


# Analysis of the relationship between surgeon procedure volume and complications after total knee arthroplasty using a propensity-matched cohort study

Tosan Okoro <sup>1</sup>, Sebastian Tomescu,<sup>2</sup> J Michael Paterson,<sup>3</sup> Bheeshma Ravi <sup>2</sup>

**To cite:** Okoro T, Tomescu S, Paterson JM, *et al*. Analysis of the relationship between surgeon procedure volume and complications after total knee arthroplasty using a propensity-matched cohort study. *BMJ Surg Interv Health Technologies* 2021;3:e000072. doi:10.1136/bmjst-2020-000072

► Additional supplemental material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/bmjst-2020-000072>).

Received 25 November 2020  
Revised 22 February 2021  
Accepted 12 March 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

<sup>1</sup>Department of Arthroplasty, Robert Jones and Agnes Hunt Orthopaedic Hospital NHS Foundation Trust, Oswestry, UK  
<sup>2</sup>Division of Orthopedic Surgery, Sunnybrook Holland Orthopedic and Arthritic Centre, Toronto, Ontario, Canada  
<sup>3</sup>Institute of Clinical Evaluative Sciences, Toronto, Ontario, Canada

## Correspondence to

Mr Tosan Okoro;  
tosanwumi@hotmail.com

## ABSTRACT

**Objectives** This study aimed to identify a threshold in annual surgeon volume associated with increased risk of revision (for any cause) and deep infection requiring surgery following primary elective total knee arthroplasty (TKA).

**Design** A propensity score matched cohort study.

**Setting** Ontario, Canada.

**Participants** 169 713 persons who received a primary TKA between 2002 and 2016, with 3-year postoperative follow-up.

**Main outcome measures** Revision arthroplasty (for any cause), and the occurrence of deep surgical infection requiring surgery.

**Results** Based on restricted cubic spline analysis, the threshold for increased probability of revision and deep infection requiring surgery was <70 cases/year. After matching of 51 658 TKA recipients from surgeons performing <70 cases/year to TKA recipients from surgeons with greater than 70 cases/year, patients in the former group had a higher rate of revision (for any cause, 2.23% (95% Confidence Interval (CI) 1.39 to 3.07) vs 1.70% (95% CI 0.85 to 2.55); Hazard Ratio (HR) 1.33, 95% CI 1.21 to 1.47, p<0.0001) and deep infection requiring surgery (1.29% (95% CI 0.44 to 2.14) vs 1.09% (95% CI 0.24 to 1.94); HR 1.33, 95% CI 1.17 to 1.51, p<0.0001).

**Conclusions** For primary TKA recipients, cases performed by surgeons who had performed fewer than 70 TKAs in the year prior to the index TKA were at 31% increased relative risk of revision (for any cause), and 18% increased relative risk for deep surgical infection requiring surgery, at 3-year follow-up.

## BACKGROUND

Patients undergoing a variety of procedures and treatments with high volume providers, that is, physicians and hospitals that undertake the procedure with relatively high frequency, tend to achieve better outcomes.<sup>1</sup> The identification of a specific volume threshold below which the risk for complications increase serves as a useful marker

## Key messages

### What is already known about this subject?

► Increased surgeon volume is associated with reduced risk for surgical complications for total knee arthroplasty (TKA), but there is no specific threshold associated with a reduced risk.

### What are the new findings?

► This study has shown that, for a cohort of first-time TKA recipients, cases performed by surgeons who had performed 70 or fewer TKA procedures in the year prior to the index TKA were at increased risk for a major surgical complication requiring further surgery.

### How might these results affect future research or surgical practice?

► Having a healthcare policy that restricts provision of TKAs to only high-volume surgeons can have the effect of restricting access to care. Research should be focused on improving quality of care and ensuring that it is uniform across surgeons with varying volume. This has significant cost implications in terms of training, revalidation and continuing professional development.

for surgeons planning their practice, and for hospitals when allocating resources.

No accepted rigorous methodology for identification of volume thresholds with regard to total knee arthroplasty (TKA) currently exists. Most studies have defaulted to assigning arbitrary cut-offs or ranking patients by volume and splitting them into quantiles for analysis, with potential information loss and the creation of volume categories that have limited rational basis.<sup>2,3</sup> Consequently, published volume categories are remarkably inconsistent between studies.

The aim of this study was to use a restricted cubic splines methodology to visualize the relationship between surgeon volume

(defined as the number of TKAs performed by the surgeon in the year before the index arthroplasty) and the rates of revision and deep infection requiring surgery within 3 years of surgery. We further aimed to identify a threshold of surgeon volume below which the risk for complications appreciably increased, and to quantify this risk using a propensity score match analysis.

## METHODS

### Study sample

We used health administrative databases from Ontario, Canada. The main data sources were hospital discharge abstracts from the Canadian Institute for Health Information Discharge Abstract Database, physician claims from the Ontario Health Insurance Plan (OHIP), and demographic information on each physician from the Ontario Physician Human Resources Data Centre and OHIP Corporate Provider Database. Using specific procedure and diagnostic codes from the Canadian version of the 10th revision of the International Statistical Classification of Diseases and the Canadian Classification of Health Interventions (ICD-10-CA/CCI), we defined a cohort of patients who received their first primary elective TKA for osteoarthritis between April 1, 2002 and March 31, 2016. Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of our research.

### Outcomes of interest: revision within three years of surgery

The primary outcome of this study was the occurrence of a revision arthroplasty (for any cause) within 3 years of surgery. Revision arthroplasty was defined placement of new components following failure of some or all of a failed primary implant. The secondary outcome was the occurrence of deep infection requiring surgery within 3 years. Primary TKAs that involved the use of computer navigation, patient-specific instrumentation or robotics were also excluded. Revision procedures (for any cause) were identified using ICD-10-CA/CCI procedure codes accompanied by the supplementary status attribute 'R'. Deep infections requiring surgery were identified using two methods: (1) occurrence of an ICD-10-CA diagnostic code for intra-articular infection, with a confirmatory code for an irrigation and debridement and/or (2) occurrence of an OHIP code for a spacer insertion. Deep infections that led to a revision (irrigation and debridement as a first stage), as well as irrigation and debridement with implant retention were also included.

### Covariates of interest

We measured several patient and provider covariates that have been previously shown to affect the risk of occurrence of complications following knee replacement. Patient age and sex was obtained from the OHIP Registered Persons Databases.<sup>4-6</sup> Comorbidities listed on hospital discharge abstracts in the 3 years before the index admission were categorized according to an adaptation of the Charlson

Comorbidity Index.<sup>7</sup> The Johns Hopkins Adjusted Clinical Groups (ACG) 'frailty' indicator, based on diagnosis codes from hospitalizations and physician visits in the 2 years before the index admission for TKA, was used to determine the presence of 'frailty' (yes/no) at the time of the surgery.<sup>8</sup> We identified patients with a history of pre-existing coronary artery disease (CAD), diabetes, hypertension, chronic obstructive pulmonary disease (COPD), and dementia using validated algorithms (see online supplemental appendix 1<sup>9-17</sup>). This study used administrative data available at the Institute for Clinical Evaluative Sciences (ICES). As Ontario, Canada has a single-payer healthcare system, ICES is able to capture every interaction a patient has within this system. Several algorithms comprised of ICD-10 codes and provider billing codes have been used to create cohorts of patients with various comorbidities. The algorithms used in this study have been validated with chart abstraction to determine their sensitivity, specificity and positive predictive value (online supplemental appendix 1<sup>9-17</sup>). We also identified patients that had been previously counseled on smoking cessation through appropriate physician billing codes.

Neighborhood income quintile was used as a surrogate for socioeconomic status. This measure categorizes small geographic areas into five roughly equal population groups, with the lowest quintile referring to the least affluent neighborhoods.<sup>18 19</sup> The Ontario marginalization index was also assessed. This comprises four elements: ethnic concentration, residential instability, dependency, and deprivation.<sup>20</sup> Each element is sorted into fifths, arranged from least (lowest fifth) to most marginalized (highest fifth).<sup>20</sup> The index has been shown to be stable across time periods and across different geographic areas and to be associated with health outcomes including depression, smoking, alcohol consumption, and body mass index.<sup>20</sup> For each TKA, hospital volume was defined as the number of primary knee arthroplasty procedures (both primary and revision) performed at the hospital where the surgery was performed in the 365 days prior to the index procedure. Unicompartamental knee arthroplasties (UKA) were excluded, as was any revision of a UKA to a TKA. Surgeon experience was operationalized as the number of years since completion of residency/fellowship orthopedic training, as evidenced by the start of Royal College accreditation and start of independent practice, for the senior surgeon at the time of the index surgery. This includes foreign surgeons moving to Canada with substantial experience, but recent Royal College accreditation.

### Main exposure variable: surgeon volume

For each TKA, surgeon volume was defined as the number of TKA procedures (both primary and revision) performed by the senior surgeon in the 365 days prior to the index procedure. This definition allowed for an individual surgeon's volume to change dynamically over time. Only revisions of primary TKAs were included with revisions of UKAs to TKAs excluded.

## Statistical analyses

Baseline cohort characteristics were calculated using proportions and medians as appropriate, and were compared between groups using Wilcoxon rank sum tests for continuous variables and  $\chi^2$  tests for categorical variables. A propensity score for receipt of a TKA from a surgeon with the determined threshold value was calculated using a logistic regression model.<sup>21 22</sup> The covariates entered into the propensity score were sociodemographics (age, sex, income quintile, Ontario marginalization index), health status (Charlson score, frailty, hypertension, Chronic Obstructive Pulmonary Disease (COPD), Congestive Heart Failure (CHF), diabetes, Coronary artery disease (CAD), smoking status), provider characteristics (annual hospital volume), and surgeon experience.

Restricted cubic splines with four knots<sup>23</sup> were used to model the relationship between surgeon volume and the occurrence of revision and infection after controlling for patient age, sex, income quintile, comorbidities, hospital volume and surgeon experience. Restricted cubic splines are a flexible tool to model complex, non-linear relationships between a continuous variable and an outcome.<sup>23</sup> Normal regression analysis assumes a linear relationship between the predictor and outcome variables. This would suggest that the impact of an increase in surgeon volume would be the same if the increase in volume was from 10 to 30 cases/year, or from 210 to 230 cases/year. A spline, on the other hand, does not make any assumptions of linearity. It divides the relationship into smaller 'chunks', allowing for portions that are not linear. In the current study, use of a spline also allowed us to identify a threshold of volume at which the greatest benefit of surgeon volume is obtained. The non-linear relationship between surgeon volume and the risk of revision was examined to identify an inflection point, if any, which could be used to dichotomize annual volume into categories in a clinically meaningful way. If an area of inflection was observed, multivariable logistic regression was used to determine the area under the curve (AUC) for the models relating various surgeon volume cut-points to the risk of revision. The surgeon volume with the maximum AUC was selected as the cut-point to dichotomize surgeon volume.

TKA recipients from a surgeon who had operated below the selected threshold value in the preceding year were matched to those from a surgeon with more than the threshold number of cases on the logit of the propensity score using calipers of width equal to 0.2 of the SD of the logit of the propensity score.<sup>24</sup> A matching ratio of 1:1 was used.<sup>25</sup>

We estimated standardized differences for all covariates after matching, with a standardized difference of 10% or more considered indicative of imbalance.<sup>25</sup> The occurrence of complications were also compared between the two groups after matching, using methods appropriate for the analysis of matched data in estimating the treatment effect and its statistical significance. Generalized

**Table 1** Characteristics of eligible primary total knee arthroplasty recipients

Total number of patients	N=169 713
<b>Demographics</b>	
Age (years) (median (IQR))	68 (61–75)
Women (N (%))	106 070 (62.49%)
Income quintile (N (%))	
Lowest	30 057 (17.71%)
2	35 079 (20.66%)
3	34 561 (20.36%)
4	34 510 (20.33%)
Highest	35 067 (20.66%)
<b>Comorbidities (N (%))</b>	
Coronary artery disease	5443 (3.20%)
Congestive heart failure	9811 (5.78%)
Chronic obstructive pulmonary disease	32 018 (18.86%)
Diabetes	45 698 (26.92%)
Hypertension	123 875 (72.99%)
Frail	11 562 (6.81%)
Charlson score (N (%))	
0	120 151 (70.79%)
1	31 647 (18.64%)
2	11 374 (6.70%)
3 or more	6541 (3.85%)
<b>Admission characteristics</b>	
Hospital volume (median (IQR))	362 (253–526)
Surgeon volume (median (IQR))	89 (59–127)
Surgeon years in practice (median (IQR))	15 (8–24)
<b>Outcome (N (%))</b>	
Revision (for all causes) within 3 years	3199 (1.88%)
Deep infection requiring surgery within 3 years	1926 (1.13%)

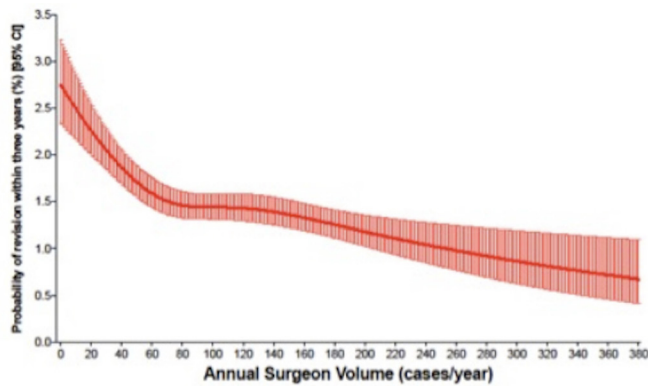
estimating equations were used to determine the increased risk (if any) in patients with surgeons whose volumes were lower than the selected threshold, after taking pair-matching into account.<sup>23</sup> All analyses were performed at the Institute for Clinical Evaluative Sciences (<http://www.ices.ca>) using SAS V.9.3 for UNIX (SAS Institute, Cary, North Carolina, USA). The type I error probability was set to 0.05 for all analyses.

## RESULTS

### Patient and characteristics

Between April 1, 2002 and March 31, 2016, there were 169 713 eligible TKA recipients (table 1), with a median age of 68, and 62.5% were of female gender. The median annual surgeon volume was 89 cases (IQR 59–127), and





**Figure 1** Probability of revision (for any cause) within 3 years vs annual surgeon volume for patients undergoing primary elective total knee arthroplasty. KEY—thick red line—mean values. Lighter vertical bars—95% CI.

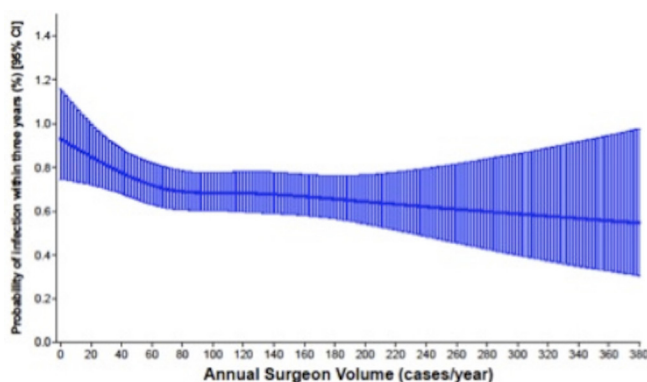
the median number of years in practice was 15 years (IQR 8–24).

### Regression splines describing the relationship between surgeon volume and risk of complications

The restricted cubic splines relating annual surgeon volume to revision and infection had similar shapes—both were negatively sloped with inflection points at approximately 70 cases/year, after which the rates of complications continued to decrease with increased surgeon volume, but at a lower rate (figures 1 and 2). The AUC was determined for various cut-points of surgeon volume (60, 65, 70, 75, 80, 85, 90). Subsequently, surgeon volume was dichotomized at 70 cases (<70 or  $\geq$ 70 cases in the 365 days prior to the procedure). A total of 56 265 procedures (33%) were performed by ‘low-volume’ surgeons over the study period.

### Matching

A total of 51 658 subjects (92%) who received a TKA from a surgeon with <70 cases in the 365 days prior to the surgery were successfully matched to a TKA recipient from a surgeon with  $\geq$ 70 cases in the 365 days prior to



**Figure 2** Probability of deep infection requiring surgery within 3 years vs annual surgeon volume for patients undergoing primary elective total knee arthroplasty. KEY—thick blue line—mean values. Lighter vertical bars—95% CI.

the surgery (table 2). There were no exclusions to the propensity-score match. Patients that did not complete follow-up to 3 years postoperatively because of death (n=4374 (2.57%)) were not excluded from the match. A patient flow diagram is illustrated in figure 3. After matching, the absolute standardized differences were less than 10% for all variables entered into the propensity score, indicating an adequate match (table 2).

### Outcomes after matching

Knee replacement recipients from a surgeon with low annual volumes (<70 cases in the year prior) had a higher rate of revision within 3 years (2.23% (95% CI 1.39 to 3.07) vs 1.70% (95% CI 0.85 to 2.55); HR 1.33, 95% CI 1.21 to 1.47,  $p<0.0001$ ) and infection requiring surgery within 3 years (1.29% (95% CI 0.44 to 2.14) vs 1.09% (95% CI 0.24 to 1.94); HR 1.33, 95% CI 1.17 to 1.51,  $p<0.0001$ ) (table 3). The number needed to harm for revision was 189 persons (95% CI 143 to 278), which translates to an absolute risk increase of 0.53% (95% CI 0.37% to 0.71%).

### DISCUSSION

The model used in this study indicated that as surgeon volume increased, the risks for complications decreased. We observed an inflection point at 70 TKAs/year, after which the rate of decrease in the risks for complications leveled off. After performing a propensity score match for surgeon volume, which also controlled for surgeon experience, we found that in patients operated on by surgeons with annual volumes <70 cases, the relative risk for revision (for any cause) increased by 31% (2.23% (95% CI 1.39 to 3.07) vs 1.70% (95% CI 0.85 to 2.55)), and the relative risk of deep surgical infection requiring surgery increased by 18% (1.29% (95% CI 0.44 to 2.14) vs 1.09% (95% CI 0.24 to 1.94)). The findings of this study indicate that for surgeons performing <70 primary TKAs per year, regardless of previous experience, there is an increased likelihood for these two complications.

Use of the above approach is lacking in other studies which define volume categories instead. One such study used stratum-specific likelihood ratio analysis, a method of analyzing receiver operating characteristic curves to show that there is a significant decrease in 90-day complication and 2-year revision rates for surgeons in higher volume categories (60–145 TKAs per year and >146 TKAs per year) vs those in lower volume categories (0–12, 13–59 TKAs per year).<sup>26</sup> A criticism of the aforementioned study is that it does not take account of changes in procedures and practices over time,<sup>27</sup> something which the cubic spline technique is able to account for, as it is a dynamic assessment of a surgeon’s practice, meaning that each individual surgeon can have different volumes across the study period.

The threshold of 70 primary TKAs per year is higher than previously reported for total hip arthroplasty (THA) (35 cases per year<sup>28</sup>). A possible reason for this difference is that the rates of early complications after TKA are

**Table 2** Comparison of primary TKA recipients, after matching

	<70 cases/year	≥70 cases/year	Standardized difference*
Number of patients	N=51 658	N=51 658	
Demographics			
Age (years) (median (IQR))	68 (61–75)	69 (62–75)	0.01
Female (N (%))	32 310 (62.54%)	32 355 (62.63%)	0
Income quintile (N (%))			
Lowest	9438 (18.27%)	9527 (18.44%)	0
2	10 727 (20.76%)	10 505 (20.33%)	0.01
3	10 556 (20.43%)	10 732 (20.77%)	0.01
4	10 552 (20.42%)	10 555 (20.43%)	0
Highest	10 385 (20.10%)	10 339 (20.01%)	0
Comorbidities (N (%))			
Coronary artery disease	1634 (3.16%)	1749 (3.38%)	0.01
Congestive heart failure	3047 (5.89%)	3120 (6.03%)	0.01
Chronic obstructive pulmonary disease	9999 (19.35%)	10 246 (19.83%)	0.01
Diabetes	13 789 (26.69%)	13 860 (26.83%)	0
Hypertension	38 040 (73.63%)	38 216 (73.97%)	0.01
Frail	3667 (7.09%)	3718 (7.19%)	0
Charlson score (N (%))			
0	36 624 (70.89%)	36 404 (70.47%)	0.01
1	9602 (18.58%)	9762 (18.89%)	0.01
2	3433 (6.64%)	3470 (6.71%)	0
3 or more	1999 (3.86%)	2022 (3.91%)	0
Admission characteristics			
Hospital volume (median (IQR))	299 (207–410)	315 (230–420)	0.07
Surgeon years in practice (median (IQR))	13 (5–23)	14 (7–21)	0.03

\*The standardized difference allows for a comparison of baseline variables (both continuous and dichotomous) between groups and is not affected by the size of the samples, as opposed to p values, which are 39. Because the standardized difference looks at the entire distribution of each sample, it is perfectly normal to see a difference for example, in median age of 1 year to be reflected in a 1% standardized difference as illustrated here.

TKA, total knee arthroplasty.

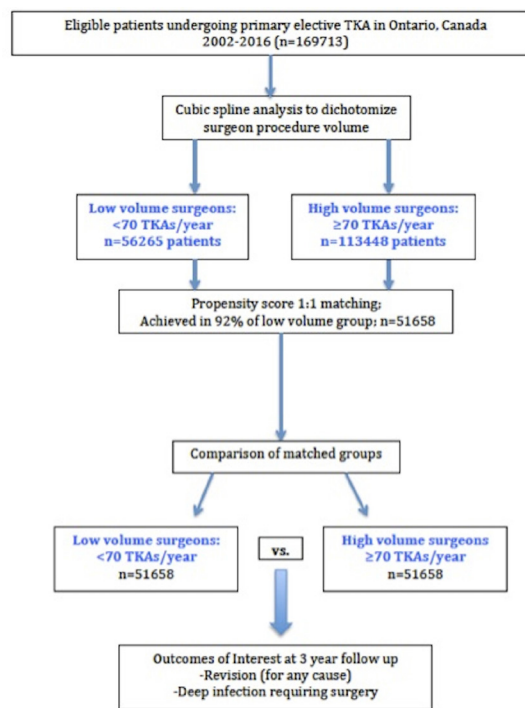
generally a little lower than after THA; as such, a higher volume threshold is required to observe a difference in outcomes. The THA study complications included dislocation, periprosthetic fracture, and a broader definition for infection. The early complications of venous thromboembolism and death within 90 days were also included in that study.

Many patients do not have the means or social support to travel to high volume TKA providers from rural areas. Having a healthcare policy that restricts provision of TKAs to only high volume surgeons can have the effect of restricting access to care. Research should be focused on improving quality of care and ensuring that it is uniform across surgeons with varying volume.<sup>27</sup> This has significant cost implications in terms of training, revalidation and continuing professional development.

Through the use of restricted cubic splines, we found that there was a noticeable decrease in likelihood of

revision (for any cause) and deep infection requiring surgery, as the surgeon's yearly primary knee arthroplasty volume increased; however, the relationship was not linear. While the relative improvement in risk of revision (for any cause) and deep infection requiring surgery with increasing surgeon volume attenuated after this point, there continued to be a downward trend in the risks for these two complications, indicating that increased surgeon volume continues to have a beneficial impact, although one that is less pronounced.

The cumulative risk of revision for TKAs in the National Joint registry for England and Wales ranges from 1.53% for cemented TKAs, to 1.80% for hybrid TKAs, and 2.09% for uncemented TKAs, within 3 years following surgery.<sup>29</sup> The absolute difference in revision rates between these implant designs is of similar magnitude (0.56) to the difference between low and high volume surgeons found in the current study (0.53).



**Figure 3** Participant flow diagram for patients undergoing analysis of outcome based on surgeon procedure volume using a propensity score matched cohort study after primary elective TKA. TKA, total knee arthroplasty.

There are limitations to the work we have presented. First, we did not have any information on patient-reported outcomes. Second, we are unable to report a subgroup analysis based on the indications for revision TKA, nor assess center effects in the modelling we have undertaken. We were also unable to control for technical aspects of the procedure, for example, severity of pre-existing deformity, the need for adjunctive soft tissue releases, or length of surgery, factors which have been linked with complication following TKA.<sup>30</sup> With regard to the analysis performed, there is a risk of introducing bias with propensity score matching.<sup>31</sup> In this study, however, we have matched more than 90% of the cases performed by low volume surgeons, suggesting a representative sample, and standardized differences for all relevant measured confounders were under 10% after matching.

This study was conducted in Ontario, Canada where it is standard practice for healthcare organizations (hospitals or hospital networks) to standardize implants for use by their surgeons. For the vast majority of surgeons therefore, implant selection would be equivalent irrespective of volume. We accept that having the detail of the types of brands may help define if there is an advantage for low volume surgeons using the best performing implants but the converse could also be the case. Unfortunately, this detail is not something we have available in our database to provide. There is no role for private practice in the provision of joint replacement; hence, all practicing surgeons are accounted for in the available data set. Surgeons registered with the College of Physicians and Surgeons of Ontario are not allowed to work across two provinces. As such, we are confident that the surgeon volumes captured in the study are not an underestimate, but we also accept that we are unable to account for the unlikely movement of surgeons in and out of the Ontario system during the period of the study.

Further research, potentially using data sources that capture this information, is recommended to confirm or refute these hypotheses. With regard to deep infection requiring surgery as an outcome, we have used a limited definition (cases requiring irrigation and debridement, liner exchange or spacer insertion), meaning that we have not captured instances of superficial surgical site infection (SSI). It has recently been demonstrated that increased surgical duration is associated with a higher risk for infection following TKA<sup>32</sup> and this is a plausible mechanism linking low volumes and infection following TKA. Infection should be viewed as a spectrum from acute to chronic.<sup>33</sup> We define an infection in our database as one that is serious enough to warrant an additional intervention such as spacer insertion or wound irrigation and debridement. The rate of superficial SSIs that are managed primarily by surgeons or primary care providers is difficult to quantify and this information is not available in our database.

This paper and the description of the study methodology provide an opportunity for a similar type of analysis to be carried out in different geographical locations/settings as it may be inappropriate to apply a number obtained from one region to another. This is especially

**Table 3** Proportion of primary TKA recipients with specific complications, after matching

Complication (N (%))	Surgeon volume <70 cases/year	Surgeon volume ≥70 cases/year	Hazard Ratio (HR)*	McNemar test (p value)
Revision (for all causes) within 3 years	1153 (2.23%; 95% CI 1.39 to 3.09)	880 (1.70%; 95% CI 0.85 to 2.55)	1.33 (95% CI 1.21 to 1.47)	<0.0001
Deep infection requiring surgery within 3 years	668 (1.29%; 95% CI 0.44 to 2.14)	566 (1.09%; 95% CI 0.24 to 1.94)	1.33 (95% CI 1.17 to 1.51)	<0.0001

\*Stratified by hospital where the surgery was performed. TKA, total knee arthroplasty.



important as the median numbers performed per year, are different between countries (median number of TKAs per surgeon in UK (n=36)<sup>29</sup>; in the USA median number of TKAs per surgeon (n=23)<sup>34</sup>).

## CONCLUSION

In a cohort of primary TKA recipients, we found that cases performed by surgeons who had performed fewer than 70 TKA procedures in the year prior to the index TKA were at 31% increased relative risk of revision (for any cause), and 18% increased relative risk for deep surgical infection requiring surgery, at 3 years of follow up.

**Acknowledgements** This study was supported by a grant from the Canadian Institutes of Health Research and by ICES, a non-profit research institute funded by the Ontario Ministry of Health and Long-Term Care. Parts of this material are based on data and information compiled and provided by the Ontario Ministry of Health and Long-Term Care and the Canadian Institute for Health Information. The opinions, results and conclusions reported in this paper are those of the authors and are independent from the data providers and funding sources; no endorsement is intended or should be inferred.

**Contributors** TO: interpretation of data for the study, drafting the study and revising it critically for important intellectual content; final approval of the version to be published; and agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. ST: interpretation of data for the study; revising the study critically for important intellectual content; final approval of the version to be published; and agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. JMP: acquisition, analysis, and interpretation of data for the study; revising the study critically for important intellectual content; final approval of the version to be published; and agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. BR: conception or design of the study; acquisition, analysis, and interpretation of data for the study; revising the study critically for important intellectual content; final approval of the version to be published; and agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

**Funding** This study was supported by a grant from the Canadian Institutes of Health Research and by ICES, a non-profit research institute funded by the Ontario Ministry of Health and Long-Term Care.

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** No data are available.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

## ORCID iDs

Tosan Okoro <http://orcid.org/0000-0003-0432-0936>

Bheeshma Ravi <http://orcid.org/0000-0003-1829-6885>

## REFERENCES

1 Morche J, Mathes T, Pieper D. Relationship between surgeon volume and outcomes: a systematic review of systematic reviews. *Syst Rev* 2016;5:204.

- 2 Altman DG, Lausen B, Sauerbrei W, *et al*. Dangers of using "optimal" cutpoints in the evaluation of prognostic factors. *J Natl Cancer Inst* 1994;86:829–35.
- 3 Kozower BD, Stukenborg GJ. Volume-Outcome relationships in thoracic surgery. *Thorac Surg Clin* 2017;27:251–6.
- 4 Weiss RJ, Kärrholm J, Rolfson O, *et al*. Increased early mortality and morbidity after total hip arthroplasty in patients with socioeconomic disadvantage: a report from the Swedish hip arthroplasty register. *Acta Orthop* 2019;90:264–9.
- 5 Santaguida PL, Hawker GA, Hudak PL, *et al*. Patient characteristics affecting the prognosis of total hip and knee joint arthroplasty: a systematic review. *Can J Surg* 2008;51:428–36.
- 6 Kralj B. *Measuring "Rurality" for Purposes of Health Care Planning: An Empirical Measure for Ontario*. Toronto: Ontario Medical Association, 2005.
- 7 Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *J Clin Epidemiol* 1992;45:613–9.
- 8 Weiner JP, Abrams C. The Johns Hopkins ACG® system: technical reference guide version 10.0. 2011.
- 9 Leong A, Dasgupta K, Bernatsky S, *et al*. Systematic review and meta-analysis of validation studies on a diabetes case definition from health administrative records. *PLoS One* 2013;8:e75256.
- 10 Hux JE, Ivis F, Flintoft V, *et al*. Diabetes in Ontario: determination of prevalence and incidence using a validated administrative data algorithm. *Diabetes Care* 2002;25:512–6.
- 11 Lipscombe LL, Hwee J, Webster L, *et al*. Identifying diabetes cases from administrative data: a population-based validation study. *BMC Health Serv Res* 2018;18:316.
- 12 Tu K, Campbell NR, Chen Z-L, *et al*. Accuracy of administrative databases in identifying patients with hypertension. *Open Med* 2007;1:e18.
- 13 Gershon AS, Wang C, Guan J, *et al*. Identifying individuals with physician diagnosed COPD in health administrative databases. *COPD* 2009;6:388–94.
- 14 Ko DT, Mamdani M, Alter DA. Lipid-Lowering therapy with statins in high-risk elderly patients: the treatment-risk paradox. *JAMA* 2004;291:1864.
- 15 Roffi M, Mukherjee D. Treatment-risk paradox in acute coronary syndromes. *Eur Heart J* 2018;39:3807–9.
- 16 Austin PC, Daly PA, Tu JV. A multicenter study of the coding accuracy of hospital discharge administrative data for patients admitted to cardiac care units in Ontario. *Am Heart J* 2002;144:290–6.
- 17 Jaakkimainen RL, Bronskill SE, Tierney MC, *et al*. Identification of Physician-Diagnosed Alzheimer's disease and related dementias in population-based administrative data: a validation study using family physicians' electronic medical records. *J Alzheimers Dis* 2016;54:337–49.
- 18 Information ClfH. *Reducing gaps in health: a focus on socio-economic status in urban Canada*. Ottawa, Ont: CIHI, 2008.
- 19 Glazier RH, Badley EM, Gilbert JE, *et al*. The nature of increased Hospital use in poor neighbourhoods: findings from a Canadian inner City. *Can J Public Health* 2000;91:268–73.
- 20 Matheson F, Dunn J, Smith K. Ontario Marginalization Index (ON-Marg): User Guide. Centre for Research in Inner City Health, St Michael's Hospital (Toronto) 2011.
- 21 White HL, Matheson FI, Moineddin R, *et al*. Neighbourhood deprivation and regional inequalities in self-reported health among Canadians: are we equally at risk? *Health Place* 2011;17:361–9.
- 22 Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res* 2011;46:399–424.
- 23 Gauthier J, Wu QV, Gooley TA. Cubic splines to model relationships between continuous variables and outcomes: a guide for clinicians. *Bone Marrow Transplant* 2020;55:675–80.
- 24 Austin PC. A tutorial and case study in propensity score analysis: an application to estimating the effect of in-hospital smoking cessation counseling on mortality. *Multivariate Behav Res* 2011;46:119–51.
- 25 Austin PC. Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharm Stat* 2011;10:150–61.
- 26 Wilson S, Marx RG, Pan T-J, *et al*. Meaningful thresholds for the volume-outcome relationship in total knee arthroplasty. *J Bone Joint Surg Am* 2016;98:1683–90.
- 27 Giori NJ. Should Only the Highest-Volume Surgeons and Centers Be Doing Primary Total Knee Arthroplasty? Commentary on an article by Sean Wilson, BA, *et al*: "Meaningful Thresholds for the Volume-Outcome Relationship in Total Knee Arthroplasty". *J Bone Joint Surg Am* 2016;98:e92.



- 28 Ravi B, Jenkinson R, Austin PC, *et al.* Relation between surgeon volume and risk of complications after total hip arthroplasty: propensity score matched cohort study. *BMJ* 2014;348:g3284.
- 29 National Joint Registry for England and Wales. 16Th annual report. Available: <https://reports.njrcentre.org.uk/Portals/0/PDFdownloads/NJR%2016th%20Annual%20Report%202019.pdf> [Accessed 11 Mar 2020].
- 30 Badawy M, Espehaug B, Fenstad AM, *et al.* Patient and surgical factors affecting procedure duration and revision risk due to deep infection in primary total knee arthroplasty. *BMC Musculoskelet Disord* 2017;18:54.
- 31 Austin PC, Jembere N, Chiu M. Propensity score matching and complex surveys. *Stat Methods Med Res* 2018;27:1240–57.
- 32 Ravi B, Jenkinson R, O'Heireamhoin S, *et al.* Surgical duration is associated with an increased risk of periprosthetic infection following total knee arthroplasty: a population-based retrospective cohort study. *EClinicalMedicine* 2019;16:74–80.
- 33 Elkins JM, Kates S, Lange J, *et al.* General assembly, diagnosis, definitions: proceedings of international consensus on orthopedic infections. *J Arthroplasty* 2019;34:S181–5.
- 34 American Joint Replacement Registry. 2018 annual report. Available: <https://www.aaos.org/registries/publications/ajrr-annual-report/> [Accessed 11 Mar 2020].



**APPENDICES****APPENDIX 1****Covariates:**

Previously validated algorithms were used to identify patients with the following diagnoses within the administrative data:

<b>Comorbidity</b>	<b>Sensitivity %</b>	<b>Specificity %</b>	<b>Reference</b>
Diabetes	86.0	97.0	9, 10, 11
Hypertension	72.0	95.0	12
Chronic Obstructive Pulmonary Disease (COPD)	57.5	95.4	13
Congestive Heart Failure (CHF)	84.8	97.0	14
Coronary Artery disease (CAD)	88.8	92.8	15, 16
Dementia	79.3	99.1	17
Frailty	The Johns Hopkins ACG® System (Version 10)		8

**References**

8. Weiner JP, Abrams C. The Johns Hopkins ACG® System: Technical Reference Guide Version 10.0. 2011.
9. Leong A, Dasgupta K, Bernatsky S, et al. Systematic review and meta-analysis of validation studies on a diabetes case definition from health administrative records. *PLoS One*. 2013 Oct 9;8(10):e75256.
10. Hux JE, Ivis F, Flintoft V, et al. Diabetes in Ontario: determination of prevalence and incidence using a validated administrative data algorithm. *Diabetes care*. 2002;25(3):512.

11. Lipscombe LL, Hwee J, Webster L, et al. Identifying diabetes cases from administrative data: a population-based validation study. *BMC Health Serv Res*. 2018 May 2;18(1):316.
12. Tu K, Campbell NR, Chen ZL, et al. Accuracy of administrative databases in identifying patients with hypertension. *Open medicine*. 2007;1(1):e18.
13. Gershon A, Wang C, Guan J, et al. Identifying individuals with physician diagnosed COPD in health administrative databases. *COPD*. [Comparative Study Research Support, Non-U.S. Gov't]. 2009 Oct;6(5):388-94.
14. Ko DT, Mamdani M, Alter DA. Lipid-lowering therapy with statins in high-risk elderly patients: the treatment-risk paradox. *Jama*. 2004;291(15):1864.
15. Roffi M, Mukherjee D. Treatment-risk paradox in acute coronary syndromes *Eur Heart J*. 2018 Nov 7;39(42):3807-3809
16. Austin PC, Daly PA, Tu JV. A multicenter study of the coding accuracy of hospital discharge administrative data for patients admitted to cardiac care units in Ontario. *Am Heart J*. 2002;144(2):290-296.
17. Jaakkimainen RL, Bronskill SE, Tierney MC, et al. Identification of Physician-Diagnosed Alzheimer's Disease and Related Dementias in Population-Based Administrative Data: A Validation Study Using Family Physicians' Electronic Medical Records. *J Alzheimers Dis*. 2016;54(1):337-349.