans Affairs Medical Center (D.A.A., K.G.V.), Children's Hospital of Philadelphia (O.E.-S., J.H.S.), and Pulsar Informatics (C.G.M., S.C., D.J.M.) — all in Philadelphia; Johns Hopkins University, Baltimore (S.V.D., A.L.S., J.T., D.M.S.); and Brigham and Women's Hospital, Boston (J.T.K., D.W.B.).

Abstract

BACKGROUND—A purpose of duty-hour regulations is to reduce sleep deprivation in medical trainees, but their effects on sleep, sleepiness, and alertness are largely unknown.

METHODS—We randomly assigned 63 internal-medicine residency programs in the United States to follow either standard 2011 duty-hour policies or flexible policies that maintained an 80-hour workweek without limits on shift length or mandatory time off between shifts. Sleep duration and morning sleepiness and alertness were compared between the two groups by means of a noninferiority design, with outcome measures including sleep duration measured with actigraphy, the Karolinska Sleepiness Scale (with scores ranging from 1 [extremely alert] to 9 [extremely sleepy, fighting sleep]), and a brief computerized Psychomotor Vigilance Test (PVT-B), with long response times (lapses) indicating reduced alertness.

RESULTS—Data were obtained over a period of 14 days for 205 interns at six flexible programs and 193 interns at six standard programs. The average sleep time per 24 hours was 6.85 hours (95% confidence interval [CI], 6.61 to 7.10) among those in flexible programs and 7.03 hours (95% CI, 6.78 to 7.27) among those in standard programs. Sleep duration in flexible programs was noninferior to that in standard programs (between-group difference, -0.17 hours per 24 hours; one-sided lower limit of the 95% confidence interval, -0.45 hours; noninferiority margin, -0.5 hours; $P = 0.02$ for noninferiority), as was the score on the Karolinska Sleepiness Scale (between-

*A complete list of the members of the iCOMPARE Research Group is provided in the Supplementary Appendix, available at [NEJM.org](https://doi.org/10.1056/NEJMoa1810641).

Address reprint requests to Dr. Basner at the Unit for Experimental Psychiatry, Division of Sleep and Chronobiology, Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, 1019 Blockley Hall, 423 Guardian Dr., Philadelphia, PA 19104-6021, or at basner@penmedicine.upenn.edu.

A data sharing statement provided by the authors is available with the full text of this article at [NEJM.org](https://doi.org/10.1056/NEJMoa1810641).

Disclosure forms provided by the authors are available with the full text of this article at [NEJM.org](https://doi.org/10.1056/NEJMoa1810641).

group difference, 0.12 points; one-sided upper limit of the 95% confidence interval, 0.31 points; non-inferiority margin, 1 point; $P < 0.001$). Noninferiority was not established for alertness according to the PVT-B (between-group difference, -0.3 lapses; one-sided upper limit of the 95% confidence interval, 1.6 lapses; noninferiority margin, 1 lapse; $P = 0.10$).

CONCLUSIONS—This noninferiority trial showed no more chronic sleep loss or sleepiness across trial days among interns in flexible programs than among those in standard programs. Noninferiority of the flexible group for alertness was not established. (Funded by the National Heart, Lung, and Blood Institute and American Council for Graduate Medical Education; [ClinicalTrials.gov](https://clinicaltrials.gov) number, NCT02274818.)

Sleep loss and night work are associated with reduced alertness and cognitive performance¹ and represent a safety challenge for all 24/7 operations, including medicine.^{2–6} To promote the safety of patients and medical trainees, in 2003 the Accreditation Council for Graduate Medical Education established resident duty-hour policies that limited resident workweeks to 80 hours and shifts to 30 hours and in 2011 further limited shifts to 16 hours for interns.⁷ Rigorous analysis of observational data revealed essentially no differences in patient mortality after the implementation of either the 2003 or 2011 policy.^{8–10} The 2011 policy changes were largely reversed in 2017, after publication of the Flexibility in Duty Hour Requirements for Surgical Trainees (FIRST) Trial,^{11,12} a national, prospective, cluster-randomized trial of surgical residency programs that showed that surgical-complication and death rates were no worse under flexible duty-hour rules (without restrictions on shift length) than under standard rules. In addition to having possible effects on patient safety, duty-hour policies are likely to affect trainee education and the sleep and alertness of trainees.

Although restricting duty hours may reduce the prevalence of acute and chronic sleep loss in interns,¹³ transitions into and out of night shifts can result in fatigue from shift-work-related sleep loss and circadian misalignment.¹⁴ Preventing interns from participating in extended shifts may reduce educational opportunities, increase handoffs, and reduce continuity of care.^{15,16} Restricting duty hours may increase the necessity of cross-coverage, contributing to work compression for both interns and more senior residents.¹⁷ Clearly, duty-hour regulations affect interns in complex ways that extend to those responsible for their training.¹⁸

In this context, we created the Individualized Comparative Effectiveness of Models Optimizing Patient Safety and Resident Education (iCOMPARE) trial, a national cluster-randomized trial involving 63 internal-medicine residency programs.¹⁹ Residency programs were randomly assigned either to maintain 2011 duty-hour standards or to use a more flexible set of duty-hour standards, noted principally for removing the 16-hour restriction on shift length for interns in a pattern similar to that in the FIRST Trial (Table S1 in the Supplementary Appendix, available with the full text of this article at [NEJM.org](https://www.nejm.org)), during the 2015–2016 academic year.

Although the primary outcome of the iCOMPARE trial focused on patient safety, as reported in a companion article in this issue of the *Journal*,²⁰ the trial was designed to simultaneously assess the effects on educational experiences, which were reported previously,²¹ and on the

sleep and alertness of interns, which we report here. The latter aim was to establish whether sleep and alertness among interns in flexible programs were noninferior to those among interns in standard programs according to pre-specified noninferiority margins. In addition to these analyses that were based on averages across all 14 days of the measurement period in each intern, a secondary goal was to investigate how sleep and alertness were affected by the different shift types in flexible and standard programs, with a focus on extended overnight shifts.

METHODS

TRIAL PROCEDURES AND PARTICIPANTS

From the 63 programs participating in the iCOMPARE parent trial, we selected 12 medium-to-large programs, 6 in the standard group and 6 in the flexible group of our trial. (See Section 3 in the Supplementary Appendix for a detailed description of the selection process and Table S2 for a comparison of selected and not-selected programs.) Interns in flexible programs were made aware of the flexible approach to duty hours before starting residency. At each program, coordinators recruited interns who were scheduled for rotations in general medicine, cardiology, or critical care units between November 5, 2015, and May 31, 2016; at programs randomly assigned to flexible rules, coordinators recruited interns who were scheduled to be on rotations that used the more flexible duty-hour standards.

After obtaining written informed consent, coordinators scheduled interns for a single 14-day measurement period, commencing on a Monday, during which the intern underwent continuous sleep–wake measurement by means of actigraphy (a wristwatch-like accelerometer; model WGT3X-BT, ActiGraph) and completed a brief survey followed by a 3-minute Psychomotor Vigilance Test (PVT-B)²² on a trial-issued smartphone every morning. Actigraphy is a well-established method for sleep–wake measurement²³ that has been used to study sleep patterns in physicians.^{13,24} Actigraphs continuously recorded wrist movements at a sample rate of 30 Hz and ambient-light intensity at a sample rate of 1 Hz. Interns were instructed to wear their actigraph continuously except during activities that might damage the device (e.g., swimming and contact sports), during situations that would disrupt the delivery of medical care, and for any reason they felt the device should not or could not be worn. They were asked to remove the actigraph for up to 2 waking hours for recharging on days 1 and 7. Interns were expected to continuously wear the actigraph for 13 measurement periods of 24 hours each between 9 p.m. on day 1 and 9 p.m. on day 14.

Each day between 6 a.m. and 9 a.m., interns were asked to complete a brief survey on the smartphone that included a question on the shift that the intern was working, a sleep log (in which they recorded sleep periods during the past 24 hours), a score for sleep quality (on a 5-point scale, from 1 [bad] to 5 [good]), a question on the experience of periods of excessive sleepiness during the past 24 hours (with instructions to check all that apply: none, 12 a.m. to 6 a.m., 6 a.m. to 12 p.m., 12 p.m. to 6 p.m., and 6 p.m. to 12 a.m.), and the score on the Karolinska Sleepiness Scale²⁵ (9-point scale, with anchors 1 [extremely alert], 3 [alert], 5 [neither alert nor sleepy], 7 [sleepy, but no effort to stay awake], and 9 [extremely sleepy, fighting sleep]).

Interns next completed the PVT-B on the smart-phone, after which they could leave comments for the trial team.²² The PVT-B is a validated measure of the stability of vigilant attention that is based on simple reaction time to visual stimuli occurring at random intervals; sustained attention is a prerequisite for more complex cognitive tasks. Extensive evidence supports the sensitivity of the PVT-B to the neurobehavioral effects of acute sleep loss (i.e., extending the wake period beyond the typical 16 hours), chronic sleep loss (i.e., consecutive nights of insufficient sleep), and circadian misalignment.²⁶ Figure S1 in the Supplementary Appendix shows screenshots of the survey and the PVT-B.

Actigraphy, survey, and PVT-B data were automatically uploaded to a secure server and checked daily for protocol adherence and problems. If problems with data acquisition were reported or no data were received, interns were contacted by the trial team. At the end of their rotation, each intern received a gift card worth \$10 for each of the days that data were received. The protocol (available at [NEJM.org](https://www.nejm.org)) was approved by the institutional review board at the University of Pennsylvania. All the authors vouch for the accuracy and completeness of the data and analyses and for the fidelity of the trial to the protocol.

DATA

Actigraphy data were classified in 1-minute epochs as wake, sleep, or unknown (see Section 5 in the Supplementary Appendix), and PVT-B data were inspected and classified as generated in a manner that was adherent with instructions, probably nonadherent, or nonadherent as judged by sleep experts who were not aware of the trial-group assignments (see Section 6 in the Supplementary Appendix). Comments that were left by interns were inspected for distractions and non-fatigue-related impairment (see Section 7 in the Supplementary Appendix).

STATISTICAL ANALYSIS

The primary sleep and alertness outcomes were the average sleep time per 24 hours (noninferiority margin, -0.5 hours), the average score on the Karolinska Sleepiness Scale (noninferiority margin, 1 point on the 9-point scale), and the average number of PVT-B performance lapses (defined as reaction times of >355 msec; a higher number of lapses indicates lower levels of alertness; noninferiority margin, 1 additional lapse), each calculated as the average across trial days for each intern. For epochs with an unknown sleep-wake state (mean, 0.76 days per intern in the flexible group and 0.64 days per intern in the standard group, out of 13 expected days), we used single imputation with stratification according to program (standard or flexible), shift type reported by the intern (e.g., day, night, or off), and time of day (1440 periods of 1 minute each) (see Section 8 in the Supplementary Appendix). For the other outcomes, missing data were not imputed. Linear mixed-effects models with random program intercepts were used for noninferiority analyses.

The noninferiority margin of -0.5 hours for sleep duration was informed by a consensus statement of the American Academy of Sleep Medicine and the Sleep Research Society that “adults should sleep 7 or more hours per night on a regular basis to promote optimal health,”^{27,28} similar risks for several outcomes in the sleep-duration categories of 6 to 7 hours and 7 to 8 hours,²⁸ and an expected sleep duration in standard programs of

approximately 7 hours.¹⁹ Noninferiority margins for the PVT-B performance lapses and the score on the Karolinska Sleepiness Scale were informed by studies on chronic sleep restriction^{29,30} and reflect the smallest meaningful changes for these outcomes.

The calculated sample size (with the assumption of 90% power, a one-sided type I error of 0.05, and a noninferiority margin of 0.5 hours of sleep time) was 290 interns. The recruitment goal was increased to 384 interns to mitigate any data loss related to nonadherence and dropouts.

We report P values and one-sided 95% confidence intervals unadjusted for multiple comparisons for our three primary comparisons. A 95% confidence interval that does not overlap with the noninferiority margin indicates noninferiority at a P value of less than 0.05. A plan for adjustment for multiple comparisons was not pre-specified in the protocol; we conducted post hoc adjustment using the Benjamini–Hochberg method³¹ to account for the multiple comparisons entailed by our use of three primary outcomes and report the effects of those adjustments on our results.

Sensitivity analyses (not prespecified in the protocol) with adjustment for intern age and sex were also performed, because both factors affect sleep duration and alertness.^{32,33} For PVT-B data only, additional sensitivity analyses were performed on the subgroup of interns classified to be adherent and the subgroup of test periods in which interns did not report distraction or non-fatigue-related impairment. For analyses stratified according to shift type, data collected on shifts of the same type were first averaged for each intern (see Section 9 in the Supplementary Appendix). Hierarchical mixed-effects models with random intercepts for programs and interns (clustered within programs) were fit for analyses according to shift type, and least-square means and their differences were calculated for the different shifts. All analyses were conducted with SAS software, version 9.4 (SAS Institute).

RESULTS

PROGRAM CHARACTERISTICS

Data from 205 interns in flexible programs (14 to 53 per program) and 193 interns in standard programs (27 to 37 per program) were available for analysis (Fig. S8 in the Supplementary Appendix). A total of 97% of interns in flexible programs reported at least one extended overnight shift. Flexible and standard programs did not differ significantly with respect to demographic characteristics or completeness of actigraphy or PVT-B data (Table S4 in the Supplementary Appendix).

NONINFERIORITY ANALYSES

Across shifts, the average sleep time per 24 hours was 6.85 hours (95% confidence interval [CI], 6.61 to 7.10) among interns in flexible programs and 7.03 hours (95% CI, 6.78 to 7.27) among those in standard programs. Sleep duration in flexible programs was noninferior to that in standard programs (between-group difference, -0.17 hours per 24 hours; one-sided lower limit of the 95% confidence interval, -0.45 hours; non-inferiority margin, -0.5 hours; $P = 0.02$ for non-inferiority) (Fig. 1A).

Across shifts, average scores on the Karolinska Sleepiness Scale were similar in flexible programs (4.8 points; 95% CI, 4.7 to 5.0) and standard programs (4.7 points; 95% CI, 4.6 to 4.9), and scores in flexible programs were noninferior to those in standard programs (between-group difference, 0.12 points; one-sided upper limit of the 95% confidence interval, 0.31 points; non-inferiority margin, 1 point; $P < 0.001$ for noninferiority) (Fig. 1B). Across shifts, the average number of PVT-B performance lapses was 5.3 (95% CI, 3.7 to 7.0) among interns in flexible programs and 5.7 (95% CI, 4.1 to 7.3) among those in standard programs. Results for alertness according to the PVT-B did not meet the criteria for noninferiority (between-group difference, -0.3 lapses; one-sided upper limit of the 95% confidence interval, 1.6 lapses; noninferiority margin, 1 lapse; $P = 0.10$ for noninferiority) (Fig. 1C).

Sleep duration and the score on the Karolinska Sleepiness Scale in flexible programs remained noninferior to those in standard programs at an alpha level of 0.05 after Benjamini–Hochberg adjustments for multiple testing (three comparisons).³¹ None of the sensitivity analyses changed the conclusions based on the main analyses (Fig. S9 in the Supplementary Appendix). Flexible programs and standard programs were similar with respect to the score for sleep quality (3.6 in flexible programs and 3.6 in standard programs), percentage of days on which at least one period of excessive sleepiness was reported (61.5% and 55.2%), percentage of days with a high score (8 or 9) on the Karolinska Sleepiness Scale (12.1% and 8.3%), and percentage of days on which the sleep duration was less than 7 hours (49.1% and 53.5%) or less than 6 hours (28.4% and 22.0%) (Table S5 in the Supplementary Appendix).

ANALYSES STRATIFIED ACCORDING TO SHIFT TYPE

Despite similarities across shifts, the sleep duration per 24 hours varied considerably among different shift types within flexible and standard programs (Table 1; see Fig. 2 for a schematic of the different shift types). Although interns in flexible programs obtained some sleep during and after extended overnight shifts (Fig. 3A), their sleep duration per 24 hours was shortest (5.12 hours) relative to all other shifts in flexible programs. The number of PVT-B performance lapses (7.8), the score on the Karolinska Sleepiness Scale (6.8), the percentage of days with a high score (8 or 9) on the Karolinska Sleepiness Scale (38.6%), and the percentage of days on which at least one episode of excessive sleepiness was reported (87.7%) were also maximal on mornings at the end of extended overnight shifts (Table 1, and Table S5 in the Supplementary Appendix). Excessive sleepiness was most often reported for the period 12 a.m. to 6 a.m. (63.2% of days; 6 a.m. to 12 p.m., 33.1%; 12 p.m. to 6 p.m., 13.0%; and 6 p.m. to 12 a.m., 21.2%). The quality of sleep during extended overnight shifts was lower than that of sleep before day shifts (score, 2.4 vs. 3.7 on a 5-point scale from bad [1] to good [5]) (Table S5 in the Supplementary Appendix).

Interns in flexible programs slept longer than those in standard programs on nights before and during the first day of extended overnight shifts (7.67 hours vs. 6.74 hours) (Fig. 3B). Point estimates also indicated that interns in flexible programs had more sleep than those in standard programs on nights before and during regular day shifts (6.89 hours vs. 6.74 hours) (Fig. 3C) and on nights before and during days off (9.05 hours vs. 8.81 hours) (Fig. 3D).

Interns in flexible programs, as compared with those in standard programs, had a lower percentage of days with a sleep duration of less than 7 hours (49.1% vs. 53.5%) and a higher percentage of days with a sleep duration of less than 6 hours (28.4% vs. 22.0%) (Table S5 in the Supplementary Appendix). Interns in standard programs had 2.23 more hours of sleep during night-shift days than interns in flexible programs during extended overnight shifts (7.35 hours vs. 5.12 hours) (Fig. 3A). Although the two groups had similar amounts of sleep during the night, interns in standard programs retired much earlier for a longer sleep period during the day after the night shift. Unexpectedly, interns in standard programs also slept longer on days when they worked night shifts than on days when they worked day shifts (7.35 hours vs. 6.74 hours), probably owing to the extra sleep while they were on night shifts (Fig. S10 in the Supplementary Appendix).

DISCUSSION

The results of our trial support the noninferiority of flexible duty-hour rules to standard rules with respect to sleep time per 24 hours across shifts and interns' assessments of their morning sleepiness. Noninferiority of the flexible rules with respect to objectively assessed alertness was not established according to the prespecified noninferiority margin.

Despite losing more than 1.5 hours of sleep per 24 hours on extended overnight shifts as compared with regular day shifts, interns in flexible programs averaged only 10 minutes less sleep per day across the 14-day trial period than interns in standard programs. A behavioral homeostatic response of sleeping when possible was reflected in the finding that interns in flexible programs slept longer than those in standard programs on nights before day shifts, before extended overnight shifts, and before days off. This sleep regimen meant that interns in flexible programs had a lower percentage of days with a sleep duration of less than 7 hours. The fact that all interns were confined by a weekly maximum of 80 hours of work, a minimum of 1 day off every 7 days, and in-house calls no more frequent than every third night probably contributed to the noninferiority of average daily sleep times in flexible programs, because these work-schedule restrictions afforded strategic sleep opportunities before and after shifts. Average sleepiness scores across shifts were in the normal range for an employed population,³⁴ but interns in both trial groups reported at least one period of excessive sleepiness on more than 50% of the trial days.

Analyses that were stratified according to shift type confirmed previous concerns about interns' reduced sleep and alertness during and after extended overnight shifts. The sleep duration was shorter on days when interns finished an overnight shift than on days when they worked day shifts, with lower levels of alertness, higher scores on the Karolinska Sleepiness Scale, and more frequent reports of excessive sleepiness, especially between 12 a.m. and 6 a.m. Countermeasures for reduced alertness during extended overnight shifts include naps; both 5-hour and 3-hour protected sleep periods are feasible during extended overnight shifts and are associated with increased sleep duration and alertness, the latter without requiring additional resident cross-coverage.^{35,36} Sleep inertia, a period of reduced alertness and performance deficits after being awakened from naps, can pose a problem for performance shortly after awakening^{13,37} but may be mitigated by caffeine intake before or after the sleep.³⁸

There were only a few instances when interns in standard programs transitioned from a day shift to a night shift during the data-collection period. The trial was not powered to determine whether sleep and alertness were affected during the first days of this transition period. However, a comparison of night shifts with day shifts in the standard programs across all days of a shift showed no signs of acute or chronic sleep loss, because both alertness and sleepiness were similar.

Strengths of this trial include the sample size, objective measurements of sleep (with actigraphy) and alertness (with the PVT-B), and a rigorous blinding through all stages of data collection and analysis. Limitations include its lack of generalizability to interns in smaller internal-medicine programs and in different specialties, although it is unlikely that sleep patterns relative to call cycles and the physiological relationship between sleep loss and alertness would differ for interns in different programs.

In conclusion, intern sleep duration and sleepiness in programs that followed flexible duty-hour rules were noninferior to those in programs that followed standard rules, whereas results for alertness did not meet prespecified noninferiority criteria. Interns in flexible programs were able to compensate for the sleep lost during extended overnight shifts by increasing sleep duration on nights before day shifts, night shifts, and days off, findings that suggest a role for fatigue-management training in residency. At the same time, this trial confirms that acute sleep loss and circadian misalignment reduce intern alertness during extended overnight shifts.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Supported by grants (U01HL125388, to Dr. Asch; and U01HL126088, to Dr. Tonascia) from the National Heart, Lung, and Blood Institute and grants from the American Council for Graduate Medical Education (to Drs. Desai, Shea, and Silber).

We thank the participating program directors, along with Amanda K. Bertram, M.S., of Johns Hopkins University; Kelsey A. Gangemi, M.P.H., of the University of Pennsylvania; and Thomas Nasca, M.D., of the Accreditation Council for Graduate Medical Education. We also thank Manqing Liu, M.H.S. (University of Pennsylvania), for preparing Table S2 in the Supplementary Appendix.

REFERENCES

1. Basner M, Rao H, Goel N, Dinges DF. Sleep deprivation and neurobehavioral dynamics. *Curr Opin Neurobiol* 2013; 23: 854–63. [PubMed: 23523374]
2. Banks S, Dinges DF. Behavioral and physiological consequences of sleep restriction. *J Clin Sleep Med* 2007; 3: 519–28. [PubMed: 17803017]
3. Barger LK, Cade BE, Ayas NT, et al. Extended work shifts and the risk of motor vehicle crashes among interns. *N Engl J Med* 2005; 352: 125–34. [PubMed: 15647575]
4. Barger LK, Ayas NT, Cade BE, et al. Impact of extended-duration shifts on medical errors, adverse events, and attentional failures. *PLoS Med* 2006; 3(12): e487. [PubMed: 17194188]
5. Ayas NT, Barger LK, Cade BE, et al. Extended work duration and the risk of self-reported percutaneous injuries in interns. *JAMA* 2006; 296: 1055–62. [PubMed: 16954484]

6. Lockley SW, Landrigan CP, Barger LK, Czeisler CA. When policy meets physiology: the challenge of reducing resident work hours. *Clin Orthop Relat Res* 2006; 449: 116–27. [PubMed: 16770285]
7. Nasca TJ, Day SH, Amis ES Jr. The new recommendations on duty hours from the ACGME Task Force. *N Engl J Med* 2010; 363(2): e3(1)–e3(6). [PubMed: 20573917]
8. Volpp KG, Rosen AK, Rosenbaum PR, et al. Mortality among patients in VA hospitals in the first 2 years following ACGME resident duty hour reform. *JAMA* 2007; 298: 984–92. [PubMed: 17785643]
9. Patel MS, Volpp KG, Small DS, et al. Association of the 2011 ACGME resident duty hour reforms with mortality and re-admissions among hospitalized Medicare patients. *JAMA* 2014; 312: 2364–73. [PubMed: 25490327]
10. Rajaram R, Chung JW, Jones AT, et al. Association of the 2011 ACGME resident duty hour reform with general surgery patient outcomes and with resident examination performance. *JAMA* 2014; 312: 2374–84. [PubMed: 25490328]
11. Bilimoria KY, Chung JW, Hedges LV, et al. National cluster-randomized trial of duty-hour flexibility in surgical training. *N Engl J Med* 2016; 374: 713–27. [PubMed: 26836220]
12. Bilimoria KY, Chung JW, Hedges LV, et al. Development of the Flexibility in Duty Hour Requirements for Surgical Trainees (FIRST) trial protocol: a national cluster-randomized trial of resident duty hour policies. *JAMA Surg* 2016; 151: 273–81. [PubMed: 26720622]
13. Basner M, Dinges DF, Shea JA, et al. Sleep and alertness in medical interns and residents: an observational study on the role of extended shifts. *Sleep* 2017; 40(4): zsx027.
14. Barger LK, Lockley SW, Rajaratnam SM, Landrigan CP. Neurobehavioral, health, and safety consequences associated with shift work in safety-sensitive professions. *Curr Neurol Neurosci Rep* 2009; 9: 155–64. [PubMed: 19268039]
15. Desai SV, Feldman L, Brown L, et al. Effect of the 2011 vs 2003 duty hour regulation-compliant models on sleep duration, trainee education, and continuity of patient care among internal medicine house staff: a randomized trial. *JAMA Intern Med* 2013; 173: 649–55. [PubMed: 23529771]
16. Levine AC, Adusumilli J, Landrigan CP. Effects of reducing or eliminating resident work shifts over 16 hours: a systematic review. *Sleep* 2010; 33: 1043–53. [PubMed: 20815185]
17. Spellberg B, Sue D, Chang D, Witt M. Change in intern calls at night after a work hour restriction process change. *JAMA Intern Med* 2013; 173: 707–9. [PubMed: 23529748]
18. Reed DA, Fletcher KE, Arora VM. Systematic review: association of shift length, protected sleep time, and night float with patient care, residents' health, and education. *Ann Intern Med* 2010; 153: 829–42. [PubMed: 21173417]
19. Shea JA, Silber JH, Desai SV, 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(9): e021711.
20. Silber JH, Bellini LM, Shea JA, et al. Patient safety outcomes under flexible and standard resident duty-hour rules. *N Engl J Med* 2019; 380: 905–14. [PubMed: 30855740]
21. Desai SV, Asch DA, Bellini LM, et al. Education outcomes in a duty-hour flexibility trial in internal medicine. *N Engl J Med* 2018; 378: 1494–508. [PubMed: 29557719]
22. Basner M, Mollicone D, Dinges DF. Validity and sensitivity of a brief Psycho-motor Vigilance Test (PVT-B) to total and partial sleep deprivation. *Acta Astronaut* 2011; 69: 949–59. [PubMed: 22025811]
23. Ancoli-Israel S, Cole R, Alessi C, Chambers M, Moorcroft W, Pollak CP. The role of actigraphy in the study of sleep and circadian rhythms. *Sleep* 2003; 26: 342–92. [PubMed: 12749557]
24. Malmberg B, Kecklund G, Karlson B, Persson R, Flisberg P, Ørbaek P. Sleep and recovery in physicians on night call: a longitudinal field study. *BMC Health Serv Res* 2010; 10: 239. [PubMed: 20712854]
25. Akerstedt T, Gillberg M. Subjective and objective sleepiness in the active individual. *Int J Neurosci* 1990; 52: 29–37. [PubMed: 2265922]
26. Basner M, Dinges DF. Maximizing sensitivity of the psychomotor vigilance test (PVT) to sleep loss. *Sleep* 2011; 34: 581–91. [PubMed: 21532951]

27. Watson NF, Badr MS, Belenky G, et al. Recommended amount of sleep for a healthy adult: a joint consensus statement of the American Academy of Sleep Medicine and Sleep Research Society. *Sleep* 2015; 38: 843–4. [PubMed: 26039963]
28. Watson NF, Badr MS, Belenky G, et al. Joint consensus statement of the American Academy of Sleep Medicine and Sleep Research Society on the recommended amount of sleep for a healthy adult: methodology and discussion. *Sleep* 2015; 38: 1161–83. [PubMed: 26194576]
29. Van Dongen HP, Maislin G, Mullington JM, Dinges DF. The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. *Sleep* 2003; 26: 117–26. [PubMed: 12683469]
30. Belenky G, Wesensten NJ, Thorne DR, et al. Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: a sleep dose-response study. *J Sleep Res* 2003; 12: 1–12. [PubMed: 12603781]
31. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc [B]* 1995; 57: 289–300.
32. Basner M, Spaeth AM, Dinges DF. Sociodemographic characteristics and waking activities and their role in the timing and duration of sleep. *Sleep* 2014; 37: 1889–906. [PubMed: 25325472]
33. Moore TM, Basner M, Nasrini J, et al. Validation of the Cognition Test Battery for spaceflight in a sample of highly educated adults. *Aerosp Med Hum Perform* 2017; 88: 937–46. [PubMed: 28923143]
34. Åkerstedt T, Hallvig D, Kecklund G. Normative data on the diurnal pattern of the Karolinska Sleepiness Scale ratings and its relation to age, sex, work, stress, sleep quality and sickness absence/illness in a large sample of daytime workers. *J Sleep Res* 2017; 26: 559–66. [PubMed: 28370590]
35. Volpp KG, Shea JA, Small DS, et al. Effect of a protected sleep period on hours slept during extended overnight in-hospital duty hours among medical interns: a randomized trial. *JAMA* 2012; 308: 2208–17. [PubMed: 23212498]
36. Shea JA, Dinges DF, Small DS, et al. A randomized trial of a three-hour protected nap period in a medicine training program: sleep, alertness, and patient outcomes. *Acad Med* 2014; 89: 452–9. [PubMed: 24448046]
37. Tassi P, Muzet A. Sleep inertia. *Sleep Med Rev* 2000; 4: 341–53. [PubMed: 12531174]
38. Van Dongen HP, Price NJ, Mullington JM, Szuba MP, Kapoor SC, Dinges DF. Caffeine eliminates psychomotor vigilance deficits from sleep inertia. *Sleep* 2001; 24: 813–9. [PubMed: 11683484]

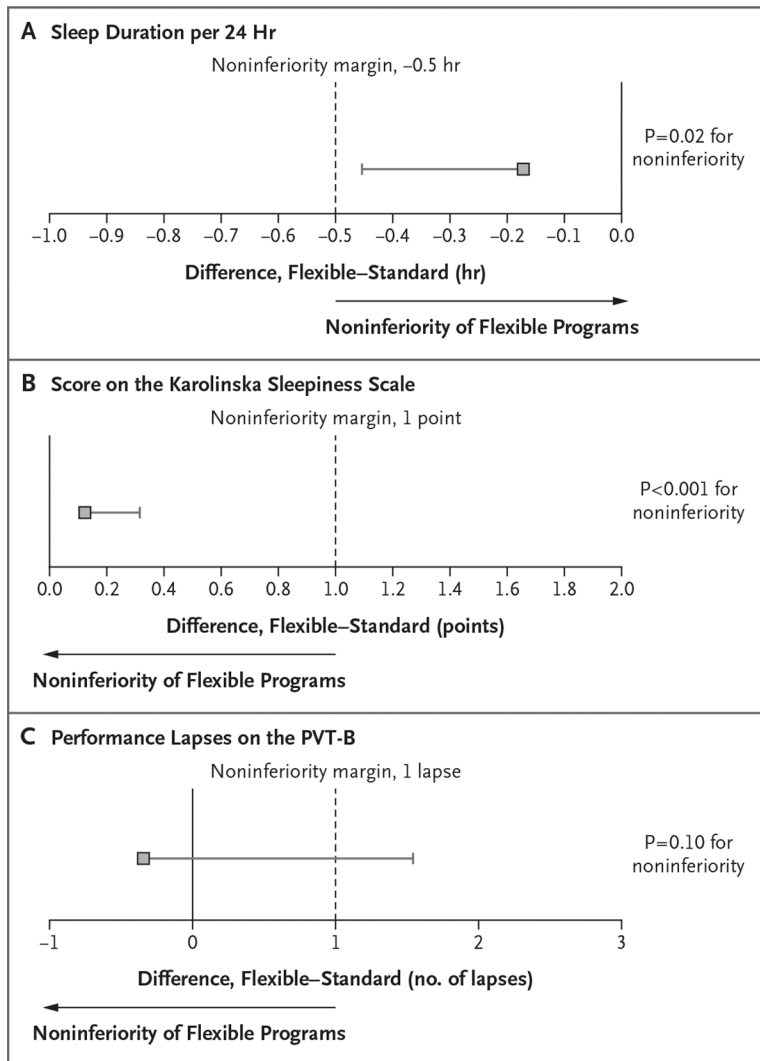


Figure 1. Noninferiority Analyses for Sleep Duration, Sleepiness, and Alertness. Shown are unadjusted 95% confidence intervals for the difference between interns in flexible programs and those in standard programs, with respect to average sleep duration per 24 hours (Panel A), average score on the Karolinska Sleepiness Scale (Panel B), and average number of performance lapses on the brief Psychomotor Vigilance Test (PVT-B) (Panel C). Scores on the Karolinska Sleepiness scale range from 1 (extremely alert) to 9 (extremely sleepy, fighting sleep); performance lapses on the PVT-B were defined as reaction times of more than 355 msec, with a higher number of lapses indicating lower levels of alertness. Noninferiority tests were one-sided, with noninferiority margins of -0.5 hour, 1 point on the Karolinska Sleepiness Scale, and 1 additional lapse on the PVT-B, respectively. Analyses indicate that the sleep duration per 24 hours and subjective ratings of sleepiness in flexible programs were noninferior to those in standard programs, whereas noninferiority was not established for objectively assessed alertness according to the PVT-B. Sleep duration and the score on the Karolinska Sleepiness Scale in flexible programs remained noninferior to those in standard programs at an alpha level of 0.05 after Benjamini-Hochberg adjustments³¹ for multiple testing (three comparisons). The 95% confidence intervals in the figure have not

been adjusted for multiple comparisons, and inferences drawn from these intervals may not be reproducible.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

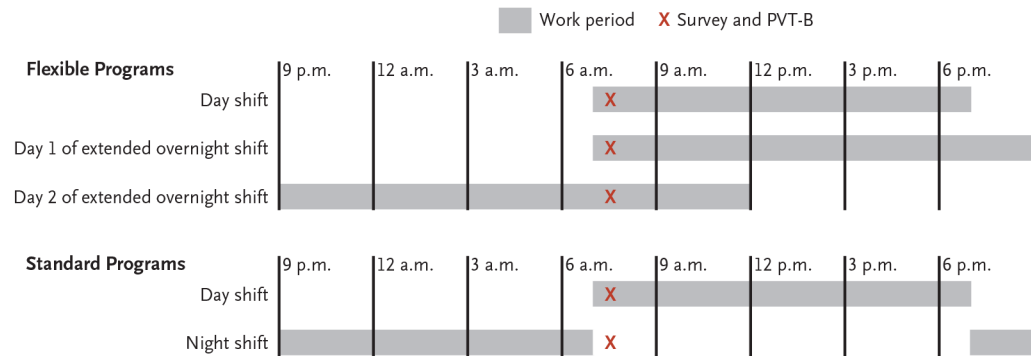


Figure 2. Different Shift Types in Standard and Flexible Programs.

Gray bars depict typical work periods for the different shift types in standard and flexible programs. Interns were instructed to take the survey and PVT-B between 6 a.m. and 9 a.m. every day. An extended overnight shift in flexible programs spanned across 2 days. The work period started in the morning of day 1 and concluded at approximately noon on day 2. In this shift, the survey and PVT-B were administered once at the beginning of the shift on day 1 and a second time 24 hours later close to the end of the shift on day 2. The actual placement and duration of shifts in standard and flexible programs may have differed for individual programs.

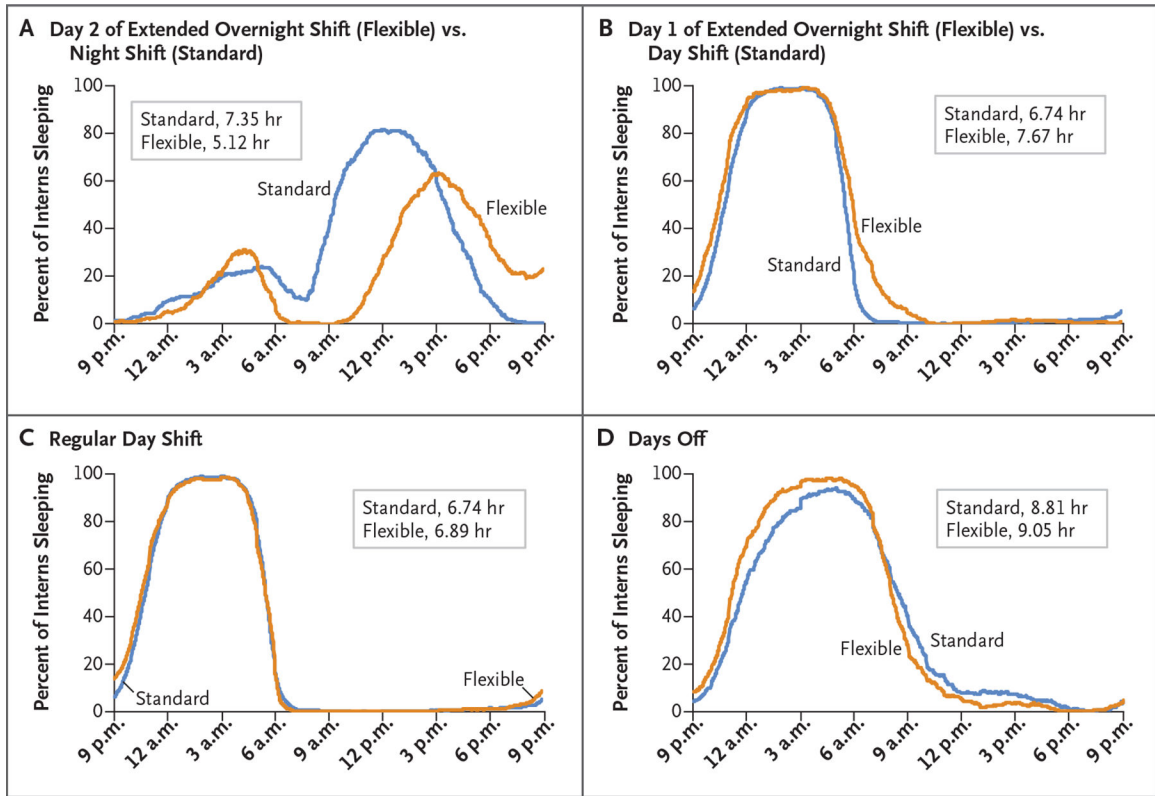


Figure 3. Percent of Interns Sleeping According to Time of Day.

The sleep duration of interns in flexible programs was more than 2 hours shorter on the second day of the extended overnight shift than that during the regular night shifts of interns in standard programs (Panel A). Interns in flexible programs slept longer before extended overnight shifts (Panel B), before day shifts (Panel C), and before days off (Panel D) than interns in standard programs.

Table 1. Sleep Duration, Sleepiness, and Alertness among Interns, According to Shift Type and Duty-Hour Group.*

Shift Type	Mean Hours of Sleep (95% CI)	Mean Score on Karolinska Sleepiness Scale (95% CI) [†]	Mean No. of Performance Lapses on PVT-B (95% CI) [‡]
Flexible programs			
Day shift	6.89 (6.61–7.16)	4.5 (4.2–4.7)	4.9 (3.3–6.4)
Day 1 of extended overnight shift	7.67 (7.39–7.95)	4.3 (4.0–4.5)	4.7 (3.1–6.2)
Day 2 of extended overnight shift	5.12 (4.84–5.40)	6.8 (6.5–7.0)	7.8 (6.3–9.4)
Day off	9.05 (8.77–9.33)	4.2 (3.9–4.4)	3.9 (2.4–5.5)
Other [§]	6.11 (5.79–6.43)	4.8 (4.2–5.4)	6.6 (4.6–8.6)
Across shifts	6.85 (6.61–7.10)	4.8 (4.7–5.0)	5.3 (3.7–7.0)
Standard programs			
Day shift	6.74 (6.46–7.02)	4.7 (4.5–5.0)	5.8 (4.2–7.3)
Night shift	7.35 (6.87–7.83)	4.9 (4.5–5.3)	5.6 (3.9–7.3)
Day off	8.81 (8.52–9.10)	4.0 (3.8–4.3)	4.4 (2.9–5.9)
Other [§]	6.78 (6.45–7.11)	4.9 (4.6–5.3)	6.3 (4.6–7.9)
Across shifts	7.03 (6.78–7.27)	4.7 (4.6–4.9)	5.7 (4.1–7.3)

* Confidence intervals have not been adjusted for multiple comparisons, and inferences drawn from the intervals may not be reproducible.

[†] Scores range from 1 (extremely alert) to 9 (extremely sleepy, fighting sleep).

[‡] Performance lapses on the brief Psychomotor Vigilance Test (PVT-B) were defined as reaction times of more than 355 msec. A higher number of lapses indicates lower levels of alertness.

[§] In flexible programs, days that had missing shift information or were classified by the interns as having a regular night shift were reclassified as “other” (see Section 8 in the Supplementary Appendix). In standard programs, days that had missing shift information or were classified by the interns as starting or finishing an extended overnight shift were reclassified as “other.”