Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: 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. DOI: 10.1056/NEJMoa1800965

Online Supplement

Education Outcomes from a Duty-Hour Flexibility Trial in Internal Medicine

Sanjay V. Desai, MD, David A. Asch, MD, MBA, Lisa M. Bellini, MD, Krisda H. Chaiyachati, MD, MPH, MSHP, Manqing Liu, MHS, Alice L. Sternberg, ScM, James Tonascia, PhD, Alyssa M. Yeager MD; Jeremy M. Asch BS, Joel T. Katz, MD, MS, Mathias Basner, MD, PhD, David W. Bates, MD, MSc, David F. Dinges PhD, Orit Even-Shoshan, MS, David M. Shade, JD, Jeffrey H Silber, MD, PhD, Dylan S. Small, PhD, Kevin G. Volpp, MD, PhD, Judy A. Shea, PhD for the iCOMPARE Research Group

Corresponding Author:
Sanjay V. Desai, MD
Johns Hopkins University School of Medicine
Baltimore, MD 21205
sanjayvdesai@jhmi.edu

Table of Contents **Pages**

iCOMPARE Research Group

Margot Boigon, MD, FACP	Abington Memorial Hospital Program
Jill Patton, DO	Advocate Lutheran General Hospital Program
Donna Astiz, MD	Atlantic Health (Morristown) Program
Cheryl O'Malley, MD	University of Arizona College of Medicine-Phoenix Internal Medicine Program
Richard J. Hamill, MD	Baylor College of Medicine Program
Michael Rosenblum, MD, FACP	Baystate Medical Center/Tufts University School of Medicine Program
Charles Smith, MD	Beth Israel Deaconess Medical Center
Joel Katz, MD	Brigham and Women's Hospital Program
Maria Yialamas, MD	Brigham and Women's Hospital Program
Dominick Tammaro , MD	Brown University Program
Jennifer Bolyard, MD	Canton Medical Education Foundation/NEOMED Program
Claudia A. Kroker-Bode, MD, PhD	Carilion Clinic-Virginia Tech Carilion School of Medicine Program
Michael J. McFarlane, MD	Case Western Reserve University (MetroHealth) Program
Keith Armitage, MD	Case Western Reserve University/University Hospitals Case Medical Center Program
Amanda Ewing, MD, FACP	Cedars-Sinai Medical Center Program
Abby L. Spencer, MD, MS, FACP	Cleveland Clinic
Tammy Wichman, MD	Creighton University School of Medicine
Richard Paluzzi, MD	Drexel University College of Medicine/Hahnemann University Hospital Program
Aimee Zaas, MD, MHS	Duke University Hospital Program
Suzanne Kraemer, MD	East Carolina University
Benjamin Goodman, MD	Eastern Virginia Medical School Program
Lorenzo Di Francesco, MD	Emory University Program
Mary Harris, MD	Geisinger Health System Program
Jillian Catalanotti, MD, MPH, FACP	George Washington University Program
Sal Pindiprolu, MD	Georgetown University Hospital/Washington Hospital Center Program
Paul Foster, MD, FACP	Greater Baltimore Medical Center Program
Sean M. Drake, MD	Henry Ford Hospital/Wayne State University Program
Sanjay Desai, MD	Johns Hopkins University Program
Erica N. Johnson, MD	Johns Hopkins University/Bayview Medical Center Program
Elizabeth Nilson, MD	Lahey Clinic Program
William D. Surkis, MD, FACP	Lankenau Medical Center
Jonathon Doroshow, MD, FACP	Lankenau Medical Center
Jatin M. Vyas, MD, PhD	Massachusetts General Hospital Program
Michael Frank, MD	Medical College of Wisconsin Affiliated Hospitals Program
Eric H. Green, MD, MSc	Mercy Catholic Medical Center Program
Cinnamon Bradley, MD	Morehouse School of Medicine Program
Soma Wali, MD	Olive View/UCLA Medical Center Program
Sapna Kuehl, MD	St. Agnes HealthCare Program
Harvey Friedman, MD, FCCP, FACP	St. Francis Hospital of Evanston Program
Ronald Witteles, MD	Stanford University Program
Marissa A. Blum, MD, MSHPR	Temple University Hospital Program

Curtis Mirkes, DO	Texas A&M College of Medicine-Scott and White Program
Michael Phy, DO	Texas Tech University (Lubbock) Program
Emily Stewart, MD, FACP	Thomas Jefferson University Program
Richard Kopelman, MD	Tufts Medical Center Program
Kari Roberts, MD	Tufts Medical Center Program
Jodi Friedman, MD	UCLA Medical Center Program
Brian Gable, MD	Robert Wood Johnson Medical School (Camden)/Cooper University Hospital Program
Eric Warm, MD	University Hospital/University of Cincinnati College of Medicine Program
Suzanne Brandenburg, MD	University of Colorado
Steven V. Angus, MD	University of Connecticut Program
Leigh M. Eck, MD	University of Kansas School of Medicine Program
Susan D. Wolfsthal, MD	University of Maryland Program
Richard M. Forster, MD, FACP	University of Massachusetts Program
James O'Dell, MD	University of Nebraska Medical Center College of Medicine Program
Debra L. Bynum, MD, MMEL	University of North Carolina Hospitals Program
Lisa Bellini, MD	University of Pennsylvania Program
Mark Pasanen, MD	University of Vermont/Fletcher Allen Health Care Program
Kenneth P. Steinberg, MD	University of Washington Program
Stephanie Ann Call, MD, MSPH	Virginia Commonwealth University Health System Program
Hal H. Atkinson, MD, MS	Wake Forest University School of Medicine Program
Melvin Blanchard, MD	Washington University/B-JH/SLCH Consortium Program
Thomas Ciesielski, MD	Washington University/B-JH/SLCH Consortium Program
Nathan Lerfald, MD	West Virginia University
Mark D. Siegel, MD	Yale - New Haven Medical Center Program

Additional details on trial sample size calculation

The trial's sample size calculation was based on the patient safety primary outcome of 30-day mortality; a 2-sample t-test for 2 independent groups was used. The primary outcome is obtained at the program level so program clustering is not an issue. The primary outcome is the difference in 30-day mortality rates, trial year rate for the program minus baseline year rate for the program, where trial year is the year of the intervention (July 2015-June 2016) and baseline year is the prior year (July 2014-June 2015). Assuming 30-day mortality of 11% in the standard group and pooled standard deviation of the paired mortality rate differences of 1.5%, non-inferiority margin of 1%, 80% power, and one-sided Type I error of 0.05, we calculated a sample size of 29 pairs of trial year vs. baseline year differences in each group (58 programs total). The 2-sample t-test compared the independent (29 calculated sample size, 32 achieved) flexible program's differences to the independent (29 calculated sample size, 31 achieved) standard programs' rate differences.

Incentives provided to promote trainee participation in iCOMPARE surveys

The nine programs with the highest response rates on the iCOMPARE end-of-year trainee survey in 2016 each received $2,500. Every two weeks, an intern and more senior resident in each program who had completed the end-of-shift survey were randomly selected to receive a $25 or $100 Amazon™ gift card.

Additional details on questions comprising content areas reported for ACGME trainee survey

The Duty Hours area includes items relating to: 80 hours per week, 1 day free in 7, in house call every 3rd night, night float duty no more than 6 nights.

The Educational Content area includes items relating to: provided goals and objectives for assignments, instructed how to manage fatigue, satisfied with opportunities for scholarly activities, appropriate balance for education, education compromised by service obligations, supervisors delegate appropriately, provided data about practice habits, see patients across a variety of settings.

The Evaluation (Assessment/Feedback) area includes items relating to: able to access evaluations, opportunity to evaluate faculty members, satisfied that evaluations of faculty are confidential, opportunity to evaluate program, satisfied that evaluations of program are confidential, satisfied that program uses evaluations to improve, satisfied with feedback after assignments.

The Faculty area includes items relating to: sufficient supervision, appropriate level of supervision, sufficient instruction, faculty and staff interested in residency education, and faculty and staff create environment of inquiry.

The Patient Safety and Teamwork area includes items relating to: culture reinforces patient safety responsibility, work in inter-professional teams, effectively work in inter-professional teams, tell patients of respective roles of faculty and residents, participated in quality improvement or patient safety activities, information lost during shift changes or patient transfers.

The Resources area includes items relating to: access to reference materials, use electronic medical records in hospital, use electronic medical records in ambulatory setting, electronic medical records integrated across settings, electronic medical records effective, provided a way to transition care when fatigued, satisfied with process to deal with problems and concerns, education compromised by other trainees, residents can raise concerns without fear.

The Overall Evaluation of Program is a single item content area.

Additional details on questions comprising content areas reported for ACGME faculty survey

The Supervision and Teaching area includes items relating to: sufficient time to supervise trainees, trainees seek supervisory guidance, interest of faculty and program director in education, evaluation after rotations and educational assignments, faulty performance evaluated.

The Educational Content area includes items relating to: worked on scholarly project with trainees, trainees see patients across a variety of settings, trainees receive education to manage fatigue, effectiveness of graduating trainees, milestone achievement of graduating trainees.

The Resources area includes items relating to: program provides a way for trainees to transition care when fatigued, trainees' workload exceeds capacity to do the work, satisfied with faculty development to supervise and educate residents, satisfied with process to deal with trainees' problems and concerns, prevent excessive reliance on trainees to provide clinical service.

The Patient Safety area includes items relating to: information lost during shift changes or patient transfers, tell patients of respective roles of faculty and trainees, culture reinforces responsibility of patient safety, trainees participate in quality improvement or patient safety activities.

The Teamwork content area includes items relating to: trainees communicate effectively when transferring clinical care, trainees effectively work in inter-professional teams, program effective in teaching teamwork skills.

The Overall Evaluation of Program is a single item content area.

Internal medicines residency programs participating in iCOMPARE.

	Program Name	Intervention*
1	Abington Memorial Hospital Program	STD
2	Advocate Lutheran General Hospital Program	STD
3	Atlantic Health (Morristown) Program	STD
4	Banner Good Samaritan Medical Center Program	FLEX
5	Baylor College of Medicine Program	STD
6	Baystate Medical Center/Tufts University School of Medicine Program	FLEX
7	Beth Israel Deaconess Medical Center Program	STD
8	Brigham and Women's Hospital Program	STD
9	Brown University Program	STD
10	Canton Medical Education Foundation/NEOMED Program	STD
11	Carilion Clinic-Virginia Tech Carilion School of Medicine Program	FLEX
12	Case Western Reserve University (MetroHealth) Program	FLEX
13	Case Western Reserve University/University Hospitals Case Medical Center Program	FLEX
14	Cedars-Sinai Medical Center Program	STD
15	Cleveland Clinic Foundation Program	FLEX
16	Creighton University Program	FLEX
17	Drexel University College of Medicine/Hahnemann University Hospital Program	STD
18	Duke University Hospital Program	FLEX
19	Eastern Virginia Medical School Program	FLEX
20	Emory University Program	FLEX
21	Geisinger Health System Program	FLEX
22	George Washington University Program	FLEX
23	Georgetown University Hospital/Washington Hospital Center Program	STD
24	Greater Baltimore Medical Center Program	STD
25	Henry Ford Hospital/Wayne State University Program	FLEX
26	Jackson Memorial Hospital/Jackson Health System Program	FLEX
27	Johns Hopkins University Program	FLEX
28	Johns Hopkins University/Bayview Medical Center Program	STD
29	Lahey Clinic Program	FLEX
30	Lankenau Medical Center Program	FLEX
31	Massachusetts General Hospital Program	STD
32	Medical College of Wisconsin Affiliated Hospitals Program	FLEX
33	Mercy Catholic Medical Center Program	STD

	Program Name	Intervention*
34	Morehouse School of Medicine Program	FLEX
35	Olive View/UCLA Medical Center Program	STD
36	Pitt County Memorial Hospital/East Carolina University Program	FLEX
37	St Agnes HealthCare Program	STD
38	St Francis Hospital of Evanston Program	FLEX
39	Stanford University Program	STD
40	Temple University Hospital Program	FLEX
41	Texas A&M College of Medicine-Scott and White Program	FLEX
42	Texas Tech University (Lubbock) Program	STD
43	Thomas Jefferson University Program	STD
44	Tufts Medical Center Program	STD
45	UCLA Medical Center Program	FLEX
46	UMDNJ Robert Wood Johnson Medical School (Camden)/Cooper University Hospital Program	STD
47	University Hospital/University of Cincinnati College of Medicine Program	STD
48	University of Colorado Denver Program	STD
49	University of Connecticut Program	STD
50	University of Kansas School of Medicine Program	FLEX
51	University of Maryland Program	FLEX
52	University of Massachusetts Program	STD
53	University of Nebraska Medical Center College of Medicine Program	STD
54	University of North Carolina Hospitals Program	STD
55	University of Pennsylvania Program	FLEX
56	University of Vermont/Fletcher Allen Health Care Program	FLEX
57	University of Washington Program	FLEX
58	UPMC Medical Education Program	STD
59	Virginia Commonwealth University Health System Program	FLEX
60	Wake Forest University School of Medicine Program	STD
61	Washington University/B-JH/SLCH Consortium Program	STD
62	West Virginia University Program	FLEX
63	Yale-New Haven Medical Center Program	FLEX

*STD = standard arm; FLEX = flexible arm

Table S1. End-of-year survey of trainees (survey questions and response options)

The survey administered to trainees at iCOMPARE programs in May 2015 is provided; the same survey, with the header updated, was administered to trainees at iCOMPARE programs in May 2016.

iCOMPARE Baseline

Baseline End of Year survey to Internal Medicine Interns and Residents

Beginning July 1, 2015, your Internal Medicine Residency program has enrolled in a study about resident duty hours and patient safety, iCOMPARE. This study is intended to inform future national duty hour policies. As part of this work, we are asking all interns and residents to take the following baseline (pre-study) survey. We estimate that it will take about 10-15 minutes to complete. The data will go directly to a secure server. Your program director and chair will never have access to your individual responses. The only identifier attached to individual responses will be the program ID. All data will be aggregated for analyses and reporting.

There are 21 questions in this survey

Opening Questions

1 [1]1. What year residency are you currently enrolled in? *

Please choose **only one** of the following:

- ○ PGY1
- ○ PGY2
- ○ PGY3
- ○ PGY4
- ○ PGY5
- ○ Other

2 [2]2. What specialty is your residency program: *

Please choose **only one** of the following:

- ○ Internal medicine (categorical, primary care, research track, etc)
- ○ Med-peds
- ○ Med-derm
- ○ Other

Intern Items

3 [3]During your most recent month on a MEDICINE FLOOR rotation, approximately how many times did you do the following? *

Only answer this question if the following conditions are met:
° Answer was 'PGY1' at question '1 [1]' (1. What year residency are you currently enrolled in?)

Please choose the appropriate response for each item:

	0 times	1-2 times	3-5 times	6-10 times	> 10 times
a. leave or miss educational conferences during a scheduled shift because of duty hour limits	○	○	○	○	○
b. handoff an active patient care issue because of duty hour limits	○	○	○	○	○
c. leave during a patient encounter because of duty hour limits	○	○	○	○	○
d. miss a patient encounter (e.g. family meeting) because of duty hour limits	○	○	○	○	○
e. work more than 16 hours continuously in house	○	○	○	○	○
f. have < 8 hours off between shifts	○	○	○	○	○

4 [3a]3a. Please indicate the reasons you worked >16 hours or had <8 hours off between shifts (yes/no for each)

Only answer this question if the following conditions are met:
°

-------- Scenario 1 --------

Answer was 'PGY1' at question '1 [1]' (1. What year residency are you currently enrolled in?) *and* Answer was '3-5 times' or '1-2 times' or '6-10 times' or '> 10 times' at question '3 [3]' (During your most recent month on a MEDICINE FLOOR rotation, approximately how many times did you do the following? (e. work more than 16 hours continuously in house))

-------- or Scenario 2 --------

Answer was '1-2 times' or '3-5 times' or '6-10 times' or '> 10 times' at question '3 [3]' (During your most recent month on a MEDICINE FLOOR rotation, approximately how many times did you do the following? (f. have < 8 hours off between shifts))

Please choose the appropriate response for each item:

	Yes	No
a. to perform routine responsibilities	○	○
b. to facilitate care transitions (e.g. signing out patients, transferring patient to ICU)	○	○
c. to stabilize critically ill patients	○	○
d. to complete an admission	○	○
e. to return to work when off-duty because my patient's condition worsened	○	○
f. to complete documentation (i.e. daily notes, discharge summaries, prescriptions,	○	○

	Yes	No
etc)		
g. to attend educational conferences or activities	○	○
h. to round with the team	○	○

5 [4]4. Overall, how do the intern duty hour regulations for this academic year (July 2014-present) at your main hospital affect: *

Only answer this question if the following conditions are met:
° Answer was 'PGY1' at question '1 [1]' (1. What year residency are you currently enrolled in?)

Please choose the appropriate response for each item:

	Positive effect	No effect	Negative effect
a. Safety of patient care	○	○	○
b. Continuity of care (ability to provide the highest level and extent of clinical care and oversight for your patients without forced interruptions or handoffs)	○	○	○
c. Ability to attend required educational conferences	○	○	○
d. Ability to acquire clinical skills	○	○	○
e. Ability to acquire clinical reasoning skills	○	○	○
f. Intern autonomy	○	○	○

	Positive effect	No effect	Negative effect
g. Number of patients interns fully evaluate on admission to the hospital	○	○	○
h. Intern availability for elective patient care encounters(e.g. family meeting)	○	○	○
i. Intern availability for urgent patient care encounters(e.g. RRTs/codes; end of life discussion)	○	○	○
j. Time to teach medical students	○	○	○
k. The relationship between interns and all other residents	○	○	○
l. Professionalism	○	○	○
m. Intern morale	○	○	○

6 [5]5. Overall, how do the intern duty hour regulations for this academic year (July 2014-present)at your main hospital affect: *

Only answer this question if the following conditions are met:
° Answer was 'PGY1' at question '1 [1]' (1. What year residency are you currently enrolled in?)

Please choose the appropriate response for each item:

	Positive effect	No effect	Negative effect
a. Your need to perform patient care related	○	○	○

	Positive effect	No effect	Negative effect
work outside of the hospital. (e.g., review medical record, read)			
b. The pace of your work day	○	○	○
c. Your ability to participate in research	○	○	○
d. Your satisfaction with your job	○	○	○
e. Your satisfaction with the decision to become a physician	○	○	○
f. Your time for family and friends	○	○	○
g. Your time for hobbies and outside interests	○	○	○
h. Your health	○	○	○
i. How well-rested you feel	○	○	○
j. Your overall wellbeing	○	○	○

7 [6]6. Please tell us whether you agree or disagree with the following statements about your main hospital: *

Only answer this question if the following conditions are met:
° Answer was 'PGY1' at question '1 [1]' (1. What year residency are you currently enrolled in?)

Please choose the appropriate response for each item:

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
a. Interns/residents	○	○	○	○	○

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
have adequate faculty supervision					
b. Interns/residents are involved in quality improvement initiatives	○	○	○	○	○
c. The culture emphasizes patient safety	○	○	○	○	○
d. Information is exchanged effectively between interns/residents during transitions in care	○	○	○	○	○
e. Interns/residents work well in interdisciplinary teams	○	○	○	○	○
f. Interns/residents are well versed in fatigue management and mitigation strategies	○	○	○	○	○

8 [7]7. Thinking back on <u>the last 6 months</u> (December 2014 to present), how satisfied were you with the following? *

Only answer this question if the following conditions are met:
° Answer was 'PGY1' at question '1 [1]' (1. What year residency are you currently enrolled in?)

Please choose the appropriate response for each item:

	Very Satisfied	Satisfied	Neutral	Dissatisfied	Very Dissatisfied
a. Continuity of care	○	○	○	○	○
b. Patient safety	○	○	○	○	○
c. Level of attending supervision	○	○	○	○	○
d. Work hours and scheduling	○	○	○	○	○
e. Quality and ease of handoffs and transitions in care	○	○	○	○	○
f. Quality of overall resident education	○	○	○	○	○
g. Time for rest	○	○	○	○	○
h. Your overall wellbeing	○	○	○	○	○
i. Your program's duty hour regulations	○	○	○	○	○
j. Your ability to follow the clinical care of the patients you admit	○	○	○	○	○
k. Number of patients you got to admit completely (ie, someone else did not start or complete the task of admitting the patient).	○	○	○	○	○

9 [8]8. Thinking back on <u>the last 6 months</u> (December 2014 to present), how often did you feel that your fatigue affected: *

Only answer this question if the following conditions are met:
° Answer was 'PGY1' at question '1 [1]' (1. What year residency are you currently enrolled in?)

Please choose the appropriate response for each item:

	Almost always	Often	Sometimes	Rarely	Never
a. Your personal safety	○	○	○	○	○
b. Patient safety	○	○	○	○	○

10 [9]9. Thinking back to <u>your last two weeks</u> of inpatient medicine, how many time did you personally witness: *

Only answer this question if the following conditions are met:
° Answer was 'PGY1' at question '1 [1]' (1. What year residency are you currently enrolled in?)

Please choose the appropriate response for each item:

	0 times	1-2 times	3-5 times	6-10 times	> 10 times
a. A patient error that resulted from intern/resident fatigue	○	○	○	○	○
b. A patient error that resulted from an inadequate handoff?	○	○	○	○	○
c. A patient error that resulted from the responding intern/resident now knowing the patient well enough?	○	○	○	○	○
d. A delay in patient discharge that was due to ineffective communication between team members?	○	○	○	○	○

11 [10]

10. The following are the current standard duty hour regulations as put forth by the ACGME:

Regulation 1. 16 hour maximum for interns

Regulation 2.8-10 hours off between shifts

Regulation 3.28 hour maximum shift for residents

Regulation 4.14 hours off after a 24 hour shift

If duty hour rules were simplified to eliminate the duty hour regulations listed above (while maintaining the 80 hour work week, one day off in 7, and, call no more frequently than every third night, all averaged over 4 weeks), what effect do you believe it would have on:

*****

Only answer this question if the following conditions are met:
° Answer was 'PGY1' at question '1 [1]' (1. What year residency are you currently enrolled in?)

Please choose the appropriate response for each item:

	Positive effect	No effect	Negative effect
a. Safety of patient care	○	○	○
b. Continuity of care	○	○	○
c. Quality of resident education	○	○	○
d. Quality of life	○	○	○

Post Intern Items

12 [11]3. During your most recent month on a MEDICINE FLOOR rotation, approximately how many times did you do the following? *

Only answer this question if the following conditions are met:
° Answer was 'PGY5' or 'Other' or 'PGY2' or 'PGY3' or 'PGY4' at question '1 [1]' (1. What year residency are you currently enrolled in?)

Please choose the appropriate response for each item:

	0 times	1-2 times	3-5 times	6-10 times	> 10 times
a. leave or miss educational conferences during a scheduled shift because of duty hour limits	○	○	○	○	○
b. handoff an active patient care issue because of duty hour limits	○	○	○	○	○
c. leave during a patient encounter because of duty hour limits	○	○	○	○	○
d. miss a patient encounter (e.g. family meeting) because of duty hour limits	○	○	○	○	○
e. return to the hospital to care for a patient on your service	○	○	○	○	○
f. work more than 28 hours continuously in house	○	○	○	○	○
g. have < 8 hours off between daily shifts	○	○	○	○	○
h. have <14 hours off after being on call	○	○	○	○	○

13 [11a]3a. Please indicate the reasons you worked >28 hours, had <8 hours off between shifts or <14 hours off after being on call (yes/no for each) *

Only answer this question if the following conditions are met:
°

-------- Scenario 1 --------

Answer was '1-2 times' or '> 10 times' or '3-5 times' or '6-10 times' at question '12 [11]' (3. During your most recent month on a MEDICINE FLOOR rotation, approximately how many times did you do the following? (f. work more than 28 hours continuously in house))

-------- or Scenario 2 --------

Answer was '1-2 times' or '3-5 times' or '6-10 times' or '> 10 times' at question '12 [11]' (3. During your most recent month on a MEDICINE FLOOR rotation, approximately how many times did you do the following? (g. have < 8 hours off between daily shifts))

-------- or Scenario 3 --------

Answer was '1-2 times' or '3-5 times' or '6-10 times' or '> 10 times' at question '12 [11]' (3. During your most recent month on a MEDICINE FLOOR rotation, approximately how many times did you do the following? (h. have <14 hours off after being on call))

Please choose the appropriate response for each item:

	Yes	No
a. to perform routine responsibilities	○	○
b. to facilitate care transitions (e.g. signing out patients, transferring patient to ICU)	○	○
c. to stabilize critically ill patients	○	○
d. to complete an admission	○	○
e. to return to work when off-duty because my patient's condition worsened	○	○
f. to complete documentation (i.e. daily notes, discharge summaries, prescriptions,	○	○

	Yes	No
etc)		
g. to attend educational conferences or activities	○	○
h. to round with the team	○	○

14 [12]4. Overall, how do the resident duty hour regulations for this academic year (July 2014-present) at your main hospital affect: *

Only answer this question if the following conditions are met:
° Answer was 'Other' or 'PGY5' or 'PGY4' or 'PGY3' or 'PGY2' at question '1 [1]' (1. What year residency are you currently enrolled in?)

Please choose the appropriate response for each item:

	Positive effect	No effect	Negative effect
a. Safety of patient care	○	○	○
b. Continuity of care (ability to provide the highest level and extent of clinical care and oversight for your patients without forced interruptions or handoffs)	○	○	○
c. Ability to attend required educational conferences	○	○	○
d. Ability to acquire clinical skills	○	○	○
e. Ability to acquire clinical reasoning skills	○	○	○
f. Resident autonomy	○	○	○

	Positive effect	No effect	Negative effect
g. Number of patients interns fully evaluate on admission to the hospital	○	○	○
h. Resident availability for elective patient care encounters(e.g. family meeting)	○	○	○
i. Resident availability for urgent patient care encounters(e.g. RRTs/codes; end of life discussion)	○	○	○
j. Time to teach medical students	○	○	○
k. The relationship between interns and all other residents	○	○	○
l. Professionalism	○	○	○
m. Resident morale	○	○	○

15 [13]5. Overall, how do the resident duty hour regulations for this academic year (July 2014-present)at your main hospital affect: *

Only answer this question if the following conditions are met:
° Answer was 'Other' or 'PGY5' or 'PGY4' or 'PGY3' or 'PGY2' at question '1 [1]' (1. What year residency are you currently enrolled in?)

Please choose the appropriate response for each item:

	Positive effect	No effect	Negative effect
a. Your need to perform patient	○	○	○

	Positive effect	No effect	Negative effect
care related work outside of the hospital. (e.g., review medical record, read)			
b. The pace of your work day	○	○	○
c. Your ability to participate in research	○	○	○
d. Your satisfaction with your job	○	○	○
e. Your satisfaction with the decision to become a physician	○	○	○
f. Your time for family and friends	○	○	○
g. Your time for hobbies and outside interests	○	○	○
h. Your health	○	○	○
i. How well-rested you feel	○	○	○
j. Your overall wellbeing	○	○	○

16 [14]6. Please tell us whether you agree or disagree with the following statements about your main hospital: *

Only answer this question if the following conditions are met:
° Answer was 'Other' or 'PGY4' or 'PGY5' or 'PGY3' or 'PGY2' at question '1 [1]' (1. What year residency are you currently enrolled in?)

Please choose the appropriate response for each item:

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
a. Interns/residents have adequate faculty supervision	○	○	○	○	○
b. Interns/residents are involved in quality improvement initiatives	○	○	○	○	○
c. The culture emphasizes patient safety	○	○	○	○	○
d. Information is exchanged effectively between interns/residents during transitions in care	○	○	○	○	○
e. Interns/residents work well in interdisciplinary teams	○	○	○	○	○
f. Interns/residents are well versed in fatigue management and mitigation strategies	○	○	○	○	○

17 [15]7. Thinking back on the last 6 months (December 2014 to present), how satisfied were you with the following? *

Only answer this question if the following conditions are met:
° Answer was 'PGY4' or 'PGY2' or 'PGY3' or 'Other' or 'PGY5' at question '1 [1]' (1. What year residency are you currently enrolled in?)

Please choose the appropriate response for each item:

	Very Satisfied	Satisfied	Neutral	Dissatisfied	Very Dissatisfied
a. Continuity of care	○	○	○	○	○
b. Patient safety	○	○	○	○	○
c. Level of attending supervision	○	○	○	○	○
d. Work hours and scheduling	○	○	○	○	○
e. Quality and ease of handoffs and transitions in care	○	○	○	○	○
f. Quality of overall resident education	○	○	○	○	○
g. Time for rest	○	○	○	○	○
h. Your overall wellbeing	○	○	○	○	○
i. Your program's duty hour regulations	○	○	○	○	○
j. Your ability to follow the clinical care of the patients you admit	○	○	○	○	○
k. Number of patients you got to admit completely (ie, someone else did not start or complete the task of admitting the patient).	○	○	○	○	○

18 [16]8. Thinking back on the last 6 months (December 2014 to present), how often did you feel that your fatigue affected *

Only answer this question if the following conditions are met:
° Answer was 'Other' or 'PGY3' or 'PGY5' or 'PGY2' or 'PGY4' at question '1 [1]' (1. What year residency are you currently enrolled in?)

Please choose the appropriate response for each item:

	Almost always	Often	Sometimes	Rarely	Never
a. Your personal safety	○	○	○	○	○
b. Patient safety	○	○	○	○	○

19 [17]9. Thinking back to your last two weeks of inpatient medicine, how many time did you personally witness: *

Only answer this question if the following conditions are met:
° Answer was 'PGY2' or 'PGY3' or 'PGY4' or 'Other' or 'PGY5' at question '1 [1]' (1. What year residency are you currently enrolled in?)

Please choose the appropriate response for each item:

	0 times	1-2 times	3-5 times	6-10 times	> 10 times
a. A patient error that resulted from intern/resident fatigue	○	○	○	○	○
b. A patient error that resulted from an inadequate handoff	○	○	○	○	○
c. A patient error that resulted from the responding intern/resident not knowing the patient well enough	○	○	○	○	○
d. A delay in patient discharge that was due to ineffective	○	○	○	○	○

	0 times	1-2 times	3-5 times	6-10 times	> 10 times
communication between team members					

20 [18]

10. The following are the current standard duty hour regulations as put forth by the ACGME:

Regulation 1. 16 hour maximum for interns

Regulation 2.8-10 hours off between shifts

Regulation 3. 28 hour maximum shift for residents

Regulation 4.14 hours off after a 24 hour shift

If duty hour rules were simplified to eliminate the duty hour regulations listed above (while maintaining the 80 hour work week, one day off in 7, and, call no more frequently than every third night, all averaged over 4 weeks), what effect do you believe it would have on:

*****

Only answer this question if the following conditions are met:
° Answer was 'PGY4' or 'PGY3' or 'PGY2' or 'PGY5' or 'Other' at question '1 [1]' (1. What year residency are you currently enrolled in?)

Please choose the appropriate response for each item:

	Positive effect	No effect	Negative effect
a. Safety of patient care	○	○	○
b. Continuity of care	○	○	○
c. Quality of resident education	○	○	○
d. Quality of life	○	○	○

Maslach Burnout Inventory

21 [MBI]Please read each statement carefully and decide if you ever feel this way *about your job*. If you have *never* had this feeling, mark a "0" (zero). If you have had this feeling, indicate *how often* you feel it by marking the number from 1 to 6 that best describes how frequently you feel that way.

How often:

M

*****

Please choose the appropriate response for each item:

	never (0)	a few times a year (1)	once a month or less (2)	a few times a month (3)	once a week (4)	a few times a week (5)	every day (6)
1. I feel emotionally drained from my work.	○	○	○	○	○	○	○
2. I feel used up at the end of the workday.	○	○	○	○	○	○	○
3. I feel fatigued when I get up in the morning and have to face another day on the job.	○	○	○	○	○	○	○
4. I can easily understand how my patients feel about things.	○	○	○	○	○	○	○
5. I feel as if I treat some patients as if they were impersonal objects.	○	○	○	○	○	○	○
6. Working with people all day is really a strain for	○	○	○	○	○	○	○

	never (0)	a few times a year (1)	once a month or less (2)	a few times a month (3)	once a week (4)	a few times a week (5)	every day (6)
me.							
7. I deal very effectively with the problems of my patients.	○	○	○	○	○	○	○
8. I feel burned out from my work.	○	○	○	○	○	○	○
9. I feel I'm positively influencing other people's lives through my work.	○	○	○	○	○	○	○
10. I've become more callous toward people since I took this job.	○	○	○	○	○	○	○
11. I worry that this job is hardening me emotionally.	○	○	○	○	○	○	○
12. I feel very energetic.	○	○	○	○	○	○	○
13. I feel frustrated by my job.	○	○	○	○	○	○	○
14. I feel I'm working too hard on my job.	○	○	○	○	○	○	○
15. I don't really care what happens to some patients.	○	○	○	○	○	○	○
16. Working with people directly puts too much stress on me.	○	○	○	○	○	○	○
17. I can easily create a relaxed atmosphere with my patients.	○	○	○	○	○	○	○
18. I feel exhilarated after	○	○	○	○	○	○	○

	never (0)	a few times a year (1)	once a month or less (2)	a few times a month (3)	once a week (4)	a few times a week (5)	every day (6)
working closely with my patients.							
19. I have accomplished many worthwhile things in this job.	○	○	○	○	○	○	○
20. I feel like I'm at the end of my rope.	○	○	○	○	○	○	○
21. In my work, I deal with emotional problems very calmly.	○	○	○	○	○	○	○
22. I feel patients blame me for some of their problems.	○	○	○	○	○	○	○

Submit your survey.
Thank you for completing this survey.

Table S2. End-of-year survey of program directors (survey questions and response options)

The survey administered to program directors at iCOMPARE programs in May 2015 is provided; the same survey, with the header updated, was administered to program directors at iCOMPARE programs in May 2016.

iCOMPARE Program Directors Baseline

Thank you for taking the time for this baseline survey. We are sending it to all program directors participating in iCOMPARE. It should take you less than 10 minutes to complete. The data will be aggregated by arms without personal identifiers.

There are 5 questions in this survey

Learning Environment

1 [1]How satisfied are you with:

Please choose the appropriate response for each item:

	Very Dissatisfied	Dissatisfied	Neutral	Satisfied	Very Satisfied
1. Intern ownership of patient care	○	○	○	○	○
2. Resident ownership of patient care	○	○	○	○	○
3. Ability of interns to manage patients they admit	○	○	○	○	○
4. Intern morale	○	○	○	○	○
5. Resident morale	○	○	○	○	○
6. Time for trainees to reflect	○	○	○	○	○
7. Effectiveness of interns in performing clinical duties	○	○	○	○	○
8. Effectiveness of residents in performing clinical duties	○	○	○	○	○
9. Ability of attending to provide real time	○	○	○	○	○

	Very Dissatisfied	Dissatisfied	Neutral	Satisfied	Very Satisfied
feedback to interns on new admissions					
10. Ability of attending to provide real time feedback to interns on patient care activities	○	○	○	○	○
11. Frequency of handoffs	○	○	○	○	○
12. Quality of handoffs	○	○	○	○	○
13. Ability of residents to work in inter-professional teams	○	○	○	○	○

Workload

2 [1]How satisfied are you with:

Please choose the appropriate response for each item:

	Very Dissatisfied	Dissatisfied	Neutral	Satisfied	Very Satisfied
1. Workload of faculty	○	○	○	○	○
2. Workload of residents	○	○	○	○	○
3. Workload of interns	○	○	○	○	○
4. Workload of program director	○	○	○	○	○
5. Opportunity for interns/residents to transition care when fatigued	○	○	○	○	○
6. Ability of	○	○	○	○	○

	Very Dissatisfied	Dissatisfied	Neutral	Satisfied	Very Satisfied
trainees to perform necessary work during the scheduled duty period					
7. Reliance on residents to provide clinical service	○	○	○	○	○

EDUCATION OPPORTUNITIES

3 [3]How satisfied are you with:

Please choose the appropriate response for each item:

	Very Dissatisfied	Dissatisfied	Neutral	Satisfied	Very Satisfied
1. Adequacy of time for bedside teaching for interns	○	○	○	○	○
2. Adequacy of time for bedside teaching for residents	○	○	○	○	○
3. Ability of interns to attend conferences while on inpatient rotations	○	○	○	○	○
4. Ability of residents to attend conferences while on inpatient rotations	○	○	○	○	○
5. Ability of interns to participate in attending	○	○	○	○	○

	Very Dissatisfied	Dissatisfied	Neutral	Satisfied	Very Satisfied
teaching rounds					
6. Ability of residents to participate in attending teaching rounds	○	○	○	○	○
7. Ability of interns to attend family meetings	○	○	○	○	○
8. Ability of residents to attend family meetings	○	○	○	○	○
9. Balance of service vs. education for interns	○	○	○	○	○
10. Balance of service vs. education for residents	○	○	○	○	○
11. Elective rotation time for housestaff	○	○	○	○	○
12. Time for housestaff to do research	○	○	○	○	○
13. Time for housestaff to engage in medical student education or quality improvement	○	○	○	○	○
14. The amount of time housestaff need to spend on night rotations	○	○	○	○	○

PROGRAM ADMINISTRATION AND ORGANIZATION

4 [4]How satisfied are you with:

Please choose the appropriate response for each item:

	Very Dissatisfied	Dissatisfied	Neutral	Satisfied	Very Satisfied
1. Financial support for non teaching services	○	○	○	○	○
2. Financial support to hire incremental allied health professionals (e.g., nurse practitioners) to cover your inpatient services at your main hospital	○	○	○	○	○
3. Financial support to hire incremental hospitalists/additional faculty members for clinical care delivery	○	○	○	○	○
4. Relationship of residency program with hospital administration	○	○	○	○	○
5. Program director morale	○	○	○	○	○
6. Effort of tracking duty hours	○	○	○	○	○

PATIENT OUTCOMES

5 [5]How satisfied are you with:

Please choose the appropriate response for each item:

	Very Dissatisfied	Dissatisfied	Neutral	Satisfied	Very Satisfied
1. Continuity of care for patients	○	○	○	○	○
2. Safety of patients	○	○	○	○	○

	Very Dissatisfied	Dissatisfied	Neutral	Satisfied	Very Satisfied
3. Graduates' preparedness for practice after residency	○	○	○	○	○

Submit your survey.
Thank you for completing this survey.

Table S3. Intervention Dose: Program Directors' Reports of How Flexible* Duty-hour Policy Was Implemented at Their Program (N=29)

	Mean	Median
No. of interns	44.9	41
No. of rotations on which Flexible* was implemented†	2.9	2
Block schedule for Flexible* (count)‡		
4 weeks	18	
3 weeks	1	
2 weeks	3	
Other (3 + 1, 4 + 1)	3	
During any block at the program's hospitals, the number of interns assigned to the following,		
All inpatient rotations §	21.6	20
All Flexible* rotations (subset of above)	14	13
Count and longest shift length for Flexible* interns		
General Medicine (n = 18)	24.3	28
Coronary Care Unit (n = 8)	28.1	28
Medical Intensive Care Unit (n = 23)	25.9	28
Cardiology (n = 12)	23.2	27
Gastroenterology (n = 2)	22.5	22.5
Infectious Disease (n = 2)	15	15
Nephrology (n = 3)	18.7	18
Pulmonary (n = 2)	15	15
Oncology (n = 6)	19.3	18
Neurology (n = 2)	9	9
Other (n = 5)	20.6	18

***** Residency programs assigned to flexible policies were allowed to waive limits on maximum shift length and mandatory time off between shifts.

†Number of rotations selected when presented a list of 13 possibilities plus 3 'other' write-in options.

‡One program director did not answer this question.

§Flexible schedule applies to inpatient rotations only. During any given block, interns could be, for example, at another hospital not using flexible scheduling, on an outpatient rotation, on an away rotation, or on vacation. On average, for the flexible programs, on any given block about half of all interns were at a hospital using a flexible schedule and of those, approximately two-thirds were on a flexible rotation.

Table S4. Intern Time by Activity: Analysis Details

Activity	Percentage of Observed Shift Time Spent in Activity*			
	Flexible† (N = 6 programs; 44 interns; 1072 hours; 96 shifts)	Standard (N = 6 programs; 36 interns; 1101 hours; 98 shifts)	Flexible versus Standard Difference‡ (95% CI)	P Value‡
Primary outcomes				
Direct patient care§				
Observed mean ± SD¶	13.0% ± 3.9%	11.8% ± 4.6%		
Marginal mean‡	13.0%	11.8%	1.2% (-0.7, 3.1)	0.21
Random effect variance‡				
Program (P value)			0.07 (0.47)	
Error			17.72	
Education‖				
Observed mean ± SD¶	7.4% ± 6.6%	7.5% ± 6.0%		
Marginal mean‡	7.3%	7.3%	-0.0% (-5.9, 5.9)	>0.99
Random effect variance‡				
Program (P value)			11.22 (0.001)	
Error			32.43	
Secondary outcomes				
Indirect patient care**				
Observed mean ± SD¶	68.3% ± 11.8%	63.8% ± 11.9%		
Marginal mean‡	67.9%	63.7%	4.2% (-6.7, 15.1)	
Random effect variance‡				
Program (P value)			37.64 (0.002)	
Error			114.79	
Handoffs				
Observed mean ± SD¶	2.6% ± 2.7%	4.0% ± 4.1%		
Marginal mean‡	2.7%	4.0%	-1.3% (-3.8, 1.1)	
Random effect variance‡				
Program (P value)			1.55 (0.04)	
Error			10.57	
Rounds				
Observed mean ± SD¶	22.4% ± 12.0%	18.8% ± 11.4%		
Marginal mean‡	22.4%	19.0%	3.4% (-7.6, 14.5)	
Random effect variance‡				
Program (P value)			39.28 (<0.001)	
Error			110.19	
Miscellaneous††				
Observed mean ± SD¶	5.6% ± 6.6%	9.7% ± 9.8%		
Marginal mean‡	5.6%	9.7%	-4.2% (-7.8, -0.6)	
Random effect variance‡				
Program (P value)			0.00 (>0.99)	
Error			67.02	

Missingness: There are no missing data; each intern observed is included in each model. If an activity was not observed during a shift, the time spent in the activity was 0 minutes.

*The sum of the percentages may exceed 100% because more than one activity can occur simultaneously.

†Residency programs assigned to flexible policies were allowed to waive limits on maximum shift length and mandatory time off between shifts.

‡The marginal mean, random effect variance, Flexible versus Standard difference, and associated P values were obtained from a mixed effects linear regression model with random intercepts (1 fixed term and 1 random term for clustering of intern outcomes within program) and an indicator term for the duty-hour policy group (1=Flexible, 0=Standard) as the fixed effect covariate in the model and the intern's mean percentage of observed shift time spent in the activity across 1-5 shifts as the outcome. A separate mixed model was fit for each activity type shown. The marginal mean is similar to an observed mean but is derived from the regression model and accounts for correlations between respondents at the same program, averaging across random effects due to variation in respondent outcomes within programs. The P value for the hypothesis that the between-program variances equaled 0 (i.e., no within programs correlations) was <0.05 for all outcomes except direct patient care and miscellaneous, reflecting large variation in how interns spent time across all programs, independent of assignment to flexible or standard.

§Direct patient evaluation or in-person patient or family communication.

¶SDs are not adjusted for correlations due to repeated measures on interns and programs.

‖Education activities include teaching or being taught (including teaching rounds), educational conferences, and reading about medicine.

**Indirect patient care activities include activities such as interacting with the electronic chart, viewing imaging, or discussing care with a consultant.

††Miscellaneous activities include activities such as eating or sleeping.

Table S5. American College of Physicians Internal Medicine In-Training Examination (ITE) Scores.

			Flexible versus Standard			
	Flexible (32 programs)	Standard (31 programs)	Difference† (95% CI)	P Value‡	Difference Adjusted for Mean Program Score in Baseline Year§ (95% CI)	P Value‡
PGY2 trainees (former Interns)	N=852	N=835				
ITE score						
Observed mean ± SD§	69.5 ± 7.9	70.5 ± 8.3				
Marginal mean†	68.9	69.4	-0.43 (-2.38, 1.52)	0.06	0.64 (-0.56, 1.84)	<0.001
Random effect variance†§						
Program (P value)			13.08 (<0.001)		3.29 (<0.001)	
Error			53.13		53.16	
All trainees	N=2233	N=2283				
ITE score						
Observed mean ± SD¶	67.3 ± 9.4	68.9 ± 9.4				
Marginal mean†	67.1	68.1	-1.05 (-2.92, 0.82)	0.16	-0.08 (-0.76, 0.60)	<0.001
Random effect variance†§						
Program (P value)			13.01 (<0.001)		0.66 (0.002)	
Error			76.79		76.83	

ITE score is the percentage of questions answered correctly and varies from 0 to 100. The examination is taken in the fall.

Missingness in the trial year: Interns: 1228 Flexible group interns trained at iCOMPARE programs; 852 (69.4% of count of interns at Flexible programs) were included in the analysis and 376 (30.6% of count of interns at Flexible programs) were not included. 1300 Standard group interns trained at iCOMPARE programs; 835 (64.2% of count of interns at Standard programs) were included in the analysis and 465 (35.8% of count of interns at Standard programs) were not included. All trainees: 3099 Flexible group trainees trained at iCOMPARE programs; 2233 (72.1% of count of trainees at Flexible programs) were included in the analysis and 866 (27.9% of count of trainees at Flexible programs) were not included. 3214 Standard group trainees trained at iCOMPARE programs; 2283 (71.0% of count of trainees at Standard programs) were included in the analysis and 931 (29.0% of count of trainees at Standard programs) were not included.

*Residency programs assigned to flexible policies were allowed to waive limits on maximum shift length and mandatory time off between shifts.

†The marginal mean, random effect variance, Flexible versus Standard difference, and Program P value were obtained from a mixed effects linear regression model with random intercepts (1 fixed term and 1 random term for clustering of trainee outcomes within program) and an indicator term for the duty-hour policy group (1=Flexible, 0=Standard) as the fixed effect covariate in the model and the trainee's ITE score as the outcome. The marginal mean is similar to an observed mean but is derived from the regression model and accounts for correlations between respondents at the same program, averaging across random effects due to variation in respondent outcomes within programs. Note the P value for the hypothesis that the between-program variances equaled 0 (i.e., no within programs correlations) was <0.05 for both the intern only analysis and the all trainee analysis, reflecting large variation in ITE score across all programs, independent of assignment to flexible or standard.

‡The P value is for the test that Flexible is not inferior to Standard with non-inferiority margin=2%. The P value is for a one-sided test of H_0: Flexible versus Standard difference did not exceed 2 percentage points and was obtained as the quantity, treatment group coefficient minus (-2).

§The marginal mean, random effect variance, Flexible versus Standard difference, and Program P value were obtained from a mixed effects linear regression model with an intercept (1 fixed term and 1 random term for clustering of trainees within program) and an indicator term for the duty-hour policy group (1=Flexible, 0=Standard) and the program's mean ITE score in the baseline year for the respondent group analyzed (interns or all trainees) as the only fixed effects covariates and the trainee's ITE score as the outcome. The marginal mean is similar to an observed mean but is derived from the regression model and accounts for correlations between respondents at the same program, averaging across random effects due to variation in respondent outcomes within programs. Note the P value for the hypothesis that the between-program variances equaled 0 (i.e., no within programs correlations) was <0.05 for both the intern only analysis and the all trainee analysis, reflecting large variation in ITE score across all programs, independent of assignment to flexible or standard.

¶SDs are not adjusted for correlations between scores for trainees at the same program.

Table S6. 2016 ACGME Survey of Residents.

Theme	Flexible* (32 programs)	Standard (31 programs)	Flexible* versus Standard: Difference† or Odds Ratio‡ (95% CI)	Flexible* versus Standard: P Value†	Flexible* versus Standard: Difference§ or Odds Ratio¶ Adjusted for Program Response in Baseline Year (95% CI)	Flexible* versus Standard: P Value§
Primary Outcome (difference)‖						
Appropriate balance for education						
Marginal mean†	3.88	3.85	0.02 (-0.12, 0.17)	0.74	0.03 (-0.17, 0.14)	0.55
Random effect variance†						
Program (P value)			0.08 (<0.001)		0.04 (<0.001)	
Error			0.71		0.71	
Secondary Outcomes (odds ratio)**						
Duty hours: marginal % noncompliant	16.9%	17.1%	0.99 (0.71, 1.38)		0.92 (0.68, 1.25)	
Educational content: marginal % noncompliant	63.2%	65.5%	0.91 (0.63, 1.30)		0.95 (0.75, 1.21)	
Evaluation (assessment/feedback): marginal % noncompliant	44.1%	48.0%	0.86 (0.65, 1.12)		0.89 (0.70, 1.12)	
Faculty: marginal % noncompliant	30.1%	29.9%	1.01 (0.72, 1.42)		0.92 (0.73, 1.15)	
Patient safety and teamwork: marginal % noncompliant	23.5%	26.0%	0.88 (0.65, 1.18)		0.91 (0.73, 1.13)	
Resources: marginal % noncompliant	45.1%	50.7%	0.80 (0.58, 1.09)		0.92 (0.72, 1.17)	
Overall evaluation of program: marginal % noncompliant	3.2%	3.5%	0.93 (0.60, 1.44)		0.88 (0.59, 1.33)	

Respondents were PGY 1-3 trainees.

Missingness in the trial year: The ACGME reported overall response rate of 91% for each duty-hour policy group.

*Residency programs assigned to flexible policies were allowed to waive limits on maximum shift length and mandatory time off between shifts.

†The marginal mean, Flexible versus Standard difference in score, Random effect variance, and associated P values were obtained from a mixed effects linear regression model with an intercept (1 random term for clustering of trainees within program) and an indicator term for the duty-hour policy group (1=Flexible, 0=Standard) as the fixed effect covariate in the model and the trainee's ordinal score as the outcome. The marginal mean is similar to the observed mean but is derived from the regression model and accounts for correlations between respondents at the same program, averaging across random effects due to variation in respondent outcomes within programs. The P value for the random effects variance tests the hypothesis that the between-program variance equaled 0 (i.e., no within programs correlations); P <0.05 reflects large variation in how program directors and core faculty responded across all programs, independent of assignment to flexible or standard.

‡The marginal percent and Flexible versus Standard odds ratio of a noncompliant response were obtained from a logistic regression model with generalized estimating equations and robust variance estimation using an independent working correlation matrix to account for the correlations between responses from respondents at the same program; the model included an indicator term for the duty-hour policy group (1=Flexible, 0=Standard) as the only covariate and the trainee's dichotomized response for the theme as the outcome. A separate model was fit for each theme. The marginal percentage is similar to the observed percentage but is derived from the regression model and accounts for correlations between respondents at the same program.

§The marginal mean, Flexible versus Standard difference in score, Random effect variance, and associated P values were obtained from a mixed effects linear regression model with an intercept (1 random term for clustering of trainees within program) and an indicator term for the duty-hour policy group (1=Flexible, 0=Standard) and the program's mean response in the baseline year as the only fixed effects covariates and the trainee's ordinal score as the outcome. The marginal mean is similar to the observed mean but is derived from the regression model and accounts for correlations between respondents at the same program, averaging across random effects due to variation in respondent outcomes within programs. The P value for the random effect variance tests the hypothesis that the between-program variance equaled 0 (i.e., no within programs correlations); P <0.05 reflects large variation in how trainees responded across all programs, independent of assignment to flexible or standard.

¶ The marginal percent and Flexible versus Standard odds ratio of a noncompliant response were obtained from a logistic regression model with generalized estimating equations and robust variance estimation using an independent working correlation matrix to account for the correlations between responses from respondents at the same program; the model included an indicator term for the duty-hour policy group (1=Flexible, 0=Standard) and the program's mean response in the baseline year as the only covariates and the trainee's dichotomized response for the theme as the outcome. A separate model was fit for each theme. The marginal percentage is similar to the observed percentage but is derived from the regression model and accounts for correlations between respondents at the same program.

‖The primary outcome for the hypothesis about trainee satisfaction was a single question from the ACGME residents survey: "Major assignments provide an appropriate balance between education and other clinical demands." Response options were Never (scored 1), Rarely (scored 2), Sometimes (scored 3), Often (scored 4), and Very Often (scored 5) and the ordinal score was analyzed.

**Additional ACGME trainee survey measures were secondary outcomes. Each secondary outcome is the trainee's dichotomized response (noncompliant versus compliant) for the specified theme. Each theme comprises 1-8 survey questions. Each survey item had 5 response choices which were dichotomized into a binary response for the item. The trainee's dichotomized responses across the survey items comprising the theme were pooled to provide a theme-level binary response of noncompliant versus compliant. The response for a theme is noncompliant if the respondent provided a noncompliant response to any of the questions comprising the theme. See online supplement for details regarding questions included in each theme.

Table S7. End-of-Year Survey: Details and Expanded Analyses of Intern Responses

	Flexible* (N = 30 programs; 638 interns)	Standard (N = 31 programs; 608 interns)	Flexible versus Standard	
			Odds Ratio† (95% CI)	Odds Ratio Adjusted for Program Response in Baseline Year‡ (95% CI)
Interns having perception of negative effect¶ of institutional duty hours on:				
Safety of patient care (q4a)				
Observed %	15.7%	6.3%		
Marginal %†	14.1%	5.7%	2.71 (1.65, 4.48)	2.67 (1.63, 4.38)
Random (program) variance (P value) †‡			0.28 (0.002)	0.22 (0.007)
Continuity of care (q4b)				
Observed %	13.0%	29.7%		
Marginal %†	11.0%	25.9%	0.35 (0.22, 0.57)	0.36 (0.23, 0.58)
Random (program) variance (P value) †‡			0.43 (<0.001)	0.32 (<0.001)
Ability to acquire clinical skills (q4d)				
Observed %	10.6%	11.4%		
Marginal %†	10.0%	9.9%	1.01 (0.62, 1.63)	1.01 (0.61, 1.66)
Random (program) variance (P value) †‡			0.29 (0.005)	0.27 (0.007)
Intern autonomy (q4f)				
Observed %	6.5%	9.4%		
Marginal %†	5.8%	8.3%	0.68 (0.40, 1.14)	0.68 (0.39, 1.20)
Random (program) variance (P value) †‡			0.26 (0.02)	0.28 (0.02)
Availability for urgent patient care encounters (q4i)				
Observed %	7.1%	10.1%		
Marginal %†	6.9%	9.8%	0.68 (0.44, 1.05)	0.71 (0.45, 1.13)
Random (program) variance (P value) †‡			0.07 (0.23)	0.06 (0.25)
Availability for elective patient care encounters (q4h)				
Observed %	17.6%	12.1%		

	Flexible* (N = 30 programs; 638 interns)	Standard (N = 31 programs; 608 interns)	Flexible versus Standard	
			Odds Ratio† (95% CI)	Odds Ratio Adjusted for Program Response in Baseline Year‡ (95% CI)
Marginal %†	15.1%	10.8%	1.47 (0.94, 2.30)	1.44 (0.93, 2.25)
Random (program) variance (P value)			0.28 (<0.001)	0.22 (0.005)
Ability to attend required educational conferences (q4c)				
Observed %	23.1%	10.4%		
Marginal %†	21.6%	9.5%	2.61 (1.77, 3.85)	2.58 (1.74, 3.82)
Random (program) variance (P value) †‡			0.14 (0.008)	0.09 (0.07)
Relationship between interns and residents (q4k)				
Observed %	8.4%	4.0%		
Marginal %†	7.2%	3.4%	2.21 (1.20, 4.06)	1.94 (1.03, 3.67)
Random (program) variance (P value) †‡			0.34 (0.01)	0.28 (0.03)
Time for teaching medical students (q4j)				
Observed %	23.1%	17.4%		
Marginal %†	22.4%	16.6%	1.44 (1.02, 2.04)	1.43 (0.99, 2.07)
Random (program) variance (P value) †‡			0.12 (0.01)	0.14 (0.01)
Need to perform patient care related work outside of the hospital (q5a)				
Observed %	26.0%	26.4%		
Marginal %†	24.3%	25.5%	0.94 (0.65, 1.36)	0.93 (0.64, 1.34)
Random (program) variance (P value) †‡			0.21 (<0.001)	0.17 (0.002)
Ability to participate in research (q5c)				
Observed %	26.2%	10.9%		
Marginal %†	25.6%	10.7%	2.88 (2.01, 4.12)	2.52 (1.70, 3.73)
Random (program) variance (P value) †‡			0.09 (0.06)	0.08 (0.10)
Professionalism (q4l)				
Observed %	8.7%	2.8%		
Marginal %†	7.7%	2.5%	3.26 (1.72, 6.16)	3.59 (1.87, 6.89)
Random (program) variance (P value) †‡			0.26 (0.05)	0.16 (0.14)

	Flexible* (N = 30 programs; 638 interns)	Standard (N = 31 programs; 608 interns)	Flexible versus Standard	
			Odds Ratio† (95% CI)	Odds Ratio Adjusted for Program Response in Baseline Year‡ (95% CI)
Job satisfaction (q5d)				
Observed %	23.2%	6.5%		
Marginal %†	21.3%	5.9%	4.32 (2.72, 6.85)	4.45 (2.77, 7.15)
Random (program) variance (P value) †‡			0.21 (0.002)	0.15 (0.02)
Satisfaction with career choice (q5e)				
Observed %	20.1%	5.5%		
Marginal %†	17.6%	4.8%	4.26 (2.52, 7.20)	4.49 (2.56, 7.85)
Random (program) variance (P value) †‡			0.32 (<0.001)	0.29 (0.002)
Intern morale (q4m)				
Observed %	26.4%	4.3%		
Marginal %†	24.0%	3.7%	8.14 (4.65, 14.26)	8.90 (4.90, 16.19)
Random (program) variance (P value) †‡			0.36 (<0.001)	0.29 (0.001)
Time with family and friends (q5f)				
Observed %	35.0%	7.7%		
Marginal % †	32.6%	7.4%	6.11 (3.76, 9.91)	5.88 (3.67, 9.42)
Random (program) variance (P value) †‡			0.37 (<0.001)	0.27 (<0.001)
Time for hobbies and outside interests (q5g)				
Observed %	33.7%	8.4%		
Marginal %†	31.0%	8.0%	5.20 (3.17, 8.54)	5.01 (3.07, 8.17)
Random (program) variance (P value) †‡			0.42 (<0.001)	0.32 (<0.001)
Health (q5h)				
Observed %	32.3%	7.5%		
Marginal %†	29.2%	6.9%	5.53 (3.32, 9.20)	4.70 (2.77, 7.97)
Random (program) variance (P value) †‡			0.43 (<0.001)	0.33 (<0.001)
Ability to acquire clinical reasoning skills (q4e)				
Observed %	9.2%	6.9%		
Marginal %†	8.5%	5.9%	1.47 (0.87, 2.51)	1.47 (0.85, 2.55)

	Flexible* (N = 30 programs; 638 interns)	Standard (N = 31 programs; 608 interns)	Flexible versus Standard	
			Odds Ratio† (95% CI)	Odds Ratio Adjusted for Program Response in Baseline Year‡ (95% CI)
Random (program) variance (P value) †‡			0.30 (0.02)	0.28 (0.02)
Pace of intern's work day (q5b)				
Observed %	20.3%	19.4%		
Marginal %†	18.9%	17.0%	1.14 (0.76, 1.71)	1.13 (0.74, 1.73)
Random (program) variance (P value) †‡			0.25 (<0.001)	0.24 (<0.001)
Intern's overall well-being (q5j)				
Observed %	28.4%	6.5%		
Marginal %†	25.8%	6.2%	5.27 (3.22, 8.64)	4.97 (3.08, 8.02)
Random (program) variance (P value) †‡			0.33 (<0.001)	0.21 (0.005)
Interns reporting dissatisfaction§ with:				
Overall quality of resident education (q7f)				
Observed %	14.7%	9.1%		
Marginal %†	12.7%	8.0%	1.67 (1.02, 2.73)	1.67 (1.03, 2.72)
Random (program) variance (P value)†‡			0.33 (<0.001)	0.22 (0.001)
Overall well-being (q7h)				
Observed %	31.5%	15.8%		
Marginal %†	30.2%	14.9%	2.47 (1.67, 3.65)	2.13 (1.42, 3.18)
Random (program) variance (P value)†‡			0.23 (<0.001)	0.18 (0.002)
Patient safety (q7b)				
Observed %	5.8%	4.2%		
Marginal %†	5.8%	4.2%	1.40 (0.83, 2.36)	1.29 (0.74, 2.24)
Random (program) variance (P value) †‡			0.00 (>0.99)	0.00 (>0.99)
Continuity of care (q7a)				
Observed %	6.1%	8.1%		
Marginal % †	5.4%	6.7%	0.80 (0.46, 1.41)	0.63 (0.34, 1.17)
Random (program) variance (P value) †‡			0.29 (0.03)	0.29 (0.03)

	Flexible* (N = 30 programs; 638 interns)	Standard (N = 31 programs; 608 interns)	Flexible versus Standard	
			Odds Ratio† (95% CI)	Odds Ratio Adjusted for Program Response in Baseline Year‡ (95% CI)
Quality and ease of handoffs and transitions in care (q7e)				
Observed %	6.6%	7.4%		
Marginal %†	6.2%	6.9%	0.89 (0.54, 1.46)	0.71 (0.40, 1.28)
Random (program) variance (P value) †‡			0.15 (0.12)	0.11 (0.19)
Duty-hour regulations of the program (q7i)				
Observed %	14.5%	5.7%		
Marginal %†	13.2%	5.2%	2.78 (1.69, 4.57)	2.44 (1.49, 3.99)
Random (program) variance (P value) †‡			0.22 (0.02)	0.11 (0.17)
Work hours and scheduling (q7d)				
Observed %	22.4%	11.8%		
Marginal %†	21.1%	10.8%	2.21 (1.45, 3.37)	2.24 (1.43, 3.52)
Random (program) variance (P value) †‡			0.23 (0.001)	0.23 (0.002)
Time for rest (q7g)				
Observed %	35.3%	18.5%		
Marginal %†	33.7%	17.3%	2.43 (1.62, 3.63)	2.06 (1.39, 3.04)
Random (program) variance (P value) †‡			0.28 (<0.001)	0.17 (0.005)
Level of attending supervision (q7c)				
Observed %	2.9%	2.0%		
Marginal %†	2.9%	2.0%	1.45 (0.69, 3.03)	1.51 (0.69, 3.28)
Random (program) variance (P value) †‡			0.00 (>0.99)	0.00 (>0.99)
Ability to follow the clinical care of the patients the intern admits (q7j)				
Observed %	5.0%	6.1%		
Marginal %†	4.9%	5.8%	0.83 (0.49, 1.39)	0.75 (0.43, 1.31)
Random (program) variance (P value) †‡			0.05 (0.36)	0.02 (0.44)

	Flexible* (N = 30 programs; 638 interns)	Standard (N = 31 programs; 608 interns)	Flexible versus Standard	
			Odds Ratio† (95% CI)	Odds Ratio Adjusted for Program Response in Baseline Year‡ (95% CI)
The number of patients the intern got to admit completely (i.e., someone else did not start of complete the task of admitting the patient) (q7k)				
Observed %	5.3%	10.8%		
Marginal %†	4.5%	8.9%	0.48 (0.27, 0.85)	0.51 (0.29, 0.88)
Random (program) variance (P value) †‡			0.36 (0.003)	0.22 (0.04)
Interns perceiving that their fatigue‖:				
Almost always or often affected their personal safety (q8a)				
Observed %	15.6%	8.7%		
Marginal %†	15.5%	8.4%	2.01 (1.28, 3.14)	2.00 (1.25, 3.23)
Random (program) variance (P value) †‡			0.21 (0.01)	0.23 (0.009)
Almost always or often affected patient safety (q8b)				
Observed %	11.7%	7.6%		
Marginal %†	11.6%	7.4%	1.64 (1.06, 2.52)	1.61 (1.00, 2.57)
Random (program) variance (P value) †‡			0.10 (0.20)	0.14 (0.12)
Interns reporting at least 1 occurrence during their most recent month on a medicine floor rotation:**				
Left during a patient encounter because of duty hour limits (q3c)				
Observed %	5.2%	5.1%		
Marginal %†	4.7%	4.8%	0.98 (0.56, 1.73)	1.12 (0.63, 2.01)
Random (program) variance (P value) †‡			0.18 (0.14)	0.14 (0.21)
Missed a patient encounter because of duty hour limits (q3d)				
Observed %	22.6%	15.8%		
Marginal %†	18.1%	14.6%	1.29 (0.80, 2.09)	1.32 (0.88, 1.99)

	Flexible* (N = 30 programs; 638 interns)	Standard (N = 31 programs; 608 interns)	Flexible versus Standard	
			Odds Ratio† (95% CI)	Odds Ratio Adjusted for Program Response in Baseline Year‡ (95% CI)
Random (program) variance (P value) †‡			0.43 (<0.001)	0.19 (0.003)
Handed off an active patient care issue because of duty hour limits (q3b)				
Observed %	31.4%	33.9%		
Marginal %†	27.9%	32.4%	0.81 (0.53, 1.22)	0.77 (0.53, 1.13)
Random (program) variance (P value) †‡			0.37 (<0.001)	0.21 (<0.001)
Left or missed educational conferences during a scheduled shift because of duty hour limits (q3a)				
Observed %	31.9%	27.0%		
Marginal %†	30.6%	26.6%	1.22 (0.87, 1.71)	1.13 (0.79, 1.61)
Random (program) variance (P value) †‡			0.17 (0.002)	0.14 (0.01)
Worked more than 16 hours continuously in house (q3e)				
Observed %	61.1%	34.7%		
Marginal %†	57.8%	31.2%	3.02 (1.79, 5.10)	3.39 (2.23, 5.17)
Random (program) variance (P value) †‡			0.74 (<0.001)	0.34 (<0.001)
Had less than 8 hours off between daily shifts (q3f)				
Observed %	33.4%	36.2%		
Marginal %†	29.8%	32.1%	0.90 (0.56, 1.44)	0.86 (0.59, 1.26)
Random (program) variance (P value) †‡			0.55 (<0.001)	0.22 (<0.001)

Trainees were asked to score 40 aspects of their trainee experiences. This table reports results only for interns; Table S4 reports results for all trainees combined. For each item, the response choices were dichotomized into a binary response as indicated for each question theme (response choices for each question theme are indicated in the footnote on the theme heading).

Missingness in the trial year: 1228 Flexible group interns were sent the survey; 622 (50.7%) answered every question, 16 (1.3%) answered 1 or more but not all questions, and 590 (48.0%) answered no questions. 1300 Standard group interns were sent the survey; 594 (45.7%) answered every question, 14 (1.1%) answered 1 or more but not all questions, and 692 (53.2%) answered no questions. Observed percentages are the percentages of interns answering the question.

*Residency programs assigned to flexible policies were allowed to waive limits on maximum shift length and mandatory time off between shifts.

†The marginal percents, random (program) variance, Flexible versus Standard odds ratio, and associated P value were obtained from a mixed effects logistic regression model with an intercept (1 random term for clustering of respondents within program) and an indicator term for the duty-hour policy group (1=Flexible, 0=Standard) as the fixed effect covariate in the model and the respondent's dichotomized response as the outcome. A separate mixed model was fit for each survey question shown. The marginal percentage is similar to the observed percentage but is derived from the regression model and accounts for correlations between respondents at the same program, averaging across random effects due to variation in respondent outcomes within programs. Note the P value for the hypothesis that the between-programs variance equaled 0 (i.e., no within programs correlations) was <0.05 for all outcomes except availability for urgent patient care encounters, ability to participate in research, professionalism, dissatisfaction with patient safety, quality and ease in patient handoffs and transitions, dissatisfaction with level of attending supervision, ability to follow clinical care of patients admitted by the intern, fatigue affecting patient safety, left patient encounter because of duty-hour limits, reflecting large variation in how interns perceive duty-hours impact across all programs, independent of assignment to flexible or standard.

‡The random (program) variance, Flexible versus Standard odds ratio, and associated P value were obtained from a mixed effects logistic regression model with an intercept (1 random term for clustering of respondents within programs) and an indicator term for the duty-hour policy group (1=Flexible, 0=Standard) and the program's percent responding negatively in 2015 as the only fixed effects covariates in the model and the respondent's dichotomized response for the item as the outcome. A separate mixed model was fit for each survey question shown. The marginal percentage is similar to the observed percentage but is derived from the regression model and accounts for correlations between respondents at the same program, averaging across random effects due to variation in respondent outcomes within programs. Note the P value for the hypothesis that the between-programs variance equaled 0 (i.e., no within programs correlations) was <0.05 for all outcomes except availability for urgent patient care encounters, ability to attend required educational conferences, ability to participate in research, professionalism, dissatisfaction with patient safety, quality and ease in patient handoffs and transitions, dissatisfaction with duty-hour regulations of the program, dissatisfaction with level of attending supervision, ability to follow clinical care of patients admitted by the intern, fatigue affecting patient safety, and left patient encounter because of duty-hour limits, reflecting large variation in how interns perceive duty-hours impact across all programs, independent of assignment to flexible or standard.

§Very dissatisfied or dissatisfied versus neutral, satisfied or very satisfied.

¶Negative effect versus no effect or positive effect.

‖Always or often versus sometimes, rarely or never.

**1 or more occurrences in the past month versus no occurrence.

Table S8. Maslach Burnout Inventory Scores: Details and Expanded Analyses of Intern Scores and All Trainee Scores.

	Flexible* (30 programs)	Standard (31 programs)	Flexible versus Standard	
			Odds Ratio† (95% CI)	Odds Ratio Adjusted for Program Response in Baseline Year‡ (95% CI)
Interns only	N=594	N=563		
Emotional exhaustion subscale (scored 0-54)				
Observed score, mean ± SD§	25.9 ± 11.7	24.7 ± 12.0		
Scoring high or moderate (≥ 17)				
Observed %	78.3%	71.9%		
Marginal %†	78.9%	72.3%	1.43 (0.96, 2.13)	1.52 (1.07, 2.17)
Random (program) variance (P value)†‡			0.27 (<0.001)	0.13 (0.02)
Depersonalization subscale (scored 0-30)				
Observed score, mean ± SD§	11.9 ± 6.8	11.3 ± 6.9		
Scoring high or moderate (≥ 7)				
Observed %	75.1%	72.3%		
Marginal %†	75.3%	72.1%	1.18 (0.81, 1.71)	1.29 (0.93, 1.80)
Random (program) variance (P value)†‡			0.22 (0.001)	0.09 (0.05)
Personal accomplishment subscale (scored 0-48)				
Observed score, mean ± SD§	33.5 ± 8.4	34.2 ± 8.1		
Scoring low or moderate (0-38)				
Observed %	70.7%	68.2%		
Marginal %†	71.0%	68.6%	1.12 (0.84, 1.49)	1.13 (0.84, 1.53)
Random (program) variance (P value) †‡			0.06 (0.09)	0.06 (0.10)
All trainees	N=1345	N=1313		
Emotional exhaustion subscale (scored 0-54)				
Observed score, mean ± SD§	23.9 ± 11.8	22.7 ± 12.2		
Scoring high or moderate (≥ 17)				
Observed %	70.9%	65.5%		
Marginal %†	70.8%	64.8%	1.32 (0.93, 1.86)	1.18 (0.92, 1.51)

	Flexible* (30 programs)	Standard (31 programs)	Flexible versus Standard	
			Odds Ratio† (95% CI)	Odds Ratio Adjusted for Program Response in Baseline Year‡ (95% CI)
Random (program) variance (P value) †‡			0.32 (<0.001)	0.08 (0.001)
Depersonalization subscale (scored 0-30)				
Observed score, mean ± SD§	11.1 ± 6.7	10.8 ± 6.9		
Scoring high or moderate (≥ 7)				
Observed %	70.7%	68.3%		
Marginal %†	70.2%	67.1%	1.16 (0.82, 1.63)	1.10 (0.87, 1.38)
Random (program) variance (P value) †‡			0.32 (<0.001)	0.06 (0.005)
Personal accomplishment subscale (scored 0-48)				
Observed score, mean ± SD§	33.9 ± 8.5	34.3 ± 8.7		
Scoring low or moderate (0-38)				
Observed %	67.6%	63.4%		
Marginal %†	67.9%	64.1%	1.18 (0.97, 1.44)	1.16 (0.96, 1.41)
Random (program) variance (P value) †‡			0.04 (0.02)	0.02 (0.09)

The Maslach Burnout Inventory – Human Services (MBI) version is a 22-item scale assessing how persons in the human services, or helping professionals, view their job and the people with whom they work closely.[1] Each item is a statement (e.g., I feel emotionally drained from my work). The respondent rates how often each item (statement) is true, using the scores 0 to 6 where 0=never and 6=every day. Three subscales are scored: Emotional exhaustion (9 items; scored 0-54 where higher scores indicate greater emotional exhaustion), Depersonalization (5 items; scored 0-30 where higher scores indicate greater depersonalization; 5 items), and Personal accomplishment (8 items; scored 0-48 where higher scores indicate greater personal accomplishment).

Missingness in the trial year: Interns: 1228 Flexible group interns were sent the survey; 594 (48.4%) completed all MBI items and 634 (51.6%) did not complete any MBI items. 1300 Standard group interns were sent the survey; 563 (43.3%) completed all MBI items and 737 (56.7%) did not complete any MBI items. All trainees: 3099 Flexible group trainees were sent the survey; 1340 (43.2%) completed all MBI items, 5 (0.2%) trainees provided some but not all MBI items, and 1754 (56.6%) did not complete any MBI items. 3214 Standard group trainees were sent the survey; 1312 (40.8%) completed all MBI items, 1 (0.0%) trainee provided some but not all MBI items, and 1901 (59.1%) did not complete any MBI items. Analysis for a subscale included respondents who provided at least one item for that subscale.

*Residency programs assigned to flexible policies were allowed to waive limits on maximum shift length and mandatory time off between shifts.

†The marginal percents, random (program) variance, Flexible versus Standard odds ratio, and associated P value were obtained from a mixed effects logistic regression model with a random intercept (1 random term for clustering of respondents within program) and an indicator term for the duty-hour policy group (1=Flexible, 0=Standard) as the only fixed effect covariate in the model and the respondent's dichotomized subscale score (worst 2 categories versus best) as the outcome. A separate model was fit for each subscale. The marginal percentage is similar to the observed percentage but is derived from the regression model and accounts for correlations between respondents at the same program, averaging across random effects due to variation in respondent outcomes within programs. Note the P value for the hypothesis that the between-programs variance equaled 0 (i.e., no within programs correlations) was <0.05 for all outcomes except personal accomplishment (interns only analysis), reflecting large variation in how interns perceive duty-hours impact across all programs, independent of assignment to flexible or standard.

‡The random (program) variance, Flexible versus Standard odds ratio, and associated P values were obtained from a mixed effects logistic regression model with an intercept (1 random term for clustering of respondents within program) and an indicator term for the duty-hour policy group (1=Flexible, 0=Standard) and the program's percent of respondents analyzed (interns only or all trainees) scoring in the worst 2 categories in the baseline year as the only fixed effects covariates and the respondent's dichotomized subscale score (worst 2 categories versus best) as the outcome. A separate model was fit for each subscale. The marginal percentage is similar to the observed percentage but is derived from the regression model and accounts for correlations between respondents at the same program, averaging across random effects due to variation in respondent outcomes within programs. Note the P value for the hypothesis that the between-programs variance equaled 0 (i.e., no within programs correlations) was <0.05 for all outcomes except for depersonalization (interns only) and personal accomplishment (both interns only and all trainees), reflecting large variation in how interns perceive duty-hours impact across all programs, independent of assignment to flexible or standard.

Table S9. End-of-Year Survey of All Trainees.

	Flexible* (N = 30 programs; 1435 trainees)	Standard (N = 31 programs; 1411 trainees)	Flexible versus Standard	
			Odds Ratio† (95% CI)	Odds Ratio Adjusted for Program Response in Baseline Year‡ (95% CI)
Trainees having perception of negative effect¶ of institutional duty hours on:				
Safety of patient care (q4a)				
Observed %	14.0%	7.8%		
Marginal %†	13.1%	7.3%	1.92 (1.36, 2.71)	1.98 (1.40, 2.81)
Random (program) variance (P value) †‡			0.18 (<0.001)	0.15 (<0.001)
Continuity of care (q4b)				
Observed %	14.9%	33.6%		
Marginal %†	12.6%	28.0%	0.37 (0.25, 0.55)	0.38 (0.26, 0.54)
Random (program) variance (P value) †‡			0.43 (<0.001)	0.28 (<0.001)
Ability to acquire clinical skills (q4d)				
Observed %	9.9%	13.9%		
Marginal %†	9.5%	11.5%	0.81 (0.55, 1.19)	0.82 (0.55, 1.21)
Random (program) variance (P value) †‡			0.29 (<0.001)	0.25 (<0.001)
Trainee autonomy (q4f)				
Observed %	6.0%	8.8%		
Marginal %†	5.6%	7.8%	0.69 (0.46, 1.04)	0.71 (0.47, 1.07)
Random (program) variance (P value) †‡			0.24 (<0.001)	0.21 (0.001)
Availability for urgent patient care encounters (q4i)				
Observed %	8.1%	9.6%		
Marginal %†	7.9%	9.4%	0.83 (0.62, 1.11)	0.86 (0.63, 1.19)
Random (program) variance (P value) †‡			0.05 (0.13)	0.05 (0.15)
Availability for elective patient care encounters (q4h)				
Observed %	15.3%	15.5%		
Marginal %†	12.8%	13.9%	0.91 (0.63, 1.30)	0.95 (0.69, 1.31)
Random (program) variance (P value)			0.27 (<0.001)	0.16 (<0.001)

	Flexible* (N = 30 programs; 1435 trainees)	Standard (N = 31 programs; 1411 trainees)	Flexible versus Standard	
			Odds Ratio† (95% CI)	Odds Ratio Adjusted for Program Response in Baseline Year‡ (95% CI)
Ability to attend required educational conferences (q4c)				
Observed %	20.1%	12.7%		
Marginal %†	17.9%	11.2%	1.72 (1.24, 2.38)	1.61 (1.18, 2.18)
Random (program) variance (P value) †‡			0.20 (<0.001)	0.12 (<0.001)
Relationship between interns and residents (q4k)				
Observed %	8.7%	5.5%		
Marginal %†	8.1%	5.1%	1.64 (1.14, 2.36)	1.52 (1.02, 2.26)
Random (program) variance (P value) †‡			0.13 (0.004)	0.10 (0.02)
Time for teaching medical students (q4j)				
Observed %	20.6%	18.3%		
Marginal %†	19.3%	16.9%	1.17 (0.87, 1.58)	1.15 (0.86, 1.55)
Random (program) variance (P value) †‡			0.16 (<0.001)	0.12 (<0.001)
Need to perform patient care related work outside of the hospital (q5a)				
Observed %	22.7%	24.8%		
Marginal %†	21.7%	23.0%	0.92 (0.69, 1.23)	0.89 (0.68, 1.18)
Random (program) variance (P value) †‡			0.17 (<0.001)	0.12 (<0.001)
Ability to participate in research (q5c)				
Observed %	21.4%	11.1%		
Marginal %†	20.6%	10.8%	2.14 (1.60, 2.87)	1.84 (1.30, 2.60)
Random (program) variance (P value) †‡			0.13 (<0.001)	0.12 (<0.001)
Professionalism (q4l)				
Observed %	8.0%	3.6%		
Marginal %†	7.7%	3.5%	2.33 (1.61, 3.37)	2.50 (1.72, 3.63)
Random (program) variance (P value) †‡			0.05 (0.18)	0.03 (0.28)
Job satisfaction (q5d)				

	Flexible* (N = 30 programs; 1435 trainees)	Standard (N = 31 programs; 1411 trainees)	Flexible versus Standard	
			Odds Ratio† (95% CI)	Odds Ratio Adjusted for Program Response in Baseline Year‡ (95% CI)
Observed %	20.4%	7.5%		
Marginal %†	19.0%	6.9%	3.17 (2.30, 4.37)	3.18 (2.30, 4.40)
Random (program) variance (P value) †‡			0.14 (<0.001)	0.07 (0.03)
Satisfaction with career choice (q5e)				
Observed %	16.9%	6.3%		
Marginal %†	16.1%	6.0%	3.01 (2.20, 4.11)	3.16 (2.27, 4.39)
Random (program) variance (P value) †‡			0.09 (0.02)	0.07 (0.06)
Trainee morale (q4m)				
Observed %	24.7%	4.8%		
Marginal %†	21.9%	4.2%	6.37 (4.12, 9.86)	6.72 (4.40, 10.26)
Random (program) variance (P value) †‡			0.36 (<0.001)	0.22 (<0.001)
Time with family and friends (q5f)				
Observed %	28.9%	7.1%		
Marginal % †	27.3%	6.9%	5.07 (3.57, 7.20)	4.43 (3.05, 6.43)
Random (program) variance (P value) †‡			0.22 (<0.001)	0.17 (<0.001)
Time for hobbies and outside interests (q5g)				
Observed %	27.7%	7.5%		
Marginal %†	25.6%	7.4%	4.32 (2.98, 6.27)	3.66 (2.48, 5.40)
Random (program) variance (P value) †‡			0.28 (<0.001)	0.20 (<0.001)
Health (q5h)				
Observed %	26.1%	6.7%		
Marginal %†	24.1%	6.5%	4.60 (3.16, 6.69)	3.70 (2.44, 5.61)
Random (program) variance (P value) †‡			0.27 (<0.001)	0.19 (<0.001)
Ability to acquire clinical reasoning skills (q4e)				
Observed %	8.2%	8.6%		
Marginal %†	7.6%	7.2%	1.05 (0.71, 1.57)	1.06 (0.71, 1.57)
Random (program) variance (P value) †‡			0.25 (<0.001)	0.20 (<0.001)

	Flexible* (N = 30 programs; 1435 trainees)	Standard (N = 31 programs; 1411 trainees)	Flexible versus Standard	
			Odds Ratio† (95% CI)	Odds Ratio Adjusted for Program Response in Baseline Year‡ (95% CI)
Pace of trainee's work day (q5b)				
Observed %	18.7%	24.2%		
Marginal %†	16.8%	19.8%	0.81 (0.57, 1.17)	0.80 (0.57, 1.11)
Random (program) variance (P value) †‡			0.31 (<0.001)	0.20 (<0.001)
Trainee's overall well-being (q5j)				
Observed %	23.3%	5.8%		
Marginal %†	21.5%	5.6%	4.65 (3.18, 6.80)	4.15 (2.76, 6.23)
Random (program) variance (P value) †‡			0.25 (<0.001)	0.19 (<0.001)
Trainees reporting dissatisfaction§ with:				
Overall quality of resident education (q7f)				
Observed %	13.4%	7.7%		
Marginal %†	11.1%	6.6%	1.75 (1.11, 2.75)	1.53 (1.04, 2.27)
Random (program) variance (P value)†‡			0.44 (<0.001)	0.20 (0.001)
Overall well-being (q7h)				
Observed %	26.3%	14.6%		
Marginal %†	25.3%	14.2%	2.04 (1.48, 2.81)	1.76 (1.28, 2.42)
Random (program) variance (P value)†‡			0.21 (<0.001)	0.14 (<0.001)
Patient safety (q7b)				
Observed %	5.2%	3.8%		
Marginal %†	5.0%	3.5%	1.44 (0.94, 2.20)	1.38 (0.87, 2.18)
Random (program) variance (P value) †‡			0.13 (0.07)	0.13 (0.08)
Continuity of care (q7a)				
Observed %	5.8%	8.5%		
Marginal % †	5.4%	7.7%	0.69 (0.48, 1.01)	0.59 (0.40, 0.85)
Random (program) variance (P value) †‡			0.14 (0.02)	0.10 (0.05)

	Flexible* (N = 30 programs; 1435 trainees)	Standard (N = 31 programs; 1411 trainees)	Flexible versus Standard	
			Odds Ratio† (95% CI)	Odds Ratio Adjusted for Program Response in Baseline Year‡ (95% CI)
Quality and ease of handoffs and transitions in care (q7e)				
Observed %	6.1%	7.1%		
Marginal %†	5.9%	6.8%	0.85 (0.61, 1.20)	0.70 (0.48, 1.02)
Random (program) variance (P value) †‡			0.07 (0.09)	0.20 (0.25)
Duty-hour regulations of the program (q7i)				
Observed %	13.3%	5.5%		
Marginal %†	12.3%	5.0%	2.66 (1.79, 3.95)	2.64 (1.79, 3.88)
Random (program) variance (P value) †‡			0.24 (<0.001)	0.13 (0.005)
Work hours and scheduling (q7d)				
Observed %	20.0%	11.3%		
Marginal %†	19.1%	10.8%	1.95 (1.41, 2.70)	1.91 (1.37, 2.68)
Random (program) variance (P value) †‡			0.19 (<0.001)	0.15 (<0.001)
Time for rest (q7g)				
Observed %	29.8%	17.4%		
Marginal %†	28.4%	16.8%	1.96 (1.44, 2.66)	1.70 (1.26, 2.30)
Random (program) variance (P value) †‡			0.20 (<0.001)	0.12 (<0.001)
Level of attending supervision (q7c)				
Observed %	2.8%	2.2%		
Marginal %†	2.7%	2.1%	1.27 (0.75, 2.14)	1.39 (0.79, 2.44)
Random (program) variance (P value) †‡			0.14 (0.21)	0.16 (0.20)
Ability to follow the clinical care of the patients the trainee admits (q7j)				
Observed %	5.3%	6.7%		
Marginal %†	5.1%	6.1%	0.83 (0.57, 1.22)	0.81 (0.55, 1.18)
Random (program) variance (P value) †‡			0.11 (0.03)	0.05 (0.21)

	Flexible* (N = 30 programs; 1435 trainees)	Standard (N = 31 programs; 1411 trainees)	Flexible versus Standard	
			Odds Ratio† (95% CI)	Odds Ratio Adjusted for Program Response in Baseline Year‡ (95% CI)
The number of patients the trainee got to admit completely (i.e., someone else did not start of complete the task of admitting the patient) (q7k)				
Observed %	5.3%	9.8%		
Marginal %†	4.5%	7.9%	0.56 (0.35, 0.89)	0.64 (0.40, 1.02)
Random (program) variance (P value) †‡			0.39 (<0.001)	0.32 (<0.001)
Trainees perceiving that their fatigue‖:				
Almost always or often affected their personal safety (q8a)				
Observed %	14.0%	10.6%		
Marginal %†	14.4%	10.7%	1.40 (0.99, 2.00)	1.40 (1.02, 1.93)
Random (program) variance (P value) †‡			0.23 (<0.001)	0.13 (0.001)
Almost always or often affected patient safety (q8b)				
Observed %	10.4%	7.4%		
Marginal %†	10.4%	7.5%	1.44 (1.02, 2.02)	1.37 (0.97, 1.94)
Random (program) variance (P value) †‡			0.15 (0.007)	0.11 (0.02)
Trainees reporting at least 1 occurrence during their most recent month on a medicine floor rotation:**				
Left during a patient encounter because of duty hour limits (q3c)				
Observed %	7.5%	6.4%		
Marginal %†	6.9%	6.2%	1.12 (0.76, 1.65)	1.11 (0.75, 1.63)
Random (program) variance (P value) †‡			0.21 (0.001)	0.12 (0.04)

	Flexible* (N = 30 programs; 1435 trainees)	Standard (N = 31 programs; 1411 trainees)	Flexible versus Standard	
			Odds Ratio† (95% CI)	Odds Ratio Adjusted for Program Response in Baseline Year‡ (95% CI)
Missed a patient encounter because of duty hour limits (q3d)				
Observed %	27.1%	20.1%		
Marginal %†	20.6%	18.8%	1.12 (0.75, 1.67)	1.10 (0.80, 1.50)
Random (program) variance (P value) †‡			0.42 (<0.001)	0.17 (<0.001)
Handed off an active patient care issue because of duty hour limits (q3b)				
Observed %	35.5%	36.9%		
Marginal %†	31.7%	36.2%	0.82 (0.59, 1.13)	0.81 (0.65, 1.03)
Random (program) variance (P value) †‡			0.27 (<0.001)	0.08 (<0.001)
Left or missed educational conferences during a scheduled shift because of duty hour limits (q3a)				
Observed %	33.3%	29.5%		
Marginal %†	31.6%	29.9%	1.08 (0.82, 1.43)	1.03 (0.81, 1.32)
Random (program) variance (P value) †‡			0.17 (<0.001)	0.08 (<0.001)
Worked more than 16 hours continuously in house (q3e)				
Observed %	47.8%	25.5%		
Marginal %†	41.9%	23.0%	2.41 (1.51, 3.86)	2.40 (1.77, 3.24)
Random (program) variance (P value) †‡			0.70 (<0.001)	0.18 (<0.001)
Had less than 8 hours off between daily shifts (q3f)				
Observed %	28.9%	30.1%		
Marginal %†	26.1%	26.5%	0.98 (0.67, 1.44)	0.92 (0.70, 1.20)
Random (program) variance (P value) †‡			0.41 (<0.001)	0.12 (<0.001)

Trainees were asked to score 40 aspects of their trainee experiences. This table reports results for all trainees; Table 4 reports results for interns only. For each item, the response choices were dichotomized into a binary response as indicated for each question theme (response choices for each question theme are indicated in the footnote on the theme heading).

Missingness in the trial year: 3099 Flexible group trainees were sent the survey; 1402 (45.2%) answered every survey question, 33 (1.1%) answered 1 or more but not all questions, and 1664 (53.7%) answered no questions. 3214 Standard group trainees were sent the survey; 1380 (42.9%) answered every survey question, 31 (1.0%) answered 1 or more but not all questions, and 1803 (56.1%) answered no questions. Observed percentages are the percentages of trainees answering the question.

*Residency programs assigned to flexible policies were allowed to waive limits on maximum shift length and mandatory time off between shifts.

†The marginal percent, random (program) variance, Flexible versus Standard odds ratio, and associated P value were obtained from a mixed effects logistic regression model with a random intercept (1 random term for clustering of respondents within program) and an indicator term for the duty-hour policy group (1=Flexible, 0=Standard) as the fixed effect covariate in the model and the respondent's dichotomized response as the outcome. A separate mixed model was fit for each survey question shown. The marginal percentage is similar to the observed percentage but is derived from the regression model and accounts for correlations between respondents at the same program, averaging across random effects due to variation in respondent outcomes within programs. Note the P value for the hypothesis that the between-programs variance equaled 0 (i.e., no within programs correlations) was <0.05 for all outcomes except for availability for urgent patient care encounters, professionalism, dissatisfaction with patient safety, quality and ease of patient handoffs and transitions in care, and level of attending supervision, reflecting large variation in how trainees perceive duty-hours impact across all programs, independent of assignment to flexible or standard.

‡The random (program) variance, Flexible versus Standard odds ratio, and associated P value were obtained from a mixed effects logistic regression model with a random intercept (1 random term for clustering of respondents within programs) and an indicator term for the duty-hour policy group (1=Flexible, 0=Standard) and the program's percent responding noncompliant in 2015 as the fixed effects covariates in the model and the respondent's dichotomized response for the item as the outcome. A separate mixed model was fit for each survey question shown. The marginal percentage is similar to the observed percentage but is derived from the regression model and accounts for correlations between respondents at the same program, averaging across random effects due to variation in respondent outcomes within programs. Note the P value for the hypothesis that the between-programs variance equaled 0 (i.e., no within programs correlations) was <0.05 for all outcomes except for availability for urgent patient care encounters, professionalism, satisfaction with career choice, dissatisfaction with patient safety, dissatisfaction with continuity of care, quality and ease of patient handoffs and transitions in care, level of attending supervision, and ability to follow the clinical care of the patients the trainee admits, reflecting large variation in how trainees perceive duty-hours impact across all programs, independent of assignment to flexible or standard.

§Very dissatisfied or dissatisfied versus neutral, satisfied or very satisfied.

¶Negative effect versus no effect or positive effect.

‖Always or often versus sometimes, rarely or never.

**1 or more occurrences in the past month versus no occurrence.

Table S10. End-of-Shift Surveys of Trainees' Experience with Education, Sense of Ownership, Work Intensity, and Continuity.

	Flexible* (N = 29 programs)	Standard (N = 31 programs)	Flexible versus Standard Difference† (95% CI)
Interns	N=877	N=869	
Time for educational conference and related activities			
Observed mean ± SD‡	1.64 ± 0.41	1.61 ± 0.40	
Marginal mean†	1.65	1.62	0.03 (-0.04, 0.11)
Random effect variance†			
Program (P value)			0.016 (<0.001)
Error			0.15
Sense of ownership of patients			
Observed mean ± SD‡	1.91 ± 0.28	1.91 ± 0.28	
Marginal mean†	1.91	1.91	-0.00 (-0.03, 0.03)
Random effect variance†			
Program (P value)			0.0016 (0.002)
Error			0.08
Work intensity			
Observed mean ± SD¶	2.24 ± 0.38	2.21 ± 0.36	
Marginal mean†	2.25	2.21	0.03 (-0.02, 0.08)
Random effect variance†			
Program (P value)			0.0039 (<0.001)
Error			0.13
Continuity of care			
Observed mean ± SD‡	1.84 ± 0.30	1.83 ± 0.30	
Marginal mean†	1.84	1.84	0.00 (-0.03, 0.04)
Random effect variance†			
Program (P value)			0.0012 (0.02)
Error			0.09
All trainees	N=1958	N=1942	
Time for educational conference and related activities			
Observed mean ± SD‡	1.64 ± 0.41	1.65 ± 0.40	
Marginal mean†	1.66	1.65	0.01 (-0.05, 0.07)
Random effect variance†			
Program (P value)			0.012 (<0.001)
Error			0.15
Sense of ownership of patients			
Observed mean ± SD‡	1.92 ± 0.27	1.93 ± 0.28	
Marginal mean†	1.92	1.93	-0.01 (-0.04, 0.02)
Random effect variance†			
Program (P value)			0.0017 (<0.001)
Error			0.07
Work intensity			

	Flexible* (N = 29 programs)	Standard (N = 31 programs)	Flexible versus Standard Difference† (95% CI)
Observed mean ± SD‡	2.21 ± 0.38	2.19 ± 0.36	
Marginal estimate†	2.22	2.19	0.02 (-0.02, 0.06)
Random effect variance†			
Program (P value)			0.0036 (<0.001)
Error			0.13
Continuity of care			
Observed mean ± SD‡	1.85 ± 0.30	1.84 ± 0.30	
Marginal estimate†	1.85	1.85	-0.00 (-0.03, 0.03)
Random effect variance†			
Program (P value)			0.0014 (<0.001)
Error			0.09

Every 2 weeks from September 2015 through April 2016 (16 cycles), each trainee was asked to rate the queried aspect of their experience as too little, just right, or too much (scored 1, 2, 3, respectively). The quantity analyzed was the trainee's mean rating of the aspect over all cycles in which the trainee participated.

Missingness: Interns: 1228 Flexible group interns were sent the surveys; 877 (71.4%) responded to at least 1 survey and answered all 4 survey questions, 1 (0.1%) responded to the education and ownership questions only, and 350 (28.5%) did not complete any survey questions. 1300 Standard group interns were sent the surveys; 869 (66.8%) responded to at least 1 survey and answered all 4 survey questions, 3 (0.2%) responded to the education and ownership questions only, and 428 (32.9%) did not complete any survey questions. All trainees: 3017 Flexible group trainees were sent the surveys; 1936 (64.2%) responded to at least 1 survey and answered all 4 survey questions, 6 (0.2%) responded to the education and ownership questions only, and 1075 (35.6%) did not complete any survey questions. 3214 Standard group trainees were sent the surveys; 1951 (60.7%) responded to at least 1 survey and answered all 4 survey questions, 7 (0.2%) responded to the education and ownership questions only, and 1256 (39.1%) did not complete any survey questions.

*Residency programs assigned to flexible policies were allowed to waive limits on maximum shift length and mandatory time off between shifts.

†The marginal mean, random effect variance, Flexible versus Standard difference, and associated P value were obtained from a mixed effects linear regression model with random intercepts (1 fixed term and 1 random term for clustering of respondents within program) and an indicator term for the duty-hour policy group (1=Flexible, 0=Standard) as the fixed effect covariate in the model and the respondent's mean rating over the survey cycles in which the respondent participated as the outcome. A separate mixed model was fit for each question. The marginal mean is similar to an observed mean but is derived from the regression model and accounts for correlations between respondents at the same program, averaging across random effects due to variation in respondent outcomes within programs. Note the P value for the hypothesis that the between-programs variance equaled 0 (i.e., no within programs correlations) was <0.05 for all outcomes, reflecting large variation in respondents' ratings across all programs, independent of assignment to flexible or standard.

‡SDs are not adjusted for correlations between scores for respondents at the same program.

Table S11. 2016 ACGME Survey of Program Directors and Core Faculty.

Theme	Flexible*	Standard	Flexible* versus Standard: Difference† or Odds Ratio‡ (95% CI)	P Value †	Flexible* versus Standard: Difference§ or Odds Ratio¶ Adjusted for Program Response in Baseline Year (95% CI)	P Value§
Primary Outcome (difference)‖						
Trainees' workload exceeds capacity to do the work§						
Marginal mean†	4.22	4.18	0.04 (-0.07, 0.16)	0.46	0.06 (-0.03, 0.16)	0.19
Random effect variance†						
Program (P value)			0.03 (<0.001)		0.01 (0.01)	
Error			0.29		0.29	
Secondary Outcomes (odds ratio)**						
Supervision and teaching: marginal % noncompliant	22.6%	19.8%	1.19 (0.76, 1.86)		1.10 (0.74, 1.65)	
Educational content: marginal % noncompliant	21.4%	20.4%	1.06 (0.74, 1.53)		1.05 (0.79, 1.39)	
Resources: marginal % noncompliant	8.0%	7.5%	1.08 (0.62, 1.90)		0.98 (0.56, 1.70)	
Patient safety: marginal % noncompliant	24.5%	25.5%	0.95 (0.64, 1.40)		0.94 (0.67, 1.32)	
Teamwork: marginal % noncompliant	2.9%	3.1%	0.93 (0.35, 2.48)		0.89 (0.35, 2.29)	
Overall evaluation of program: marginal % noncompliant	0.4%	0.4%	0.94 (0.10, 9.08)		--††	

Respondents were program directors and faculty at 32 Flexible programs and 31 Standard programs.

Missingness in the trial year: The ACGME reported response rates of 91% flexible and 91% standard.

*Residency programs assigned to flexible policies were allowed to waive limits on maximum shift length and mandatory time off between shifts.

†The marginal mean, Random effect variance, Flexible versus Standard difference in score, and associated P values were obtained from a mixed effects linear regression model with an intercept (1 random term for clustering of respondents within program) and an indicator term for the duty-hour policy group (1=Flexible, 0=Standard) as the only fixed effect covariate in the model and the program director's or faculty member's ordinal score as the outcome. The marginal mean is similar to the observed mean but is derived from the regression model and accounts for correlations between respondents at the same program, averaging across random effects due to variation in respondent outcomes within

programs. The P value for the random effect variance tests the hypothesis that there was no clustering effect of programs on the outcome. Note that P value for between-program variance <0.05 reflects large variation in how program directors and core faculty responded across all programs, independent of assignment to flexible or standard.

‡The marginal percent and Flexible versus Standard odds ratio of a noncompliant response were obtained from a logistic regression model with generalized estimating equations and robust variance estimation using an independent working correlation matrix to account for the correlations between responses from respondents at the same program; the model included an indicator term for the duty-hour policy group (1=Flexible, 0=Standard) as the covariate and the program director's or faculty member's dichotomized response for the theme as the outcome. A separate model was fit for each theme. The marginal percentage is similar to the observed percentage but is derived from the regression model and accounts for correlations between respondents at the same program.

§The marginal mean, Random effect variance, Flexible versus Standard difference in score, and associated P values were obtained from a mixed effects linear regression model with an intercept (1 random term for clustering of respondents within program) and an indicator term for the duty-hour policy group (1=Flexible, 0=Standard) and the program's mean response in the baseline year as the fixed effects covariates and the program director's or faculty member's ordinal score as the outcome. The marginal mean is similar to the observed mean but is derived from the regression model and accounts for correlations between respondents at the same program, averaging across random effects due to variation in respondent outcomes within programs. The P value for the random effect variance tests the hypothesis that there was no clustering effect of programs on the outcome. Note that P value for between-program variance <0.05 reflects large variation in how program directors and core faculty responded across all programs, independent of assignment to flexible or standard.

¶ The marginal percent and Flexible versus Standard odds ratio of a noncompliant response were obtained from a logistic regression model with generalized estimating equations and robust variance estimation using an independent working correlation matrix to account for the correlations between responses from respondents at the same program; the model included an indicator term for the duty-hour policy group (1=Flexible, 0=Standard) and the program's mean response in the baseline year as the only covariates and the respondent's dichotomized response for the theme as the outcome. A separate model was fit for each theme. The marginal percentage is similar to the observed percentage but is derived from the regression model and accounts for correlations between respondents at the same program.

‖The primary outcome for the hypothesis about program director and faculty satisfaction was a single question from the ACGME faculty survey: "Residents' clinical workload exceeds their capacity to do the work." Response options were Very Often (scored 1), Often (scored 2), Sometimes (scored 3), Rarely (scored 4), and Never (scored 5) and the ordinal score was analyzed.

**Additional ACGME faculty survey measures were secondary outcomes. Each secondary outcome is the respondent's dichotomized response (noncompliant versus compliant) for the specified theme. Each theme comprises 1-5 survey questions. Each survey item had 5 response choices which were dichotomized into a binary response for the item. The respondent's dichotomized responses across the survey items comprising the

theme were pooled to provide a theme-level binary response of noncompliant versus compliant. The response for a theme is noncompliant if the respondent provided a noncompliant response to any of the questions comprising the theme. See online supplement for details regarding questions included in each theme.

††Model did not converge.

Table S12. End-of-Year Survey of Program Directors: Details and Expanded Analyses.

	Flexible* Observed % (32 programs)	Standard Observed % (30 programs)	Flexible versus Standard: Odds Ratio† (95% CI)	Flexible versus Standard: Odds Ratio Adjusted for Response in Baseline Year‡ (95% CI)
Dissatisfaction with learning environment				
1. Intern ownership of patient care	0%	23.3%	0.08 (0.00, 0.57)	0.25 (0.00, 2.43)
2. 1Resident ownership of patient care	0%	10.0%	0.23 (0.00, 2.22)	0.40 (0.00, 5.27)
3. Ability of interns to manage the patients they admit	0%	13.3%	0.16 (0.00, 1.36)	0.18 (0.00, 1.85)
4. Intern morale	3.1%	26.7%	0.09 (0.00, 0.77)	0.13 (0.00, 1.00)
5. Resident morale	9.4%	20.0%	0.42 (0.06, 2.22)	0.40 (0.01, 5.77)
6. Time for trainees to reflect	25.0%	56.7%	0.26 (0.07, 0.84)	0.31 (0.06, 1.37)
7. Effectiveness of interns in performing clinical duties	0%	10.0%	0.23 (0.00, 2.22)	0.48 (0.00, 6.36)
8. Effectiveness of residents in performing clinical duties	0%	3.3%	0.94 (0.00, 36.6)	Not Calculable§
9. Ability of attending to provide real time feedback to interns on patient care activities	3.1%	63.3%	0.02 (0.00, 0.15)	0.03 (0.00, 0.18)
10. Ability of attending to provide real time feedback to residents on patient care activities	3.1%	43.3%	0.04 (0.00, 0.34)	0.06 (0.00, 0.40)
11. Frequency of handoffs	6.3%	66.7%	0.04 (0.00, 0.19)	0.06 (0.00, 0.35)
12. Quality of handoffs	12.5%	40.0%	0.22 (0.04, 0.87)	0.25 (0.02, 1.64)
13. Ability of residents to work in inter-professional teams	3.1%	3.3%	0.94 (0.01, 75.9)	1.12 (0.01, 93.2)
Dissatisfaction with workload				
1. Workload of faculty	29.0%	40.0%	0.62 (0.18, 2.02)	1.04 (0.21, 5.20)
2. Workload of residents	9.7%	33.3%	0.22 (0.03, 1.00)	0.44 (0.04, 3.34)
3. Workload of interns	6.5%	36.7%	0.12 (0.01, 0.66)	0.40 (0.03, 2.94)
4. Workload of program director	29.0%	43.3%	0.54 (0.16, 1.74)	0.74 (0.15, 3.39)
5. Opportunity for residents to transition care when fatigued	12.9%	13.8%	0.93 (0.15, 5.55)	1.26 (0.08, 20.1)

	Flexible* Observed % (32 programs)	Standard Observed % (30 programs)	Flexible versus Standard: Odds Ratio† (95% CI)	Flexible versus Standard: Odds Ratio Adjusted for Response in Baseline Year‡ (95% CI)
6. Ability of trainees to perform necessary work during the scheduled duty period	9.7%	55.2%	0.09 (0.01, 0.40)	0.05 (0.00, 0.46)
7. Reliance of residents to provide clinical service	29.0%	31.0%	0.91 (0.26, 3.18)	1.49 (0.30, 7.70)
Dissatisfaction with education opportunities				
1. Adequacy of time for bedside teaching for interns	12.9%	55.2%	0.13 (0.03, 0.49)	0.14 (0.02, 0.78)
2. Adequacy of time for bedside teaching of residents	12.9%	51.7%	0.14 (0.03, 0.56)	0.18 (0.03, 0.98)
3. Ability of interns to attend conferences while on inpatient rotations	19.4%	55.2%	0.20 (0.05, 0.70)	0.24 (0.03, 1.37)
4. Ability of residents to attend conferences while on inpatient rotations	16.7%	37.9%	0.33 (0.08, 1.27)	0.39 (0.05, 2.42)
5. Ability of interns to participate in attending teaching rounds	6.7%	31.0%	0.16 (0.02, 0.91)	0.14 (0.00, 1.36)
6. Ability of residents to participate in attending teaching rounds	6.7%	17.2%	0.35 (0.03, 2.37)	0.33 (0.01, 4.41)
7. Ability of interns to attend family meetings	6.7%	17.2%	0.35 (0.03, 2.37)	0.20 (0.00, 1.92)
8. Ability of residents to attend family meetings	3.3%	6.9%	0.47 (0.01, 9.54)	0.50 (0.00, 6.50)
9. Balance of service vs. education for interns	16.7%	34.5%	0.39 (0.09, 1.49)	0.97 (0.14, 6.54)
10. Balance of service vs. education for residents	16.7%	24.1%	0.63 (0.14, 2.71)	1.40 (0.19, 11.4)
11. Elective rotation time for housestaff	16.7%	21.4%	0.74 (0.15, 3.36)	1.68 (0.15, 25.0)
12. Time for housestaff to do research	13.3%	27.6%	0.41 (0.08, 1.79)	0.33 (0.00, 3.52)
13. Time for housestaff to engage in medical student education or quality improvement	30.0%	17.2%	2.03 (0.51, 9.01)	1.49 (0.21, 11.8)
14. The amount of time housestaff need to spend on night rotations	3.3%	27.6%	0.09 (0.00, 0.79)	0.14 (0.00, 1.35)

	Flexible* Observed % (32 programs)	Standard Observed % (30 programs)	Flexible versus Standard: Odds Ratio† (95% CI)	Flexible versus Standard: Odds Ratio Adjusted for Response in Baseline Year‡ (95% CI)
Dissatisfaction with program administration and organization				
1. Financial support for nonteaching services	26.7%	57.1%	0.28 (0.08, 0.93)	0.39 (0.07, 1.88)
2. Financial support to hire incremental allied health professionals (e.g.--nurse practitioners) for clinical care delivery	50.0%	72.4%	0.39 (0.11, 1.27)	0.27 (0.05, 1.34)
3. Financial support to hire incremental hospitalists/additional faculty members for clinical care delivery	43.3%	65.5%	0.41 (0.12, 1.30)	0.28 (0.05, 1.34)
4. Relationship of residency program with hospital administration	3.3%	17.2%	0.17 (0.00, 1.67)	0.37 (0.00, 4.90)
5. Program director morale	6.7%	25.0%	0.22 (0.02, 1.31)	0.14 (0.00, 1.07)
6. Effort of tracking duty hours	33.3%	55.2%	0.41 (0.12, 1.31)	0.17 (0.02, 0.92)
Dissatisfaction with patient outcomes				
1. Continuity of care for patients	6.7%	51.7%	0.07 (0.01, 0.36)	0.09 (0.00, 0.81)
2. Safety of patients	0.0%	17.2%	0.13 (0.00, 0.98)	0.29 (0.00, 2.98)
3. Graduates' preparedness for practice after residency	0.0%	13.8%	0.17 (0.00, 1.40)	0.30 (0.00, 3.04)

Program directors were asked to score 43 aspects of the educational environment of their internal medicine training program on a scale of 1 to 5 where 1=Very Dissatisfied, 2=Dissatisfied, 3=Neutral, 4=Satisfied, and 5=Very Satisfied. For each item, these response choices were dichotomized into a binary response of very dissatisfied or dissatisfied versus other.

Missingness during the trial year: 32 Flexible group program directors were sent the survey; 30 (93.8%) program directors answered every survey question and 2 (6.3%) program directors answered at least 1 question but not all. 31 Standard group program directors were sent the survey; 26 (83.9%) program directors answered every survey question, 4 (12.9%) answered at least 1 question but not all, and 1 (3.2%) did no answer any questions. Percentages shown are the observed percentages of program directors answering the question.

*Residency programs assigned to flexible policies were allowed to waive limits on maximum shift length and mandatory time off between shifts.

†The Flexible versus Standard odds ratio was obtained from an exact logistic regression model with an indicator term for duty-hour policy group (1=Flexible, 0=Standard) as the only covariate and the program director's dichotomized trial year response for the item as the outcome. The conditional maximum likelihood estimate of the odds ratio is provided.

‡The Flexible versus Standard odds ratio was obtained from an exact logistic regression model with an indicator term for duty-hour policy group (1=Flexible, 0=Standard) and the program director's dichotomized baseline year response as the only covariates and the program director's dichotomized trial year response for the item as the outcome. The conditional maximum likelihood estimate of the odds ratio is provided. Because of a data acquisition error, baseline year survey data were available for 20 Flexible programs and 19 Standard programs, and the adjusted analysis is based on data from the 19 Flexible and 18 Standard programs for which both 2015 and 2016 data were available.

§All Flexible group responses in 2016 were negative and all Flexible group responses in 2015 were negative.

References cited in Online Supplement

1. Maslach C, Jackson SE, MP L. Maslach Burnout Inventory Manual. 3 ed. Palo Alto, CA: Consulting Psychologists Press; 1996.