Annals of Surgery Open

OPEN

Patient Surgical Outcomes When Surgery Residents Are the Primary Surgeon by Intensity of Surgical Attending Supervision in Veterans Affairs Medical Centers

T. Michael Kashner, PhD, JD, MPH,*† Paul B. Greenberg, MD, MPH,‡§ Andrea D. Birnbaum, MD, PhD,*‖ John M. Byrne, DO,*† Karen M. Sanders, MD,*¶ Mark A. Wilson, MD, PhD,# and Marjorie A. Bowman, MD, MPA*,**††§§

Objective: Using health records from the Department of Veterans Affairs (VA), the largest healthcare training platform in the United States, we estimated independent associations between the intensity of attending supervision of surgical residents and 30-day postoperation patient outcomes.

Background: Academic leaders do not agree on the level of autonomy from supervision to grant surgery residents to best prepare them to enter independent practice without risking patient outcomes.

Methods: Secondary data came from a national, systematic 1:8 sample of n = 862,425 teaching encounters where residents were listed as primary surgeon at 122 VA medical centers from July 1, 2004, through September 30, 2019. Independent associations between whether attendings had scrubbed or not scrubbed on patient 30-day all-cause mortality, complications, and 30-day readmission were estimated using generalized linear-mixed models. Estimates were tested for any residual confounding biases, robustness to different regression models, stability over time, and validated using moderator and secondary factors analyses.

Results: After accounting for potential confounding factors, residents supervised by scrubbed attendings in 733,997 nonemergency surgery encounters had fewer deaths within 30 days of the operation by 14.2% [0.3%, 29.9%], fewer case complications by 7.9% [2.0%, 14.0%], and fewer readmissions by 17.5% [11.2%, 24.2%] than had attendings not scrubbed. Over the 15 study years, scrubbed surgery attendings may have averted an estimated 13,700 deaths, 43,600 cases with complications, and 73,800 readmissions.

Conclusions: VA policies on attending surgeon supervision have protected patient safety while allowing residents in selected teaching encounters to have limited autonomy from supervision.

Keywords: Graduate Medical Education in Surgery, Attending Surgeon Supervision of Surgical Residents, Outcomes of Teaching Surgery Encounters, Independent Associations

Supervision is central to GME's apprenticeship educational model where residents progress toward higher levels of autonomy from direct attending supervision, as they gain clinical competencies.[1–4] For surgical specialties, however, the advent of maximum hours of educational activities per week,[5] increasingly complex procedures, and more intensive supervision with less autonomy have caused concerns among academic leaders that graduating surgery residents may not be adequately trained to enter independent practice.[6–9] To address these concerns, the American Board of Surgery supports the explicit use and tracking of Entrustable Professional Activities, where the mastery of knowledge and skills for a given clinical activity is matched to progressive levels of resident supervision: "observation only," "performance under direct supervision," "indirect supervision," "independent practice" and "supervising others."[10]

From the *Office of Academic Affiliations, Department of Veterans Affairs, Washington, DC; †Department of Medicine, Loma Linda University Medical School, Loma Linda, CA; ‡VA Providence Healthcare System, Providence, RI; §Department of Surgery (Ophthalmology), The Warren Alpert Medical School of Brown University, Providence, RI; ‖Department of Ophthalmology, Northwestern University Feinberg School of Medicine, Chicago, IL; ¶Department of Medicine, Virginia Commonwealth University School of Medicine, Richmond, VA; #Department of Veterans Affairs, National Director of Surgery, National Office of Surgery (11SURG), Washington, DC; **Department of Family Medicine and Community Health, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; ††Department of Family Medicine, Wright State University Boonshoft School of Medicine, Dayton, OH; and §§ Chief Academic Affiliations Officer, Department of Veterans Affairs, Washington, DC..

This work is supported by Department of Veterans Affairs' Office of Academic Affiliations (self- funded).

All statements and descriptions expressed herein do not necessarily reflect the opinions or positions of the Department of Veterans Affairs or its affiliated institutions.

IRB Status: The study was exempt under 45 C.F.R. §§ 46.104(d)(4)(ii) & 46.101(c).

The authors declare that they have nothing to disclose.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal`s Web site (www.annalsofsurgery.com).

Reprints: T. Michael Kashner, PhD, JD, MPH, Department of Veterans Affairs, 810 Vermont Ave., NW., Washington, DC 20420. E-mail: michael.kashner@va.gov.

Annals of Surgery Open (2023) 4:e351

Received: 2 September 2023; Accepted 25 September 2023

Published online 14 November 2023

DOI: 10.1097/AS9.0000000000000351

The empirical question is how to supervise residents so they may progress toward independence from direct supervision without adversely impacting patient safety or quality of care. This question is important to patients who want safe and effective care and residents who need patient care opportunities to learn proper technique, evidence-based protocols, sound clinical judgement, and professionalism.[1–3] To address this question, academic leaders often cite studies reporting patients receive better medical[11,12] and surgical care[13,14] in teaching than nonteaching clinical settings. Surgery encounters involving a scrubbed resident have been found to have no worse outcome than those that either did not involve a resident or the resident was not scrubbed with respect to mortality,[8,9,15–18] complications,[8,9,15,19–22] readmissions and reoperations,[15,17–19,23] patient satisfaction,[21] or pain scores.[22]

Comparing the outcomes of teaching with nonteaching encounters offer little guidance on how intensive supervision should be once a resident is involved in an operation. Two recent Department of Veterans Affairs [VA] studies estimated the effect of the intensity of resident supervision on patient outcomes using data from the VA Surgical Quality Improvement Program [VASQIP].[24] Intensity was measured based on whether the surgical attending was scrubbed or not scrubbed. A surgical attending not scrubbed for the operation implies a lower level of supervision when they are in the room supervising, but unavailable to do any of the surgery directly with immediacy. In the first study, Oliver et al[8] used VASQIP encounter data between July 2004 through September 2019 and propensity score matched 137,749 from among 138,750 VASQIP encounters where the attending did not scrub [ANS] to 137,749 of the 871,546 encounters where the attending had scrubbed [AS]. The AS match served to estimate the counterfactual outcome for each ANS encounter. In a similar study, Tonelli et al[25] used VASQIP data and propensity score matched 11,181 ANS core general surgery procedure cases between January 2005 through December 2021 to 11,181 of 98,526 AS encounters. In both studies, outcomes between ANS and their respective matched AS encounter were not statistically different with respect to all-cause 30-day mortality or complication rates. Oliver et al also reported no significant difference in reoperation rates. As noted in Terhune recent commentary,[26] the lack of association between supervision intensity and patient outcomes may suggest attending surgeons were making the correct choices when not to scrub.

The two VA studies estimated how patient outcomes would have changed had not scrubbed attendings in ANS encounters scrubbed. They offer no information about how outcomes would change had scrubbed attendings in AS encounters not scrubbed. The latter is important to academic leaders deliberating over how to further progress surgery resident autonomy without adversely impacting patient outcomes.

In this paper, we address the unresolved question using an equivalent VASQIP combined ANS and AS subsamples of surgery teaching encounters. Our purpose is to compute the impact on 30-day all-cause mortality, and secondarily on complications and readmission rates, had scrubbed attendings in AS encounters not scrubbed.

METHODS

To answer our research question using VASQIP data is difficult because there are too few ANS encounters to serve as a counterfactual match to each AS encounter. We therefore applied control functions, a data-analytic method[27–29] outlined in the Supplement, see http://links.lww.com/AOSO/A267, where teaching encounter outcomes are regressed on attending scrubbed status and other covariates on the combined AS and ANS VASQIP sample. The strategy is to select covariates so that the estimated association between attending's scrubbed status and patient outcomes is independent of all confounding biases.

The authors are federal employees with approval from their respective agency heads to conduct these secondary analyses on VA collected and deidentified datasets. The study is exempt from institutional review per 45 C.F.R. §§ 46.104(d)(4)(ii) & 46.101(c).

Study Setting, Sample, and Data Sources

The study setting is the surgery services at VA teaching medical centers. In the United States, VA is the second largest funder of GME behind the Centers for Medicare and Medicaid Services.[30] VA hosted 10,071 surgery residents from among 47,521 physician residents and fellows in 2019 to 2020.[31] VA's clinical delivery system is under consistent management and uniform policies governing resident supervision. Patients do not face financial access barriers including reimbursement for travel expenses.

Data came from the Veterans Affairs Surgical Quality Improvement Program[24] [VASQIP] files providing information about inpatient and outpatient surgery encounters plus 30-day follow-ups. VASQIP surgery encounters are systematically sampled with an 8-day cycle. A template for the American College of Surgeons' National Surgical Quality Improvement Program, VASQIP encounter data contains 200 variables including 78 manually abstracted fields by local, trained nurses who report on patient demographic, health status, surgery procedures, risk factors, resident status of primary surgeon, and surgery outcomes. Additional data on program size came from VA's Office of Academic Affiliations.

The pre-COVID19 sample included all VASQIP teaching encounters beginning July 1, 2004, through September 30, 2019, excluding unplanned emergency surgeries and deceased patients whose organs were being removed for donation or cases with missing data. A VASQIP surgery is considered a "teaching" encounter whenever an appointed Operating Room nurse classifies a scrubbed resident as the primary surgeon.

Table 1 lists VASQIP and VA study variables. The binary resident supervision factor [OSup] describes the surgical attending as either scrubbed [AS] or not scrubbed [ANS]. AS means the attendings were performing the operation or otherwise in the operating room scrubbed. ANS means the attendings were in the operating room but not scrubbed or were not in the operating room but immediately available in the suite.

For this study, three measures were computed from VASQIP encounter-level common procedure terminology [CPT] procedures and visit codes that were reported in the health record and valued by a relative value unit [rvu] from the National Physician Fee Schedule Relative Value Files.[32] An encounter-level surgery complexity score [SRVU] was computed by summing rvu's of all procedures assigned to the encounter. Higher scores indicate more complex surgeries. Surgery Teaching Intensity Rates [FIntensity] quantified residents' participation in VA surgery care by specialty, facility, and academic year, and calculated as the sum of rvu's in teaching encounters divided by total rvu's of all surgery encounters in the teaching facility, but only for specialties where the facility reported at least one teaching encounter in that specialty during the academic year. Surgery Procedure Variety Rates [FVariety] quantified the extent residents were exposed to different surgery procedures computed by specialty, facility, and academic year, and calculated as a coefficient of variation in procedure rvu.

VA's 5-point, ordinal scale measures the complexity of care at the medical center [FComplex] based on care activities, services offered, size and diversity of health professions training programs, and research activities.[33] VA's Office of Academic Affiliations' provided counts of funded residency positions by

TABLE 1.

Description of Covariates and Outcome Variables by Attending Physician Subsamples

	Scrubbed		Not Scrubbed	
	n/ value	%/ (SD)	n/ value	%/ (SD)
Total sample size	733,997	100.0%	128,428	100.0%
Number of medical centers*	121		91	
Race—Black or African American	151,555	20.6%	26,773	20.8%
Race—White	568,117	77.4%	99,376	77.4%
Race—Other	14.325	2.0%	2,279	1.8%
Gender—female	49,771	6.8%	6,987	5.4%
Age (SD)				
American Society Anesthesiology Physical Status Classification				
Healthy patient	13,397	1.8%	2,834	2.2%
Mild systemic disease	168,125	22.9%	32,163	25.0%
Severe systemic disease	471,891	64.3%	80,720	62.9%
Constant threat to life	78,632	10.7%	12,485	9.7%
Moribund patient	1,952	0.3%	226	0.2%
Brain dead patient, organs removed	0	0.0%	0	0.0%
Smoking past year, self-report				
Yes	247,932	33.8%	42,351	33.0%
No	486,065	66.2%	86,007	67.0%
Alcohol consumption, self-report				
2 oz. or more per day past year	54,636	7.5%	9,017	7.0%
Otherwise	678,920	92.4%	119,349	92.9%
missing	441	0.1%	65	0.1%
Activities daily living 30 days prior				
Dependent	14,070	1.9%	2,580	2.0%
Partially dependent	47,811	6.5%	10,496	8.2%
Independent	671,948	91.6%	115,324	89.8%
Missing	168	0.0%	28	0.0%
Hospital length of stay				
Before surgery mean days (SD)†	2.26	(36.48)	2.14	(30.88)
After surgery mean days (SD)†	4.54	(10.10)	3.26	(9.62)
Surgery length of stay				
Before surgery mean days (SD)†	0.56	(6.66)	0.46	(3.62)
After surgery mean days (SD)†	4.07	(8.25)	2.85	(7.04)
Complexity surgery procedure‡	19.01 rvu	(13.19)	13.62 rvu	(11.31)
Number surgeries by specialty				
General surgery	250,946	34.2%	28,300	22.0%
Gynecology	4,689	0.6%	239	0.2%
Neuro surgery	45,875	6.3%	4,952	3.9%
Ophthalmology	490	0.1%	59	0.0%
Orthopedic surgery	158,381	21.6%	42,063	32.8%
Otolaryngology	32,985	4.5%	6,275	4.9%
Plastic surgery	18,225	2.5%	2,124	1.7%
Thoracic surgery	22,273	3.0%	878	0.7%
Urology	97,537	13.3%	35,556	27.7%
Oral surgery	2,740	0.4%	271	0.2%
Podiatry	8,498	1.2%	465	0.4%
Peripheral vascular surgery	89,278	12.2%	7,130	5.6%
Other	2,110	0.3%	116	0.1%
Number surgeries by year (calendar year)§				
2004 (July 1–December 31)	16,534	2.3%	3,883	3.0%
2005	43,421	5.9%	10,069	7.8%
2006	44,252	6.0%	10,060	7.8%
2007	47,381	6.5%	9,969	7.8%
2008	50,539	6.9%	9,592	7.5%
2009	52,620	7.2%	9,086	7.1%
2010	53,220	7.3%	8,662	6.7%
2011	54,746	7.5%	8,296	6.5%
2012	54,911	7.5%	8,402	6.5%
2013	51,407	7.0%	7,601	5.9%
2014	51,584	7.0%	7,249	5.6%
2015	49,445	6.7%	7,259	5.7%

(Continued)

TABLE 1.
(Continued)

	Scrubbed		Not Scrubbed	
	n/ value	%/ (SD)	n/ value	%/ (SD)
2016	41,233	5.6%	13,369	10.4%
2017	44,754	6.1%	5,825	4.5%
2018	44,209	6.0%	5,383	4.2%
2019 (January 1–September 30)	33,741	4.6%	3,723	2.9%
Facility complexity classification‖				
1a-High	425,077	57.9%	72,131	56.2%
1b-High	169,510	23.1%	35,196	27.4%
1c-High	124,804	17.0%	19,992	15.6%
2-Medium	14,597	2.0%	1,109	0.9%
3-Low	9	0.0%	0	0.0%
Surgery teaching intensity rate‖	0.789	(0.181)	0.813	(0.172)
Procedure variety‖	0.691	(0.073)	0.692	(0.073)
Funded residency programs‖	32.1	(11.5)	32.3	(9.6)
All residency positions PGY1-2‖	49.6	(21.8)	49.5	(20.3)
All residency positions PGY3-4‖	53.6	(23.5)	54.3	(21.5)
All residency positions PGY5‖	15.9	(8.7)	16.0	(7.8)
All residency positions PGY6+‖	7.0	(5.2)	7.1	(5.2)
Surgery residency positions PGY1-2‖	5.9	(3.7)	6.1	(3.7)
Surgery residency positions PGY3-4‖	5.7	(3.2)	6.0	(3.2)
Surgery residency positions PGY5‖	3.5	(2.4)	3.7	(2.5)
Surgery residency positions PGY6+‖	1.9	(2.1)	2.0	(2.1)
Number residents PGY 1	36,645	5.0%	2,987	2.3%
Number residents PGY 2	81,072	11.0%	10,658	8.3%
Number residents PGY 3	131,517	17.9%	25,398	19.8%
Number residents PGY 4	125,520	17.1%	20,494	16.0%
Number residents PGY 5	275,725	37.6%	55,467	43.2%
Number residents PGY 6	50,657	6.9%	9,439	7.3%
Number residents PGY 7	28,059	3.8%	3,367	2.6%
Number residents PGY 8+	4,802	0.7%	618	0.5%
Number residents PGY 4 or lower	374,751	51.1%	59,537	46.4%
Number residents PGY 5 or higher	359,243	48.9%	68,891	53.6%
Supervision level-attending surgeon				
In O.R. doing operation (A)	34,218	4.7%	0	0.0%
In O.R. scrubbed (B)	699,779	95.3%	0	0.0%
In O.R. not scrubbed (C)	0	0.0%	116,444	90.7%
In suite and available (D)	0	0.0%	11,984	9.3%
Contacted, emergency service (E)	0	0.0%	0	0.0%
Scrubbed (levels ‘A’ or ‘B’)—AS	733,997	100.0%	0	0.0%
Not Scrubbed (levels “C” or “D”)—ANS	0	0.0%	128,428	100.0%
30-day all-cause mortality outcome				
Death¶	10,670	1.5%	1,581	1.2%
Alive	723,327	98.5%	126,847	98.8%
30-day case complications				
Yes#	66,443	9.1%	8,275	6.4%
No	667,554	90.9%	120,153	93.6%
30-day readmission rate				
Yes**	49,768	6.8%	8,107	6.3%
No	684,229	93.2%	120,321	93.7%

Unless otherwise noted, all data came from US Department of Veterans Affairs, Veterans Affairs Surgery Quality Improvement Program (VASQIP). Means and standard deviation estimates describe positive continuous value variables. Frequencies describe categorial, ordinal, and binary variables.

*Teaching encounters in the final sample came from a total 122 medical centers.

†Outpatient surgery is assigned zero days stayed leading to a bi-modal distribution.

‡Encounter-level complexity surgery procedure and facility-level surgery teaching intensity rates and procedure variety rates were computed from VASQIP data based on rvu's assigned to procedures listed in the patient's health record for the surgery encounter.

§The reference centered year of surgery is a negative integer variable computed by subtracting year of surgery by the reference year 2019: $\left[\widetilde{\text{SYear}}\right]$ = [SYear]–2019.

‖Based on data supplied by the Department of Veterans Affairs and VA Office of Academic Affiliations.

¶Unadjusted difference in mortality: OR% = –25.4%, χ (1) = 38.7, $P < 0.001$.

#Unadjusted difference in case complication rates: OR% = –46.4%, χ (1) = 940.2, $P < 0.001$.

**Unadjusted difference in 30-day readmission rates: OR% = –8.5%, χ (1) = 38.2, $P < 0.001$.

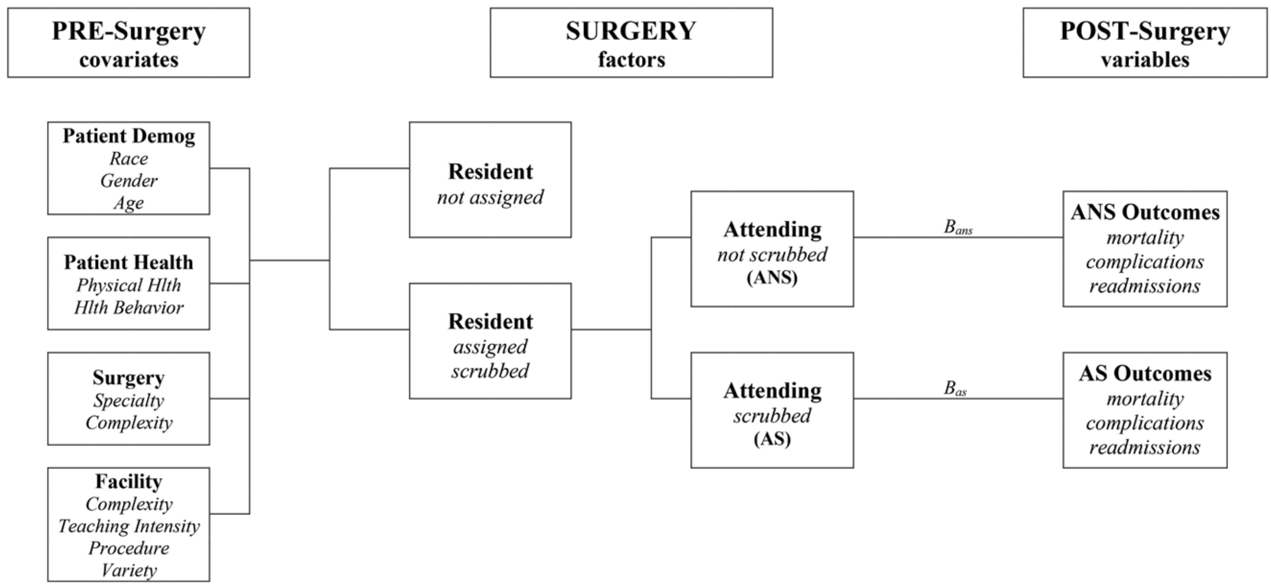

FIGURE 1. Graduate Medical Education—Surgery Model explaining relationships between resident supervision and patient outcomes.

surgery specialty, accredited program, resident PGY level, facility, and fiscal year.

Analyses

Our research question is to determine how outcomes change had attending surgeons in AS encounters not scrubbed. Using control functions outlined in the Supplement, see http://links.lww.com/AOSO/A267, we regressed encounter outcomes on whether the attending scrubbed, plus the resident's PGY level and other covariates using a generalized linear-mixed model on the combined AS and ANS sample (Table 1). Covariates were selected so that estimates of associations between attending scrubbed status and patient outcomes are independent of all confounding variables including those not otherwise observed in the data. To identify appropriate covariates, Figure 1 describes the sequence of events during a surgery encounter beginning with (1) presurgery when the patient arrives at the hospital and patient demographic, initial health status, surgery type, and facility characteristics are determined, (2) surgery when a resident at a given PGY level serves as the primary surgeon and the attending supervises the resident by either scrubbing at some point before or during the surgery or remaining not scrubbed throughout the surgery, and (3) postsurgery for 30 days after the operation when surgery outcomes are determined. Per Figure 1, B_{ans} is the estimated change in outcome had attendings in ANS encounters scrubbed. Oliver[8] and Toneli[25] results suggest B_{ans} is zero. We estimated B_{as} as the change in outcomes had attendings in AS encounters not scrubbed.

Once the control regression model is estimated, we evaluated the estimated association between the attending's scrubbed status and patient outcomes for: (i) independence (factor conditional exogeneity) to determine if the included covariates had controlled for the variance of all confounders (α = 0.10), (ii) robustness to determine if estimates varied across different specifications of the core model (α = 0.10), (iii) time stability to determine whether estimates are stable over time and thus be relevant for future policy decisions, and (iv) validation to determine if estimates varied by surgery specialty and facility complexity. Association estimates are time stable if they are time invariant (α = 0.05), or if positive estimates increased with time or negative estimates decreased with time. To account for the long observation period, association estimates were based on 2019 as the referent year.

RESULTS

Study Sample

The VASQIP sample of 1,737,662 surgery encounters from 144 VA medical centers were reduced by 388,656 (22.4%) encounters in nonteaching settings, 242 (0.0%) emergency surgery cases, 121 (0.0%) with missing facility identifier, and 176,148 with missing values from covariates in the core regression model, leaving 1,172,495 usable encounters. Further analysis revealed most missing cases were due to incomplete patient demographic information [174,690 (99.1%) of 176,148]. Compared with included encounters, missing cases had lower mortality at 2205 (1.3%) of 176,148 missing versus 15,986 (1.4%) of 1,172,495 nonmissing cases (χ (1) = 14.3, P < 0.001), fewer complications at 13,275 (7.5%) missing versus 111,106 (8.2%) nonmissing cases (χ (1) = 132.1, P < 0.001), fewer readmissions at 9752 (5.5%) missing versus 76,477 (6.5%) nonmissing cases (χ (1) = 248.9, P < 0.001), and better ASA physical rating of "mild systemic disease" or "healthy patient" in 57,708 (33.1%) of 174,477 missing versus 300,201 (25.6%) of 1,172,495 nonmissing cases (χ (4) = 4988.1, P < 0.001). Usable cases were further reduced by 310,070 (73.6%) nonteaching encounters for a final sample of 862,425 teaching encounters from 122 medical centers.

Table 1 describes both surgical attendings scrubbed (AS, 733,997 (85.1%) of 862,425) and not scrubbed (ANS, 128,428 (14.9%) of 862,425) subsamples. Differences between AS and ANS were numerically negligible, with ANS encounters having slightly shorter hospital stays, less complex surgery procedure, patients in better health, residents with more experience, and patients with statistically significantly better surgery outcomes.

Primary and Secondary Outcome Associations

From Table 2, attendings scrubbed status in the combined AS and ANS sample were independently associated with decreases in 30-day mortality by 14.2% [0.3%, 29.9%], case complications by 7.9% [2.0%, 14.0%], and readmissions by 17.5% [11.2%, 24.2%]. From these estimates, we can infer that for every 1000 surgery patients from among the 733,997 AS encounters, had the scrubbed attending in fact not scrubbed an estimated 2.03 [0.04, 4.26] deaths would have been added to the reported 14.54 deaths per 1000 patients, 6.46 [1.64, 11.38] cases with complications would have been added to the

TABLE 2.

Associations Between Resident Supervision, Senior Resident Status, and Patient Outcomes

				Test Criteria*	
	OR%	**95% CI**	**±**	**Method**	**Statistic**
Resident supervision†‡					
Mortality	−14.2%	[−29.9, −0.3]	+	Sampling significance	t = 2.004, *P* = 0.045
			+	Conditional exogeneity	χ (1) = 0.096, *P* = 0.757
			+	Robustness	χ (4) = 0.942, *P* = 0.919
			+	Stability over time	t = 0.101, *P* = 0.920, Δ = −0.1%/yr, 95% CI [−1.4, 1.3]
Complications	−7.9%	[−14.0, −2.0]	+	Sampling significance	t = 2.656, *P* = 0.008
			+	Conditional exogeneity	χ (1) = 0.486, *P* = 0.486
			+	Robustness	χ (4) = 0.484, *P* = 0.975
			+	Stability over time	t = 3.684, *P* < 0.001, Δ = −1.1%/yr, 95% CI [−1.7, −0.5]
Readmissions	−17.5%	[−24.2, −11.2]	+	Sampling significance	t = 5.786, *P* < 0.001
			+	Conditional exogeneity	χ (1) = 0.486, *P* = 0.486
			+	Robustness	χ (4) = 0.484, *P* = 0.975
			+	Stability over time	t = 0.406, *P* = 0.685, Δ = −0.1%/yr, 95% CI [−0.7, 0.5]
Senior resident†					
Mortality	−1.7%	[−11.0. 7.2]	−	Sampling significance	t = 0.388, *P* = 0.698
			+	Conditional exogeneity	χ (1) = 0.036, *P* = 0.849
			+	Robustness	χ (4) = 4.107, *P* = 0.392
			+	Stability over time	t = 2.404, *P* = 0.016, Δ = −1.1%/yr, 95% CI [−2.0, −0.2]

*Test criteria are detailed in the Supplement, see http://links.lww.com/AOSO/A267.

†Core models are detailed in Supplement (see http://links.lww.com/AOSO/A267) with variables described in Tables 1.

‡Total 30-day, all-cause mortality, case complications, and 30-day readmissions that would have resulted had attendings who scrubbed had not scrubbed was computed across all VA surgery patients over the 15-year time period: 13,713 deaths, 43,637 complications and 73,832 readmissions = (1,737,662 sampled surgeries—247 emergencies) × (0.77633395 teaching settings/encounter) × (0.73554685 teaching encounters/teaching setting) × (0.85108502 scrubbed attendings/teaching encounter) × (8:1 sampling ratio) × (0.00203 [0.00004, 0.00426] deaths/scrubbed encounter | 0.00646 [0.00164, 0.01138] complications/scrubbed encounter | 0.01093 [0.00703–0.01505] readmissions/scrubbed encounter).

reported 90.52 cases per 1000 patients, and 10.93 [7.03, 15.05] cases with 30-day readmissions would have been added to the reported 67.80 readmissions per 1000 patients.

Secondary Factor Associations

From Table 2, there was no independent association between senior resident status and mortality outcomes. The result neither validates nor invalidates the primary association.

Moderator Analyses

Surgery specialty (F(8, 862,388) = 9.993, *P* < 0.001) (Table 3) and facility complexity (*F*(3, 862,389) = 4.415, *P* = 0.004) (Table 4) had moderated the size of the independent association between attendings having scrubbed and patient mortality. Tests for robustness was not determined because the alternate models would have to be constructed by replacing moderator variables with other covariates. The largest drop in mortality from having scrubbed attendings versus not scrubbed were for thoracic, peripheral vascular, and orthopedic surgeries. Only urological surgery saw mortality rates increase when attending surgeons scrubbed. The most (−23.7% [−91.9, −7.8]) and least (−73.3% [−199.5, −0.2]) complex facilities tended to have the largest strength of association. More research that goes beyond the scope of this paper is needed to better explain these findings.

Trend Analyses

For confirmatory purposes, we compared how attending scrubbed status and patient outcomes changed over time after adjusting for presurgery covariates and the resident's senior status level. The likelihood that surgical attendings in teaching encounters were scrubbed rose annually by 5.1%/yr ([4.9%, 5.2%], t = 53.9, *P* < 0.001, n = 862,425), while 30-day mortality rates declined by 7.3%/yr ([5.9%, 8.5%], t = 11.0, *P* < 0.001, n = 862,425), case complication rates declined by 6.9%/yr ([6.3%, 7.5%], t = 23.6, *P* < 0.001, n = 862,425), and 30-day readmission rates declined by 3.9%/yr ([3.4%, 4.5%], t = 13.6, *P* < 0.001, n = 862,425). Underscoring these trends is the decline in the proportion of patients admitted to a teaching surgery service who go on to be assigned a resident by 0.8%/yr ([0.7%, 1.0%], t = 10.6, *P* < 0.001, n = 1,172,495) after adjusting for changes in the core presurgery covariates.

DISCUSSION

As one of the largest healthcare training platforms in the United States, VA's official GME policy is to: "… provide appropriate supervision for the patient's evaluation, including management decisions involving the patient's medical condition and procedures performed…" by relying on judgments of attending physicians to supervise according to the resident's demonstrated competence, the patient's clinical condition, and care complexity.[34] Under this policy, VA has seen over time an increase in the percent of teaching encounters where surgical attendings had scrubbed and a corresponding decrease in mortality, case complications, and readmission rates.

We assessed this policy by applying control functions to a systematic sample of 862,425 surgery teaching encounters from the Department of Veterans Affairs from July 1, 2004, through September 30, 2019. We know from prior studies[8,25] that there is no evidence from the 128,428 (14.9%) encounters where surgical attendings did not scrub [ANS] that patient outcomes would have improved had those attendings scrubbed. Our study focuses on the larger AS group of 733,997 (85.1%) encounters and determined that patient 30-day mortality, case complications, and readmission rates would have increased had scrubbed attendings not scrubbed. The cumulative impact across VA over the 15 years ending in 2019 had those scrubbed attendings not

TABLE 3.

Impact of Surgery Specialty on Associations Between Resident Supervision and Patient Mortality Outcomes

	OR%	95% CI	±	Test Criteria* Method	Statistic
Thoracic surgery	–75.6%	[–147.9, –24.4]	+	Sampling significance:	$t = 3.197$, $P = 0.001$
			+	Conditional exogeneity:	$X(1)=1.007$, $P = 0.316$
			+	Stability over Time:	$t = 0.501$, $P = 0.616$, $\Delta = 0.30\%$/yr, 95% CI [–1.01, 1.61]
Peripheral vascular surgerys	–55.7%	[–85.7, –30.7]	+	Sampling significance:	$t = 4.940$, $P < 0.001$
			+	Conditional exogeneity:	$X(1) = 0.037$, $P = 0.848$
			+	Stability over time:	$t = 0.624$, $P = 0.532$, $\Delta = 0.40\%$/yr, 95% CI [–0.90, 1.71]
Orthopedic surgery	–38.0%	[–62.3, –17.4]	+	Sampling Significance:	$t = 3.887$, $P < 0.001$
			+	Conditional Exogeneity:	$X(1)=0.038$, $P = 0.845$
			+	Stability over Time:	$t = 0.177$, $P = 0.860$, $\Delta = 0.10\%$/yr, 95% CI [–1.21, 1.41]
NeuroSurgery	–21.4%	[–67.5, 3.1]	–	Sampling Significance:	$t = 1.744$, $P = 0.081$
			+	Conditional Exogeneity:	$X(1)=0.081$, $P = 0.776$
			+	Stability over Time:	$t = 0.303$, $P = 0.762$, $\Delta = 0.20\%$/yr, *95% CI*[–1.11, 1.51]
Otolaryngology surgery	–11.5%	[–63.6, 31.4]	–	Sampling Significance:	$t = 0.560$, $P = 0.576$
			+	Conditional Exogeneity:	$X(1)=0.014$, $P = 0.905$
			+	Stability over Time:	$t = 0.273$, $P = 0.785$, $\Delta = 0.20\%$/yr, *95% CI*[–1.11, 1.51]
Plastic Surgery	8.2%	[–96.0, 129.3]	–	Sampling Significance:	$t = 0.205$, $P = 0.837$
			+	Conditional Exogeneity:	$X(1)=0.165$, $P = 0.684$
			+	Stability over Time:	$t = 0.251$, $P = 0.801$, $\Delta = 0.20\%$/yr, 95% CI [–1.11, 1.51]
General Surgery	10.6%	[–6.2, 30.0]	–	Sampling Significance:	$t = 1.228$, $P = 0.219$
			+	Conditional Exogeneity:	$X(1)=0.763$, $P = 0.382$
			+	Stability over Time:	$t = 0.164$, $P = 0.869$, $\Delta = 0.10\%$/yr, 95% CI [–1.21, 1.41]
Urology Surgery	25.9%	[3.3, 53.3]	+	Sampling Significance:	$t = 2.275$, $P = 0.023$
			+	Conditional Exogeneity:	$X(1)=1.189$, $P = 0.275$
			+	Stability over Time:	$t = 0.202$, $P = 0.840$, $\Delta = -0.10\%$/yr, 95% CI [–1.51, 1.21]

a - $F(8, 862388)=9.993$, $P < 0.001$, test for robustness was not possible since moderators were substituted out to create alternate models, surgery specialty was included if the number of total encounters were 20,000 or greater.

TABLE 4.

Impact of Facility Complexity on Associations Between Resident Supervision and Patient Mortality Outcomes

				Test Criteria*	
	O.R.%	**95%C.I.**	**±**	**Method**	**Statistic**
1a—High	–23.7%	[–91.9, –7.8]	+	Sampling Significance:	t = 3.015, P = 0.003
			+	Conditional Exogeneity:	χ (1)=0.549, P = 0.459
			+	Stability over Time:	t = 0.201, P = 0.840, Δ =–0.10%/yr, 95% CI [–1.41, 1.21]
1b—High	–4.5%	[–22.1, 12.0]	–	Sampling Significance:	t = 0.544, P = 0.586
			+	Conditional Exogeneity:	χ (1)=0.002, P = 0.962
			+	Stability over Time:	t = 0.144, P = 0.886, Δ =–0.10%/yr, 95% CI [–1.41, 1.21]
1c - High	3.4%	[–17.8, 26.0]	–	Sampling Significance:	t = 0.329, P = 0.742
			+	Conditional Exogeneity:	χ (1)=0.226, P = 0.634
			+	Stability over Time:	t = 0.137, P = 0.891, Δ =–0.10%/yr, 95% CI [–1.41, 1.21]
2—Medium/Low	–73.3%	[–199.5, –0.2]	+	Sampling Significance:	t = 1.968, P = 0.049
			+	Conditional Exogeneity:	χ (1) = 0.005, P = 0.943
			+	Stability over Time:	t = 0.091, P = 0.927, Δ =–0.10%/yr, 95% CI [–1.41, 1.31]

*F(3, 862389)=4.415, P = 0.004; test for robustness was not possible since moderators were substituted out to create alternate models.

scrubbed would have added an estimated 13,700 [270–28,800] deaths, 43,600 [11,200–76,900] cases with complications, and 73,800 [47,500–101,700] hospital readmissions (Table 2, note c). These estimates were theory-driven, were tested for sampling error, biases from unaccounted for confounding variables, robustness for model specification, stable over time for future applicability, and valid based on predictable moderators of association size. These findings are important to academic leaders considering opening clinical learning opportunities by expanding resident autonomy from direct supervision.

Our findings are consistent with Tonelli et al[25] logistic regression of a comparable VASQIP sample showing that not scrubbed versus scrubbed surgical attendings were associated with an increased risk of any postoperative complication after adjusting for patient demographic, clinical, facility, procedure, and surgery covariates. These findings are also consistent with Farnan et al[35] review of the literature who concluded supervision did impact a resident's use of resources, compliance with protocol, and incidence of complications of performing diagnostic and therapeutic procedures, though the authors noted most studies were limited by small sample sizes, study design, and poor measures of supervision. Finally, these results are also consistent with Fallon et al[36] medical audit study that reported the presence of an attending surgeon during surgery encounters was associated with lower mortality and case complication rates when clinical procedures were stratified by elective versus non-elective surgeries.

This study also demonstrated the value of a theory-driven, tested, and validated control function approach to estimate GME outcomes. The theory-driven core model added to the list of fixed effects resident PGY status at the encounter level, teaching intensity, procedure variety, and GME program volume at the surgery specialty, facility, and academic year level, medical center complexity at the facility level, and facility identity as a random effect. These added covariates have been shown to be associated with care outcomes[14,37] and resident satisfaction with their clinical learning environment.[38] Our approach also tested for outstanding confounder variance and model robustness. This is important to avoid adding terms to control functions or propensity scores that may amplify any remaining confounding biases when such terms are more associated with resident supervision than patient outcomes.[39] Validating the underlying GME-S theory helps to avoid biases created when mediating and collider variables are inappropriately added as adjusters to control functions.[40]

Some caution in applying these findings to GME policy is advised in light of the growing concerns that graduates may be leaving residency programs unprepared to enter independent practice.[6–23,25] One approach is to change how residents are supervised to promote greater contacts with surgical attendings while maintaining the resident's increasing autonomy during the procedure.[7] On the other hand, a separate study using electronic health records of VA medical centers between 2004 and 2019 showed clinical productivity of surgery residents net of supervision had increased over time consistent with a progressive independence hypothesis.[41] More qualitative and quantitative research on approaches to measure resident supervision using a continuous scale is critically needed to optimize patient care, contribution to clinical workload, and resident education outcomes.

Our study has limitations. Findings may not generalize outside VA where policies on resident supervision may differ, or apply to other medical specialties outside surgery such as general medical[42] and emergency medicine.[43] Assessing resident supervision by whether surgical attendings were scrubbed does not fully quantify the extent residents acted independently to produce a procedure or the amount of the procedure that was done by the attending.[41] Our tests for estimate robustness and factor conditional exogeneity are one-way assessing whether biases exist. Data were not available on the clinical and supervisory skills of attending surgeons. Facilities may interpret PGY level, identify the primary surgeon, and define the role of the attending surgeon differently. For example, PGY may include years spent in research training. Minimum operative procedure time requirements were not standardized when the attending surgeons supervision status was determined.

In conclusion, this study suggests that VA policies on surgery resident supervision have protected patient safety and care effectiveness while allowing residents in selected teaching encounters to have limited autonomy from supervision to gain skills to progress toward independent practice. However, binary scales of resident supervision may lack precision to assess at the encounter level how much autonomy from direct supervision surgery residents should be allowed to optimize both patient outcomes and resident learning. More studies are needed to apply validated and continuous measures of resident supervision to evaluate new ways to achieve patient safety, care quality, and resident education progress in academic medical centers.

REFERENCES

1. Kashner TM, Byrne JM, Henley SS, et al. Measuring progressive independence with the resident supervision index: theoretical approach. *J Grad Med Educ*. 2010;2:8–16.
2. Kashner TM, Byrne JM, Chang BK, et al. Measuring progressive independence with the resident supervision index: empirical approach. *J Grad Med Educ*. 2010;2:17–30.

3. Byrne JM, Kashner TM, Gilman SC, et al. Measuring the intensity of resident supervision in the department of veterans affairs: the resident supervision index. *Acad Med*. 2010;85:1171–1181.
4. Patel M, Bhullar JS, Subhas G, et al. Present status of autonomy in surgical residency—a program director's perspective. *Am Surg*. 2015;81:786–790.
5. Fabricant PD, Dy CJ, Dare DM, et al. A narrative review of surgical resident duty hour limits: where do we go from here? *J Grad Med Educ*. 2013;5:19–24.
6. Hashimoto DA, Bynum WE, Lillemoe KD, et al. See more, do more, teach more: surgical resident autonomy and the transition to independent practice. *Acad Med*. 2016;91:757–760.
7. Haber LA, Lau CY, Sharpe BA, et al. Effects of increased overnight supervision on resident education, decision-making, and autonomy. *J Hosp Med*. 2012;7:606–610.
8. Oliver JB, Kunac A, McFarlane JL, et al. Association between operative autonomy of surgical residents and patient outcomes. *JAMA Surg*. 2022;157:211.
9. Kunac A, Oliver JB, McFarlane JL, et al. General surgical resident operative autonomy vs patient outcomes: are we compromising training without net benefit to hospitals or patients? *J Surg Educ* 2021;78:e174–e182.
10. Stahl CC, Collins E, Jung SA, et al. Implementation of entrustable professional activities into a general surgery residency. *J Surg Educ*. 2020;77:739–748.
11. Burke LG, Frakt AB, Khullar D, et al. Association between teaching status and mortality in US hospitals. *JAMA*. 2017;317:2105–2113.
12. Kupersmith J. Quality of care in teaching hospitals: a literature review. *Acad Med*. 2005;80:458–466.
13. Silber JH, Rosenbaum PR, Romano PS, et al. Hospital teaching intensity, patient race, and surgical outcomes. *Arch Surg*. 2009;144:113–120.
14. Dimick JB, Cowan JA Jr, Colletti LM, et al. Hospital teaching status and outcomes of complex surgical procedures in the United States. *Arch Surg*. 2004;139:137–141.
15. Giordano L, Oliviero A, Peretti GM, et al. The presence of residents during orthopedic operation exerts no negative influence on outcome. *Br Med Bull*. 2019;130:65–80.
16. Itani KMF, DePalma RG, Schifftner T, et al. Surgical resident supervision in the operating room and outcomes of care in veterans affairs hospitals. *Am J Surg*. 2005;190:725–731.
17. Bruno VD, Chivasso P, Hayat A, et al. Propensity-matched analysis of outcomes after mitral valve surgery between trainees and consultants (institutional report). *Interact Cardiovasc Thorac Surg*. 2018;26:443–447.
18. Mazur P, Litwinowicz R, Krzych T, et al. Absence of perioperative excessive bleeding in on-pump coronary artery bypass grafting cases performed by residents. *Interact Cardiovasc Thorac Surg*. 2019;29:836–843.
19. 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.
20. Patel SP, Gauger PG, Brown DL, et al. Resident participation does not affect surgical outcomes despite introduction to new techniques. *J Am Coll Surg*. 2010;211:540–545.
21. Wojcik BM, McKinley SK, Fong ZV, et al. The resident-run minor surgery clinic: a four-year analysis of patient outcomes, satisfaction, and resident education. *J Surg Educ*. 2021;78:1838–1850.
22. Nguyen CT, Hernandez AV, Gao T, et al. Office based vasectomy can be performed by supervised urological residents with patient pain and morbidity comparable to those of a staff surgeon procedure. *J Urol*. 2008;180:1451–1454.
23. Christophersen C, Fonnes S, Andresen K, et al. Risk of reoperation for recurrence after elective primary groin and ventral hernia repair by supervised residents. *JAMA Surg*. 2023;158:359. Available at: https://jamanetwork.com. Accessed June 2, 2023.
24. Massarweh NN, Kaji AH, Itani KM. Practical guide to surgical data sets: Veterans Affairs Surgical Quality Improvement Program (VASQIP). *JAMA Surg*. 2018;153:768–769.
25. Tonelli CM, Cohn T, Abdelsattar Z, et al. Association of resident independence with short-term clinical outcome in core general surgery procedures. *JAMA Surg*. 2023;158:302. Available at: https://jamanetwork.com. Accessed September 2, 2023.
26. Terhune K. Supervision and transparency in resident training (invited commentary). *JAMA Surg*. 2023;158:367. Available at: https://jamanetwork.com. Accessed June 2, 2023.
27. Kashner TM, Henley SS, Golden RM, et al. Making causal inferences about treatment effect sizes from observational datasets. *Biostat Epidemiol*. 2020;4:48–83.
28. Heckman J, Navarro-Lozano S. Using matching, instrumental variables, and control functions to estimate economic choice models. *Rev Econ Stat*. 2004;86:30–57.
29. Lu X, White H. Robustness checks and robustness tests in applied economics. *J Econom*. 2014;178:194–206.
30. Heisler EJ, Mendez BHP, Mitchell A, Panangala SV, Villagrana MA. Federal Support for Graduate Medical Education: An Overview (R44376). Congressional Research Service, December 27, 2018. Available at: https://crsreports.congress.gov. Accessed May 23, 2021.
31. Bernett DS, Kashner TM. Annual Report of the Health Services Training Survey Findings for Academic Year 2019-2020 (VHA-OAA Report#0001). Office of Academic Affiliations, Department of Veterans Affairs, Washington, DC, April 20, 2021. Available at: https://www.va.gov/oaa.
32. American Medical Association. CPT New Codes. Available at: https://www.ama-assn.org/topi9cs/cpt-new-codes. Accessed June 14, 2020.
33. Veterans Health Administration (VHA), Office of Productivity, Efficiency, and Staffing (OPES). Veterans Health Administration (VHA) Facility Complexity Model. Available at: http://opes.vassc.med.va.gov/Pages/Facility-Complexity-Model.aspx. Accessed March 30, 2021.
34. Department of Veterans Affairs, Veterans Health Administration VHA Directive 1440.01. November 7, 2019. Available at: https://www.va.gov/OPTOMETRY/docs/1400_01_D_20191107.pdf. Accessed April 21, 2022.
35. Farnon JM, Petty LA, Georgitis E, et al. A systematic review: the effect of clinical supervision on patient and residency education outcomes. *Acad Med*. 2012;28:428–442.
36. Fallon WF, Jr, Wears RI, Tepas JJ, 3rd. Resident supervision in the operating room: Does this impact on outcome. *J Trauma*. 1993;35:556–561.
37. Haider AD, Scott VK, Rehman KA, et al. Racial disparities in surgical care and outcomes in the United States: a comprehensive review of patient, provider and systemic factors. *J Am Coll Surg*. 2013;216:482–492.
38. Kantor O, Schneider AB, Rojnica M, et al. Implementing a resident acute care surgery service: Improving resident education and patient care. *Surgery*. 2017;161:876–883.
39. Pearl J. Invited Commentary: Understanding bias amplification. *Am J Epidemiol*. 2011;174:1223–1227.
40. Pearl J, Glymour M, Jewell NP. *Causal Inference in Statistics: A Primer*. John Wiley & Sons; 2016.
41. Kashner TM, Greenberg PB, Henley SS, et al. Assessing physician resident contributions to outpatient clinical workload. *Med Care*. 2022;60:709–717.
42. Finn KM, Metlay JP, Chang Y, et al. Effect of increased inpatient attending physician supervision on medical errors, patient safety, and resident education: a randomized clinical trial. *JAMA Intern Med*. 2018;178:952–959.
43. Van Leer PE, Lavine EK, Rabrich JS, et al. Resident supervision and patient safety: do different levels of resident supervision affect the rate of morbidity and mortality cases? *J Emerg Med*. 2015;49:944–948.