How much is enough? Finding the minimum annual surgical volume threshold for total knee replacement

Per-Henrik Randsborg, Amanda C Chen

There is compelling evidence that the outcome of total knee replacement (TKR) depends on the annual caseload of the surgeon and the institution. The fact that high-volume centres have better outcomes is so well known that patients themselves increasingly seek treatment at high-volume institutions.

But how do we determine the surgical volume threshold that constitutes an acceptable risk?

In the recent article by Okoro et al, data from nearly 170,000 TKR recipient in Ontario, Canada, were analysed using a state of the art restricted cubic spline (RCS) analysis. The authors identified an inflection point for increased probability for early revision or infection at 70 annual cases, after which the risk of complications plateaued. However, risk reduction continued even beyond an annual caseload of 70, highlighting the value of treatment by high-volume surgeons.

This method of identifying the surgical volume threshold differs from previous studies conducted in the last 10 years. While some papers fail to mention their methods for defining the volume groups, others have largely split the data by quartiles. This results in more or less arbitrary thresholds that are more dependent on local healthcare structure and population size than providing clinical meaningful volume categories. Due to varying methods for defining the volume groups, others have largely split the data by quartiles. This results in more or less arbitrary thresholds that surgeons and institutions can aim to achieve, the volume categories should not be defined a priori. This methodological flaw is corrected by the RCS analysis, which uses the dataset to identify the inflection point to identify the actual caseload threshold where the risk of complications is reduced.

However, it is unlikely that the threshold of 70 annual cases to reduce risk of complications is universal. Like any surgery, TKR surgery is composed of teamwork and environmental factors, such as healthcare structure and population demographics, which influence the outcome. Okoro et al suggest that all communities with available real-world data perform similar RCS analysis to identify the caseload threshold relevant to their population. It should therefore be mentioned that Yu et al published a similar RCS analysis from Taiwan in 2019. They found the inflection point to be somewhat lower in their population, at 50 cases per year, proving the point that the caseload threshold varies between populations. Differences in specialty training practices and best practices across nations affect volume threshold estimates. We predict that future research will produce new and different surgical volume thresholds to reduce complications, and it is unlikely that a definite answer will be possible to find. It is more likely that the true annual caseload threshold varies over time, as surgical training, implant design and population characteristics change.

Another challenge in determining acceptable caseload thresholds is that different complications will have different volume thresholds. Existing literature on the effect of surgical volume has used a spectrum of different outcomes, including 30-day readmission, revision rate, radiological implant alignment, surgical site infection and length of hospital stay. Different complications will have different etiologies, not all related to the experience of the surgeon. Hospital environment, population characteristics, postoperative care and rehabilitation facilities all play a part in securing a good outcome after arthroplasty surgery. For example, length of stay is dependent on...
Table 1  A literature review of publications in English from 2011 to 2021, with a minimum of 20 000 patients included, evaluating the effect of surgeon and/or hospital volume on total knee replacement outcome

<table>
<thead>
<tr>
<th>Study / Author year</th>
<th>N of TKRs</th>
<th>Setting</th>
<th>Volume thresholds</th>
<th>Definition of threshold (Method)</th>
<th>Outcome</th>
<th>Result</th>
<th>OR (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>de la Torre et al. 2019</td>
<td>36316</td>
<td>Catalan Arthroplasty Registry, Spain</td>
<td>&lt;125               &gt;125</td>
<td>Previous studies</td>
<td>Revision</td>
<td>Higher risk for low volume</td>
<td>1.29 (1.16 to 1.44)</td>
<td>NR</td>
</tr>
<tr>
<td>Badawy et al. 2013</td>
<td>26698</td>
<td>Norwegian Arthroplasty Registry, Norway</td>
<td>&lt;25               25–49               50–99               100–149               ≥150</td>
<td>Previous studies</td>
<td>Revision</td>
<td>No difference</td>
<td>0.81* (0.68)</td>
<td></td>
</tr>
<tr>
<td>D’apuzzo et al. 2017</td>
<td>377705</td>
<td>Statewide Planning and Research Cooperative System (SPARCS), New York, USA</td>
<td>1–89               90–235               236–644               ≥645</td>
<td>SSLR</td>
<td>Readmission</td>
<td>Higher risk for low volume</td>
<td>1.32 (1.16 to 1.51)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Kurtz et al. 2016</td>
<td>952593</td>
<td>Medicare 100% claims, USA</td>
<td>150–299               300–499               450–599               ≥600</td>
<td>Not explained</td>
<td>Readmission</td>
<td>Higher readmission for low volume</td>
<td>51.2%†</td>
<td></td>
</tr>
<tr>
<td>Meyer et al. 2011</td>
<td>43180</td>
<td>Krankenhaus–Infektions-Surveillance System, Germany</td>
<td>1–50               51–99               &gt;100</td>
<td>Previous studies</td>
<td>Surgical site infections</td>
<td>Higher risk for low volume</td>
<td>2.04</td>
<td>NR</td>
</tr>
<tr>
<td>Namba et al. 2013a</td>
<td>64017</td>
<td>Total Joint Replacement Registry, USA</td>
<td>&lt;100               100–199               ≥200</td>
<td>Not explained</td>
<td>Aseptic revision</td>
<td>No difference</td>
<td>0.93 (0.59 to 1.48)</td>
<td>0.769</td>
</tr>
<tr>
<td>Namba 2013b</td>
<td>56216</td>
<td>Total Joint Replacement Registry, USA</td>
<td>&lt;100               100–199               ≥200</td>
<td>Not explained</td>
<td>Surgical site infection</td>
<td>Lower risk for low volume</td>
<td>0.33 (0.12 to 0.90)</td>
<td>0.030</td>
</tr>
<tr>
<td>Pamilo et al. 2015</td>
<td>59696</td>
<td>Perfect Knee Replacement database, Finland</td>
<td>1–99               100–249               250–449               ≥450</td>
<td>chosen arbitrarily</td>
<td>LOS, Readmission, MUA, revision</td>
<td>Higher readmission risk for low volume, no difference for revision</td>
<td>Readmission: 1.11 (1.03 to 1.19)</td>
<td>NR</td>
</tr>
</tbody>
</table>

Continued
Table 1

<table>
<thead>
<tr>
<th>Study</th>
<th>Setting</th>
<th>Data source, country</th>
<th>Volume thresholds</th>
<th>Low surgical volume</th>
<th>Result</th>
<th>Definition of threshold</th>
<th>Outcome</th>
<th>Type</th>
<th>CR and revision</th>
<th>OR (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singh et al. 2011</td>
<td>Pennsylvania Healthcare Cost Containment Council database, USA</td>
<td>Low: 1–25</td>
<td>Higher risk for low volume</td>
<td>Low</td>
<td>1.07 (0.80 to 1.43)</td>
<td>Not explained</td>
<td>No association</td>
<td>No of TKRs</td>
<td>readmission</td>
<td>0.435</td>
<td></td>
</tr>
<tr>
<td>Wilson 2016</td>
<td>Statewide Planning and Research Cooperative System (SPARCS), New York, USA</td>
<td>Medium: 26–100</td>
<td>Higher risk for low volume</td>
<td>Medium</td>
<td>0.97 (0.84 to 1.12)</td>
<td>CR and MR</td>
<td>No association</td>
<td>No of TKRs</td>
<td>readmission</td>
<td>0.435</td>
<td></td>
</tr>
<tr>
<td>Yu et al. 2019</td>
<td>Taiwan National Health Insurance Research Database, Taiwan</td>
<td>High: 101–200</td>
<td>Higher risk for low volume</td>
<td>High</td>
<td>1.32 (1.20 to 1.45)</td>
<td>SSLR</td>
<td>No association</td>
<td>No of TKRs</td>
<td>readmission</td>
<td>0.435</td>
<td></td>
</tr>
<tr>
<td></td>
<td>&gt;200</td>
<td>Very high: &gt;200</td>
<td>Higher risk for low volume</td>
<td>Very high</td>
<td>1.47 (1.33 to 1.64)</td>
<td>SSLR</td>
<td>No association</td>
<td>No of TKRs</td>
<td>readmission</td>
<td>0.435</td>
<td></td>
</tr>
<tr>
<td></td>
<td>&gt;235</td>
<td>&gt;235</td>
<td>Higher risk for low volume</td>
<td>&gt;235</td>
<td>1.65 (1.50 to 1.82)</td>
<td>SSLR</td>
<td>No association</td>
<td>No of TKRs</td>
<td>readmission</td>
<td>0.435</td>
<td></td>
</tr>
</tbody>
</table>

Multiple factors, not least the hospital capacity and the proximity of a potential rehabilitation facility or the home of the patient. A longer hospital stay in a rural hospital with low surgical volume may therefore be an indication of poor quality. Furthermore, not all of these complications necessarily constitute a clinically meaningful difference for the patient (eg, implant alignment may not have clinical relevance). As such, the revision rate is probably the best outcome measure, representing a hard end-point with a clear consequence for the patient. However, even revision rates are subject to individual variation thresholds by the revision surgeon, and local traditions such as patella resurfacing or not. Defining the minimal surgical volume threshold is clearly not an easy task.

However, the true surgical volume threshold is perhaps not that important. The point is that there is little doubt that surgical volume matters, and the crucial question is what to do with this knowledge. Some authors have suggested that low-volume surgeons should either stop doing the procedure, or do more of it.10 There are many arguments for centralisation, as TKR are nearly always an elective procedure that may be postponed. However, travelling for hours to receive care is not always possible, and removing TKR service from local hospitals reduces both the availability and the quality of care. Caring for patients operated by other surgeons is difficult, especially if the patient has had a type of surgery you do not perform yourself, and interrupts the continuity of a patient’s care.

Furthermore, patients treated at high-volume institutions differ from those of low-volume institutions,11 meaning that changing the population flow might affect the caseload thresholds, which are probably constantly changing anyway.

Still, there is no denying the effect of surgical volume on outcome quality. The improved, and likely more accurate, caseload thresholds estimated by Okoro et al and Yu et al provide valuable information that can guide decision-makers when organising the arthroplasty service in the community. An effort to reach a minimum of caseload in most centres should be combined with efforts to improve the quality of the care in low-volume institutions, so that uniform healthcare and equity can be reached while preserving access to care.

Twitter Per-Henrik Randsborg @randsborg

Contributors P-HP wrote the initial draft and screened the literature abstracts. ACC performed the literature search, completed the table and revised the manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Commissioned; internally peer reviewed.

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
REFERENCES


