Comparison of postoperative outcomes among patients treated by male and female surgeons: a population based matched cohort study

Christopher JD Wallis resident\(^1\), Bheeshma Ravi surgeon and assistant professor\(^3\), Natalie Coburn surgeon and associate professor\(^4\), Robert K Nam surgeon and professor\(^1\), Allan S Detsky internist and professor\(^2\)\(^5\), Raj Satkunasivam surgeon and assistant professor\(^1\)\(^6\)

\(^1\)Division of Urology, Sunnybrook Health Sciences Centre, University of Toronto, ON M4N 3M5, Canada; \(^2\)Institute of Health Policy, Management, and Evaluation, University of Toronto; \(^3\)Division of Orthopedic Surgery, Sunnybrook Health Sciences Centre; \(^4\)Division of General Surgery, Sunnybrook Health Sciences Centre; \(^5\)Department of Medicine, Mount Sinai Hospital, University Health Network, University of Toronto; \(^6\)Department of Urology and Center for Outcomes Research, Houston Methodist Hospital, Houston, TX, USA.

Abstract

Objective To examine the effect of surgeon sex on postoperative outcomes of patients undergoing common surgical procedures.

Design Population based, retrospective, matched cohort study from 2007 to 2015.

Setting Population based cohort of all patients treated in Ontario, Canada.

Participants Patients undergoing one of 25 surgical procedures performed by a female surgeon were matched by patient age, patient sex, comorbidity, surgeon volume, surgeon age, and hospital to patients undergoing the same operation by a male surgeon.

Interventions Sex of treating surgeon.

Main outcome measure The primary outcome was a composite of death, readmission, and complications. We compared outcomes between groups using generalised estimating equations.

Results 104 630 patients were treated by 3314 surgeons, 774 female and 2540 male. Before matching, patients treated by female doctors were more likely to be female and younger but had similar comorbidity, income, rurality, and year of surgery. After matching, the groups were comparable. Fewer patients treated by female surgeons died, were readmitted to hospital, or had complications within 30 days (5810 of 52 315, 11.1%, 95% confidence interval 10.9% to 11.4%) than those treated by male surgeons (6046 of 52 315, 11.6%, 11.3% to 11.8%; adjusted odds ratio 0.96, 0.92 to 0.99, P=0.02). Patients treated by female surgeons were less likely to die within 30 days (adjusted odds ratio 0.88; 0.79 to 0.99, P=0.04), but there was no significant difference in readmissions or complications. Stratified analyses by patient, physician, and hospital characteristics did not significant modify the effect of surgeon sex on outcome. A retrospective analysis showed no difference in outcomes by surgeon sex in patients who had emergency surgery, where patients do not usually choose their surgeon.

Conclusions After accounting for patient, surgeon, and hospital characteristics, patients treated by female surgeons had a small but statistically significant decrease in 30 day mortality and similar surgical outcomes (length of stay, complications, and readmission), compared with those treated by male surgeons. These findings support the need for further examination of the surgical outcomes and mechanisms related to physicians and the underlying processes and patterns of care to improve mortality, complications, and readmissions for all patients.

Introduction

Successful surgical practice has four core components: knowledge, communication skills, judgment, and technical proficiency. The acquisition and maintenance of technical skills distinguishes surgeons from many other doctors, and these skills are directly associated with short term postoperative outcomes. Women and men practise medicine differently, although little research exists on the differences in learning styles, acquisition of skills, or outcomes for female and male surgeons. Tsugawa et al found that beneficiaries of US Medicare that were treated by female general internists in hospital had lower rates of 30 day mortality and readmission than those treated by male internists. Suggested mechanisms for this difference include female doctors being more likely to use a patient centred approach and to follow evidence based guidelines. Surgery, however, has a major technical component, so there is less reason to expect a difference in outcomes between female and male surgeons.

Surgical disciplines are disproportionately male despite increasing numbers of female medical students. Gender equity
in the surgical profession, including disparities in compensation and promotion, is a growing concern. Assessing outcomes for female and male surgeons is important for combating implicit bias and gender schemas that might perpetuate current inequalities. We sought to determine whether postoperative outcomes for patients having one of 25 surgical procedures (emergent and elective) differ between male and female operating surgeons using a large, population based cohort.

**Methods**

**Overview**

We conducted a population based, retrospective cohort study of patients having surgical procedures in Ontario, Canada, between 1 January 2007 and 31 December 2015 to assess the hypothesis that sex of the operating surgeon would significantly affect 30 day postoperative outcomes. The study sample included patients who had one of 25 surgical interventions, selected by multidisciplinary discussion with consultants from all surgical specialties. We selected surgical procedures using the following criteria: inclusion of all surgical subspecialties with female surgeons and either frequently performed in Ontario or having an increased likelihood of complications. We selected coronary artery bypass grafting, femoral-popliteal bypass, abdominal aortic aneurysm repair, appendectomy, cholecystectomy, gastric bypass, colon resection, liver resection, hysterectomy, anterior or posterior spinal decompression, anterior or posterior spinal arthrodesis, craniootomy for brain tumour, total knee replacement, total hip replacement, open repair of femoral neck or shaft fracture, total thyroidectomy, neck dissection, lung resection, radical cystectomy, radical prostatectomy, transurethral resection of prostate, carpal tunnel release, and breast reduction.

Despite using administrative data sources, we performed a sample size calculation to ensure that the proposed analysis was feasible. Based on previous estimates of the study primary outcome (composite event rate of death, readmission, and postoperative complications; 17%) and effect size (4%, or odds ratio 0.96), we calculated a necessary sample size of 94 270 (see appendix). We designed and conducted this study according to STROBE (strengthening the reporting of observational studies in epidemiology) guidelines and the RECORD (reporting of studies conducted using observational routinely collected health data) statement. Sunnybrook Health Sciences Centre Research Ethics Board approved this study (project identification number 375-2016).

**Data sources**

We linked the following datasets using encrypted patient identifiers: the Ontario Health Insurance Plan (OHIP) database, which tracks claims paid for physician billings, laboratories, and out-of-province providers; the Canadian Institute for Health Information Discharge Abstract Database (CIHI-DAD), which contains records for hospital admissions; the CIHI National Ambulatory Care Reporting System, which contains records for emergency department visits; the Registered Persons Database for demographic information; and the Corporate Provider Database for surgeon data.

**Cohort derivation**

We identified all patients who had one of the 25 index procedures in the study period (n=1 534 592). We excluded patients treated by physicians whose primary declared specialty was non-surgical (n=8753) and patients under 18 (n=29 158). We included only the first one of the 25 procedures for each patient, thus excluding 282 399 procedures in patients already in the cohort and ensuring that each patient was included only once. We excluded patients for whom the date of death preceded the date of surgery (n=332). We think these cases represent coding errors as the majority died several years before the date of surgery. A sensitivity analysis including the 10 patients who died one day before surgery did not alter the outcome of the primary analysis (see appendix). To capture data on hospital factors, we excluded patients for whom the treating institution could not be identified (n=54 263). After these exclusions, the study sample comprised 1 159 687 patients who had surgery between 2007 and 2015.

**Outcomes**

Our primary outcome was a composite of death, complications, or readmission (to any hospital in the province of Ontario) in the 30 days after surgery, as previously described. We included deaths that occurred in the initial postoperative period in hospital. We think this outcome best captures the overall burden of short term postoperative complications. Our secondary outcomes were the individual components of the primary outcome (death, complications, and readmission within 30 days of surgery), and length of stay in hospital. We used a definition of surgical complications that has previously been used and represents major morbidity. These outcomes were ascertained from health administrative data using a combination of procedural and diagnostic codes that are uniformly collected for all hospitals and patients in Ontario.

**Covariates**

We collected patient age at surgery, geographic location (local health integration networks), sex, socioeconomic status (based on geographic location), rurality, and general comorbidity according to the Johns Hopkins aggregate disease group score (ADG) from linked administrative databases. The ADG score discriminates better than the Charlson score. We identified data on the treating surgeons including age, sex, years in practice, specialty, and surgical volume. Surgeon sex refers to biological sex, as reported by physicians at the time of registering for licenses to practise medicine in the province of Ontario. Patients’ biological sex is captured in the Registered Persons Database, which we used for demographic information. We were, therefore, unable to identify transgendered patients or surgeons. To determine surgical volume, for each patient and for the specific procedure they had, we identified the number of identical procedures their operating surgeon performed in
the previous year. We defined this variable categorically into quarters for each procedure performed, among the whole cohort. We collected hospital institution identifiers to account for variability between institutions. We defined a surgical procedure as emergent or elective using the CIHI-DAD database admission variables, which denote urgent, emergent, or admission from the emergency department before surgery. We considered all same day surgery procedures to be elective.

Matching
Surgeon age, years in practice, operative volumes, surgical specialty, and patient demographics differed between participants treated by female and male surgeons, so we conducted a matched analysis. Variables such as surgeon age could potentially confound the relationship between surgeon sex and outcomes, as male surgeons were older on average. Far fewer patients were treated by female surgeons, so we identified these first and matching them to patients treated by male surgeons. We matched patients 1:1 using a hard match comprising procedural fee code (a unique identifier for each procedure), surgeon volume (separated into quarters based on the number of index procedures performed by each surgeon in the year before index), surgeon age (±3 years), hospital identifier, patient age (±5 years), patient sex, and patient comorbidity using ADG score (categorical). Surgeon years in practice was collinear with surgeon age, so we selected only age for matching. We selected all covariates for matching based on a priori identification without variable selection.

Statistical analysis
In all analyses we examined surgeon sex as an exposure potentially associated with postoperative outcomes. Descriptive statistics were used to compare patients treated by male and female physicians, as well as to compare the characteristics of the male and female surgeons. Given the large sample size, traditional statistical measures were likely to show statistically significant differences where no clinically important differences exist. So we compared groups using standardised differences—the difference in the mean of a variable between two groups divided by an estimate of the standard deviation of that variable among both groups.\(^{23}\) We defined a clinically important standardised difference as greater than 0.10.\(^{23}\) Matched data are inherently correlated, so we used multivariable generalised estimating equations (GEE) with a logit link to estimate the association between surgeon sex and the outcomes. We used Poisson regression to examine the association between surgeon sex and hospital length of stay.

We conducted subgroup analyses to assess for an interaction between patient (age, sex, and comorbidity), surgeon (specialty, age, years in practice, and annual volume), and hospital characteristics and to assess the association between surgeon sex and outcomes. To minimise the effect of patients selecting their physician, we performed a retrospective stratified analysis of the matched cohort according to whether the surgery was elective or emergent.

Matching considerably reduced our sample size, which might adversely affect the applicability of our findings. Thus, we examined the entire cohort (n=1 159 687) using regression with multivariable GEE models with a logit link, rather than matching, to account for differences in patient, surgeon, and hospital factors between patients treated by female and male surgeons. The GEE model accounted for correlation in outcomes based on surgeon, hospital, and procedural fee code. Statistical significance was set at P<0.05 based on a two tailed comparison. We used SAS Enterprise Guide 6.1 (SAS Institute Inc., Cary, NC, USA) for all analyses.

Patient involvement
No patients were involved in setting the research question or the outcome measures, nor were they involved in developing plans for implementation of the study. No patients were asked to advise on interpretation or writing up of results. There are no plans to disseminate the results of the research to study participants or the relevant patient community. Individual patient consent was not sought given the use of anonymised administrative data.

Results
We identified 1 159 687 eligible patients who had one of the 25 index procedures in the study period. Surgery was performed by 3314 surgeons, 774 (23.4%) of whom were female, and 2540 (76.6%) were male (table 1). Of the 1 159 687 patients, 144 119 (12.4%) were treated by female surgeons, and 1 015 568 (87.6%) were treated by male surgeons. Patients treated by female physicians were more likely to be female and younger (table 1). We found no difference in year of surgery, comorbidity, income, rurality, or region of residence between groups. Female surgeons were younger and had fewer years in practice. Surgeon volume (based on the number of identical surgical procedures that surgeon had performed in the preceding year) was lower for female surgeons. Female surgeons performed proportionally more operations than men in general surgery, obstetrics and gynaecology, and plastic surgery. Before matching, patients treated by female surgeons had statistically significantly lower rates of the composite endpoint (15 731 of 144 119, 10.9%; 10.8% to 11.1%) than those treated by male surgeons (122 468 of 1 015 568, 12.1%; 12.0% to 12.1%; unadjusted odds ratio 0.94, 0.91 to 0.97, P<0.001).

After matching, the two groups were well balanced with respect to patient age, sex, comorbidity, rurality, and income; surgeon age, years in practice, annual volume of the index procedure, and specialty; and hospital type. More patients treated by female surgeons had their operation in later calendar years, although the difference was not clinically important (standardized difference <0.10).

Differences in surgical specialty, procedural volume, and age of female surgeons meant that patients who were successfully matched to male surgeons (and, thus, included in the primary analytical cohort) were treated by younger surgeons with less experience and lower annual volumes than those who were excluded (supplementary table 2). Included patients were more likely to have had general surgery or an obstetrics and gynaecology related procedure and less likely to have had neurosurgery, orthopaedic surgery, or a urology related procedure. Included patients were younger and more likely to be female; comorbidity did not differ (supplementary table 2).

In the matched cohort, the primary outcome of death, readmission, or complications occurred in 3810 of 52 315 (11.1%, 10.9% to 11.4%) patients treated by female surgeons and in 6046 of 52 315 (11.6%, 11.3% to 11.8%) patients treated by male surgeons (table 2), translating to an absolute difference of 0.43% (number needed to treat to prevent one event=230).

In GEE models accounting for the correlation between matched individuals, patients treated by female surgeons were significantly less likely to experience the composite outcome (adjusted odds ratio 0.96, 0.92 to 0.99, P=0.02). Among the
secondary outcomes, patients treated by female surgeons had a significantly lower likelihood of death within 30 days of surgery (adjusted odds ratio 0.88, 0.79 to 0.99, P=0.04) and comparable likelihood of readmission to hospital (0.96, 0.91 to 1.02, P=0.20) or complications (0.96, 0.92 to 1.01, P=0.10). Median length of stay was two days (interquartile range 0-4 days), regardless of the sex of the treating physician. As the mean duration of hospital admission differed among patients treated by female surgeons and those treated by male surgeons (mean 3.9 days (SD 10.0) and mean 4.0 days (SD 11.1), respectively), this was statistically significant in adjusted models (table 2).

We performed a retrospective sensitivity analysis of all outcomes in the matched cohort, stratifying by elective or emergent surgery. Patients who had emergent procedures were less likely to be female and more likely to have these procedures performed by younger surgeons with lower surgical volumes who had practiced for a shorter period of time (supplementary table 3). Specialties wherein emergent procedures are common (such as general surgery) were proportionately more represented than those where it is uncommon (such as obstetrics and gynaecology and plastic surgery), as expected. Emergent procedures were not more or less likely to be performed by a female surgeon. Using GEE models accounting for the correlation between matched individuals, patients treated by female surgeons were significantly less likely to experience the composite outcome when the procedure was elective (adjusted odds ratio 0.94, 0.89 to 0.98, P=0.007) but not emergent (adjusted odds ratio 1.01, 0.96 to 1.08, P=0.636; Pinteraction=0.048). We did not find any evidence of statistically significant effect modification by the urgency of the procedure for secondary outcomes, except length of stay (supplementary table 4).

We found no significant evidence of effect modification when analyses were stratified according to surgical specialty, surgeon age, surgeon years in practice, annual surgical volume of the index procedure, hospital status (academic or community), patient sex, patient age, or patient comorbidity (fig 1.). Although the effect estimates remained relatively consistent, many comparisons were no longer statistically significant in these subgroup analyses; for example, differences based on surgical specialty were not statistically significant (P=0.17). Given the dramatic effect of surgeon sex on outcomes of plastic surgical procedures (fig 1); GEE regression model adjusted odds ratio 0.24, 0.17 to 0.36, supplementary table 5), we repeated our primary analysis excluding plastic surgery. The direction and magnitude of the adjusted relative effects were unchanged (adjusted odds ratio of composite outcome 0.963, 0.928 to 0.999, P=0.045; adjusted odds ratio of mortality 0.886, 0.787 to 0.998, P=0.0465).

Matching considerably reduced our sample size, which might adversely affect the extrapolation of these findings to our entire patient cohort. In regression models based to account for confounding, patients treated by female surgeons had a comparably lower likelihood of the composite primary outcome (adjusted odds ratio 0.96, 0.93 to 0.99, P=0.006); the other findings, including mortality, were also comparable to the matched analysis (supplementary table 5). Notably, the year of surgery was not significantly associated with primary outcome in the regression analysis. To explore the degree to which case-mix variation and surgical volume affected the study conclusions, we examined outcomes in regression models with sequentially greater numbers of covariates. We clustered observations based on the procedural fee code for each model to compare patients undergoing the same operation. Comparing patients treated by female surgeons to those treated by male surgeons, in unadjusted models, the crude odds ratio of the composite endpoint was 0.94 (0.91 to 0.97, P=0.0004). It remained 0.94 (0.90 to 0.97, P=0.0002) after accounting for surgeon age. Finally, we accounted for case-mix and all variables except for surgical volume, and the difference between female and male surgeons became non-significant (adjusted odds ratio 0.97, 0.94 to 1.00, P=0.08).

Discussion

Principal findings

In this population based cohort, we found small but significant differences in postoperative complications (mortality, complications, and readmission) between patients treated by male and female surgeons, using an α criterion of 0.05. Patients treated by female surgeons had lower rates of 30 day mortality. We matched patients for age, sex, and general comorbidity that had the same surgical procedure by a female or male surgeon of the same age with comparable annual surgical volume in the same hospital. Male surgeons in this cohort had higher average procedural volumes, which mitigates some of the effect of surgeon sex. Outcomes adjusted for case-mix (without adjustment for surgical volume) were comparable for patients treated by female and male surgeons. Retrospective analyses showed that the effect of surgeon sex was restricted to patients who had elective operations; outcomes did not differ among patients who had emergent operations. This finding must be interpreted within the limitations of retrospective analyses.24 A large proportion of patients were excluded because of the inability to find a suitable match. This was most prominent among patients who had procedures in specialties dominated by men. But regression models accounting for baseline differences with adjustment rather than matching gave comparable results for the same outcomes of interest. This indicates that the matched analytic approach might provide results that can be generalised.

We found an association between improved 30 day postoperative outcomes and female surgeons, but whether these differences would contribute to long term survival is unclear. We focused on short term outcomes, as others have previously done,14,15 because technical factors related to the surgeon are more likely to contribute to complications in the immediate postoperative period rather than to long term surgical outcomes.9 10

We don't know the mechanism that underlies better outcomes for patients treated by female surgeons, although it might be related to delivery of care that is more congruent with guidelines, more patient centred, and involves superior communication.15 We found that improved postoperative outcomes for patients treated by female surgeons were restricted to patients who had elective operations, which might reflect better patient selection for surgery or residual confounding. Other possible explanations include a willingness to collaborate by female surgeons, which might avert scenarios that could otherwise result in the “failure to rescue” phenomenon.25 In a relatively small cohort of medical students (n=342), Lou et al found that female students completed tests of basic surgical skills more efficiently and had significantly higher scores on tests of theoretical surgical knowledge than men.26 Further study is required to understand how differences in medical practice, clinical acumen, technical skills,27 or risk taking behaviour28 could underlie the sex based differences in mortality that we found, with a view to improving outcomes for patients treated by physicians of either sex. The confidence intervals of the effect estimate approach might represent residual confounding or be due to chance alone. Our
retrospective analyses of emergent operations, where patients are less likely able to select or be biased towards a surgeon of a particular sex, partly support the possibility of confounding. We found that the benefit in composite outcome for patients treated by female surgeons was restricted to elective procedures. By contrast, patients treated by female surgeons did not have a lower rate of 30 day mortality for either elective or emergent procedures, although this subanalysis had lower power than our main analysis. Regardless, outcomes for patients treated by female surgeons were never worse than for those treated by male surgeons.

The Canadian healthcare system is entirely publicly funded, which must be considered when interpreting these results. Patients can freely choose their operating surgeon, so factors such as lower marketing at patients might lower the likelihood of patients “shopping around” for surgeons compared with other healthcare systems. Whether factors such as this affect the likelihood of selection bias whereby “favourable” patients preferentially seek female surgeons in Canada is unknown. Women undertaking a surgical career experience barriers in training and in practice. Previous work has shown that the “surgical personality,” surgical culture, and sex based discrimination, as well as lifestyle factors and workload, are deterrents for women considering a career in surgery. Difficulty identifying mentors, owing to the lack of women in leadership positions in surgery, might exacerbate this problem. These barriers might create a higher standard for women to gain entrance into the surgical workforce than men, resulting in the selection of a cohort of women that are proportionately more skilled, motivated, and harder working.

Strengths and limitations of this study
We were able to comprehensively capture patients undergoing the selected surgical procedures because of single payer healthcare in Ontario. Unlike Tsugawa et al, who were limited to Medicare beneficiaries aged 65 years and older, we identified all adult patients who had the procedures of interest. This is particularly important because patients treated by female physicians were younger than those treated by male surgeons. Moreover, the nature of the healthcare system enabled us to capture readmissions and complications occurring anywhere in the province, including at hospitals other than where the index surgery was performed.

The observational nature of our study carries some limitations. Primarily, we are unable to account for unmeasured (and unmeasurable) residual confounding. Differences in socioeconomics and comorbidity, which were not measured by our data sources, might have contributed to our findings. We used hard matching to directly compare outcomes between patients treated by female and male surgeons, but within group heterogeneity from the categorical definitions of surgical volume and comorbidity might have caused residual confounding. Finally, we did not have information on the severity of the diseases for which surgery was performed (for example, cancer stage) or case complexity, though procedures were matched directly.

Comparison with other studies
The association between sex of treating physician and patient mortality has now been assessed in three settings—after surgery, among older patients admitted to an internal medicine service, and in outpatient care. Jerant et al found no association between the treating physician’s sex and mortality in an outpatient setting, but their study was conducted in a smaller cohort (roughly 20,000) of relatively healthy patients. By contrast, Tsugawa et al found small, but meaningful, differences in mortality based on sex of the treating physician.

Conclusions and implications
Using a comprehensive sample of patients undergoing a broad range of surgeries in all relevant specialties in Ontario, Canada, we found small differences in postoperative outcomes between patients treated by female and male surgeons, with those treated by female surgeons having a statistically significantly lower rate of 30 day mortality.

Although the relative difference in mortality is quite modest (4%), it has potentially significant clinical implications. Tsugawa et al found a 4% relative risk reduction in mortality for patients treated by female internists, compared with those treated by male internists. In our primary matched analysis, we found a similar 4% relative lower likelihood of the composite outcome (adjusted odds ratio 0.96) and a 12% lower likelihood of 30 day mortality (adjusted odds ratio 0.88) among patients treated by female surgeons.

That physician sex might have a greater effect on mortality rates in patients having surgery than those treated by internists is not surprising. Outcomes after surgery are probably more dependent on individual physician actions than those admitted to general internal medicine departments, which, proportionally, might depend more on patient characteristics. Moreover, much of the care provided by an inpatient internal medicine service is performed by a team. Tsugawa et al identified the primary treating physician as the individual who billed the largest proportion of services, which, on average, represented half of each patient’s billings. By contrast, although residents, fellows, hospitalists, and nurses participate in the operative and postoperative care of patients having surgery, the primary surgeon has a direct effect on patient outcomes. This has been found when assessing the effect of surgeon volume, specialisation, and skill. Surgeons are not completely independent of teams, so the differences we found could not only reflect differences in surgeon sex but also how they use the resources available to them (such as other consultants or intensive care units). Because we compared surgeons at the same hospital, who shared the same resources, mortality differences are less likely to be related to system or team based factors. The buffering effect of surgical teams should result in a bias towards the null. So the effect we observed might be an underestimate of the true effect. These results do not support the preferential selection of a surgeon of either sex in clinical practice. Rather they indicate the need for further study in independent cohorts using different research methods. In addition, they support the examination of surgical outcomes and mechanisms related to physicians and the underlying processes and patterns of care to improve mortality, complications, and readmissions for all patients. Our findings have important implications for supporting sex equality and diversity in a traditionally male dominated profession.

We thank Refik Saskin and Lisa Ellison for their methodological and analytical support.

Contributors and sources: CJDW and RS conceived the study and were responsible for the design and development of the data analysis. CJDW, BR, ASD and RS were responsible for data assembly, collection, and analysis. CJDW, BR, NC, RKN, ASD, and RS were responsible for data interpretation. CJDW wrote the first draft. All authors provided significant revisions for important intellectual content and approved the final version. CJDW and RS affirm that the manuscript is an honest,
accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

Funding: This study received no support from any organisation. CJDW is supported by the Canadian Institute of Health Research Banting and Best Doctoral Award. NGC is supported by the Sherif and MaryLou Hanna Chair in Surgical Oncology. RKN is supported by the Ajmera Family Chair in Urologic Oncology.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Data sharing: Statistical code is available from the RS (raj.satkunasivam@gmail.com). Given the administrative nature of the data, patients did not give informed consent for data sharing but all data are fully anonymised, and risk of identification is low.

Disclaimer: This study made use of de-identified data from the ICES Data Repository, which is managed by the Institute for Clinical Evaluative Sciences with support from its funders and partners: Canada’s Strategy for Patient-Oriented Research (SPOR), the Ontario SPOR Support Unit, the Canadian Institutes of Health Research and the Government of Ontario. The opinions, results and conclusions reported are those of the authors. No endorsement by ICES of any of its funders or partners is intended or should be inferred.

## Tables

### Table 1 | Baseline characteristics of patients included in the analysis

<table>
<thead>
<tr>
<th>Baseline characteristics of patients included in the analysis</th>
<th>Before matching</th>
<th>After matching</th>
<th>St diff</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Treated by female surgeon</td>
<td>Treated by male surgeon</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Number of patients</td>
<td>144 119</td>
<td>1 015 568</td>
</tr>
<tr>
<td>Surgeon characteristics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age, years:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>44.2 (8.1)</td>
<td>49.3 (9.6)</td>
<td>0.57</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>43 (38-50)</td>
<td>49 (41-57)</td>
<td>0.55</td>
</tr>
<tr>
<td>Years in practice:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>11.6 (8.0)</td>
<td>15.4 (8.4)</td>
<td>0.46</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>10 (5-19)</td>
<td>17 (8-22)</td>
<td>0.45</td>
</tr>
<tr>
<td>Surgical volume among all patients, n (%):</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First quarter (lowest)</td>
<td>52 730 (36.6)</td>
<td>232 356 (22.9)</td>
<td>0.22</td>
</tr>
<tr>
<td>Second quarter</td>
<td>43 264 (30.0)</td>
<td>256 528 (25.3)</td>
<td>0.26</td>
</tr>
<tr>
<td>Third quarter</td>
<td>29 203 (20.3)</td>
<td>257 512 (25.4)</td>
<td>0.14</td>
</tr>
<tr>
<td>Fourth quarter (highest)</td>
<td>18 902 (13.1)</td>
<td>269 172 (26.5)</td>
<td>0.36</td>
</tr>
<tr>
<td>Surgical specialty:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardiothoracic surgery</td>
<td>5727 (3.7)</td>
<td>63 494 (6.3)</td>
<td>0.36</td>
</tr>
<tr>
<td>General surgery</td>
<td>57 814 (40.1)</td>
<td>326 911 (32.2)</td>
<td>0.28</td>
</tr>
<tr>
<td>Neurosurgery</td>
<td>2370 (1.6)</td>
<td>44 264 (4.4)</td>
<td>0.02</td>
</tr>
<tr>
<td>Obstetrics/Gynaecology</td>
<td>48 301 (33.5)</td>
<td>86 641 (8.5)</td>
<td>0.13</td>
</tr>
<tr>
<td>Orthopaedic surgery</td>
<td>11 400 (7.9)</td>
<td>319 634 (31.5)</td>
<td>0.49</td>
</tr>
<tr>
<td>Otolaryngology</td>
<td>2879 (2.0)</td>
<td>23 529 (2.3)</td>
<td>0.03</td>
</tr>
<tr>
<td>Plastic surgery</td>
<td>14 123 (9.8)</td>
<td>52 719 (5.2)</td>
<td>0.39</td>
</tr>
<tr>
<td>Thoracic surgery</td>
<td>1079 (0.7)</td>
<td>11 075 (1.1)</td>
<td>0.02</td>
</tr>
<tr>
<td>Urology</td>
<td>862 (0.6)</td>
<td>83 456 (8.2)</td>
<td>0.02</td>
</tr>
<tr>
<td>Vascular surgery</td>
<td>20 (0)</td>
<td>3845 (0.4)</td>
<td>0.03</td>
</tr>
<tr>
<td>Patient characteristics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age, years:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>52.7 (16.5)</td>
<td>59.7 (17.2)</td>
<td>0.42</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>51 (41-65)</td>
<td>61 (48-73)</td>
<td>0.44</td>
</tr>
<tr>
<td>Sex, n (%):</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>111 705 (77.5)</td>
<td>584 042 (57.5)</td>
<td>0.44</td>
</tr>
<tr>
<td>Male</td>
<td>32 414 (22.5)</td>
<td>431 526 (42.5)</td>
<td>0.44</td>
</tr>
<tr>
<td>Comorbidity, n (%):</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADG 0-5</td>
<td>39 761 (27.6)</td>
<td>270 369 (26.6)</td>
<td>0.02</td>
</tr>
<tr>
<td>ADG 5-7</td>
<td>35 984 (25.0)</td>
<td>243 640 (24.0)</td>
<td>0.02</td>
</tr>
<tr>
<td>ADG 8-10</td>
<td>43 520 (30.2)</td>
<td>303 605 (29.9)</td>
<td>0.02</td>
</tr>
<tr>
<td>ADG ≥11</td>
<td>24 854 (17.2)</td>
<td>197 954 (19.5)</td>
<td>0.02</td>
</tr>
<tr>
<td>Rurality, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urban</td>
<td>125 309 (88.9)</td>
<td>858 825 (84.6)</td>
<td>0.07</td>
</tr>
<tr>
<td>Rural</td>
<td>18 680 (13.0)</td>
<td>155 864 (15.3)</td>
<td>0.02</td>
</tr>
<tr>
<td>Missing</td>
<td>130 (0.1)</td>
<td>859 (0.01)</td>
<td>0.01</td>
</tr>
<tr>
<td>Income, n (%):</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First fifth (lowest)</td>
<td>26 184 (18.2)</td>
<td>190 796 (18.8)</td>
<td>0.01</td>
</tr>
<tr>
<td>Second fifth</td>
<td>28 094 (19.5)</td>
<td>202 670 (20.0)</td>
<td>0.02</td>
</tr>
<tr>
<td>Third fifth</td>
<td>28 762 (20.0)</td>
<td>202 843 (20.0)</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>Before matching</td>
<td>After matching</td>
<td></td>
</tr>
<tr>
<td>--------------------</td>
<td>-----------------</td>
<td>----------------</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Treated by female surgeon</td>
<td>Treated by male surgeon</td>
<td>St diff</td>
</tr>
<tr>
<td>Fourth fif (5)th</td>
<td>30 833 (21.0)</td>
<td>211 167 (20.8)</td>
<td>11 109 (21.2)</td>
</tr>
<tr>
<td>Fifth fif (5)th (highest)</td>
<td>29 698 (20.6)</td>
<td>203 627 (20.1)</td>
<td>10 358 (19.8)</td>
</tr>
<tr>
<td>Missing</td>
<td>548 (0.4)</td>
<td>4465 (0.4)</td>
<td>206 (0.4)</td>
</tr>
</tbody>
</table>

Hospital characteristics

Practice setting, \(n\) (%):

| Academic       | 54 801 (38.0) | 344 027 (33.9) | 20 878 (39.9) | 20 878 (39.9) | 0.00 |
| Community       | 89 318 (62.0) | 671 541 (66.1) | 31 437 (60.1) | 31 437 (60.1) | 0.00 |

Year of surgery, \(n\) (%):

| 2007            | 14 481 (10.0) | 128 186 (12.6) | 5230 (10.0)   | 6906 (13.2)   | 0.05 |
| 2008            | 15 224 (10.6) | 121 318 (11.9) | 5346 (10.2)   | 6292 (12.0)   | 0.05 |
| 2009            | 15 432 (10.7) | 117 545 (11.6) | 5579 (10.7)   | 5675 (11.2)   | 0.05 |
| 2010            | 15 363 (10.7) | 112 699 (11.1) | 5467 (10.5)   | 5690 (10.9)   | 0.05 |
| 2011            | 15 265 (10.6) | 111 693 (11.0) | 5382 (10.3)   | 5624 (10.8)   | 0.05 |
| 2012            | 15 759 (10.9) | 108 527 (10.7) | 5550 (10.6)   | 5210 (10.0)   | 0.05 |
| 2013            | 16 688 (11.6) | 108 251 (10.7) | 6122 (11.7)   | 5694 (10.9)   | 0.05 |
| 2014            | 17 356 (12.0) | 104 967 (10.3) | 6564 (12.5)   | 5502 (10.5)   | 0.05 |
| 2015            | 18 551 (12.9) | 102 382 (10.1) | 7075 (13.5)   | 5522 (10.6)   | 0.05 |

ADG=aggregate disease group; IQR=interquartile range; SD=standard deviation; St diff=standardized difference.

*Because of ICES (Institute for Clinical Evaluative Sciences) privacy regulations, data fields with fewer than six events are not shown, along with cells that would allow for derivation of these small cells.
Table 2 Outcomes in the matched study cohort, n (%), 95% CI unless otherwise stated

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Patients treated by female surgeon (n=52 315)</th>
<th>Patients treated by male surgeon (n=52 315)</th>
<th>Absolute difference</th>
<th>Adjusted odds ratio (95%CI; P value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary outcome (death, readmission, or complication within 30 days)</td>
<td>5819 (11.1, 10.9 to 11.4)</td>
<td>6046 (11.6, 11.3 to 11.8)</td>
<td>0.43%</td>
<td>0.96 (0.92 to 0.99; 0.02)</td>
</tr>
<tr>
<td>Death within 30 days</td>
<td>480 (0.9, 0.8 to 1.0)</td>
<td>543 (1.0, 1.0 to 1.1)</td>
<td>0.12%</td>
<td>0.88 (0.78 to 0.99; 0.04)</td>
</tr>
<tr>
<td>Readmission within 30 days</td>
<td>2433 (4.7, 4.5 to 4.8)</td>
<td>2518 (4.8, 4.6 to 5.0)</td>
<td>0.16%</td>
<td>0.96 (0.91 to 1.02; 0.20)</td>
</tr>
<tr>
<td>Complication within 30 days</td>
<td>3543 (6.8, 6.6 to 7.0)</td>
<td>3674 (7.0, 6.8 to 7.2)</td>
<td>0.25%</td>
<td>0.96 (0.92 to 1.01; 0.10)</td>
</tr>
<tr>
<td>Hospital length of stay, median (IQR)</td>
<td>2 (0 to 4)</td>
<td>2 (0 to 4)</td>
<td>0</td>
<td>0.97 (0.94 to 0.99; 0.01)*</td>
</tr>
</tbody>
</table>

*Adjusted relative rate rather than adjusted odds ratio.
Figure

**Subgroups**
- **Specialty**
  - Cardiothoracic surgery
  - General surgery
  - Neurosurgery
  - Obstetrics and gynaecology
  - Orthopaedic surgery
  - Otolaryngology
  - Plastic surgery
  - Thoracic surgery
  - Urology
  - Vascular surgery

**Surgeon age**
- ≤40
- 41-50
- 51-60
- ≥61

**Years in practice**
- ≤5
- 5.1-10
- 10.1-15
- ≥15

**Hospital status**
- Academic
- Community

**Patient gender**
- Female
- Male

**Patient age**
- 18-35
- 36-64
- ≥65

**Patient comorbidity**
- 0-5
- 6-7
- 8-10
- ≥11

**Surgeon volume (quartiles)**
- 0 - low
- 1
- 2
- 3 - high

**Odds ratio (95% CI)**

- 0.91 (0.82 to 1.01)
- 0.98 (0.93 to 1.03)
- 0.95 (0.69 to 2.32)
- 0.96 (0.87 to 1.05)
- 1.00 (0.89 to 1.12)
- 1.20 (0.81 to 1.79)
- 0.52 (0.35 to 0.79)
- 0.88 (0.60 to 1.29)
- 0.86 (0.59 to 1.26)
- 1.00 (1.00 to 1.00)
- 0.94 (0.88 to 0.99)
- 1.00 (0.94 to 1.06)
- 0.94 (0.86 to 1.02)
- 0.79 (0.60 to 1.03)
- 0.94 (0.88 to 1.00)
- 0.96 (0.89 to 1.04)
- 0.76 (0.68 to 0.85)
- 1.06 (1.00 to 1.13)
- 0.98 (0.93 to 1.03)
- 0.94 (0.90 to 0.99)
- 0.96 (0.92 to 1.01)
- 0.95 (0.90 to 1.01)
- 0.99 (0.88 to 1.11)
- 0.98 (0.93 to 1.04)
- 0.92 (0.87 to 0.97)
- 0.94 (0.87 to 1.01)
- 1.01 (0.93 to 1.10)
- 0.94 (0.88 to 1.01)
- 0.95 (0.88 to 1.02)
- 0.96 (0.90 to 1.02)
- 0.99 (0.92 to 1.06)
- 0.95 (0.87 to 1.04)
- 0.92 (0.84 to 1.01)

**Fig 1** Likelihood of adverse postoperative outcomes (death, readmission, or complications) among patients treated by female and male surgeons, stratified by physician, patient, and hospital factors.