Operating Room Times For Teaching and Nonteaching Cases are Converging: Less Time for Learning?

Jordan D. Bohnen, MD, MBA,[*] David C. Chang, PhD, MPH, MBA,[*,†] and Brian C. George, MD, MAEd[‡]

[*]Department of Surgery, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts; [†]Codman Center for Clinical Effectiveness in Surgery, Massachusetts General Hospital, Boston, Massachusetts; and [‡]Department of Surgery, University of Michigan, Ann Arbor, Michigan

OBJECTIVE: To compare differences in operating room (OR) times between teaching and nonteaching cases across calendar years. We hypothesize that time devoted to intraoperative resident education is decreasing, therefore, OR times for teaching and nonteaching cases will be converging.

BACKGROUND: Teaching cases take longer than similar nonteaching cases, in part due to intraoperative resident education. Pressures to improve OR efficiency and patient safety may threaten resident education and leave less time for intraoperative learning; however, the magnitude of impact is unknown.

SETTING/PARTICIPANTS: National Surgical Quality Improvement Program (NSQIP) deidentified national databases from 2006 to 2012, queried for 30 most common General surgery procedures and case teaching status (i.e., teaching vs. nonteaching cases).

DESIGN: The NSQIP database was retrospectively reviewed to identify the 30 most common General Surgery procedures. Teaching cases included all operations in which a resident participated. Multivariable regression analyses were constructed to determine the impact of resident involvement on OR times, controlling for year, resident participation, procedure, and patient demographics and comorbidities. Difference-in-difference analysis was performed to assess OR time differences between teaching and nonteaching cases across calendar years and within subpopulations.

RESULTS: A total of 693,223 cases met inclusion criteria. Average overall OR times were 98.89 minutes (teaching) vs. 74.22 minutes (nonteaching), with a difference of 24.67 minutes (95% confidence interval [CI] 24.34-24.99 minutes, $p < 0.001$). In multivariable analyses, the difference between teaching and nonteaching cases was 21.94 minutes (95% CI = 21.11-22.76) in 2006 and 13.95 minutes (95% CI = 10.62-17.28) in 2012, with a difference-in-difference of 7.99 minutes per case. A similar trend was observed across individual PGYs and several individual procedures.

CONCLUSIONS: OR times for teaching and nonteaching cases converged by approximately 8 minutes per general surgery procedure during the 7-year study period, representing a 36% reduction in the difference between groups. We must seek to better understand the source of this convergence, and in doing so ensure to preserve and enhance the intraoperative learning experience of surgical trainees. (J Surg Ed 78:148−159. © 2020 Published by Elsevier Inc. on behalf of Association of Program Directors in Surgery.)

KEY WORDS: autonomy, difference-in-difference analysis, intraoperative resident education, NSQIP, OR times, surgical education

COMPETENCIES: Patient Care, Systems-Based Practice, Practice-Based Learning and Improvement

INTRODUCTION

There is growing concern that surgical residents may not be reaching readiness for independent practice by the time they graduate.[1] These concerns arise at a time when resident education faces numerous systemic challenges, including an erosion of trainee autonomy[2-7]; rising patient safety

This research did not receive any specific funding support from funding agencies in the public, commercial, or not-for-profit sectors.

A preliminary version of this research was presented as a podium presentation at the 2015 American College of Surgeons Clinical Congress on October 5th 2015, in Chicago, IL, where it received the ACS Excellence in Research Award.

Correspondence: Inquiries to Jordan D. Bohnen, MD, MBA, Department of Surgery, Massachusetts General Hospital, 15 Parkman Street, Boston, MA 02114.; e-mail: jbohnen@partners.org

  1931-7204/$30.00
https://doi.org/10.1016/j.jsurg.2020.06.029

considerations[7-19]; persistent medicolegal and malpractice concerns[5,20]; duty hours restrictions[2,21,22]; and increasing demands on faculty to be more efficient. Teaching cases are well known to take longer than nonteaching cases[15-19,23,24] and in an environment increasingly focused on productivity, pressures on attending surgeons to "finish a case" may therefore adversely affect resident education.

Unfortunately, there remains a paucity of directly-observed, longitudinal data to quantify changes in intraoperative resident education over time. Prior work has investigated trends in resident case volumes,[22,25-27] intraoperative trainee roles,[28] and perceptions of the education system more broadly,[22] but large prospective studies quantifying trends intraoperative resident learning do not exist. Recent efforts to collect and analyze these metrics at scale may someday be useful for this purpose,[29] but cannot yet provide longitudinal trend data beyond the last few years.

We set out to address this gap by quantifying whether the amount of time devoted to intraoperative resident education is changing over time. We sought to do so through an innovative use of data from the National Surgical Quality Improvement Program (NSQIP). NSQIP has collected data for the purpose of surgical quality improvement since the mid-2000s.[30] Since its inception, NSQIP has incorporated operating room (OR) times data for each recorded case. For several years NSQIP also included postgraduate year (PGY) of trainees involved in teaching cases. In our current study, we compare differences in OR times between teaching and nonteaching cases in the NSQIP database in order to establish if these differences are changing over time. We leverage the fact that teaching cases take longer than nonteaching cases, and that differences between teaching and nonteaching cases serve as indirect measures of intraoperative resident education. We hypothesized that the proportion of OR time devoted to teaching is decreasing and consequently that OR times for teaching and nonteaching cases are converging.

METHODS

Using the NSQIP database, we retrospectively calculated differences in OR times between teaching and non-teaching cases for thirty of the most common procedures performed by general surgery trainees, stratified by calendar year. Multivariable regression models were constructed to determine OR time differences between teaching and nonteaching cases in each year. A difference-in-difference analysis was performed to assess trends in OR time differences between teaching and nonteaching cases over time.

Database – NSQIP

NSQIP is a nationally recognized surgical outcomes database that collects data on pre-specified preoperative, intraoperative, and postoperative variables in a systematic and prospective fashion by trained and dedicated staff. The NSQIP data collection methodology has been previously described and validated.[30] Standard NSQIP data collection includes the following variables of interest in the present study: (1) resident involvement in the operating room (including PGY), (2) procedure type, (3) OR times, and (4) year of surgery. National deidentified NSQIP Participant Use Data Files (PUF) are currently available for the 2005 to 2018 calendar years.[31] Of these, PGY data are available from 2006 to 2012; our analyses are therefore restricted to those years.

Study Population: Inclusion and Exclusion Criteria

The national, deidentified NSQIP databases were electronically combined and queried for the 30 most common procedures performed by general surgeons, by Current Procedural Terminology (CPT) code. All patients who underwent one of the 30 most common procedures, by a general surgeon, were eligible for inclusion. Patients with missing "resident involvement" and "OR time" data were excluded (see next 2 subsections below), constituting the study population.

Resident Involvement – Teaching Versus Nonteaching Groups

The study population was divided into two groups. The "teaching" group included all cases in which a resident of any level participated. The "nonteaching" group included all cases in which only a surgical attending participated, without resident involvement. These groups were sorted electronically according to the NSQIP variable "PGY," where PGY = 0 represents cases with no resident (i.e., nonteaching cases), and PGY = 1-10 represents the highest PGY of the resident who scrubbed for the case (i.e., teaching cases).[32] Cases with missing data for the variable PGY were excluded.

Procedure Type

The 30 most common procedures were grouped by CPT code into 8 groups of related procedures. The groups were Appendectomy (44970, 44950); Bariatric (43644, 43770); Bowel/Adhesions/Exploratory laparotomy (49000, 44204, 44140, 44120, 44160, 44145, 44005, 44207); Breast (19125, 19303, 19301, 19120, 19307); Cholecystectomy (47562, 47563, 47600); Foregut (43280); Hernia (49505, 49560, 49585, 49650, 49587, 49561); Thyroid/Parathyroid (60240, 60500, 60220).

Operative Times

OR times were recorded as the total operation time in minutes per case, per NSQIP data abstraction criteria

(NSQIP variable "OPTIME").[32] Cases with missing data for this variable were excluded.

Primary Outcome

The primary outcome measure was the difference-in-difference between OR times of teaching and nonteaching cases across calendar years, expressed in minutes. Difference-in-difference analysis allows for the comparison of differences between groups over time, controlling for background changes that occur over time (i.e., secular trends).[33,34] Figure 1 provides a schematic illustration of a difference-in-difference analysis (not based on study data). The primary analysis was performed on the entire study population, including all PGY levels and all 30 procedure types combined.

Subpopulation Analyses

Subpopulation analyses were performed on uniform subsets of cases, stratified by PGY level (i.e., PGY1 through PGY5) and by individual groups of related procedures, to determine if the pattern observed for the entire study population was consistent among subgroups.

Statistical Analysis

Unadjusted comparison of categorical variables was performed using the chi-square test, and using *t*-tests for continuous variables. Multivariable regression analyses were performed using linear regression to determine the impact of resident involvement on operative times after controlling for procedure-related variables (e.g., CPT code, emergency status), patient demographics (e.g., age, gender), and comorbidities (e.g., hypertension, smoking, steroid use, cardiac, pulmonary, stroke, and renal comorbidities). Difference-in-difference analyses were performed using the univariate and multivariable regression model results to assess trends in OR time differences between teaching and nonteaching groups across calendar years. Subpopulation analyses assessed the difference-in-difference between OR times of teaching and nonteaching cases after stratifying by PGY level and procedure type. Multivariable regression models for the subpopulation analyses were constructed using the same variables as above for the entire cohort. Data from 2012 were excluded from subpopulation analyses due to small sample sizes of subpopulations in this year. Statistical significance was defined as $p < 0.05$.

RESULTS

Study Population

A total of 1,446,663 operative cases were performed by general surgeons during the study period, of which *937,471* (64.80%) underwent one of the 30 most common procedures; of these, 693,223 cases were categorized into either the teaching (n = 414,804, 59.84%) or nonteaching (n = 278,419, 40.16%) group (Fig. 2). There were no clinically meaningful differences between groups in terms of patient demographics, comorbidities, and preoperative characteristics (Table 1). Table 2 describes the distribution of procedures in the teaching and nonteaching groups.

Unadjusted Analysis

In unadjusted analysis, average OR time was 24.67 minutes longer for teaching cases than nonteaching cases across all calendar years (95% confidence interval [CI] 24.34-24.99 minutes, $p < 0.001$). Teaching cases were on average 29.06 minutes longer than nonteaching cases in 2006 (95% CI = 28.06-30.05, $p < 0.001$) and 14.57 minutes longer than nonteaching cases in 2012 (95% CI = 10.59-18.56, $p < 0.001$). Using the unadjusted difference-in-difference, the gap in operative times between teaching and

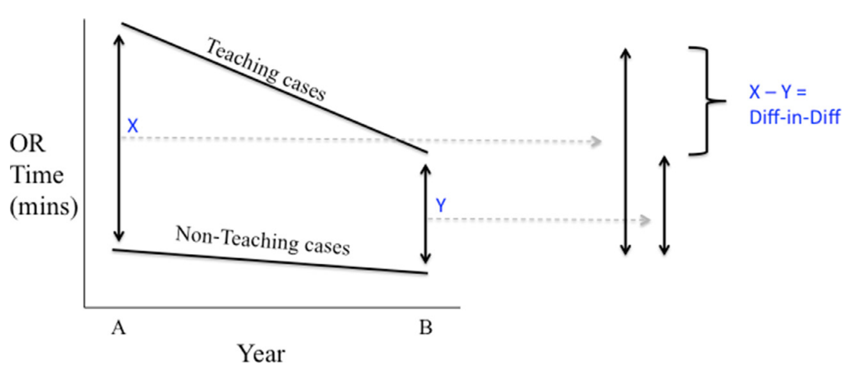

FIGURE 1. Schematic diagram of difference-in-difference analysis for trends in OR times between teaching and nonteaching cases. *X = difference in OR times for teaching vs. nonteaching cases at Time A. Y = Difference in OR times for teaching vs. nonteaching cases at Time B. X-Y (i.e., difference-in-difference) represents the magnitude of change in the difference in OR times between teaching and nonteaching cases from Year A to Year B. X > Y represents a decrease in the difference in OR times between teaching and nonteaching cases; X < Y represents an increase in the difference. * This schematic diagram is intended to illustrate the concept of difference-in-difference analysis, and is not based on actual study data.

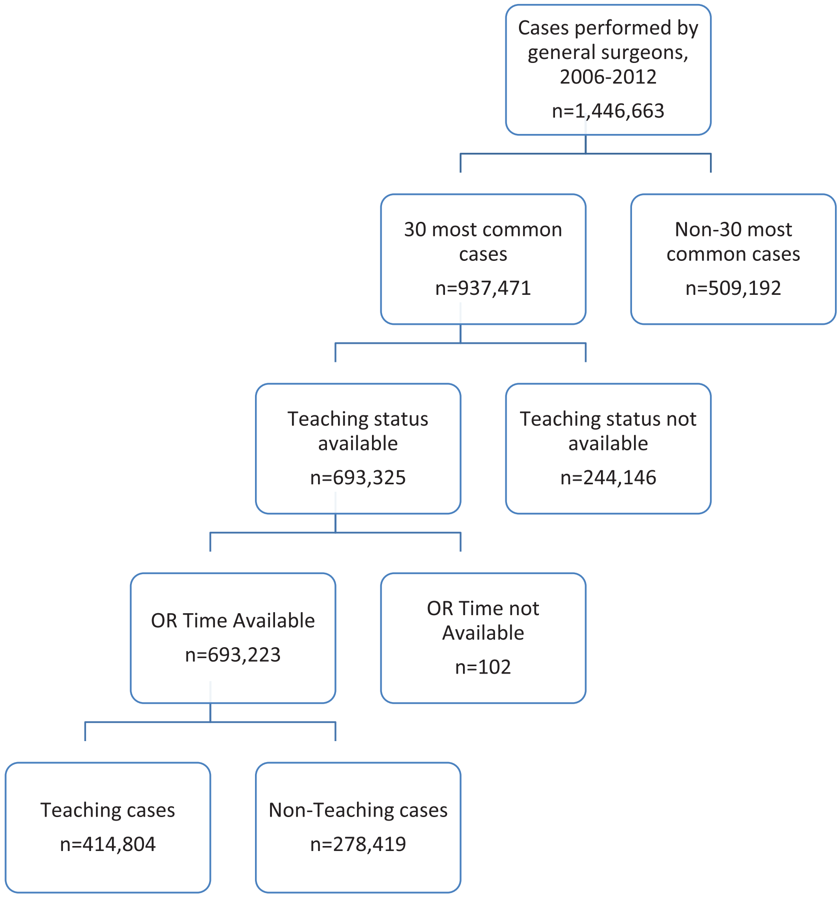

FIGURE 2. Flow diagram of study population.

nonteaching cases decreased by 14.49 minutes per general surgery case ($p < 0.05$; Table 3).

Multivariable Analysis

In multivariable analyses, average OR time was 20.45 minutes longer for teaching cases than nonteaching cases across all calendar years (95% CI = 20.18-20.72, $p < 0.001$). In 2006, teaching cases were on average 21.94 minutes longer than nonteaching cases (95% CI = 21.11-22.76, $p < 0.001$) whereas in 2012 teaching cases were 13.95 minutes longer than nonteaching cases (95% CI = 10.62-17.28, $p < 0.001$). The difference-in-difference between teaching and nonteaching cases was 7.99 minutes per general surgery case ($p < 0.05$; Fig. 3), representing a 36% (7.99/21.94 = 0.36) convergence in the difference between OR times of teaching and nonteaching cases during the study period.

Subpopulation Analyses

When stratified into uniform subpopulations by PGY and procedure type, teaching cases were on average longer than nonteaching cases in each of the 5 individual PGY levels, and across all of the procedure types included. Several PGY levels and procedure groups demonstrated the same trend as seen in the overall cohort, with a statistically significant convergence in OR times between teaching vs. nonteaching cases over time: the trend was statistically significant for PGY1, PGY2, and PGY5 after adjusting for differences in case-mix and other confounding variables, with directionally similar trends also observed for the PGY3 and PGY4 cohorts (Fig. 4a-e). Similarly, when stratified by procedure type, the Appendectomy, Bariatric, Cholecystectomy, and Hernia groups each experienced a statistically significant convergence in OR times between teaching and nonteaching cases over time (Fig. 5a-d). The remainder of procedure groups had heterogeneous and nonsignificant trends.

DISCUSSION

In a cohort of 693,325 General Surgery cases from the NSQIP national database, operating room times for teaching and nonteaching cases converged significantly over a 7-year period. During this time, the risk-adjusted

TABLE 1. Demographics, Comorbidities and Preoperative Characteristics of "Teaching" vs. "Nonteaching" Cases.

Patient Characteristic, n (%)	Total (%) N = 693,223	Teaching (%) N = 414,804	Nonteaching (%) N = 278,419	p-Value
Female gender	428,475/691,548 (61.96)	253,422 (61.21)	175,053 (63.08)	<0.001
Age, years; mean (95% CI)	52.34 (52.30-52.38)	52.08 (52.03-52.13)	52.73 (52.67-52.79)	<0.001
Race				<0.001
White	334,044 (76.94)	199,157 (75.49)	134,887 (79.20)	
Black	48,425 (11.15)	33,927 (12.86)	14,498 (8.51)	
Hispanic	35,956 (8.28)	20,545 (7.79)	15,411 (9.05)	
Asian/Pacific Islander	11,595 (2.67)	7,199 (2.73)	4,396 (2.58)	
American Indian/Alaska Native	4,120 (0.95)	3,006 (1.14)	1,114 (0.65)	
ASA class				<0.001
1- normal health	90,038 (13.03)	53,107 (12.83)	36,931 (13.33)	
2- mild comorbidity	365,393 (52.86)	218,024 (52.66)	147,369 (53.17)	
3- severe comorbidity	212,689 (30.77)	127,983 (30.91)	84,706 (30.56)	
4- life-threatening comorbidity	21,890 (3.17)	14,053 (3.39)	7,837 (2.83)	
5- moribund	1,193 (0.17)	882 (0.21)	311 (0.11)	
BMI category				<0.001
<18.5	12,421 (1.85)	7,806 (1.95)	4,615 (1.70)	
18.5-24.9	176,182 (26.21)	107,193 (26.76)	68,989 (25.39)	
25-29.9	197,231 (29.34)	118,936 (29.69)	78,295 (28.82)	
30-34.9	121,625 (18.09)	72,837 (18.18)	48,788 (17.96)	
35-39.9	69,022 (10.27)	40,038 (9.99)	28,984 (10.67)	
>40	95,819 (14.25)	53,813 (13.43)	42,006 (15.46)	
Emergency status	110,408/693,221 (15.93)	69,857 (16.84)	40,551 (14.56)	<0.001
Bleeding disorder	20,963/693,218 (3.02)	13,137 (3.17)	7,826 (2.81)	<0.001
Cardiac comorbidity	49,428/693,223 (7.13)	29,826 (7.19)	19,602 (7.04)	0.017
Diabetes	84,143/693,222 (12.14)	49,816 (12.01)	34,327 (12.33)	0.001
Hypertension	273,429/693,217 (39.44)	161,896 (39.03)	111,533 (40.06)	0.001
Pulmonary comorbidity	23,170/693,223 (3.34)	13,813 (3.33)	9,357 (3.36)	0.485
Renal comorbidity	7,578/693,223 (1.09)	5,194 (1.25)	2,384 (0.86)	0.001
Smoking	124,189/693,208 (17.92)	75,681 (18.25)	48,508 (17.42)	0.001
Steroid use	15,061/693,218 (2.17)	10,402 (2.51)	4,659 (1.67)	0.001
Stroke-related comorbidity	25,278/693,223 (3.65)	15,237 (3.67)	10,041 (3.61)	0.145
Year of operation				0.001
2006	59,816 (8.63)	37,627 (9.07)	22,189 (7.97)	
2007	107,617 (15.52)	63,424 (15.29)	44,193 (15.87)	
2008	127,745 (18.43)	75,707 (18.25)	52,038 (18.69)	
2009	148,847 (21.47)	82,116 (19.80)	66,731 (23.97)	
2010	151,131 (21.80)	82,394 (19.86)	68,737 (24.69)	
2011	66,655 (9.62)	43,635 (10.52)	23,020 (8.27)	
2012	31,412 (4.53)	29,901 (7.21)	1,511 (0.54)	

IQR, interquartile range; ASA, American Society of Anesthesiologists; BMI, body mass index; NS, not significant; p > 0.05.

TABLE 2. Procedure Frequency by Teaching Status

Procedure Group	(CPT Code)	Procedure Name	Total (%) n = 693,223	Teaching (%) n = 414,804	Nonteaching (%) n = 278,419
Appendectomy	44970	Laparoscopic appendectomy	75,201 (10.85)	46,166 (11.13)	29,035 (10.43)
	44950	Open appendectomy	13,607 (1.96)	8,127 (1.96)	5,480 (1.97)
Bariatric	43644	Laparoscopic gastric bypass	41,651 (6.01)	21,363 (5.15)	20,288 (7.29)
	43770	Laparoscopic adjustable gastric band	22,141 (3.19)	11,093 (2.67)	11,048 (3.97)
Bowel/Adhesions/ Ex-Lap	44140	Partial colectomy with anastomosis	24,223 (3.49)	14,960 (3.61)	9,263 (3.33)
	44204	Laparoscopic partial colectomy with anastomosis	18,486 (2.67)	10,850 (2.62)	7,636 (2.74)
	44120	Small Bowel Resection	18,037 (2.60)	12,893 (3.11)	5,144 (1.85)
	49000	Exploratory laparotomy	11,607 (1.67)	8,156 (1.97)	3,451 (1.24)
	44160	Partial colectomy with removal of terminal ileum and ileocolostomy	11,466 (1.65)	8,244 (1.99)	3,222 (1.16)
	44145	Partial colectomy with low pelvic anastomosis	9,651 (1.39)	6,947 (1.67)	2,704 (0.97)
	44005	Enterolysis (freeing of intestinal adhesion)	9,607 (1.39)	6,393 (1.54)	3,214 (1.15)
	44207	Laparoscopic partial colectomy with low pelvic anastomosis	8,115 (1.17)	5,562 (1.34)	2,553 (0.92)
Breast	19303	Simple mastectomy	23,824 (3.44)	14,171 (3.42)	9,653 (3.47)
	19301	Partial mastectomy	23,207 (3.35)	12,469 (3.01)	10,738 (3.86)
	19125	Excision of breast lesion with needle localization	22,747 (3.28)	11,397 (2.75)	11,350 (4.08)
	19120	Excision of cyst, fibroadenoma, or other benign or malignant tumor	19,655 (2.84)	8,774 (2.12)	10,881 (3.91)
	19307	Modified radical mastectomy including axillary lymph nodes	14,074 (2.03)	7,709 (1.86)	6,365 (2.29)
Cholecystectomy	47562	Laparoscopic cholecystectomy	84,452 (12.18)	52,417 (12.64)	32,035 (11.51)
	47563	Laparoscopic cholecystectomy with cholangiography	28,809 (4.16)	14,410 (3.47)	14,399 (5.17)
	47600	Open cholecystectomy	10,252 (1.48)	7,469 (1.80)	2,783 (1.00)
Foregut	43280	Laparoscopic esophagogastric fundoplasty (e. g., Nissen, Toupet)	11,450 (1.65)	7,551 (1.82)	3,899 (1.40)
Hernia	49505	Inguinal hernia repair, initial, reducible	51,037 (7.36)	30,173 (7.27)	20,864 (7.49)
	49560	Incisional or ventral hernia repair, initial, reducible	29,688 (4.28)	18,110 (4.37)	11,578 (4.16)
	49585	Umbilical hernia repair, reducible	25,171 (3.63)	13,737 (3.31)	11,434 (4.11)
	49650	Laparoscopic inguinal hernia repair, initial	16,321 (2.35)	8,920 (2.15)	7,401 (2.66)
	49587	Umbilical hernia repair, incarcerated or strangulated	11,096 (1.60)	5,513 (1.33)	5,583 (2.01)
	49561	Incisional or ventral hernia repair, initial, incarcerated or strangulated	10,659 (1.54)	6,087 (1.47)	4,572 (1.64)
Thyroid/ Parathyroid	60240	Total thyroidectomy	18,241 (2.63)	13,574 (3.27)	4,667 (1.68)
	60500	Parathyroidectomy	17,614 (2.54)	13,745 (3.31)	3,869 (1.39)
	60220	Thyroid lobectomy with or without isthmusectomy	11,134 (1.61)	7,824 (1.89)	3,310 (1.19)

CPT, Current Procedural Terminology.

TABLE 3. Average OR Times for Teaching vs. Nonteaching Cases, Unadjusted Results 2006-2012

Year	Teaching Cases Average Time, Minutes (95% CI)	Nonteaching cases Average Time, Minutes (95% CI)	Difference between Teaching and Nonteaching Cases Minutes (95% CI)
2006	95.85 (95.19-96.50)	66.79 (66.14-67.45)	29.06 (28.06-30.05)
2007	99.72 (99.17-100.28)	73.31 (72.78-73.83)	26.41 (25.62-27.21)
2008	99.01 (98.49-99.52)	75.51 (75.00-76.02)	23.50 (22.74-24.25)
2009	98.24 (97.74-98.74)	74.79 (74.33-75.24)	23.46 (22.77-24.15)
2010	98.84 (98.33-99.36)	73.88 (73.43-74.34)	24.96 (24.26-25.66)
2011	99.50 (98.81-100.20)	78.78 (77.90-79.65)	20.72 (19.57-21.87)
2012	101.67 (100.79-102.54)	87.09 (83.39-90.79)	14.57 (10.59-18.56)
Overall	**98.89 (98.67-99.11)**	**74.22 (74.00-74.45)**	**24.67 (24.34-24.99)**

CI, confidence interval.

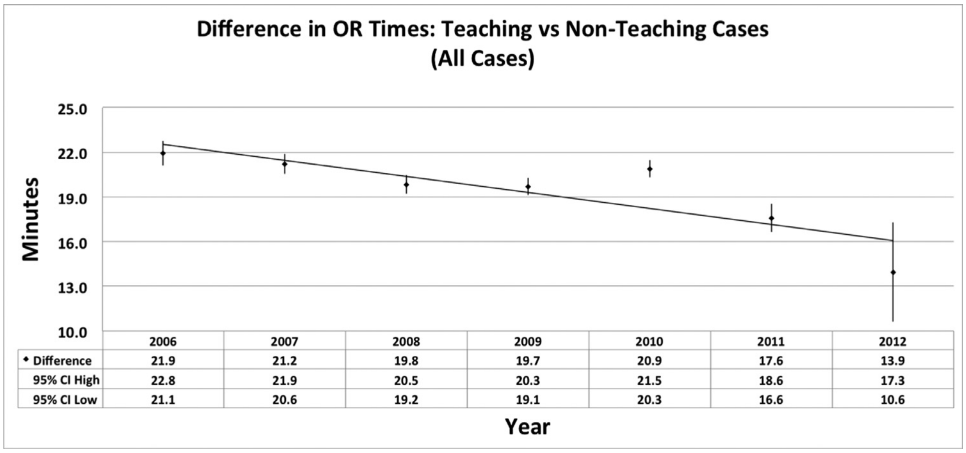

FIGURE 3. Annual difference in operative times between teaching and nonteaching cases, difference-in-difference analysis, multivariable results, 2006-2012[++]·Annual linear regression models were run independently to determine annual difference in OR times between all teaching and non-teaching cases, controlling for teaching group status; procedure type; emergency status; patient age, BMI, gender, ASA, diabetes, HTN, cardiac risk, pulmonary risk, stroke risk, bleeding risk, renal risk, steroid use, smoking. 2006: n = 57,708 observations; 2007: n = 103,978 observations; 2008: n = 123,487 observations; 2009: n = 143,971 observations; 2010: n = 145,262 observations; 2011: n = 64,357 observations; 2012: n = 30,031 observations

difference in OR times between teaching and nonteaching cases decreased by approximately 36%, or 8 minutes, per case. Our analyses suggest that these trends are unlikely to be attributable to differences in patient acuity and/or case-mix.

This study provides evidence of a previously unrecognized convergence in OR times between teaching and nonteaching cases, adding evidence to the ongoing conversation about the impact of systemic forces on trainee education. Whereas other studies have shown that teaching cases take longer than nonteaching cases[15-19, 23,24] and that increased resident autonomy prolongs operative times,[35] our trend analyses suggest that, from an operative times standpoint, teaching cases are increasingly resembling nonteaching cases. In subpopulation analyses, we found this convergence was statistically significant for PGY1, PGY2, and PGY5 groups with directionally similar results for PGY3 and PGY4s, and also across several procedure types, suggesting a systemic cause.

These findings have implications for trainees and surgical educators. In a healthcare system increasingly focused on maximizing productivity and efficiency,[2,21] ensuring patient safety, and controlling trainee work-hours,[5,22] trainees' educational experiences are under continuous pressure and risk being compromised.[1,36] Unfortunately, large prospective studies quantifying trends intraoperative resident teaching and the provision of autonomy are not available. Until such studies are performed, surrogate or indirect measures, such as those used in the present study, may help to estimate trends in intraoperative resident education. Our study provides an initial attempt at calculating the magnitude of change

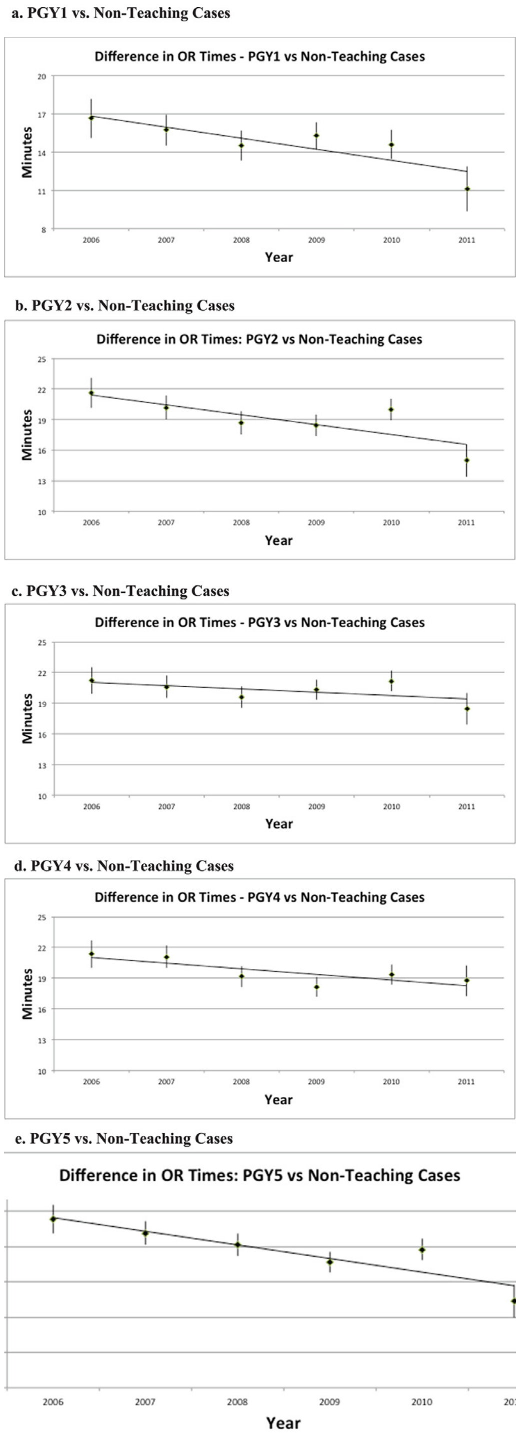

FIGURE 4a-e. Subpopulation difference-in-difference analysis: Difference in operative times between teaching and nonteaching cases across calendar years by PGY, Multivariable Results, 2006-2011. (a-e) Annual linear regression models were run independently for all procedures combined, controlling for PGY level (4a: PGY1, 4b: PGY2, 4c: PGY3, 4d: PGY4, 4e: PGY5). Each model included the following variables: OR time in minutes; Teaching group status; procedure type; emergency status; patient age, BMI, gender, ASA, diabetes, HTN, cardiac risk, pulmonary risk, stroke risk, bleeding risk, renal risk, steroid use, smoking.

over time using a large, nationally recognized, dataset with established evidence of validity (i.e., NSQIP). The magnitude of convergence observed here is noteworthy in large part because of the underlying trend it represents. While it may be difficult to establish an exact numeric threshold to determine when a statistically significant convergence might be considered clinically "meaningful," we believe that the double-digit percentage change deserves attention and, if left unchecked, has the potential to be clinically impactful, regardless of the underlying cause. We hope this effort provides both baseline data and a stimulus for future investigations aimed at better understanding intraoperative trends in resident education.

We initially hypothesized that OR times for teaching and nonteaching cases would be converging due to a relative decrease in the proportion of OR time devoted to teaching, over time. Given the many systemic forces that threaten the amount of time available for intraoperative resident education, our data offer indirect evidence in support of this hypothesis. However, several other potential explanations must be acknowledged. For example, these same findings could be seen if the "average" resident in recent years is more capable of operating quickly than the "average" resident in earlier years, after adjusting for differences in resident experience (e.g., PGY, case-mix), patient factors, attending surgeon factors, and other potential confounders. Indeed, the gradual integration of surgical simulation into residency programs may have accelerated surgical skills acquisition amongst trainees during the study period, enabling them to perform more expeditiously in the OR. Whether for simple tasks such as basic suturing and knot tying, or for more complex tasks such as construction of bowel anastomoses or performance of 2-handed laparoscopic techniques, surgical simulation allows trainees to enter the OR better prepared to assist or perform the operation. In one randomized controlled trial by Zendehas et al.,[37] residents who participated in an immersive simulation training program outperformed residents who did not participate in this program by 13 minutes per laparoscopic totally extraperitoneal inguinal hernia repair. In another study of outcomes associated with simulation-based learning by Griswold-Theodorson et al.,[38] the authors reported an overall trend toward decreased operative times in simulation (vs. nonsimulation) groups across 14 studies. Interestingly, one of the studies included in this latter paper, by Hogle et al.,[39] found that residents who participated in a simulation program took *longer* on average to perform laparoscopic cholecystectomy when compared to residents who did not participate in the simulation program. The authors did acknowledge several limitations that call into question the generalizability of these findings.

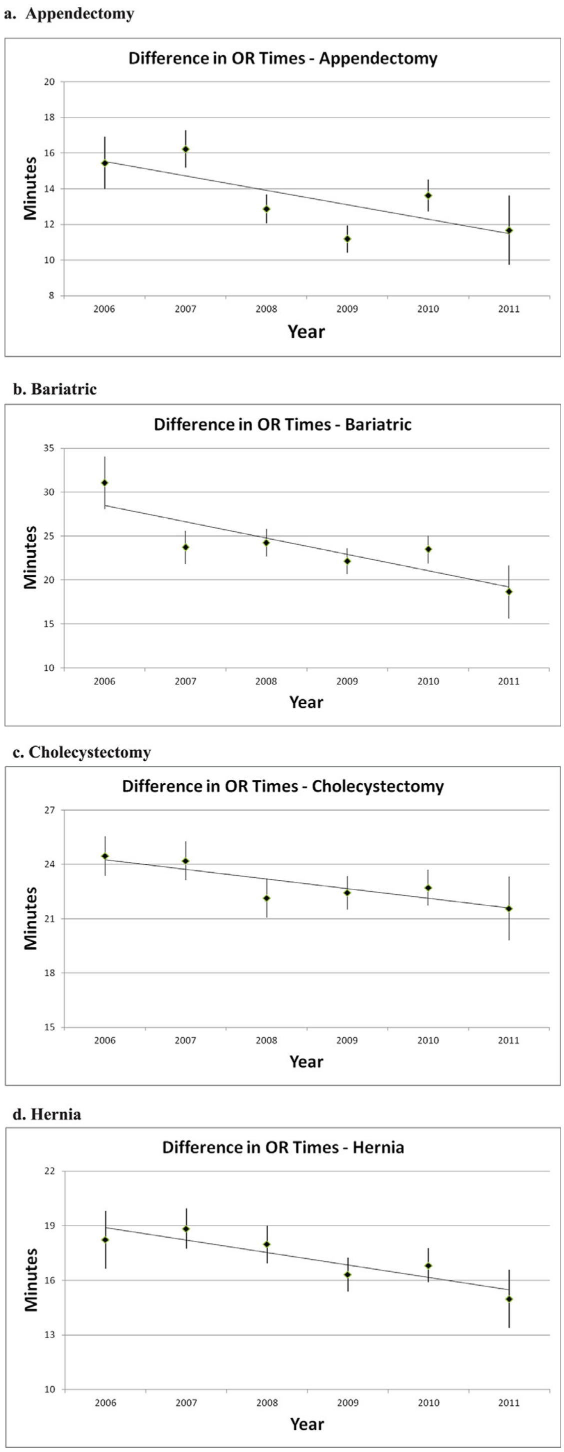

FIGURE 5a-d. Subpopulation difference-in-difference analysis: Difference in operative times between teaching and nonteaching cases across calendar years by procedure type, multivariable results, 2006-2011. (a-d) Annual linear regression models were run independently for all PGYs combined, controlling for procedure (a: Appendectomy, b: Bariatric, c: Cholecystectomy, d: Hernia). Each model included the following variables: OR time in minutes; Teaching group status; procedure type; emergency status; patient age, BMI, gender, ASA, diabetes, HTN, cardiac risk, pulmonary risk, stroke risk, bleeding risk, renal risk, steroid use, smoking.

In a related but alternative potential explanation for the convergence of OR times, increased emphasis on minimally invasive surgery during the study period may have led residents (adjusted for PGY-level) to be more proficient in minimally invasive procedures as the study progressed. In this case, residents in more recent years may have served as more capable and expeditious assistants or operating surgeons. A review of our data lends support to this potential explanation, finding that the majority (3 of 4) of procedure types in which a statistically significant convergence in OR times was observed – namely (1) appendectomy, (2) bariatric, and (3) cholecystectomy (Fig. 5a-d) are heavily weighted toward laparoscopic procedures.

The possibility of increased representation of fellows (i.e., >PGY5) in our database over time may also help to explain this convergence, since fellows are more likely to operate quickly than residents with less experience. We anticipate this effect, if present, to be small because fellow cases comprised only a small fraction of the teaching cases included in the study, and because the convergence was also observed in several of the PGY1 through PGY5 subpopulations.

Lastly, OR times may converge if attending surgeons in non-teaching cases are becoming disproportionately “slower” over time than attending surgeons in teaching cases. Interestingly, our results do demonstrate a trend towards longer procedure times during the study period in *both* the teaching and non-teaching groups, in unadjusted analyses. Yet, these increases were observed only in unadjusted analyses and should be interpreted with caution. This study was not designed to quantify or to explain secular trends in OR times; instead, it was designed to control for them. Factors that may contribute to a progressive prolongation of operative times among attendings include introduction of new technologies (e.g., minimally invasive tools including the operative robot); prioritization of intraoperative patient safety policies; differences in surgeon-mix over time; increasing case-complexity; and evolving case mix; among others. However, these factors would generally be expected to impact teaching and nonteaching cases similarly. We have otherwise controlled for differences in case complexity and case mix in our multivariable regression models.

Future studies should attempt to better understand the source of convergence of OR times between teaching and nonteaching cases. While NSQIP is a widely used and readily available database, it is not ideally suited, at least not alone, to determine the cause of convergence. NSQIP does not explicitly focus on education-related variables, which often require direct observation. The deidentified NSQIP database also does not allow us to identify individual hospitals, and it is therefore challenging to correlate our findings with any changes in simulation practices or other systems changes at participating hospitals during the study period. An ideal dataset would longitudinally collect intraoperative education-related variables for each case (for example, PGY of trainee, resident vs. fellow, quantity/quality of intraoperative teaching, trainee autonomy achieved, simulation experience of the trainee on that specific procedure and in general) and link these data to more traditional surgical outcomes data (such as those in NSQIP) to better assess the aggregate impact of education-related *and* other (i.e., patient-related, provider-related, team-related, system-related) variables on outcomes of interest, in this case operative times.

This study has several limitations. First, it is a retrospective analysis and subject to potential bias from unmeasured confounders. Second, the deidentified national NSQIP database does not include hospital level data and we, therefore, cannot control for changes in the distribution of teaching vs. nonteaching cases in individual hospitals over time. This limitation was the motivation for using a pseudo-randomization technique such as difference in difference, which examines differences over time and in effect uses one population as the control for the other. We believe this limitation is further minimized by the fact that the OR times used in this study include only "surgical time" (incision to close) and not "nonsurgical times" (e.g., operating room turnover time, anesthesia time). Third, this study was not designed to explain the cause of convergence in OR times between teaching and nonteaching cases. Several potential explanations have been proposed, and prospectively collected data are likely needed with attention to the various potential explanations put forth above to more directly and definitively explain our findings. Fourth, the PGY variable was retired from the NSQIP database in 2013,[40] and nonteaching case data from the 2012 year are limited. We, therefore, restricted subpopulation analyses to a 6-year cohort (2006-2011) to ensure adequate sample size in each year. This was a conservative approach, and excluding 2012 data from subpopulation analyses likely led us to underestimate the magnitude of trend across many of the subgroups studied.

CONCLUSIONS

OR times for teaching and nonteaching cases converged by approximately 8 minutes per General Surgery procedure during the seven-year study period, representing a 36% reduction in the difference between groups. We must seek to better understand the source of this convergence while preserving and enhancing the intraoperative learning experience for surgical trainees into the future.

TYPE OF STUDY

Retrospective, Longitudinal analysis

REFERENCES

1. George BC, Bohnen JD, Williams RG, et al. Readiness of US general surgery residents for independent practice. *Ann Surg*. 2017;266:582–594.
2. Sandhu G, Teman NR, Minter RM. Training autonomous surgeons: more time or faculty development? *Ann Surg*. 2015;261:843–845.
3. Dauphinee WD. Self regulation must be made to work. *BMJ*. 2005;330:1385–1387.
4. Halpern SD, Detsky AS. Graded autonomy in medical education—managing things that go bump in the night. *N Engl J Med*. 2014;370:1086–1089.
5. Ten Cate O, Hart D, Ankel F, et al. Entrustment Decision Making in Clinical Training. *Acad Med*. 2016;91:191–198.
6. Hashimoto DA, 4th Bynum WE, Lillemoe KD, Sachdeva AK. See more, do more, teach more: surgical resident autonomy and the transition to independent practice. *Acad Med*. 2016;91:757–760.
7. George BC, Dunnington GL, DaRosa DA. Trainee Autonomy and Patient Safety. *Ann Surg*. 2018;267:820–822.
8. Fallon WF Jr, Wears RL, Tepas JJ 3rd. Resident supervision in the operating room: does this impact on outcome? *J Trauma*. 1993;35:556–560. discussion 560-1.
9. Gawande AA, Zinner MJ, Studdert DM, Brennan TA. Analysis of errors reported by surgeons at three teaching hospitals. *Surgery*. 2003;133:614–621.
10. Wojcik BM, Fong ZV, Patel MS, et al. The resident-run minor surgery clinic: a pilot study to safely increase operative autonomy. *J Surg Educ*. 2016;73:e142–e149.
11. Flynn T. Resident supervision. *ACGME Bulletin*. 2005: 15–17.
12. Regenbogen SE, Greenberg CC, Studdert DM, Lipsitz SR, Zinner MJ, Gawande AA. Patterns of technical error among surgical malpractice claims: an analysis of strategies to prevent injury to surgical patients. *Ann Surg*. 2007 Nov;246:705–711.
13. Singh H, Thomas EJ, Petersen LA, Studdert DM. Medical errors involving trainees: a study of closed

malpractice claims from 5 insurers. *Arch Intern Med*. 2007;167:2030–2036.

14. Nagpal K, Vats A, Lamb B, et al. Information transfer and communication in surgery: a systematic review. *Ann Surg*. 2010;252:225–239.

15. Advani V, Ahad S, Gonczy C, Markwell S, Hassan I. Does resident involvement effect surgical times and complication rates during laparoscopic appendectomy for uncomplicated appendicitis? An analysis of 16,849 cases from the ACS-NSQIP. *Am J Surg*. 2012;203:347–351. discussion 351-2.

16. Kiran RP, Ahmed Ali U, Coffey JC, Vogel JD, Pokala N, Fazio VW. Impact of resident participation in surgical operations on postoperative outcomes: National Surgical Quality Improvement Program. *Ann Surg*. 2012;256:469–475.

17. Scarborough JE, Bennett KM, Pappas TN. Defining the impact of resident participation on outcomes after appendectomy. *Ann Surg*. 2012;255:577–582.

18. Iannuzzi JC, Rickles AS, Deeb AP, Sharma A, Fleming FJ, Monson JR. Outcomes associated with resident involvement in partial colectomy. *Dis Colon Rectum*. 2013;56:212–218.

19. Kasotakis G, Lakha A, Sarkar B, et al. Trainee participation is associated with adverse outcomes in emergency general surgery: an analysis of the National Surgical Quality Improvement Program database. *Ann Surg*. 2014;260:483–490. discussion 490-3.

20. Kachalia A, Studdert DM. Professional liability issues in graduate medical education. *JAMA*. 2004;292:1051–1056.

21. Teman NR, Gauger PG, Mullan PB, Tarpley JL, Minter RM. Entrustment of general surgery residents in the operating room: factors contributing to provision of resident autonomy. *J Am Coll Surg*. 2014;219:778–787.

22. Ahmed N, Devitt KS, Keshet I, et al. A systematic review of the effects of resident duty hour restrictions in surgery: impact on resident wellness, training, and patient outcomes. *Ann Surg*. 2014;259:1041–1053.

23. Papandria D, Rhee D, Ortega G, et al. Assessing trainee impact on operative time for common general surgical procedures in ACS-NSQIP. *J Surg Educ*. 2012;69:149–155.

24. Vinden C, Malthaner R, McGee J, et al. Teaching surgery takes time: the impact of surgical education on time in the operating room. *Can J Surg*. 2016;59:87–92.

25. Kairys JC, McGuire K, Crawford AG, et al. Cumulative operative experience is decreasing during General Surgery residency: a worrisome trend for surgical trainees? *J Am Coll Surg*. 2008;206:804–813.

26. Watson DR, Flesher TD, Ruiz O, Chung JS. Impact of the 80-hour workweek on surgical case exposure within a general surgery residency program. *J Surg Educ*. 2010;67:283–289.

27. Malangoni MA, Biester TW, Jones AT, Klingensmith ME, Jr Lewis FR. Operative experience of surgery residents: trends and challenges. *J Surg Educ*. 2013;70:783–788.

28. Cortez AR, Winer LK, Kassam AF, et al. See none, do some, teach none: an analysis of the contemporary operative experience as nonprimary surgeon. *J Surg Educ*. 2019;7:e92–e101.

29. Bohnen JD, George BC, Williams RG, et al. The feasibility of real-time intraoperative performance assessment with SIMPL (System for Improving and Measuring Procedural Learning): early experience from a multi-institutional trial. *J Surg Educ*. 2016;73:e118–e130.

30. 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;204:1089–1102.

31. ACS NSQIP Participant Use Data File, Accessed 22 April 2020. Available at: https://www.facs.org/quality-programs/acs-nsqip/participant-use

32. ACS NSQIP, User guide for the 2012 ACS NSQIP Participant Use Data File. 2012.

33. Dimick JB, Ryan AM. Methods for evaluating changes in health care policy: the difference-in-differences approach. *JAMA*. 2014;312:2401–2402.

34. Zhou H, Taber C, Arcona S, Li Y. Difference-in-differences method in comparative effectiveness research: utility with unbalanced groups. *Appl Health Econ Health Policy*. 2016;14:419–429.

35. Siam B, Al-Kurd A, Simanovsky N, et al. Comparison of Appendectomy Outcomes Between Senior General Surgeons and General Surgery Residents. *JAMA Surg*. 2017;152:679–685.

36. Mattar SG, Alseidi AA, Jones DB, et al. General surgery residency inadequately prepares trainees for fellowship: results of a survey of fellowship program directors. *Ann Surg*. 2013;258:440–449.

37. Zendejas B, Cook DA, Bingener J, et al. Simulation-based mastery learning improves patient outcomes in laparoscopic inguinal hernia repair: a randomized controlled trial. *Ann Surg*. 2011;254:502–511.

38. Griswold-Theodorson S, Ponnuru S, Dong C, et al. Beyond the simulation laboratory: a realist synthesis review of clinical outcomes of simulation-based mastery learning. *Acad Med.* 2015;90:1553–1560.

39. Hogle NJ, Chang L, Strong VE, et al. Validation of laparoscopic surgical skills training outside the operating room: a long road. *Surg Endosc*. 2009;23:1476–1482.

40. ACS NSQIP, User guide for the 2014 ACS NSQIP participant use data file. 2014.