Supplementary Appendix

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

Supplement to: Bilimoria KY, Chung JW, Hedges LV, et al. National cluster-randomized trial of duty-hour flexibility in surgical training. N Engl J Med 2016;374:713-27. DOI: 10.1056/NEJMoa1515724

SUPPLEMENTARY APPENDIX

National Cluster-Randomized Trial of Duty Hour Flexibility in Surgical Training

CONTENTS **PAGE**

A. STUDY DESIGN – ADDITIONAL DETAILS

Additional Information on ACS NSQIP

The American College of Surgeons National Surgical Quality Improvement Program (ACS NSQIP; www.acsnsip.org) is the largest and most prominent multispecialty surgical outcomes assessment program.[1,2] The structure of ACS NSQIP, including sampling strategy, data abstraction procedures, variables collected, risk-adjustment methodology, and outcomes collected, have been described extensively.[1-15] Briefly, the program collects detailed data regarding patient demographics, preoperative comorbidities and other risk factors, laboratory values, and operative details to allow comprehensive risk adjustment for hospital quality comparisons on more than 30 postoperative outcomes and some process measures.[1]

To standardize data collection across institutions, ACS NSQIP data are abstracted at each site by trained, certified, and audited surgical clinical reviewers (SCRs) who use highly standardized data definitions.[11] Patients are followed for complications for 30 days after the *index* operation irrespective of whether an inpatient, discharged to their home or another facility, or readmitted to another hospital.[16] The SCRs examine inpatient records, review outpatient physician office charts, and even contact patients directly to accurately assess postoperative outcomes.[13,17] ACS NSQIP data are clinical data collected by a trained abstractor for quality improvement, and these data are generally more accurate than administrative data for assessing postoperative complications.[18-20] The data have been shown to have excellent inter-rater reliability. Participating hospitals must ensure complete follow-up for 95% of cases. ACS NSQIP performs regular and event-driven audits to assess data integrity.

Most data points (except labs) are required before a hospital can submit a case to ACS NSQIP, thus there is very little missing data – typically 0% for non-lab values. All patient outcomes are required, so there are no missing patient outcome data. ACS NSQIP imputes data for 38 important patient characteristics and laboratory values. Missing data were imputed by ACS NSQIP using Buck's method, and this has been shown to be comparable to multiple imputation when modeling postoperative outcomes using ACS NSQIP data.[14]

For the purposes of the trial, we did not perform additional data audits, as the standard ACS NSQIP data audit process is thorough, well-tested, and thought to be sufficient.

1. Hall BL, Hamilton BH, Richards K, Bilimoria KY, Cohen ME, Ko CY. Does surgical quality improve in the American College of Surgeons National Surgical Quality Improvement Program: an evaluation of all participating hospitals. Ann Surg 2009;250:363-76.
2. Hall BL, Richards K, Ingraham A, Ko CY. New approaches to the National Surgical Quality Improvement Program: the American College of Surgeons experience. Am J Surg 2009;198:S56-62.
3. Werner RM, Bradlow ET. Relationship between Medicare's hospital compare performance measures and mortality rates. JAMA : the journal of the American Medical Association 2006;296:2694-702.
4. Khuri SF, Daley J, Henderson W, et al. The Department of Veterans Affairs' NSQIP: the first national, validated, outcome-based, risk-adjusted, and peer-controlled program for the measurement and enhancement of the quality of surgical care. National VA Surgical Quality Improvement Program. Ann Surg 1998;228:491-507.
5. Daley J, Khuri SF, Henderson W, et al. Risk adjustment of the postoperative morbidity rate for the comparative assessment of the quality of surgical care: results of the National Veterans Affairs Surgical Risk Study. J Am Coll Surg 1997;185:328-40.
6. Khuri SF, Henderson WG, Daley J, et al. The patient safety in surgery study: background, study design, and patient populations. J Am Coll Surg 2007;204:1089-102.
7. Davis CL, Pierce JR, Henderson W, et al. Assessment of the reliability of data collected for the Department of Veterans Affairs national surgical quality improvement program. J Am Coll Surg 2007;204:550-60.

8. Khuri SF, Henderson WG, Daley J, et al. Successful implementation of the Department of Veterans Affairs' National Surgical Quality Improvement Program in the private sector: the Patient Safety in Surgery study. Ann Surg 2008;248:329-36.
9. Dimick JB, Ghaferi AA, Osborne NH, Ko CY, Hall BL. Reliability Adjustment for Reporting Hospital Outcomes With Surgery. Ann Surg 2012.
10. Raval MV, Cohen ME, Ingraham AM, et al. Improving American College of Surgeons National Surgical Quality Improvement Program risk adjustment: incorporation of a novel procedure risk score. J Am Coll Surg 2010;211:715-23.
11. Shiloach M, Frencher SK, Jr., Steeger JE, et al. Toward robust information: data quality and inter-rater reliability in the American College of Surgeons National Surgical Quality Improvement Program. J Am Coll Surg 2010;210:6-16.
12. Birkmeyer JD, Shahian DM, Dimick JB, et al. Blueprint for a new American College of Surgeons: National Surgical Quality Improvement Program. J Am Coll Surg 2008;207:777-82.
13. Cohen ME, Dimick JB, Bilimoria KY, Ko CY, Richards K, Hall BL. Risk adjustment in the American College of Surgeons National Surgical Quality Improvement Program: a comparison of logistic versus hierarchical modeling. J Am Coll Surg 2009;209:687-93.
14. Hamilton BH, Ko CY, Richards K, Hall BL. Missing data in the American College of Surgeons National Surgical Quality Improvement Program are not missing at random: implications and potential impact on quality assessments. J Am Coll Surg 2010;210:125-39 e2.
15. Ingraham AM, Richards KE, Hall BL, Ko CY. Quality improvement in surgery: the American College of Surgeons National Surgical Quality Improvement Program approach. Advances in surgery 2010;44:251-67.
16. Bilimoria KY, Cohen ME, Ingraham AM, et al. Effect of postdischarge morbidity and mortality on comparisons of hospital surgical quality. Ann Surg 2010;252:183-90.
17. Cohen ME, Ko CY, Bilimoria KY, et al. Optimizing ACS NSQIP Modeling for Evaluation of Surgical Quality and Risk: Patient Risk Adjustment, Procedure Mix Adjustment, Shrinkage Adjustment, and Surgical Focus. J Am Coll Surg 2013.
18. Lawson EH, Hall BL, Louie R, et al. Association Between Occurrence of a Postoperative Complication and Readmission: Implications for Quality Improvement and Cost Savings. Ann Surg 2013;258:10-8.
19. Lawson EH, Louie R, Zingmond DS, et al. A comparison of clinical registry versus administrative claims data for reporting of 30-day surgical complications. Ann Surg 2012;256:973-81.
20. Haut ER, Pronovost PJ, Schneider EB. Limitations of administrative databases. JAMA : the journal of the American Medical Association 2012;307:2589; author reply -90.

Planned Interim Analysis

The FIRST Trial carried out a planned interim analysis in January 2015 to assess whether any differences could be observed in patient outcomes between study arms. Interim evaluation was based on a comparison of 30-day postoperative death or serious morbidity (primary patient outcome) across study arms using ACS NSQIP data from FIRST Trial hospitals spanning surgical cases from July through August 2014, as cases are not available for analysis until at least 120 days after the date of surgery.

The FIRST Trial Data Safety and Monitoring Board reviewed a confidential Interim Report in February 2015 and determined that the FIRST Trial could safely continue.

Statistical Significance Levels

The overall level of statistical significance for the study was set at $p<0.05$. Significance levels for the final analysis of patient outcomes (as reported in this paper) were adjusted to account for one planned interim analysis of patient outcomes that was carried out halfway through the study period (discussed above). Using the

method of Lan and DeMets (1983), we adjusted the p-value for one-sided tests in final analyses to $p<0.04$ in order to maintain an overall significance level of $p<0.05$ for the entire study.

The level of statistical significance for hypothesis tests concerning resident outcomes remained $p<0.05$ because resident outcomes could not be studied in the interim analysis.

Statistical Power and Sample Size

The FIRST Trial was powered on the basis of the primary patient outcome: 30-day postoperative death or serious morbidity.

The baseline rate of 30-day postoperative death or serious morbidity was 9.94% in 2012 and 9.82% in 2013 using data on general surgery ACS NSQIP cases from sample hospitals in 2012 and 2013. The study team defined the noninferiority margin to be an absolute difference of 1.25 percentage points in rates of 30-day postoperative death or serious morbidity, based on clinical judgment, trends in this ACS NSQIP outcome measure, power calculations, and intra-cluster correlations. Variance components models of 30-day postoperative death/serious morbidity were estimated with program-level and hospital-level random effects using these data. Models were estimated with and without stratifying programs by tertiles of 2012 observed rates of death or serious morbidity. The final variance components model that informed our power calculations was a 3-level hierarchical logistic regression model that included adjustment for baseline rates (in the form of tertiles of program-level rates of 30-day postoperative death/serious morbidity) with hospital (level 2) and residency program (level 3) random intercepts. Level-2 (hospital) and Level-3 (program) variances were estimated to be 0.062 and 2.44e-12, respectively.

Given baseline rates and estimated variance components, we calculated minimum sample size requirements to be 45 programs per arm with an average of 1.1 hospitals per program and 950 patients per hospital in order to obtain at least 80% power to detect a 1.25 percentage point absolute difference between study arms in rates of 30-day postoperative death or serious morbidity with $\alpha=0.04$. Additional details regarding power calculations are in the Study Protocol.

Randomization

The unit of randomization was the residency program. To improve statistical efficiency, we implemented a stratified, cluster-randomization strategy to balance residency programs across study arms with respect to key program characteristics. Using 2013 ACS NSQIP data, we calculated program-level aggregate 30-day postsurgical death/serious morbidity rates as the average of hospital-level rates across hospitals within a residency program. Residency programs in the study were stratified into three groups based on their tertile ranking with respect to rates of 30-day postoperative death/serious morbidity. Table S1 shows the mean and standard deviation for rates of 30-day postoperative death/serious morbidity by tertile. Table S2 shows the distribution of enrolled programs, hospitals and NSQIP cases from the final analytic sample across randomization tertiles.

Each residency program was assigned a unique, randomly-generated integer between 1,000 and 9,999 using a random-number generator. Within each stratum, residency programs were ordered in ascending order according to their randomly-assigned number. We created separate lists for each stratum containing only the randomly-generated number corresponding to programs within that stratum. These blinded lists were given to two Study Team members who alternately assigned the letters “A” and “B” to each number in each list. A coin

toss determined study arm assignment of letters. A third Study Team member matched the random numbers back to program identifiers.

Residency Program Directors and Program Coordinators were notified of their study arm assignment by electronic mail on April 1, 2014.

TABLE S1. Definition of Tertiles of 2013 Rates of 30-day Postoperative Death or Serious Morbidity: Variable Used for Stratified Randomization

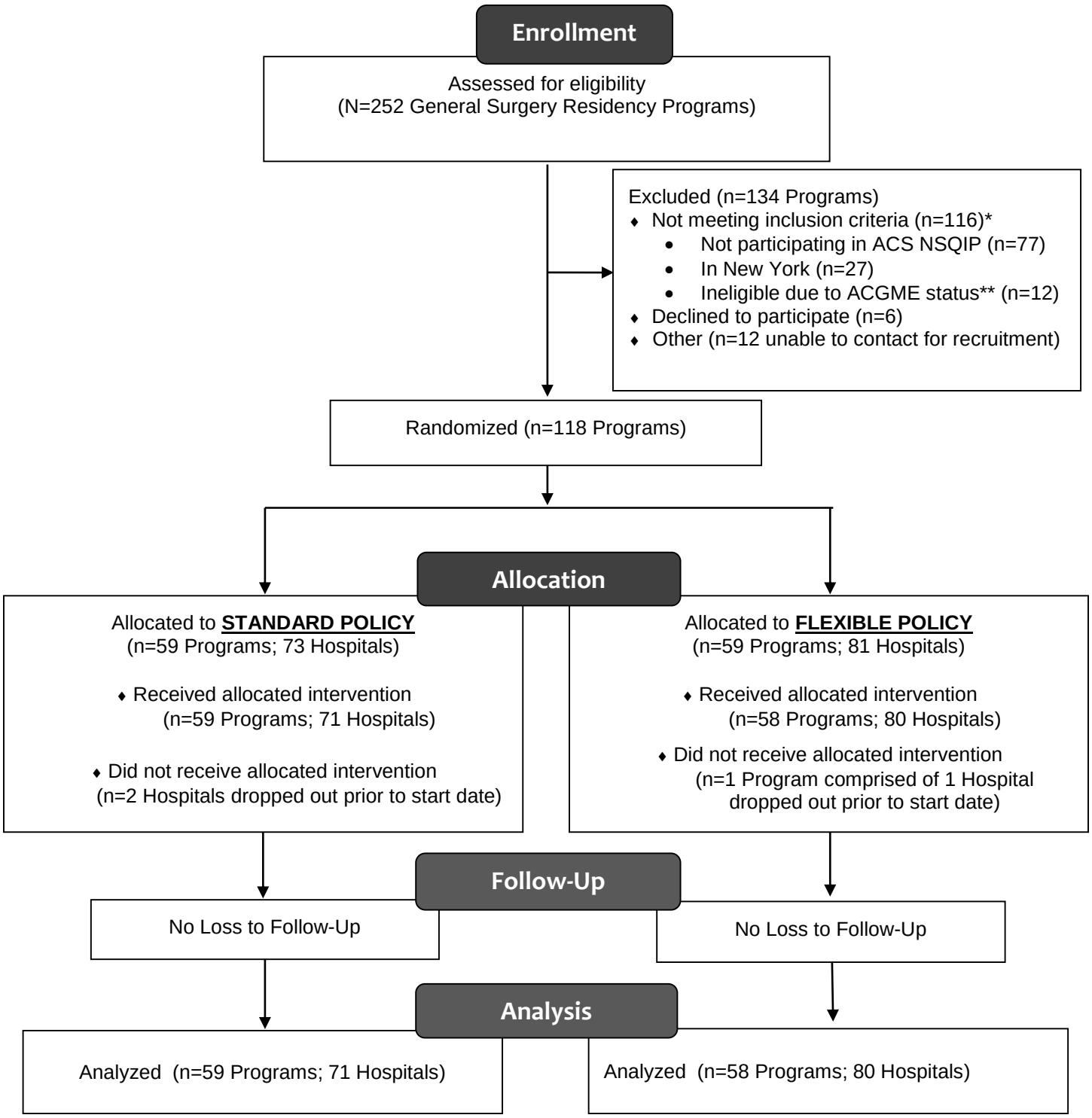
Tertile	N Programs	Observed Rate of 30-Day Postoperative Death or Serious Morbidity
		Mean (SD)
1	39	6.42% (1.40)
2	39	8.98% (0.77)
3	38	12.97% (2.48)

NOTE: The data used to define tertiles of 30-day postoperative death or serious morbidity for stratified randomization came from earlier data collected by the ACS NSQIP.

TABLE S2. Distribution of General Surgery Programs, Hospitals and ACS NSQIP Cases in Final Analytic Dataset by Tertiles of 2013 Rates of 30-day Postoperative Death or Serious Morbidity (Variable Used for Stratified Randomization)

Unit	Total N	Frequency (%)			
		No Baseline Data Available	Tertile 1	Tertile 2	Tertile 3
General Surgery Residency Programs					
Standard Policy	58	7 (12.07%)	19 (32.76%)	15 (25.86%)	17 (29.31%)
Flexible Policy	57	5 (8.77%)	17 (29.82%)	16 (28.07%)	19 (33.33%)
Total (Both Arms)	115	12 (10.43%)	36 (31.30%)	31 (26.96%)	36 (31.30%)
Hospitals					
Standard Policy	70	9 (12.86%)	23 (32.86%)	20 (28.57%)	18 (25.71%)
Flexible Policy	78	5 (6.41%)	24 (30.77%)	25 (32.05%)	24 (30.77%)
Total (Both Arms)	148	14 (9.46%)	47 (31.76%)	45 (30.41%)	42 (28.38%)
ACS NSQIP General Surgery Cases					
Standard Policy	65,849	8054 (12.23%)	22505 (34.18%)	17648 (26.80%)	17642 (26.79%)
Flexible Policy	72,842	4262 (5.85%)	21080 (28.94%)	21909 (30.08%)	25591 (35.13%)
Total (Both Arms)	138691	12316 (8.88%)	43585 (31.43%)	39557 (28.52%)	43233 (31.17%)

FIGURE S1. FIRST Trial CONSORT Diagram



* The exclusions were applied in this order, so each category is conditional on the prior category

**Ineligible due to being a new program or due to standing with ACGME (e.g., on probation due to duty hour violations)

*** In the final analysis of patient outcomes, two hospitals were excluded. One Standard Policy hospital was dropped by ACS NSQIP due to inadequate 30-day follow up of postoperative outcomes. Another hospital (Flexible Policy) changed their version of ACS NSQIP to one where the variables were no longer compatible for analysis. As this was the only hospital from that residency program, one Flexible Policy program was lost.

B. STUDY ADHERENCE

The FIRST Trial was designed and executed as a pragmatic trial, with no enforcement of adherence to study arm conditions. However, all 117 FIRST Trial general surgery residency Program Directors were surveyed in an effort to determine the extent to which programs adhered to study arm conditions.

Program adherence in the FIRST Trial was defined on the basis of Program Directors' responses to the following item in the 2015 FIRST Trial Program Directors Survey (Exhibit 2).

> *Which of the following statements are consistent with the formal duty hour policies and procedures for the general surgery residents at your institution during the FIRST Trial* [2014-2015]. *Please check all boxes that apply.*
>
> ☐ *PGY-1 resident duty periods can exceed 16 hours*
> ☐ *PGY-2 resident duty periods can exceed 28 hours (24 hours + 4 hours for transition)*
> ☐ *Residents do not require 14 hours off after continuous in-house duty of 24 hours*
> ☐ *Residents do not require 8-10 hours off between shifts*
> ☐ *None of the above apply to the formal policy at our institution*

Table S3 reports the frequency of reported departures from ACGME duty hour standards by FIRST Trial study arm. All but two programs in the Standard Policy arm adhered to the study conditions of their assigned arm. Two programs randomized to the Standard Policy arm indicated that their formal institutional duty hour policies in 2014-15 permitted residents fewer than 14 hours off following 24-hour in-house duty.

All (100%) Program Directors in Flexible Policy programs reported departures from ACGME standards regarding maximum shift length for PGY1 residents. A large proportion of program directors in Flexible Policy programs also reported departures from ACGME standards regarding maximum shift length for PGY2+ residents (84%), minimum time off after 24-hour shifts (88%), and minimum time off between shifts (81%).

TABLE S3. FIRST Trial Program Director's Survey (Summer 2015) Responses Regarding Departures from ACGME Duty Hour Standards during the 2014-2015 Academic Year

Duty Hour Requirement	Standard Policy N=59 programs	Flexible Policy (%) N=58 programs
PGY-1 resident duty periods can exceed 16 hours, n (%)	0 (0%)	58 (100%)
PGY-2 resident duty periods can exceed 28 hours (24 hours + 4 hours for transition), n (%)	0 (0%)	49 (84%)
Residents do not require 14 hours off after continuous in-house duty of 24 hours, n (%)	2 (3%)	51 (88%)
Residents do not require 8-10 hours off between shifts, n (%)	0 (0%)	47 (81%)

Table S4 reports the number of departures from ACGME duty hour standards by FIRST Trial study arm. Among programs randomized to Standard Policy, 57 (97%) reported no departures from ACGME standards while two programs (3%) reported a single departure from ACGME standards. Among programs randomized to Flexible Policy, all (100%) reported at least one departure from ACGME

standards; 3 (5%) reported one departure; 6 (10%) reported two departures; 6 (10%) reported three departures; and 43 (74%) reported four departures.

TABLE S4. Number of Departures from ACGME Duty Hour Standards during the 2014-2015 Academic Year, Based on FIRST Trial Program Director's Survey (Summer 2015)

Number of Departures from ACGME Duty Hour Standards (maximum=4)	Standard Policy N=59 programs	Flexible Policy (%) N=58 programs
Zero	57 (97%)	0 (0%)
One	2 (3%)	3 (5%)
Two	0 (0%)	6 (10%)
Three	0 (0%)	6 (10%)
Four	0 (0%)	43 (74%)

An *adherent program* was defined as a program that adhered to the study conditions of its assigned study arm. An adherent program in the Standard Policy arm was defined as a program that (a) was randomized to Standard Policy and (b) reported zero departures from current ACGME duty hour standards.

An adherent program in the Flexible Policy arm was defined as a program that (a) was randomized to Flexible Policy and (b) reported *at least one* departure from ACGME duty hour standards. Programs randomized to Flexible Policy were simply *permitted* to depart from a circumscribed set of ACGME duty hour standards (maximum shift length for PGY1 residents, maximum shift length for PGY2+ residents, minimum time off following 24-hour shifts, and/or minimum time off between shifts). Programs randomized to Flexible Policy *were not mandated* to depart from these ACGME duty hour standards – they merely had the sanctioned option to do so. Thus, a strict definition of adherence in the Flexible Policy arm identifies adherent programs as those which instituted at least one departure from ACGME standards within their institutional duty hour policies during the 2014-2015 year.

Table S5 shows program adherence status by FIRST Trial study arm assignment. Ninety-seven percent of Standard Policy programs were adherent. All programs in Flexible Policy were adherent.

TABLE S5. FIRST Trial Study Arm Adherence Rates

Adherence Status	Standard Policy N=59 programs	Flexible Policy N=58 programs	Total N=117 programs
Not Adherent	2 (3%)	0 (0%)	2 (2%)
Adherent	57 (97%)	58 (100%)	115 (98%)

C. STUDY ENDPOINTS

C.1. PATIENT ENDPOINTS

Table S6 lists the 11 patient endpoints that were studied in the FIRST Trial, along with brief definitions.

The primary patient endpoint in the FIRST Trial was the ACS NSQIP outcome, *30-Day Postoperative Death or Serious Morbidity.* The FIRST Trial was powered on the basis of this outcome.

All patient endpoints were standard ACS NSQIP outcomes, with the exception of failure-to-rescue (i.e., not reported to hospitals for quality improvement by ACS NSQIP).

Data for all patient endpoints came from ACS NSQIP data provided by hospitals in the FIRST Trial.

TABLE S6. FIRST Trial Study Endpoints – Patient Outcomes

PRIMARY/ SECONDARY ENDPOINT	ENDPOINT	DEFINITION
Primary	30-day postoperative (postop.) death or serious morbidity	All-cause mortality and/or serious morbidity within 30 days of surgical procedure (see definition of 'serious morbidity' below)
Secondary	30-day postop. death	All-cause mortality within 30 days of surgical procedure
Secondary	30-day postop. serious morbidity	Any of the following complications within 30 days of surgical procedure: organ space surgical site infection (no preoperative wound infection); wound dehiscence; stroke; myocardial infarction; cardiac arrest with cardiopulmonary resuscitation; pulmonary embolism; ventilation >48 hours (no preoperative ventilation); acute renal failure (no preoperative renal failure/dialysis); bleeding requiring transfusions >4 units; sepsis or septic shock
Secondary	30-day postop. any morbidity	Any of the following complications within 30 days of surgical procedure: any of the complications included in 'serious morbidity' or superficial or deep incisional surgical site infection (no preoperative wound infection); pneumonia (no preoperative pneumonia); unplanned intubation (no preoperative ventilation); progressive renal insufficiency (no preoperative dialysis/renal failure); urinary tract infection; deep vein thrombosis
Secondary	30-day postop. failure-to-rescue	Death in the presence of serious morbidity, within 30 days of surgical procedure (see definition of 'serious morbidity' above)
Secondary	30-day postop. pneumonia	Pneumonia (no preoperative pneumonia) within 30 days of surgical procedure
Secondary	30-day postop. renal failure	Renal failure (no preoperative renal failure or dialysis) within 30 days of surgical procedure
Secondary	30-day postop. surgery-related return to operating room	Return to operating room for reason related to index surgery within 30 days of the index surgical procedure
Secondary	30-day postop. sepsis	Sepsis or septic shock within 30 days of surgical procedure
Secondary	30-day postop. surgical site infection (SSI)	Superficial, deep incisional, or organ space surgical site infection (no preoperative wound infection) within 30 days of surgical procedure
Secondary	30-day postop. urinary tract infection (UTI)	Urinary tract infection within 30 days of surgical procedure

C.2. RESIDENT ENDPOINTS

Table S7 on the following page lists the 34 resident endpoints that were studied in the FIRST Trial, along with brief definitions.

The primary resident endpoints in the FIRST Trial were resident satisfaction with education quality, and resident satisfaction with overall wellbeing.

Data for all resident endpoints came from a resident survey administered by the American Board of Surgery (ABS) in conjunction with the January 2015 ABSITE examination. The survey was administered to all surgical residents who sat for the ABSITE examination, irrespective of whether they were training in a FIRST trial-participating program. The data were processed by the ABS and delivered to the study team for analysis.

TABLE S7. FIRST Trial Study Endpoints – Resident Outcomes

PRIMARY/ SECONDARY ENDPOINT	ENDPOINT	RESPONSE CATEGORIES	DICHOTOMIZED RESPONSE CATEGORIES
Primary	Resident satisfaction with overall education quality	5-Point Likert: Very Dissatisfied, Dissatisfied, Neutral, Satisfied, Very Satisfied	1 = Very Dissatisfied or Dissatisfied 0 = Neutral, Satisfied or Very Satisfied
Primary	Resident satisfaction with overall wellbeing	5-Point Likert: Very Dissatisfied, Dissatisfied, Neutral, Satisfied, Very Satisfied	1 = Very Dissatisfied or Dissatisfied 0 = Neutral, Satisfied or Very Satisfied
Secondary	Perceived effect of institutional duty hours on patient safety	Perceived Negative Effect; Perceived No Effect; Perceived Positive Effect	1 = Perceived Negative Effect 0 = Perceived No Effect or Positive Effect
Secondary	Perceived effect of institutional duty hours on continuity of care	Perceived Negative Effect; Perceived No Effect; Perceived Positive Effect	1 = Perceived Negative Effect 0 = Perceived No Effect or Positive Effect
Secondary	Perceived effect of institutional duty hours on conference attendance	Perceived Negative Effect; Perceived No Effect; Perceived Positive Effect	1 = Perceived Negative Effect 0 = Perceived No Effect or Positive Effect
Secondary	Perceived effect of institutional duty hours on clinical skills acquisition	Perceived Negative Effect; Perceived No Effect; Perceived Positive Effect	1 = Perceived Negative Effect 0 = Perceived No Effect or Positive Effect
Secondary	Perceived effect of institutional duty hours on resident autonomy	Perceived Negative Effect; Perceived No Effect; Perceived Positive Effect	1 = Perceived Negative Effect 0 = Perceived No Effect or Positive Effect
Secondary	Perceived effect of institutional duty hours on operative volume	Perceived Negative Effect; Perceived No Effect; Perceived Positive Effect	1 = Perceived Negative Effect 0 = Perceived No Effect or Positive Effect
Secondary	Perceived effect of institutional duty hours on availability for elective cases	Perceived Negative Effect; Perceived No Effect; Perceived Positive Effect	1 = Perceived Negative Effect 0 = Perceived No Effect or Positive Effect
Secondary	Perceived effect of institutional duty hours on availability for urgent cases	Perceived Negative Effect; Perceived No Effect; Perceived Positive Effect	1 = Perceived Negative Effect 0 = Perceived No Effect or Positive Effect
Secondary	Perceived effect of institutional duty hours on time for teaching medical students	Perceived Negative Effect; Perceived No Effect; Perceived Positive Effect	1 = Perceived Negative Effect 0 = Perceived No Effect or Positive Effect
Secondary	Perceived effect of institutional duty hours on relationship between interns/residents	Perceived Negative Effect; Perceived No Effect; Perceived Positive Effect	1 = Perceived Negative Effect 0 = Perceived No Effect or Positive Effect
Secondary	Perceived effect of institutional duty hours on professionalism	Perceived Negative Effect; Perceived No Effect; Perceived Positive Effect	1 = Perceived Negative Effect 0 = Perceived No Effect or Positive Effect
Secondary	Perceived effect of institutional duty hours on morale	Perceived Negative Effect; Perceived No Effect; Perceived Positive Effect	1 = Perceived Negative Effect 0 = Perceived No Effect or Positive Effect
Secondary	Perceived effect of institutional duty hours on ability to prepare for cases away from hospital	Perceived Negative Effect; Perceived No Effect; Perceived Positive Effect	1 = Perceived Negative Effect 0 = Perceived No Effect or Positive Effect
Secondary	Perceived effect of institutional duty hours on participation in research	Perceived Negative Effect; Perceived No Effect; Perceived Positive Effect	1 = Perceived Negative Effect 0 = Perceived No Effect or Positive Effect
Secondary	Perceived effect of institutional duty hours on job satisfaction	Perceived Negative Effect; Perceived No Effect; Perceived Positive Effect	1 = Perceived Negative Effect 0 = Perceived No Effect or Positive Effect
Secondary	Perceived effect of institutional duty hours on career choice satisfaction (decision to become a surgeon)	Perceived Negative Effect; Perceived No Effect; Perceived Positive Effect	1 = Perceived Negative Effect 0 = Perceived No Effect or Positive Effect

TABLE S7 (continued). FIRST Trial Study Endpoints – Resident Outcomes

PRIMARY/ SECONDARY ENDPOINT	ENDPOINT	RESPONSE CATEGORIES	DICHOTOMIZED RESPONSE CATEGORIES
Secondary	Perceived effect of institutional duty hours on time with family and friends	Perceived Negative Effect; Perceived No Effect; Perceived Positive Effect	1 = Perceived Negative Effect 0 = Perceived No Effect or Positive Effect
Secondary	Perceived effect of institutional duty hours on time for extracurricular activities (hobbies)	Perceived Negative Effect; Perceived No Effect; Perceived Positive Effect	1 = Perceived Negative Effect 0 = Perceived No Effect or Positive Effect
Secondary	Perceived effect of institutional duty hours on health	Perceived Negative Effect; Perceived No Effect; Perceived Positive Effect	1 = Perceived Negative Effect 0 = Perceived No Effect or Positive Effect
Secondary	Perceived effect of institutional duty hours on rest	Perceived Negative Effect; Perceived No Effect; Perceived Positive Effect	1 = Perceived Negative Effect 0 = Perceived No Effect or Positive Effect
Secondary	Resident satisfaction with continuity of care	5-Point Likert: Very Dissatisfied, Dissatisfied, Neutral, Satisfied, Very Satisfied	1 = Very Dissatisfied or Dissatisfied 0 = Neutral, Satisfied or Very Satisfied
Secondary	Resident satisfaction with patient safety	5-Point Likert: Very Dissatisfied, Dissatisfied, Neutral, Satisfied, Very Satisfied	1 = Very Dissatisfied or Dissatisfied 0 = Neutral, Satisfied or Very Satisfied
Secondary	Resident satisfaction with work hours/scheduling	5-Point Likert: Very Dissatisfied, Dissatisfied, Neutral, Satisfied, Very Satisfied	1 = Very Dissatisfied or Dissatisfied 0 = Neutral, Satisfied or Very Satisfied
Secondary	Resident satisfaction with quality/ease of handoffs/care transitions	5-Point Likert: Very Dissatisfied, Dissatisfied, Neutral, Satisfied, Very Satisfied	1 = Very Dissatisfied or Dissatisfied 0 = Neutral, Satisfied or Very Satisfied
Secondary	Resident satisfaction with time for rest	5-Point Likert: Very Dissatisfied, Dissatisfied, Neutral, Satisfied, Very Satisfied	1 = Very Dissatisfied or Dissatisfied 0 = Neutral, Satisfied or Very Satisfied
Secondary	Resident satisfaction with work hour regulations	5-Point Likert: Very Dissatisfied, Dissatisfied, Neutral, Satisfied, Very Satisfied	1 = Very Dissatisfied or Dissatisfied 0 = Neutral, Satisfied or Very Satisfied
Secondary	How often fatigue affects personal safety	5-Point Scale: Almost Always; Often; Sometimes; Rarely; Never	1 = Almost Always or Often 0 = Sometimes, Rarely or Never
Secondary	How often fatigue affects patient safety	5-Point Scale: Almost Always; Often; Sometimes; Rarely; Never	1 = Almost Always or Often 0 = Sometimes, Rarely or Never
Secondary	How many times in the last month did resident leave during an operation due to duty hour regulations	5-Point Scale: 0 Times; 1-2 Times; 3-5 Times; 6-10 Times; >10 Times	1 = 1 or More Times 0 = 0 Times
Secondary	How many times in the last month did resident miss an operation due to duty hour regulations	5-Point Scale: 0 Times; 1-2 Times; 3-5 Times; 6-10 Times; >10 Times	1 = 1 or More Times 0 = 0 Times
Secondary	How many times in the last month did resident hand off active patient care issue due to duty hour regulations	5-Point Scale: 0 Times; 1-2 Times; 3-5 Times; 6-10 Times; >10 Times	1 = 1 or More Times 0 = 0 Times

D. STATISTICAL ANALYSIS – ADDITIONAL DETAILS

D.1. ANALYSIS OF PATIENT OUTCOMES

Intent-to-Treat (ITT) Analyses. We estimated three-level hierarchical mixed-effects logistic regression models with empirical Bayes estimates of variance components. In these models, we regressed patient outcomes (Section C) on study arm assignment with controls for program-level tertile of 30-day postoperative death/serious morbidity (stratifying variable in randomization of residency programs) and hospital-level and program-level random intercepts. Point estimates reported in this paper are conditional effects, conditioning on program and hospital intercepts (and other covariates in adjusted models).

Evaluation of Noninferiority. A noninferiority margin for 30-day postoperative death or serious morbidity was set at an absolute difference (Δ) of 1.25 percentage points from baseline rates. Given a baseline rate (P_0) of 9.00%, we used the following formula to express the noninferiority margin as an odds ratio (OR):

$$OR_{\Delta} = \big((P_0 + \Delta) * (1 - P_0)\big) / \big(P_0 * (1 - P_0 - \Delta)\big)$$

Thus, for 30-day death and serious morbidity (DSM), $P_0 = 0.090$ and Δ =0.0125:

$$OR_{\Delta}^{DSM} = \big((0.090 + 0.0125) * (1 - 0.090)\big) / \big(0.090 * (1 - 0.090 - 0.0125)\big)$$

$$OR_{\Delta}^{DSM} = 1.15$$

Given a noninferiority margin of Δ=1.25 percentage points, the difference between a 9.00% baseline rate of DSM and a rate of 10.25% corresponded to an odds ratio of OR_{Δ}=1.15.

Given a 9.00% baseline rate of DSM in Standard Policy, a noninferiority margin of Δ=1.25 percentage point absolute difference between Flexible Policy and Standard Policy arms amounted to a relative difference of 13.89% (((10.25-9.00)/9.00)*100% = 13.89%) over baseline. Because we did not define noninferiority margins for secondary patient outcomes *ex ante*, we defined *ex post* (but prior to data analysis) noninferiority margins for secondary patient outcomes as a relative difference of 13.89% over Standard Policy baseline rates for each outcome. Table S8 shows Standard Policy baseline rates for each patient outcome as well as the noninferiority margin for each outcome, expressed as an odds ratio (OR_{Δ}).

For each outcome, Flexible Policy was judged to be *noninferior* to Standard Policy if the point-estimate odds ratio was below OR_{Δ}, and the upper bound of the 92% confidence interval (92%CI) was also below OR_{Δ}.

Flexible Policy was judged to be *superior* to Standard Policy if both the point estimate odds ratio and 92%CI upper bound were <1.00 and <OR_{Δ}.

Flexible Policy was judged to be *inferior* to Standard Policy if both the point estimate odds ratio and 92%CI lower bound were above 1.00 and >OR_{Δ}

The noninferiority of Flexible Policy with respect to Standard Policy was deemed *inconclusive* if the point estimate odds ratio was below OR_{Δ}, but the 92%CI upper bound was above OR_{Δ}.

TABLE S8. Patient Outcomes, Baseline Rates of Outcomes, and NonInferiority Margins

30-Day Postoperative Patient Outcome	Standard Policy Baseline Rate (P_0)	NonInferiority Margin of 13% Relative Difference (from Baseline) Expressed as an Odds Ratio ($OR_{\|}$)
Death/Serious Morbidity	9.00%	1.15
Death	1.14%	1.14
Overall Morbidity	9.18%	1.16
Serious Morbidity	8.58%	1.15
Failure-to-Rescue	8.34%	1.15
Pneumonia	1.17%	1.14
Renal Failure	0.64%	1.14
Surgery-Related Return-to-OR/Reoperation	2.66%	1.14
Sepsis	1.90%	1.14
Surgical Site Infection	4.52%	1.15
Urinary Tract Infection	1.06%	1.14

NOTE: Rates are observed rates of complications among general surgery patients in the Standard Policy arm (where general surgery patients were identified on the basis of a combination of surgeon specialty and Current Procedural Terminology™ (CPT) codes) in FIRST Trial ACS NSQIP data (2014-2015 academic year).

Sensitivity Analyses. To examine the robustness of our estimates with respect to variation in model specification, we also estimated non-hierarchical logistic regression with two-dimensional clustered standard errors. Two-level hierarchical logistic regression models were also estimated with program-level intercepts only, and also with hospital-level intercepts only.

All models were also estimated with additional adjustment for patient characteristics only, hospital characteristics only, as well as a combination of patient and hospital characteristics. Patient covariates varied across models for different outcomes and were based on standard ACS NSQIP risk-adjustment models from the *ACS NSQIP Semiannual Report (SAR) Supplement Model Reports (Released January 2015)*. Section G lists patient covariates that were used in adjusting each patient outcome. Hospital characteristics that were used for adjustment were: total admission volume, presence of a Commission on Cancer-approved cancer program, resident-to-bed ratio, ownership type, and geographic region.

Subgroup Analyses. Planned subgroup analyses were conducted for the primary patient outcome (30-day postoperative death or serious morbidity) to investigate whether there were any differential effects of assignment to Flexible Policy on outcomes by (1) emergent/urgent vs. elective surgery; (2) high-risk patients (top decile of highest predicted risk of DSM) vs. all other patients; and (3) inpatient vs outpatient operations.

In subgroup analyses, the basic hierarchical logistic regression ITT model was modified to include an interaction term between study arm assignment and the subgroup variable of interest.

Per-Protocol Analyses. Per-protocol analyses of patient outcomes were carried out on the subset of programs that were adherent to the conditions of their study arm, as defined in Section B (number of

adherent programs = 115 programs). The models we employed were the same as those described above for ITT analyses.

As-Treated Analyses. As-treated analyses investigated the effect of departures from ACGME duty hour standards on resident outcomes, regardless of study arm assignment. We modeled the effect of departures from ACGME duty hour standards in three ways.

First, we regressed patient outcomes on a count of the number of departures from ACGME standards (range: 0-4). Coefficients on this variable represented the effect of an additional departure from ACGME standards on patient outcomes.

Second, to explore whether there were any nonlinearities or non-monotonicity in the relationship between departures from ACGME standards and patient outcomes, we regressed patient outcomes on a set of dummy variables indicating the total number of departures from ACGME standards at an institution: no departures from standards (reference category), one departure, two departures, three departures, and four departures from standards.

Third, we regressed resident outcomes on a set of four dichotomous variables indicating whether or not a program deviated from the specific ACGME duty hour standards and implemented the following: a) work shifts for interns can exceed 16 hours; b) work shifts for residents can exceed 28 hours; c) residents are not required to have at least 8 hours off between shifts; d) residents are not required to have at least 14 hours off after 24 hours of continuous duty. These variables were included in regression models simultaneously to investigate whether there were any standard-specific effects on outcomes.

As-treated analyses modeled the relationship between outcomes and predictors using hierarchical logistic regression with program- and hospital-level random intercepts.

Local Average Treatment Effect Analyses. Because adherence to study arm conditions may be endogenous to factors that also affect outcomes of interest, we used instrumental variables (IV) estimates of the local average treatment effect (LATE) to estimate the effect of Flexible Policy on patient outcomes among the subset of programs that would change their institutional duty hour policies (and make the policies less restrictive) if given the opportunity to do so.

An IV is a variable that is highly predictive of the endogenous variable, but otherwise uncorrelated with outcomes of interest except through its effect on the endogenous variable. We used study arm assignment to instrument actual exposure to flexible duty hours because by construction, it was highly predictive of exposure, and uncorrelated with outcomes except through its effect on exposure.

Using two-stage regression models, we regressed exposure to departures from ACGME standards on study arm assignment in the first stage. In the second stage, patient outcomes were regressed on predicted exposure to departures from the selected ACGME duty hour standards. Thus, IV LATE analyses used variation in actual exposure to departures from ACGME duty hour standards that was "induced" by randomization to assess the effect of the Flexible Policy on outcomes

IV provides consistent estimates of causal effects if 1) the instrumental variable is strongly correlated (i.e., is strongly predictive) of actual departure from ACGME standards (i.e., actual exposure to Flexible Policy conditions), and 2) the instrument is independent of outcomes and only affects outcomes through its effect on actual exposure to deviations from ACGME duty hour standards. To test the first assumption, we examined the magnitude and significance of the coefficient on the instrument in the first

stage regression, as well as the first stage F statistic. The second assumption is satisfied because treatment assignment could not logically affect outcomes except through its effect on exposure to departures from ACGME duty hour standards.

To establish a baseline for comparison, we first estimated linear probability models in which patient outcomes were regressed on study arm assignment and controls for program-level tertile of 30-day postoperative death/serious morbidity (stratifying variable in randomization of residency programs). We estimated both ordinary least squares models with program-level clustered standard errors, as well as hierarchical linear models with program-level random intercepts.

To implement IV LATE analyses, we estimated both first- and second-stage regressions as linear probability models. For each patient outcome, we estimated two-stage least squares (TSLS) IV models with program-level clustered standard errors, as well as two-level hierarchical IV models with program-level random intercepts. We explored both generalized two-stage least squares (G2SLS) and error-component two-stage least squares (EC2SLS) estimation methods. All IV LATE models controlled for program-level tertile of 30-day postoperative death/serious morbidity (stratifying variable in randomization of residency programs).

D.2. ANALYSIS OF RESIDENT ENDPOINTS

Preliminary Analyses

Intent-to-Treat (ITT) Analyses. Resident outcomes originally measured using a 5-point scale (e.g., satisfaction outcomes, frequency outcomes) were initially modeled using two-level hierarchical ordered logistic regression with program-level random intercepts. Violation of proportional odds assumption in ordered logistic regression was assessed using Brant and Wald tests.

Resident outcomes originally measured as unordered trichotomous categorical variables (i.e., variables measuring residents' perception of the effects of duty hour policy) were initially modeled using two-level hierarchical multinomial logistic regression with program-level random intercepts.

In all models, resident outcomes were regressed on a variable indicating study arm assignment (Flexible Policy vs. Standard Policy (reference category)). All models for all outcomes were estimated with and without controls for program-level tertile of 30-day postoperative death/serious morbidity (stratifying variable in randomization of residency programs). All models were also estimated with and without adjustment for resident gender (male, female); postgraduate year (PGY1-PGY5); program type (academic, community or military); and program geographic region (Northeast, West, Southwest, Midwest, South).

To assess the sensitivity of our estimates with respect to minor variations in model specification, we also estimated non-hierarchical ordered logistic regression models with program-level clustered standard errors and multinomial and hierarchical multinomial logistic regression models for all ordered categorical outcomes. For unordered categorical outcomes, we also estimated non-hierarchical multinomial logistic regression models with program-level clustered standard errors.

Subgroup Analyses. For primary resident outcomes only, we examined whether there were any subgroup effects of assignment to Flexible Policy (vs. Standard Policy) within gender subgroups (among

females, among males), within resident PGY-level subgroups (among junior residents (PGY1 & PGY2), among senior residents (PGY3 & PGY4), among chief residents (PGY5), within geographic regions (among Northeastern programs, among Southern programs, among Midwestern programs, among Western programs), and within program types (among academic-based programs, among community or military-based programs). Subgroup analyses were carried out by interacting study arm assignment with subgroup variables.

Program Adherence to Study Arm Conditions. To explore the influence of differential adherence on our estimates of the Flexible Policy effect on resident outcomes, we undertook per-protocol, as-treated, and local average treatment effect analyses.

Per-Protocol Analyses. Per-protocol analyses of patient outcomes and resident outcomes were carried out on the subset of programs that were adherent to the conditions of their study arm (as previously described in Section B and in Section D.1). Models for per-protocol analyses were otherwise the same as those used in ITT analyses.

As-Treated Analyses. As-treated analyses investigated the effect of departures from ACGME duty hour standards on resident outcomes, regardless of study arm assignment. We modeled the effect of departures from ACGME duty hour standards in three ways.

First, we regressed resident outcomes on a count of the number of departures from ACGME standards (previously described in D.1).

Second, we regressed resident outcomes on a set of dummy variables indicating the total number of departures from ACGME standards at an institution: no departures from standards (reference category), one departure, two departures, three departures, and four departures from standards (see D.1).

Third, we regressed resident outcomes on a set of four dichotomous variables indicating whether or not a program deviated from each of the specific ACGME duty hour standards and implemented the following: a) work shifts for interns can exceed 16 hours; b) work shifts for residents can exceed 28 hours; c) residents are not required to have at least 8 hours off between shifts; d) residents are not required to have at least 14 hours off after 24 hours of continuous duty (see D.1.).

All as-treated models included controls for program-level tertile of 30-day postoperative death/serious morbidity (stratifying variable in randomization of residency programs). As-treated models were estimated using hierarchical logistic regression with program random intercepts.

Local Average Treatment Effect Analyses. As previously described in D.1., we again used IV LATE methods to estimate the effect of Flexible Policy on resident outcomes among the subset of programs that would change their institutional duty hour policies (and make the policies less restrictive) if given the opportunity to do so.

To implement IV LATE analyses, we dichotomized all measures of resident outcomes and then estimated both first- and second-stage regressions as linear probability models. For each resident outcome, we estimated two-stage least squares (TSLS) IV models with program-level clustered standard errors, as well as two-level hierarchical IV models with program-level random intercepts. We explored both generalized two-stage least squares (G2SLS) and error-component two-stage least squares (EC2SLS) estimation methods. All IV LATE models controlled for program-level tertile of 30-day postoperative death/serious morbidity (stratifying variable in randomization of residency programs).

Reported Analyses

The original study protocol discussed the possibility of dichotomizing the resident outcomes, and this was done in the final analyses. Results were comparable between hierarchical ordered and multinomial logistic regression models and the final hierarchical logistic regression models (dichotomized outcomes) shown in the paper. We dichotomized all resident outcomes (see Section D) for three reasons: (1) due to small/zero cells, (2) violation of proportional odds assumption for some (but not all) outcomes, and (3) ease of presentation. We then repeated all ITT, subgroup, per-protocol, as-treated, and IV LATE analyses as previously described using methods for dichotomous outcomes (hierarchical logistic regression).

Other Methodological Notes

No inferiority margins were defined for any of the resident outcomes. Thus, hypothesis testing was based on two-tailed tests of significance on study arm assignment (Flexible Policy vs. Standard Policy (reference)). Because resident outcomes were not examined during interim analysis, hypothesis testing for resident outcomes was conducted with α=0.05 and standard 95% confidence intervals.

We did not adjust p-values for multiple hypothesis testing because doing so would favor our hypothesis of finding no differences in resident outcomes between study arms. We also considered each resident outcome to be of specific substantive interest rather than multiple indicators of a single construct of resident outcomes.

E. RESULTS TABLES

TABLE S9. Raw Hospital-Level Rates of Postoperative 30-Day Patient Outcomes by Study Arm

	Mean Complication Rate (96%CI)			Clustered P-Value
	TOTAL	STANDARD POLICY	FLEXIBLE POLICY	
Raw Hospital-Level Postoperative Outcome Rates				
Death or serious morbidity	8.88 (8.36-9.40)	8.97 (8.17-9.76)	8.80 (8.12-9.48)	0.743
Death	1.17 (1.07-1.27)	1.14 (0.97-1.30)	1.20 (1.07-1.33)	0.522
Serious morbidity	8.43 (7.93-8.94)	8.53 (7.76-9.30)	8.35 (7.68-9.02)	0.717
Any morbidity	8.34 (7.81-8.86)	8.48 (7.69-9.27)	8.21 (7.51-8.91)	0.600
Failure-to-rescue	0.72 (0.65-0.80)	0.70 (0.59-0.81)	0.75 (0.64-0.85)	0.475
Pneumonia	1.18 (1.05-1.32)	1.21 (1.00-1.42)	1.16 (0.99-1.33)	0.682
Renal failure	0.65 (0.57-0.73)	0.61 (0.50-0.71)	0.69 (0.58-0.81)	0.248
Return to operating room	2.61 (2.41-2.80)	2.71 (2.42-3.00)	2.51 (2.24-2.79)	0.310
Postoperative sepsis	1.71 (1.50-1.92)	1.77 (1.47-2.06)	1.65 (1.36-1.94)	0.577
Surgical site infection	4.31 (3.98-4.64)	4.42 (3.93-4.91)	4.22 (3.77-4.66)	0.520
Urinary tract infection	1.01 (0.90-1.12)	1.04 (0.88-1.20)	0.97 (0.82-1.13)	0.536

NOTE: Patient outcomes were aggregated to the hospital level. This table reports sample means of hospital-level aggregated patient outcomes by study arm and for the entire sample (N=148 hospitals). 96% confidence intervals are reported because a two-tailed test for differences in means across study arms was conducted with α=0.04.

TABLE S10. Intent-to-Treat Estimate of the Effect of Assignment to Flexible Policy on 30-Day Postoperative Death or Serious Morbidity (*Primary Patient Outcome*)

REGRESSOR	ODDS RATIO (OR)	92% CONFIDENCE INTERVAL	P-VALUE
Flexible Policy (vs. Standard Policy)	0.96	0.87 – 1.06	0.443
2013 (Baseline) 30-Day Postoperative Death/Serious Morbidity			
Tertile 1	1.00	0.83 – 1.21	0.991
Tertile 2	1.29	1.07 – 1.56	0.017
Tertile 3	1.53	1.27 – 1.85	<0.001
Baseline Data Not Available	Reference	Reference	Reference
Variance Components			
Hospital	0.10	0.08 – 0.13	-----
Residency Program	5.58E-09	0 - .	-----
N Cases	138691		
N Hospitals	148		
N Residency Programs	115		

NOTE: Estimates are from a 3-level hierarchical logistic regression with hospital and program random intercepts.

TABLE S11. Intent-to-Treat Estimate of the Effect of Assignment to Flexible Policy on 30-Day Postoperative Death or Serious Morbidity (*Primary Patient Outcome)* with Adjustment for Patient Characteristics

REGRESSOR	ODDS RATIO (OR)	92% CONFIDENCE INTERVAL	P-VALUE
Flexible Policy (vs. Standard Policy)	0.96	0.90 – 1.04	0.378
2013 (Baseline) 30-Day Postoperative Death/Serious Morbidity			
Tertile 1	0.87	0.76 – 0.99	0.065
Tertile 2	0.95	0.82 – 1.08	0.470
Tertile 3	1.05	0.91 – 1.20	0.559
Baseline Data Not Available	Reference	Reference	Reference
Age Group (Reference: Age<65)			
Ages 65-74	1.19	1.14 – 1.24	<0.001
Ages 75-84	1.23	1.17 – 1.31	<0.001
Ages 85+	1.33	1.22 – 1.46	<0.001
ASA Class (Reference: Class 1)			
Class 2	1.71	1.50 – 1.95	<0.001
Class 3	2.89	2.53 – 3.30	<0.001
Classes 4-5	6.72	5.85 – 7.73	<0.001
CPT Linear Risk	2.58	2.51 – 2.64	<0.001
Emergent/Urgent Surgery (Reference: Elective)	1.33	1.26 – 1.41	<0.001
Functional Status (Reference: Independent)			
Partially Dependent	1.70	1.54 – 1.87	<0.001
Totally Dependent	1.91	1.61 – 2.26	<0.001
Male (Reference: Female)	1.09	1.05 – 1.13	<0.001
Wound Class (Reference: Clean)			
Clean/Contaminated	1.03	0.98 – 1.09	0.260
Contaminated	1.27	1.19 – 1.36	<0.001
Dirty/Infected	1.42	1.33 – 1.52	<0.001
Variance Components			
Hospital	0.04	0.02 – 0.06	-----
Residency Program	0.0043	0.0001 – 0.2212	-----
N Cases	138691		
N Hospitals	148		
N Residency Programs	115		

NOTE: Estimates are from a 3-level hierarchical logistic regression with hospital and program random intercepts. 5 integration points used in estimation.

TABLE S12. Summary of Hierarchical Logistic Regression Estimates of the Association between Assignment to Flexible Policy and Odds of Postoperative Complications, with and without Adjustment

MODEL	ODDS RATIO	92% CI	P	N		
				PROGRAMS	HOSPITALS	CASES
30-Day Postoperative Death/Serious Morbidity						
No Patient or Hospital Characteristics ("Unadjusted")	0.96	0.87 - 1.06	0.443	115	148	138691
Adjusted for Patient Characteristics	0.96	0.90 - 1.04	0.378	115	148	138691
Adjusted for Hospital Characteristics	0.99	0.90 - 1.09	0.857	112	140	133838
Adjusted for Patient and Hospital Characteristics	0.99	0.92 - 1.07	0.892	112	140	133838
30-Day Postoperative Death						
No Patient or Hospital Characteristics ("Unadjusted")	1.00	0.86 - 1.16	0.993	115	148	138691
Adjusted for Patient Characteristics	0.95	0.82 – 1.10	0.558	115	148	138691
Adjusted for Hospital Characteristics	1.04	0.89 - 1.21	0.665	112	140	133838
Adjusted for Patient and Hospital Characteristics	0.95	0.82 - 1.12	0.636	112	140	133838
30-Day Postoperative Serious Morbidity						
No Patient or Hospital Characteristics ("Unadjusted")	0.96	0.86 - 1.06	0.449	115	148	138691
Adjusted for Patient Characteristics	0.96	0.90 – 1.04	0.399	115	148	138691
Adjusted for Hospital Characteristics	0.99	0.90 - 1.09	0.848	112	140	133838
Adjusted for Patient and Hospital Characteristics	0.99	0.92 – 1.07	0.878	112	140	133838
30-Day Postoperative Any Morbidity						
No Patient or Hospital Characteristics ("Unadjusted")	0.94	0.84 - 1.06	0.392	115	148	138691
Adjusted for Patient Characteristics	0.96	0.89 – 1.04	0.388	115	148	138691
Adjusted for Hospital Characteristics	0.99	0.89 - 1.10	0.849	112	140	133838
Adjusted for Patient and Hospital Characteristics	1.00	0.92 – 1.08	0.928	112	140	133838
30-Day Postoperative Failure-to-Rescue						
No Patient or Hospital Characteristics ("Unadjusted")	1.03	0.87 - 1.23	0.730	115	148	11937
Adjusted for Patient Characteristics	1.00	0.86 – 1.18	0.966	115	148	11937
Adjusted for Hospital Characteristics	1.06	0.89 - 1.26	0.575	112	140	11623
Adjusted for Patient and Hospital Characteristics	1.01	0.86 – 1.20	0.880	112	140	11623
30-Day Postoperative Pneumonia						
No Patient or Hospital Characteristics ("Unadjusted")	0.95	0.78 - 1.14	0.603	115	148	138375
Adjusted for Patient Characteristics	0.96	0.81 – 1.14	0.669	115	148	138375
Adjusted for Hospital Characteristics	1.02	0.84 - 1.24	0.867	112	140	133531
Adjusted for Patient and Hospital Characteristics	1.00	0.84 – 1.20	0.994	112	140	133531

Note: Missing complete hospital-level characteristics for 8 hospitals (3 programs).

TABLE S12 (continued). Summary of Hierarchical Logistic Regression Estimates of the Association between Assignment to Flexible Policy and Odds of Postoperative Complications, with and without Adjustment

MODEL	ODDS RATIO	92% CI	P	N		
				PROGRAMS	HOSPITALS	CASES
30-Day Postoperative Renal Failure						
No Patient or Hospital Characteristics ("Unadjusted")	1.05	0.86 - 1.28	0.659	115	148	138596
Adjusted for Patient Characteristics	1.07	0.91 – 1.27	0.466	115	148	138596
Adjusted for Hospital Characteristics	1.11	0.91 - 1.35	0.357	112	140	133745
Adjusted for Patient and Hospital Characteristics	1.10	0.92 – 1.31	0.371	112	140	133745
30-Day Postoperative Unplanned Reoperation						
No Patient or Hospital Characteristics ("Unadjusted")	0.91	0.81 - 1.03	0.173	115	148	138691
Adjusted for Patient Characteristics	0.93	0.84 – 1.04	0.249	115	148	138691
Adjusted for Hospital Characteristics	0.97	0.86 - 1.09	0.618	112	140	133838
Adjusted for Patient and Hospital Characteristics	0.99	0.89 – 1.09	0.804	112	140	133838
30-Day Postoperative Sepsis						
No Patient or Hospital Characteristics ("Unadjusted")	0.90	0.73 - 1.10	0.363	115	148	135258
Adjusted for Patient Characteristics	0.89	0.76 – 1.03	0.166	115	148	135258
Adjusted for Hospital Characteristics	0.94	0.77 - 1.15	0.594	112	140	130482
Adjusted for Patient and Hospital Characteristics	0.93	0.79 – 1.09	0.414	112	140	130482
30-Day Postoperative Surgical Site Infection						
No Patient or Hospital Characteristics ("Unadjusted")	0.93	0.81 - 1.08	0.396	115	148	137346
Adjusted for Patient Characteristics	0.94	0.86 – 1.04	0.317	115	148	137346
Adjusted for Hospital Characteristics	0.97	0.85 - 1.11	0.731	112	140	132526
Adjusted for Patient and Hospital Characteristics	0.99	0.88 – 1.10	0.847	112	140	132526
30-Day Postoperative Urinary Tract Infection						
No Patient or Hospital Characteristics ("Unadjusted")	0.91	0.76 - 1.08	0.324	115	148	138691
Adjusted for Patient Characteristics	0.90	0.76 – 1.06	0.254	115	148	138691
Adjusted for Hospital Characteristics	0.91	0.76 - 1.09	0.376	112	140	133838
Adjusted for Patient and Hospital Characteristics	0.90	0.75 – 1.08	0.304	112	140	133838

Note: Missing complete hospital-level characteristics for 8 hospitals (3 programs).

TABLE S13. Summary of Subgroup Effects of Flexible Policy Assignment: 30-Day Postoperative Death/Serious Morbidity (*Primary Patient Outcome*)

Subgroup Comparison	UNEXPONENTIATED Coefficient (95%CI)	P-Value
Emergency vs. Non-Emergency		
Non-Emergency: Flexible Policy vs. Standard Policy	-0.047 (-0.168 – 0.074)	0.446
Emergency: Flexible Policy vs. Standard Policy	-0.008 (-0.150 – 0.134)	0.914
Subgroup Difference in Study Arm Differences	0.039 (-0.059 – 0.137)	0.433
Overall F-Test for Significant Interaction	-----	0.560
Inpatient vs. Outpatient		
Outpatient: Flexible Policy vs. Standard Policy	-0.161 (-0.316 - -0.006)	0.041
Inpatient: Flexible Policy vs. Standard Policy	0.005 (-0.102 – 0.112)	0.928
Subgroup Difference in Study Arm Differences	0.166 (0.039 – 0.293)	0.010
Overall F-Test for Significant Interaction	-----	0.036†
High-Risk vs. Not High Risk		
Not High Risk: Flexible Policy vs. Standard Policy	-0.052 (-0.148 – 0.043)	0.282
High Risk: Flexible Policy vs. Standard Policy	-0.064 (-0.179 – 0.052)	0.279
Subgroup Difference in Study Arm Differences	-0.011 (-0.101 – 0.078)	0.803
Overall F-Test for Significant Interaction	-----	0.494

N Programs = 115 programs; N Hospitals = 148 hospitals; N NSQIP Cases = 138691 cases

NOTE: Estimates are from 3-level hierarchical logistic regression models that regress 30-day death/serious morbidity on an interaction between study arm assignment (Flexible Policy vs. Standard Policy) and subgroup variable. Models control for program-level tertile of 2013 rates of 30-day postoperative death/serious morbidity (stratifying variable used in randomization of programs) and include both program-level and hospital-level random intercepts. "High-risk" group defined by being in the top decile of ACS NSQIP CPT-based risk predictor for DSM. †This interaction is not statistically significant at the $p<0.04$ level after Bonferroni adjustment for multiple tests (number of tests = 2, because there are two subgroups), and there is no difference between Flexible Policy and Standard Policy for either group.

TABLE S14. Summary of Per-Protocol Analyses of the Association between Trial Arm Assignment and Patient Outcomes

OUTCOME	Assignment to Flexible Policy OR (92%CI) P-Value			N Programs	N Hospitals	N ACS NSQIP Cases
	Model 1	Model 2	Model 3			
Primary Outcomes						
30-Day Postoperative Death or Serious Morbidity	0.97 (0.87 – 1.07) 0.543	0.98 (0.89 – 1.08) 0.709	0.97 (0.87 – 1.07) 0.543	113	146	136319
Secondary Outcomes						
30-Day Postoperative Death	1.00 (0.87 – 1.16) 0.979	0.98 (0.86 – 1.13) 0.841	1.00 (0.87 – 1.16) 0.979	113	146	136319
30-Day Postoperative Serious Morbidity	0.96 (0.87 – 1.07) 0.547	0.98 (0.89 – 1.08) 0.758	0.96 (0.87 – 1.07) 0.547	113	146	136319
30-Day Postoperative Any Morbidity	0.95 (0.85 – 1.07) 0.478	0.97 (0.87 – 1.08) 0.659	0.95 (0.85 – 1.07) 0.478	113	146	136319
30-Day Postoperative Failure to Rescue	1.02 (0.87 – 1.21) 0.797	1.01 (0.85 – 1.19) 0.957	1.02 (0.87 – 1.21) 0.797	113	146	11688
30-Day Postoperative Pneumonia	0.93 (0.77 – 1.13) 0.524	0.98 (0.82 – 1.16) 0.812	0.93 (0.77 – 1.13) 0.524	113	146	136005
30-Day Postoperative Renal Complications	1.05 (0.86 – 1.28) 0.644	1.12 (0.94 – 1.35) 0.258	1.05 (0.86 – 1.28) 0.644	113	146	136224
30-Day Postoperative Return to Operating Room	0.92 (0.82 – 1.04) 0.236	0.93 (0.83 – 1.04) 0.254	0.92 (0.82 – 1.02) 0.166	113	146	136319
30-Day Postoperative Sepsis	0.92 (0.75 – 1.13) 0.481	0.97 (0.80 – 1.16) 0.754	0.91 (0.75 – 1.11) 0.396	113	146	132955
30-Day Postoperative Surgical Site Infection	0.95 (0.82 – 1.09) 0.496	0.97 (0.85 – 1.11) 0.697	0.95 (0.82 – 1.09) 0.496	113	146	134990
30-Day Postoperative Urinary Tract Infection	0.91 (0.76 – 1.08) 0.338	0.91 (0.77 – 1.09) 0.367	0.91 (0.77 – 1.08) 0.346	113	146	136319

NOTE: Estimates are from 3-level hierarchical logistic regression models that regress 30-day death/serious morbidity study on arm assignment (Flexible Policy vs. Standard Policy). Models control for program-level tertile of 2013 rates of 30-day postoperative death/serious morbidity (stratifying variable used in randomization of programs) and include both program-level and hospital-level random intercepts. **Model 1** was a 3-level hierarchical logistic regression model with program- and hospital-level random intercepts. **Model 2** was a 2-level hierarchical logistic regression model with program random intercepts. **Model 3** was a 2-level hierarchical logistic regression model with hospital random intercepts. Per-protocol analyses were run on a subset of the FIRST Trial sample comprised of programs that were adherent to study arm conditions. Adherent Standard Policy programs were those in which program directors reported zero policy deviations from 2015 ACGME duty hour standards. Adherent Flexible Policy programs were those in which program directors reported one or more of the following deviations from 2015 ACGME duty standards, per FIRST Trial protocol: 1) PGY1 shifts can exceed 16 hours, 2) PGY2+ shifts can exceed 28 hours, 3) 14 hours off after 24-hour shifts not required, 4) 8-10 hours off between shifts not required.

TABLE S15. As-Treated Analysis of the Effect of Actual Exposure to Flexible Policy on Patient Outcomes: 30-Day Postoperative Death/Serious Morbidity (*Primary Patient Outcome*)

Exposure Measure	Odds Ratio (92%CI) P-Value		
	Model 1	Model 2	Model 3
Number of departures from ACGME standards	0.99 (0.97 – 1.02) 0.606	-----	-----
No (Zero) departures	-----	Reference	-----
One departure	-----	1.04 (0.79 – 1.36) 0.820	-----
Two departures	-----	0.98 (0.81 – 1.20) 0.891	-----
Three departures	-----	0.88 (0.70 – 1.11) 0.348	-----
Four departures	-----	0.98 (0.88 – 1.09) 0.726	-----
PGY1 duty hours can exceed 16 hours	-----	-----	0.83 (0.66 – 1.04) 0.152
PGY2+ duty hours can exceed 28 hours	-----	-----	1.11 (0.89 – 1.37) 0.408
Residents don't require 14 hours off after 24-hour duty	-----	-----	1.22 (0.97 – 1.54) 0.126
Residents don't require 8-10 hours off between shifts	-----	-----	0.86 (0.70 – 1.05) 0.185

N Programs = 115 programs; N Hospitals = 148 hospitals; N NSQIP Cases = 138691 cases.

NOTE: Estimates are from 3-level hierarchical logistic regression models that regress 30-day death/serious morbidity on study arm assignment (Flexible Policy vs. Standard Policy) and measures of actual program-level exposure to deviations from ACGME duty hour standards. All models control for program-level tertile of 2013 rates of 30-day postoperative death/serious morbidity (stratifying variable used in randomization of programs) and include both program-level and hospital-level random intercepts. In **Model 1**, the single exposure variable in the model was a count from 0-4 of the number of the following deviations from ACGME duty hour standards that were implemented at an institution (regardless of study arm assignment): 1) PGY1 shifts can exceed 16 hours, 2) PGY2+ shifts can exceed 28 hours, 3) 14 hours off after 24-hour shifts not required, 4) 8-10 hours off between shifts not required. In **Model 2**, the number of deviations in 2015 ACGME standards implemented at an institution were entered as separate categorical exposure variables to capture any nonlinear cumulative effect of departure from 2015 ACGME standards on resident perceptions. In **Model 3**, all four deviations from ACGME standards were entered in the model simultaneously to investigate whether departures from specific 2015 standards were associated with resident perceptions.

TABLE S16. Instrumental Variables (IV) Estimates of the Local Average Treatment Effect (LATE): 30-Day Postoperative Death or Serious Morbidity (*Primary Patient Outcome*)

Model	Coefficient (92% CI)	P-Value
Linear Probability Models (LPM) (for Comparison)		
Model 1: 2-Level Hierarchical LPM with Program Random Effects	-0.003 (-0.010 - 0.005)	0.551
Model 2: 2-Level Hierarchical LPM with Hospital Random Effects	-0.004 (-0.012 – 0.004)	0.350
Model 3: Non-hierarchical LPM with Program Clustered Robust Standard Errors	-0.003 (-0.010 – 0.005)	0.490
Model 4: Non-Hierarchical LPM with Hospital Clustered Robust Standard Errors	-0.003 (-0.011 – 0.005)	0.492
Instrumental Variables (IV) Models		
Model 5: TSLS IV with Program Clustered Robust Standard Errors	-0.003 (-0.011 – 0.005)	0.487
Model 6: 2-Level Hierarchical IV with Program Random Effects (G2SLS, Swamy-Arora variance components)	-0.003 (-0.010 – 0.005)	0.544
Model 7: 2-Level Hierarchical IV with Program Random Effects (EC2SLS, Swamy-Arora variance components)	-0.003 (-0.010 – 0.005)	0.562
Model 8: 2-Level Hierarchical IV with Program Random Effects (EC2SLS, Baltagi-Chang variance components)	-0.003 (-0.010 – 0.005)	0.562
Model 9: TSLS IV with Hospital Clustered Robust Standard Errors	-0.003 (-0.011 – 0.005)	0.490
Model 10: 2-Level Hierarchical IV with Hospital Random Effects (G2SLS, Swamy-Arora variance components)	-0.004 (-0.012 – 0.004)	0.348
Model 11: 2-Level Hierarchical IV with Hospital Random Effects (EC2SLS, Swamy-Arora variance components)	-0.004 (-0.012 – 0.004)	0.423
Model 12: 2-Level Hierarchical IV with Hospital Random Effects (EC2SLS, Baltagi-Chang variance components)	-0.004 (-0.012 – 0.004)	0.423

N Programs = 115 programs; N NSQIP Cases = 138691 cases.

Model 1 is a 2-level hierarchical linear probability model regressing patient outcomes on assignment to Flexible Policy (vs. Standard Policy) with program-level random intercepts and controls for program-level tertile of 2013 rates of 30-day postoperative death/serious morbidity (stratifying variable used in randomization of programs). **Model 2** is a 2-level hierarchical linear probability model regressing patient outcomes on assignment to Flexible Policy (vs. Standard Policy) with program-level random intercepts and controls for program-level tertile of 2013 rates of 30-day postoperative death/serious morbidity (stratifying variable used in randomization of programs). **Model 3** is a non-hierarchical linear probability model regressing outcomes on study arm assignment and controls for program-level tertile of 2013 rates of 30-day postoperative death/serious morbidity (stratifying variable used in randomization of programs) with program-level clustered standard errors. **Model 4** is a non-hierarchical linear probability model regressing outcomes on study arm assignment and controls for program-level tertile of 2013 rates of 30-day postoperative death/serious morbidity (stratifying variable used in randomization of programs) with hospital-level clustered standard errors. **Model 5** is a non-hierarchical two-stage least squares instrumental variables model where actual receipt of Flexible Policy (regardless of study arm) is instrumented by study arm assignment with program clustered robust SEs. Second stage estimate of instrumented exposure presented. **Models 6** is a hierarchical generalized two-stage least squares (G2SLS) second stage estimate of the effect of Flexible Policy instrumented by study arm assignment with program random intercepts. Second stage estimate of instrumented exposure presented. **Model 7** is a hierarchical error-corrected two-stage least squares (EC2SLS) model with program random intercepts using Swamy-Arora method for estimating variance components. Second stage estimate of instrumented exposure presented. **Model 8** is a hierarchical EC2SLS model with program random intercepts and Baltagi-Chang method for estimating variance components. Second stage estimate of instrumented exposure presented. **Models 9-12** are the same as Models 5-8 but include hospital-level random intercepts rather than program-level intercepts. TSLS first stage coefficients on study arm assignment (instrumenting for actual exposure to deviations from ACGME duty hour standards) were all > 0.95, all p<0.001. All first-stage F statistics were >1900.00, all p<0.001.

TABLE S17. Unadjusted Intent-to-Treat Estimate of the Effect of Assignment to Flexible Policy on Resident Dissatisfaction and Wellbeing (*Primary Resident Outcomes*)

REGRESSOR	ODDS RATIO (95% CONFIDENCE INTERVAL)	
	(Very) Dissatisfied: Overall Education Quality	(Very) Dissatisfied: Overall Wellbeing
Flexible Policy (vs. Standard Policy)	1.08 (0.77-1.52)	1.31 (0.99-1.74)
2013 (Baseline) 30-Day Postoperative Death/Serious Morbidity		
Tertile 1	1.11 (0.61-2.00)	1.30 (0.78-2.19)
Tertile 2	0.81 (0.44-1.49)	1.21 (0.71-2.04)
Tertile 3	0.91 (0.50-1.65)	1.20 (0.72-2.01)
Baseline Data Not Available	Reference	Reference
Constant	0.10 (0.06-0.18)	0.10 (0.06-0.16)
Variance Components		
Residency Program	0.44 (0.26-0.73)	0.28 (0.16-0.49)
N Residents	3642	3645
N Residency Programs	117	117

NOTE: Estimates are from a 3-level hierarchical logistic regression with hospital and program random intercepts. 5 integration points used in estimation. Dependent variables are coded '1' for residents reporting being 'very dissatisfied' or 'dissatisfied' and '0' for residents reporting being 'neutral,' 'satisfied' or 'very satisfied.' OR>1.00 indicate a higher odds of dissatisfaction among Flexible Policy arm residents.

TABLE S18. Adjusted Intent-to-Treat Estimate of the Effect of Assignment to Flexible Policy on Resident Dissatisfaction with Education Quality and Wellbeing (*Primary Resident Outcomes*) with Inclusion of Program-Level Covariates

REGRESSOR	ODDS RATIO (95% CONFIDENCE INTERVAL) P-VALUE	
	(Very) Dissatisfied: Education Quality	(Very) Dissatisfied: Wellbeing
Flexible Policy (vs. Standard Policy)	1.00 (0.72-1.41) 0.988	1.20 (0.91-1.59) 0.199
2013 (Baseline) 30-Day Postoperative Death/Serious Morbidity		
Tertile 1	0.84 (0.46-1.55) 0.579	1.11 (0.65-1.88) 0.703
Tertile 2	0.62 (0.33-1.15) 0.127	1.05 (0.61-1.78) 0.869
Tertile 3	0.69 (0.38-1.27) 0.233	1.04 (0.62-1.76) 0.876
Baseline Data Not Available	Reference	Reference
Female (Reference: Male)	1.26 (1.02-1.57) 0.036	1.37 (1.13-1.67) 0.002
Postgraduate Year (Reference: PGY1)		
PGY2	0.92 (0.68-1.23) 0.569	0.83 (0.64-1.08) 0.167
PGY3	0.83 (0.60-1.15) 0.268	0.73 (0.55-0.97) 0.030
PGY4	0.93 (0.67-1.29) 0.659	0.63 (0.46-0.85) 0.003
PGY5	0.58 (0.40-0.84) 0.004	0.50 (0.36-0.70) <0.001
Program Type (Reference: Academic)		
Community-Based	0.88 (0.62-1.25) 0.467	0.91 (0.68-1.22) 0.539
Military	0.31 (0.05-1.89) 0.206	0.74 (0.19-2.83) 0.656
Geographic Region (Reference: Northeast)		
Southeast	0.78 (0.48-1.25) 0.297	0.85 (0.58-1.26) 0.429
Midwest	0.54 (0.35-0.84) 0.006	0.58 (0.41-0.84) 0.004
Southwest	0.51 (0.28-0.94) 0.032	0.66 (0.40-1.09) 0.102
West	0.88 (0.51-1.51) 0.639	1.08 (0.69-1.67) 0.745
Constant	0.21 (0.11-0.40) <0.001	0.17 (0.10-0.31) <0.001
Variance Components		
Residency Program	0.35 (0.20-0.62)	0.20 (0.10-0.40)
N Residents	3642	3645
N Residency Programs	117	117

NOTE: Estimates are from a 3-level hierarchical logistic regression with hospital and program random intercepts. 7 integration points used in estimation. Dependent variables are coded '1' for residents reporting being 'very dissatisfied' or 'dissatisfied' and '0' for residents reporting being 'neutral,' 'satisfied' or 'very satisfied.' OR>1.00 indicate a higher odds of dissatisfaction among Flexible Policy arm residents.

TABLE S19. Primary Resident Outcomes, Summary of Subgroup Analyses

Subgroup Comparison	(Very) Dissatisfied: Education Quality	(Very) Dissatisfied: Wellbeing
Flexible Policy X Gender Subgroups	No Significant Interaction (p=0.537)	No Significant Interaction (p=0.097)
Flexible Policy X Resident Level (PGY1 vs. PGY2-3 vs. PGY4-5)	No Significant Interaction (p=0.018)†	No Significant Interaction (p=0.235)
Flexible Policy X Geographic Region	No Significant Interaction (p=0.877)	No Significant Interaction (p=0.162)
Flexible Policy X Program Type	No Significant Interaction (p=0.148)	No Significant Interaction (p=0.050)

NOTE: Subgroup analyses were conducted by including an interaction term between Flexible Policy assignment variable and subgroup variables in logistic regression models with program-level clustered standard errors and controls for program-level tertile of 30-day postoperative death/serious morbidity (stratifying variable in randomization of residency programs). We report the p-value on the joint test for significant interactions across all subgroup interactions. †Not significant at the 0.05 or 0.01 level after Bonferroni correction for multiple subgroup tests (3 subgroups)

TABLE S20. Per-Protocol Estimate of the Effect of Assignment to Flexible Policy on Resident Dissatisfaction with Education Quality and Wellbeing (*Primary Resident Outcomes*)

REGRESSOR	ODDS RATIO (95% CONFIDENCE INTERVAL) P-VALUE	
	(Very) Dissatisfied: Education Quality	(Very) Dissatisfied: Wellbeing
Flexible Policy (vs. Standard Policy)	1.09 (0.77-1.54) 0.623	1.35 (1.02-1.80) 0.039
2013 (Baseline) 30-Day Postoperative Death/Serious Morbidity		
Tertile 1	1.14 (0.61-2.12) 0.681	1.30 (0.77-2.22) 0.327
Tertile 2	0.83 (0.44-1.58) 0.577	1.21 (0.70-2.06) 0.496
Tertile 3	0.94 (0.50-1.75) 0.843	1.15 (0.67-1.95) 0.616
Baseline Data Not Available	Reference	Reference
Constant	0.10 (0.06-0.18) <0.001	0.10 (0.06-0.16) <0.001
Variance Components		
Residency Program	0.45 (0.27-0.75)	0.27 (0.15-0.48)
N Residents	3590	3592
N Residency Programs	115	115

NOTE: Estimates are from a 3-level hierarchical logistic regression with hospital and program random intercepts. 5 integration points used in estimation. Dependent variables are coded '1' for residents reporting being 'very dissatisfied' or 'dissatisfied' and '0' for residents reporting being 'neutral,' 'satisfied' or 'very satisfied.' OR>1.00 indicate a higher odds of dissatisfaction among Flexible Policy arm residents.

TABLE S21. As-Treated Estimate of the Effect of Assignment to Flexible Policy on Odds of Resident Being (Very) Dissatisfied with Education Quality (*Primary Resident Outcome*)

REGRESSOR	ODDS RATIO (95% CONFIDENCE INTERVAL) P-VALUE		
	Model 1	Model 2	Model 3
Sum of departures from ACGME standards	1.01 (0.92-1.10) 0.838	-----	-----
One departure	-----	1.58 (0.68-3.66) 0.283	-----
Two departures	-----	1.00 (0.44-2.29) 0.993	-----
Three departures	-----	1.21 (0.55-2.67) 0.633	-----
Four departures	-----	1.05 (0.73-1.51) 0.813	-----
No (Zero) departures	-----	Reference	-----
PGY1 duty hours can exceed 16 hours	-----	-----	1.37 (0.64-2.93) 0.415
PGY2+ duty hours can exceed 28 hours	-----	-----	1.04 (0.49-2.22) 0.922
Residents don't require 14 hours off after 24-hour duty	-----	-----	0.86 (0.38-1.96) 0.714
Residents don't require 8-10 hours off between shifts	-----	-----	0.85 (0.39-1.83) 0.676
2013 (Baseline) 30-Day Postoperative Death/Serious Morbidity			
Tertile 1	1.11 (0.92-1.10) 0.730	1.16 (0.64-2.11) 0.619	1.08 (0.58-2.00) 0.810
Tertile 2	0.81 (0.44-1.49) 0.506	0.83 (0.45-1.53) 0.548	0.79 (0.43-1.46) 0.452
Tertile 3	0.92 (0.51-1.66) 0.774	0.92 (0.51-1.68) 0.796	0.90 (0.49-1.65) 0.724
Baseline Data Not Available	Reference	Reference	Reference
Constant	0.11 (0.06-0.18) <0.001	0.10 (0.06-0.17) <0.001	0.11 (0.06-0.19) <0.001
Variance Components			
Residency Program	0.44 (0.26-0.73)	0.43 (0.26-0.72)	0.43 (0.26-0.72)
N Residents	3642	3642	3642
N Residency Programs	117	117	117

NOTE: Estimates are from a 3-level hierarchical logistic regression with hospital and program random intercepts. 5 integration points used in estimation. Dependent variables are coded '1' for residents reporting being 'very dissatisfied' or 'dissatisfied' and '0' for residents reporting being 'neutral,' 'satisfied' or 'very satisfied.' OR>1.00 indicate a higher odds of dissatisfaction among Flexible Policy arm residents.

TABLE S22. As-Treated Estimate of the Effect of Assignment to Flexible Policy on Odds of Resident Being (Very) Dissatisfied with Wellbeing (*Primary Resident Outcome*)

REGRESSOR	ODDS RATIO (95% CONFIDENCE INTERVAL) P-VALUE		
	Model 1	Model 2	Model 3
Sum of departures from ACGME standards	1.07 (0.99-1.15) 0.091	-----	-----
One departure	-----	2.21 (1.13-4.32) 0.020	-----
Two departures	-----	1.45 (0.75-2.80) 0.264	-----
Three departures	-----	0.69 (0.33-1.46) 0.331	-----
Four departures	-----	1.38 (1.03-1.86) 0.033	-----
No (Zero) departures	-----	Reference	-----
PGY1 duty hours can exceed 16 hours	-----	-----	1.40 (0.74-2.65) 0.301
PGY2+ duty hours can exceed 28 hours	-----	-----	0.92 (0.48-1.77) 0.812
Residents don't require 14 hours off after 24-hour duty	-----	-----	0.97 (0.49-1.93) 0.924
Residents don't require 8-10 hours off between shifts	-----	-----	1.03 (0.54-1.99) 0.920
2013 (Baseline) 30-Day Postoperative Death/Serious Morbidity			
Tertile 1	1.07 (0.99-1.15) 0.091	1.35 (0.81-2.225) 0.249	1.31 (0.76-2.25) 0.327
Tertile 2	1.31 (0.78-2.19) 0.314	1.29 (0.77-2.15) 0.333	1.21 (0.71-2.06) 0.482
Tertile 3	1.20 (0.72-2.02) 0.483	1.18 (0.71-1.96) 0.522	1.20 (0.71-2.04) 0.491
Baseline Data Not Available	Reference	Reference	Reference
Constant	0.10 (0.06-0.16) <0.001	0.10 (0.06-0.15) <0.001	0.10 (0.06-0.16) <0.001
Variance Components			
Residency Program	0.28 (0.16-0.49)	0.25 (0.14-0.44)	0.28 (0.16-0.49)
N Residents	3645	3645	3645
N Residency Programs	117	117	117

NOTE: Estimates are from a 3-level hierarchical logistic regression with hospital and program random intercepts. 5 integration points used in estimation. Dependent variables are coded '1' for residents reporting being 'very dissatisfied' or 'dissatisfied' and '0' for 'neutral', 'satisfied,' or 'very satisfied.' OR>1.00 indicate a higher odds of dissatisfaction among Flexible Policy arm residents.

TABLE S23. Instrumental Variables (IV) Estimates of the Local Average Treatment Effect on Resident Dissatisfaction with Education Quality (Primary Resident Outcome)

Model	Coefficient (95% CI)	P-Value
Linear Probability Models (LPM) (for Comparison)		
Model 1: Non-hierarchical LPM with Program Clustered Robust Standard Errors	0.006 (-0.027 - 0.038)	0.722
Model 2: 2-Level Hierarchical LPM with Program Random Effects	0.009 (-0.023 – 0.040)	0.600
Instrumental Variables (IV) Models		
Model 3: TSLS IV with Program Clustered Robust Standard Errors	0.006 (-0.027 – 0.039)	0.720
Model 4: 2-Level Hierarchical IV with Program Random Effects (G2SLS, Swamy-Arora variance components)	0.009 (-0.025 – 0.043)	0.602
Model 5: 2-Level Hierarchical IV with Program Random Effects (EC2SLS, Swamy-Arora variance components)	0.007 (-0.027 – 0.041)	0.682
Model 6: 2-Level Hierarchical IV with Program Random Effects (EC2SLS, Baltagi-Chang variance components)	0.007 (-0.026 – 0.040)	0.678

N Programs = 117 programs; N Residents = 3642 cases

Model 1 is an OLS linear probability model regressing resident outcomes on assignment to Flexible Policy (vs. Standard Policy) and tertiles of program-level observed rates of 2013 30-day postoperative death/serious morbidity with program-level clustered robust SEs (stratifying variable in randomization process). The reported coefficient corresponds to the effect of assignment to Flexible Policy on the probability of reporting dissatisfied or very dissatisfied vs. neutral, satisfied, or very satisfied (or a negative effect vs. a positive effect, or disagreement/strong disagreement vs. neutral, agreement or strong agreement). **Model 2** is a hierarchical linear probability model with program random intercepts. **Model 3** is a non-hierarchical two-stage least squares instrumental variables second-stage estimate where actual receipt of Flexible Policy (regardless of study arm) is instrumented by study arm assignment with program clustered robust SEs. **Model 4** is a hierarchical generalized two-stage least squares (G2SLS) second stage estimate of the effect of Flexible Policy instrumented by study arm assignment with program random intercepts. **Model 5** is a hierarchical error-corrected two-stage least squares (EC2SLS) second stage estimate of Flexible Policy instrumented by study arm assignment with program random intercepts using Swamy-Arora method for estimating variance components. **Model 6** is a hierarchical EC2SLS second stage estimate of Flexible Policy instrumented by study arm assignment with program random intercepts and Baltagi-Chang method for estimating variance components. The TSLS first stage study assignment coefficients were all > 0.95, all $p<0.001$. All first-stage F statistics were >2300.00, all $p<0.001$. All models controlled for program-level tertile of 2013 rates of 30-day postoperative death/serious morbidity (stratifying variable used in randomization of programs).

TABLE S24. Instrumental Variables (IV) Estimates of the Local Average Treatment Effect on Resident Dissatisfaction with Wellbeing (Primary Resident Outcome)

Model	Coefficient (95% CI)	P-Value
Linear Probability Models (LPM) (for Comparison)		
Model 1: Non-hierarchical LPM with Program Clustered Robust Standard Errors	0.029 (-0.006 – 0.063)	1.00
Model 2: 2-Level Hierarchical LPM with Program Random Effects	0.031 (-0.001 – 0.063)	0.056
Instrumental Variables (IV) Models		
Model 3: TSLS IV with Program Clustered Robust Standard Errors	0.030 (-0.005 – 0.064)	0.094
Model 4: 2-Level Hierarchical IV with Program Random Effects (G2SLS, Swamy-Arora variance components)	0.032 (-0.001 – 0.066)	0.059
Model 5: 2-Level Hierarchical IV with Program Random Effects (EC2SLS, Swamy-Arora variance components)	0.030 (-0.003 – 0.064)	0.076
Model 6: 2-Level Hierarchical IV with Program Random Effects (EC2SLS, Baltagi-Chang variance components)	0.030 (-0.002 – 0.063)	0.070

N Programs = 117 programs; N Residents = 3645 cases

Model 1 is an OLS linear probability model regressing resident outcomes on assignment to Flexible Policy (vs. Standard Policy) and tertiles of program-level observed rates of 2013 30-day postoperative death/serious morbidity with program-level clustered robust SEs (stratifying variable in randomization process). The reported coefficient corresponds to the effect of assignment to Flexible Policy on the probability of reporting dissatisfied or very dissatisfied vs. neutral, satisfied, or very satisfied (or a negative effect vs. a positive effect, or disagreement/strong disagreement vs. neutral, agreement or strong agreement). **Model 2** is a hierarchical linear probability model with program random intercepts. **Model 3** is a non-hierarchical two-stage least squares instrumental variables second-stage estimate where actual receipt of Flexible Policy (regardless of study arm) is instrumented by study arm assignment with program clustered robust SEs. **Model 4** is a hierarchical generalized two-stage least squares (G2SLS) second stage estimate of the effect of Flexible Policy instrumented by study arm assignment with program random intercepts. **Model 5** is a hierarchical error-corrected two-stage least squares (EC2SLS) second stage estimate of Flexible Policy instrumented by study arm assignment with program random intercepts using Swamy-Arora method for estimating variance components. **Model 6** is a hierarchical EC2SLS second stage estimate of Flexible Policy instrumented by study arm assignment with program random intercepts and Baltagi-Chang method for estimating variance components. The TSLS first stage study assignment coefficients were all > 0.95, all $p<0.001$. All first-stage F statistics were >2300.00, all $p<0.001$. All models controlled for program-level tertile of 2013 rates of 30-day postoperative death/serious morbidity (stratifying variable used in randomization of programs).

F. NOTE ON ABSITE RESIDENT SURVEY RESPONSE RATES

The 2015 ABSITE Resident Survey was administered to all general surgery residents who took the January 2015 ABSITE examination (irrespective of whether training in a FIRST Trial-participating program), of which 4,330 were residents in the 117 general surgery residency programs that participated in the FIRST Trial. Of these 4,330 residents, 2220 residents were in the 59 programs randomized to Standard Policy, and 2110 residents were in the 58 programs randomized to Flexible Policy.

Of the 4,330 respondents in the FIRST Trial resident survey sample, 585 (13.05%) had missing values on all 34 survey items reported in the tables below.

Overall Response Rates

Overall Response Rates. Table S25 shows item-level missing-value frequencies and response rates for the FIRST Trial ABSITE Resident Survey sample for the resident outcomes we studied. Response rates for survey items ranged from 84% to 87%.

Response Rates by Gender. Table S26 presents item-level response rates by gender for outcomes we studied. There were no statistically significant gender differences in response rates for any of our primary or secondary endpoints in the FIRST Trial.

Overall Response Rates by Postgraduate Year. Item-level response rates by resident's postgraduate year (PGY) are reported in Table S27. PGY1, PGY4 and PGY5 residents had response rates exceeding 90%, but PGY2 and PGY3 response rates ranged in the 70-79% range (program-level cluster-corrected chi-square p-values all <0.01).

Response Rates by Program Type. Table S28 presents item-level response rates by program type. Community programs had the highest response rates (90%+ range), followed by military programs (approximately 89%) and academic programs (approximately 81%, program-level cluster-corrected chi-square p-values all <0.01).

Response Rates by Geographic Region. Table S29 presents item-level response rates by geographic region of program location. We found no regional differences in response rates for any of the survey items.

Response Rates by Study Arm

Table S30 shows item-level response rates and frequencies for missing values for the FIRST Trial ABSITE Resident Survey sample, by study arm. This table also provides p-values for chi-square tests of association between study arm and non-response for each item. We found no differences in response rates across study arms for any of the outcomes we studied.

In our Statistical Analysis Plan (§F.3), we made provisions for various approaches to treat missing data. However, we chose not to impute values (singly or multiply) for the analyses we undertook in this report because all of the limited number of resident variables (e.g., PGY and gender) in this analysis were used as dependent variables. While the inclusion of outcome variables in imputation models for missing independent variables (Xs) is standard practice, to our knowledge, regression on imputed values in the dependent variable (Ys) is not supported strongly, if at all (some articles argue against it) due to noise (e.g. White et al. 2011; von Hippel 2007). Thus, we did not impute missing values in our analyses of resident outcomes. Moreover, nearly all missing data are due to residents not taking any part of the survey, thus all of the data are missing for that resident. Thus, a complete case analysis was performed where item response rates ranged from 84% to 87%.

TABLE S25. 2015 ABSITE Resident Survey Overall Item Response Rates for FIRST Trial Resident Sample

SURVEY ITEM	MISSING	NON-MISSING	TOTAL	ITEM RESPONSE RATE
Perceived effect of duty hours on: Patient safety	657	3,673	4,330	85%
Perceived effect of duty hours on: Continuity of care	652	3,678	4,330	85%
Perceived effect of duty hours on: Attendance at required educational conferences	664	3,666	4,330	85%
Perceived effect of duty hours on: Ability to acquire clinical skills	665	3,665	4,330	85%
Perceived effect of duty hours on: Ability to acquire operative skills	664	3,666	4,330	85%
Perceived effect of duty hours on: Resident autonomy	660	3,670	4,330	85%
Perceived effect of duty hours on: Number of operations performed	665	3,665	4,330	85%
Perceived effect of duty hours on: Availability for elective cases	660	3,670	4,330	85%
Perceived effect of duty hours on: Availability for urgent cases	657	3,673	4,330	85%
Perceived effect of duty hours on: Time to teach medical students	661	3,669	4,330	85%
Perceived effect of duty hours on: Relationship between interns/residents	656	3,674	4,330	85%
Perceived effect of duty hours on: Professionalism	659	3,671	4,330	85%
Perceived effect of duty hours on: Morale	656	3,674	4,330	85%
Perceived effect of duty hours on: Ability to prepare for cases away from the hospital	662	3,668	4,330	85%
Perceived effect of duty hours on: Ability to participate in research	662	3,668	4,330	85%
Perceived effect of duty hours on: Job Satisfaction	660	3,670	4,330	85%
Perceived effect of duty hours on: Satisfaction with decision to become a surgeon	666	3,664	4,330	85%
Perceived effect of duty hours on: Time with family and friends	663	3,667	4,330	85%
Perceived effect of duty hours on: Time for hobbies/extracurricular activities	665	3,665	4,330	85%
Perceived effect of duty hours on: Own health	669	3,661	4,330	85%
Perceived effect of duty hours on: Rest	662	3,668	4,330	85%
Satisfaction with: Continuity of care	685	3,645	4,330	84%
Satisfaction with: Patient safety	685	3,645	4,330	84%
Satisfaction with: Work hours/scheduling	689	3,641	4,330	84%
Satisfaction with: Quality/ease of handoffs and transitions	691	3,639	4,330	84%
Satisfaction with: Resident education quality	688	3,642	4,330	84%

TABLE S25 (continued). 2015 ABSITE Resident Survey Overall Item Response Rates for FIRST Trial Resident Sample

SURVEY ITEM	MISSING	NON-MISSING	TOTAL	ITEM RESPONSE RATE
Satisfaction with: Time for rest	687	3,643	4,330	84%
Satisfaction with: Wellbeing	685	3,645	4,330	84%
Satisfaction with: Work hour regulations	686	3,644	4,330	84%
How often fatigue affected: personal safety	678	3,652	4,330	84%
How often fatigue affected: patient safety	678	3,652	4,330	84%
Frequency in last month: hand off active patient care issue because of duty hour limits	565	3,765	4,330	87%
Frequency in last month: leave during operation because of duty hour limits	565	3,765	4,330	87%
Frequency in last month: miss operation because of duty hour limits	565	3,765	4,330	87%

TABLE S26. 2015 ABSITE Resident Survey Overall Item Response Rates for FIRST Trial Resident Sample, by Gender

SURVEY ITEM	ITEM RESPONSE RATE (%)		P-VALUE†
	MALE	FEMALE	
Perceived effect of duty hours on: Patient safety	85.62	83.65	0.30
Perceived effect of duty hours on: Continuity of care	85.69	83.82	0.32
Perceived effect of duty hours on: Attendance at required educational conferences	85.46	83.48	0.30
Perceived effect of duty hours on: Ability to acquire clinical skills	85.35	83.59	0.36
Perceived effect of duty hours on: Ability to acquire operative skills	85.38	83.59	0.35
Perceived effect of duty hours on: Resident autonomy	85.50	83.65	0.34
Perceived effect of duty hours on: Number of operations performed	85.35	83.59	0.36
Perceived effect of duty hours on: Availability for elective cases	85.54	83.59	0.32
Perceived effect of duty hours on: Availability for urgent cases	85.58	83.71	0.33
Perceived effect of duty hours on: Time to teach medical students	85.54	83.53	0.30
Perceived effect of duty hours on: Relationship between interns/residents	85.58	83.77	0.34
Perceived effect of duty hours on: Professionalism	85.58	83.59	0.30
Perceived effect of duty hours on: Morale	85.54	83.82	0.37
Perceived effect of duty hours on: Ability to prepare for cases away from the hospital	85.38	83.71	0.38
Perceived effect of duty hours on: Ability to participate in research	85.35	83.77	0.41
Perceived effect of duty hours on: Job Satisfaction	85.38	83.82	0.41
Perceived effect of duty hours on: Satisfaction with decision to become a surgeon	85.23	83.71	0.43
Perceived effect of duty hours on: Time with family and friends	85.31	83.77	0.42
Perceived effect of duty hours on: Time for hobbies/extracurricular activities	85.27	83.71	0.42
Perceived effect of duty hours on: Own health	83.15	83.65	0.43
Perceived effect of duty hours on: Rest	85.35	83.77	0.41
Satisfaction with: Continuity of care	84.73	83.36	0.49
Satisfaction with: Patient safety	84.73	83.36	0.49
Satisfaction with: Work hours/scheduling	84.65	83.25	0.48
Satisfaction with: Quality/ease of handoffs and transitions	84.65	83.13	0.44
Satisfaction with: Resident education quality	84.69	83.25	0.47
Satisfaction with: Time for rest	84.69	83.30	0.49
Satisfaction with: Wellbeing	84.73	83.36	0.49
Satisfaction with: Work hour regulations	84.73	83.30	0.48
How often fatigue affected: personal safety	84.88	83.53	0.50
How often fatigue affected: patient safety	84.88	83.53	0.50
Frequency in last month: hand off active patient care issue because of duty hour limits	87.58	86.01	0.38
Frequency in last month: leave during operation because of duty hour limits	87.58	86.01	0.38
Frequency in last month: miss operation because of duty hour limits	87.58	86.01	0.38

†Program-level cluster-corrected chi-square p-values

TABLE S27. 2015 ABSITE Resident Survey Overall Item Response Rates for FIRST Trial Resident Sample, by Postgraduate Year (PGY)

SURVEY ITEM	ITEM RESPONSE RATE (%)					P-VALUE†
	PGY1	PGY2	PGY3	PGY4	PGY5	
Perceived effect of duty hours on: Patient safety	94.64	71.79	75.11	93.47	94.60	<0.01
Perceived effect of duty hours on: Continuity of care	94.55	72.06	75.00	93.79	94.94	<0.01
Perceived effect of duty hours on: Attendance at required educational conferences	94.29	71.60	74.89	93.47	94.77	<0.01
Perceived effect of duty hours on: Ability to acquire clinical skills	94.12	71.97	94.54	93.47	94.77	<0.01
Perceived effect of duty hours on: Ability to acquire operative skills	94.12	71.79	75.00	93.63	94.44	<0.01
Perceived effect of duty hours on: Resident autonomy	94.20	72.06	74.89	9379	94.44	<0.01
Perceived effect of duty hours on: Number of operations performed	94.20	71.69	74.77	93.63	94.60	<0.01
Perceived effect of duty hours on: Availability for elective cases	94.29	71.97	74.77	93.63	94.77	<0.01
Perceived effect of duty hours on: Availability for urgent cases	94.38	71.88	74.89	93.95	94.77	<0.01
Perceived effect of duty hours on: Time to teach medical students	94.38	71.88	74.77	93.47	94.77	<0.01
Perceived effect of duty hours on: Relationship between interns/residents	94.38	71.97	74.89	93.79	94.94	<0.01
Perceived effect of duty hours on: Professionalism	94.29	71.88	74.77	93.79	94.94	<0.01
Perceived effect of duty hours on: Morale	94.29	71.97	75.00	93.79	94.94	<0.01
Perceived effect of duty hours on: Ability to prepare for cases away from the hospital	94.38	71.79	75.11	93.15	94.60	<0.01
Perceived effect of duty hours on: Ability to participate in research	94.38	71.88	75.00	93.31	94.44	<0.01
Perceived effect of duty hours on: Job satisfaction	94.55	71.79	75.00	93.47	94.44	<0.01
Perceived effect of duty hours on: Satisfaction with decision to become a surgeon	94.29	71.79	74.89	92.99	94.60	<0.01
Perceived effect of duty hours on: Time with family and friends	94.29	71.79	74.89	93.47	94.60	<0.01
Perceived effect of duty hours on: Time for hobbies/extracurricular activities	94.20	71.79	75.00	93.31	94.44	<0.01
Perceived effect of duty hours on: Own health	94.38	71.51	74.66	93.31	94.44	<0.01
Perceived effect of duty hours on: Rest	94.38	71.88	74.89	93.31	94.60	<0.01
Satisfaction with: Continuity of care	93.08	71.60	74.66	92.99	94.44	<0.01
Satisfaction with: Patient safety	93.08	71.60	74.66	92.99	94.44	<0.01
Satisfaction with: Work hours/scheduling	92.91	71.60	74.66	92.99	94.10	<0.01
Satisfaction with: Quality/ease of handoffs and transitions	92.91	71.32	74.54	92.99	94.44	<0.01
Satisfaction with: Resident education quality	92.99	71.51	74.66	92.99	94.27	<0.01

†Program-level cluster-corrected chi-square p-values for tests of association between PG year and response rate

TABLE S27 (continued). 2015 ABSITE Resident Survey Overall Item Response Rates for FIRST Trial Resident Sample, by Postgraduate Year (PGY)

SURVEY ITEM	ITEM RESPONSE RATE (%)					P-VALUE†
	PGY1	PGY2	PGY3	PGY4	PGY5	
Satisfaction with: Time for rest	93.08	71.60	74.54	92.99	94.27	<0.01
Satisfaction with: Wellbeing	93.17	71.60	74.66	92.83	94.44	<0.01
Satisfaction with: Work hour regulations	93.17	71.60	74.54	92.83	94.44	<0.01
How often fatigue affected: personal safety	93.60	71.60	74.77	92.99	94.44	<0.01
How often fatigue affected: patient safety	93.60	71.60	74.77	92.99	94.44	<0.01
What effect would a change in duty hour policy have on: Safety of patient care?	92.99	71.32	74.66	92.99	94.10	<0.01
What effect would a change in duty hour policy have on: Continuity of care?	92.99	71.32	74.66	92.99	94.10	<0.01
What effect would a change in duty hour policy have on: Quality of resident education?	92.91	71.14	74.66	92.83	94.10	<0.01
Frequency in last month: hand off active patient care issue because of duty hour limits	97.32	73.36	76.38	96.34	97.13	<0.01
Frequency in last month: leave during operation because of duty hour limits	97.32	73.36	76.38	96.34	97.13	<0.01
Frequency in last month: miss operation because of duty hour limits	97.32	73.36	76.38	96.34	97.13	<0.01

†Program-level cluster-corrected chi-square p-values

TABLE S28. 2015 ABSITE Resident Survey Overall Item Response Rates for FIRST Trial Resident Sample, by Program Type

SURVEY ITEM	ITEM RESPONSE RATE (%)			P-VALUE†
	ACADEMIC	COMMUNITY	MILITARY	
Perceived effect of duty hours on: Patient safety	82.18	92.57	88.64	<0.01
Perceived effect of duty hours on: Continuity of care	82.37	92.47	88.64	<0.01
Perceived effect of duty hours on: Attendance at required educational conferences	82.02	92.38	88.64	<0.01
Perceived effect of duty hours on: Ability to acquire clinical skills	81.99	92.47	86.36	<0.01
Perceived effect of duty hours on: Ability to acquire operative skills	81.99	92.47	88.64	<0.01
Perceived effect of duty hours on: Resident autonomy	82.18	92.29	88.64	<0.01
Perceived effect of duty hours on: Number of operations performed	82.02	92.29	88.64	<0.01
Perceived effect of duty hours on: Availability for elective cases	82.12	92.57	86.36	<0.01
Perceived effect of duty hours on: Availability for urgent cases	82.18	92.57	88.64	<0.01
Perceived effect of duty hours on: Time to teach medical students	82.06	92.57	88.64	<0.01
Perceived effect of duty hours on: Relationship between interns/residents	82.21	92.57	88.64	<0.01
Perceived effect of duty hours on: Professionalism	82.15	92.47	88.64	<0.01
Perceived effect of duty hours on: Morale	82.21	92.57	88.64	<0.01
Perceived effect of duty hours on: Ability to prepare for cases away from the hospital	82.09	92.38	88.64	<0.01
Perceived effect of duty hours on: Ability to participate in research	82.09	92.38	88.64	<0.01
Perceived effect of duty hours on: Job satisfaction	82.18	92.29	88.64	<0.01
Perceived effect of duty hours on: Satisfaction with decision to become a surgeon	82.06	92.10	88.64	<0.01
Perceived effect of duty hours on: Time with family and friends	82.12	92.29	86.36	<0.01
Perceived effect of duty hours on: Time for hobbies/extracurricular activities	82.02	92.29	88.64	<0.01
Perceived effect of duty hours on: Own health	81.99	92.01	88.64	<0.01
Perceived effect of duty hours on: Rest	82.12	92.29	88.64	<0.01
Satisfaction with: Continuity of care	81.50	92.01	88.64	<0.01
Satisfaction with: Patient safety	81.50	92.01	88.64	<0.01
Satisfaction with: Work hours/scheduling	81.40	91.91	88.64	<0.01
Satisfaction with: Quality/ease of handoffs and transitions	81.34	91.91	88.64	<0.01
Satisfaction with: Resident education quality	81.43	91.91	88.64	<0.01

† Program-level cluster-corrected chi-square p-values

TABLE S28 (continued). 2015 ABSITE Resident Survey Overall Item Response Rates for FIRST Trial Resident Sample, by Program Type

SURVEY ITEM	ITEM RESPONSE RATE (%)			P-VALUE†
	ACADEMIC	COMMUNITY	MILITARY	
Satisfaction with: Time for rest	81.50	91.82	88.64	<0.01
Satisfaction with: Wellbeing	81.50	92.01	88.64	<0.01
Satisfaction with: Work hour regulations	81.46	92.01	88.64	<0.01
How often fatigue affected: personal safety	81.71	92.01	88.64	<0.01
How often fatigue affected: patient safety	81.71	92.01	88.64	<0.01
Frequency in last month: hand off active patient care issue because of duty hour limits	84.45	94.24	90.91	<0.01
Frequency in last month: leave during operation because of duty hour limits	84.45	94.24	90.91	<0.01
Frequency in last month: miss operation because of duty hour limits	84.45	94.24	90.91	<0.01

†Program-level cluster-corrected chi-square p-values

TABLE S29. 2015 ABSITE Resident Survey Overall Item Response Rates for FIRST Trial Resident Sample, by Geographic Region

SURVEY ITEM	ITEM RESPONSE RATE (%)					P-VALUE†
	NORTHEAST	SOUTHEAST	MIDWEST	SOUTHWEST	WEST	
Perceived effect of duty hours on: Patient safety	82.82	88.21	85.05	87.60	80.71	0.11
Perceived effect of duty hours on: Continuity of care	82.98	88.21	85.23	87.40	81.07	0.14
Perceived effect of duty hours on: Attendance at required educational conferences	82.57	88.21	85.05	87.21	80.36	0.09
Perceived effect of duty hours on: Ability to acquire clinical skills	82.74	87.66	85.05	87.40	80.54	0.14
Perceived effect of duty hours on: Ability to acquire operative skills	82.49	87.99	85.14	87.21	80.71	0.12
Perceived effect of duty hours on: Resident autonomy	82.65	88.10	85.05	87.40	80.89	0.13
Perceived effect of duty hours on: Number of operations performed	82.65	87.99	84.86	87.21	80.71	0.12
Perceived effect of duty hours on: Availability for elective cases	82.65	87.99	85.32	87.40	80.54	0.12
Perceived effect of duty hours on: Availability for urgent cases	82.82	88.10	85.23	87.21	80.89	0.14
Perceived effect of duty hours on: Time to teach medical students	82.65	87.99	85.23	87.02	80.89	0.14
Perceived effect of duty hours on: Relationship between interns/residents	82.74	88.21	85.23	87.40	80.89	0.12
Perceived effect of duty hours on: Professionalism	82.74	87.99	85.23	87.40	80.71	0.12
Perceived effect of duty hours on: Morale	82.82	88.21	85.14	87.40	80.89	0.12
Perceived effect of duty hours on: Ability to prepare for cases away from the hospital	82.57	88.21	85.05	87.40	80.54	0.09
Perceived effect of duty hours on: Ability to participate in research	82.65	88.10	85.14	87.21	80.54	0.11
Perceived effect of duty hours on: Job satisfaction	82.74	88.21	85.05	87.02	80.89	0.13
Perceived effect of duty hours on: Satisfaction with decision to become a surgeon	82.49	88.10	85.05	87.21	80.36	0.09

† Program-level cluster-corrected chi-square p-values

TABLE S29 (continued). 2015 ABSITE Resident Survey Overall Item Response Rates for FIRST Trial Resident Sample, by Geographic Region

SURVEY ITEM	ITEM RESPONSE RATE (%)					P-VALUE†
	NORTHEAST	SOUTHEAST	MIDWEST	SOUTHWEST	WEST	
Perceived effect of duty hours on: Time with family and friends	82.74	87.99	85.05	87.02	80.71	0.13
Perceived effect of duty hours on: Time for hobbies/extracurricular activities	82.57	88.21	84.95	87.02	80.54	0.12
Perceived effect of duty hours on: Own health	82.49	88.10	84.86	86.82	80.54	0.11
Perceived effect of duty hours on: Rest	82.65	88.10	85.05	87.21	80.71	0.12
Satisfaction with: Continuity of care	82.17	87.77	84.50	86.82	79.64	0.10
Satisfaction with: Patient safety	82.17	87.77	84.41	86.82	79.82	0.11
Satisfaction with: Work hours/scheduling	81.92	87.77	84.50	86.63	79.64	0.09
Satisfaction with: Quality/ease of handoffs and transitions	81.92	87.66	84.41	86.43	79.82	0.11
Satisfaction with: Resident education quality	82.17	87.66	84.41	86.82	79.46	0.10
Satisfaction with: Time for rest	82.17	87.55	84.50	86.82	79.64	0.11
Satisfaction with: Wellbeing	82.17	87.77	84.41	86.82	79.82	0.11
Satisfaction with: Work hour regulations	82.08	87.66	84.50	86.82	79.82	0.11
How often fatigue affected: personal safety	82.25	87.77	84.77	86.82	80.18	0.13
How often fatigue affected: patient safety	82.25	87.77	84.77	86.82	80.18	0.13
Frequency in last month: hand off active patient care issue because of duty hour limits	84.61	90.28	86.94	89.92	83.93	0.09
Frequency in last month: leave during operation because of duty hour limits	84.61	90.28	86.94	89.92	83.93	0.09
Frequency in last month: miss operation because of duty hour limits	84.61	90.28	86.94	89.92	83.93	0.09

† Program-level cluster-corrected chi-square p-values

TABLE S30. 2015 ABSITE Resident Survey Item Response Rates for FIRST Trial Resident Sample by Study Arm

SURVEY ITEM	Standard Policy				Flexible Policy				Chi-Square P-Value†
	MISSING	NON-MISSING	TOTAL	ITEM RESPONSE RATE	MISSING	NON-MISSING	TOTAL	ITEM RESPONSE RATE	
Perceived effect of duty hours on: Patient safety	329	1,891	2220	85%	328	1,782	2110	84%	0.70
Perceived effect of duty hours on: Continuity of care	328	1,892	2220	85%	324	1,786	2110	85%	0.75
Perceived effect of duty hours on: Attendance at required educational conferences	334	1,886	2220	85%	330	1,780	2110	84%	0.75
Perceived effect of duty hours on: Ability to acquire clinical skills	332	1,888	2220	85%	333	1,777	2110	84%	0.66
Perceived effect of duty hours on: Ability to acquire operative skills	335	1,885	2220	85%	329	1,781	2110	84%	0.79
Perceived effect of duty hours on: Resident autonomy	332	1,888	2220	85%	328	1,782	2110	84%	0.76
Perceived effect of duty hours on: Number of operations performed	333	1,887	2220	85%	332	1,778	2110	84%	0.69
Perceived effect of duty hours on: Availability for elective cases	331	1,889	2220	85%	329	1,781	2110	84%	0.72
Perceived effect of duty hours on: Availability for urgent cases	330	1,890	2220	85%	327	1,783	2110	85%	0.74
Perceived effect of duty hours on: Time to teach medical students	332	1,888	2220	85%	329	1,781	2110	84%	0.73
Perceived effect of duty hours on: Relationship between interns/residents	328	1,892	2220	85%	328	1,782	2110	84%	0.68
Perceived effect of duty hours on: Professionalism	329	1,891	2220	85%	330	1,780	2110	84%	0.66
Perceived effect of duty hours on: Morale	328	1,892	2220	85%	328	1,782	2110	84%	0.68
Perceived effect of duty hours on: Ability to prepare for cases away from the hospital	333	1,887	2220	85%	329	1,781	2110	84%	0.75
Perceived effect of duty hours on: Ability to participate in research	332	1,888	2220	85%	330	1,780	2110	84%	0.72
Perceived effect of duty hours on: Job satisfaction	332	1,888	2220	85%	328	1,782	2110	84%	0.75
Perceived effect of duty hours on: Satisfaction with decision to become a surgeon	333	1,887	2220	85%	333	1,777	2110	84%	0.68
Perceived effect of duty hours on: Time with family and friends	332	1,888	2220	85%	331	1,779	2110	84%	0.69

†Program-level cluster-corrected chi-square test for association between study arm and missing value

TABLE S30 (continued). 2015 ABSITE Resident Survey Item Response Rates for FIRST Trial Resident Sample by Study Arm

SURVEY ITEM	Standard Policy				Flexible Policy				Chi-Square P-Value†
	MISSING	NON-MISSING	TOTAL	ITEM RESPONSE RATE	MISSING	NON-MISSING	TOTAL	ITEM RESPONSE RATE	
Perceived effect of duty hours on: Time for hobbies/extracurricular activities	334	1,886	2220	85%	331	1,779	2110	84%	0.74
Perceived effect of duty hours on: Own health	337	1,883	2220	85%	332	1,778	2110	84%	0.77
Perceived effect of duty hours on: Rest	333	1,887	2220	85%	329	1,781	2110	84%	0.75
Satisfaction with: Continuity of care	344	1,876	2220	85%	341	1,769	2110	84%	0.73
Satisfaction with: Patient safety	345	1,875	2220	84%	340	1,770	2110	84%	0.77
Satisfaction with: Work hours/scheduling	346	1,874	2220	84%	343	1,767	2110	84%	0.73
Satisfaction with: Quality/ease of handoffs and transitions	347	1,873	2220	84%	344	1,766	2110	84%	0.73
Satisfaction with: Resident education quality	346	1,874	2220	84%	342	1,768	2110	84%	0.75
Satisfaction with: Time for rest	345	1,875	2220	84%	342	1,768	2110	84%	0.73
Satisfaction with: Wellbeing	344	1,876	2220	85%	341	1,769	2110	84%	0.73
Satisfaction with: Work hour regulations	344	1,876	2220	85%	342	1,768	2110	84%	0.72
How often fatigue affected: personal safety	342	1,878	2220	85%	336	1,774	2110	84%	0.79
How often fatigue affected: patient safety	342	1,878	2220	85%	336	1,774	2110	84%	0.79
Frequency in last month: hand off active patient care issue because of duty hour limits	276	1944	2220	88%	289	1821	2110	86%	0.47
Frequency in last month: leave during operation because of duty hour limits	276	1944	2220	88%	289	1821	2110	86%	0.47
Frequency in last month: miss operation because of duty hour limits	276	1944	2220	88%	289	1821	2110	86%	0.47

†Program-level cluster-corrected chi-square test for association between study arm and missing value

G. PATIENT OUTCOMES RISK ADJUSTMENT VARIABLES

30-Day Postoperative Death/Serious Morbidity (Primary Outcome)

List of patient characteristics included in adjusted models:

- Age (age <65 (reference), 65-74, 75-84, 85+)
- American Society of Anesthesiologists (ASA) Physical Status Classification (normal (reference), mild systemic disease, severe systemic disease, life-threatening systemic disease/moribund)
- ACS NSQIP outcome-specific CPT-based linear predictor for 30-day postoperative death/serious morbidity
- Emergency/urgent surgery (reference: not emergency/urgent)
- Preoperative functional status (independent (reference), partially dependent, totally dependent)
- Male sex (reference: female)
- Wound class (clean (reference), clean/contaminated, contaminated, dirty/infected)

Note: Using an expanded list of covariates (from ACS NSQIP model for 30-day postoperative mortality) did not alter our results

30-Day Postoperative Death (Secondary Outcome)

List of patient characteristics included in adjusted models:

- Preoperative sepsis (none (reference), systemic inflammatory response syndrome (SIRS), sepsis, septic shock
- American Society of Anesthesiologists (ASA) Physical Status Classification (normal (reference), mild systemic disease, severe systemic disease, life-threatening systemic disease/moribund)
- ACS NSQIP outcome-specific CPT-based linear predictor for 30-day postoperative death
- Age (in years)
- Preoperative albumin
- Disseminated cancer (reference: none)
- Serum glutamic oxidase transaminase >40 (reference: ≤40)
- Preoperative functional status (independent (reference), partially dependent, totally dependent)
- Ascites (reference: none)
- Preoperative platelet count<150 (reference: ≥150)
- Dyspnea (none (reference), at rest, exertional)
- Body mass index classification (normal (reference), Class I obese, Class II obese, Class III obese, overweight, underweight)
- Inpatient surgery setting (reference: outpatient)
- Preoperative ventilator dependence (reference: none)
- Alkaline phosphatase>125 (reference: ≤125)
- Emergency/urgent surgery (reference: not emergency/urgent)
- Chronic obstructive pulmonary disease (reference: none)
- Preoperative prothrombin time>35 (reference: ≤35)
- Preoperative weight loss >10% (reference: none)
- Congestive heart failure (reference: none)
- Preoperative blood urea nitrogen>40 (reference: ≤40)
- Steroid use (reference: none)
- Preoperative sodium >145 (reference: ≤145)
- Transfer status (admit from home (reference), transfer from acute care, transfer from chronic care, transfer from other facility, transfer from outside emergency department)
- Current smoker (reference: not current smoker)
- Preoperative hematocrit >45 (reference: ≤45)
- Preoperative dialysis (reference: none)
- Patient race (White (reference), Black/African American, American Indian/Native Alaskan, Native Hawaiian/Pacific Islander, Asian, Unknown)
- Preoperative white blood cell count ≤4.5 (reference: >4.5)

30-Day Postoperative Serious Morbidity (Secondary Outcome)

List of patient characteristics included in adjusted models:

- Age (age <65 (reference), 65-74, 75-84, 85+)
- American Society of Anesthesiologists (ASA) Physical Status Classification (normal (reference), mild systemic disease, severe systemic disease, life-threatening systemic disease/moribund)
- ACS NSQIP outcome-specific CPT-based linear predictor for 30-day postoperative serious morbidity
- Emergency/urgent surgery (reference: not emergency/urgent)
- Preoperative functional status (independent (reference), partially dependent, totally dependent)
- Male sex (reference: female)
- Wound class (clean (reference), clean/contaminated, contaminated, dirty/infected)

30-Day Postoperative Any Morbidity (Secondary Outcome)

List of patient characteristics included in adjusted models:

- ACS NSQIP outcome-specific CPT-based linear predictor for 30-day postoperative any morbidity
- American Society of Anesthesiologists (ASA) Physical Status Classification (normal (reference), mild systemic disease, severe systemic disease, life-threatening systemic disease/moribund)
- Inpatient surgery setting (reference: outpatient)
- Preoperative sepsis (none (reference), systemic inflammatory response syndrome (SIRS), sepsis, septic shock)
- Preoperative albumin
- Preoperative creatinine >1.2 (reference: ≤1.2)
- Body mass index classification (normal (reference), Class I obese, Class II obese, Class III obese, overweight, underweight)
- Age
- Current smoker (reference: none)
- Work relative value unit
- Hispanic (no (reference), yes, unknown), preoperative functional status (independent (reference), partially dependent, totally dependent)
- Preoperative ventilator dependence (reference: none)
- Bleeding disorder (reference: none)
- Serum glutamic oxidase transaminase >40 (reference: ≤40)
- Steroid use (reference: none)
- Preoperative blood urea nitrogen>40 (reference: ≤40)
- Dyspnea (none (reference), at rest, exertional)
- Emergency/urgent surgery (reference: not emergency/urgent)
- Chronic obstructive pulmonary disease (reference: none)
- Preoperative prothrombin time>35 (reference: ≤35)
- Congestive heart failure (reference: none)
- Patient race (White (reference), Black/African American, American Indian/Native Alaskan, Native Hawaiian/Pacific Islander, Asian, Unknown)
- Wound class (clean (reference), clean/contaminated, contaminated, dirty/infected)
- Hypertension (reference: none)
- Disseminated cancer (reference: none)
- Diabetes (none (reference), insulin, oral)
- Preoperative dialysis (reference: none)
- Male sex (reference: female)
- Preoperative alkaline phosphatase>125 (reference: ≤125)
- Preoperative transfusion (reference: none)
- Ascites (reference: none)

30-Day Postoperative Any Morbidity (continued)

- Preoperative weight loss >10% (reference: none)
- Preoperative hematocrit >45 (reference: ≤45)
- Preoperative platelet count (<150, 150-400 (reference), >400)
- Preoperative renal failure (reference: none)
- Preoperative white blood cell count >11 (reference: ≤11)
- Preoperative sodium (<135, 135-145 (reference), >145)

30-Day Postoperative Failure-to-Rescue (Secondary Outcome)

List of patient characteristics included in adjusted models:

- Age (age <65 (reference), 65-74, 75-84, 85+)
- American Society of Anesthesiologists (ASA) Physical Status Classification (normal (reference), mild systemic disease, severe systemic disease, life-threatening systemic disease/moribund)
- ACS NSQIP outcome-specific CPT-based linear predictor for 30-day postoperative failure-to-rescue
- Emergency/urgent surgery (reference: not emergency/urgent)
- Preoperative functional status (independent (reference), partially dependent, totally dependent)
- Male sex (reference: female)
- Wound class (clean (reference), clean/contaminated, contaminated, dirty/infected)

30-Day Postoperative Pneumonia (Secondary Outcome)

List of patient characteristics included in adjusted models:

- American Society of Anesthesiologists (ASA) Physical Status Classification (normal (reference), mild systemic disease, severe systemic disease, life-threatening systemic disease/moribund)
- ACS NSQIP outcome-specific CPT-based linear predictor for 30-day postoperative pneumonia
- Preoperative sepsis (none (reference), systemic inflammatory response syndrome (SIRS), sepsis, septic shock)
- Inpatient surgery (reference: outpatient)
- Age
- Current smoker (reference: not current smoker)
- Preoperative functional status (independent (reference), partially dependent, totally dependent)
- Chronic obstructive pulmonary disease (reference: none)
- Male sex (reference: female)
- Hispanic (no (reference), yes, unknown)
- Preoperative albumin
- Preoperative creatinine >1.2 (reference: ≤1.2)
- Work relative value unit, dyspnea (none (reference), at rest, exertional)
- Bleeding disorder (reference: none)
- Patient race (White (reference), Black/African American, American Indian/Native Alaskan, Native Hawaiian/Pacific Islander, Asian, Unknown)
- Preoperative transfusion (reference: none)
- Preoperative hematocrit >45 (reference: ≤45)
- Ascites (reference: none)
- Steroid use (reference: none)
- Preoperative dialysis (reference: none)

30-Day Postoperative Renal Failure (Secondary Outcome)

List of patient characteristics included in adjusted models:

- Preoperative sepsis (none (reference), systemic inflammatory response syndrome (SIRS), sepsis, septic shock)
- ACS NSQIP outcome-specific CPT-based linear predictor for 30-day postoperative renal failure
- Preoperative creatinine >1.2 (reference: ≤1.2)
- American Society of Anesthesiologists (ASA) Physical Status Classification (normal (reference), mild systemic disease, severe systemic disease, life-threatening systemic disease/moribund)
- Inpatient surgery setting (reference: outpatient)
- Hypertension (reference: none)
- Preoperative albumin
- Male sex (reference: female)
- Body mass index classification (normal (reference), Class I obese, Class II obese, Class III obese, overweight, underweight)
- Serum glutamic oxidase transaminase >40 (reference: ≤40)
- Dyspnea (none (reference), at rest, exertional)
- Current smoker (reference: not current smoker)
- Preoperative platelet count <150 (reference: ≥150)
- Work relative value unit
- Ascites (reference: none)
- Preoperative weight loss >10% (reference: none)
- Disseminated cancer (reference: none)
- Age

30-Day Postoperative Return to Operating Room (Secondary Outcome)

List of patient characteristics included in adjusted models:

- ACS NSQIP outcome-specific CPT-based linear predictor for 30-day postoperative return to operating room
- American Society of Anesthesiologists (ASA) Physical Status Classification (normal (reference), mild systemic disease, severe systemic disease, life-threatening systemic disease/moribund)
- Current smoker (reference: not current smoker)
- Preoperative sepsis (none (reference), systemic inflammatory response syndrome (SIRS), sepsis, septic shock)
- Male sex (reference: female)
- Body mass index classification (normal (reference), Class I obese, Class II obese, Class III obese, overweight, underweight)
- Wound class (clean (reference), clean/contaminated, contaminated, dirty/infected)
- Work relative value unit
- Hypertension (reference: none)
- Serum glutamic oxidase transaminase >40 (reference: ≤40)
- Bleeding disorder (reference: none)
- Preoperative albumin
- Patient race (White (reference), Black/African American, American Indian/Native Alaskan, Native Hawaiian/Pacific Islander, Asian, Unknown)
- Chronic obstructive pulmonary disease (reference: none)
- Preoperative dialysis (reference: none)
- Preoperative weight loss >10% (reference: none)
- Steroid use (reference: none)
- Preoperative sodium <135 (reference: ≥135)
- Preoperative prothrombin time>35 (reference: ≤35)

30-Day Postoperative Sepsis (Secondary Outcome)

List of patient characteristics included in adjusted models:

- ACS NSQIP outcome-specific CPT-based linear predictor for 30-day postoperative sepsis
- American Society of Anesthesiologists (ASA) Physical Status Classification (normal (reference), mild systemic disease, severe systemic disease, life-threatening systemic disease/moribund)
- Inpatient surgery setting (reference: outpatient)
- Preoperative albumin
- Male sex (reference: female)
- Preoperative functional status (independent (reference), partially dependent, totally dependent)
- Preoperative white blood cell count >11 (reference: ≤11)
- Wound class (clean (reference), clean/contaminated, contaminated, dirty/infected)
- Chronic obstructive pulmonary disease (reference: none)
- Work relative value unit
- Preoperative alkaline phosphatase>125 (reference: ≤125)
- Steroid use (reference: none)
- Preoperative weight loss >10% (reference: none)
- Hypertension (reference: none)
- Disseminated cancer (reference: none)
- Current smoker (reference: not current smoker)
- Body mass index classification (normal (reference), Class I obese, Class II obese, Class III obese, overweight, underweight)
- Preoperative platelet count >400 (reference: ≤400)
- Preoperative ventilator dependence (reference: none)
- Preoperative creatinine >1.2 (reference: ≤1.2)
- Transfer status (admit from home (reference), transfer from acute care, transfer from chronic care, transfer from other facility, transfer from outside emergency department)
- Ascites (reference: none)
- Preoperative hematocrit <38 (reference: ≥38)
- Bleeding disorder (reference: none)
- Emergency/urgent surgery (reference: not emergency/urgent)
- Patient race (White (reference), Black/African American, American Indian/Native Alaskan, Native Hawaiian/Pacific Islander, Asian, Unknown)
- Age

30-Day Postoperative Surgical Site Infection (Secondary Outcome)

List of patient characteristics included in adjusted models:

- ACS NSQIP outcome-specific CPT-based linear predictor for 30-day postoperative surgical site infection
- Body mass index classification (normal (reference), Class I obese, Class II obese, Class III obese, overweight, underweight)
- Inpatient surgery setting (reference: outpatient)
- American Society of Anesthesiologists (ASA) Physical Status Classification (normal (reference), mild systemic disease, severe systemic disease, life-threatening systemic disease/moribund)
- Patient race (White (reference), Black/African American, American Indian/Native Alaskan, Native Hawaiian/Pacific Islander, Asian, Unknown)
- Current smoker (reference: not current smoker)
- Wound class (clean (reference), clean/contaminated, contaminated, dirty/infected)
- Preoperative sepsis (none (reference), systemic inflammatory response syndrome (SIRS), sepsis, septic shock)
- Work relative value unit
- Steroid use (reference: none)
- Disseminated cancer (reference: none)
- Diabetes (none (reference), insulin, oral)
- Hispanic (no (reference), yes, unknown)
- Preoperative platelet count >400 (reference: ≤400)
- Preoperative alkaline phosphatase>125 (reference: ≤125)
- Male sex (reference: female)
- Preoperative creatinine >1.2 (reference: ≤1.2)
- Preoperative sodium <135 (reference: ≥135)
- Emergency/urgent surgery (reference: not emergency/urgent)
- Preoperative ventilator dependence (reference: none)
- Preoperative renal failure (reference: none)
- Chronic obstructive pulmonary disease (reference: none)
- Preoperative dialysis (reference: none)
- Preoperative weight loss >10% (reference: none)
- Age
- Bleeding disorder (reference: none)
- Preoperative white blood cell count >11 (reference: ≤11)

30-Day Postoperative Urinary Tract Infection (Secondary Outcome)

List of patient characteristics included in adjusted models:

- ACS NSQIP outcome-specific CPT-based linear predictor for 30-day postoperative UTI
- Male sex (reference: female)
- Age
- American Society of Anesthesiologists (ASA) Physical Status Classification (normal (reference), mild systemic disease, severe systemic disease, life-threatening systemic disease/moribund)
- Preoperative functional status (independent (reference), partially dependent, totally dependent)
- Inpatient surgery setting (reference: outpatient)
- Patient race (White (reference), Black/African American, American Indian/Native Alaskan, Native Hawaiian/Pacific Islander, Asian, Unknown)
- Wound class (clean (reference), clean/contaminated, contaminated, dirty/infected)
- Steroid use (reference: none)
- Diabetes (none (reference), insulin, oral)
- Chronic obstructive pulmonary disease (reference: none)
- Preoperative dialysis (reference: none)
- Bleeding disorder (reference: none)
- Emergency/urgent surgery (reference: not emergency/urgent)
- Preoperative sodium <135 (reference: ≥135)
- Preoperative hematocrit <38 (reference: ≥38)
- Preoperative sepsis (none (reference), systemic inflammatory response syndrome (SIRS), sepsis, septic shock)

H. COMPARISON OF ENROLLED VERSUS NON-ENROLLED PROGRAMS AND HOSPITALS

TABLE S31. Characteristics of Programs Enrolled vs. Not Enrolled in the FIRST Trial

	Total	Non-Enrolled	Enrolled	P	Valid N (Per Data Item)		
					Total	Non-Enrolled	Enrolled
Program Type, n (%)					251	134	117
Academic	118 (47.01)	48 (35.82)	70 (59.83)	0.001			
Community	124 (49.40)	79 (58.96)	45 (38.46)				
Military	9 (3.59)	7 (5.22)	2 (1.71)				
Geographic Region, n (%)					251	134	117
Northeast	85 (33.86)	51 (38.06)	34 (29.06)	0.214			
Southeast	49 (19.52)	23 (17.16)	26 (22.22)				
Midwest	58 (23.11)	25 (18.66)	33 (28.21)				
Southwest	25 (9.96)	14 (10.45)	11 (9.40)				
West	34 (13.55)	21 (15.67)	13 (11.11)				
Number of Slots per year, mean (SD)	4.92 (2.11)	4.46 (1.93)	5.46 (2.18)	<0.001	251	134	117
Program Size (5-Year Average Number of QE Examinees), mean (SD)	4.16 (2.15)	3.67 (1.99)	4.71 (2.21)	<0.001	251	134	117
Proportion Residents: Male, mean (SD)	0.70 (0.11)	0.71 (0.12)	0.69 (0.11)	0.150	238	125	113
Proportion Residents: International Medical Graduates, mean (SD)	0.22 (0.21)	0.25 (0.23)	0.17 (0.17)	0.003	238	125	113
QE 1st Attempt Pass Rate (2009-2013), mean (SD)	85.92 (12.31)	83.79 (14.60)	88.30 (8.57)	0.004	239	126	113
CE 1st Attempt Pass Rate (2009-2013), mean (SD)	80.61 (13.94)	78.52 (15.43)	82.91 (11.72)	0.015	238	125	113

QE: American Board of Surgery Qualifying Examination (Written Boards)
CE: American Board of Surgery Certifying Examination (Oral Boards)
Data are program-level aggregate data from the American Board of Surgery.

TABLE S32. Characteristics of Hospitals Enrolled and Not Enrolled in the FIRST Trial

Hospital Characteristic	Valid N per Data Item		NOT IN FIRST	IN FIRST	P-Value
	NOT IN FIRST†	IN FIRST			
Total Bed Size, mean (SD)	1044	148	322.45 (257.93)	577.87 (286.94)	<0.001[a]
Total Surgical Volume in 1000s, mean (SD) †	1044	148	9.84 (9.90)	23.39 (16.09)	<0.001[a]
Nurse-to-Bed Ratio, mean (SD)	1044	148	2.01 (1.18)	2.56 (0.90)	<0.001[a]
Resident-to-Bed Ratio [source: CMS 2014], mean (SD)†	698	141	0.17 (0.22)	0.39 (0.26)	<0.001[a]
CMS Case Mix Index [source: CMS 2014], mean (SD)	698	141	1.64 (0.25)	1.90 (0.24)	<0.001[a]
COTH Membership, frequency (%)	1044	148	209 (20.02%)	106 (71.62%)	<0.001[b]
CBSA Type, frequency (%)	1044	148			0.052[b]
Metro/Division			979 (93.77%)	146 (98.65%)	
Micro			55 (5.27%)	2 (1.35%)	
Rural			10 (0.96%)	0 (0.00%)	
Total Admission Volume in 1000s, mean (SD)	1044	148	13.47 (12.30)	28.73 (14.20)	<0.001[a]
Level 1 Trauma Center, frequency (%)	1044	148	200 (19.16%)	91 (61.49%)	<0.001[b]
Geographic Region, frequency (%)	1043	148			0.789[b]
Midwest			266 (25.50%)	36 (24.32%)	
West			168 (16.11%)	27 (18.24%)	
South			202 (19.37%)	23 (15.54%)	
Northeast			238 (22.82%)	36 (24.32%)	
Southeast			169 (16.20%)	26 (17.57%)	
Performs Any Transplants	1044	148	240 (22.99%)	106 (71.62%)	<0.001[b]

N Hospitals NOT IN FIRST = 1044 hospitals. N Hospitals IN FIRST = 148 hospitals.

Unless otherwise noted, data are from American Hospital Association Fiscal Year 2013 Annual Survey of Hospitals. CMS 2014 data are from Centers for Medicare and Medicaid Services 2014 Payment Update Impact File. **For comparability, we compared FIRST Trial hospitals to those hospitals that reported any residency training approval by ACGME to AHA. †IMPORTANT NOTE: Hospitals in comparison group may or may not be affiliated with a General Surgery residency program.** There is no reliable roster of hospitals in the U.S. affiliated with specific ACGME residency programs. Hospitals in Guam, Puerto Rico, Virgin Islands (US), American Samoa, and Mariana Islands not included. [a] Two-tailed t-test [b] Chi-square test of association

I. NUMBER OF VALID OBSERVATIONS FOR PROGRAM, HOSPITAL, AND PATIENT DATA

TABLE S33A. Number of Valid Observations for Program, Hospital, and Patient Outcomes Data

CHARACTERISTICS	NON-MISSING N			N MISSING		
	ALL PROGRAMS	STANDARD POLICY	FLEXIBLE POLICY	ALL PROGRAMS	STANDARD POLICY	FLEXIBLE POLICY
Residency Program Characteristics						
Program Type	117 (100%)	59 (100%)	58 (100%)	0 (0%)	0 (0%)	0 (0%)
Geographic Region	117 (100%)	59 (100%)	58 (100%)	0 (0%)	0 (0%)	0 (0%)
Number of Residents per Program	117 (100%)	59 (100%)	58 (100%)	0 (0%)	0 (0%)	0 (0%)
5-Year Average Number of ABS QE Examinees	117 (100%)	59 (100%)	58 (100%)	0 (0%)	0 (0%)	0 (0%)
Proportion Male Residents	113 (97%)	58 (98%)	55 (95%)	4 (3%)	1 (2%)	3 (5%)
Proportion International Medical Graduate	113 (97%)	58 (98%)	55 (95%)	4 (3%)	1 (2%)	3 (5%)
Qualifying Exam % Pass 1st Attempt (2009-2013)	113 (97%)	58 (98%)	55 (95%)	4 (3%)	1 (2%)	3 (5%)
Certifying Exam % Pass 1st Attempt (2009-2014)	113 (97%)	58 (98%)	55 (95%)	4 (3%)	1 (2%)	3 (5%)
Hospital Characteristics						
Total Bed Size	148 (100%)	70 (100%)	78 (100%)	0 (0%)	0 (0%)	0 (0%)
Total Surgical Volume	148 (100%)	70 (100%)	78 (100%)	0 (0%)	0 (0%)	0 (0%)
Nurse-to-Bed Ratio	148 (100%)	70 (100%)	78 (100%)	0 (0%)	0 (0%)	0 (0%)
Resident-to-Bed Ratio	141 (95%)	66 (94%)	75 (96%)	7 (5%)	4 (6%)	3 (4%)
CMS Case Mix Index	141 (95%)	66 (94%)	75 (96%)	7 (5%)	4 (6%)	3 (4%)
COTH Membership	148 (100%)	70 (100%)	78 (100%)	0 (0%)	0 (0%)	0 (0%)
ANCC Nursing Magnet Status	148 (100%)	70 (100%)	78 (100%)	0 (0%)	0 (0%)	0 (0%)
Prior Year (2013) Rate of 30-Day Postop. DSM	143 (97%)	69 (99%)	74 (95%)	5 (3%)	1 (1%)	4 (5%)
Patient Outcomes						
30-Day Postoperative Death/Serious Morbidity	138691 (100%)	65849 (100%)	72842 (100%)	0 (0%)	0 (0%)	0 (0%)
30-Day Postoperative Death	138691 (100%)	65849 (100%)	72842 (100%)	0 (0%)	0 (0%)	0 (0%)
30-Day Postoperative Serious Morbidity	138691 (100%)	65849 (100%)	72842 (100%)	0 (0%)	0 (0%)	0 (0%)
30-Day Postoperative Any Morbidity	138691 (100%)	65849 (100%)	72842 (100%)	0 (0%)	0 (0%)	0 (0%)
30-Day Postoperative Failure-to-Rescue	11937 (100%)	5649 (100%)	6288 (100%)	0 (0%)	0 (0%)	0 (0%)
30-Day Postoperative Pneumonia	137375 (100%)	65719 (100%)	72656 (100%)	0 (0%)	0 (0%)	0 (0%)
30-Day Postoperative Renal Failure	138596 (100%)	65805 (100%)	72791 (100%)	0 (0%)	0 (0%)	0 (0%)
30-Day Postoperative Unplanned Reoperation	138691 (100%)	65849 (100%)	72842 (100%)	0 (0%)	0 (0%)	0 (0%)
30-Day Postoperative Sepsis	135258 (100%)	64237 (100%)	71021 (100%)	0 (0%)	0 (0%)	0 (0%)
30-Day Postoperative SSI	137346 (100%)	65180 (100%)	72166 (100%)	0 (0%)	0 (0%)	0 (0%)
30-Day Postoperative Urinary Tract Infection	138691 (100%)	65849 (100%)	72842 (100%)	0 (0%)	0 (0%)	0 (0%)

No significant differences in rates of missing values across study arms in program, hospital or patient characteristics. Differences in patient outcome Ns are not due to missing data, but differences in the denominator (some outcomes such as pneumonia, sepsis patients based on conditions present upon admission). All patient outcomes are required by ACS NSQIP in order to complete case abstraction, so there are no missing data.

TABLE S33B. Number of Valid Observations for Imputed Patient Characteristics

IMPUTED VARIABLES	NON-MISSING N	MISSING N	PERCENT MISSING
Surgical Specialty	138,691	0	0%
Age	138,691	0	0%
Sex	138,691	0	0%
Body mass index (BMI)	136,389	2,302	1.7%
Diabetes	138,691	0	0%
Dyspnea	138,691	0	0%
Preoperative ventilator dependence	138,691	0	0%
History of COPD	138,691	0	0%
History of CHF	138,691	0	0%
Ascites	138,691	0	0%
Renal failure	138,691	0	0%
Dialysis dependent	138,691	0	0%
Disseminated cancer	138,691	0	0%
Preoperative wound infection	138,691	0	0%
Steroid use	138,691	0	0%
Weight loss	138,691	0	0%
Bleeding disorder	138,691	0	0%
Preoperative functional status	138,317	374	0.3%
Emergency surgery	138,691	0	0%
Smoking	138,691	0	0%
Wound class	138,691	0	0%
ASA Class	138,552	139	0.1%
Preoperative sepsis	138,691	0	0%
Hypertension requiring medication	138,691	0	0%
Preoperative transfusion	138,691	0	0%
Elective surgery	138,608	83	0.1%
Serum sodium	115,530	23,161	16.7%
BUN	113,865	24,826	17.9%
Creatinine	116,223	22,468	16.2%
Albumin	86,682	52,009	37.5%
Bilirubin	87,375	51,316	37.0%
SGOT	87,791	50,900	36.7%
Alkaline phosphatase	87,653	51,038	36.8%
WBC	115,946	22,745	16.4%
Hematocrit	117,610	21,081	15.2%
Platelets	116,223	22,468	16.2%
PTT	38,001	100,690	72.6%
Operative time	138,649	42	0.0%

Note: ACS NSQIP requires most variables to be completed in order to finalize case abstraction and transmit the case from the hospital to ACS NSQIP, except laboratory values. Thus, missing data are infrequent. If missing, ACS NSQIP imputes the 38 variables above using Buck's method.

J. COMPARISON OF POPULATION AVERAGED AND CONDITIONAL ESTIMATES FOR PATIENT AND RESIDENT OUTCOMES

TABLE S34. Intent-to-Treat (ITT) Estimates of the Association between Study Arm Assignment and Patient Outcomes: Population-Averaged Estimates and Conditional Estimates

PATIENT OUTCOME	Non-Inferiority Threshold (Δ)	ODDS RATIO (92% CONFIDENCE INTERVAL) P-VALUE			
		UNADJUSTED ANALYSES		ADJUSTED ANALYSES	
		MODEL 1 Logistic Regression POPULATION AVERAGED ESTIMATE	MODEL 2 Hierarchical Logistic Regression CONDITIONAL ESTIMATE	MODEL 3 Logistic Regression POPULATION AVERAGED ESTIMATE	MODEL 4 Hierarchical Logistic Regression CONDITIONAL ESTIMATE
30-Day Postoperative Death/Serious Morbidity ***Primary Outcome*	1.15	0.96 (0.88-1.06) 0.489	0.96 (0.87-1.06) 0.443	0.98 (0.90-1.05) 0.571	0.96 (0.90-1.04) 0.378
30-Day Postoperative Death	1.14	0.97 (0.85-1.10) 0.661	1.00 (0.86-1.16) 0.993	0.94 (0.82-1.07) 0.393	0.95 (0.82-1.10) 0.558
30-Day Postoperative Serious Morbidity	1.15	0.97 (0.88-1.06) 0.516	0.96 (0.86-1.06) 0.449	0.97 (0.90-1.05) 0.566	0.96 (0.90-1.04) 0.399
30-Day Postoperative Morbidity	1.16	0.97 (0.88-1.08) 0.617	0.94 (0.84-1.06) 0.392	0.98 (0.91-1.06) 0.640	0.96 (0.89-1.04) 0.388
30-Day Postoperative Failure to Rescue	1.15	1.01 (0.86-1.18) 0.940	1.03 (0.87-1.23) 0.730	0.98 (0.84-1.14) 0.818	1.00 (0.86-1.18) 0.966
30-Day Postoperative Pneumonia	1.14	1.01 (0.84-1.22) 0.933	0.95 (0.78-1.14) 0.603	0.96 (0.89-1.04) 0.401	0.96 (0.81-1.14) 0.669
30-Day Postoperative Renal Failure	1.14	1.04 (0.87-1.24) 0.703	1.05 (0.86-1.28) 0.659	1.05 (0.90-1.23) 0.565	1.07 (0.91-1.27) 0.466
30-Day Postoperative Unplanned Reoperation	1.14	0.90 (0.81-1.01) 0.105	0.91 (0.81-1.03) 0.173	0.93 (0.84-1.03) 0.227	0.93 (0.84-1.04) 0.249
30-Day Postoperative Sepsis	1.14	0.90 (0.74-1.10) 0.370	0.90 (0.73-1.10) 0.363	0.90 (0.78-1.04) 0.199	0.89 (0.76-1.03) 0.166
30-Day Postoperative SSI	1.15	0.99 (0.87-1.13) 0.911	0.93 (0.81-1.08) 0.396	0.99 (0.90-1.09) 0.822	0.94 (0.86-1.04) 0.317
30-Day Postoperative UTI	1.14	0.94 (0.79-1.11) 0.496	0.91 (0.76-1.08) 0.324	0.94 (0.79-1.11) 0.520	0.90 (0.76-1.06) 0.254

N patients varies across outcomes. N residency programs = 115 programs. Unadjusted models (Models 1&2) regress outcomes on study arm assignment. Adjusted models (Models 3&4) regress outcomes on study arm assignment and patient characteristics (see Section G for list of covariates). All models adjusted for program-level rates of 2013 30-day postoperative death/serious morbidity (stratifying variable used in randomization). Models 1&3 are population-averaged models estimated using logistic regression with program-level clustered standard errors. Models 2&4 are conditional estimates obtained from 3-level hierarchical logistic regression models that included program-level random intercepts.

TABLE S35. Intent-to-Treat (ITT) Estimates of the Association between Study Arm Assignment and Resident Outcomes: Population-Averaged Estimates and Conditional Estimates

RESIDENT OUTCOME	ODDS RATIO (95% CONFIDENCE INTERVAL) P-VALUE			
	UNADJUSTED ANALYSES		ADJUSTED ANALYSES	
	MODEL 1 Logistic Regression POPULATION AVERAGED ESTIMATE	MODEL 2 Hierarchical Logistic Regression CONDITIONAL ESTIMATE	MODEL 3 Logistic Regression POPULATION AVERAGED ESTIMATE	MODEL 4 Hierarchical Logistic Regression CONDITIONAL ESTIMATE
PRIMARY OUTCOMES				
(Very) dissatisfied with education quality	1.06 (0.76-1.48) 0.722	1.08 (0.77-1.52) 0.640	0.96 (0.69-1.33) 0.819	1.00 (0.72-1.41) 0.988
(Very) dissatisfied with personal wellbeing	1.28 (0.96-1.71) 0.093	1.31 (0.99-1.74) 0.062	1.14 (0.89-1.46) 0.297	1.20 (0.91-1.59) 0.199
SECONDARY OUTCOMES				
Perceived negative effect of duty hours on: patient safety	0.40 (0.32-0.50) <0.001	0.40 (0.32-0.51) <0.001	0.39 (0.31-0.49) <0.001	0.39 (0.31-0.50) <0.001
Perceived negative effect of duty hours on: continuity of care	0.18 (0.14-0.24) <0.001	0.16 (0.12-0.21) <0.001	0.17 (0.13-0.21) <0.001	0.15 (0.11-0.20) <0.001
Perceived negative effect of duty hours on: conference attendance	0.47 (0.37-0.61) <0.001	0.47 (0.36-0.62) <0.001	0.45 (90.35-0.58) <0.001	0.43 (0.33-0.58) <0.001
Perceived negative effect of duty hours on: clinical skills acquisition	0.26 (0.21-0.33) <0.001	0.24 (0.19-0.31) <0.001	0.24 (0.19-0.29) <0.001	0.22 (0.18-0.28) <0.001
Perceived negative effect of duty hours on: operative skills acquisition	0.24 (0.19-0.30) <0.001	0.22 (0.17-0.27) <0.001	0.22 (0.18-0.28) <0.001	0.21 (0.16-0.26) <0.001
Perceived negative effect of duty hours on: resident autonomy	0.28 (0.21-0.36) <0.001	0.26 (0.20-0.34) <0.001	0.25 (0.19-0.32) <0.001	0.24 (0.18-0.31) <0.001
Perceived negative effect of duty hours on: operative volume	0.24 (0.18-0.30) <0.001	0.22 (0.17-0.28) <0.001	0.22 (0.18-0.28) <0.001	0.21 (0.16-0.26) <0.001

N residents varies across outcomes. N residency programs = 117 programs. Dissatisfaction outcomes coded: 1=Very Dissatisfied/Dissatisfied, 0=Neutral, Satisfied or Very Satisfied. Perceived effect outcomes coded: 1=Perceived Negative Effect, 0=Perceived Positive Effect or No Effect. Disagreement outcomes coded: 1=Strongly Disagree/Disagree, 0=Neutral, Agree, Strongly Agree. Fatigue outcomes coded: 1=Always/Often, 0=Sometimes, Rarely, Never. Unadjusted models (Models 1&2) regress outcomes on study arm assignment. Adjusted models (Models 3&4) regress outcomes on study arm assignment, resident gender, postgraduate year, program type and geographic region (geographic region not included in Disagreement or Fatigue models). All models adjusted for program-level rates of 2013 30-day postoperative death/serious morbidity (stratifying variable used in randomization). Models 1&3 are population-averaged models estimated using logistic regression with program-level clustered standard errors. Models 2&4 are conditional estimates obtained from 2-level hierarchical logistic regression models that included program-level random intercepts.

TABLE S35 (continued). Intent-to-Treat (ITT) Estimates of the Association between Study Arm Assignment and Resident Outcomes: Population-Averaged Estimates and Conditional Estimates

RESIDENT OUTCOME	ODDS RATIO (95% CONFIDENCE INTERVAL) P-VALUE			
	UNADJUSTED ANALYSES		ADJUSTED ANALYSES	
	MODEL 1 Logistic Regression POPULATION AVERAGED ESTIMATE	MODEL 2 Hierarchical Logistic Regression CONDITIONAL ESTIMATE	MODEL 3 Logistic Regression POPULATION AVERAGED ESTIMATE	MODEL 4 Hierarchical Logistic Regression CONDITIONAL ESTIMATE
SECONDARY OUTCOMES (continued)				
Perceived negative effect of duty hours on: availability for elective cases	0.32 (0.26-0.41) <0.001	0.30 (0.24-0.39) <0.001	0.30 (0.24-0.39) <0.001	0.29 (0.23-0.37) <0.001
Perceived negative effect of duty hours on: availability for urgent cases	0.21 (0.17-0.27) <0.001	0.20 (0.16-0.25) <0.001	0.21 (0.16-0.26) <0.001	0.19 (0.15-0.25) <0.001
Perceived negative effect of duty hours on: time for teaching medical students	0.45 (0.36-0.56) <0.001	0.45 (0.37-0.56) <0.001	0.41 (0.33-0.49) <0.001	0.41 (0.33-0.50) <0.001
Perceived negative effect of duty hours on: relationship between interns/residents	0.36 (0.28-0.46) <0.001	0.38 (0.29-0.49) <0.001	0.34 (0.27-0.44) <0.001	0.36 (0.27-0.47) <0.001
Perceived negative effect of duty hours on: professionalism	0.63 (0.48-0.83) 0.001	0.65 (0.49-0.87) 0.003	0.60 (0.44-0.81) 0.001	0.61 (0.45-0.84) 0.002
Perceived negative effect of duty hours on: morale	1.05 (0.83-1.33) 0.686	1.09 (0.85-1.40) 0.513	1.00 (0.79-1.27) 0.993	1.05 (0.81-1.36) 0.727
Perceived negative effect of duty hours on: ability to prepare for cases	3.13 (2.41-4.06) <0.001	3.37 (2.54-4.47) <0.001	3.36 (2.55-4.43) <0.001	3.60 (2.68-4.84) <0.001
Perceived negative effect of duty hours on: ability to participate in research	2.66 (2.06-3.44) <0.001	2.81 (2.12-3.73) <0.001	2.70 (2.09-3.50) <0.001	2.87 (2.14-3.85) <0.001
Perceived negative effect of duty hours on: job satisfaction	0.91 (0.71-1.16) 0.448	0.94 (0.73-1.23) 0.666	0.95 (0.74-1.21) 0.675	0.98 (0.75-1.29) 0.898
Perceived negative effect of duty hours on: satisfaction with decision to become a surgeon	1.02 (0.79-1.30) 0.905	1.03 (0.79-1.33) 0.843	1.00 (0.77-1.30) 0.989	1.01 (0.77-1.32) 0.955

N residents varies across outcomes. N residency programs = 117 programs. Dissatisfaction outcomes coded: 1=Very Dissatisfied/Dissatisfied, 0=Neutral, Satisfied or Very Satisfied. Perceived effect outcomes coded: 1=Perceived Negative Effect, 0=Perceived Positive Effect or No Effect. Disagreement outcomes coded: 1=Strongly Disagree/Disagree, 0=Neutral, Agree, Strongly Agree. Fatigue outcomes coded: 1=Always/Often, 0=Sometimes, Rarely, Never. Unadjusted models (Models 1&2) regress outcomes on study arm assignment. Adjusted models (Models 3&4) regress outcomes on study arm assignment, resident gender, postgraduate year, program type and geographic region (geographic region not included in Disagreement or Fatigue models). All models adjusted for program-level rates of 2013 30-day postoperative death/serious morbidity (stratifying variable used in randomization). Models 1&3 are population-averaged models estimated using logistic regression with program-level clustered standard errors. Models 2&4 are conditional estimates obtained from 2-level hierarchical logistic regression models that included program-level random intercepts.

TABLE S35 (continued). Intent-to-Treat (ITT) Estimates of the Association between Study Arm Assignment and Resident Outcomes: Population-Averaged Estimates and Conditional Estimates

RESIDENT OUTCOME	ODDS RATIO (95% CONFIDENCE INTERVAL) P-VALUE			
	UNADJUSTED ANALYSES		ADJUSTED ANALYSES	
	MODEL 1 Logistic Regression POPULATION AVERAGED ESTIMATE	MODEL 2 Hierarchical Logistic Regression CONDITIONAL ESTIMATE	MODEL 3 Logistic Regression POPULATION AVERAGED ESTIMATE	MODEL 4 Hierarchical Logistic Regression CONDITIONAL ESTIMATE
SECONDARY OUTCOMES (continued)				
Perceived negative effect of duty hours on: time with family and friends	3.39 (2.56-4.49) <0.001	3.66 (2.70-4.97) <0.001	3.46 (2.60-4.62) <0.001	3.79 (2.75-5.23) <0.001
Perceived negative effect of duty hours on: time for hobbies/extracurricular activities	3.48 (2.65-4.59) <0.001	3.81 (2.84-5.11) <0.001	3.53 (2.68-4.66) <0.001	3.87 (2.85-5.25) <0.001
Perceived negative effect of duty hours on: own health	3.07 (2.30-4.09) <0.001	3.22 (2.37-4.36) <0.001	3.00 (2.26-3.97) <0.001	3.20 (2.33-4.38) <0.001
Perceived negative effect of duty hours on: rest	3.49 (2.71-4.50) <0.001	3.85 (2.88-5.15) <0.001	3.62 (2.80-4.70) <0.001	4.04 (2.99-5.47) <0.001
Perceived negative effect of duty hours on: wellbeing	2.12 (1.59-2.83) <0.001	2.26 (1.64-3.11) <0.001	2.07 (1.56-2.76) <0.001	2.23 (1.61-3.08) <0.001
(Very) dissatisfied: continuity of care	0.44 (0.32-0.60) <0.001	0.44 (0.32-0.60) <0.001	0.42 (0.31-0.57) <0.001	0.41 (0.29-0.58) <0.001
(Very) dissatisfied: patient safety	0.85 (0.56-1.29) 0.442	0.85 (0.55-1.31) 0.458	0.74 (0.49-1.12) 0.156	0.74 (0.47-1.16) 0.192
(Very) dissatisfied: work hours/scheduling	0.96 (0.72-1.28) 0.772	0.95 (0.71-1.27) 0.725	0.88 (0.67-1.16) 0.361	0.89 (0.66-1.20) 0.453
(Very) dissatisfied: quality/ease of handoffs/care transitions	0.68 (0.52-0.90) 0.008	0.70 (0.52-0.92) 0.011	0.64 (0.49-0.83) 0.001	0.64 (0.48-0.85) 0.002
(Very) dissatisfied: time for rest	1.32 (1.00-1.76) 0.052	1.41 (1.06-1.89) 0.020	1.20 (0.94-1.54) 0.146	1.32 (0.98-1.76) 0.066

N residents varies across outcomes. N residency programs = 117 programs. Dissatisfaction outcomes coded: 1=Very Dissatisfied/Dissatisfied, 0=Neutral, Satisfied or Very Satisfied. Perceived effect outcomes coded: 1=Perceived Negative Effect, 0=Perceived Positive Effect or No Effect. Disagreement outcomes coded: 1=Strongly Disagree/Disagree, 0=Neutral, Agree, Strongly Agree. Fatigue outcomes coded: 1=Always/Often, 0=Sometimes, Rarely, Never. Unadjusted models (Models 1&2) regress outcomes on study arm assignment. Adjusted models (Models 3&4) regress outcomes on study arm assignment, resident gender, postgraduate year, program type and geographic region (geographic region not included in Disagreement or Fatigue models). All models adjusted for program-level rates of 2013 30-day postoperative death/serious morbidity (stratifying variable used in randomization). Models 1&3 are population-averaged models estimated using logistic regression with program-level clustered standard errors. Models 2&4 are conditional estimates obtained from 2-level hierarchical logistic regression models that included program-level random intercepts.

TABLE S35 (continued). Intent-to-Treat (ITT) Estimates of the Association between Study Arm Assignment and Resident Outcomes: Population-Averaged Estimates and Conditional Estimates

RESIDENT OUTCOME	ODDS RATIO (95% CONFIDENCE INTERVAL) P-VALUE			
	UNADJUSTED ANALYSES		ADJUSTED ANALYSES	
	MODEL 1 Logistic Regression POPULATION AVERAGED ESTIMATE	MODEL 2 Hierarchical Logistic Regression CONDITIONAL ESTIMATE	MODEL 3 Logistic Regression POPULATION AVERAGED ESTIMATE	MODEL 4 Hierarchical Logistic Regression CONDITIONAL ESTIMATE
SECONDARY OUTCOMES (continued)				
(Very) dissatisfied: work hour regulations	0.95 (0.68-1.32) 0.748	0.99 (0.71-1.40) 0.972	0.94 (0.68-1.29) 0.692	0.98 (0.68-1.41) 0.912
Always/often: fatigue affects personal safety	1.14 (0.90-1.44) 0.270	1.15 (0.91-1.47) 0.247	1.15 (0.91-1.46) 0.228	1.17 (0.92-1.48) 0.212
Always often: fatigue affects patient safety	1.18 (0.91-1.53) 0.207	1.18 (0.91-1.53) 0.208	1.20 (0.93-1.54) 0.164	1.20 (0.92-1.55) 0.173
At least once in recent typical month: handed off active patient care issue due to duty hour limits	0.54 (0.46-0.64) <0.001	0.53 (0.45-0.63) <0.001	0.53 (0.44-0.63) <0.001	0.51 (0.43-0.62) <0.001
At least once in recent typical month: left during an operation due to duty hour limits	0.49 (0.35-0.68) <0.001	0.46 (0.32-0.65) <0.001	0.51 (0.35-0.74) <0.001	0.47 (0.32-0.68) <0.001
At least once in recent typical month: missed an operation due to duty hour limits	0.58 (0.47-0.71) <0.001	0.56 (0.45-0.69) <0.001	0.56 (0.45-0.69) <0.001	0.53 (0.42-0.67) <0.001

N residents varies across outcomes. N residency programs = 117 programs. Dissatisfaction outcomes coded: 1=Very Dissatisfied/Dissatisfied, 0=Neutral, Satisfied or Very Satisfied. Perceived effect outcomes coded: 1=Perceived Negative Effect, 0=Perceived Positive Effect or No Effect. Disagreement outcomes coded: 1=Strongly Disagree/Disagree, 0=Neutral, Agree, Strongly Agree. Fatigue outcomes coded: 1=Always/Often, 0=Sometimes, Rarely, Never. Unadjusted models (Models 1&2) regress outcomes on study arm assignment. Adjusted models (Models 3&4) regress outcomes on study arm assignment, resident gender, postgraduate year, program type and geographic region (geographic region not included in Disagreement or Fatigue models). All models adjusted for program-level rates of 2013 30-day postoperative death/serious morbidity (stratifying variable used in randomization). Models 1&3 are population-averaged models estimated using logistic regression with program-level clustered standard errors. Models 2&4 are conditional estimates obtained from 2-level hierarchical logistic regression models that included program-level random intercepts.

K. INSTITUTIONAL REVIEW BOARD OFFICE DETERMINATION

Institutional Review Board Office
Northwestern University

Biomedical IRB
750 North Lake Shore Drive
Suite 700
Chicago, Illinois 60611
312-503-9338

Social and Behavioral Sciences IRB
600 Foster Street
Chambers Hall, Second Floor
Evanston, Illinois 60208
847-467-1723

NORTHWESTERN UNIVERSITY

Form for Determining Whether a Project Involves Human Subjects Research

Version: 3.0
Date: 5/10/2013

The Northwestern University IRB is required to review and approve all research involving human subjects. This application is intended to help you determine if your project requires IRB approval. If you require written documentation from the IRB Office, complete the entire form, and email the signed form and any relevant supporting documents (i.e., grant, protocol, consent forms) to irb@northwestern.edu. You should receive an IRB response within 10 business days.

Current Status of the Project

Has the project already been conducted (i.e., data has already been collected and analyzed)?
☐ Yes ☒ No

SECTION I: Activities Determined by the NU IRB Office Not to Represent Human Subjects Research

A. ☐ **Case Report:** The project consists of a case report or series which describes an interesting treatment, presentation or outcome. A critical component is that nothing was done to the patient(s) with prior "research" intent.

NOTE: For case reports, HIPAA requires that the disclosure of an individual's protected health information must be authorized by that individual. If a case report contains any of the 18 protected health information identifiers as defined by the HIPAA regulations, a signed authorization (using the authorization form from the entity that holds the record) to disclose this information must be obtained from the individual(s) whose information is being disclosed.

B. ☐ **Course-Related Activities:** The project is limited to course-related activities designed specifically for educational or teaching purposes where data are collected from and about students as part of a routine class exercise or assignment and is not intended for use outside of the classroom.
NOTE: IRB approval **is** required if a student is involved in an activity designed to teach research methodologies and the instructor or student wishes to conduct further investigation and analyses in order to contribute to scholarly knowledge.

C. ☐ **Decedents:** The project involves research that is limited to death records, autopsy materials, or cadaver specimens. If the project involves the use and/or collection of Protected Health Information (PHI), HIPAA regulations apply to decedent research. As the Privacy Board, the IRB Office requires that you confirm the following conditions as set forth in the Privacy Rule at 45 CFR 164.512(i)(ii)(iii), have been met.

i. ☐ the use will be solely for research on the information of a decedent; and
ii. ☐ the Principal Investigator has documentation of the death of the individual about whom information is being sought, and
iii. ☐ the information sought is for the purposes of the research

Note, however, that this exception may not be available for decedent Information that contains Psychotherapy Notes or Information relating to HIV, mental health, genetic testing, or drug or alcohol abuse

D. ☐ **Journalism/Documentary Activities:** The activities are limited to investigations and interviews that focus on specific events, views, etc., and that lead to publication in any medium (including electronic), documentary production, or are part of training that is explicitly linked to journalism. There is no intent to test a hypothesis.

NOTE: IRB approval may be required when journalists conduct activities normally considered scientific research intended to produce generalizable knowledge (e.g., systematic research, surveys, and/or interviews that are intended to test theories or develop models).

E. ☐ **Oral History:** The project is limited to oral history activities, such as open ended interviews, that only document a specific historical event or the experiences of individuals without the intent to draw conclusions or generalize findings.
NOTE: IRB approval is required when the oral history activities are intended to produce generalizable conclusions (e.g., that serve as data collection intended to test economic, sociological, or anthropological models/theories).

F. ☐ **Program evaluation /Quality Improvement/Quality Assurance Activities:** The project is limited to program evaluation, quality improvement or quality assurance activities designed specifically to assess or improve performance within the department, hospital or classroom setting. The intention of the project is not to generate conclusions that can be applied universally, outside of the immediate environment where the project occurred.
Note: Investigators who plan to conduct a QI/QA project, should ensure that they have received approval from any applicable committees within their department or the site in which the activity will occur.

G. ☐ **Public Use Datasets:** The project is limited to analyzing de-identified data contained within a publicly available dataset. Below are examples of data sources that qualify as not-human subjects research (unless the researcher has received the restricted use data):

☐ Data files downloaded from the ICPSR (Interuniversity Consortium for Political and Social Research): http://www.icpsr.umich.edu/icpsrweb/ICPSR/ or the Roper Center for Public Opinion Research http://www.ropercenter.uconn.edu.

☐ American National Election Studies, (ANES) 1948-2010 http://www.electionstudies.org/

☐ Bureau of Economic Analysis: http://www.bea.gov/

☐ Bureau of Labor Statistics (BLS): http://www.bls.gov/

☐ Center for Disease Control (CDC): http://www.cdc.gov/

☐ Consumer expenditure Survey: http://www.bls.gov/cex/

☐ Current Population Survey: http://www.bls.gov/cps/

☐ FBI Uniform Crime Reporting Program: http://www.fbi.gov/about-us/cjis/ucr/ucr or National Archive of Criminal justice data: http://www.icpsr.umich.edu/icpsrweb/NACJD/index.jsp

☐ General Social Survey: http://www3.norc.org/GSS+Website/

☐ National Center for Education Statistics (NCES): http://nces.ed.gov/

☐ National Longitudinal Surveys (NLS): http://www.bls.gov/nls/

☐ Survey of Income and Program Participation: http://www.census.gov/sipp/

☐ Government sites that bring data files together: Data.gov (http://www.data.gov/); FedStats (http://www.fedstats.gov/); and USA.gov (http://www.usa.gov/Topics/Reference_Shelf/Data.shtml)

NOTE: IRB review is required if the publicly available data set contains identifiers, or if the merging of multiple data sets might result in identification of the subjects. In both cases, Exempt Category #4 may apply.

H. ☒ **Coded* Private Information and/or Human Biological Specimens:** The project is limited to the use of existing and/or prospectively collected coded private information and/or human biological specimens (hereafter referred to as "specimens"). IRB Approval is not required if all of the following conditions apply to the project:

i. ☒ (1) The private information or specimens were/are not collected specifically for the currently proposed research project through an interaction or intervention with living individuals; **and**

☒ (2) The investigator(s)** cannot readily ascertain the identity of the individual(s) to whom the coded private information or specimens pertain because, for example:

☒ (a) the investigators and the holder of the key enter into an agreement prohibiting the release of the key to the investigators under any circumstances, until the individuals are deceased (note that the HHS regulations do not require the IRB to review and approve this agreement);

☐ (b) there are IRB-approved written policies and operating procedures for a repository or data management center that prohibit the release of the key to the investigators under any circumstances, until the individuals are deceased; or

☐ (c) there are other legal requirements prohibiting the release of the key to the investigators, until the individuals are deceased, **and**

ii. ☒ Specimens are <u>not</u> being used to test the effectiveness of a medical device or as a control in an investigation of an investigational device and the results of the activity are to be submitted to the FDA or held for inspection by the FDA, **and**

iii. ☒ The records/images/charts that are being collected for this study are <u>not</u> from individuals who are or will become recipients of an FDA regulated product (approved or experimental) or act as a control as directed by a research protocol and not by medical practice, and the results are to be submitted to the FDA or held for inspection by the FDA.

From the Office for Human Research Protections (OHRP) guidance document dated October 16, 2008:
**Coded* means that: (1) identifying information (such as name or social security number) that would enable the investigator to readily ascertain the identity of the individual to whom the private information or specimens pertain has been replaced with a number, letter, symbol, or combination thereof (i.e., the code); and (2) a key to decipher the code exists, enabling linkage of the identifying information to the private information or specimens.

***Investigator* includes anyone involved in conducting the research. The act of solely providing coded private information or specimens (for example, by a tissue repository) does not constitute involvement in the conduct of the research. If the individuals who provide coded information or specimens collaborate on other activities related to the conduct of this research with the investigators who receive such information or specimens, then the IRB would consider such additional activities to constitute involvement in the conduct of the research. Examples of such additional activities include, but are not limited to: (1) the study, interpretation, or analysis of the data resulting from the coded information or specimens; and (2) authorship of presentations or manuscripts related to the research.

l. ☐ **De-Identified Private Information or Human Biological Specimens:** The project is limited to the use of existing and/or prospectively collected de-identified private information and/or human biological specimens (hereafter referred to as "specimens"). IRB Approval is not required if you can confirm the following:

i. ☐ The private information or specimens were/are not collected specifically for the currently proposed research project through an interaction or intervention with living individuals; **and**

ii. ☐ The investigator can confirm that the use of the private information or specimens is not in violation of the terms of use under which the information or specimens were/will be collected; **and**

iii. ☐ The investigator will only receive information or specimens that are fully de-identified. De-identified means that the materials to be studied are devoid of any of the 18 Protected Health Information elements set forth in the Privacy Rule, as well as any codes that would enable linkage of the information or specimens to individual identifiers. Note: To be considered de-identified, nobody, including individuals who are not involved in the conduct of the project, should be able to link the information or specimens back to

identifiers. **and**

iv. ☐ Specimens are not being used to test the effectiveness of a medical device or as a control in an investigation of an investigational device and the results of the activity are to be submitted to the FDA or held for inspection by the FDA, **and**

v. ☐ The records/images/charts that are being collected for this study are not from individuals who are or will become recipients of an FDA regulated product (approved or experimental) or act as a control as directed by a research protocol and not by medical practice, and the results are to be submitted to the FDA or held for inspection by the FDA.

Instructions: If your activity did not fall into the categories described in Section I, continue to Section II and III to assess if you are engaged in human subjects research per the regulations set forth by the Department of Health and Human Services (HHS) and the Food and Drug Administration (FDA).

SECTION II. Activities subject to HHS human subject research regulations (45 CFR 46)

1. **Is the activity RESEARCH: a systematic investigation designed to contribute to generalizable knowledge?**

   TIP: If the investigation characterized by order, planning, and methodology and the intention of the investigation is to generate conclusions that can be applied universally, outside of the immediate environment where the investigation occurred (i.e., the classroom, hospital, department), then the activity meets the definition of research.

   ☐ Yes, Go to #2 ☐ No, Go to FDA section III

2. **Does the research involve obtaining information about LIVING individuals?**
   ☐ Yes, Go to #3 ☐ No, Go to FDA section III

3. **Does the research involve collecting data through intervention (i.e., physical procedures or manipulation of the environment) or interaction (i.e., communication or interpersonal contact between investigator and person) with the individuals?**
   ☐ Yes, IRB review required. ☐ No, Go to #4
   Go to FDA section III to assess if FDA regulations apply to your study.

4. **Does the research involve collecting identifiable information (i.e., the identity of the subject is or may readily be ascertained by the investigator or associated with the information)?**
   ☐ Yes, Go to #5 ☐ No, Go to FDA section III

5. **Is the information private? (About behavior that occurs in a context in which an individual can reasonably expect that no observation or recording is taking place, or provided for specific purposes by an individual and which the individual can reasonably expect will not be made public)**
   ☐ Yes, IRB review required ☐ No, Go to FDA section III
   Go to FDA section III to assess if FDA regulations apply to your study.

SECTION III. Activities subject to FDA human subject regulations: If your answer is "yes" to any of the 3 questions below, IRB approval is required and the FDA regulations apply to your study.

1. Is this is an experiment that involves a test article * and one or more human subjects, and the results of which are intended to be later submitted to, or held for inspection by, the FDA as part of an application for a research or marketing permit? A subject is an individual (either health or a patient) who is a recipient of the test article or a control.

   *Test article *Test article* means any drug (including a biological product for human use), medical device for human use, human food additive, color additive, electronic product, or any other article subject to regulation under the Food, Drug, and Cosmetic Act.

   ☐ Yes , IRB review required ☐ No

2. Is this is a clinical investigation or research involving one or more human subjects to determine the safety or effectiveness of a device? A subject is an individual (healthy or has a medical condition or disease) on whom or on whose specimen an investigational device is used, or who participates as a control.

   ☐ Yes, IRB review required ☐ No

3. Is this an experiment in which a drug is administered or dispensed to, or used involving, one or more human subjects? This excludes the use of a marketed drug in the course of medical practice. A human subject is an individual (healthy or patient with a disease) that participates either as a recipient of the investigational new drug or as a control.

☐ Yes, IRB review required ☐ No

Instructions:
If IRB Review is required, you will need to submit NEW STUDY application eIRB.

SECTION IV: Complete this section if you have determined that your activities do not constitute human subjects research and you require written confirmation of this determination from the IRB Office. E-mail the signed form and any relevant supporting documents (i.e., grant, protocol, consent forms) to irb@northwestern.edu.

Investigator Information

Name (Last, First)	Degree(s)	University Status/Title
BILIMORIA, KARL Y.	**MD, MS**	**ASSISTANT PROFESSOR**
Department **DEPARTMENT OF SURGERY**		College **FEINBERG**
Phone Number **+1 (312) 695-4853**		E-mail Address **k-bilimoria@northwestern.edu**

Project Information

Project Title
Modification of Surgical Resident Duty Hours Study: A Cluster-Randomized Pragmatic Trial

Name of Funding Source (i.e., Department, NIH, Foundation)
Self-Funded

Grant Number (if applicable)

Project Description (describe the aims of the study and any activities involving interaction, intervention with human subjects, and/or their information or specimens)

Summary of Study Design Features That May Influence Whether Proposed Study Constitutes Human Subjects Research

- Aim is to evaluate the effect of a **policy change** (resident duty hour restriction policies) on patient outcomes and resident perceptions/wellbeing
- **Units of randomization are organizations (residency programs)**, not individual persons
- **Interventions are at the level of organizations (residency programs)**, and will involve a change of resident duty hour restrictions

- The intervention will involve a change in selected ACGME resident duty hour restrictions. Half of the participating residency programs will be randomized to usual care (current duty hour restrictions). Half of the participating residency programs in this study will be randomized to the intervention arm (relaxed duty hour restrictions).
- Apart from an informational webinar to recruit residency programs, there will be no direct interaction between Study Team members and participating residency programs/organizations.
- There will be **no direct contact/interaction/intervention between Study Team and human subjects**
- This study has been sanctioned, supported and approved by the Accreditation for Graduate Medical Education, the American Board of Surgery, and the American College of Surgeons
- Data for evaluation will come from the American College of Surgeons National Surgical Quality Improvement Program (ACS NSQIP). These data are collected as part of an ongoing process through hospitals' prior and ongoing participation in ACS NSQIP (participation is independent of the proposed study). The Study Team will only have access to **coded private information** for the purposes statistical analyses. The coded dataset will not contain any direct patient identifiers or dates. These data will be made available in a coded, private dataset to the Study Team by the ACS NSQIP.
- Additional data for evaluation will come from a module to be added to the American Board of Surgery In-Training Examination (ABSITE). The ABSITE is a compulsory examination administered annually to all residents in the month of January. The ABS will include a special module in the January 2015 ABSITE that asks residents questions regarding their perception of the effectiveness of training, as well as their satisfaction with training. The Study Team will only have access to **coded private information** for the purposes of statistical analyses. The data will be collected by the ABS, and then made available to the Study Team in the form of a coded, private dataset that does not include any direct individual-level identifiers.

Signature of Investigator: ___  _____

Date: __11/19/2013_____

SECTION V: IRB Determination (to be completed by IRB Office)*

The activities as described in the ☒ submitted protocol and/or ☒ materials and description of activities provided by the investigator,

☒ **Do not constitute research with human subjects in accordance with 45 CFR 46 and 21 CFR 50 & 56. IRB approval is not required.**

☐ **For activities involving decedents and their Protected Health Information (PHI), the conditions as set forth in the Privacy Rule at 45 CFR 164.512(i)(l)(iii), have been met.**

- the use or disclosure sought is solely for research on the protected health information of decedents;
- documentation can be provided, at the request of the covered entity, of the death of such individuals; and
- the protected health information for which use or disclosure is sought is necessary for the research purposes.

Authorized IRB Personnel Printed Name: _Kathleen E. Murphy, PhD, CIP_____

Authorized IRB Personnel Signature: Kathleen E Murphy, PhD, CIP

Title: Manager, Social and Behavioral IRB, Northwestern University, Evanston, IL_____

Date: _11-21-2013_____

*If any activities completed were or possibly were not in compliance with federal regulations regarding prior IRB review, please forward the form to the IRB Compliance Manager for review. For example, the investigator reports activities which are already completed but initially required IRB approval.

L. ACKNOWLEDGEMENT OF ALL FIRST TRIAL PROGRAM DIRECTORS, PROGRAM COORDINATORS, SURGEON CHAMPIONS, AND SURGICAL CLINICAL REVIEWERS

HOSPITAL NAME	RESIDENCY PROGRAM	PROGRAM DIRECTOR	PROGRAM COORDINATOR	SURGEON CHAMPION	SURGICAL CLINICAL REVIEWER
University of Alabama at Birmingham	University of Alabama Medical Center Program	John R. Porterfield, MD	Regina Lynn Hough	Mary Hawn, MD	Ashley L Webster Beth Faust
Mayo Clinic Arizona	Mayo Clinic Arizona Program	Richard J. Gray, MD	Carolyn Pence-Smith	David Etzioni, MD Richard Fowl, MD	Sharon Black Twila Lobitz Yolanda Nichols
Riverside County Regional Medical Center	Riverside County Regional Medical Center Program	Afshin Molkara, MD	Jenni Shieck	Yong-Kwon Lee, MD	Sheila Gilbert
Santa Barbara Cottage Hospital	Santa Barbara Cottage Program	Jeffrey Gauvin, MD	Cynthia Stoddard	Pamela Lee, MD	Caroline C. Finet
Kaiser Permanente Santa Clara	Stanford University Program	Marc Melcher, MD, PhD	Anita Hagan	John Stevenson, MD	X
Stanford Hospital and Clinics	Stanford University Program	Marc Melcher, MD, PhD	Anita Hagan	David Spain, MD	Candy McKinley
UCLA Medical Center	UCLA Medical Center Program	Oscar J. Hines, MD	Chi Quach	Oscar Joe Hines, MD	Hallie Chung
Kaiser Permanente Oakland Medical Center	UCSF (East Bay) Program	Terrence H. Liu, MD	Martha George	Olakunle Ajayi, MD Christopher Grimsrud, MD	Ann Conroy
UC San Francisco	University of California (San Francisco) Program	Linda M. Reilly, MD	Rachelle Bresnahan	Mary McGrath, MD	Tennille Parsons Yanina Stanislavskaya
Sutter West Bay Hospitals dba CPMC (California Pacific Medical Center)	University of California (San Francisco) Program	Linda M. Reilly, MD	Rachelle Bresnahan	Peter C. Richards, MD	Linda Ono Marissa Luber Yali Shu
Kaiser Permanente San Francisco	University of California (San Francisco) Program	Linda M. Reilly, MD	Rachelle Bresnahan	James Constant, MD	Millie Barnett
Kaiser Foundation Sacramento	University of California Davis Program	Joseph Galante, MD	Bryan Fandrich, MD	Damon Herr, MD	Kathryn Unger Margaret Chabot
University of California Davis Medical Center	University of California Davis Program	Joseph Galante, MD	Juanita Braxton	James Holcroft, MD	Anne Marder Kimberly Brink-Capps Roxanne Hyke
UC Irvine Medical Center	University of California Irvine Program	Matthew O. Dolich, MD	Tania Saba	Ninh Nguyen, MD	Sidney Diniz
Exempla St. Joseph Hospital	Exempla Saint Joseph Hospital Program	John T. Moore, MD Deborah Davis-Merrit, MD	Laurie Cooper	Margaret Schrieber, MD	Jill L. Grivetti Rhonda Simpson
University of Colorado	University of Colorado Program	Mark R. Nehler, MD	Claire Travis	Robert Meguid, MD David Kuwayama, MD	Nora Hennecken Sandra Espinoza
Danbury Hospital	Danbury Hospital Program	Royd Fukumoto, MD	Meryl Bennett	Keith Zuccala, MD	Christie Good

HOSPITAL NAME	RESIDENCY PROGRAM	PROGRAM DIRECTOR	PROGRAM COORDINATOR	SURGEON CHAMPION	SURGICAL CLINICAL REVIEWER
Saint Mary's Hospital	Saint Mary's Hospital Program	John A. Palesty, MD	Joan Reeser	Philip R. Corvo, MD	Kim O'Meara Sheila Staib
Stamford Hospital	Stamford Hospital/Columbia University College of Physicians and Surgeons Program	Kevin Dwyer, MD	Carla Rennie	Kevin M. Dwyer, MD	Suman Chaudhry
Hartford Hospital	University of Connecticut School of Medicine Program	Brian Shames, MD	Patricia Reilly	Orlando Kirton, MD	Jay Encarnacion
Hospital of Central Connecticut	University of Connecticut School of Medicine Program	Brian Shames, MD	Patricia Reilly	Michael Posner, MD	Cynthia Ross-Richardson
John Dempsey Hospital	University of Connecticut School of Medicine Program	Brian Shames, MD	Patricia Reilly	Stephen J. Lahey, MD	Jessica Bernard
Saint Francis Hospital and Medical Center	University of Connecticut School of Medicine Program	Brian Shames, MD	Patricia Reilly	Scott Ellner, MD	Gail Gruszczynski
George Washington University	George Washington University Program	Paul Lin, MD Juliet Lee, MD	Robert Pakan	Khashayar Vaziri, MD Stanley Knoll, MD	Kara Coullard Christina Junker
Christiana Care	Christiana Care Health Services Program	Frederick Giberson, MD	Sandy DelCoglin	Gerard Fulda, MD	Eileen Przybylek Jenny Marowski Jo Ann Beddow Rocco DeMaio
Orlando Regional Medical Center	Orlando Health Program	Michael Cheatham, MD	Joann Whittington	Matthew Lube, MD	Gayle C. Amberson Lisa M Allen
UF Health Jacksonville	University of Florida College of Medicine- Jacksonville Program	Michael Nussbaum, MD	Patricia Edwards	Joseph Tepas, MD Michael Nussbaum, MD	Jhun A Devilla
Tampa General Hospital	University of South Florida Morsani Program	John Y. Cha, MD	Wendy McCrorey	Victor Velanovich, MD David Smith, MD	Dena Waskiewicz
Dwight David Eisenhower Army Medical Center	Dwight David Eisenhower Army Medical Center Program	David Kauvar, MD	X	Dominic Gallo, MD James D. Frizzi, MD	Clifette Johnson
Emory University Hospital	Emory University Program	Keith A. Delman, MD	Susan Ratliff	John Sweeney, MD	Amy Newell Judy Lewis
Medical Center of Central Georgia	Medical Center of Central Georgia/ Mercer University School of Medicine Program	Benjie Christie, MD	Irma Miranda	Kim Thompson, MD	Michelle Chapman
Memorial University Medical Center	Mercer University School of Medicine (Savannah Campus) Program	Christopher Senkowski, MD	Debbie Wells	Carl Boyd, MD	Maureen Davis Sonja Marcey Soeffner
Queen's Medical Center	University of Hawaii Program	Danny Takanishi, MD	Gary Belcher	Kathleen Mah, MD Whitney Limm, MD	Ruby Adams Stacy H Ujimori Wanda M Muranaka

HOSPITAL NAME	**RESIDENCY PROGRAM**	**PROGRAM DIRECTOR**	**PROGRAM COORDINATOR**	**SURGEON CHAMPION**	**SURGICAL CLINICAL REVIEWER**
Straub Hospital and Clinic	University of Hawaii Program	Danny Takanishi, MD	Gary Belcher	Bradely Sakaguchi, MD Scott Crawford, MD	Kevin Speyer Thomas Yamashita
Kapioloani Medical Center for Women and Children	University of Hawaii Program	Danny Takanishi, MD	Gary Belcher	Richard McCartin, MD Russell Woo, MD	Deborah A Martyniuk
Iowa Methodist Medical Center	Iowa Methodist General Surgery Program	Richard A. Sidwell, MD	Paula Rasmussen	Frederick Nuss, MD	Paul Van Ryswyk
Mercy Medical Center-Des Moines	Mercy Medical Center- Des Moines Program	Charles Goldman, MD	Lori Wahman	Charles Goldman, MD	X
University of Iowa Hospitals and Clinics	University of Iowa Hospitals and Clinics Program	William J. Sharp, MD	Michael Healy	Timothy Kresowik, MD	Belding-Schmitt Mary Nancy Krutzfield
Carle Foundation Hospital	Carle Foundation Hospital Program	Michelle M. Olson, MD	X	Kimberly Cradock, MD	Jan Bice Lori Fossier
Northwestern Memorial Hospital	McGaw Medical Center of Northwestern University Program	Jonathan Fryer, MD Shari Meyerson, MD	Leslie McSpadden	Karl Bilimoria, MD MS	Kara J. Nelis Kathryn Paredes Nancy Tomaska
Rush University Medical Center	Rush University Medical Center Program	Norman L Wool, MD	Delores Austin	Jonathan Myers, MD	Patrick O'Brien
Memorial Medical Center	Southern Illinois University Program	John D. Mellinger, MD	Nikki Workman	Jan Rakinic, MD	Laura Antenan
OSF St. Francis Medical Center	University of Illinois College of Medicine-Peoria Program	Norman C. Estes, MD	Marnie Koeppel	Norman Estes, MD	Gail Sexton Karen Doty Linda Cooper
Indiana University Health (IUH)	Indiana University Program	Jennifer Choi, MD	Brianne Nickel	Eugene Ceppa, MD	X
IU Health Methodist	Indiana University Program	Jennifer Choi, MD	Brianne Nickel	Christopher Bearden, MD	Elizabeth "Betty" Roberts
St. Vincent Hospital Indianapolis	St. Vincent Hospitals and Health Care Center Program	Paul Nelson, MD Jonathan Saxe, MD	Lisa Stuart	Juliana Meyer, MD	Eileen Mclnnes
University of Kansas Hospital	University of Kansas School of Medicine Program	Kurt Schropp, MD	Kelly Dale	Chris Haller, MD	Jaime Davis-Thomas
University of Kentucky	University of Kentucky College of Medicine Program	Eric D. Endean, MD	Pamela Creech	Patrick C. McGrath, MD	Devauna Riley Roseanna Adair
Ochsner Clinic Foundation	Ochsner Clinic Foundation Program	George M. Fuhrman, MD	Denise Pinkston	X	Angela Teagle Brenda Falanga
Baystate Medical Center	Baystate Medical Center/Tufts University School of Medicine Program	Neal Seymour, MD	Joy Isotti	Jay Kuhn, MD	Christine Anderson Jane Stauber-Wilson Jodi Kashouh Linda Burgess Patricia Humiston

HOSPITAL NAME	RESIDENCY PROGRAM	PROGRAM DIRECTOR	PROGRAM COORDINATOR	SURGEON CHAMPION	SURGICAL CLINICAL REVIEWER
Beth Israel Deaconess Medical Center	Beth Israel Deaconess Medical Center Program	Tara Kent, MD	Kelly Barnes	Richard Whyte, MD	Mary Beth Cotter Mary F. Ward Valentina Lavarias
Brigham and Womens Faulkner Hospital	Brigham and Women's Hospital Program	Douglas S. Smink, MD, MPH	Pardon R Kenney, MD	Pardon Kenney, MD	Alexandra Koffman Evelyn Haas Felix O Akinbami Jill Steinberg Tess Panizales
Brigham and Womens Hospital	Brigham and Women's Hospital Program	Douglas S. Smink, MD, MPH	Sara Broughton Herd	Dennis Orgill, MD	Evelyn Haas Jill Steinberg Maria Theresa (Tess) Panizales
Lahey Hospital and Medical Center	Lahey Clinic General Surgery Program	Harold Welch, MD	Susan Downer	Rocco Ricciardi, MD	Lynne Crawford Nancy Manfredi Therese Golden
Massachusetts General Hospital	Massachusetts General Hospital Program	John T. Mullen, MD	Barbara Wolf	Matthew Hutter, MD	Kathy Swierzewski Lynn Devaney Shaun Sutcliffe
Newton-Wellesley Hospital	Massachusetts General Hospital Program	John T. Mullen, MD	Sheila Partridge, MD	Frederick Millham, MD Sheila Partridge, MD	Linda Burr
Tufts Medical Center	Tufts Medical Center Program	Jeffrey T. Cooper, MD	Annette Cerulli	William C. Mackey, MD	Rita Estey
Umass Memorial Health Care	University of Massachusetts Program	Anne Larkin, MD	Jeannine Bottis	W. Brian Sweeney, MD	Gail Butcher Joanne Pascarelli-
John Hopkins Hospital	John Hopkins University Program	Pamela A. Lipsett, MD	L. Robin Newcomb Kimberly Duncan	Martin Makary, MD	Jennifer Castellani Regina Morton
Sinai Hospital of Baltimore	Sinai Hospital of Baltimore Program	Mark Katlic, MD	Jean Sturdivant	Thomas Genuit, MD	Karen Sweeney Mary L Garland
University of Maryland	University of Maryland Program	Stephen M Kavic, MD	Sarah Kidd-Romero	X	Lewelyn Cuevas Cariaga
Maine Medical Center	Maine Medical Center Program	James F. Whiting, MD	Jennifer Perros	Brad Cushing, MD	Kimberly A. Newman Robert Cormier
Spectrum Health Butterworth	Grand Rapids Medical Education Partners/Michigan State University Program	Mathew Chung, MD Stanley Sherman, MD Paul Kemmeter, MD Jeremiah Awori Hayanga, MD	Marc Schlatter, MD	Ashraf Mansour, MD	Stephanie Laird
Henry Ford Hospital	Henry Ford Hospital/Wayne State University Program	Ann Woodward, MD	Grace Pacini	J.H. Patton, MD	Jennifer Ritz Misty Desimpelaere Rita Straub

HOSPITAL NAME	RESIDENCY PROGRAM	PROGRAM DIRECTOR	PROGRAM COORDINATOR	SURGEON CHAMPION	SURGICAL CLINICAL REVIEWER
Sparrow Hospital	Michigan State University Integrated Program	Michael K. McLeod, MD	Lisa Rendall	Michael McLeod, MD	Anita Kassel Jori Smith MaryAnn Taylor
St. Joseph Mercy Hospital	St Joseph Mercy Hospital Program	Edward Kreske, MD	Erin Madden	Wallace Arneson Jr., MD	James Vandewarker Sally A. Knight
Bronson Methodist Hospital	Western Michigan University School of Medicine Program	Earl Norman, MD	Cynthia Shattuck	Mark Dittenbir, MD	Deborah Rozewicz Erica Nagra
Beaumont Health System (Grosse Pointe)	William Beaumont Hospital Program	Felicia A. Ivascu, MD	Larry Lloyd, MD	Larry Lloyd, MD	Julie Pelton Karen Reeder
William Beaumont Hospital	William Beaumont Hospital Program	Felicia A. Ivascu, MD	Kathy Janowski	Robert Welsh, MD	Catherine Shuell Elizabeth Gates Julie Pelton Patricia Ciofu-Smith
Hennepin County Medical Center	Hennepin County Medical Center Program	Joan M Van Camp, MD	Phyllis Squiers	Jon Krook, MD	Megan Oberle Sheri Dodd
Mayo Clinic Methodist	Mayo Clinic College of Medicine (Rochester) Program	Stephanie F. Heller, MD	Judith Cook	X	Diane Tyndale Mary Roubik Sharon Nehring
Mayo Clinic Saint Marys Hospital	Mayo Clinic College of Medicine (Rochester) Program	Stephanie F. Heller, MD	Judith Cook	Sean Dowdy, MD	Kim Giehtbrock Sharon Nehring
University of Minnesota Medical Center	University of Minnesota Program	Jeffrey G. Chipman, MD	Cathryn Larson	Mary Kwaan, MD	Alyssia Mills-Hokenson Stacy Jo Carda
Saint Louis University	St Louis University School of Medicine Program	Catherine Wittgen, MD	Carol Kamp	Donald Jacobs, MD	Martha Antal
University of Missouri - Columbia	University of Missouri-Columbia Program	Arthur Rawlings, MD	Bethany Bennett	X	Linda Hanley
Truman Medical Center	University of Missouri-Kansas City Program	Mark Friedell, MD	X	Mark Friedell, MD	X
Barnes Jewish West County Hospital	Washington University/B-JH/SLCH Consortium Program	Paul Wise, MD	Michelle Tuetken	Sam Bhayani, MD	Mary Johnson
Barnes-Jewish Hospital	Washington University/B-JH/SLCH Consortium Program	Paul Wise, MD	Michelle Tuetken	Bruce Hall, MD	Carmen Broccard Louise H. Schrama Mitzi Hirbe
Carolinas Medical Center	Carolinas Medical Center Program	John M. Green, MD	Jessica Roof	Michael H. Thomason, MD	Meredith Moore
Duke University Hospital	Duke University Hospital Program	John Migaly, M.D.	Tammy Watson	Christopher Mantyh, MD	Monica R Walter Pat Tucker Yvonne Acker
Womack Army Medical Center	Dwight David Eisenhower Army Medical Center Program	David Kauvar, MD	Raymond Sanders	Steven Khoo, MD	X

HOSPITAL NAME	RESIDENCY PROGRAM	PROGRAM DIRECTOR	PROGRAM COORDINATOR	SURGEON CHAMPION	SURGICAL CLINICAL REVIEWER
New Hanover Regional Medical Center	New Hanover Regional Medical Center Program	Thomas Clancy, MD	Kathy Radley	William Hope, MD	Pam Moore
UNC Hospitals	University of North Carolina Hospitals Program	Michael O. Meyers, MD	Kathie Patterson	Mark Koruda, MD	Lynn Flagg Marcia Prince
Vidant Medical Center	Vidant Medical Center/East Carolina University Program	Claudia Goettler, MD	Sue West	Claudia Goettler, MD	X
Wake Forest Baptist Health	Wake Forest University School of Medicine Program	John Stewart, MD	Mollie Draughon	Perry Shen, MD	Mary Ann Mealor Pamela Eversole
Alegent Creighton Health, Creighton University Medical Center	Creighton University Program	Jeffrey T. Sugimoto, MD	Rhonda Peavy	Sumeet Mittal, MD	Christina Graf Lisa Schuster
Nebraska Medical Center	University of Nebraska Medical Center Program	Chandrakanth Are, MD	Danielle Brown	Eugene Waltke, MD	Andrea Paxton Jocelyn Pearson
Dartmouth-Hitchcock Medical Center	Dartmouth-Hitchcock Medical Center Program	Paul Kispert, MD	Karen Lee	Philip Goodney, MD	Erin L. Boettcher Mary Menduni
Morristown Medical Center	Atlantic Health System Program	Eric L. Lazar, MD	Catherine Nitto	Brian Siegel, MD	Joanne Pawar Patricia Vorel
Cooper University Hospital	Cooper Medical School Program of Rowan University	James B. Alexander, MD	Cathy Cooney	Francis Spitz, MD	Catherine Cristofalo David Spurrier Dawn Stepnowski Mary Buddle
Newark Beth Israel Medical Center	Monmouth Medical Center Program	Mark K. Hirko, MD	Donna Turovac	Adam Kopelan, MD	Constance McKoy-Holt
Hackensack University Medical Center	Rutgers New Jersey Medical School Program	Michael Shapiro, MD	Michelle Jimenez	Massimo Napolitano, MD	Inia Estima Magdalena Sudol
University Hospital - Rutgers	Rutgers New Jersey Medical School Program	Michael Shapiro, MD	Michelle Jimenez	Aziz Merchant, MD Adam Fox, MD Michael Curi, MD	Kimberly B Nester Roxanne M Poon
MetroHealth Medical Center	Case Western Reserve University Program	Jeremy Lipman, MD	Jennifer Lastic	Natalie Joseph, MD	Judi Spath Maria Opris
Cleveland Clinic	Cleveland Clinic Foundation Program	Allan E. Siperstein, MD	Janine Keough	Allan Siperstein, MD	Jeanne Shewchik Meryl Insler Susan Bohne Susan M. Rydzinski Nancy Anzlovar
Ohio State University Wexner Medical Center	Ohio State University Wexner Medical Center Program	Mark Arnold, MD	Beth Hanson	Steven Steinberg, MD	Erica Porter Judi Michalek
The Jewish Hospital	The Jewish Hospital Program	Carrie Ogg, MD	Amy Broughton	S. Russell Vester, MD Cari Ogg, MD	X

HOSPITAL NAME	RESIDENCY PROGRAM	PROGRAM DIRECTOR	PROGRAM COORDINATOR	SURGEON CHAMPION	SURGICAL CLINICAL REVIEWER
Good Samaritan - TriHealth	TriHealth (Good Samaritan Hospital) Program	Kevin J. Grannan, MD	Teresa Arnold	George Kerlakian, MD	Donna M Werth
The Christ Hospital	University of Cincinnati Medical Center/College of Medicine Program	Bradley R. Davis, MD	Gilda Young	Ian Paquette, MD	Emily Gatch Priyanka Prakash
Kaiser Permanente Sunnyside	Oregon Health & Science University Program	Karen Brasel, MD	Robin Alton	Waleed L. Lutfiyya, MD	Andrea M Calarco Juliann Breen
Legacy Emanuel Medical Center	Oregon Health & Science University Program	Karen Brasel, MD	Robin Alton	Nathan Kemalyan, MD	Becky Swick
Legacy Good Samaritan Medical Center	Oregon Health & Science University Program	Karen Brasel, MD	Robin Alton	Blayne Standage, MD	X
Oregon Health and Science University Hospital	Oregon Health & Science University Program	Karen Brasel, MD	Robin Alton	Brett Sheppard, MD	Fouad Attia
Providence Portland Hospital	Oregon Health & Science University Program	Karen Brasel, MD	Robin Alton	Karen Zink, MD Kelvin Yu, MD	Annette Bruer Kristina Loudon
Providence St. Vincent Hospital	Oregon Health & Science University Program	Karen Brasel, MD	Robin Alton	Ali Khaki, MD	Annette Bruer Scott Kato
Abington Memorial Hospital	Abington Memorial Hospital Program	Kenric M. Murayama, MD	Rebecca Augustine	John Kukora, MD	Cynthia Brophy Karen Beem
Hahnemann University Hospital	Drexel University/Hahnemann University Hospital Program	Andres E. Castellanos, MD	Allison Stein	David Stein, MD	Patricia Fisher
Robert Packer Hospital	Guthrie/Robert Packer Hospital Program	Thomas J. VanderMeer, MD	Laura Warner	Thomas VanderMeer, MD	Laurie Kinsman Nicole Teeter
Hospital of the University of Pennsylvania	Hospital of the University of Pennsylvania Program	Jon B. Morris, MD	Laura Huth	Rachel Kelz, MD	Susan Kreider
Penn State Milton S Hershey Medical Center	Penn State Milton S Hershey Medical Center Program	David Han, MD	Jessica Moyer	Matthew Indeck, MD	Gail Ortenzi Linda Burgess
Temple University Hospital	Temple University Hospital Program	Amy J. Goldberg, MD	Kiesba Herrin	Eric Choi, MD	Cynthia Brophy Kathleen Campbell
Thomas Jefferson University Hospital	Thomas Jefferson University Program	Karen Chojnacki, MD	Donna Guinto	Scott Cowan, MD Stacey Milan, MD Herbert Cohn, MD	Randi Altmark
Pennsylvania Hospital	University of Pennsylvania Program	Jon B. Morris, MD	Laura Huth	Dahlia Sataloff, MD	Jessica Stevens John Regan
UPMC Presbyterian Hospital	UPMC Medical Education Program	Kenneth K. Lee, MD	Maggie Mrozinski	Kevin O. Garrett, MD	Denise (Dee) Burkhart
York Hospital	York Hospital Program	Richard B. Damewood, MD	Mark Neal	John Castronuovo, MD	Kelly Gemmill Pamela Emig Susan Diehl
Rhode Island Hospital	Brown University Program	David T. Harrington, MD	Pamela Richardson	David Harrington, MD	X

HOSPITAL NAME	RESIDENCY PROGRAM	PROGRAM DIRECTOR	PROGRAM COORDINATOR	SURGEON CHAMPION	SURGICAL CLINICAL REVIEWER
The Miriam Hospital	Brown University Program	David T. Harrington, MD	Pamela Richardson	David Coultier, MD	Mary Valencia Pat Sullivan
Medical University of South Carolina	Medical University of South Carolina Program	Chris Streck, MD	Sue Wetherholt	Karl Byrne, MD	Christie Merritt Deborah R. Lorris
Bristol Regional Medical Center	East Tennessee State University Program	William Browder, MD	Julie B. Simerly	Benjamin S. Scharfstein, MD	Leilani M Evans
University of Tennessee Medical Center at Knoxville	University of Tennessee at Knoxville Program	Brian Daley, MD	Ginger Miya	Blaine Enderson, MD	Debra A Vittetoe
Erlanger Hospital	University of Tennessee College of Medicine at Chattanooga Program	Heath Giles, MD	Cindy Schultz Rudolph	Joseph Cofer, MD	Amy Harvey Patricia Spangler
Baptist Memorial Hospital	University of Tennessee Program	Frances E. Pritchard, MD	Cynthia Tooley	Stephen Behrman, MD	Jackie Cibulka Kay Loyd
Methodist Healthcare	University of Tennessee Program	Frances E. Pritchard, MD	Cynthia Tooley	Alexander Mathew, MD Martin D Fleming, MD	Bobbie Hale
Regional Medical Center	University of Tennessee Program	Frances E. Pritchard, MD	Cynthia Tooley	George O. Maish, MD	Sandy Long
St. Francis Hospital - Memphis	University of Tennessee Program	Frances E. Pritchard, MD	Cynthia Tooley	Joshua Katz, MD	Cindy Wylie
St. Thomas West Hospital	Vanderbilt University Program	John Tarpley, MD	Stephanie Burnham	Raymond S. Martin, MD	X
Vanderbilt Medical Center	Vanderbilt University Program	John Tarpley, MD Kyla Terhune, MD	Stephanie Burnham	Oscar Guillamondegui, MD	Barbara Martin Sherree Levering
John Peter Smith- Tarrant County	Baylor University Medical Center Program	Robert Goldstein, MD	Sandy Fishman	David McReynolds, MD	Brenda Ellis Julie Chenoweth
Baylor University Medical Center	Baylor University Medical Center Program	Robert Goldstein, MD	Sandy Fishman	Ernest Franklin, MD	Jacqueline Wohadlo
Houston Methodist	Houston Methodist Program	Sherilyn Gordon Burroughs, MD	Myriam Gandy	Barbara Bass, MD	X
Scott and White Healthcare	Scott and White Healthcare Program	J. Scott Thomas, MD	Lynn Botts	Harry T. Papaconstantinou, MD	Bonnie Hodges Christie C Cummings Glenda Goolsby Nancy A Bowman
Memorial Hermann Southwest	University of Texas at Houston Program	Donald P. Lesslie, DO	Angel Lopez	Tammy Lee, MD George Peterkin, MD	X

HOSPITAL NAME	RESIDENCY PROGRAM	PROGRAM DIRECTOR	PROGRAM	SURGEON CHAMPION	SURGICAL CLINICAL

			COORDINATOR		REVIEWER
University Hospital - San Antonio	University of Texas Health Science Center at San Antonio Program	Daniel L. Dent, MD	Eileen M. Kleffner	Ronald M. Stewart, MD	Kristi Hill- Herrera
University of Texas M.D. Anderson	University of Texas at Houston Program	Donald P. Lesslie, DO	Angel Lopez	X	Annie Z. Philip Constance R Curtis Lavinia Zanaj Maria VictoriaTiu Melony Levy
Hermann Memorial TMC	University of Texas at Houston Program	Donald P. Lesslie, DO	Angel Lopez	Todd Wilson, MD Erik Wilson, MD	Ira Martin Lucia Flores Regina Essex
University of Texas Medical Branch	University of Texas Medical Branch Program	Kristene Gugliuzza, MD	Erica Ruiz	Dennis Gore, MD	Michelle Gonzalez Theresa Speich
Parkland Hospital	University of Texas Southwestern Medical School Program	Daniel J. Scott, MD	Lisa Bailey	Jennifer Rabaglia, MD Michael Choti, MD	Emily J. Kent-Street Reina Duhon
University Hospital UT Southwest	University of Texas Southwestern Medical School Program	Daniel J. Scott, MD	Lisa Bailey	X	Nisha Jose
Intermountain Medical Center	University of Utah Program	Daniel Vargo, MD	Janell Clements	Mark Ott, MD Ute Gawlick, MD	Brett Bulloch
University of Utah Hospital	University of Utah Program	Daniel Vargo, MD	Lori Bybee	Robert Glasgow, MD	Judy Larsen Karie Cluff Linchee Cheong Natalie Turner
Inova Fairfax Hospital	Inova Fairfax Program	Jonathan Dort, MD	Diann Carreker	H. David Reines, MD	Jean Donovan
Naval Medical Center Portsmouth, Virginia	Naval Medical Center (Portsmouth) Program	Angela S. Earley, MD	Dovie I Loud	Robert Strange, MD	Laurie Erskine
University of Virginia	University of Virginia Program	Bruce D. Schirmer, MD	Kristen Dudley	Traci Hedrick, MD John Hanks, MD	Beth Turrentine Lynn Murray
Winchester Medical Center	Virginia Commonwealth University Program	Brian J. Kaplan, MD	Cindi Phares	Erich Bruhn, MD	Edward Damico
Carilion Roanoke Memorial Hospital	Virginia Tech Carilion School of Medicine Program	Charles Chuck Paget, MD	Tina Toms	Sandy L. Fogel, MD	Debbie Copening James Jones Lisa Turner Patti Shorner
Fletcher Allen Health Care	University of Vermont/ Fletcher Allen Health Care Program	Julie Adams, MD	Diantha Langmaid	Paul Penar, MD	Brenda Murphy Joanne Rheaume Joey Larson

HOSPITAL NAME	RESIDENCY PROGRAM	PROGRAM DIRECTOR	PROGRAM	SURGEON CHAMPION	SURGICAL CLINICAL

			COORDINATOR		REVIEWER
Gundersen Lutheran Medical Center	Gunderson Lutheran Medical Foundation Program	Benjamin T. Jarman, MD	Colette O'Heron	Travis Smith, MD	Julie Trussoni Pam Lambert
University of Washington Hospitals	University of Washington Program	Karen Horvath, MD Lisa McIntyre, MD	Gina Coluccio	Zoe Parr, MD David Flum, MD	Alex Ruiz Joshua Matlock
University of Wisconsin	University of Wisconsin Program	Eugene Foley, MD	Mara Snyder	Gregory D. Kennedy, MD	Barbara Braunger Karen Armstrong
Meriter Hospital	University of Wisconsin Program	Eugene Foley, MD	Mara Snyder	Jacquelynn Arbuckle, MD	Loretta Herfel Wendy L. McManners
West Virginia University	West Virginia University Program	Jon Cardinal, MD	Linda Shaffer	Matthew Loos, MD Richard Vaughan, MD	Brittany L Brooks Keri L. Orlando Michael E Jude Stephanie Kish