In 2003, the Accreditation Council for Graduate Medical Education (ACGME) implemented a single duty-hour standard nationwide. The evidence to date suggests that this neither improved nor worsened patient outcomes. In June 2010, the ACGME proposed a new set of duty-hour standards for implementation in July 2011. The main disadvantage of this approach is that there is no ability to determine whether different standards would have worked better to reduce resident fatigue while improving patient safety. Many unanswered questions remain about how to design duty-hour standards, but relatively little evidence exists. In addition, the same approach may not work in all specialties and all hospitals. A more flexible, dynamic policy that emphasizes ongoing testing and evaluation would be more likely to achieve improvements in clinical and educational outcomes.

In 2003, the Accreditation Council for Graduate Medical Education (ACGME) adopted duty-hour standards for all specialties. Five years later, the Institute of Medicine (IOM) convened a Committee on Optimizing Graduate Medical Trainee (Resident) Hours and Work Schedules to Improve Patient Safety, whose report indicated that there was a paucity of data about how to optimize duty hours for physicians in training (1). Several studies by our research team and others evaluated the effect of the 2003 duty-hour rules on patient outcomes, including deaths, readmissions, prolonged hospital stays, and complications, and found no consistent evidence of statistically significant improvement or worsening in outcomes (2–7). However, these studies could retrospectively evaluate only the single set of duty-hour standards that was implemented nationwide. These studies also could not address the inherent variability in training methods and implementation challenges across specialties and types of programs.

At the request of Congress in 2007, the IOM requested the Committee on Optimizing Graduate Medical Trainee (Resident) Hours and Work Schedules to Improve Patient Safety to “evaluate current evidence on the topic and to develop strategies to optimize work schedules . . . while maintaining the necessary educational experience to ensure long-term patient safety after trainees are on their own” (1). The committee report offered a well-balanced and thoughtful assessment of the status quo and suggested rapid implementation of its recommendations. Although strong evidence from laboratory studies links sleep deprivation to impaired cognitive performance and epidemiologic evidence links postcall fatigue to vehicular crashes and occupational injuries (8, 9), the IOM report highlighted continuing uncertainty about how to translate these findings into health care delivery and training settings. For example, the centerpiece of the IOM committee recommendations—that residents on extended-duty overnight shifts be given a “mandatory nap period” of 5 hours or more—has been the subject of very limited testing and has never been shown to be feasible or effective in health care settings (10, 11). The benefits of shorter shifts, in terms of decreased fatigue, may or may not outweigh the risks of more handoffs and cross-coverage, even in light of the recent research on how to make these handoffs safer (12, 13). Finally, reducing work hours may compromise the resident training experience, leading to less well-trained physicians or an increase in the duration of training. Ericsson (14) has described the importance of “deliberate practice” to improve professional performance; this concept entails giving trainees tasks with well-defined goals, providing motivation to improve, offering timely feedback, and ensuring ample repetition and reinforcement.

In June 2010, the ACGME proposed new duty-hour standards for implementation in July 2011 (15). Although the proposed ACGME recommendations steer clear of mandating naps, intern shifts are limited to a total of 16 hours. This recommendation is largely based on evidence from a single trial that indicated that eliminating 24-hour shifts improved patient safety in an intensive care unit (16). The generalizability of these findings is uncertain because the hazards of discontinuity of care were probably mitigated by a nurse-to-patient ratio of 1:1 or 1:2 in the intensive care environment, whereas the nurse-to-patient ratio may be as high as 1:8 on regular medical or surgical units. In the latter setting, the detrimental effects of handoffs and cross-coverage (17) may outweigh the benefits of reduced fatigue among housestaff. It is also unknown whether 16-hour shifts would affect quality of care differentially across specialties, perhaps due to differences in the temporal distribution of admissions and discharges and the nature of the tasks required (that is, cognitive vs. procedural). Finally, it is noteworthy that the innovative work schedule at Brigham and Women’s Hospital, Boston, Massachusetts, was deliberately discontinued when the experiment ended, so its long-term efficacy cannot be determined (1).

Our purpose is not to critique the choice of 16-hour shifts for interns. Although we acknowledge the difficulty in defining duty-hour standards for the entire profession,
we are concerned that a similar one-size-fits-all approach, as in 2003, is being taken without adequate consideration of alternatives that might achieve similar objectives. If one set of duty-hour standards is implemented nationwide with no testing of alternative approaches, then 5 years from now we will be able to assess only whether patient outcomes changed after implementation of the new standards. Any such comparison will be confounded by contemporaneous changes in clinical practice that affect teaching hospitals. We will not know whether the new duty-hour standards were better than potential alternatives at improving patient outcomes, resident training, and quality of life. As we have previously argued, not enough is known for any single alternative to be universally embraced as the “optimal” approach to duty-hour reduction in all settings (18). It is important that any changes be critically assessed, with an emphasis on designing interventions to allow careful evaluation of their relative costs and benefits.

We fully accept the principles set forth by the ACGME’s task force: Patients must be safe and receive excellent care; outstanding education needs to be delivered today to ensure that trainees are capable of providing unsupervised care in the future; and a humanistic educational environment should nurture professionalism and the effacement of self-interest (19). However, these principles could be translated into standards in several ways while recognizing the risks of sleep deprivation and circadian misalignment and also attending to concerns about continuity, workload, and other factors that affect safety. For example, a recent survey of 429 program directors in medicine, pediatrics, and surgery reported that 56% strongly disagree and 23% moderately disagree with the proposal to limit duty periods for interns to 16 hours, with striking differences of opinion between surgical educators and educators in medicine and pediatrics (20). Therefore, a preferred alternative to the one-size-fits-all approach would be for the ACGME to offer a few acceptable alternatives for which there is suggestive evidence. Examples of potential alternatives could include 16-hour shifts for interns, which were shown to reduce errors in the medical intensive care setting but proved locally unsustainable (16); 5-hour mandatory naps on extended duty overnight shifts, similar to what was recommended in the IOM report (1); flexibility in the number of consecutive hours of duty, based on specialty- and program-specific workflows (that is, in some surgical specialties, emergency admissions are rare, teamwork is especially critical, and immediate perioperative care may extend longer than 16 hours); and no change from the present.

Each residency review committee could be given a choice of either adopting 16-hour shifts (if that is agreed on) or encouraging residency programs in its specialty to accept randomization to 1 of these alternatives for the next 3 years. Nonrandom allocation to alternative schemes for resident work-hour management could also be considered, although we would highly recommend some form of randomization. Nonrandom assignment would have the obvious disadvantage of producing weaker evidence on effectiveness. However, concerns about randomization should not be a barrier to providing the flexibility needed for training programs to test reasonable alternatives. The main advantage to a system that encourages the rigorous testing of alternatives would be that we will have more information 5 years from now about which alternative was optimal in which circumstances. Medicare and all-payer data could be used to evaluate risk-adjusted patient outcomes, with either programs or residency review committees required to collect standardized measures of case volume, educational outcomes, occupational hazards for residents, relevant patient experience, and resident quality of life.

Given the cost and very limited evidence of benefit of the work-hour rules that were adopted in 2003, we suggest that the medical profession seriously consider a plan that allows for systematic testing of several alternatives going forward. Schedule reform, just as any other therapeutic intervention, should be implemented widely only after robust pilot testing suggests that a particular plan is better than competing alternatives. Other approaches that specific residency review committees view as viable alternatives may exist and should be considered as well. Although the proposed ACGME standards were approved in September 2010 for planned implementation in July 2011, we are optimistic that this will be an iterative process with a robust dialogue involving all stakeholders, in which it will never be too late for good ideas to be considered. For example, the Association of American Medical Colleges has already expressed a similar interest in rigorous, multiinstitutional evaluation studies to provide an appropriate evidence base for evolving standards (21). The American Hospital Association has emphasized the importance of phasing in the implementation of any new standards to provide time for necessary planning and budgeting (22).

The proposed ACGME plan does not provide any incentives for needed innovation; indeed, we fear that its regulatory mandates will stifle innovation. The alternatives we suggest would not be any more expensive than what the ACGME has proposed, so additional incentives for accredited programs to participate would not be required, but resources would be needed to fund rigorous evaluation by third parties other than the ACGME. As the IOM recommended in 2008, “the ACGME should convene a meeting of stakeholders and potential funders to set priorities for research and evaluation . . . the Centers for Medicare and Medicaid Services, Agency for Healthcare Research and Quality, National Institutes of Health, Department of Defense, Department of Veterans Affairs, and other funders should support this work as a high priority.”

Five years ago, with the specter of federal legislation looming, the ACGME implemented duty-hour standards nationally without any data that the rules adopted would be likely to improve patient safety. The preponderance of the evidence suggests that the 2003 rules neither worsened nor improved quality of care and patient safety (2–7), even
as new evidence from related disciplines has reinforced concerns about the risk for burnout and the effects of fatigue on physician performance and safety. New approaches to the problem are therefore necessary. However, implementing a single approach nationally means that only 1 alternative can be evaluated. As a society, we will be better off 5 years from now if several different initiatives are tested and rigorously evaluated using the questions highlighted in the IOM report as a conceptual foundation. As in many domains of clinical practice, a more flexible, dynamic policy that emphasizes creativity, innovation, and ongoing evaluation seems like the path most likely to achieve substantial improvements in clinical and educational outcomes.

From Center for Health Equity Research and Promotion, Philadelphia Veterans Affairs Medical Center; Center for Health Incentives, Leonard Davis Institute of Health Economics, University of Pennsylvania; University of Pennsylvania School of Medicine; The Wharton School of the University of Pennsylvania; and Center for Outcomes Research, Children’s Hospital of Philadelphia, Philadelphia, Pennsylvania; University of Florida, Gainesville, Florida; Center for Healthcare Policy and Research, University of California, Davis, School of Medicine, Sacramento, California; Center for Organization, Leadership and Management Research, Veterans Affairs Boston Healthcare System, and Boston University School of Public Health, Boston, Massachusetts.

Grant Support: By grant R01 HL082637 from the National Heart, Lung, and Blood Institute.

References

Current Author Addresses: Dr. Volpp: 423 Guardian Drive, 1232 Blockley Hall, Philadelphia, PA 19104.
Dr. Friedman: University of Florida, PO Box 100265, MBI University of Florida, Gainesville, FL 32610.
Dr. Romano: Division of General Medicine, University of California, Davis, 4150 V Street, PSSB Suite 2400, Sacramento, CA 95817.
Dr. Rosen: Center for Health Quality, Outcomes and Research, Bedford Veterans Affairs Medical Centers, Building 70 (152), Bedford, MA 01730.
Dr. Silber: Center for Outcomes Research, The Children’s Hospital of Philadelphia, 3535 Market Street, Suite 1029, Philadelphia, PA 19104.

Author Contributions: Conception and design: K.G. Volpp, W. Friedman, P.S. Romano, A. Rosen, J.H. Silber.
Analysis and interpretation of the data: J.H. Silber.
Drafting of the article: K.G. Volpp, W. Friedman, P.S. Romano, A. Rosen, J.H. Silber.
Critical revision of the article for important intellectual content: K.G. Volpp, P.S. Romano, J.H. Silber.
Final approval of the article: K.G. Volpp, W. Friedman, P.S. Romano, A. Rosen, J.H. Silber.
Provision of study materials or patients: J.H. Silber.
Administrative, technical, or logistic support: J.H. Silber.
Collection and assembly of data: J.H. Silber.