UC Davis
UC Davis Previously Published Works

Title

A National Comparison of Operative Outcomes of New and Experienced Surgeons.

Authors

Kelz, Rachel R
Sellers, Morgan M
Niknam, Bijan A
et al.

A National Comparison of Operative Outcomes of New and Experienced Surgeons

Rachel R. Kelz, MD, MSCE, MBA, FACS[*,†], **Morgan M. Sellers, MD**[*], **Bijan A. Niknam, BS**[‡], **James E. Sharpe, MS**[‡], **Paul R. Rosenbaum, PhD**[†,§], **Alexander S. Hill, BS**[‡], **Hong Zhou, MS**[‡], **Lauren L. Hochman, BA**[‡], **Karl Y. Bilimoria, MD, MS**[¶], **Kamal Itani, MD**[∥,#], **Patrick S. Romano, MD, MPH, FACP, FAAP**[**], **Jeffrey H. Silber, MD, PhD**[†,‡,††,‡‡,§§]

[*]Department of Surgery, Center for Surgery and Health Economics, The University of Pennsylvania Perelman School of Medicine, Philadelphia, PA;

[†]The Leonard Davis Institute of Health Economics, The University of Pennsylvania, Philadelphia, PA;

[‡]Center for Outcomes Research, Children's Hospital of Philadelphia, Philadelphia, PA;

[§]Department of Statistics, The Wharton School, The University of Pennsylvania, Philadelphia, PA;

[¶]Surgical Outcomes and Quality Improvement Center (SOQIC), Department of Surgery and Center for Healthcare Studies, Northwestern Medicine, Chicago IL;

[∥]VA Boston Health Care System, Boston, MA;

[#]Department of Surgery, Boston University School of Medicine, Boston, MA;

[**]Division of General Medicine and Center for Healthcare Policy and Research, University of California Davis School of Medicine, Sacramento, CA;

[††]The Departments of Pediatrics, The University of Pennsylvania Perelman School of Medicine, Philadelphia, PA;

[‡‡]Department of Anesthesiology and Critical Care, The University of Pennsylvania Perelman School of Medicine, Philadelphia, PA;

[§§]Department of Health Care Management, The Wharton School, The University of Pennsylvania, Philadelphia, PA.

Abstract

Objective: To determine whether outcomes achieved by new surgeons are attributable to inexperience or to differences in the context in which care is delivered and patient complexity.

Background: Although prior studies suggest that new surgeon outcomes are worse than those of experienced surgeons, factors that underlie these phenomena are poorly understood.

Methods: A nationwide observational tapered matching study of outcomes of Medicare patients treated by new and experienced surgeons in 1221 US hospitals (2009–2013). The primary outcome studied is 30-day mortality. Secondary outcomes were examined.

Results: In total, 694,165 patients treated by 8503 experienced surgeons were matched to 68,036 patients treated by 2119 new surgeons working in the same hospitals. New surgeons' patients were older (25.8% aged 85 vs 16.3%, $P<0.0001$) with more emergency admissions (53.9% vs 25.8%, $P<0.0001$) than experienced surgeons' patients. Patients of new surgeons had a significantly higher baseline 30-day mortality rate compared with patients of experienced surgeons (6.2% vs 4.5%, $P<0.0001$; OR 1.42 (1.33, 1.52)). The difference remained significant after matching the types of operations performed (6.2% vs 5.1%, $P<0.0001$; OR 1.24 (1.16, 1.32)) and after further matching on a combination of operation type and emergency admission status (6.2% vs 5.6%, $P=0.0007$; OR 1.12 (1.05, 1.19)). After matching on operation type, emergency admission status, and patient complexity, the difference between new and experienced surgeons' patients' 30-day mortality became indistinguishable (6.2% vs 5.9%, $P=0.2391$; OR 1.06(0.97, 1.16)).

Conclusions: Among Medicare beneficiaries, the majority of the differences in outcomes between new and experienced surgeons are related to the context in which care is delivered and patient complexity rather than new surgeon inexperience.

Society relies on the entry of new surgeons to serve the needs of the population over time. In the United States, a surgeon shortfall of between 20,700 and 30,500 surgeons is anticipated by 2030.[1] With approximately 1800 general and orthopedic surgeons entering the workforce each year,[2] it is important to understand more about the relative performance of new surgeons. Most studies examine new surgeon performance on single operations or conditions[3–7] or the experience of single surgeons or hospitals.[3,4,7] Others conflate experience with surgeon age or procedural volume.[8–17] More recently, we demonstrated that new surgeon performance when compared with that of experienced surgeons was not negatively affected by duty hour reform or the accompanied changes in the nature of surgical care.[18] Outcomes of patients treated by new surgeons were slightly worse than those of experienced surgeons regardless of an experiential training model or an outcomes-based model.

The transition to practice is a critical time in the development of new surgeons. In surgery— where a single surgeon performs the majority of the operation, assumes responsibility for patient selection, and directs the preoperative and postoperative care[19] —early wins are essential and poor outcomes can be devastating for both the patient and the surgeon. There is scant information available regarding new and experienced surgeons' practice patterns or the outcomes of the patients they treat. This information is critical for optimizing the entrance of new surgeons into independent practice and for assessing surgeon performance. Thus, our objectives were 1) to examine differences between new and experienced surgeons' practices

and 2) to compare the postoperative outcomes achieved by new and experienced surgeons using tapered matching to understand whether observed outcome deficiencies achieved by new surgeons are due to inexperience or to differences in the context in which care is delivered and the complexity of patient needs.

METHODS

This research protocol was approved by the institutional review board of the University of Pennsylvania.

Patients and Surgeons

Fee-for-service Medicare beneficiaries aged 65.5 or older who underwent orthopedic or general surgery from 2009 through 2013 were identified using the international classification of diseases (ICD)-9 principal procedure field of the Medicare inpatient file, and assigned to the operating physician using the Part B file (see Supplemental Digital Content eTable 1, eTable 2 and eSection 1, http://links.lww.com/SLA/B668 for lists of principal procedures and algorithm for assignment of patients to surgeons). We assigned each patient to a procedure category based on the ICD-9 principal procedure code and the current procedural terminology bill submitted by the surgeon. All operations studied required an incision. If a patient had multiple qualifying admissions, a random admission was selected.

Surgeon information was derived from the American Medical Association's Physician Masterfile.[20] Surgeons were considered new during the first 3 years of independent practice. Surgeons missing information on the residency completion date were classified as new if they graduated medical school in 2003 or later and billed for an operation performed at least 5 years after the graduation date. Experienced surgeons had at least 10 years of experience in independent practice and completed residency in or after 1970.

Surgeon specialty (general or orthopedic) was classified using the preponderance of their operations. Surgeon operative volume was calculated. Surgeons were assigned to the hospital where they performed the majority of their operations over the study period. To qualify for study inclusion, surgeons had to bill for at least 10 qualifying specialty-specific operations in a single hospital. To match new and experienced surgeons in the same hospital, we only included surgeons at hospitals with both a new and experienced surgeon (see Supplemental Digital Content eTable 3, http://links.lww.com/SLA/B668).

We obtained hospital characteristics from Medicare cost reports[21] and the Medicare Provider of Services (POS) files.[22]

Matching Covariates

We defined each patient's age at admission, admission year, sex, emergent or transfer status, admission in the previous 6 months, 30 comorbidities,[23–25] number of comorbidities per patient, and secondary operative procedure status.[26] Transfer status was assigned to patients with claims that appeared within 1 day prior to the index hospitalization of interest issued by an alternative hospital. To be included as a transfer, a patient's claim needed to have both a different claim ID and different hospital ID from the index hospitalization.[27] We further

defined a propensity score for treatment by a new surgeon and an externally estimated risk score for 30-day mortality (see Supplemental Digital Content eSection 2 for a list of all matching covariates and eTable 4 for risk models, http://links.lww.com/SLA/B668).

Clinical Outcomes

The primary outcome was 30-day all-location mortality. We also studied failure to rescue,[25,28,29] readmission or death within 30-days of discharge, rate of ICU use, anesthesia time, and rate of prolonged length of stay.[30,31] See Supplemental Digital Content eSection 3, http://links.lww.com/SLA/B668 for the anesthesia time algorithm. A prolonged stay is a length of stay longer than the point at which the rate of discharge begins to decrease (see Supplemental Digital Content eTable 5, http://links.lww.com/SLA/B668). We further studied length of stay (LOS) and resource utilization-based cost (see Supplemental Digital Content eSection 4 for costing algorithm, http://links.lww.com/SLA/B668).[32]

STATISTICAL ANALYSIS

Matching Methodology

The Surgeon Match—We randomly sampled 10 patients per new surgeon, to prevent the study from emphasizing new surgeons with the largest practices, who are presumably the most experienced of the new surgeons. Using these 10 patients, new surgeons were matched to experienced surgeons practicing within the same hospital based on the types of operations they performed. Thus, we used nearly all of the new surgeons, and drew from their patients in a manner that was representative of the population of new surgeons who cared for Medicare beneficiaries.

The Patient Matches

Four matched controls groups were constructed, with each successive match controlling for additional covariates, a process known as tapered matching.[33–35] Specifically, for the 10 new surgeon patients, 4 sets of 10 matched controls were selected from the patients of an experienced surgeon. For each match, the 10 new surgeon patients remained fixed while the 10 matched patients from the experienced surgeon changed according to the variables included in the match. The "Baseline" analysis served as an unadjusted analysis, picking 10 experienced surgeon patients who had operations in the same year as the matched new surgeons' patients. The "Operation" analysis picked 10 experienced surgeon patients best matched for the individual operation type in the same year. The "Emergent Status" analysis picked the 10 best matched patients of the paired experienced surgeon based on the operation types, the emergent status of the operations and the year. Finally, the "Risk Factor" analysis paired patients based on the previously described covariates and additional patient factors including the patient demographics, age, comorbid conditions, predicted risk of 30-day mortality, and others.[36] See Supplemental Digital Content eSection 2, http://links.lww.com/SLA/B668 for a detailed description of the matching algorithm and a complete list of matching covariates.

Matches between new and experienced surgeons and their patients were accomplished using the rcbsubset[37] package; exterior matches were performed using the ExteriorMatch package. All matching was performed in R, version 3.2.1.[38]

Statistical Tests

Differences in binary outcomes were compared using the McNemar statistic.[39] M-statistics for matched pairs[40–43] were used to compare continuous outcomes. Comparisons between different groups of experienced surgeon controls were performed by applying the McNemar and Wilcoxon signed rank tests to an exterior match that constructed nonoverlapping control groups from 2 given sets of matched controls.[44] All outcome testing was performed using statistical packages run on the R programming platform, version 3.2.1.[38]

RESULTS

Study Population and Setting

We identified 2119 new and 8503 experienced surgeons. Overall 762,201 qualifying patients were treated, 68,036 by new surgeons and 694,165 by experienced surgeons, at 1221 study hospitals. Pairing exactly on the hospital, the surgeon match retained 1820 eligible new surgeons (85.9%) and 1820 experienced surgeons. After matching surgeons, new surgeons had an average of 1.62 years of experience at the time of the operation while experienced surgeons had an average of 21.30 years of experience ($P<0.0001$).

Table 1 describes the hospital practice settings of new and experienced surgeons before and after matching. Before matching, new surgeons more frequently practiced in nonteaching hospitals (46.2% vs 39.4%, $P<0.001$). They also typically operated in smaller hospitals (mean bed size 358.7 vs 421.3, $P<0.001$), and less frequently operated in hospitals with high-level technology (61.8% vs67.8%, $P<0.001$). After matching new surgeons to experienced surgeons in the same hospital, the hospital practice settings were identical.

Quality of the Matches

Table 2 shows the characteristics of the new surgeon patients and those of the experienced surgeons prior to matching (outer dark gray columns), after the surgeons were paired (inner light gray columns), and for each of the groups of experienced surgeon patients matched sequentially to the patients of the new surgeons (unshaded). As indicated by the staircase line in Table 2, moving from right to left entails matching for additional covariates, making the groups more and more alike. After pairing the surgeons, there were noteworthy differences in the patients treated by new and experienced surgeons. For example there were differences in the operations they performed. Within general surgery, cholecystectomy comprised 29.8% of new surgeon procedures compared with 25.4% for experienced surgeons, and within orthopedic surgery, knee replacement comprised 16.8% of new surgeon procedures compared with 47.3% for experienced surgeons. There were also differences in the presentation of the patients to the new and experienced surgeons. For example, new surgeons' patients presented emergently nearly twice as often as experienced surgeons' patients (53.9% vs 25.8%, $P<0.0001$); likewise, new surgeons operated on transferred patients more often (2.5% vs 1.5%, $P<0.0001$). New surgeons more frequently operated on

patients aged 85 and older (25.8% vs 16.3%, $P<0.0001$). Patients of new surgeons also had a higher probability of 30-day mortality on admission than patients of experienced surgeons (5.6% vs 3.0%, $P<0.0001$).

The 4 matched groups of patients treated by experienced surgeons sequentially removed certain characteristics while leaving others intact. This method was designed so that in each given match, differences in unmatched variables could be examined to provide insights into how they may relate to the residual differences outcomes of new surgeons. In all matches, the variables deliberately controlled in the match were closely balanced, with no standardized difference exceeding 0.05 SDs. Here, we see that the match for operation type alone controlled for many of the differences between the new and experienced surgeon patients with a few remaining anticipated differences. For example, after matching for the operations, 53.9% new surgeon patients presented emergently compared with 43.0% of the experienced surgeon patients. The emergency admission status match then removed the difference in emergency admission presentation, choosing experienced surgeon patients who presented emergently in 53.9% of cases—identical to the new surgeon rate (53.9%), while still controlling for year and operation-type. The risk factor match then removed the remaining difference in the additional patient risk factors, achieving nearly identical distributions of each factor across patients of new and experienced surgeons. As such, we can examine the relative contribution of each of these elements to the disparate outcomes achieved by the new surgeons. See Supplemental Digital Content eTable 6, http://links.lww.com/SLA/B668 for the full list of covariates and eTables 7 to 10 for the quality of balance in each match.

Outcomes

Table 3 reports the differences in the outcomes achieved in new surgeons' patients and in experienced surgeons' patients after matching on the year in which the operation was performed (Baseline Match); the Year and Operation; the Year, Operation and Emergency Admission Status; and finally, matching on Year, Operation, Emergency Admission Status, and Patient Risk Factors.

30-day All-location Mortality Rate

Patients of new surgeons had a significantly higher 30-day all-location mortality rate (6.2%) than patients of experienced surgeons (4.5%, $P<0.0001$; OR 1.42 (1.33, 1.52)). After additional matching to control for differences in operations performed, the 30-day mortality rate of new surgeon patients remained significantly higher than that of matched experienced surgeon patients (6.2% vs 5.1%, $P<0.0001$; OR 1.24 (1.16, 1.32)). Then, after further matching on operation type and emergency admission status, the difference in 30-day mortality between new and experienced surgeons narrowed(6.2% vs 5.6%, $P=0.0007$; OR 1.12 (1.05, 1.19)). Finally, after matching on operation type, emergency admission status, and patient complexity, the difference in the 30-day mortality rate of new surgeon patients was no longer significantly higher than that of the experienced surgeons patients (6.2% vs 5.9%, $P=0.2391$; OR1.06 (0.97, 1.16)). See Supplemental Digital Content eTable 11 for outcomes within specific surgical specialties, http://links.lww.com/SLA/B668.

Figure 1 displays the contribution of the operation type, emergency admission status, and patient complexity to the 30-day all-location mortality rate of the experienced surgeons in comparison with the new surgeons is demonstrated. The operative mix of the experienced surgeons is significantly associated with the difference in 30-day all-location mortality rate (*P*=0.0004). Similarly, emergency status (*P*=0.0001) is also significantly associated with the 30-day all-location mortality rate achieved by the experienced surgeons whereas, after controlling for operative mix and emergency status, patient complexity is not significantly associated with the experienced surgeon performance in 30-day all location mortality (*P* = 0.43).

Secondary Clinical Outcomes

Table 3 also reports differences in secondary clinical outcomes across the tapered match. For 30-day readmission or death, the pattern was similar to that observed for 30-day all location mortality (19.4% vs 18.6%, *P* = 0.0305; OR 1.06 (1.01, 1.12)). For prolonged length of stay and anesthesia time, while each step of the tapered match resulted in a reduction in the difference between outcomes of patients treated by new and experienced surgeons, a greater proportion of new surgeon patients experienced prolonged length of stay than did experienced surgeons patients (56.9% vs 54.5%, *P*<0.0001;OR 1.14 (1.08, 1.19)) and new surgeons patients required a longer anesthesia time than that required for experienced surgeon patients (155.4 min vs 137.6 min, *P*<0.0001; a paired difference of 18.7 min (17.4, 20.1)). Notably, there was no difference in failure-to-rescue rates between patients treated by new and experienced surgeons after controlling for patients' hospital setting, year, operation-type, and emergent status (9.5% vs 9.0%, *P* = 0.1127; 1.07(0.99, 1.16)).

In Figure 1, the contribution of each factor to the individual clinical outcome differences between new and experienced surgeon patients is demonstrated graphically. Here, we see that differences in 30-day readmission or death, prolonged length of stay and anesthesia time, appear to be explained to a great extent by the operations required by the surgeons' patients and their emergency admission status. Patient risk factors do not appear to be contributing as much to the differences after matching on the operations and admission status.

Secondary Utilization Outcomes

Table 3 also describes differences in utilization measures. Before matching and at each step of the tapered match, patients of new surgeons had significantly higher rates for all utilization measures. Before matching, there was a +3.9% (3.1%–4.8%) difference in ICU usage between new and experienced surgeons which declined to +3.3% (2.4%–4.1%), +2.7% (1.9%–3.6%), and +2.2% (1.4%–3.1%) after matching operation type, operation type plus emergency admission and operation type, emergency admission and patient risk factors, respectively. Differences in length of stay were real but small with a baseline matched difference of +0.6 (0.5,+0.7) which declined to a clinically insignificant difference of +0.2 (+0.1,+0.3) after matching on operation type and emergency status. The LOS difference did not change with the addition of patient risk factors to the match (+0.2 (+0.1, 0.2)). Differences in 30-day resource costs from +2466 in the baseline match to +1888, +1340, and

+1257 after matching operation type, operation type emergency admission and operation type, emergency admission, and patient risk factors, respectively.

In Figure 2, we see that differences in length of stay and 30-day resources costs between new and experienced surgeons appear to be explained to a great extent by the operations required by the patients and the emergency admission status. After having already matched on these covariates, additional matching for patient risk factors does not appear to affect the differences as much.

DISCUSSION

Surgeons are generally thought to improve with experience,[8] up to certain limits.[10,12,45] Therefore, new surgeons are often thought to have maturing skills and poorer outcomes. However, surgical outcomes reflect a combination of patient risk, surgeon skill across a variety of domains, and hospital quality. This study examines differences in practice composition and assesses clinical and utilization outcomes between new and experienced surgeons who treated a nationwide cohort of Medicare beneficiaries. Testing the hypothesis that outcomes of new surgeons are different from those of experienced surgeons, we found vast differences not only in outcomes but also in salient patient and procedural factors. After controlling for these differences and the hospitals in which care was delivered, new surgeons generally achieved outcomes that were almost similar to experienced surgeons.

For the first time to our knowledge, we document differences in surgeon practice composition by experience level. In a nationwide Medicare cohort, we demonstrate that new surgeons treat a disproportionate portion of emergency referrals. Accordingly, the typical operative practices of new and experienced surgeons are different presumably due to differences in their patients' characteristics. Finally, new surgeons' patients are more often older and higher risk than those of experienced surgeons who practice in the same hospital.

Next, we show surgical outcomes by experience level. At first glance, the baseline results suggest that patients of new surgeons fare substantially worse than patients of experienced surgeons. For the baseline in this study, new surgeon patients had an odds ratio of 1.42 (1.33, 1.52) for death (or 42% higher odds) following the operation relative to experienced surgeon patients despite receiving treatment in the same hospitals. After controlling for differences in operative mix, the odds ratio declined to 1.24 (1.16, 1.32), a reduction of almost half the excess risk. Then, after further controlling for emergency admission status, the odds ratio declined to 1.12 (1.05,1.19). Finally, after further controlling for patient risk factors, the odds ratio declined to just 1.06 (0.97, 1.16). Taken together, the evidence demonstrates that the great majority of the baseline mortality difference between new and experienced surgeons can be explained by differences in patient case-mix rather than surgeon inexperience. These patterns were consistent across almost all of the outcomes examined except failure-to-rescue.

The difference in failure-to rescue followed a slightly different pattern than the other measures studied, as it appears to be primarily due to the emergency admission status of the patients. This is reassuring as there may be some residual confounding by indication due to

physiological derangements common to this population. After matching for emergency status, there was no residual difference in the failure-to-rescue rates between new and experienced surgeons. This is encouraging because failure-to-rescue is a well-established measure of hospital quality and our surgeon and patient pairs were matched within hospitals specifically to control for between-hospital differences. Second, while it is possible that patients of new and experienced surgeons might have received differential care in the post-operative phase based on the status of the surgeon, the finding of near-equivalent failure-to-rescue suggests that the quality of care delivered to patients of new surgeons approximates that provided to patients of experienced surgeons.

Placing these findings in the context of the literature, we find that prior studies of interventionists including surgeons have only examined single procedures[46–48] or subsets of non-elective procedures.[8] As such, these studies have not been able to examine the totality of differences in practice patterns by experience level, which may influence the observed outcomes at the population level. The existing studies have yielded mixed results, with some finding no difference in outcomes,[4,7,11,12] while others report benefits of experience.[3,5,6,8,9,16] Furthermore, heterogenous definitions of experienced surgeon have limited the interpretation and generalizability of the results.[8,46–48]

In the present study, we restricted the definition of new surgeon status to the treatment they provided in their first 3 years of practice, deliberately sampling new surgeons to capture the most vulnerable period just after the transition to practice, and compared outcomes to surgeons with at least 10 years of experience. Additionally, we examined a comprehensive set of procedures including elective and nonelective cases performed across specialties in over 1200 hospitals; and, the matching design permitted the examination of a broad array of clinical and financial outcomes on the same sets of matched pairs.

Despite the small magnitude of the remaining differences in the adjusted outcomes, including mortality, it is important to consider the possible explanations for the observed findings. First, it is plausible that the differences simply reflect unobserved severity of illness that cannot be measured using administrative data. Alternatively, it is reasonable to believe that surgeon inexperience confers a modest risk of adverse outcome. Whether the risk is a function of surgeon inexperience or unmeasured severity of illness, strategies that provide additional support for new surgeons, and their complex patients should be considered.

The majority of the reduction in outcome differences came from matching on the operation, suggesting that the operations required by the patients of new surgeons are associated with other high-risk characteristics. For example, an emergency cholecystectomy, often performed by a new surgeon, has been found in the literature to be associated with an inherently high risk of death or serious morbidity (6.4%–16.0%).[49,50] As such, the new surgeon patients may merit more attention from experienced surgeons prior to and during the operation, to guide both surgical judgement (eg, when to operate and which procedure to perform) and operative technique. This concept aligns with the principles applied by programs such as the American College of Surgeons Transition to Practice Program.[51]

Furthermore, guidelines that encourage new surgeons to discuss high-risk cases with an experienced colleague may further reduce the small observed difference.

There are several limitations to this study. Our claims data lacked information on physiologic factors or the full complexity of the procedure. However, we did adjust for the presence of secondary operative procedures on the day of surgery. This approach has been shown to be a useful proxy for operative complexity.[52] Moreover, a nationwide data source with physiologic data does not presently exist. Second, because we matched within hospitals to control for the hospital effects, the analysis was limited to hospitals that credential new and experienced surgeons. This might limit the generalizability of the experienced surgeon control performance, as there are many experienced surgeons who practice at hospitals without new surgeons. However, as we retained 85.9% of the new surgeons in the match, the impact of this limitation is not likely to be important for our specific question. This study examined outcomes only for 2 highly representative specialties in the Medicare population, namely general surgery, a broad and primary care field, and orthopedic surgery, a leader of surgical volume. Therefore, it is not known whether the conclusions are generalizable to other specialties or to younger patient populations.

Our study has several notable strengths. The majority of operations in the United States are performed on older patients and this study examined a national cohort of Medicare beneficiaries. We also matched new and experienced surgeons within the same hospital. As such, we were able to overcome biases that may have resulted from differences in resource availability, coding practices, or other confounding hospital factors. Beyond yielding new knowledge on new surgeon performance, this study demonstrates that the small gap in performance between new and experience surgeons can be quantified using patient outcomes. This confirms the utility of clinical and utilization outcomes data in measuring one dimension of surgical education; namely, new surgeon performance. As medical education has moved to an outcomes-based system of accreditation, this finding is important, as a robust audit and feedback system for graduate medical education is presently lacking.

CONCLUSIONS

New and experienced surgeons are called upon to treat different patient populations, with new surgeons typically treating older and sicker patients and more frequently in emergency settings. After controlling for hospital, operation type, emergency admission status and patient risk factors, patients of the newest surgeons have only slightly worse outcomes across multiple measures than patients of experienced surgeons. These differences appear to be explained primarily by the types of operations required by the new surgeons' patients and emergency admission status rather than the inexperience of the new surgeons. Because new and experienced surgeons were matched within the same hospital, differences are not explained by the hospitals in which they practice, their coding practices, or the quality of care within the hospital in which care is delivered. When preparing surgeons to handle the excessive disease burden common to new practices, enhanced support in practice may be useful.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

REFERENCES

1. The Complexities of Physician Supply and Demand: Projections from 2016 to 2030: Final Report, 2018 Update. Prepared by IHS Markit Ltd for Association of American Medical Colleges. 3 2018 Available at: https://aamc-black.-global.ssl.fastly.net/production/media/filer_public/bc/a9/bca9725e-3507-4e35-87e3-d71a68717d06/aamc_2018_workforce_projections_update_april_11_2018.pdf. Accessed August 8, 2018.

2. American College of Surgeons Health Policy Research Institute. The Surgical Workforce in the United States: Profile and Recent Trends. Published April 2010. Available at: http://www.acshpri.org/pubs.html. Accessed August 8, 2018.

3. Elbardissi AW, Duclos A, Rawn JD, et al. Cumulative team experience matters more than individual surgeon experience in cardiac surgery. J Thorac Cardiovasc Surg. 2013;145:328–333. [PubMed: 23083794]

4. Scherl S, Mehra S, Clain J, et al. The effect of surgeon experience on the detection of metastatic lymph nodes in the central compartment and the pathologic features of clinically unapparent metastatic lymph nodes: what are we missing when we don't perform a prophylactic dissection of central compartment lymph nodes in papillary thyroid cancer? Thyroid. 2014;24:1282–1288. [PubMed: 24787362]

5. Cahill PJ, Pahys JM, Asghar J, et al. The effect of surgeon experience on outcomes of surgery for adolescent idiopathic scoliosis. J Bone Joint Surg Am. 2014;96:1333–1339. [PubMed: 25143493]

6. Mahmoudi E, Lu Y, Chang SC, et al. The associations of hospital volume, surgeon volume, and surgeon experience with complications and 30-day rehospitalization after free tissue transfer: A national population study. Plast Reconstr Surg. 2017;140:403–411. [PubMed: 28746290]

7. Yuksel V, Ozdemir AC, Huseyin S, et al. Impact of surgeon experience during carotid endarterectomy operation and effects on perioperative outcomes. Braz J Cardiovasc Surg. 2016;31:444–448. [PubMed: 28076622]

8. Tsugawa Y, Jena AB, Orav EJ, et al. Age and sex of surgeons and mortality of older surgical patients: observational study. BMJ. 2018;361:k1343. [PubMed: 29695473]

9. Sheetz KH, Ibrahim AM, Regenbogen SE, et al. Surgeon experience and Medicare expenditures for laparoscopic compared to open colectomy. Ann Surg. 2018;268:1036–1042. [PubMed: 28549007]

10. Hashimoto DA, Bababekov YJ, Mehtsun WT, et al. Is annual volume enough? The role of experience and specialization on inpatient mortality after hepatectomy. Ann Surg. 2017;266:603–609. [PubMed: 28692470]

11. Dubois L, Allen B, Bray-Jenkyn K, et al. Higher surgeon annual volume, but not years of experience, is associated with reduced rates of postoperative complications and reoperations after open abdominal aortic aneurysm repair. J Vasc Surg. 2018;67:1717–1726. e1715. [PubMed: 29248240]

12. Anderson BR, Wallace AS, Hill KD, et al. Association of surgeon age and experience with congenital heart surgery outcomes. Circ Cardiovasc Qual Outcomes. 2017;10:pii: e003533.

13. Yeo HL, Abelson JS, Mao J, et al. Surgeon annual and cumulative volumes predict early postoperative outcomes after rectal cancer resection. Ann Surg. 2017;265:151–157. [PubMed: 28009740]

14. Forte ML, Virnig BA, Eberly LE, et al. Provider factors associated with intramedullary nail use for intertrochanteric hip fractures. J Bone Joint Surg Am. 2010;92:1105–1114. [PubMed: 20439655]

15. Ulmer WD, Prasad SM, Kowalczyk KJ, et al. Factors associated with the adoption of minimally invasive radical prostatectomy in the United States. J Urol. 2012;188:775–780. [PubMed: 22818134]

16. Suh DH, Cho HY, Kim K, et al. Matched-case comparisons in a single institution to determine critical points for inexperienced surgeons' successful performances of laparoscopic radical hysterectomy versus abdominal radical hysterectomy in stage IA2-IIA cervical cancer. PLoS One. 2015;10:e0131170. [PubMed: 26110866]

17. McArdle CS, Hole DJ. Influence of volume and specialization on survival following surgery for colorectal cancer. Br J Surg. 2004;91:610–617. [PubMed: 15122614]

18. Kelz RR, Niknam BA, Sellers MM, et al. Duty hour reform and the outcomes of patients treated by new surgeons. Ann Surg. 2019.

19. Hoyt D, Ko C. Chapter 2: Team Based Care: The Surgeons as Leader in Each Phase of Care In: Hoyt DB, Ko CY, eds. Optimal Resources for Surgical Quality and Safety. Chicago, IL: American College of Surgeons; 2017.

20. American Medical Association. AMA Physician Masterfile. Available at: https://www.ama-assn.org/life-career/ama-physician-masterfile. Accessed October 15, 2015.

21. CMS.gov. Cost Reports. 4 2017 Available at: https://www.cms.gov/research-statistics-data-and-systems/downloadable-public-use-files/cost-reports/. Accessed June 9, 2017.

22. Centers for Medicare & Medicaid Services. Provider of Services Current Files. 4 2017 Available at: https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Provider-of-Services/.Accessed June 9, 2017.

23. Silber JH, Rosenbaum PR, McHugh MD, et al. Comparison of the value of nursing work environments in hospitals across different levels of patient risk. JAMA Surg. 2016;151:527–536. [PubMed: 26791112]

24. Silber JH, Rosenbaum PR, Kelz RR, et al. Examining causes of racial disparities in general surgical mortality: Hospital quality versus patient risk. Med Care. 2015;53:619–629. [PubMed: 26057575]

25. Center for Outcomes Research, Children's Hospital of Philadelphia. National Quality Forum Outcomes Measures Maintained by the Center for Outcomes Research. 2015 Available at: http://cor.research.chop.edu/node/26. Accessed April 6, 2018.

26. Agency for Healthcare Research and Quality. Clinical Classifications Software (CCS) for ICD-9-CM. Health Care Utilization Project. 3 2017 Available at: www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp. Accessed April 20, 2018.

27. Yale New Haven Health Services Corporation/Center for Outcomes Research &, Evaluation., Prepared for Centers for Medicare & Medicaid, Services., 2014 Procedure-Specific Readmission Measures Updates and Specifications, Report. Elective Primary Total Hip Arthroplasty (THA) and/or Total Knee Arthroplasty (TKA) – Version, 30, 3 2014.

28. Silber JH, Williams SV, Krakauer H, et al. Hospital and patient characteristics associated with death after surgery: a study of adverse occurrence and failure to rescue. Med Care. 1992;30:615–629. [PubMed: 1614231]

29. Ghaferi AA, Birkmeyer JD, Dimick JB. Variation in hospital mortality associated with inpatient surgery. N Engl J Med. 2009;361:1368–1375. [PubMed: 19797283]

30. Silber JH, Rosenbaum PR, Zhang X, et al. Estimating anesthesia and surgical procedure times from Medicare anesthesia claims. Anesthesiology. 2007;106:346–355. [PubMed: 17264730]

31. Silber JH, Rosenbaum PR, Rosen AK, et al. Prolonged hospital stay and the resident duty hour rules of 2003. Med Care. 2009;47:1191–1200. [PubMed: 19786912]

32. Silber JH, Rosenbaum PR, Kelz RR, et al. Medical and financial risks associated with surgery in the elderly obese. Ann Surg. 2012;256:79–86. [PubMed: 22566017]

Author Manuscript

33. Daniel SR, Armstrong K, Silber JH, et al. An algorithm for optimal tapered matching, with application to disparities in survival. J Comput Graph Stat. 2008;17:914–924. [PubMed: 25580069]

34. Silber JH, Rosenbaum PR, Clark AS, et al. Characteristics associated with differences in survival among black and white women with breast cancer. JAMA. 2013;310:389–397. [PubMed: 23917289]

35. Silber JH, Rosenbaum PR, Ross RN, et al. Racial disparities in colon cancer survival. A matched cohort study. Ann Intern Med. 2014;161:845–854. [PubMed: 25506853]

36. Hansen BB. The prognostic analogue of the propensity score. Biometrika. 2008;95:481–488.

37. Pimentel SD. Optimal Subset Matching with Refined Covariate Balance. Package "rcbsubset". Version 1.1.3. 11 12, 2017 Available at: https://cran.r-project.org/web/packages/rcbsubset/rcbsubset.pdf. Accessed April 20, 2018.

38. R Development Core Team. R: A Language and Environment for Statistical Computing. 2018 Available at: http://www.R-project.org. Accessed April 20, 2018.

39. Bishop YMM, Fienberg SE, Holland PW. Discrete Multivariate Analysis: Theory and Practice. Cambridge, MA: The MIT Press; 1975.

40. Rosenbaum PR. Sensitivity analysis for m-estimates, tests, and confidence intervals in matched observational studies. Biometrics. 2007;63:456–464 (R package sensitivitymv and sensitivitymw). [PubMed: 17688498]

41. Maritz JS. A note on exact robust confidence intervals for location. Biometrika. 1979;66:163–166.

42. Rosenbaum PR. Two R packages for sensitivity analysis in observational studies. Obs Stud. 2015;1:1–17.

43. Huber PJ. Chapter 3. The Basic Types of Estimates. In: Robust Statistics. Hoboken, NJ: John Wiley & Sons; 1981:43–55.

44. Rosenbaum PR, Silber JH. Using the exterior match to compare two entwined matched control groups. Am Stat. 2013;67:67–75.

45. Waljee JF, Greenfield LJ, Dimick JB, et al. Surgeon age and operative mortality in the United States. Ann Surg. 2006;244:353–362. [PubMed: 16926561]

46. O'Neill L, Lanska DJ, Hartz A. Surgeon characteristics associated with mortality and morbidity following carotid endarterectomy. Neurology. 2000;55:773–781. [PubMed: 10993995]

47. Hartz AJ, Kuhn EM, Pulido J. Prestige of training programs and experience of bypass surgeons as factors in adjusted patient mortality rates. Med Care. 1999;37:93–103. [PubMed: 10413397]

48. Khatana SAM, Fiorilli PN, Groeneveld PW, et al. Abstract 006: association between percutaneous coronary intervention outcomes and physician education and board certification in New York State. Circ Cardiovasc Qual Outcomes. 2017;10.

49. Ingraham AM, Cohen ME, Bilimoria KY, et al. Comparison of 30-day outcomes after emergency general surgery procedures: potential for targeted improvement. Surgery. 2010;148:217–238. [PubMed: 20633727]

50. Melloul E, Denys A, Demartines N, et al. Percutaneous drainage versus emergency cholecystectomy for the treatment of acute cholecystitis in critically ill patients: does it matter? World J Surg. 2011;35:826–833. [PubMed: 21318431]

51. American College of Surgeons. Transition to Practice. From Resident to General Surgeon. Available at: https://www.facs.org/education/program/ttp. Accessed April 21, 2018.

52. Simmons KD, Hoffman RL, Kuo LE, et al. Is a colectomy always just a colectomy? Additional procedures as a proxy for operative complexity. J Am Coll Surg. 2015;221:862–870. e861–862. [PubMed: 26228014]
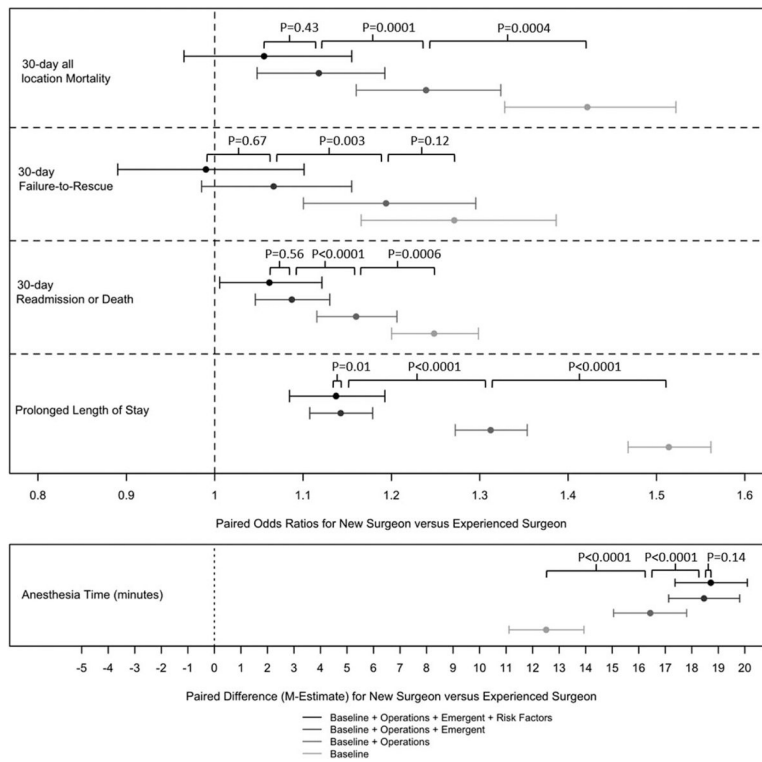
FIGURE 1.

Clinical outcomes. The point estimates and the 95% CIs concern the difference in performance of new and experienced surgeons as the adjustment for the difference in their patient mix becomes more extensive. The *P* values were calculated using the exterior match. Each *P* value indicates whether adjustment for a characteristic alters the relative performance of new and experienced surgeons: Did the adjustment make a difference? For example, in the 30-day all-location mortality row, the $P = 0.0004$ tells us that the operation mix performed by experienced surgeons is a significant component of the superior outcomes achieved by the experienced surgeons. Similarly, in the same row, we can see that emergency status ($P = 0.0001$) is also significant component of the superior performance of experienced surgeons: experienced surgeons treat fewer emergent cases. However, controlling for operative mix and emergency status, patient complexity is not a significant component of the difference between new and experienced surgeon performance ($P = 0.43$).
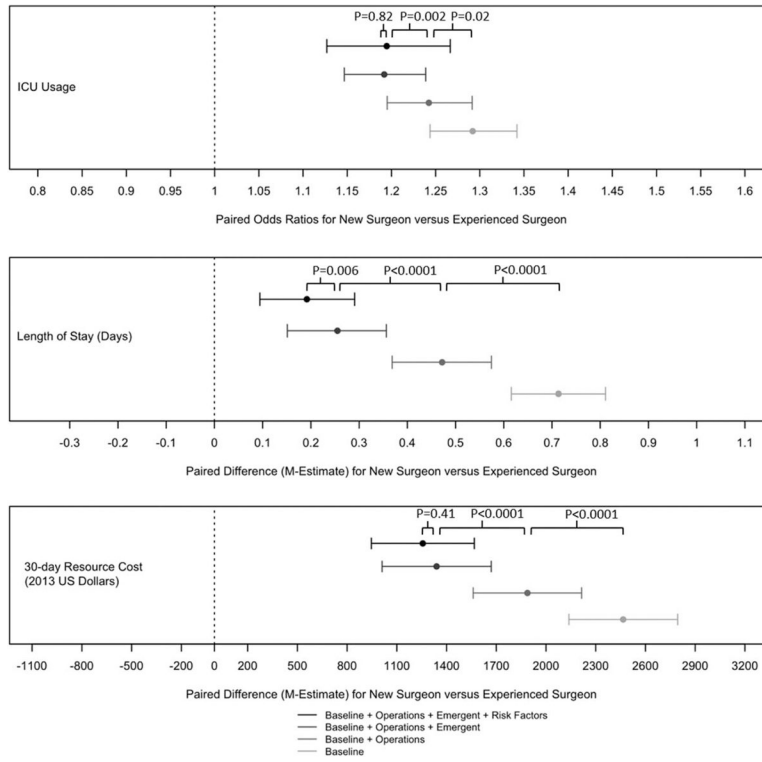
FIGURE 2.

Utilization outcomes. The point estimates and the 95% CIs concern the difference in performance of new and experienced surgeons as the adjustment for the difference in their patient mix becomes more extensive. The *P* values were calculated using the exterior match. As in Figure 1, each *P* value indicates whether a characteristic explains a significant component of the difference in performance of new and experienced surgeons on the indicated outcome measure.

TABLE 1.

Characteristics of the Surgeons and Hospitals in Which New and Experienced Surgeons Practice[‡]

Covariate (% Unless Noted)	Surgeon-patients Before Sampling or Matching			Surgeon-patients After Sampling and Matching		
	New	Experienced	P Value	New	Experienced	P Value
N surgeons	2119	8503		1820	1820	
N patients	68,036	694,165		18,200	18,200	
Surgeon characteristics						
Years of practice at date of surgery	1.78	21.61	<0.0001	1.62	21.30	<0.0001
Operative volume per surgeon (mean)	32.1	81.6	<0.0001	31.2	111.6	<0.0001
Hospital characteristics						
% in major teaching hospitals[*]	17.9	20.7	<0.0001	19.5	19.5	1.0000
% in nonteaching hospitals	46.2	39.4	<0.0001	45.3	45.3	1.0000
% in hospitals with nurse-to-bed 1.0	75.5	79.1	<0.0001	76.0	76.0	1.0000
% with high technology level[†]	61.8	67.8	<0.0001	61.4	61.4	1.0000
mean bed size	358.7	421.3	<0.0001	371.0	371.0	1.0000

[*] Major teaching hospitals are defined as those with resident-to-bed ratios 0.25. Nonteaching hospitals are those with RTB = 0.

[†] High technology level is defined as the provision of cardiothoracic surgery or organ transplantation services.

[‡] Note. The left set of columns represents the practice distribution of all eligible new and experienced surgeons who had at least 10 cases and were practicing in the same hospitals. The right set of columns represents the random selection of 10 patients per new surgeon, and the experienced surgeon patients they were matched to. All covariates are weighted at the patient level.

TABLE 2.

Quality of the Matches for Representative Covariates

| Selected Covariates (percent unless noted) | All New Surgeon Patients (Patients not Matched) | Matched New Surgeon Patients | Experienced Surgeon Patients Matched for: | | | | All Experienced Surgeon Patients After Surgeon Pairing (patients not matched) | All Experienced Surgeon Patients Before Surgeon Pairing (patients not matched) |
			Baseline + Operation + Emerg. + Risk Factors	Baseline + Operation + Emerg.	Baseline + Operation	Baseline		
Number of Surgeons	2,119	1,820	1,820	1,820	1,820	1,820	1,820	8,503
Number of Patients	68,036	18,200	18,200	18,200	18,200	18,200	203,113	694,165
Year of admission (mean)	**2011.4**	**2011.5**	**2011.4**	**2011.5**	**2011.5**	**2011.5**	2011.0	2011.0
Principal Procedure Categories								
Appendectomy	5.1	5.3	5.3	5.3	5.3	3.9	3.9	3.7
Cholecystectomy	31.0	29.8	29.8	29.8	29.8	24.5	25.4	23.1
Hernia Groin - Open	3.4	3.6	3.6	3.6	3.6	3.4	3.2	3.1
Mastectomy	1.3	1.4	1.4	1.4	1.4	3.0	2.6	3.9
Pancreatectomy	0.7	0.7	0.7	0.7	0.7	1.7	1.8	2.5
Small Bowel Resection	8.3	8.3	8.3	8.3	8.3	6.9	6.6	6.1
Stomach Anti-Reflux	0.4	0.3	0.3	0.3	0.3	0.7	1.0	0.9
Thyroidectomy Partial	0.2	0.3	0.3	0.3	0.3	0.6	0.6	1.1
Hip Repair	23.2	29.7	29.7	29.7	29.7	17.1	10.4	9.0
Hip Replacement	46.8	50.1	50.1	50.1	50.1	37.4	35.5	33.4
Hip Replacement Revision	3.7	2.0	2.0	2.0	2.0	3.0	3.5	3.2
Knee Replacement	23.2	16.8	16.8	16.8	16.8	39.8	47.3	51.0
Knee Replacement Revision	3.1	1.5	1.5	1.5	1.5	2.8	3.4	3.5
Emergent Admission	**47.2**	**53.9**	**53.9**	**53.9**	**43.0**	34.7	25.8	23.0
Risk Factors & Patient Demographics								
Age at Admission (years, mean)	78.7	79.0	79.1		78.2	77.2	77.0	76.7
Age 65–69	18.1	17.2	16.9		19.0	21.6	21.6	22.7
Age 85+	24.6	25.8	26.7	24.1	21.9	17.8	16.3	15.0
Past Admission in Last 6 Months	24.0	25.6	24.1	24.5	24.1	22.9	19.3	18.0
Probability of 30-day Death	4.8	5.6	5.5	5.2	4.7	4.2	3.0	2.7
Comorbidity Count (mean)	4.0	4.1	4.0	4.0	4.0	3.9	3.5	3.5

Selected Covariates (percent unless noted)	All New Surgeon Patients (Patients not Matched)	Matched New Surgeon Patients	Experienced Surgeon Patients Matched for:				All Experienced Surgeon Patients After Surgeon Pairing (patients not matched)	All Experienced Surgeon Patients Before Surgeon Pairing (patients not matched)
			Baseline + Operation + Emerg. + Risk Factors	Baseline + Operation + Emerg.	Baseline + Operation	Baseline		
Comorbidities								
Congestive Heart Failure	20.5	22.2	21.4	20.5	19.9	18.2	15.2	14.0
Past Myocardial Infarction	9.9	10.4	9.1	9.9	10.0	9.5	8.3	8.0
Past Arrhythmia	29.2	30.5	29.8	30.3	30.0	29.2	26.9	26.7
Diabetes	31.8	32.5	32.9	31.9	32.0	32.2	30.1	29.8
Renal Dysfunction	22.7	25.2	24.8	22.6	21.5	20.8	16.4	15.5
Renal Failure	5.0	5.6	5.0	5.2	5.1	4.8	3.7	3.5
Cancer	33.6	36.0	37.9	37.4	37.6	38.7	34.4	34.4
Dementia	19.9	21.2	20.5	18.4	16.7	13.8	11.7	10.7
Stroke	14.7	15.9	14.4	15.1	14.1	13.0	11.0	10.5

See eTable 6 for complete list of covariates and eTables 7–10 for the balance of the individual matches.

Outcomes

TABLE 3.

Outcome (Percent Unless Noted)	Cohort	Point Estimate	Odds Ratio or Paired Difference (95% CI)	New Versus Exp. *P* Value
30-d all location mortality				
	New Surgeons	6.2	-	-
	Exp. Surg Baseline + Opx + Emerg + Risk Factors	5.9	1.06 (0.97, 1.16)	0.2391
	Exp. Surg Baseline + Opx + Emerg	5.6	1.12 (1.05, 1.19)	0.0007
	Exp. Surg Baseline + Opx	5.1	1.24 (1.16, 1.32)	<0.0001
	Exp. Surg Baseline	4.5	1.42 (1.33, 1.52)	<0.0001
30-d failure-to-rescue				
	New Surgeons	9.5	-	-
	Exp. Surg Baseline + Opx + Emerg + Risk Factors	9.3	0.99 (0.89, 1.10)	0.8725
	Exp. Surg Baseline + Opx + Emerg	9.0	1.07 (0.99, 1.16)	0.1127
	Exp. Surg Baseline + Opx	8.4	1.19 (1.10, 1.30)	<0.0001
	Exp. Surg Baseline	7.9	1.27 (1.17, 1.39)	<0.0001
30-d readmission or death				
	New Surgeons	19.4	-	-
	Exp. Surg Baseline + Opx + Emerg + Risk Factors	18.6	1.06 (1.01, 1.12)	0.0305
	Exp. Surg Baseline + Opx + Emerg	18.2	1.09 (1.05, 1.13)	<0.0001
	Exp. Surg Baseline + Opx	17.3	1.16 (1.12, 1.21)	<0.0001
	Exp. Surg Baseline	16.3	1.25 (1.20, 1.30)	<0.0001
Prolonged length of stay				
	New Surgeons	56.9	-	-
	Exp. Surg Baseline + Opx + Emerg + Risk Factors	54.5	1.14 (1.08, 1.19)	<0.0001
	Exp. Surg Baseline + Opx + Emerg	53.8	1.14 (1.11, 1.18)	<0.0001
	Exp. Surg Baseline + Opx	50.6	1.31 (1.27, 1.35)	<0.0001
	Exp. Surg Baseline	47.1	1.51 (1.47, 1.56)	<0.0001
Anesthesia time (min, m-est.)				
	New Surgeons	155.4	-	-
	Exp. Surg Baseline + Opx + Emerg + Risk Factors	137.6	18.7 (17.4, 20.1)	<0.0001
	Exp. Surg Baseline + Opx + Emerg	134.9	18.5 (17.1, 19.8)	<0.0001

Outcome (Percent Unless Noted)	Cohort	Point Estimate	Odds Ratio or Paired Difference (95% CI)	New Versus Exp. P Value
	Exp. Surg Baseline + Opx	137.7	16.4 (15.1, 17.8)	<0.0001
	Exp. Surg Baseline	142.6	12.5 (11.1, 13.9)	<0.0001
ICU usage	New Surgeons	22.8	-	-
	Exp. Surg Baseline + Opx + Emerg + Risk Factors	20.5	1.19 (1.13, 1.27)	<0.0001
	Exp. Surg Baseline + Opx + Emerg	20.1	1.19 (1.15, 1.24)	<0.0001
	Exp. Surg Baseline + Opx	19.5	1.24 (1.20, 1.29)	<0.0001
	Exp. Surg Baseline	18.9	1.29 (1.24, 1.34)	<0.0001
Length of stay (d, m-est.)	New Surgeons	7.3	-	-
	Exp. Surg Baseline + Opx + Emerg + Risk Factors	7.2	0.2 (0.1, 0.2)	0.0002
	Exp. Surg Baseline + Opx + Emerg	7.1	0.2 (0.1, 0.3)	<0.0001
	Exp. Surg Baseline + Opx	6.9	0.4 (0.3, 0.5)	<0.0001
	Exp. Surg Baseline	6.6	0.6 (0.5, 0.7)	<0.0001
30-day resource cost ($, m-est.)	New Surgeons	22,404	-	-
	Exp. Surg Baseline + Opx + Emerg + Risk Factors	20,998	1257 (946, 1568)	<0.0001
	Exp. Surg Baseline + Opx + Emerg	20,857	1340 (1011, 1670)	<0.0001
	Exp. Surg Baseline + Opx	20,166	1888 (1561, 2215)	<0.0001
	Exp. Surg Baseline	19,477	2466 (2138, 2795)	<0.0001

See eTable 11 for Outcomes by Specialty.

Number of matched pairs = 18,200. Baseline = matched for year; Opx = matched for operation; Emerg = matched for emergent status; Risk factors = matched for patient demographics comorbidity risk factors.

Binary outcomes were compared using the McNemar test while continuous outcomes were tested using M-statistics.

Cost is measured in 2013 USD.