Universiteit Leiden
The Netherlands

What makes the best performing hospital? the IQ Joint study

Schie, P. van

Citation
Schie, P. van. (2023, November 8). *What makes the best performing hospital?: the IQ Joint study*. Retrieved from https://hdl.handle.net/1887/3656771

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/3656771

Note: To cite this publication please use the final published version (if applicable).

WHAT MAKES THE BEST PERFORMING HOSPITAL?

The IQ Joint study

Peter van Schie

What makes the best performing hospital?

The IQ Joint study

Peter van Schie

Cover design and layout: Peter van Schie

Lay-out & Printing: ProefschriftMaken | proefschriftmaken.nl

ISBN: 978-94-6469-653-0

Copyright© 2023 by Peter van Schie

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means without permission in writing from the author. The copyright of the articles has been transferred to the respective journals.

The conduction of the research included in this thesis was carried out at the Departments of Orthopedics & Biomedical Data Sciences, Medical Decision Making of the Leiden University Medical Centre, Leiden, the Netherlands.

The studies were financed by a grant from the Van Rens Foundation (VRF 2018-001). The Dutch Orthopaedic Association and the Anna Fund in Leiden have contributed financially to the printing and distribution of this thesis.

What makes the best performing hospital?

The IQ Joint study

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof.dr.ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op woensdag 8 november 2023
klokke 15:00 uur

door
Peter van Schie
geboren te Leiden
in 1988

Promotor

Prof. dr. R.G.H.H. Nelissen

Copromotor

Dr. P.J. Marang-van de Mheen

Leden promotiecommissie
Prof. dr. D.H. Biesma
Prof. dr. M. Levi (Amsterdam UMC & Nederlandse Organisatie voor Wetenschappelijk Onderzoek)
Prof. dr. A. Lübbeke-Wolff (University of Oxford)
Prof. dr. R.W. Poolman
Prof. dr. M.W.J.M. Wouters

Table of contents

Chapter 1	General introduction and outline of the thesis	9
Chapter 2	International comparison of variation in performance between hospitals for total hip and knee arthroplasties: Is it even possible? A systematic review including 33 studies and 8 arthroplasty register reports. *EFORT Open Rev. 2022 Apr 21;7(4):247-263*	29
Chapter 3	Between-hospital variation in revision rates after Total Hip and Knee Arthroplasty in the Netherlands. Directing quality-improvement initiatives. *J Bone Joint Surg Am. 2020 Feb 19;102(4):315-324*	89
Chapter 4	Monitoring hospital performance after total hip and knee arthroplasty using statistical process control: How much earlier can we detect worsening performance? *J Bone Joint Surg Am. 2020 Dec 2;102(23):2087-2094*	109
Chapter 5	A more comprehensive evaluation of quality of care after total hip and knee arthroplasty: Combining 4 indicators in an ordered composite outcome. *Acta Orthop. 2022 Jan 3;93:138-145*	129
Chapter 6	Is postoperative improvement in PROM scores overestimated due to selective non-response? Examining associations with clinical outcomes can provide insights. *Submitted*	149
Chapter 7	Awareness of performance on outcomes after total hip and knee arthroplasty among Dutch orthopaedic surgeons: How to improve feedback from arthroplasty registries. *Acta Orthop. 2021 Feb;92(1):54-61*	173
Chapter 8	Effectiveness of a multifaceted quality improvement intervention to improve patient outcomes after total hip and knee arthroplasty: A registry nested cluster randomized controlled trial. *BMJ Qual Saf. 2023 Jan;32(1):34-46*	199

Chapter 9 Summary, general discussion, and future perspectives 245

Chapter 10 Appendices 267

Dutch summary (Nederlanse samenvatting) 269

Acknowledgment (Dankwoord) 277

Bibliography 281

Curriculum vitae 285

Chapter 1

General introduction and outline of the thesis

Background

Total hip and knee arthroplasties (THA and TKA) are the most successful treatments for end-stage hip and knee osteoarthritis and are among the most common surgeries performed worldwide.(1-8) In the Netherlands, more than 33,000 THA and 34,000 TKA are performed yearly. The number of procedures is expected to increase exponentially in the coming decades due to the ageing population and the increasing prevalence of obesity.(9-12) As a result, the number of adverse events such as revisions will rise, increasing not only the burden to patients but also the burden on healthcare systems and thus society.(10,13-17) It thus becomes even more critical to improve clinical outcomes for arthroplasty surgery, such as revision, readmission, complications, and length of stay (long LOS) by delivering high-quality care to these patients.

Since the earliest recorded attempt at total hip and knee arthroplasty by Professor Glück in 1891, the most significant improvement in clinical outcome rates were achieved by the ongoing process of implant improvement.(18,19) For that matter, the first hip implants made of ivory were unsuccessful due to severe postoperative pain and high revision rates due to prosthesis loosening; these implants were replaced by a hollow ball of glass that fits over the femoral head in 1923.(20,21) However, the glass could not withstand the hip joint forces and shattered. Subsequently, experiments with materials for hip implants were done with several materials (e.g., Bakelite, Pyrex, Viscaloid, Vitallium, and metal).(22,23) Professor Sir Charnley was the first who successfully performed a THA in 1960, the basic principles he used still apply today (Figure 1).(24) For TKA, Insall and Burstein in the US and simultaneously Yamamoto in Japan developed the prototype of a new knee implant concept in 1971, resurfacing all three knee joint articular surfaces (i.e., femur, tibia, and patella) (Figure 2).(25) However, even though clinical outcomes have improved considerably since these first successful THA and TKA prototypes, clinical outcome after surgery does not only depend on implant design but also the quality of care delivered to the patient.

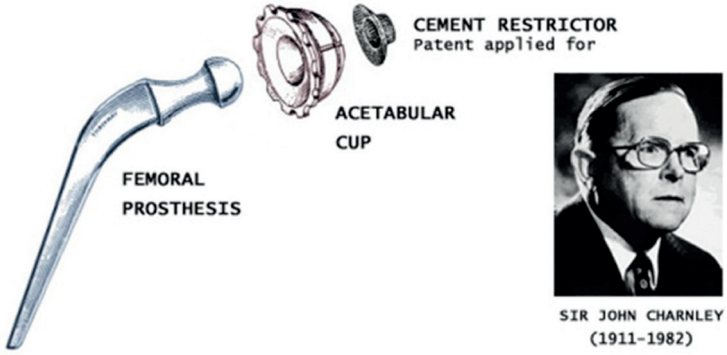


Figure 1 Charnley hip replacement
Charnley's design consisted of a metal (stainless steel) femoral component and a Teflon acetabular component; both were fixed to the bone using bone cement (acrylic).

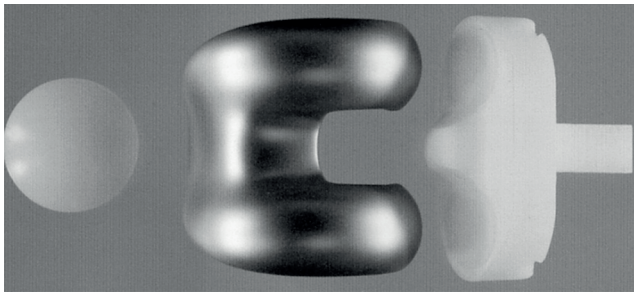


Figure 2 Insall-Burstein knee implant
Insall and Burstein's total condylar implant. The patellar button is shown at the top, the femur component in the middle, and the tibial component at the bottom.

In orthopaedics, revision surgery is the most commonly reported clinical outcome measure following THA and TKA because of the dramatic consequences for the patient and the considerable cost involved.(26,27) During such revision surgery, some or all of the primary joint components are replaced by new parts. Furthermore, revision surgery is a longer and more complex procedure than primary arthroplasty surgery, with less favourable outcomes such as higher infection rates and worse function compared with the primary procedure.(28,29) A revision within one year is a widely used clinical outcome to monitor the quality of care delivered, as it is close to the primary surgery performed and, as such likely related to the quality of care delivered during and after that primary surgery. In contrast, a revision within 5 years

is less likely to be related to the quality of care delivered but more relevant to track implant survival.

Dutch arthroplasty register

The Dutch Arthroplasty Register (LROI), established in 2007 by the Netherlands Orthopaedic Association (NOV), started with the registration of hip and knee implants.(9) In 2014, the implants for the shoulder, elbow, and ankle, and Patient Reported Outcome Measures (PROMs) for hip, knee, and shoulder were added. In 2016 the register was further expanded with data collection for wrists and fingers. The initial primary purpose of arthroplasty registries was to compare performance, defined as survival, of different types of implants and detect worse-performing implants earlier to promote patient safety. A well-known example of the latter was the detection of the metal-on-metal (MoM) hip arthroplasty disaster. The latter was identified by the Australian Orthopaedic Association National Joint Replacement Registry (AOANJRR) as having an outlier performance in 2007.(30) Twenty percent of patients had to undergo a revision within ten years, compared with four percent in the "classic" metal-on-polyethylene arthroplasties.(31,32) The mortality risk increased by 8.5% (95%-CI: 5.8%-11.2%) due to these implants.(33) These MoM implants were withdrawn from the market in 2010, showing the value of registry data to ensure safety. Another advantage, besides the ability of registries to flag bad-performing implants, is that implants can be traced back to the patient in case of a calamity as the implant number of each implant component and a personal encrypted security number for each patient are collected.

In recent years, registries have also been used to monitor the quality of care delivered by orthopaedic groups/clinics by tracking clinical outcomes (e.g., revision) of hospitals, with the aim to improve performance by providing hospitals with feedback on their outcomes.(9,34-36) The LROI provides feedback on case-mix adjusted revision rates, Patient-Reported Outcome Measurement (PROM) difference scores, and patient characteristics at the hospital level, which are reported on a secured web-based dashboard annually. Connected to that feedback, the LROI and the NOV started a "Quality commission" in 2017 that actively approaches poorly performing hospitals (i.e., negative outliers) to discuss quality improvement initiatives (QII) and create action plans to improve.(37) The commission includes a team of clinical experts visiting hospitals with unfavourable results to advise and help them start improving the quality of care delivered.(37)

Registry data can be used reliably for monitoring the quality of care delivered if coverage (i.e., the proportion of hospitals reporting to the register), completeness (i.e., the proportion of included patients in the register), and validity of data are good. The LROI coverage is 100% as all hospitals upload their data. Data completeness is checked against Hospital Electronic Health Records and currently exceeds 99% for primary procedures and 97% for revisions, meaning that >99% of primary procedures are included.(38) To increase validity, the LROI has implemented several steps, such as mandatory boxes in the web-based registration form and automatically generated reports when one or more variables are missing or inconsistent. The validity is currently 93% for THA and 96% for TKA.(39) The patient's vital status (dead or alive) is obtained from the Dutch insurance healthcare database Vektis, which is needed to calculate implant survival. The opt-out system for informed consent is applied, whereby patients must actively object not to be included.

Feedback using registry data

Arthroplasty registries, including the LROI, provide feedback to orthopeadic surgeons, which is intended to improve the quality of care delivered.(40,41) The variation in performance between hospitals on clinical outcomes is usually reported compared to a reference standard (i.e., the benchmark) that indicates whether performance in that specific clinic is comparable to the predefined benchmark or deviates from it. The total population of interest (i.e., national average on revision for TKA or THA etc.) with a specific norm is often chosen as the benchmark for comparison. The most commonly used clinical outcomes for feedback are revision, readmission, and complications, as these indicators are considered reliable, actionable, and fit for purpose.(42-47) Providing feedback is based on the belief that orthopaedic surgeons are prompted to modify their practice when performance feedback shows that their clinical practice outcome measures deviate from the desirable benchmark target. Feedback is not only aimed at underperforming hospitals but also at average or high-performing (i.e., best-practices) hospitals with the rationale that there is always room to improve further. In addition, hospitals can learn from better-performing hospitals by engaging and adopting items from these best practices. However, there may be less incentive to improve further for hospitals among the best-performing hospitals in their own country. These hospitals may be interested in comparing their outcomes with hospitals from other countries or healthcare systems to stimulate further improvement. The latter is only possible if there is consistency in the clinical outcome definitions and methods used to collect data across countries or healthcare systems, as these will determine the frequency of occurrence of end-point definitions (i.e., events) of the benchmark. Furthermore, the occurrence of clinical outcomes

should be adjusted for differences in patient characteristics that determine the risk of these outcomes, to achieve a fair hospital comparison. For example, healthy patients (e.g., ASA I patients with osteoarthritis) are expected to have a lower frequency of adverse clinical outcomes (e.g., revision) than patients with multiple comorbidities (e.g., ASA IV patients with congenital hip deformities).(48-51) Therefore, in case of hospital comparisons across countries, it will have to be assessed whether the same patient characteristics are available in all countries for risk adjustment.

Funnel plots with control limits are commonly used as a graphical tool to show between-hospitals variation for clinical outcomes (Figure 3). Hospitals plotted between the two control limits have a performance that is not statistically different from what is expected based on their patient characteristics. Hospitals plotted above the upper control limit have more events observed than expected and are negative outliers when considering adverse outcomes like revision; vice versa, hospitals plotted under the lower control limit have fewer events observed than expected and are positive outliers. Nowadays, multiple years of clinical outcome data (e.g., revision, readmission, and complications) are usually combined to obtain detectable and reliable differences in hospital performance due to low event rates.(9,34-36,52-57) The LROI and other arthroplasty registries typically combine three years of data. Therefore, it may take long before deteriorating performance is noticed, resulting in late action plans to improve care. However, if the reliability of ranking hospitals using single years of data is acceptable, a reliable earlier signal can be given when performance deteriorates. In addition, hospitals may be underperforming for all-cause revision, which may be due to various underlying causes. For example, hospitals may have higher infection, dislocation, implant loosening, or technical failure rates. To connect with subsequent targeted QII, it is relevant whether hospitals are an outlier for specific revision indications rather than only reporting outlier status on the all-cause revision rates. Assessing whether hospital performance can be reliably distinguished for revision indications may therefore be a step forward.

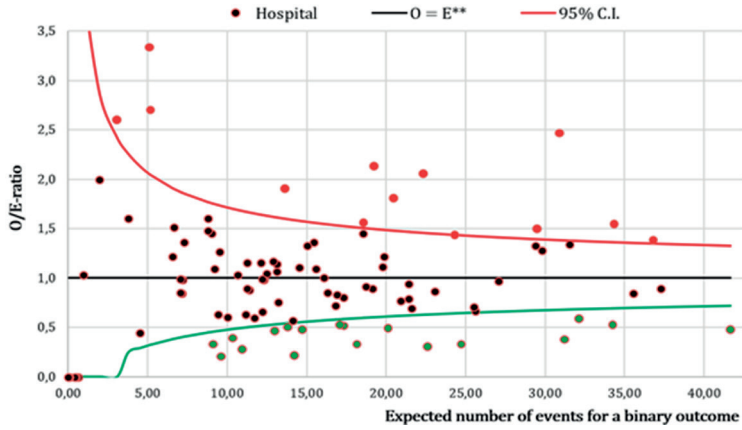


Figure 3 Example of a funnel plot
Each point represents a hospital. The O/E-ratio is described on the y-axis, and the expected number of events per hospital on the x-axis. The red line is the upper control limit, and the green line is the lower control limit. When O/E-ratio is equal to 1, the observed number equals the expected number of events in a hospital.
O/E=observed number of revisions divided by the expected number of revisions.

A second opportunity to detect underperforming hospitals earlier is using statistical approaches other than the traditional funnel plot, where the aggregate performance over a period is shown rather than the development over time. Several clinical studies have led to growing interest in Statistical Process Control charts such as the Shewhart-p-chart and cumulative sum (CUSUM) chart, which can distinguish between an "in-control" process, showing only random (chance) variation over time and an "out-of-control" process, showing systematic (special-cause) variation by generating a signal (alert) at a specific point in time when the control limit is reached (Figure 4).(58-69) The Shewhart p-chart is considered an accessible chart, especially concerning implementation and straightforward interpretation.(70) However, the CUSUM chart has superior performance in detecting small (<10%) and significant (>10%) increases in event rates.(65,70-72) These two SPC charts thus seem logical alternatives for the funnel plot. The CUSUM chart is used to identify higher complication rates after THA and TKA in Scotland.(69) However, that study did not report how much earlier a signal was generated and with what accuracy compared with the traditional funnel plot presentation of events according to the benchmark.

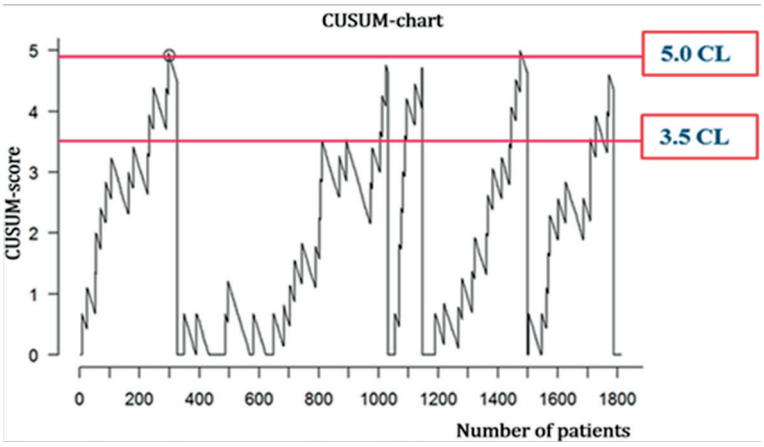


Figure 4 Example of a CUSUM chart
The CUSUM chart for a single hospital is shown. The observed minus expected probability for an event is plotted for every consecutive patient. When the score goes upward, this means that the observed performance for that patient is worse than expected, vice versa when going down. A single (alert) is generated when crossing the CL. A higher CL (5.0) means increasing certainty that this is a valid signal with fewer false-positive signals but could miss cases of worse performance. The opposite applies to lower CL (3.5).
CL=control limit; CUSUM=cumulative sum.

A third opportunity to obtain an earlier signal is to combine multiple relevant clinical outcomes into one composite outcome measure to achieve a higher number of events per hospital, thereby increasing the accuracy by which hospital performance is estimated. Feedback by arthroplasty registries is primarily provided on single clinical outcomes (e.g., revision, readmission, or complications). Using a composite outcome may take less time to differentiate between hospitals in their performance reliably.(73-76) Furthermore, single outcomes provide only a partial view of the quality of care delivered, as a hospital may have a high score on one outcome but may need to improve on another.(9,34-36,52,77-79) Because of these limitations, there is growing interest in composite measures, in which multiple relevant outcomes are combined to provide a more comprehensive overview of the delivered quality of care. However, existing composite outcomes often represent a binary all-or-none concept. These measures are less useful for quality improvement since they give equal weight to all clinical outcomes. In contrast, from a patient's perspective, revision is probably more important to avoid than, for example, upper-quartile LOS. In addition, these all-or-none composite outcomes do not provide feedback on where to improve (i.e., for which combination of outcomes), and do not take into account the possible interrelationship between individual clinical outcomes.(73,76,80,81)

Sofar, the focus on providing feedback in arthroplasty registries has mainly been concentrated on "passive" reporting clinical outcomes that measure the unintended effects rather than the intended outcomes of pain reduction, improvement in functionality, and improved health-related quality of life. Ten and 20% of patients following THA and TKA, respectively, are not satisfied with postoperative results, mainly related to persistent pain and disability.(82,83) The proportion of patients achieving a clinically relevant improvement is not routinely used as hospital feedback in most arthroplasty registries due to low response rates that vary between hospitals. Most arthroplasty registries collect PROMs, but response rates (i.e., patients who completed the PROM questionnaire) are often low compared with above 95% completeness of data for clinical outcomes.(84-89) It is unclear in which direction missing PROM scores bias results on clinically relevant improvements. In addition, despite using Patient-Reported Outcome Measures (PROMs) for a long-time, it remains unclear whether clinically relevant improvement in PROM scores reflects the quality of care delivered as measured by well-known clinical outcomes (i.e., revision, readmission, complications, and upper-quartile LOS).(90) For example, a readmission may affect postoperative PROM scores and thereby the improvement achieved, which would suggest that hospitals reducing their readmission rates may also increase the proportion of patients receiving clinically relevant improvements. If these are unrelated, improvement of PROM scores should be achieved in a different direction, e.g., in management of pre-operative expectations.

Improving the effectiveness of feedback using registry data

Arthroplasty registries have been used worldwide to give surgeons and hospitals feedback on their performance, aiming to improve the quality of care delivered. (9,34-36,52,77-79,91) The LROI provides such feedback on a secured web-based dashboard at the hospital level, which is updated annually. However, feedback is only effective when surgeons are aware of their performance. The awareness of overall defined performance on surgical procedures among orthopaedic surgeons performing THA and TKA in the Netherlands is unknown, as well as factors associated with their awareness of this overall performance. Reasons might be that it is too time-consuming for orthopaedic surgeons to log into the LROI dashboard, so that feedback should be sent differently (e.g., by email or presentation), or surgeons may need education to (mis)interpret the graphs (i.e., funnel plots). The latter also prevents wrong interpretation and thus conclusions regarding their clinic's performance data. They may also be interested in other clinical outcomes than revision (e.g., readmission, complications, and upper-quartile LOS) or require the feedback to be tailored to their surgeon group rather than making selections themselves in an online dashboard.

Surgeons often overestimate their own performance, assuming that performance is good even when there is room for improvement, which may limit the introduction of QII.(92)

Even though a necessary first step, awareness is not sufficient to make feedback effective. The effectiveness of feedback was evaluated in a Cochrane review, including 49 studies from multiple fields, and showed a median absolute improvement of 4.3% (interquartile range (IQR): 0.5% to 16.0%).(41) Although the median effect size is relatively small, the 75te percentile effect size suggests that A&F, when optimally designed and used in the proper context, can play an essential role in improving clinical practice. The current literature shows that A&F is more effective when provided proactively in a multifaceted way, including for instance an action implementation toolbox to facilitate actions undertaken instead of a "passive" single element (e.g., feedback or education alone).(41,93-95) However, a previous study including such an action implementation toolbox only showed improvement in process indicators but not in patient outcomes, whereas the ultimate goal is to improve patient outcomes.(95) In addition, it should be mentioned that even a relatively small improvement for THA and TKA will have a significant societal impact considering the sizeable annual number of surgeries performed (I.e., approximately 70,000 THA and TKA only in the Netherlands annually).(9)

To support continuous quality improvement initiatives arising from feedback, the *Plan-Do-Study-Act (PDSA)* cycle is mostly used.(96) The cycle consists of four repetitive steps in which the main focus is the interpretation of feedback results, the introduction of QII, and the evaluation of its effect. Because the cycle can only be appropriately used when up-to-date feedback is available, it seems well-suited to be used by hospitals using registry-based feedback. The following four steps are distinguished within the PDSA cycle (Figure 5):

1. Detect opportunities for improvement based on feedback, and plan QII to improve care (*Plan*);

2. Implement the QII (*Do*);

3. Review whether the QII was implemented as planned, analyse the effect, and identify what is learned (*Study*);

4. If the QII did not work as planned, adapt the QII and go through the cycle again with a revised plan. If the QII was successful, incorporate what is learned. Use what is learned to plan a new QII, beginning the cycle again and continue further improvement of outcomes.

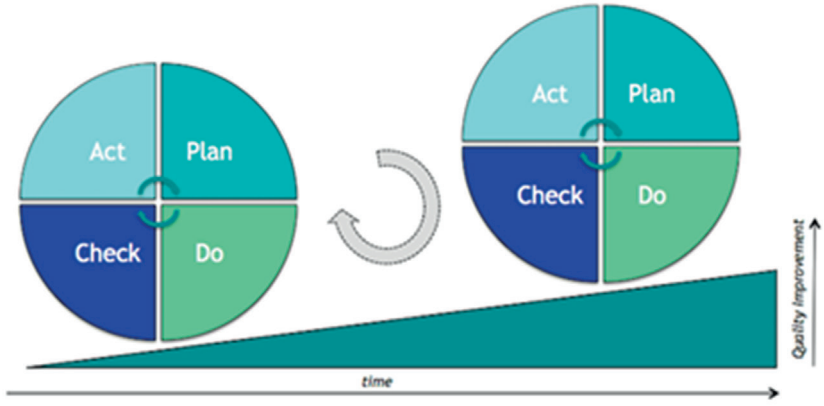


Figure 5 Plan-Do-Check-Act cycle

Outline of this thesis

This thesis aimed to study how arthroplasty registries can improve their feedback to surgical groups and its individual surgeons to give direction to QII that improve care of THA and TKA, and to test the effectiveness of such improved "active" feedback to improve patient outcomes.

In **chapter 2,** the literature was reviewed to assess the international variation between hospitals for revision, readmission, and complications. We focused on definitions used, data collection methods, and for which patient characteristics clinical outcomes were adjusted, to interpret the variation between hospitals. This will also show whether it is feasible for hospitals to compare their outcomes with other countries to stimulate further improvement.

To improve current registry-based feedback to surgical groups, **chapter 3** investigates the between-hospital variation for 1-year revision and specific indications for revision (i.e., infection, dislocation, implant loosening, and technical failure) to give direction to improvement initiatives and assess the reliability of ranking hospitals on their performance using 3 years or single years of data, with the latter enabling an earlier signal of deteriorating performance.

In **chapter 4**, the Shewhart-p-chart and CUSUM chart were tested to assess how much earlier and with what accuracy worsening hospital's performance in 1-year

revision rates can be detected, compared with the traditional funnel plots using three years of data.

In **chapter 5,** an ordinal composite outcome measure was developed with all combinations of clinical outcomes (i.e., revision, readmission, complications, and upper-quartile LOS) ranked from best to worst according to the patient's perspective, indicating more specifically where improvement is possible while also potentially increasing the ability to reliably differentiate between hospitals in their performance with fewer years of data. The reliability of ranking was therefore calculated for the composite and single outcomes.

Chapter 6 aimed to provide insight in how improvement in PROM scores may be under- or overestimated relative to all patients who received a THA or TKA. This was done by comparing PROM questionnaire respondents and non-respondents on their clinical outcome rates (i.e., revision, readmission, complications, and upper-quartile LOS). Furthermore, is was assessed whether patients experiencing an adverse clinical outcome had different improvement in PROMS scores as well as whether hospitals with better clinical outcome rates showed different improvement in PROM scores and PROMS response rates.

Improving feedback is only effective when it is seen and correctly interpreted by orthopaedic surgeons; only then do surgeons become aware of their performance and the need for quality improvement. **Chapter 7**, therefore, assessed the awareness of Dutch orthopaedic surgeons regarding their performance on revision and factors associated with this awareness, such as whether they could recall their revision rate, log in to the LROI dashboard at least once a year and the ability to interpret data presentation by a funnel plot correctly. In addition, it was investigated what surgeons wanted to be improved of the LROI feedback, such as additional clinical outcomes, mode of receiving feedback, and preferred frequency.

The knowledge obtained from all previous chapters combined with contemporary theory for providing effective feedback was incorporated into a multifaceted quality improvement intervention, including improved feedback on performance and interactive education combined with an action implementation toolbox containing evidence-based QII. This intervention was tested in a cluster randomised controlled trial on its effectiveness to improve patient outcomes, the results of which are reported in **chapter 8**. **Chapter 9** Includes a summary, general discussion, and future perspectives.

References

1. OECD. Health at a Glance 20192019.
2. Maradit Kremers H, Larson DR, Crowson CS, Kremers WK, Washington RE, Steiner CA, et al. Prevalence of Total Hip and Knee Replacement in the United States. J Bone Joint Surg Am. 2015;97(17):1386-97.
3. Nelson AE, Allen KD, Golightly YM, Goode AP, Jordan JM. A systematic review of recommendations and guidelines for the management of osteoarthritis: The chronic osteoarthritis management initiative of the U.S. bone and joint initiative. Semin Arthritis Rheum. 2014;43(6):701-12.
4. Association) NOVDO. Richtlijn Totale Heup & Richtlijn Totale Knie (Total Hip and Knee Protocol). 2014.
5. Bachmeier CJ, March LM, Cross MJ, Lapsley HM, Tribe KL, Courtenay BG, et al. A comparison of outcomes in osteoarthritis patients undergoing total hip and knee replacement surgery. Osteoarthritis Cartilage. 2001;9(2):137-46.
6. Pivec R, Johnson AJ, Mears SC, Mont MA. Hip arthroplasty. Lancet. 2012;380(9855):1768-77.
7. Learmonth ID, Young C, Rorabeck C. The operation of the century: total hip replacement. Lancet. 2007;370(9597):1508-19.
8. Wilson RA, Gwynne-Jones DP, Sullivan TA, Abbott JH. Total Hip and Knee Arthroplasties Are Highly Cost-Effective Procedures: The Importance of Duration of Follow-Up. J Arthroplasty. 2021.
9. Dutch Arthroplasty Register (LROI). Online Annual Report. https://www.lroi-report.nl/.
10. Kurtz S, Ong K, Lau E, Mowat F, Halpern M. Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030. J Bone Joint Surg Am. 2007;89(4):780-5.
11. Kurtz SM, Ong KL, Lau E, Widmer M, Maravic M, Gomez-Barrena E, et al. International survey of primary and revision total knee replacement. Int Orthop. 2011;35(12):1783-9.
12. Kurtz SM, Lau E, Ong K, Zhao K, Kelly M, Bozic KJ. Future young patient demand for primary and revision joint replacement: national projections from 2010 to 2030. Clin Orthop Relat Res. 2009;467(10):2606-12.
13. Kurtz SM, Ong KL, Lau E, Bozic KJ. Impact of the economic downturn on total joint replacement demand in the United States: updated projections to 2021. J Bone Joint Surg Am. 2014;96(8):624-30.
14. Kurtz SM, Lau E, Watson H, Schmier JK, Parvizi J. Economic burden of periprosthetic joint infection in the United States. J Arthroplasty. 2012;27(8 Suppl):61-5.e1.
15. Vanhegan IS, Malik AK, Jayakumar P, Ul Islam S, Haddad FS. A financial analysis of revision hip arthroplasty: the economic burden in relation to the national tariff. J Bone Joint Surg Br. 2012;94(5):619-23.
16. Khan M, Osman K, Green G, Haddad FS. The epidemiology of failure in total knee arthroplasty: avoiding your next revision. Bone Joint J. 2016;98-b(1 Suppl A):105-12.
17. Crowe JF, Sculco TP, Kahn B. Revision total hip arthroplasty: hospital cost and reimbursement analysis. Clin Orthop Relat Res. 2003(413):175-82.
18. Hernigou P. Earliest times before hip arthroplasty: from John Rhea Barton to Themistocles Glück. Int Orthop. 2013;37(11):2313-8.
19. Brand R, Mont M, Manring M. Biographical Sketch: Themistocles Gluck (1853–1942). Clin Orthop Relat Res. 2011;6:1525-7.
20. Glück T. Referat uber die durch das moderne chirurgische Experiment gewonnenen positiven Resultate, betreffend die Naht und den Ersatz von Defecten ho¨herer Gewebe, sowie u¨ber die

Verwethung resorbirbarer und lebendiger Tampons in der Chirurgie. Arch klin chir. 1891;41:187-239.
21. Hernigou P. Smith-Petersen and early development of hip arthroplasty. Int Orthop. 2014;38(1):193-8.
22. Smith-Petersen MN. Evolution of mould arthroplasty of the hip joint. J Bone Joint Surg Br. 1948;30b(1):59-75.
23. McKee G. Total hip replacement - past, present and future. Biomaterials. 1982;3(3):130-5.
24. Charnley J. Arthroplasty of the hip. A new operation. Lancet. 1961;1(7187):1129-32.
25. Insall J, Tria AJ, Scott WN. The total condylar knee prosthesis: the first 5 years. Clin Orthop Relat Res. 1979(145):68-77.
26. Premkumar A, Kolin DA, Farley KX, Wilson JM, McLawhorn AS, Cross MB, et al. Projected Economic Burden of Periprosthetic Joint Infection of the Hip and Knee in the United States. J Arthroplasty. 2021;36(5):1484-9.e3.
27. Lavernia C, Lee DJ, Hernandez VH. The increasing financial burden of knee revision surgery in the United States. Clin Orthop Relat Res. 2006;446:221-6.
28. Mahomed NN, Barrett JA, Katz JN, Phillips CB, Losina E, Lew RA, et al. Rates and outcomes of primary and revision total hip replacement in the United States medicare population. J Bone Joint Surg Am. 2003;85(1):27-32.
29. Weber M, Renkawitz T, Voellner F, Craiovan B, Greimel F, Worlicek M, et al. Revision Surgery in Total Joint Replacement Is Cost-Intensive. Biomed Res Int. 2018;2018:8987104.
30. de Steiger RN, Hang JR, Miller LN, Graves SE, Davidson DC. Five-year results of the ASR XL Acetabular System and the ASR Hip Resurfacing System: an analysis from the Australian Orthopaedic Association National Joint Replacement Registry. J Bone Joint Surg Am. 2011;93(24):2287-93.
31. Drummond J, Tran P, Fary C. Metal-on-Metal Hip Arthroplasty: A Review of Adverse Reactions and Patient Management. Journal of functional biomaterials. 2015;6(3):486-99.
32. Pfuntner A, Wier LM, Stocks C. Most Frequent Procedures Performed in U.S. Hospitals, 2011: Statistical Brief #165. Healthcare Cost and Utilization Project (HCUP) Statistical Briefs. Rockville (MD): Agency for Healthcare Research and Quality (US); 2006.
33. Pijls BG, Meessen JM, Schoones JW, Fiocco M, van der Heide HJ, Sedrakyan A, et al. Increased Mortality in Metal-on-Metal versus Non-Metal-on-Metal Primary Total Hip Arthroplasty at 10 Years and Longer Follow-Up: A Systematic Review and Meta-Analysis. PLoS One. 2016;11(6):e0156051.
34. Swedish Knee Arthroplasty Register (SKAR). Annual Report 2020. http://myknee.se/pdf/SVK_2020_Eng_1.0.pdf.
35. Danish Hip Arthroplasty Register (DHAR). Annual report 2020. http://danskhoftealloplastikregister.dk/wp-content/uploads/2020/11/dhr-aarsrapport-2020_til_offentliggoerelse-1.pdf.
36. Danish Knee Arthroplasty Register (DKAR). Annual Report 2020. https://www.sundhed.dk/content/cms/99/4699_dkr-arsrapport-2020_offentliggorelse.pdf.
37. CommisionQuality. Commision Quality (Collaboration between Dutch Orthopaedic Association and Dutch Arthroplasty Registry). Protocol: Quality procedure, 2017. www.orthopeden.org.
38. Dutch Arthroplasty Register (LROI). Data completeness. https://www.lroi-report.nl/data-quality/completeness/.
39. Dutch Arthroplasty Register (LROI). Data validity. https://www.lroi-report.nl/data-quality/validity/.
40. Brehaut JC, Eva KW. Building theories of knowledge translation interventions: use the entire menu of constructs. Implement Sci. 2012;7:114.

41. Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, French SD, et al. Audit and feedback: effects on professional practice and healthcare outcomes. Cochrane Database Syst Rev. 2012(6):Cd000259.
42. Bozic KJ, Grosso LM, Lin Z, Parzynski CS, Suter LG, Krumholz HM, et al. Variation in hospital-level risk-standardized complication rates following elective primary total hip and knee arthroplasty. J Bone Joint Surg Am. 2014;96(8):640-7.
43. Thirukumaran CP, McGarry BE, Glance LG, Ying M, Ricciardi BF, Cai X, et al. Impact of Hospital Readmissions Reduction Program Penalties on Hip and Knee Replacement Readmissions: Comparison of Hospitals at Risk of Varying Penalty Amounts. J Bone Joint Surg Am. 2020;102(1):60-7.
44. Courtney M, Darrith B, Bohl DD, Frisch NB, Valle CJD. Reconsidering the affordable care act's restrictions on physician-owned hospitals analysis of CMS data on total hip and knee arthroplasty. Journal of Bone and Joint Surgery - American Volume. 2017;99(22):1888-94.
45. Van Schie P, Van Bodegom-Vos L, Zijdeman TM, Nelissen R, Marang-Van De Mheen PJ. Awareness of performance on outcomes after total hip and knee arthroplasty among Dutch orthopedic surgeons: how to improve feedback from arthroplasty registries. Acta Orthop. 2020:1-8.
46. Barbazza E, Klazinga NS, Kringos DS. Exploring the actionability of healthcare performance indicators for quality of care: a qualitative analysis of the literature, expert opinion and user experience. BMJ quality & safety. 2021.
47. Fischer C, Lingsma HF, Marang-van de Mheen PJ, Kringos DS, Klazinga NS, Steyerberg EW. Is the readmission rate a valid quality indicator? A review of the evidence. PLoS One. 2014;9(11):e112282.
48. Lenguerrand E, Whitehouse MR, Beswick AD, Kunutsor SK, Burston B, Porter M, et al. Risk factors associated with revision for prosthetic joint infection after hip replacement: a prospective observational cohort study. Lancet Infect Dis. 2018;18(9):1004-14.
49. Prokopetz JJ, Losina E, Bliss RL, Wright J, Baron JA, Katz JN. Risk factors for revision of primary total hip arthroplasty: a systematic review. BMC Musculoskelet Disord. 2012;13:251.
50. Jasper LL, Jones, C. A., Mollins, J., Pohar, S. L., Beaupre, L. A.. Risk factors for revision of total knee arthroplasty: a scoping review. BMC Musculoskelet Disord. 2016;17:182.
51. Signorini DF, Weir NU. Any variability in outcome comparisons adjusted for case mix must be accounted for. BMJ. 1999;318(7176):128.
52. Swedish Hip Arthroplasty Register (SHAR). Annual Report 2018. https://registercentrum.blob.core.windows.net/shpr/r/Arsrapport_2018_Hoftprotes_ENG_26mars_Final-rJepCXNsLI.pdf.
53. Australian Orthopaedic Association National Joint Replacement Registry (AOANJRR). Annual report 2019. https://aoanjrr.sahmri.com/.
54. van Dishoeck AM, Koek MB, Steyerberg EW, van Benthem BH, Vos MC, Lingsma HF. Use of surgical-site infection rates to rank hospital performance across several types of surgery. Br J Surg. 2013;100(5):628-36; discussion 37.
55. van Dishoeck AM, Lingsma HF, Mackenbach JP, Steyerberg EW. Random variation and rankability of hospitals using outcome indicators. BMJ quality & safety. 2011;20(10):869-74.
56. Adab P, Rouse AM, Mohammed MA, Marshall T. Performance league tables: the NHS deserves better. BMJ. 2002;324(7329):95-8.
57. Cram P, Cai X, Lu X, Vaughan-Sarrazin MS, Miller BJ. Total knee arthroplasty outcomes in top-ranked and non-top-ranked orthopedic hospitals: an analysis of Medicare administrative data. Mayo Clin Proc. 2012;87(4):341-8.

58. Woodall WH. The use of control charts in health-care and public health surveillance. J Qual Technology. 2006;26:89-104.
59. Cecil E, Bottle, A., Esmail, A., Wilkinson, S., Vincent, C., Aylin, P. P.. Investigating the association of alerts from a national mortality surveillance system with subsequent hospital mortality in England: an interrupted time series analysis. BMJ quality & safety. 2018;27(12):965-73.
60. Cecil E, Wilkinson, S., Bottle, A., Esmail, A., Vincent, C., Aylin, P. P.. National hospital mortality surveillance system: a descriptive analysis. BMJ quality & safety. 2018;27(12):974-81.
61. Dyrkorn OA, Kristoffersen, M., Walberg, M.. Reducing post-caesarean surgical wound infection rate: an improvement project in a Norwegian maternity clinic. BMJ quality & safety. 2012;21(3):206-10.
62. Benning A, Ghaleb, M., Suokas, A., Dixon-Woods, M., Dawson, J., Barber, N., Franklin, B. D., Girling, A., Hemming, K., Carmalt, M., Rudge, G., Naicker, T., Nwulu, U., Choudhury, S., Lilford, R.. Large scale organisational intervention to improve patient safety in four UK hospitals: mixed method evaluation. BMJ. 2011;342:d195.
63. Nicolay CR, Purkayastha, S., Greenhalgh, A., Benn, J., Chaturvedi, S., Phillips, N., Darzi, A.. Systematic review of the application of quality improvement methodologies from the manufacturing industry to surgical healthcare. Br J Surg. 2012;99(3):324-35.
64. Woodall WH, Fogel, S.L., Steiner, S.H.. The monitoring and improvement of surgical-outcome quality. J Qual Technology. 2015(47):383-99.
65. Grigg O, Farewell, V.. An overview of risk-adjusted charts. J R Stat Soc Ser A Stat Soc. 2004;167:523-39.
66. Curran E, Harper, P., Loveday, H., Gilmour, H., Jones, S., Benneyan, J., Hood, J., Pratt, R.. Results of a multicentre randomised controlled trial of statistical process control charts and structured diagnostic tools to reduce ward-acquired meticillin-resistant Staphylococcus aureus: the CHART Project. J Hosp Infect. 2008;70(2):127-35.
67. Cohen ME, Liu, Y., Ko, C. Y., Hall, B. L.. Improved Surgical Outcomes for ACS NSQIP Hospitals Over Time: Evaluation of Hospital Cohorts With up to 8 Years of Participation. Ann Surg. 2016;263(2):267-73.
68. Hollesen RVB, Johansen, R. L. R., Rorbye, C., Munk, L., Barker, P., Kjaerbye-Thygesen, A.. Successfully reducing newborn asphyxia in the labour unit in a large academic medical centre: a quality improvement project using statistical process control. BMJ quality & safety. 2018;27(8):633-42.
69. Macpherson GJ, Brenkel, I. J., Smith, R., Howie, C. R.. Outlier analysis in orthopaedics: use of CUSUM: the Scottish Arthroplasty Project: shouldering the burden of improvement. J Bone Joint Surg Am. 2011;93 Suppl 3:81-8.
70. Neuburger J, Walker, K., Sherlaw-Johnson, C., van der Meulen, J., Cromwell, D. A.. Comparison of control charts for monitoring clinical performance using binary data. BMJ quality & safety. 2017;26(11):919-28.
71. Montgomery DC. Introduction to Statistical Quality Control. 6th ed New York: Wiley & Sons, 2009.
72. Spiegelhalter D, Sherlaw-Johnson C, Bardsley M. Statistical methods for healthcare regulation: rating, screening and surveillance. J R Stat Soc Ser A Stat Soc. 2012;175:1-47.
73. Hofstede SN, Ceyisakar IE, Lingsma HF, Kringos DS, Marang-van de Mheen PJ. Ranking hospitals: do we gain reliability by using composite rather than individual indicators? BMJ quality & safety. 2019;28(2):94-102.

74. Karthaus EG, Lijftogt N, Busweiler LAD, Elsman BHP, Wouters M, Vahl AC, et al. Textbook Outcome: A Composite Measure for Quality of Elective Aneurysm Surgery. Ann Surg. 2017;266(5):898-904.
75. Dimick JB, Birkmeyer NJ, Finks JF, Share DA, English WJ, Carlin AM, et al. Composite measures for profiling hospitals on bariatric surgery performance. JAMA Surg. 2014;149(1):10-6.
76. Lingsma HF, Bottle A, Middleton S, Kievit J, Steyerberg EW, Marang-van de Mheen PJ. Evaluation of hospital outcomes: the relation between length-of-stay, readmission, and mortality in a large international administrative database. BMC Health Serv Res. 2018;18(1):116.
77. Norwegian Arthroplasty Register (NAR). Annual Report 2020. http://nrlweb.ihelse.net/eng/Rapporter/Report2020_english.pdf.
78. Finnish Arthroplasty Register (FAR). Online Annual Report. https://www.thl.fi/far/#index.
79. Swiss Arthroplasty Register (SAR). Annual Report 2020. https://www.siris-implant.ch/de/Downloads&category=16.
80. Barclay M, Dixon-Woods M, Lyratzopoulos G. The problem with composite indicators. BMJ quality & safety. 2019;28(4):338-44.
81. Hofstede SN, van Bodegom-Vos L, Kringos DS, Steyerberg E, Marang-van de Mheen PJ. Mortality, readmission and length of stay have different relationships using hospital-level versus patient-level data: an example of the ecological fallacy affecting hospital performance indicators. BMJ quality & safety. 2018;27(6):474-83.
82. Rolfson O, Kärrholm J, Dahlberg LE, Garellick G. Patient-reported outcomes in the Swedish Hip Arthroplasty Register: results of a nationwide prospective observational study. J Bone Joint Surg Br. 2011;93(7):867-75.
83. Dunbar MJ, Richardson G, Robertsson O. I can't get no satisfaction after my total knee replacement: rhymes and reasons. Bone Joint J. 2013;95-b(11 Suppl A):148-52.
84. Bohm ER, Kirby S, Trepman E, Hallstrom BR, Rolfson O, Wilkinson JM, et al. Collection and Reporting of Patient-reported Outcome Measures in Arthroplasty Registries: Multinational Survey and Recommendations. Clin Orthop Relat Res. 2021;479(10):2151-66.
85. Rolfson O, Eresian Chenok K, Bohm E, Lübbeke A, Denissen G, Dunn J, et al. Patient-reported outcome measures in arthroplasty registries. Acta Orthop. 2016;87 Suppl 1(Suppl 1):3-8.
86. Johnston B, Patrick D, Devji T, al e. Chapter 18: Patient_reported outcomes. Cochrane Handbook for Systematic Reviews of Interventions. 2019.
87. Makhni EC. Meaningful Clinical Applications of Patient-Reported Outcome Measures in Orthopaedics. J Bone Joint Surg Am. 2021;103(1):84-91.
88. FDA. Guidance for Indusctry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. 2009.
89. Black N. Patient reported outcome measures could help transform healthcare. BMJ. 2013;346:f167.
90. Rolfson O, Wissig S, van Maasakkers L, Stowell C, Ackerman I, Ayers D, et al. Defining an International Standard Set of Outcome Measures for Patients With Hip or Knee Osteoarthritis: Consensus of the International Consortium for Health Outcomes Measurement Hip and Knee Osteoarthritis Working Group. Arthritis Care Res (Hoboken). 2016;68(11):1631-9.
91. National Joint Registry (NJR). Website: http://www.njrclinicianfeedback.org.uk/.
92. Gude WT, Roos-Blom MJ, van der Veer SN, Dongelmans DA, de Jonge E, Francis JJ, et al. Health professionals' perceptions about their clinical performance and the influence of audit and feedback on their intentions to improve practice: a theory-based study in Dutch intensive care units. Implement Sci. 2018;13(1):33.

93. Soong C, Shojania KG. Education as a low-value improvement intervention: often necessary but rarely sufficient. BMJ quality & safety. 2020;29(5):353-7.
94. Gude WT, Roos-Blom MJ, van der Veer SN, de Jonge E, Peek N, Dongelmans DA, et al. Electronic audit and feedback intervention with action implementation toolbox to improve pain management in intensive care: protocol for a laboratory experiment and cluster randomised trial. Implement Sci. 2017;12(1):68.
95. Roos-Blom MJ, Gude WT, de Jonge E, Spijkstra JJ, van der Veer SN, Peek N, et al. Impact of audit and feedback with action implementation toolbox on improving ICU pain management: cluster-randomised controlled trial. BMJ quality & safety. 2019;28(12):1007-15.
96. Taylor MJ, McNicholas C, Nicolay C, Darzi A, Bell D, Reed JE. Systematic review of the application of the plan-do-study-act method to improve quality in healthcare. BMJ quality & safety. 2014;23(4):290-8.

Chapter 2

International comparison of variation in performance between hospitals for THA and TKA: Is it even possible?

A systematic review including 33 studies and 8 arthroplasty register reports

Peter van Schie[1,2], Shaho Hasan[1], Leti van Bodegom-Vos[2], Jan W. Schoones[3], Rob G.H.H. Nelissen[1], Perla J. Marang- van de Mheen[2]

1. Department of Orthopaedics, Leiden University Medical Centre, Leiden, The Netherlands;
2. Department of Biomedical Data Sciences, Medical Decision Making, Leiden University Medical Centre, Leiden, The Netherlands;
3. Walaeus Library, Leiden University Medical Centre, Leiden, Netherlands.

EFORT Open Rev. 2022 Apr 21;7(4):247-263.

Abstract

Purpose

To improve care for total hip and knee arthroplasties (THA/TKA), hospitals may want to compare their performance with hospitals in other countries. Pooling data across countries also enables earlier detection of infrequently occurring safety issues. We therefore aimed to assess the between-hospital variation and definitions used for revision, readmission, and complications across countries.

Methods

PubMed, Embase, Web of Science, Cochrane library, Emcare, and Academic Search Premier were searched from January 2009 to August 2020 for studies reporting on: 1) Primary THA/TKA; 2) Revision, readmission, or complications; 3) Between-hospital variation. Most recent registry reports of Network of Orthopedic Registries of Europe (NORE) members were also reviewed. Two reviewers independently screened records, extracted data, and assessed the risk of bias (RoB) using the Integrated quality Criteria for the Review Of Multiple Study designs (ICROMS) tool for studies and relevant domains for registries. We assessed agreement for the following domains: 1) Outcome definition; 2) Follow-up and starting point; 3) Case-mix adjustment; 4) Type of patients and hospitals included.

Results

Between-hospital variation was reported in 33 (1 high-quality, 13 moderate-quality, and 19 low-quality) studies and 8 registry reports. The range of variation for revision was 0%-33% for THA and 0%-27% for TKA varying between assessment within hospital admission until 10 years of follow-up; for readmission 0%-40% and 0%-32%, respectively; and for complications 0%-75% and 0%-50%, respectively. Indicator definitions and methodological variables varied considerably across domains.

Conclusion

The large heterogeneity in definitions and methods used likely explains the considerable variation in between-hospital variation reported for revision, readmission, and complications , making it impossible to benchmark hospitals across countries or pool data for earlier detection of safety issues. It is necessary to collaborate internationally and strive for more uniformity in indicator definitions and methods in order to achieve reliable international benchmarking in the future.

Introduction

Arthroplasty registries were originally established to monitor safety and compare the survival of different types of implants. In recent years, however, registries have also been used to show between-hospital variation for various quality indicators and provide hospitals and surgeons with feedback on their performance, usually compared with a reference standard (i.e., the benchmark) which is mostly the national average.(1-8) Most registries give feedback through annual reports intended to encourage quality improvement initiatives in low-performing hospitals and learn from high-performing hospitals by adopting best practices.(1-8) In addition, scientific articles are published for quality improvement purposes; for example, hospitals are benchmarked, ranked, or (statistical) methods are compared to monitor the quality of care delivered.(9-13) The most commonly used quality indicators in this context are implant revision, readmission, and complications, as these indicators are considered reliable, actionable, and fit for purpose.(14-19) However, the reliability of hospital rankings has been shown to be affected by e.g., minor registration incompleteness in the outcome and low event rates, with particularly low volume providers being less likely to become an outlier in funnel plots.(11,13)

The rationale for benchmarking is that if another hospital treating comparable patients achieves better outcomes, there is potential to improve the underlying quality of care processes and patient outcomes. However, there may be less incentive to improve further for hospitals that are among the best performing hospitals in their own country. These hospitals may have an interest to compare their outcomes with hospitals from other countries or healthcare systems to stimulate further improvement. In addition, pooling of data across countries would also enable to detect any safety issues that occur with low frequency much earlier. Both of these are only possible if there is consistency in the indicator definitions and methods used to collect data, as these will determine the frequency of occurrence. For example, a previous study showed that a change in definition within the same surgical context increased the occurrence of adverse outcomes from 7% to 27%.(20) Similarly, data from one study where the complication rate is defined as the case-mix adjusted proportion of complications within 14 days post-surgery cannot be pooled with another study where it is defined as non-case-mix adjusted proportion of complications within 30 days post-surgery.(9,21)

The present study, therefore, aims to systematically assess the between-hospital variation and definitions used for revision, readmission, and complications after total hip and knee arthroplasties (THA and TKA) across countries, including both

scientific papers published in the past decade and the most recent arthroplasty registry reports from the Network of Orthopedic Registries of Europe (NORE).

Materials and Methods

This systematic review was registered at inception with PROSPERO (CRD42019122779) and conducted according to the PRISMA 2020 statements. (22) The authors received a grant from the Van Rens Foundation (VRF2018-001) to perform this study.

Search strategy

PubMed, Embase, Web of Science, Cochrane library, Emcare, and Academic Search Premier were searched for publications from January 2009 to August 2020 using a systematic search created by a librarian (JS). The search consisted of three components: 1) Primary THA/TKA; 2) Revision, readmission, complication, length-of-stay (LOS), and mortality; 3) Between-hospital variation (Appendix I). LOS and mortality were included as secondary outcomes. LOS was included because it indicates the severity/complexity of patients treated or more time to identify complications during admission, both of which may influence the need for subsequent readmission. Prolonged LOS may also be a proxy for a complicated disease course, even without these complications being reported. Therefore between-hospital variation in LOS can act as a proxy for between-hospital variation in complications within a given healthcare system. Mortality was included because this is a highly undesirable outcome.

Study and report selection

Titles and abstracts were screened independently by two reviewers (PvS/SH), and discrepancies resolved through discussion. Senior researchers (PM/RN) were available if consensus could not be reached. Inclusion criteria were studies reporting on 1) Primary THA and/or TKA; 2) National or regional between-hospital variation for revision, readmission, complication, LOS, or mortality with at least 2 hospitals included. All studies using registry, administrative, claim, or audit data were directly included for full-text screening, as these are usually national or regional studies that are likely to report between-hospital variations even if not included in the title and abstract. Reviews and study protocols were excluded. Studies in English, Dutch, German, French, and Danish were eligible for inclusion and were translated by both reviewers (PvS/SH). Authors were contacted if the full text could not be found.

In parallel, all most recent registry reports of NORE members including registries in and outside Europe, were reviewed in full-text on reporting between-hospital variation for the same indicators.(23)

Data extraction

Data were extracted independently by two reviewers (PvS/SH) using a prespecified SPSS file (Version 26, IBM Corp). Data extracted were first author, title, year of publication, country of the first author, and type of implant (i.e., THA and/or TKA). For arthroplasty reports, the first author was replaced by the country or region of origin. In addition, data sources, data collection period, and data completeness were collected, and the number of patients and hospitals included. The between-hospital variations as reported for the outcomes were collected in the original unit, including mean, standard deviation (sd), standard error (se), 95%-confidence interval (95%-CI), median, interquartile range (IQR), and range. If between-hospital variation was not reported in the text, but hospital outcomes were reported individually, hospital variation was calculated using the individual hospital outcomes. If the variation was only reported in a graph, the values were derived from the chart. Outcome definitions and any adjustment for case-mix were also collected and the type of patients and/or hospitals included. In addition, we documented for what purpose the between-hospital variation was reported (e.g. pay for performance or quality improvement) and whether it was reported using one overall estimate (i.e., mean (sd), median (IQR) or range) or whether also individual hospitals outcomes were shown (e.g., in funnel plots or forest plots).

Definition of outcomes

All outcomes were reviewed on the following domains: 1) Outcome definition (i.e., what constitutes a revision, readmission, or complications); 2) Follow-up and starting point (e.g., post-discharge or post-surgery); 3) Case-mix adjustment (yes/no); 4) Type of patients (e.g., osteoarthritis or trauma) and hospitals (e.g., hospital type or size) included. For each outcome, it was assessed how often perfect agreement was reached across all these domains, which would be needed to allow for the pooling of data. In addition to documenting case-mix adjustment or not, it was assessed for which confounding factors the between-hospital variation was adjusted.

Data analysis

The between-hospital variation for revision, readmission, complications, LOS, and mortality was reported separately for THA, TKA and THA&TKA combined and plotted in a forest plot. When available, the mean, median, and range were plotted, and when both 95%-CI and IQR were available, only the IQR was plotted. When mean and se were available, we calculated the 95%-CI. If only the sd was available,

the se was calculated by dividing the sd by the square root of the number of hospitals included.(24) If variation for an outcome was longitudinally reported multiple times, the most recent variation was reported and plotted. Data were not pooled as there was considerable heterogeneity, in which case it is recommended to refrain from pooling as the resulting estimate will be rather unreliable.(25)

Risk of bias assessment

The Integrated Quality Criteria for Review of Multiple Study Designs (ICROMS) was used to assess the risk of bias (RoB) independently by both reviewers (PvS/SH).(26) The ICROMS is a comprehensive tool to evaluate the quality of multiple study designs and includes a set of universally applicable and study-specific quality criteria for each study design. Every study design must meet a minimum score and mandatory criteria to be included in the review. The specific criteria for cohort studies and controlled before-after studies were addressed as these were the study designs included in this review (Table 1A). We included all studies independent of the ICROMS score and reported the RoB for every study, with the rationale that RoB could be taken into account when weighting study results, whereas excluding studies with medium or low RoB could result in the loss of potentially valuable information. Studies scoring at least 18 points out of the total of 26 points for cohort studies or at least 18 of the 28 points for controlled-before-after studies and meeting the mandatory criteria were classified as high-quality (HQ) studies.(26) Studies scoring at least 18 points for both study designs but failing to meet the mandatory criteria were classified as moderate-quality (MQ) studies. Studies scoring less than 18 points for both study designs were classified as low-quality (LQ) studies.

Since there is no tool available to assess the RoB for registry reports, we tailored the RoB assessment to our research question, that is those factors that could potentially bias the between-hospital variation as reported in registries (Table 1B). Consistent with the ICROMS tool, each item could get 0-2 points, resulting in a range of 0-14 points. No reports were excluded based on the RoB assessment, but the RoB could be considered when weighing the reports' results.

Results

Studies and reports

The search identified 1,643 records, including 1621 scientific papers and 22 registry reports. After removing duplicates, 943 remained (i.e., 921 studies and 22 reports). Title and abstract screening excluded 625 studies, as 157 did not involve primary THA or TKA, 373 did not report between-hospital variation, 38 did not report at least one

of the outcomes, 54 were reviews, and 3 were study protocols. All registry reports were directly selected for full-text screening, resulting in 22 reports and 296 papers to be further assessed. During the full-text review, 14 reports and 263 papers were excluded as 3 did not involve primary THA or TKA, 270 did not report between-hospital variation, and 4 did not report at least one indicator outcome, leaving 33 papers and 8 arthroplasty reports to be included (Figure 1).

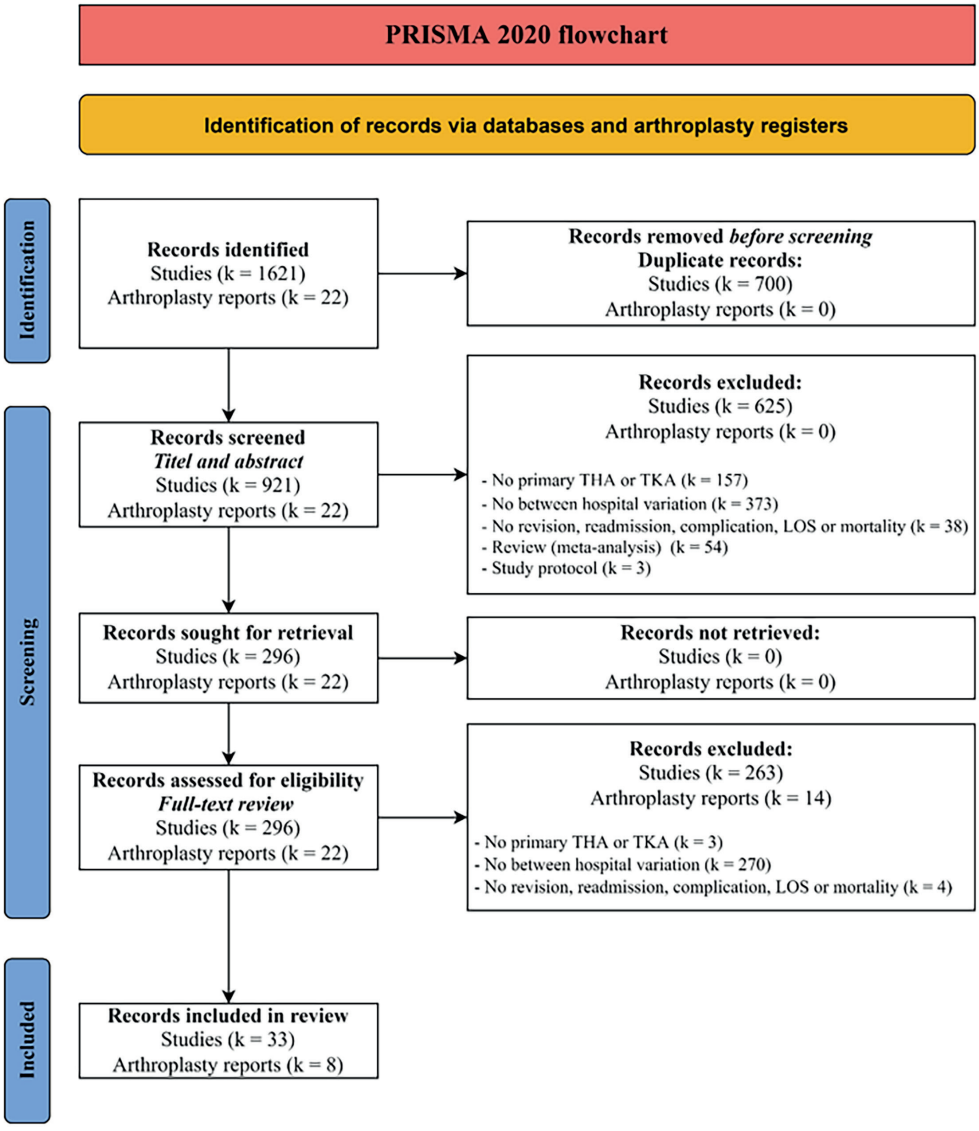


Figure 1 PRISMA 2020 flowchart
THA=total hip arthroplasty; TKA=total knee arthroplasty.

Table 1A Risk of Bias (RoB) studies

Study number	Author	Year	Design	1A*	2D*	2E*	3B*	3C*	3E	3F	3G*	4C*	5A	5B	F3	7A	7B	7C	7D	7E	ICROMS score
1	Bozic(14)	2014	CS	2		2			0	2	2	2		1	2	2	2	2	1	2	22
2	Thirukumaran(15)	2020	CBA	2	2		2	1	0	2		0	2		2	2	2	2	2	0	21
3	Courtney(39)	2018	CS	2		1			0	2	2	2		1	2	1	2	2	1	2	20
4	van Schie(30)	2020	CS	2		1			0	2	2	0		1	2	2	2	2	1	2	19
5	Graham(21)	2019	CS	2		1			0	2	2	0		1	2	2	2	2	1	2	19
6	Sheetz(9)	2019	CS	2		1			0	2	2	0		1	2	2	2	2	1	2	19
7	Bottle(36)	2018	CS	1		1			0	1	2	2		1	2	2	2	2	1	2	19
8	Padegimas(34)	2018	CS	2		1			0	2	2	0		1	2	2	2	2	1	2	19
9	Marang-vd Mheen(50)	2017	CS	2		1			0	2	2	0		1	2	2	2	2	1	2	19
10	Hollis(40)	2017	CS	2		1			0	2	2	0		1	2	2	2	2	1	2	19
11	Qian(42)	2013	CS	2		1			0	2	2	0		1	2	2	2	2	1	2	19
12	Voorn(10)	2017	CS	2		1			0	1	2	0		1	2	2	2	2	1	2	18
13	Cram(43)	2012	CS	2		1			0	2	2	0		1	1	2	2	2	1	2	18
14	Cai(49)	2012	CS	2		1			0	1	2	0		1	2	2	2	2	1	2	18
15	Padegimas(31)	2017	CS	2		1			0	2	2	0		1	2	2	2	2	1	0	17
16	Chen(52)	2017	CS	2		1			0	2	2	0		1	2	2	2	2	1	0	17
17	Courtney(16)	2017	CS	2		0			0	1	2	1		1	1	2	2	2	1	2	17
18	Husni(54)	2010	CS	2		0			0	2	1	2		1	1	2	2	1	1	2	17
19	Hofstede(41)	2018	CS	2		1			0	1	2	0		1	2	2	2	2	1	0	16
20	Pross(27)	2017	CS	2		1			0	1	2	0		1	2	2	2	2	1	0	16
21	Calderwood(44)	2013	CS	2		1			0	2	2	2		1	1	2	0	2	1	0	16
22	Kurtz(38)	2016	CS	2		1			0	1	1	0		1	2	2	2	2	1	0	15
23	Kurtz(37)	2016	CS	2		1			0	1	1	0		1	2	2	2	2	1	0	15
24	Jergesen(48)	2016	CS	2		1			0	1	1	0		1	1	2	2	1	1	2	15
25	Makela(32)	2011	CS	2		1			0	1	2	2		1	2	2	0	1	1	0	15
26	Mittal(28)	2018	CS	2		1			0	1	1	0		1	1	2	2	2	1	0	14
27	Skufca(45)	2017	CS	2		1			0	1	2	0		1	0	2	2	2	1	0	14
28	Asaid(51)	2013	CS	2		0			0	2	1	0		1	1	2	2	2	1	0	14

Table 1A Continued

Study number	Author	Year	Design	1A*	2D*	2E*	3B*	3C*	3E	3F	3G*	4C*	5A	5B	F3	7A	7B	7C	7D	7E	ICROMS score
29	**Dailey(46)**	2009	CS	**2**		**1**			0	2	**2**	**0**		1	1	2	0	2	1	0	**14**
30	**Martino(53)**	2018	CS	**2**		**1**			0	1	**2**	**0**		1	1	2	0	1	0	2	**13**
31	**Singh(33)**	2017	CS	**2**		**1**			0	1	**1**	**0**		1	0	2	2	1	1	0	**12**
32	**Husted(29)**	2010	CS	**2**		**0**			0	2	**1**	**0**		1	1	2	0	2	1	0	**12**
33	**Lopez-Contreras(47)**	2012	CS	**1**		**1**			0	1	**2**	**0**		1	1	2	0	1	1	0	**11**

32 cohort studies (CS) and 1 controlled before-and-after study were included. Following domains were assessed: 1A*) Clear statement of the aims of the research; 2D*) Intervention and control group selection designed to protect against systematic difference/selection bias; 2E*) Comparability of groups; 3B*) Baseline measurement-protection against selection bias; 3C*) Protection against contamination; 3E) Protection against detection bias: blinded assessment of primary outcome; 3F) Reliable primary outcome measure; 3G*) Comparability of outcomes; 4C*) Incomplete outcome data addressed; 5A) Protection against detection bias: Intervention unlikely to affect data collection; 5B) Protection against information bias; 6C) Analysis sufficiently rigorous/free from bias; 7A) Free of selective outcome reporting; 7B) Limitations addressed; 7C) Conclusions clear and justified; 7D) Free of other bias; 7E) Ethics issues addressed.

Scores for each domain were assigned as follows: 0=Did not fulfil the criteria; 1=Unclear if criteria are fulfilled; 2=Did meet the criteria.

*Indicates the mandatory criteria, and these criteria are darker coloured and in bold.

The green highlighted study has an ICROMS score ≥18, and fulfils the mandatory criteria, and was therefore classified as low RoB/high quality. The grey highlighted studies have an ICROMS ≥18 points, but do not fulfil the mandatory criteria and were therefore classified as moderate RoB/moderate quality. The red highlighted studies have an ICROMS <18 points, and do not fulfil the mandatory criteria, and were therefore classified as high RoB/low quality.

CBA=controlled before-after study; CS=cohort study; ICROMS=integrated quality criteria for review of multiple study designs.

Table 1B Risk of Bias (RoB) arthroplasty reports

Report code	Arthroplasty report	Year	1	2*	3	4	5**	6***	7****	Total RoB score for reports
A	**Norwegian Arthroplasty Register(1)**	2020	2	2	2	2	1	2	0	**11**
B	**Dutch Arthroplasty Register(2)**	2020	2	2	0	2	2	0	2	**10**
C	**Swedish Knee Arthroplasty Register(3)**	2020	2	2	2	2	2	2	0	**12**
D	**Danish Hip Arthroplasty Register(4)**	2020	2	2	2	1	0	2	0	**9**
E	**Swedish Hip Arthroplasty Register(5)**	2018	2	2	2	2	2	0	0	**10**
F	**Danish Knee Arthroplasty Register(6)**	2020	2	2	2	1	0	0	0	**7**
G	**Finnish Arthroplasty Register(7)**	2020	2	2	2	0	0	1	0	**7**
H	**Swiss Arthroplasty Register(8)**	2020	1	2	2	0	0	2	0	**7**

Since there is no tool available to assess RoB for registry reports, we tailored the ICROMs to our research question, i.e. those factors that could potentially bias the between-hospital variation as reported in registries. Following domains were assessed: 1) Patients could be traced when treated in another hospital; 2) Data completeness was reported for THA and TKA separately; 3) Data completeness was reported for single hospitals; 4) Indicator outcomes were validated for at least a part of the data; 5) Indicator outcomes were adjusted for covariates; 6) Missing data for covariates were reported; 7) Missing values for covariates were imputed.
Scores for each criterium were assigned as follows: 0 (red-color)=Did not fulfil the criteria; 1 (yellow-color)=Unclear if criteria are fulfilled; 2 (green-color)=Did fulfil the criteria.
*For this domain. 0=No; 1=Yes, for THA and TKA combined; 2=Yes, for THA and TKA separately. **For this domain. 0=No; 1=Yes, for age and gender; 2=Yes, for age, gender and comorbidities. ***For this domain. 0=Did not fulfil the criteria; 1=for at least one covariate; 2=Did fulfil the criteria. ****For this domain. 0=Data were not imputed; 1=Unclear if criteria are fulfilled; 2=Data were imputed.

Risk of bias

Thirty-two cohort studies and 1 controlled before-and-after study were included. One study was classified as a HQ study, 13 as MQ, and 19 as LQ. The median ICROMS score was 17 points (IQR:15-19). Most studies did not meet the mandatory criteria, often involving the comparability of groups (Table 1A, domain 2E) and incomplete outcome data addressed (Table 1A, domain 4C).

The RoB for registry reports ranged from 7 for the Swiss Arthroplasty Register to 11 for the Norwegian Arthroplasty Register (out of the maximum of 14). The median score was 9 points (IQR:7-10). Most variation was in the covariates used to adjust outcomes (Table 1B, domain 5) and whether missing values for covariates were imputed (Table 1B, domain 7).

Characteristics

Seven of the 33 studies included THA, 4 TKA, 12 both THA and TKA, and 10 studies combined THA&TKA as one group (Appendix IIA). Six studies reported the between-hospital variation for revision (18%), 13 for readmission (39%), and 20 for complications (61%). The studies included numbers of patients ranging from 122 to 524,892 for THA, from 84 to 952,593 for TKA, and from 1596 to 878,098 for THA&TKA combined. Four studies did not report the number of patients included. (15,27-29) A wide range was found for the number of hospitals included for THA, TKA, and THA&TKA combined (i.e., 2 to 3479). Data completeness on THA/TKA included was reported in 8 of the 33 studies and was at least 75% for 7 studies. Twenty studies used administrative data and 13 clinical data. Data were routinely collected for 23 studies and clinician-reported for 10 studies (Appendix IIA).

Two of the 8 registry reports included THA, 2 TKA, and 4 both THA and TKA. All reports included between-hospital variation for revision, 2 for readmission, and 3 for complications. Reports showed smaller variation in the number of patients (i.e., 7161 to 33248) and hospitals included (i.e., 47 to 152) compared with studies. All reports stated their overall data completeness in THA/TKA included to be at least 94.9%, but was only reported for individual hospitals by 7 reports (Appendix IIB).

Between-hospitals variation and indicator definitions

Revision

From the 6 studies reporting revision rates, the between-hospital variation was reported in 5 studies (1 MQ, 4 LQ) for THA(27,30-33) and 4 studies (2 MQ, 2 LQ) for TKA(30,32-34). Notable differences were seen with regard to the extent of between-hospital variation across studies, as shown in Figure 2. This is likely due to the large variety in definitions used, both to indicate what constituted a revision, the follow-up, patient selections, and whether hospital differences were adjusted for case-mix (Table 2). Revision within 1 year was mainly reported (33% of studies reporting revision), but even then, the variation remained large (Figure 2).

Revisions were reported in all 8 registry reports, but the between-hospital variation was reported in 6 reports for both THA(2,4,5,7,8,35) and TKA(2,3,6-8,35). Again, there were notable differences in the between-hospital variation across registries (Figure 2). As for the included studies, we found a large variety in definitions used, both in what constituted a revision, the follow-up, type of patients and hospitals selected, and adjustment for case-mix (Table 2). The only aspect where all reports were consistent was that follow-up started post-surgery. Revision within 5 years was most reported regarding follow-up (50% of reports reporting revision), followed by 1 year (38%) and 2 year (38%) revision.

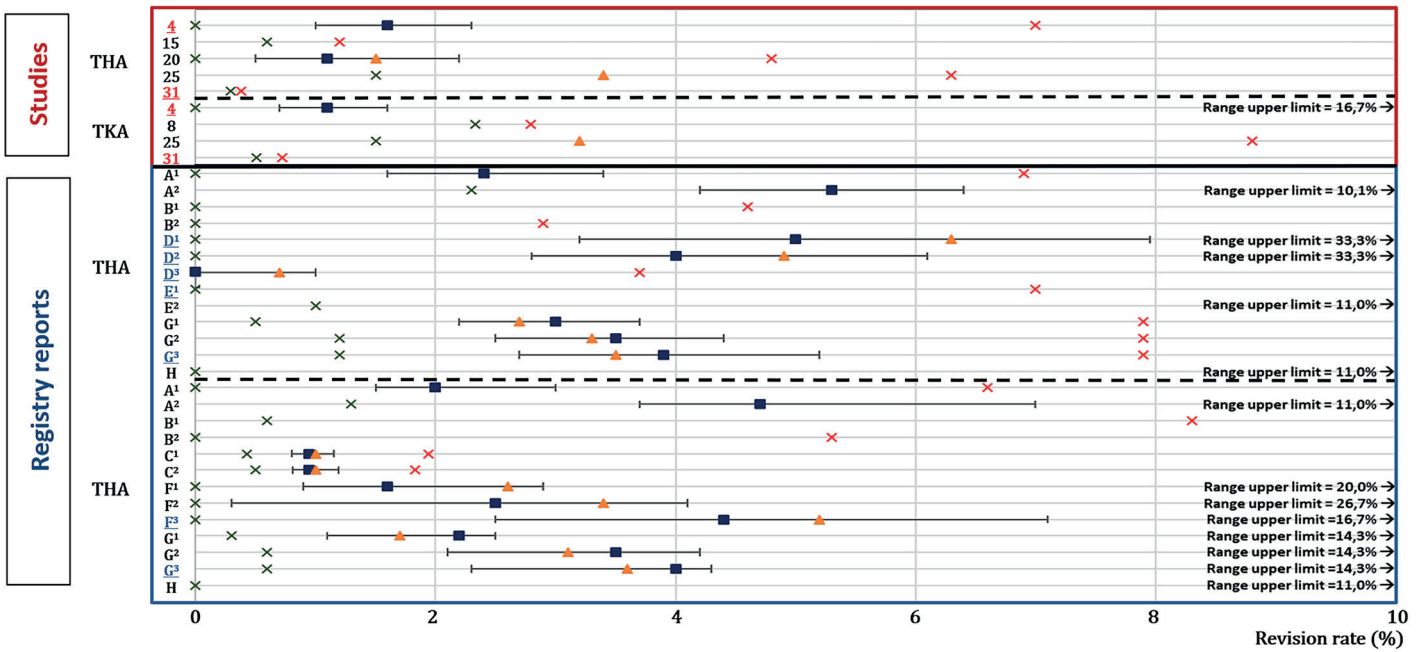


Figure 2 Between-hospital variation for revision

The numbers on the y-axis correspond with the study numbers from Table 1A and the letters on the y-axis with the report codes from Table 1B. A letter in superscript was added to a study number or a number in superscript to a report code when the revision rate was reported more than once with different definitions. Study numbers were underlined and red coloured when revision within 1 year was reported and registry reports were underlined and blue coloured when revision within 5 years was reported. The green and red cross represent the lower and upper range, respectively. The blue square represents the median and the yellow triangle the mean. The interquartile range is shown in a solid line through the median. Results of outlier procedures are shown below.

THA studies: 4=No:13, Po:18, Others:66. **TKA studies:** 4=No:7, Po:14, Others:77. **THA registry reports:** A^2=No:2, Po:8, Others:28; B^1=No:19, Po:31, Others:45; 2^B=No:18, Po:19, Others:58; D^1=No:4, Po:3, Others:36; D^2=No:3, Po:3, Others:36); D3=No:3, Po:3, Others:36; E^1=No:7, Po:12, Others:56); E^2=No:8, Po:14, Others:58; H=No:9, Po:5, Others:138. **TKA registry reports:** A^2=No:6, Po:6, Others:21; B^1=No:20, Po:24, Others:53; B^2=No:22, Po:19, Others:56; H=No:12, Po:13, Others:128.

No=negative outlier; Po=positive outlier; THA=total hip arthroplasty; TKA=total knee arthroplasty.

Readmission

From the 13 studies reporting readmission rates, the between-hospital variation was reported in 5 studies (3 MQ, 2 LQ) for THA(9,21,31,36,37), 4 studies (3 MQ, 1 LQ) for TKA(21,34,36,38) and 6 studies (3 MQ, 3 LQ) for THA&TKA combined(15,16,28,39-41). Ten studies reported the variation for readmission more than once with different indicator definitions in all domains except for the type of patients selected (Table 3). Figure 3 shows large differences in the between-hospital variation across studies and the reported means and medians, likely due (at least in part) to variety in how readmissions were defined and which patients were included (Table 3). Studies combining THA&TKA in a single group were mostly case-mix adjusted, whereas studies reporting only THA and/or TKA separately were often unadjusted. Readmission within 30 days was the most often used definition (85% of studies reporting readmission).

Overall readmissions were reported in 2 registry reports (25% of reports), but the between-hospital variation was only given in 1 report with 3 different patient selections (i.e., all patients, only with osteoarthritis or with a fracture) for THA(4) and in 1 report for TKA(6) (Figure 3 and Table 3). All-cause readmission within 30 days post-surgery was reported for THA and readmission of at least 2 days within 30 days after discharge for TKA. No adjustments for case-mix were performed for these data.

Complications

From the 20 studies reporting complication rates, the between-hospital variation was reported in 11 studies (5 MQ, 6 LQ) for THA(9,10,21,27,42-48), 8 studies (4 MQ, 4 LQ) for TKA(10,21,34,45-49) and 8 studies (1 HQ, 3 MQ, 4 LQ) for THA&TKA combined(14-16,33,39,48,50,51). Eight studies reported the variation more than once with different outcome definition, follow-up, and type of hospitals selected (Table 4). Again, large differences were found in the between-hospital variation, which is (at least) partly explained by the different definitions used (Figure 4 and Table 4). Two studies used the same dataset and reported comparable between-hospital variations.(16,39) Studies varied particularly in the type of complications included, such as reoperations, surgical site infections, blood transfusions, and deep venous thrombosis. There were also large differences in follow-up, type of patients and hospitals selected, and whether between-hospital variation was adjusted for case-mix. Complications were mostly defined as occurring within 30 days (15% of studies reporting complications).

Table 2 Definitions to report between-hospital variation for revision

Revision	Studies (*n*=6)		Registry reports (n=8)	
	THA (*n*=5) [4,15,20,25,31]	TKA (*n*=4) [4,8,25,31]	THA (*n*=13) [A^1,A^2,B^1,B^2,D^1,D^2,D^3, E^1,E^2,G^1,G^2,G^3,H]	TKA (*n*=13) [A^1,A^2,B^1,B^2,C^1,C^2,F^1, F^2,F^3,G^1,G^2,G^3,H]
1) Outcome definition				
• Exchange, removal, or addition of any component	[4,25]	[4,25]	[A^1,A^2,B^1]	[A^1,A^2]
• Revision of at least acetabulum or femur component	---	---	[B^2]	---
• Revision of at least femur or tibia component	---	---	---	[B^1]
• Parts or the whole prosthesis is changed or extracted	---	---	[E^1, E^2]	[B^2]
• Parts or complete of the primary implant are replaced	---	---	[H]	[H]
• One or more of the components were exchanged, removed, or added, including soft tissue procedures	---	---	[G^1-G^3]	[G^1-G^3]
			---	[C^1,C^2]
• One or more of the components are exchanged, removed, or added, including arthrodesis or amputation	---	---		---
			---	
• Due to infection	[31]	[31]		---
• Due to aseptic loosening	---	---	[D^3]	
• Not specified	[15,20]	[8]	[D^1,D^2]	[F^1-F^3]
2a) Follow-up				
• Within hospital admission	[20]	---	---	---
• Within 90 days	[15]	[8]	---	---
• Within 1 year	[4,31]	[4,31]	[B^1,B^2,G^1]	[F^1, G^1]
• Within 2 years	---	---	[A^1,H]	[A^1,F^2,H]
• Within 3 years	---	---	[G^2]	[B^1,B^2,G^2]
• Within 5 years	---	---	[D^1-E^1,G^3]	[F^3,G^3]
• Within 10 years	---	---	[A^2,E^2]	[A^2]
• Any revision within a period of time	[25]	[25]	---	[C^1,C^2]
2b) Follow-up starting point				
• Post-surgery	[4,31]	[4,31]	[A^1-H]	[A^1-H]
• Not specified	[15,25]	[8,25]	---	---
• Not applicable	[20]	---	---	---
3) Case-mix adjusted				
• Yes	[4,25]	[4,7]	[B^1,B^2,H]	[B^1-C^2,H]
• No	[15,20,25]	[8,31]	[A^1,A^2,D^1-G^3]	[A^1,A^2,F^1-G^3]

Table 2 Continued

Revision	Studies (*n*=6)		Registry reports (n=8)	
	THA (*n*=5) [4,15,20,25,31]	**TKA (*n*=4)** [4,8,25,31]	**THA (*n*=13)** [A^1,A^2,B^1,B^2,D^1,D^2,D^3, E^1,E^2,G^1,G^2,G^3,H]	**TKA (*n*=13)** [A^1,A^2,B^1,B^2,C^1,C^2,F^1, F^2,F^3,G^1,G^2,G^3,H]
4a) Type of patient selected				
• Age selection(s)	---	---	[A^1,A^2]	[A^1,A^2]
• Osteoarthritis	[25]	[25]	[A^1,A^2,D^2]	[A^1,A^2,C^1,C^2]
• No trauma patients	[15]	[8]	---	---
• Comorbidity score selection	[15]	[8]	[A^1,A^2]	[A^1,A^2]
• Matching of patient groups	[15]	[8]	---	---
• No selections, all patients included	[4,20,31]	[4,31]	[B^1-D^1,D^3-H]	[B^1,B^2,F^1-H]
4b) Type of hospitals selected				
• Number of procedures limit	---	---	[A^1,A^2,E^1,E^2]	[A^1,A^2,C^1,C^2]
• Completeness of data limit	---	---	[A^1,A^2]	[A^1,A^2]
• No selections, all hospitals included	[4-31]	[4-31]	[B^1-D^3,G^1-H]	[B^1,B^2,F^1-H]

The definitions for revision were defined for 6 domains. The numbers in brackets correspond to the study number from Table 1A and the letters in brackets to the report codes from Table 1B.

No=negative outlier; Po=positive outlier; THA=total hip arthroplasty; TKA=total knee arthroplasty.

Complications were reported in 3 reports (38% of reports), with between-hospital variation reported in 2 reports for THA(4,5) and 1 report for TKA(3). All reports reported the variation more than once with different outcome definition, follow-up, and type of patients selected (Table 4). As with data reported from studies, a large between-hospital variation was found, although less variation in the type of complications was present, but more variation in the type of patients selected (Figure 4 & Table 4).

Length-of-stay and mortality

Nine studies (27% of studies)(10,29,31,34,41,50,52-54) reported the between-hospital variation for LOS and only 1 study (3% of studies)(9) represented mortality. Between-hospital variation for LOS was given by 1 report (13% of reports)(6), and for mortality by 2 reports (25% of reports).(3,5) (appendixes III-VI).

Perfect agreement

Given the heterogeneity in definitions used across studies and registry reports, none of the outcomes had perfect agreement across all 6 domains (i.e., what constituted a revision, readmission or complications, follow-up, and starting point, case-mix adjustment, and patient- and hospital selections) for both THA, TKA, and THA&TKA combined (Tables 2-4).

Variables used for case-mix adjustment

Both studies and reports varied whether rates were case-mix adjusted and which variables were used for case-mix adjustment (Tables 2-4). Revision rates, when adjusted, were always adjusted for age and gender. Considerable variation was observed with regard to additional case-mix adjustments: American Society of Anaesthesiology (ASA) score (2,8,30), diagnosis (osteoarthritis versus other)(2,30), body mass index (BMI) (8,30), Charnley score(8,30), smoking status(30), use of patellar button(3), previous contralateral arthroplasty(32), bilaterally of the operation(32), heart disease(32), hypertension(32), cancer(32), alcoholism(32), dementia(32), depression(32), Parkinson's disease(32), mental disorders(32), degenerative brain diseases(32) and atherosclerosis(32).

For readmissions, also wide variation in case-mix adjustments: for age(9,15,21,39), gender(9,15,21), ethnicity(21), functional status(21), ASA score(21), history of acute myocardial infarction(21), history of peripheral vascular disease(21), depression(21), diabetes mellitus(21), surgical time(21), work relative value unit(21), emergency surgery(21), patient comorbidities (16,28,39), Elixhauser comorbidities(9) procedure (THA/TKA)(16), demographics(40), healthcare use(40), comorbidities selected by veteran affair surgical quality improvement programme (VASQIP) nurses(40) and clinical comorbidity(15).

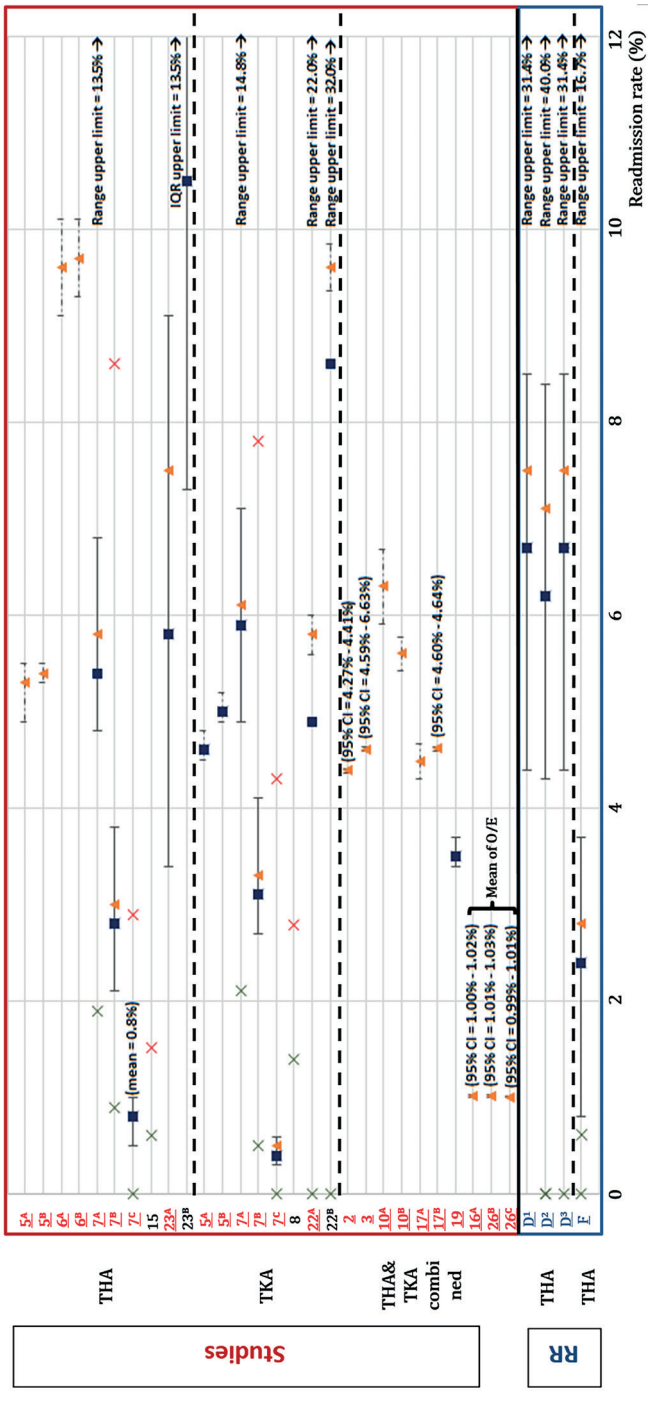


Figure 2 Between-hospital variation for readmission

The numbers on the y-axis correspond with the study numbers from Table 1A and the letters on the y-axis with the report codes from Table 1B. A letter in superscript was added to a study number or a number in superscript to a report code when the readmission rate was reported more than once with different definitions. Study numbers and report codes were underlined and red/blue coloured when readmission within 30 days was reported. The green and red cross represent the upper and lower range, respectively. The blue square represents the median and the yellow triangle the mean. The interquartile range is shown in a solid line through the median. The 95% confidence interval is shown with a dashed line through the mean. Results of outlier procedures are shown below. **THA registry reports:** D1=No:2, Po:8, Others:30; D2=No:0, Po:9, Others:31; D3=No:3, Po:2, Others:35. No=negative outlier; O/E=observed divided by expected; Po=positive outlier; RR=Registry reports; THA=total hip arthroplasty; TKA=total knee arthroplasty; 95% CI=95% confidence interval.

Table 3 Definitions to report between-hospital variation for readmission

Readmission	Studies (*n*=13)			Registry reports (*n*=2)	
	THA (*n*=10) [5[A],5[B],6[A],6[B],7[A],7[B],7[C],15,23[A],23[B]]	**TKA (*n*=8)** [5[A],5[B],7[A],7[B],7[C],8,22[A],22[B]]	**THA&TKA (*n*=10)** [2,3,10[A],10[B],17[A],17[B],19,26[A],26[B],26[C]]	**THA (*n*=3)** [D[1],D[2],D[3]]	**TKA (*n*=1)** [F]
1) Outcome definition					
• All-cause	[5[B],6[A],6[B],7[A],15]	[5[B],7[A],8]	[2-10[B],26[A]-26[C]]	[D[1]-D[3]]	---
• Emergency only	---	---	[19]	---	---
• Related to surgery	[5[A],7[B]]	[5[A],7[B]]	---	---	---
• Return to theatre	[7[C]]	[7[C]]	---	---	---
• Specific composition	[23[A],23[B]]	[22[A],22[B]]	[17[A],17[B]]	---	[F]
2a) Follow-up					
• Within 30-days	[5[A]-7[C],23[A]]	[5[A]-7[C],22[A]]	[2-26[C]]	[D[1]-D[3]]	[F]
• Within 90-days	[15,23[B]]	[8,22[B]]	---	---	---
2b) Fu time starting point					
• Post-surgery	---	---	---	[D[1]-D[3]]	---
• Post-discharge	[5[A],5[B],6[A],6[B],7[A],7[B],7[C],23[A],23[B]]	[5[A]-7[C],22[A],22[B]]	[2-26[C]]	---	[F]
• Not specified	[15]	[8]	---	---	---
3) Case-mix adjusted					
• Yes	[5[A]-6[B]]	[5[A]-5[B]]	[2,3,10[B]-17[B],26[A]-26[C]]	---	---
• No	[7[A]-23[B]]	[7[A]-22[B]]	[10[A],19]	[D[1]-D[3]]	[F]
4a) Type of patient selected					
• Age selection(s)	[6[A],6[B],23[A],23[B]]	[22[A],22[B]]	[2,3,26[A]-26[C]]	---	---
• Osteoarthritis	---	---	[19]	[D[2]]	[F]
• No trauma patients	[15]	[8]	---	---	---
• Medicare patients	[6[A],6[B],23[A],23[B]]	[22[A],22[B]]	[2,3,17[A],17[B],26[A]-26[C]]	---	---
• Elective surgery	[7[A],7[B],7[C]]	[7[A]-7[C]]	---	---	---
• If LOS ≥2 days	[5[A],5[B]]	[5[A],5[B]]	[10[A],10[B]]	---	---
• Minimum LOS of readmission	---	---	---	---	[F]
• Matching of patient groups	[15]	[8]	---	---	---
• Fracture patients	---	---	---	[D[3]]	---
• No selections	---	---	---	[D[1]]	---

Table 3 Continued

	Studies (*n*=13)			Registry reports (*n*=2)	
Readmission	**THA (*n*=10)** [5[A],5[B],6[A],6[B],7[A],7[B],7[C],15,23[A],23[B]]	**TKA (*n*=8)** [5[A],5[B],7[A],7[B],7[C],8,22[A],22[B]]	**THA&TKA (*n*=10)** [2,3,10[A],10[B],17[A],17[B],19,26[A],26[B],26[C]]	**THA (*n*=3)** [D[1],D[2],D[3]]	**TKA (*n*=1)** [F]
4b) Type of hospitals selected					
• Number of procedures limit	---	---	[3-17[B]]	---	---
• Veteran Affairs Hospitals	[5[A],5[B]]	[5[A],5[B]]	[10[A],10[B]]	---	---
• Government hospitals	---	---	[26[A]]	---	---
• Proprietary hospitals	---	---	[26[B]]	---	---
• Non-profit hospitals	---	---	[26[C]]	---	---
• Honor roll hospitals	[6[A]]	---	---	---	---
• Affiliated honor roll hospitals	[6[B]]	---	---	---	---
• Physician-owned	---	---	17[A]	---	---
• Non-physician owned	---	---	17[B]	---	---
• No selections	[7[A]-23[B]]	[7[A]-22[B]]	[2,19]	---	---

The definitions for readmission were defined for 6 domains. The numbers in brackets correspond to the study number from Table 1A and the letters in brackets to the report codes from Table 1B.

Fu=Follow-up; LOS=length-of-stay; THA=total hip arthroplasty; TKA=total knee arthroplasty. No=negative outlier; Po=positive outlier.

For complications, between-hospital variations were case-mix adjusted for age(9,10,14-16,21,27,42,43,49,51), gender(9,10,14,15,21,27,42,43,49,51), ethnicity(21,42,43), Elixhauser comorbidities(9,42,49), patient comorbidities(14,16,27,39,43), ASA score(10,21), procedure (THA or TKA) when THA&TKA are combined(14,16,51), payer(42), admission status(42), functional status(21), history of acute myocardial infarction(21), history of peripheral vascular disease(21), depression(21), diabetes mellitus(21), surgical time(21), work relative value unit(21), emergency surgery(21), BMI(10), smoking(10), smoking status(10), diagnosis (osteoarthritis versus other) (10), preoperative Hb(10), clinical comorbidities(15) and bilateral surgery (14).

Context for benchmarking hospitals

Between-hospital variation was generated mostly as feedback for quality improvement purposes(9,10,21,30,32,33,42,45-47), but also to assess variation by structural hospital characteristics (e.g., ownership structure or teaching status)(28,31,34,48,51-53), to assess outcome associations between specialisms and hospitals(40) and to assess the impact of coding schemes(43). Regardless of the purpose of the studies, 19 studies (58% of studies) informed individual hospitals about their performance (i.e., 5 for revision(30-34), 6 for readmission(9,21,28,31,34,40) and 12 for complications(9,10,21,33,34,42,43,45-48,51)). The remaining 14 studies (42% of studies) reported the variation in one overall estimate (i.e., mean (sd), median (IQR) or range) from which hospitals are unable to infer how they are performing compared with other hospitals.(14-16,27,29,36-39,41,44,49,50,54) All registry reports gave outcomes at the individual hospital level when outcomes were collected (i.e., 8 for revision(2-8,35), 2 for readmission(4,6), 3 for complications(3-5), 1 for LOS(6) and 2 for mortality(3,5)).

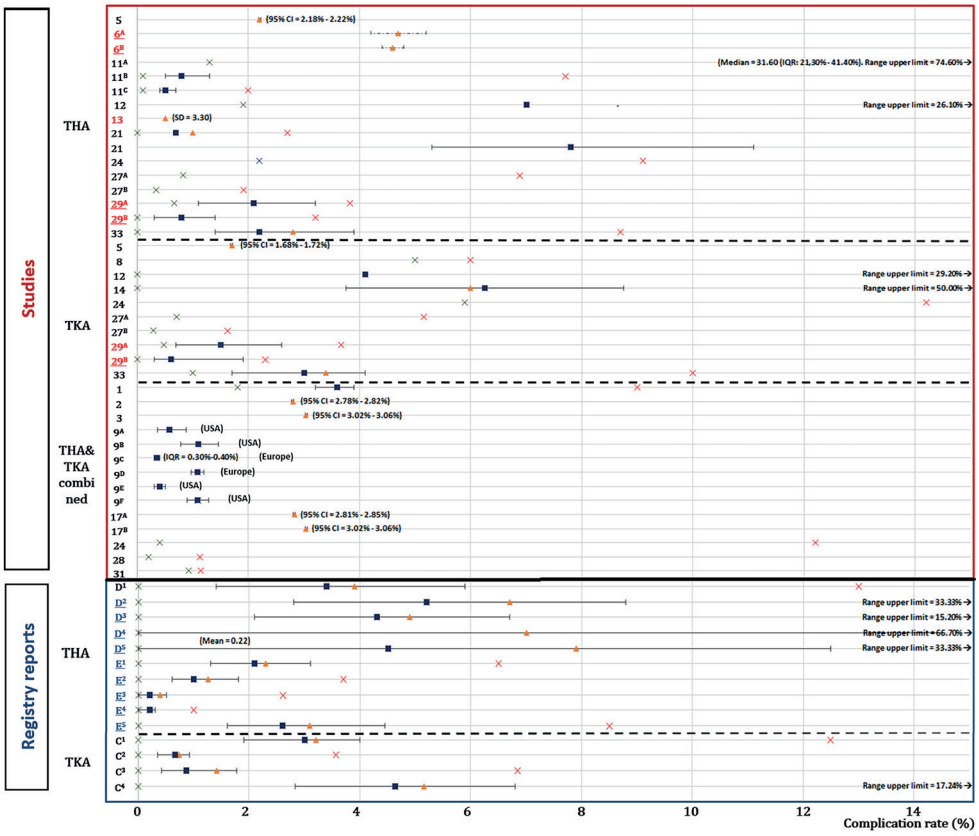


Figure 4 Between-hospital variation for complications

The numbers on the *y*-axis correspond with the study numbers from Table 1A and the letters on the *y*-axis with the report codes from Table 1B. A letter in superscript was added to a study number or a number in superscript to a report code when the complication rate was reported more than once with different definitions. Study numbers were underlined and red coloured when complications within 30 days was reported and registry reports were underlined and blue coloured when complications within 2 years was reported. The green and red cross represent the upper and lower range, respectively. The blue square represents the median and the yellow triangle the mean. The interquartile range is shown in a solid line through the median. The 95% confidence interval is shown with a dashed line through the mean. Results of outlier procedures are shown below.

THA studies: 12=No:2, Po:0, Others:21; 27^{A}=No:5, Po:4, Others:8; 27^{A}=No:4, Po:4, Others:9; 29^{A}=No:1, Po:2, Others:7; 27^{B}=No:1, Po:4, Others:5. **TKA studies:** 27^{A}=No:4, Po:2, Others:12; 27A=No:3, Po:1, Others:14; 29^{A}=No:0, Po:2, Others:7; 29^{B}=No:0, Po:3, Others:6. **THA reports:** D^{1}=No:0, Po:22, Others:16; D^{2}=No:4, Po:4, Others:31; D^{3}=No:4, Po:5, Others:29; D^{4}=No:0, Po:1, Others:19; D^{5}=No:0, Po:3, Others:33; E^{6}=No:14, Po:14, Others:48; E^{7}=No:0, Po:18, Others:58; E^{8}=No:6, Po:6, Others:57; E^{9}=No:7, Po:3, Others:47; E^{10}=No:2, Po:8, Others:46; E^{11}=No:2, Po:6, Others:48.

No=negative outlier; O/E=observed divided by expected; Po=positive outlier; THA=total hip arthroplasty; TKA=total knee arthroplasty; 95% CI=95% confidence interval.

Table 4 Definitions to report between-hospital variation for complications

Complications	Studies (*n*=20)			Registry reports (*n*=3)	
	THA (*n*=16) [5,6[A],6[B],11[A],11[B],11[C],12,13,20,21,24,27[A],27[B],29[A],29[B],33]	**TKA (*n*=10)** [5,8,12,14,24,27[A],27[B],29[A],29[B],33]	**THA&TKA (*n*=14)** [1,2,3,9[A],9[B],9[C],9[D],9[E],9[F],17[A],17[B],24,28,31]	**THA (*n*=16)** [D[1],D[2],D[3],D[4],D[5],E[1],E[2],E[3],E[4],E[5],E[6],E[7],E[8],E[9],E[10],E[11]]	**TKA (*n*=4)** [C[1],C[2],C[3],C[4]]
1) Outcome definition					
• NQF complication rate*	---	---	[1-3,17[A],17[B]]	---	---
• VASQIP complication**	[5]	[5]	---	---	---
• Study/report specific composite	[6[A],6[B],24]	[14,24]	[24]	---	---
• Early prosthetic joint infections	---	---	[9[A],9[C],9[E]]	---	---
• Late prosthetic joint infections	---	---	[9[B],9[D],9[F]]	---	---
• Blood transfusion	---	---	---	[D[1]]	---
• Blood transfusion (red blood cells)	[11[A]]	[12]	---	---	---
• Blood transfusion (fresh frozen plasma)	[11[B]]	---	---	---	---
• Blood transfusion (platelets)	[11[C]]	---	---	---	---
• Allogeneic transfusion	[12]	---	---	---	---
• DVT and/or PE	[13]	---	---	---	---
• Reoperation	[20]	---	---	[D[2]-E[1], E[5]]	[C[1]]
• Reoperation due to deep infection	---	---	---	[E[2]]	---
• Reoperation due to dislocation	---	---	---	[E[3]]	---
• Reoperation due to a fracture	---	---	---	[E[4]]	---
• Surgical site infection	[21,27[A],29[A],33]	[27[A],29[A],33]	[28,31]	---	---
• Deep Surgical site infection	[27[B],29[B]]	[27[B],29[B]]	---	---	---
• Cardiovascular events	---	---	---	---	[C[2]]
• May be related to the surgery	---	---	---	---	[C[3]]
• All adverse events, including death	---	---	---	---	[C[4]]
• Not specified	---	[8]	---	---	---

Table 4 Continued

Complications	Studies (*n*=20)				
	THA (*n*=16) [5,6[A],6[B],11[A],11[B],11[C],12,13,20,21,24,27[A],27[B],29[A],29[B],33]	**TKA (*n*=10)** [5,8,12,14,24,27[A],27[B],29[A],29[B],33]	**THA&TKA (*n*=14)** [1,2,3,9[A],9[B],9[C],9[D],9[E],9[F],17[A],17[B],24,28,31]	**THA (*n*=16)** [D[1],D[2],D[3],D[4],D[5],E[1],E[2],E[3],E[4],E[5],E[6],E[7],E[8],E[9],E[10],E[11]]	**TKA (*n*=4)** [C[1],C[2],C[3],C[4]]
2a) Follow-up					
Within hospital stay	[11[A]-11[C],12,20]	[12]	---	---	---
Within 7 days	---	---	---	[D[1]]	---
Within 14 days	[5]	[5]	---	---	---
Within 4 weeks	---	---	[9[A],9[C],9[E]]	---	---
Within 30 days	[6[A],6[B],13,29[A],29[B]]	[29[A],29[B]]	---	---	---
Within 90 days	---	[14]	---	---	[C[1]-C[4]]
Within 1 year	[21,33]	[33]	[28,31]	---	---
Within 2 years	[24]	[24]	[9[B],9[D],9[F],24]	[D[2]-E[5]]	---
Composite	---	---	[1-3,17[A],17[B]]	---	---
Not specified	[27[A],27[B]]	[8,27[A],27[B]]	---	---	---
2b) Follow-up starting point					
Post-admission	[13]	---	[1-17[B]]	---	---
Post-operative	[6[A],6[B],21,24,29[A]-33]	[14,24,29A,29B,33]	[24-31]	[D[1]-E[5]]	[C[1]-C[4]]
Post-discharge	[5]	[5]	---	---	---
Not specified	[27[A],27[B]]	[8,27[A],27[B]]	---	---	---
Not applicable	[11[A]-11[C],12,20]	[12]	---	---	---
3) Case-mix adjusted					
Yes	[5-20]	[5,12,14]	[1-3,17[A],17[B],28]	---	---
No	[21-33]	[8, 24-33]	[9[A]-9[F], 24,31]	[D[1]-E[5]]	[C[1]-C[4]]

Table 4 Continued

Complications	Studies (*n*=20)			Registry reports (*n*=3)	
	THA (*n*=16) [5,6[A],6[B],11[A],11[B],11[C],12,13,20,21,24,27[A],27[B],29[A],29[B],33]	**TKA (*n*=10)** [5,8,12,14,24,27[A],27[B],29[A],29[B],33]	**THA&TKA (*n*=14)** [1,2,3,9[A],9[B],9[C],9[D],9[E],9[F],17[A],17[B],24,28,31]	**THA (*n*=16)** [D[1],D[2],D[3],D[4],D[5],E[1],E[2],E[3],E[4],E[5],E[6],E[7],E[8],E[9],E[10],E[11]]	**TKA (*n*=4)** [C[1],C[2],C[3],C[4]]
4a) Type of patient selected					
Age selection(s)	[6[A],6[B]]	[14]	[1-29]	---	---
Osteoarthritis	---	---	---	[D[1],D[3]]	---
No trauma patients	[13]	---	[28]	---	---
Fracture	---	---	---	[D[4], E[5]]	---
Proximal femoral fracture	---	---	---	[D[5]]	---
Medicare patients	[6[A],6[B],21]	[14]	[1-29,17[A],17[B]]	---	---
Elective surgery	[12,33]	[12,33]	[1]	---	---
Matching of patient groups	---	[8]	---	[D[2]]	---
No selections, all patients included	[5,11[A]-11[C],20,24-29[B]]	[5,24-29[B]]	[9[A]-9[F],24,31]	[E[1]-E[4]]	[C[1]-C[4]]
4b) Type of hospitals selected					---
Number of procedures limit	[33]	[33]	[1,3,17[A],17[B]]	[D[1]-E[5]]	---
Academic hospitals	---	---	[9[A]-9[F]]	---	---
Academic and affiliated hospitals	[11[A]-11[C]]	---	---	---	---
Non-academic hospitals	[12]	[12]	---	---	---
Honor roll hospitals	[6[A]]	---	---	---	---
Affiliated honor roll hospitals	[6[B]]	---	---	---	---
Veteran Affairs Hospitals	[5]	[5]	---	---	---
Physician owned	---	---	17[A]	---	---
Non-physician owned	---	---	17[B]	---	---
No selections, all patients included	[13-29[B]]	[8,14-29[B]]	[2,24-31]	---	[C[1]-C[4]]

The definitions for complications were defined for 6 domains. The numbers in brackets correspond to the study number from Table 1A, and the letters in brackets correspond to the report codes from Table 1B.

*National Quality Forum (NQF)-endorsed hospital-level risk-standardized complication rate developed by the Centres for Medicare and Medicaid Services. **Veterans Affairs Surgical Quality Improvement Programme (VASQIP) nurse-identified postoperative complications.

DVT=deep venous thrombosis; No=negative outlier; PE=pulmonary embolism; Po=positive outlier; THA=total hip arthroplasty; TKA=total knee arthroplasty.

Discussion

The present study showed that between-hospital variation for revision, readmission, and complications is often reported in arthroplasty cohort studies and registry reports, with considerable differences between hospitals present for both THA and TKA. Large heterogeneity was found in definitions of variables and methods used, which likely explains at least part of the variation but obscures the ability to compare results and pool data. For revision, most studies reported revision within one year and most registry reports revision within 5 years. Most studies and reports reported on readmission within 30 days. As for complications, most studies reported complications within 30 days, with reports evaluating complications up to 2 years. The between-hospital variation was generally reported in the context of quality improvement purposes, but also the association with structural characteristics like ownership or teaching status.

Data currently available in literature and registry reports therefore do not facilitate an international comparison between hospital outcomes for THA and TKA, due to heterogeneity in definitions and methods used and it is impossible to pool data to enable, for example, earlier detection of safety issues. A well-known example where earlier detection would have prevented many patients from unnecessary suffering was the metal-on-metal hip arthroplasty disaster, in which 20% of patients had to undergo a revision within 10 years, compared with 4% in metal-on-polyethylene arthroplasties.(55,56) The Australian Orthopaedic Association National Joint Replacement Registry (AOANJRR) identified these implants as having an outlier performance in 2007, more than three years before retraction from the market.(57) In addition, the mortality risk increased by 8.5% (95%-CI: 5.8%-11.2%) due to these implants.(58) To pool data and enable international comparison of between-hospital variation, two steps must be taken.

First, worldwide agreement on definitions is needed for the outcome, follow-up (starting time), case-mix adjustment, and patients/hospitals that should be selected. An example of this on a smaller scale is, the Nordic Arthroplasty Register Association (NARA). They previously merged revision data with matching definitions to identify differences in revision rates between Sweden, Denmark, Norway, and Finland in 2014.(59) However, as shown in the present study, the definitions in their published annual reports do not match exactly when patient- and hospital selections are considered. A collaboration of arthroplasty registries such as the International Society of Arthroplasty Registries and NORE (EFORT) could play a leading role in assessing the feasibility of a unified global system to evaluate delivered care and benchmark hospital performance using the same definitions.(60) Since 2012, the International Consortium of Orthopaedic Registries (ICOR) has been working to implement a

global surveillance system for monitoring medical devices throughout their life. They already have several tools available to facilitate collaboration at different stages.(61,62) In this context, it is essential to distinguish between suitable indicators for monitoring quality of care or implant survival. Revision of an implant within 1 year, for example, gives a better reflection of the quality of care delivered as it is closer to (and therefore more likely to be related with) the surgery performed, whereas a revision within 5 years is highly relevant to monitor implant survival. Even if definitions match in the future, it will often remain difficult to compare hospitals from different healthcare systems in a fair way. For example, differences in LOS and readmissions between hospitals in different healthcare systems can be caused by the availability of outpatient clinics, hospitalization shorter than 24 hours imposed by health insurance policies, cooperation agreements with general practitioners, and other financial incentives.

Second, to allow for fair hospital comparison between hospitals, it is important to adjust for differences in case-mix.(63) Hospitals that tend to treat mainly patients without comorbidities (e.g., ASA I patient with osteoarthritis and no hip deformities) are expected to have a lower frequency of adverse events (e.g., revision, infection) than hospitals treating patients with multiple comorbidities (e.g., ASA III and congenital hip deformities).(64-66) As shown in this study, there is no consensus on whether or not to adjust for case-mix, let alone for which patient characteristics should be adjusted. Adjustments were made for 35 different patient characteristics, mainly for age and gender, followed by ethnicity, BMI, ASA score, and Elixhauser comorbidities; these patient characteristics are readily available in routinely collected data. In 3 studies and 1 report, hospital variation was adjusted for surgery- or hospital-specific determinants (e.g., hospital and surgeon volume) in addition to patient characteristics. However, these determinants could also be a proxy for experience and thereby an intermediate variable in the causal pathway to achieve good patient outcomes, that should not be adjusted for.

Consensus in data definitions and case-mix adjustment definitions enables international hospital comparison, such that (global) feedback can be given in relation to others as this has been shown effective to improve care. A previous study showed a 0.89% (95%-CI:0.83%-0.96%) reduction in serious adverse events for THA and TKA when hospitals receiving feedback were compared with control hospitals.(67) In addition, a Cochrane review showed a median absolute improvement of 4.3% associated with audit and feedback (IQR:0.5%-16.0%).(68) Studies have also shown that feedback is more effective when given monthly in an active way by a senior colleague, both verbal and written, with specific goals and actions planned rather than in a passive way (e.g., registry reports).(68-71) Feedback is often reactive and only targeted at underperforming hospitals (i.e., negative outliers), sometimes

with financial consequences. Feedback could, however, be more effective if not only underperforming hospitals feel addressed, but if normal or good performing hospitals are also actively motivated to improve further, which could be achieved by international comparisons.

Several limitations should be noted. First, completeness of data was reported for only 8 studies (24% of studies), making it impossible to assess whether selection bias affected hospital outcomes and thus generalizability of our results (Appendix IIIA&B). To allow for a correct interpretation, it is therefore essential to state the variation in completeness of data across hospitals in a study or report. Second, when indicator outcomes occurred but in another hospital, this underestimates the outcome in the first hospital and also the variation between hospitals if this happens systematically for some hospitals. However, this does not apply to registries included in this study because they use a unique personal code, linking outcomes in other hospitals to the primary procedure. Third, between-hospital variation may have been overestimated when outcomes were not adjusted for case-mix or only by a limited number of variables, resulting in possible residual confounding which is now attributed to the hospital. Finally, some studies and reports have not reported the definitions across all 4 domains so that the agreement may have been higher for some domains (Table 2, 3, and 4).

Conclusion

To benchmark hospital performance or pool data for earlier detection of safety problems across countries, it is necessary to collaborate internationally and strive for more uniformity in indicator definitions and methods used.

References

1. Norwegian Arthroplasty Register (NAR). Annual Report 2020. http://nrlweb.ihelse.net/eng/Rapporter/Report2020_english.pdf.
2. Dutch Arthroplasty Register (LROI). Online Annual Report. https://www.lroi-report.nl/.
3. Swedish Knee Arthroplasty Register (SKAR). Annual Report 2020. http://myknee.se/pdf/SVK_2020_Eng_1.0.pdf.
4. Danish Hip Arthroplasty Register (DHAR). Annual report 2020. http://danskhoftealloplastikregister.dk/wp-content/uploads/2020/11/dhr-aarsrapport-2020_til_offentliggoerelse-1.pdf.
5. Swedish Hip Arthroplasty Register (SHAR). Annual Report 2018. https://registercentrum.blob.core.windows.net/shpr/r/Arsrapport_2018_Hoftprotes_ENG_26mars_Final-rJepCXNsLI.pdf.
6. Danish Knee Arthroplasty Register (DKAR). Annual Report 2020. https://www.sundhed.dk/content/cms/99/4699_dkr-arsrapport-2020_offentliggorelse.pdf.
7. Finnish Arthroplasty Register (FAR). Online Annual Report. https://www.thl.fi/far/#index.
8. Swiss Arthroplasty Register (SAR). Annual Report 2020. https://www.siris-implant.ch/de/Downloads&category=16.
9. Sheetz KH, Ibrahim AM, Nathan H, Dimick JB. Variation in Surgical Outcomes Across Networks of the Highest-Rated US Hospitals. JAMA Surg. 2019;154(6):510-5.
10. Voorn VMA, Marang-van de Mheen PJ, van der Hout A, So-Osman C, van den Akker-van Marle ME, Koopman-van Gemert A, et al. Hospital variation in allogeneic transfusion and extended length of stay in primary elective hip and knee arthroplasty: a cross-sectional study. BMJ Open. 2017;7(7):e014143.
11. van Schie P, van Steenbergen LN, van Bodegom-Vos L, Nelissen R, Marang-van de Mheen PJ. Between-Hospital Variation in Revision Rates After Total Hip and Knee Arthroplasty in the Netherlands: Directing Quality-Improvement Initiatives. J Bone Joint Surg Am. 2020;102(4):315-24.
12. van Schie P, van Bodegom-Vos L, van Steenbergen LN, Nelissen R, Marang-van de Mheen PJ. Monitoring Hospital Performance with Statistical Process Control After Total Hip and Knee Arthroplasty: A Study to Determine How Much Earlier Worsening Performance Can Be Detected. J Bone Joint Surg Am. 2020;102(23):2087-94.
13. Ranstam J, Wagner P, Robertsson O, Lidgren L. Health-care quality registers: outcome-orientated ranking of hospitals is unreliable. J Bone Joint Surg Br. 2008;90(12):1558-61.
14. Bozic KJ, Grosso LM, Lin Z, Parzynski CS, Suter LG, Krumholz HM, et al. Variation in hospital-level risk-standardized complication rates following elective primary total hip and knee arthroplasty. J Bone Joint Surg Am. 2014;96(8):640-7.
15. Thirukumaran CP, McGarry BE, Glance LG, Ying M, Ricciardi BF, Cai X, et al. Impact of Hospital Readmissions Reduction Program Penalties on Hip and Knee Replacement Readmissions: Comparison of Hospitals at Risk of Varying Penalty Amounts. J Bone Joint Surg Am. 2020;102(1):60-7.
16. Courtney M, Darrith B, Bohl DD, Frisch NB, Valle CJD. Reconsidering the affordable care act's restrictions on physician-owned hospitals analysis of CMS data on total hip and knee arthroplasty. Journal of Bone and Joint Surgery - American Volume. 2017;99(22):1888-94.
17. Van Schie P, Van Bodegom-Vos L, Zijdeman TM, Nelissen R, Marang-Van De Mheen PJ. Awareness of performance on outcomes after total hip and knee arthroplasty among Dutch orthopedic surgeons: how to improve feedback from arthroplasty registries. Acta Orthop. 2020:1-8.

18. Barbazza E, Klazinga NS, Kringos DS. Exploring the actionability of healthcare performance indicators for quality of care: a qualitative analysis of the literature, expert opinion and user experience. BMJ quality & safety. 2021.
19. Fischer C, Lingsma HF, Marang-van de Mheen PJ, Kringos DS, Klazinga NS, Steyerberg EW. Is the readmission rate a valid quality indicator? A review of the evidence. PLoS One. 2014;9(11):e112282.
20. Veen EJ, Janssen-Heijnen ML, Leenen LP, Roukema JA. The registration of complications in surgery: a learning curve. World J Surg. 2005;29(3):402-9.
21. Graham LA, Mull HJ, Wagner TH, Morris MS, Rosen AK, Richman JS, et al. Comparison of a Potential Hospital Quality Metric With Existing Metrics for Surgical Quality-Associated Readmission. JAMA Netw Open. 2019;2(4):e191313.
22. Page MJ, McKenzie J, Bossuyt P, Boutron I, Hoffman T, Mulrow C, et al. The PRISMA 2020 statement: an upgrated guideline for reporting systematic reviews. MetaArXiv. 2020.
23. Network of Orthopaedic Registries of Europe (NORE). https://www.efort.org/about-us/nore/research/ (visited on December 31, 2020).
24. Weir CJ, Butcher I, Assi V, Lewis SC, Murray GD, Langhorne P, et al. Dealing with missing standard deviation and mean values in meta-analysis of continuous outcomes: a systematic review. BMC Med Res Methodol. 2018;18(1):25.
25. Cochrane Handbook for Systematic Reviews. Version 6.2, 2021. https://training.cochrane.org/handbook/current.
26. Zingg W, Castro-Sanchez E, Secci FV, Edwards R, Drumright LN, Sevdalis N, et al. Innovative tools for quality assessment: integrated quality criteria for review of multiple study designs (ICROMS). Public Health. 2016;133:19-37.
27. Pross C, Busse R, Geissler A. Hospital quality variation matters - A time-trend and cross-section analysis of outcomes in German hospitals from 2006 to 2014. Health Policy. 2017;121(8):842-52.
28. Mittal M, Wang CE, Goben AH, Boyd AD. Proprietary management and higher readmission rates: A correlation. PLoS One. 2018;13(9):e0204272.
29. Husted H, Hansen HC, Holm G, Bach-Dal C, Rud K, Andersen KL, et al. What determines length of stay after total hip and knee arthroplasty? A nationwide study in Denmark. Arch Orthop Trauma Surg. 2010;130(2):263-8.
30. van Schie P, van Steenbergen LN, van Bodegom-Vos L, Nelissen R, Marang-van de Mheen PJ. Between-Hospital Variation in Revision Rates After Total Hip and Knee Arthroplasty in the Netherlands: Directing Quality-Improvement Initiatives. J Bone Joint Surg Am. 2019.
31. Padegimas EM, Kreitz TM, Zmistowski BM, Girden AJ, Hozack WJ, Chen AF. Comparison of Short-Term Outcomes After Total Hip Arthroplasty Between an Orthopedic Specialty Hospital and General Hospital. J Arthroplasty. 2017;32(8):2347-52.
32. Makela KT, Peltola M, Sund R, Malmivaara A, Hakkinen U, Remes V. Regional and hospital variance in performance of total hip and knee replacements: a national population-based study. Ann Med. 2011;43 Suppl 1:S31-S8.
33. Singh S, Reddy S, Shrivastava R. Does laminar airflow make a difference to the infection rates for lower limb arthroplasty: a study using the National Joint Registry and local surgical site infection data for two hospitals with and without laminar airflow. Eur J Orthop Surg Traumatol. 2017;27(2):261-5.
34. Padegimas EM, Kreitz TM, Zmistowski B, Teplitsky SL, Namdari S, Purtill JJ, et al. Short-term Outcomes of Total Knee Arthroplasty Performed at an Orthopedic Specialty Hospital. Orthopedics. 2018;41(1):e84-e91.

35. The Norwegian National Arthroplasty Register. Annual Report 2020. http://nrlweb.ihelse.net/eng/Rapporter/Report2020_english.pdf.
36. Bottle A, Loeffler MD, Aylin P, Ali AM. Comparison of 3 Types of Readmission Rates for Measuring Hospital and Surgeon Performance After Primary Total Hip and Knee Arthroplasty. J Arthroplasty. 2018;33(7):2014-9.
37. Kurtz SM, Lau EC, Ong KL, Adler EM, Kolisek FR, Manley MT. Hospital, Patient, and Clinical Factors Influence 30- and 90-Day Readmission After Primary Total Hip Arthroplasty. J Arthroplasty. 2016;31(10):2130-8.
38. Kurtz SM, Lau EC, Ong KL, Adler EM, Kolisek FR, Manley MT. Which Hospital and Clinical Factors Drive 30- and 90-Day Readmission After TKA? J Arthroplasty. 2016;31(10):2099-107.
39. Courtney PM, Frisch NB, Bohl DD, Della Valle CJ. Improving Value in Total Hip and Knee Arthroplasty: The Role of High Volume Hospitals. J Arthroplasty. 2018;33(1):1-5.
40. Hollis RH, Graham LA, Richman JS, Morris MS, Mull HJ, Wahl TS, et al. Hospital Readmissions after Surgery: How Important Are Hospital and Specialty Factors? J Am Coll Surg. 2017;224(4):515-23.
41. Hofstede SN, Ceyisakar IE, Lingsma HF, Kringos DS, Marang-van de Mheen PJ. Ranking hospitals: do we gain reliability by using composite rather than individual indicators? BMJ quality & safety. 2018.
42. Qian F, Osler TM, Eaton MP, Dick AW, Hohmann SF, Lustik SJ, et al. Variation of blood transfusion in patients undergoing major noncardiac surgery. Ann Surg. 2013;257(2):266-78.
43. Cram P, Ibrahim SA, Lu X, Wolf BR. Impact of alternative coding schemes on incidence rates of key complications after total hip arthroplasty: a risk-adjusted analysis of a national data set. Geriatr Orthop Surg Rehabil. 2012;3(1):17-26.
44. Calderwood MS, Kleinman K, Bratzler DW, Ma A, Bruce CB, Kaganov RE, et al. Use of Medicare claims to identify US hospitals with a high rate of surgical site infection after hip arthroplasty. Infect Control Hosp Epidemiol. 2013;34(1):31-9.
45. Skufca J, Ollgren J, Virtanen MJ, Huotari K, Lyytikainen O. Interhospital Comparison of Surgical Site Infection Rates in Orthopedic Surgery. Infect Control Hosp Epidemiol. 2017;38(4):423-9.
46. Dailey L, Van GH, Peterson A. Two years of surgical site infection surveillance in Western Australia: Analysing variation between hospitals. Healthcare Infection. 2009;14(2):51-60.
47. Lopez-Contreras J, Limon E, Matas L, Olona M, Salles M, Pujol M. Epidemiology of surgical site infections after total hip and knee joint replacement during 2007-2009: a report from the VINCat Program. Enferm Infecc Microbiol Clin. 2012;30 Suppl 3:26-32.
48. Jergesen HE, Yi PH. Early Complications in Hip and Knee Arthroplasties in a Safety Net Hospital vs a University Center. J Arthroplasty. 2016;31(4):754-8.
49. Cai X, Cram P, Vaughan-Sarrazin M. Are African American patients more likely to receive a total knee arthroplasty in a low-quality hospital? Clin Orthop Relat Res. 2012;470(4):1185-93.
50. Marang-van de Mheen PJ, Bragan Turner E, Liew S, Mutalima N, Tran T, Rasmussen S, et al. Variation in Prosthetic Joint Infection and treatment strategies during 4.5 years of follow-up after primary joint arthroplasty using administrative data of 41397 patients across Australian, European and United States hospitals. BMC Musculoskelet Disord. 2017;18(1):207.
51. Asaid R, Williams I, Hyde D, Tiang T. Infection rates following hip and knee joint arthroplasty: large referral centre versus a small elective-only hospital. Eur J Orthop Surg Traumatol. 2013;23(2):165-8.
52. Chen AF, Pflug E, O'Brien D, Maltenfort MG, Parvizi J. Utilization of Total Joint Arthroplasty in Physician-Owned Specialty Hospitals vs Acute Care Facilities. J Arthroplasty. 2017;32(7):2060-4.

53. Martino J, Peterson B, Thompson S, Cook JL, Aggarwal A. Day of Week and Surgery Location Effects on Stay Length and Cost for Total Joint Arthroplasty: Academic versus Orthopaedic-Specific Hospital. J Knee Surg. 2018;31(9):815-21.
54. Husni ME, Losina E, Fossel AH, Solomon DH, Mahomed NN, Katz JN. Decreasing medical complications for total knee arthroplasty: effect of critical pathways on outcomes. BMC Musculoskelet Disord. 2010;11:160.
55. Drummond J, Tran P, Fary C. Metal-on-Metal Hip Arthroplasty: A Review of Adverse Reactions and Patient Management. Journal of functional biomaterials. 2015;6(3):486-99.
56. Pfuntner A, Wier LM, Stocks C. Most Frequent Procedures Performed in U.S. Hospitals, 2011: Statistical Brief #165. Healthcare Cost and Utilization Project (HCUP) Statistical Briefs. Rockville (MD): Agency for Healthcare Research and Quality (US); 2006.
57. de Steiger RN, Hang JR, Miller LN, Graves SE, Davidson DC. Five-year results of the ASR XL Acetabular System and the ASR Hip Resurfacing System: an analysis from the Australian Orthopaedic Association National Joint Replacement Registry. J Bone Joint Surg Am. 2011;93(24):2287-93.
58. Pijls BG, Meessen JM, Schoones JW, Fiocco M, van der Heide HJ, Sedrakyan A, et al. Increased Mortality in Metal-on-Metal versus Non-Metal-on-Metal Primary Total Hip Arthroplasty at 10 Years and Longer Follow-Up: A Systematic Review and Meta-Analysis. PLoS One. 2016;11(6):e0156051.
59. Mäkelä KT, Matilainen M, Pulkkinen P, Fenstad AM, Havelin LI, Engesaeter L, et al. Countrywise results of total hip replacement. An analysis of 438,733 hips based on the Nordic Arthroplasty Register Association database. Acta Orthop. 2014;85(2):107-16.
60. International Society of Arthroplasty Registries (ISAR). https://www.isarhome.org/home.
61. Sedrakyan A, Paxton E, Graves S, Love R, Marinac-Dabic D. National and international postmarket research and surveillance implementation: achievements of the International Consortium of Orthopaedic Registries initiative. J Bone Joint Surg Am. 2014;96 Suppl 1:1-6.
62. Sedrakyan A, Paxton EW, Marinac-Dabic D. Stages and tools for multinational collaboration: the perspective from the coordinating center of the International Consortium of Orthopaedic Registries (ICOR). J Bone Joint Surg Am. 2011;93 Suppl 3:76-80.
63. Signorini DF, Weir NU. Any variability in outcome comparisons adjusted for case mix must be accounted for. BMJ. 1999;318(7176):128.
64. Lenguerrand E, Whitehouse MR, Beswick AD, Kunutsor SK, Burston B, Porter M, et al. Risk factors associated with revision for prosthetic joint infection after hip replacement: a prospective observational cohort study. Lancet Infect Dis. 2018;18(9):1004-14.
65. Prokopetz JJ, Losina E, Bliss RL, Wright J, Baron JA, Katz JN. Risk factors for revision of primary total hip arthroplasty: a systematic review. BMC Musculoskelet Disord. 2012;13:251.
66. Jasper LL, Jones, C. A., Mollins, J., Pohar, S. L., Beaupre, L. A.. Risk factors for revision of total knee arthroplasty: a scoping review. BMC Musculoskelet Disord. 2016;17:182.
67. Duclos A, Chollet F, Pascal L, Ormando H, Carty MJ, Polazzi S, et al. Effect of monitoring surgical outcomes using control charts to reduce major adverse events in patients: cluster randomised trial. BMJ. 2020;371:m3840.
68. Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, French SD, et al. Audit and feedback: effects on professional practice and healthcare outcomes. Cochrane Database Syst Rev. 2012(6):Cd000259.
69. Soong C, Shojania KG. Education as a low-value improvement intervention: often necessary but rarely sufficient. BMJ quality & safety. 2020;29(5):353-7.

70. Gude WT, Roos-Blom MJ, van der Veer SN, de Jonge E, Peek N, Dongelmans DA, et al. Electronic audit and feedback intervention with action implementation toolbox to improve pain management in intensive care: protocol for a laboratory experiment and cluster randomised trial. Implement Sci. 2017;12(1):68.
71. Roos-Blom MJ, Gude WT, de Jonge E, Spijkstra JJ, van der Veer SN, Peek N, et al. Impact of audit and feedback with action implementation toolbox on improving ICU pain management: cluster-randomised controlled trial. BMJ quality & safety. 2019;28(12):1007-15.

Supplemental data

Appendix I Literature search strategy

1. PubMed (http://www.ncbi.nlm.nih.gov/pubmed?otool=leiden)

((("tha"[ti] OR "total hip"[ti] OR "Arthroplasty, Replacement, Hip"[majr] OR "Hip Prosthesis"[majr] OR "Hip Prosthesis"[ti] OR "Hip Replacement"[ti] OR "Hip Arthroplasty"[ti] OR "Prosthetic Hip"[ti] OR "Prosthetic Hip"[ti] **OR** "tka"[ti] OR "total knee"[ti] OR "Arthroplasty, Replacement, Knee"[majr] OR "Knee Prosthesis"[majr] OR "Knee Prosthesis"[ti] OR "Knee Replacement"[ti] OR "Knee Arthroplasty"[ti] OR "Prosthetic Knee"[ti] OR "Prosthetic Knees"[ti] OR (("Arthroplasty, Replacement"[majr] OR "Joint Prosthesis"[majr] OR "Joint Arthroplasty"[ti] OR "Joint Replacement"[ti] OR "Joint Prosthesis"[ti] OR "Prosthetic Joint"[ti] OR "Prosthetic Joints"[ti]) AND ("Hip"[majr] OR "Hip Joint"[majr] OR "hip"[ti] OR "hips"[ti] OR "Knee"[majr] OR "Knee Joint"[majr] OR "knee"[ti] OR "knees"[ti]))) AND ("revision rate"[ti] OR "revision rates"[ti] OR (("revision"[ti] OR "revisions"[ti]) AND ("rate"[ti] OR "rates"[ti] OR "median"[ti] OR "mean"[ti] OR "percentage"[ti] OR "percent"[ti] OR percent*[ti])) OR "revision"[ti] OR "Repeat Surgery"[ti] OR "Joint Revision"[ti] OR "Reoperation"[majr] OR "Reoperation"[ti] OR Reoperat*[ti] OR "Re-operation"[ti] OR Re-operat*[ti] OR "Length of Stay"[majr] OR "length of stay"[ti] OR "lengths of stay"[ti] OR "length of stays"[ti] OR "lengths of stays"[ti] OR "stay length"[ti] OR "stay lengths"[ti] OR "Patient Readmission"[majr] OR "patient readmission"[ti] OR "hospital readmission"[ti] OR "patient re-admission"[ti] OR "hospital re-admission"[ti] OR "readmission"[ti] OR "re-admission"[ti] OR readmit*[ti] OR re-admit*[ti] OR "Mortality"[majr] OR "mortality"[Subheading] OR "mortality"[ti] OR mortalit*[ti] OR "death"[ti] OR "deaths"[ti] OR "Cause of Death"[ti] OR "fatality rate"[ti] OR "fatality rates"[ti] OR "Fatal Outcome"[ti] OR "Fatal Outcomes"[ti] OR "Survival Rate"[majr] OR "complications"[Subheading:noexp] OR "Postoperative Complications"[majr:noexp] OR "Prosthesis Failure"[majr] OR "Prosthesis-Related Infections"[majr] OR "Surgical Wound Infection"[majr] OR "Prosthesis Failure"[ti] OR "Prosthesis-Related Infections"[ti] OR "Prosthesis-Related Infection"[ti] OR "Prosthesis Infections"[ti] OR "Prosthesis Infection"[ti] OR "Surgical Wound Infection"[ti] OR "Surgical Infection"[ti] OR "Surgical Wound Infections"[ti] OR "Surgical Infections"[ti] OR "surgical injury"[ti] OR "surgical injuries"[ti] OR "complication"[ti] OR "complications"[ti] OR "Joint Dislocations"[majr] OR "Dislocations"[ti] OR "Dislocation"[ti] OR dislocat*[ti] OR Subluxat*[ti] OR "Prosthesis-Related Infections"[majr] OR "Prosthesis-Related Infection"[ti] OR "Prosthesis-Related Infections"[ti] OR "Infections"[majr:noexp] OR "infection"[ti]

OR "infections"[ti] OR "infected"[ti] OR "loosening"[ti] OR "malalignment"[ti] OR "malalignments"[ti] OR "malaligned"[ti] OR "Joint Instability"[majr] OR "Instability"[ti] OR "Instabilities"[ti] OR "Patellar Dislocation"[majr] OR "Patellar Dislocation"[ti] OR "Patella Dislocation"[ti] OR "Patellar Dislocations"[ti] OR "Patella Dislocations"[ti]) AND ((("variation"[tw] OR "variations"[tw] OR "difference"[tw] OR "differences"[tw]) AND ("Hospitals"[mesh] OR "hospital"[tw] OR "hospitals"[tw])) OR "hospital characteristics"[tw] OR "hospital outcome"[tw] OR "hospital outcomes"[tw] OR "Hospitals/statistics and numerical data"[mesh] OR "Practice Patterns, Physicians'/statistics and numerical data"[mesh] OR "international variation"[tw] OR "international variations"[tw] OR "ranking hospitals"[tw] OR "hospital rank"[tw] OR "hospital ranking"[tw] OR "hospital rankings"[tw] OR "hospital ranks"[tw])) **OR** (("tha"[tiab] OR "total hip"[tiab] OR "Arthroplasty, Replacement, Hip"[mesh] OR "Hip Prosthesis"[mesh] OR "Hip Prosthesis"[tiab] OR "Hip Replacement"[tiab] OR "Hip Arthroplasty"[tiab] OR "Prosthetic Hip"[tiab] OR "Prosthetic Hip"[tiab] **OR** "tka"[tiab] OR "total knee"[tiab] OR "Arthroplasty, Replacement, Knee"[mesh] OR "Knee Prosthesis"[mesh] OR "Knee Prosthesis"[tiab] OR "Knee Replacement"[tiab] OR "Knee Arthroplasty"[tiab] OR "Prosthetic Knee"[tiab] OR "Prosthetic Knees"[tiab] OR (("Arthroplasty, Replacement"[mesh] OR "Joint Prosthesis"[mesh] OR "Joint Arthroplasty"[tiab] OR "Joint Replacement"[tiab] OR "Joint Prosthesis"[tiab] OR "Prosthetic Joint"[tiab] OR "Prosthetic Joints"[tiab]) AND ("Hip"[mesh] OR "Hip Joint"[mesh] OR "hip"[tiab] OR "hips"[tiab] OR "Knee"[mesh] OR "Knee Joint"[mesh] OR "knee"[tiab] OR "knees"[tiab]))) AND ("revision rate"[tw] OR "revision rates"[tw] OR (("revision"[tw] OR "revisions"[tw]) AND ("rate"[tw] OR "rates"[tw] OR "median"[tw] OR "mean"[tw] OR "percentage"[tw] OR "percent"[tw] OR percent*[tw])) OR "revision"[tw] OR "Repeat Surgery"[tw] OR "Joint Revision"[tw] OR "Reoperation"[mesh] OR "Reoperation"[tw] OR Reoperat*[tw] OR "Re-operation"[tw] OR Re-operat*[tw] OR "Length of Stay"[mesh] OR "length of stay"[tw] OR "lengths of stay"[tw] OR "length of stays"[tw] OR "lengths of stays"[tw] OR "stay length"[tw] OR "stay lengths"[tw] OR "Patient Readmission"[mesh] OR "patient readmission"[tw] OR "hospital readmission"[tw] OR "patient re-admission"[tw] OR "hospital re-admission"[tw] OR "readmission"[tw] OR "re-admission"[tw] OR readmit*[tw] OR re-admit*[tw] OR "Mortality"[mesh] OR "mortality"[Subheading] OR "mortality"[tw] OR mortalit*[tw] OR "death"[tw] OR "deaths"[tw] OR "Cause of Death"[tw] OR "fatality rate"[tw] OR "fatality rates"[tw] OR "Fatal Outcome"[tw] OR "Fatal Outcomes"[tw] OR "Survival Rate"[mesh] OR "complications"[Subheading:noexp] OR "Postoperative Complications"[mesh:noexp] OR "Prosthesis Failure"[mesh] OR "Prosthesis-Related Infections"[mesh] OR "Surgical Wound Infection"[mesh] OR "Prosthesis Failure"[tw] OR "Prosthesis-Related Infections"[tw] OR "Prosthesis-Related Infection"[tw] OR "Prosthesis Infections"[tw] OR "Prosthesis Infection"[tw]

OR "Surgical Wound Infection"[tw] OR "Surgical Infection"[tw] OR "Surgical Wound Infections"[tw] OR "Surgical Infections"[tw] OR "surgical injury"[tw] OR "surgical injuries"[tw] OR "complication"[tw] OR "complications"[tw] OR "Joint Dislocations"[mesh] OR "Dislocations"[tw] OR "Dislocation"[tw] OR dislocat*[tw] OR Subluxat*[tw] OR "Prosthesis-Related Infections"[mesh] OR "Prosthesis-Related Infection"[tw] OR "Prosthesis-Related Infections"[tw] OR "Infections"[mesh:noexp] OR "infection"[tw] OR "infections"[tw] OR "infected"[tw] OR "loosening"[tw] OR "malalignment"[tw] OR "malalignments"[tw] OR "malaligned"[tw] OR "Joint Instability"[mesh] OR "Instability"[tw] OR "Instabilities"[tw] OR "Patellar Dislocation"[mesh] OR "Patellar Dislocation"[tw] OR "Patella Dislocation"[tw] OR "Patellar Dislocations"[tw] OR "Patella Dislocations"[tw]) AND ((("variation"[ti] OR "variations"[ti] OR "difference"[ti] OR "differences"[ti]) AND ("Hospitals"[majr] OR "hospital"[ti] OR "hospitals"[ti])) OR "hospital characteristics"[ti] OR "hospital outcome"[ti] OR "hospital outcomes"[ti] OR "Hospitals/statistics and numerical data"[majr] OR "Practice Patterns, Physicians'/statistics and numerical data"[majr] OR "international variation"[ti] OR "international variations"[ti] OR "ranking hospitals"[ti] OR "hospital rank"[ti] OR "hospital ranking"[ti] OR "hospital rankings"[ti] OR "hospital ranks"[ti]))) AND ("2009/01/01"[PDAT] : "3000/12/31"[PDAT])

2. Embase (http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=main&MODE=ovid&D=oemezd)

((("tha".ti OR "total hip".ti OR exp *"Hip Replacement"/ OR exp *"hip arthroplasty"/ OR exp *"Hip Prosthesis"/ OR "Hip Prosthesis".ti OR "Hip Replacement".ti OR "Hip Arthroplasty".ti OR "Prosthetic Hip".ti OR "Prosthetic Hip".ti **OR** "tka".ti OR "total knee".ti OR exp *"Knee Replacement"/ OR exp *"knee arthroplasty"/ OR exp *"Knee Prosthesis"/ OR "Knee Prosthesis".ti OR "Knee Replacement".ti OR "Knee Arthroplasty".ti OR "Prosthetic Knee".ti OR "Prosthetic Knees".ti OR ((*"Arthroplasty"/ OR *"Joint Prosthesis"/ OR "Joint Arthroplasty".ti OR "Joint Replacement".ti OR "Joint Prosthesis".ti OR "Prosthetic Joint".ti OR "Prosthetic Joints".ti) AND (exp *"Hip"/ OR "hip".ti OR "hips".ti OR exp *"Knee"/ OR "knee".ti OR "knees".ti))) AND ("revision rate".ti OR "revision rates".ti OR (("revision".ti OR "revisions".ti) AND ("rate".ti OR "rates".ti OR "median".ti OR "mean".ti OR "percentage".ti OR "percent".ti OR percent*.ti)) OR "revision".ti OR "Repeat Surgery".ti OR "Joint Revision".ti OR *"Reoperation"/ OR "Reoperation".ti OR Reoperat*.ti OR "Re-operation".ti OR "Re-operat*".ti OR *"Length of Stay"/ OR "length of stay".ti OR "lengths of stay".ti OR "length of stays".ti OR "lengths of stays".ti OR "stay length".ti OR "stay lengths".ti OR *"Hospital Readmission"/ OR "patient readmission".ti OR "hospital readmission".ti OR "patient re-admission".ti OR "hospital re-admission".ti OR "readmission".ti OR "re-admission".ti OR

readmit*.ti OR "re-admit*".ti OR exp *"Mortality"/ OR "mortality".ti OR mortalit*.ti OR "death".ti OR "deaths".ti OR "Cause of Death".ti OR "fatality rate".ti OR "fatality rates".ti OR "Fatal Outcome".ti OR "Fatal Outcomes".ti OR *"Survival Rate"/ OR *"complication"/ OR "Postoperative Complication"/ OR exp *"Prosthesis Complication"/ OR "Surgical Infection"/ OR *"infectious complication"/ OR "Prosthesis Failure".ti OR "Prosthesis-Related Infections".ti OR "Prosthesis-Related Infection".ti OR "Prosthesis Infections".ti OR "Prosthesis Infection".ti OR "Surgical Wound Infection".ti OR "Surgical Infection".ti OR "Surgical Wound Infections".ti OR "Surgical Infections".ti OR *"Infection"/ OR "infection".ti OR "infections".ti OR "infected".ti OR "surgical injury".ti OR "surgical injuries".ti OR "complication".ti OR "complications".ti OR *"Joint Dislocation"/ OR "Dislocations".ti OR "Dislocation".ti OR dislocat*.ti OR *"subluxation"/ OR Subluxat*.ti OR *"prosthesis loosening"/ OR "loosening".ti OR "malalignment".ti OR "malalignments".ti OR "malaligned".ti OR exp *"Joint Instability"/ OR "Instability".ti OR "Instabilities".ti OR *"Patella Dislocation"/ OR "Patellar Dislocation".ti OR "Patella Dislocation".ti OR "Patellar Dislocations".ti OR "Patella Dislocations".ti) AND ((("variation".mp OR "variations".mp OR "difference".mp OR "differences".mp) ADJ5 ("hospital".mp OR "hospitals".mp)) OR (("variation".mp OR "variations".mp OR "difference".mp OR "differences".mp) AND exp "Hospital"/) OR "hospital characteristics".mp OR "hospital outcome".mp OR "hospital outcomes".mp OR "international variation".mp OR "international variations".mp OR "ranking hospitals".mp OR "hospital rank".mp OR "hospital ranking".mp OR "hospital rankings".mp OR "hospital ranks".mp)) **OR** (("tha".ti,ab OR "total hip".ti,ab OR exp "Hip Replacement"/ OR exp "hip arthroplasty"/ OR exp "Hip Prosthesis"/ OR "Hip Prosthesis".ti,ab OR "Hip Replacement".ti,ab OR "Hip Arthroplasty".ti,ab OR "Prosthetic Hip".ti,ab OR "Prosthetic Hip".ti,ab **OR** "tka".ti,ab OR "total knee".ti,ab OR exp "Knee Replacement"/ OR exp "knee arthroplasty"/ OR exp "Knee Prosthesis"/ OR "Knee Prosthesis".ti,ab OR "Knee Replacement".ti,ab OR "Knee Arthroplasty".ti,ab OR "Prosthetic Knee".ti,ab OR "Prosthetic Knees".ti,ab OR ((*"Arthroplasty"/ OR *"Joint Prosthesis"/ OR "Joint Arthroplasty".ti OR "Joint Replacement".ti OR "Joint Prosthesis".ti OR "Prosthetic Joint".ti OR "Prosthetic Joints".ti) AND (exp *"Hip"/ OR "hip".ti OR "hips".ti OR exp *"Knee"/ OR "knee".ti OR "knees".ti))) AND ("revision rate".mp OR "revision rates".mp OR (("revision".mp OR "revisions".mp) ADJ5 ("rate".mp OR "rates".mp OR "median".mp OR "mean".mp OR "percentage".mp OR "percent".mp OR percent*.mp)) OR "revision".mp OR "Repeat Surgery".mp OR "Joint Revision".mp OR "Reoperation"/ OR "Reoperation".mp OR Reoperat*.mp OR "Re-operation".mp OR "Re-operat*".mp OR "Length of Stay"/ OR "length of stay".mp OR "lengths of stay".mp OR "length of stays".mp OR "lengths of stays".mp OR "stay length".mp OR "stay lengths".mp OR "Hospital Readmission"/ OR "patient readmission".mp OR "hospital readmission".mp OR "patient re-admission".

mp OR "hospital re-admission".mp OR "readmission".mp OR "re-admission".mp OR readmit*.mp OR "re-admit*".mp OR exp "Mortality"/ OR "mortality".mp OR mortalit*.mp OR "death".mp OR "deaths".mp OR "Cause of Death".mp OR "fatality rate".mp OR "fatality rates".mp OR "Fatal Outcome".mp OR "Fatal Outcomes".mp OR "Survival Rate"/ OR *"complication"/ OR "Postoperative Complication"/ OR exp "Prosthesis Complication"/ OR "Surgical Infection"/ OR "infectious complication"/ OR "Prosthesis Failure".mp OR "Prosthesis-Related Infections".mp OR "Prosthesis-Related Infection".mp OR "Prosthesis Infections".mp OR "Prosthesis Infection".mp OR "Surgical Wound Infection".mp OR "Surgical Infection".mp OR "Surgical Wound Infections".mp OR "Surgical Infections".mp OR "Infection"/ OR "infection".mp OR "infections".mp OR "infected".mp OR "surgical injury".mp OR "surgical injuries". mp OR "complication".mp OR "complications".mp OR "Joint Dislocation"/ OR "Dislocations".mp OR "Dislocation".mp OR dislocat*.mp OR "subluxation"/ OR Subluxat*.mp OR "prosthesis loosening"/ OR "loosening".mp OR "malalignment". mp OR "malalignments".mp OR "malaligned".mp OR exp "Joint Instability"/ OR "Instability".mp OR "Instabilities".mp OR "Patella Dislocation"/ OR "Patellar Dislocation".mp OR "Patella Dislocation".mp OR "Patellar Dislocations".mp OR "Patella Dislocations".mp) AND ((("variation".ti OR "variations".ti OR "difference". ti OR "differences".ti) AND (exp *"Hospital"/ OR "hospital".ti OR "hospitals".ti)) OR "hospital characteristics".ti OR "hospital outcome".ti OR "hospital outcomes". ti OR "international variation".ti OR "international variations".ti OR "ranking hospitals".ti OR "hospital rank".ti OR "hospital ranking".ti OR "hospital rankings". ti OR "hospital ranks".ti)) AND (2009 OR 201* OR 202*).yr

3. Web of Science (http://isiknowledge.com/wos)

((ti=("tha" OR "total hip" OR "Hip Replacement" OR "hip arthroplasty" OR "Hip Prosthesis" OR "Hip Prosthesis" OR "Hip Replacement" OR "Hip Arthroplasty" OR "Prosthetic Hip" OR "Prosthetic Hip" **OR** "tka" OR "total knee" OR "Knee Replacement" OR "knee arthroplasty" OR "Knee Prosthesis" OR "Knee Prosthesis" OR "Knee Replacement" OR "Knee Arthroplasty" OR "Prosthetic Knee" OR "Prosthetic Knees" OR (("Arthroplasty" OR "Joint Prosthesis" OR "Joint Arthroplasty" OR "Joint Replacement" OR "Joint Prosthesis" OR "Prosthetic Joint" OR "Prosthetic Joints") AND ("Hip" OR "hip" OR "hips" OR "Knee" OR "knee" OR "knees"))) AND ti=("revision rate" OR "revision rates" OR (("revision" OR "revisions") AND ("rate" OR "rates" OR "median" OR "mean" OR "percentage" OR "percent" OR percent*)) OR "revision" OR "Repeat Surgery" OR "Joint Revision" OR "Reoperation" OR "Reoperation" OR Reoperat* OR "Re-operation" OR "Re-operat" OR "Length of Stay" OR "length of stay" OR "lengths of stay" OR "length of stays" OR "lengths of stays" OR "stay length" OR "stay lengths" OR "Hospital Readmission" OR "patient readmission" OR "hospital readmission" OR "patient re-admission" OR "hospital

re-admission" OR "readmission" OR "re-admission" OR readmit* OR "re-admit" OR "Mortality" OR "mortality" OR mortalit* OR "death" OR "deaths" OR "Cause of Death" OR "fatality rate" OR "fatality rates" OR "Fatal Outcome" OR "Fatal Outcomes" OR "Survival Rate" OR "complication" OR "Postoperative Complication" OR "Prosthesis Complication" OR "Surgical Infection" OR "infectious complication" OR "Prosthesis Failure" OR "Prosthesis-Related Infections" OR "Prosthesis-Related Infection" OR "Prosthesis Infections" OR "Prosthesis Infection" OR "Surgical Wound Infection" OR "Surgical Infection" OR "Surgical Wound Infections" OR "Surgical Infections" OR "Infection" OR "infection" OR "infections" OR "infected" OR "surgical injury" OR "surgical injuries" OR "complication" OR "complications" OR "Joint Dislocation" OR "Dislocations" OR "Dislocation" OR dislocat* OR "subluxation" OR Subluxat* OR "prosthesis loosening" OR "loosening" OR "malalignment" OR "malalignments" OR "malaligned" OR "Joint Instability" OR "Instability" OR "Instabilities" OR "Patella Dislocation" OR "Patellar Dislocation" OR "Patella Dislocation" OR "Patellar Dislocations" OR "Patella Dislocations") AND ts=((("variation" OR "variations" OR "difference" OR "differences") NEAR5 ("hospital" OR "hospitals")) OR (("variation" OR "variations" OR "difference" OR "differences") AND "Hospital") OR "hospital characteristics" OR "hospital outcome" OR "hospital outcomes" OR "international variation" OR "international variations" OR "ranking hospitals" OR "hospital rank" OR "hospital ranking" OR "hospital rankings" OR "hospital ranks")) **OR** (ts=("tha" OR "total hip" OR "Hip Replacement" OR "hip arthroplasty" OR "Hip Prosthesis" OR "Hip Prosthesis" OR "Hip Replacement" OR "Hip Arthroplasty" OR "Prosthetic Hip" OR "Prosthetic Hip" **OR** "tka" OR "total knee" OR "Knee Replacement" OR "knee arthroplasty" OR "Knee Prosthesis" OR "Knee Prosthesis" OR "Knee Replacement" OR "Knee Arthroplasty" OR "Prosthetic Knee" OR "Prosthetic Knees" OR (("Arthroplasty" OR "Joint Prosthesis" OR "Joint Arthroplasty" OR "Joint Replacement" OR "Joint Prosthesis" OR "Prosthetic Joint" OR "Prosthetic Joints") AND ("Hip" OR "hip" OR "hips" OR "Knee" OR "knee" OR "knees"))) AND ts=("revision rate" OR "revision rates" OR (("revision" OR "revisions") NEAR/5 ("rate" OR "rates" OR "median" OR "mean" OR "percentage" OR "percent" OR percent*)) OR "revision" OR "Repeat Surgery" OR "Joint Revision" OR "Reoperation" OR "Reoperation" OR Reoperat* OR "Re-operation" OR "Re-operat" OR "Length of Stay" OR "length of stay" OR "lengths of stay" OR "length of stays" OR "lengths of stays" OR "stay length" OR "stay lengths" OR "Hospital Readmission" OR "patient readmission" OR "hospital readmission" OR "patient re-admission" OR "hospital re-admission" OR "readmission" OR "re-admission" OR readmit* OR "re-admit" OR "Mortality" OR "mortality" OR mortalit* OR "death" OR "deaths" OR "Cause of Death" OR "fatality rate" OR "fatality rates" OR "Fatal Outcome" OR "Fatal Outcomes" OR "Survival Rate" OR "complication" OR "Postoperative Complication" OR "Prosthesis

Complication" OR "Surgical Infection" OR "infectious complication" OR "Prosthesis Failure" OR "Prosthesis-Related Infections" OR "Prosthesis-Related Infection" OR "Prosthesis Infections" OR "Prosthesis Infection" OR "Surgical Wound Infection" OR "Surgical Infection" OR "Surgical Wound Infections" OR "Surgical Infections" OR "Infection" OR "infection" OR "infections" OR "infected" OR "surgical injury" OR "surgical injuries" OR "complication" OR "complications" OR "Joint Dislocation" OR "Dislocations" OR "Dislocation" OR dislocat* OR "subluxation" OR Subluxat* OR "prosthesis loosening" OR "loosening" OR "malalignment" OR "malalignments" OR "malaligned" OR "Joint Instability" OR "Instability" OR "Instabilities" OR "Patella Dislocation" OR "Patellar Dislocation" OR "Patella Dislocation" OR "Patellar Dislocations" OR "Patella Dislocations") AND ti=((("variation" OR "variations" OR "difference" OR "differences") AND ("Hospital" OR "hospital" OR "hospitals")) OR "hospital characteristics" OR "hospital outcome" OR "hospital outcomes" OR "international variation" OR "international variations" OR "ranking hospitals" OR "hospital rank" OR "hospital ranking" OR "hospital rankings" OR "hospital ranks"))) AND py=(2009 OR 2010 OR 2011 OR 2012 OR 2013 OR 2014 OR 2015 OR 2016 OR 2017 OR 2018 OR 2019 OR 2020)

4. Cochrane library (http://www.cochranelibrary.com/)

(("tha" OR "total hip" OR "Hip Replacement" OR "hip arthroplasty" OR "Hip Prosthesis" OR "Hip Prosthesis" OR "Hip Replacement" OR "Hip Arthroplasty" OR "Prosthetic Hip" OR "Prosthetic Hip" **OR** "tka" OR "total knee" OR "Knee Replacement" OR "knee arthroplasty" OR "Knee Prosthesis" OR "Knee Prosthesis" OR "Knee Replacement" OR "Knee Arthroplasty" OR "Prosthetic Knee" OR "Prosthetic Knees" OR ((*"Arthroplasty" OR "Joint Prosthesis" OR "Joint Arthroplasty" OR "Joint Replacement" OR "Joint Prosthesis" OR "Prosthetic Joint" OR "Prosthetic Joints") AND ("Hip" OR "hip" OR "hips" OR "Knee" OR "knee" OR "knees"))):ti AND ("revision rate" OR "revision rates" OR (("revision" OR "revisions") AND ("rate" OR "rates" OR "median" OR "mean" OR "percentage" OR "percent" OR percent*)) OR "revision" OR "Repeat Surgery" OR "Joint Revision" OR "Reoperation" OR "Reoperation" OR Reoperat* OR "Re-operation" OR "Re-operat*" OR "Length of Stay" OR "length of stay" OR "lengths of stay" OR "length of stays" OR "lengths of stays" OR "stay length" OR "stay lengths" OR "Hospital Readmission" OR "patient readmission" OR "hospital readmission" OR "patient re-admission" OR "hospital re-admission" OR "readmission" OR "re-admission" OR readmit* OR "re-admit*" OR "Mortality" OR "mortality" OR mortalit* OR "death" OR "deaths" OR "Cause of Death" OR "fatality rate" OR "fatality rates" OR "Fatal Outcome" OR "Fatal Outcomes" OR "Survival Rate" OR "complication" OR "Postoperative Complication" OR "Prosthesis Complication" OR "Surgical Infection" OR "infectious complication" OR "Prosthesis Failure" OR "Prosthesis-Related Infections" OR "Prosthesis-Related

Infection" OR "Prosthesis Infections" OR "Prosthesis Infection" OR "Surgical Wound Infection" OR "Surgical Infection" OR "Surgical Wound Infections" OR "Surgical Infections" OR "Infection" OR "infection" OR "infections" OR "infected" OR "surgical injury" OR "surgical injuries" OR "complication" OR "complications" OR "Joint Dislocation" OR "Dislocations" OR "Dislocation" OR dislocat* OR "subluxation" OR Subluxat* OR "prosthesis loosening" OR "loosening" OR "malalignment" OR "malalignments" OR "malaligned" OR "Joint Instability" OR "Instability" OR "Instabilities" OR "Patella Dislocation" OR "Patellar Dislocation" OR "Patella Dislocation" OR "Patellar Dislocations" OR "Patella Dislocations"):ti AND ((("variation" OR "variations" OR "difference" OR "differences") NEAR/5 ("hospital" OR "hospitals")) OR (("variation" OR "variations" OR "difference" OR "differences") AND "Hospital") OR "hospital characteristics" OR "hospital outcome" OR "hospital outcomes" OR "international variation" OR "international variations" OR "ranking hospitals" OR "hospital rank" OR "hospital ranking" OR "hospital rankings" OR "hospital ranks"):ti,ab,kw) **OR** (("tha" OR "total hip" OR "Hip Replacement" OR "hip arthroplasty" OR "Hip Prosthesis" OR "Hip Prosthesis" OR "Hip Replacement" OR "Hip Arthroplasty" OR "Prosthetic Hip" OR "Prosthetic Hip" **OR** "tka" OR "total knee" OR "Knee Replacement" OR "knee arthroplasty" OR "Knee Prosthesis" OR "Knee Prosthesis" OR "Knee Replacement" OR "Knee Arthroplasty" OR "Prosthetic Knee" OR "Prosthetic Knees" OR ((*"Arthroplasty" OR "Joint Prosthesis" OR "Joint Arthroplasty" OR "Joint Replacement" OR "Joint Prosthesis" OR "Prosthetic Joint" OR "Prosthetic Joints") AND ("Hip" OR "hip" OR "hips" OR "Knee" OR "knee" OR "knees"))):ti,ab,kw AND ("revision rate" OR "revision rates" OR (("revision" OR "revisions") NEAR/5 ("rate" OR "rates" OR "median" OR "mean" OR "percentage" OR "percent" OR percent*)) OR "revision" OR "Repeat Surgery" OR "Joint Revision" OR "Reoperation" OR "Reoperation" OR Reoperat* OR "Re-operation" OR "Re-operat*" OR "Length of Stay" OR "length of stay" OR "lengths of stay" OR "length of stays" OR "lengths of stays" OR "stay length" OR "stay lengths" OR "Hospital Readmission" OR "patient readmission" OR "hospital readmission" OR "patient re-admission" OR "hospital re-admission" OR "readmission" OR "re-admission" OR readmit* OR "re-admit*" OR "Mortality" OR "mortality" OR mortalit* OR "death" OR "deaths" OR "Cause of Death" OR "fatality rate" OR "fatality rates" OR "Fatal Outcome" OR "Fatal Outcomes" OR "Survival Rate" OR "complication" OR "Postoperative Complication" OR "Prosthesis Complication" OR "Surgical Infection" OR "infectious complication" OR "Prosthesis Failure" OR "Prosthesis-Related Infections" OR "Prosthesis-Related Infection" OR "Prosthesis Infections" OR "Prosthesis Infection" OR "Surgical Wound Infection" OR "Surgical Infection" OR "Surgical Wound Infections" OR "Surgical Infections" OR "Infection" OR "infection" OR "infections" OR "infected" OR "surgical injury" OR "surgical injuries" OR "complication" OR "complications" OR "Joint Dislocation"

OR "Dislocations" OR "Dislocation" OR dislocat* OR "subluxation" OR Subluxat* OR "prosthesis loosening" OR "loosening" OR "malalignment" OR "malalignments" OR "malaligned" OR "Joint Instability" OR "Instability" OR "Instabilities" OR "Patella Dislocation" OR "Patellar Dislocation" OR "Patella Dislocation" OR "Patellar Dislocations" OR "Patella Dislocations"):ti,ab,kw AND ((("variation" OR "variations" OR "difference" OR "differences") NEAR/5 ("Hospital" OR "hospital" OR "hospitals")) OR "hospital characteristics" OR "hospital outcome" OR "hospital outcomes" OR "international variation" OR "international variations" OR "ranking hospitals" OR "hospital rank" OR "hospital ranking" OR "hospital rankings" OR "hospital ranks"):ti)

AND py=(2009 OR 2010 OR 2011 OR 2012 OR 2013 OR 2014 OR 2015 OR 2016 OR 2017 OR 2018 OR 2019 OR 2020)

5. Emcare (http://ovidsp.ovid.com/ovidweb.cgi?T=JS&NEWS=n&CSC=Y&PAGE=main&D=emcr)

(((("tha".ti OR "total hip".ti OR exp *"Hip Replacement"/ OR exp *"hip arthroplasty"/ OR exp *"Hip Prosthesis"/ OR "Hip Prosthesis".ti OR "Hip Replacement".ti OR "Hip Arthroplasty".ti OR "Prosthetic Hip".ti OR "Prosthetic Hip".ti **OR** "tka".ti OR "total knee".ti OR exp *"Knee Replacement"/ OR exp *"knee arthroplasty"/ OR exp *"Knee Prosthesis"/ OR "Knee Prosthesis".ti OR "Knee Replacement".ti OR "Knee Arthroplasty".ti OR "Prosthetic Knee".ti OR "Prosthetic Knees".ti OR ((*"Arthroplasty"/ OR *"Joint Prosthesis"/ OR "Joint Arthroplasty".ti OR "Joint Replacement".ti OR "Joint Prosthesis".ti OR "Prosthetic Joint".ti OR "Prosthetic Joints".ti) AND (exp *"Hip"/ OR "hip".ti OR "hips".ti OR exp *"Knee"/ OR "knee".ti OR "knees".ti))) AND ("revision rate".ti OR "revision rates".ti OR (("revision".ti OR "revisions".ti) AND ("rate".ti OR "rates".ti OR "median".ti OR "mean".ti OR "percentage".ti OR "percent".ti OR percent*.ti)) OR "revision".ti OR "Repeat Surgery".ti OR "Joint Revision".ti OR *"Reoperation"/ OR "Reoperation".ti OR Reoperat*.ti OR "Re-operation".ti OR "Re-operat*".ti OR *"Length of Stay"/ OR "length of stay".ti OR "lengths of stay".ti OR "length of stays".ti OR "lengths of stays".ti OR "stay length".ti OR "stay lengths".ti OR *"Hospital Readmission"/ OR "patient readmission".ti OR "hospital readmission".ti OR "patient re-admission".ti OR "hospital re-admission".ti OR "readmission".ti OR "re-admission".ti OR readmit*.ti OR "re-admit*".ti OR exp *"Mortality"/ OR "mortality".ti OR mortalit*.ti OR "death".ti OR "deaths".ti OR "Cause of Death".ti OR "fatality rate".ti OR "fatality rates".ti OR "Fatal Outcome".ti OR "Fatal Outcomes".ti OR *"Survival Rate"/ OR *"complication"/ OR "Postoperative Complication"/ OR exp *"Prosthesis Complication"/ OR "Surgical Infection"/ OR *"infectious complication"/ OR "Prosthesis Failure".ti OR "Prosthesis-Related Infections".ti OR "Prosthesis-Related

Infection".ti OR "Prosthesis Infections".ti OR "Prosthesis Infection".ti OR "Surgical Wound Infection".ti OR "Surgical Infection".ti OR "Surgical Wound Infections".ti OR "Surgical Infections".ti OR *"Infection"/ OR "infection".ti OR "infections".ti OR "infected".ti OR "surgical injury".ti OR "surgical injuries".ti OR "complication".ti OR "complications".ti OR *"Joint Dislocation"/ OR "Dislocations".ti OR "Dislocation".ti OR dislocat*.ti OR *"subluxation"/ OR Subluxat*.ti OR *"prosthesis loosening"/ OR "loosening".ti OR "malalignment".ti OR "malalignments".ti OR "malaligned".ti OR exp *"Joint Instability"/ OR "Instability".ti OR "Instabilities".ti OR *"Patella Dislocation"/ OR "Patellar Dislocation".ti OR "Patella Dislocation".ti OR "Patellar Dislocations".ti OR "Patella Dislocations".ti) AND ((("variation".mp OR "variations".mp OR "difference".mp OR "differences".mp) ADJ5 ("hospital".mp OR "hospitals".mp)) OR (("variation".mp OR "variations".mp OR "difference".mp OR "differences".mp) AND exp "Hospital"/) OR "hospital characteristics".mp OR "hospital outcome".mp OR "hospital outcomes".mp OR "international variation".mp OR "international variations".mp OR "ranking hospitals".mp OR "hospital rank".mp OR "hospital ranking".mp OR "hospital rankings".mp OR "hospital ranks".mp)) **OR** (("tha".ti,ab OR "total hip".ti,ab OR exp "Hip Replacement"/ OR exp "hip arthroplasty"/ OR exp "Hip Prosthesis"/ OR "Hip Prosthesis".ti,ab OR "Hip Replacement".ti,ab OR "Hip Arthroplasty".ti,ab OR "Prosthetic Hip".ti,ab OR "Prosthetic Hip".ti,ab **OR** "tka".ti,ab OR "total knee".ti,ab OR exp "Knee Replacement"/ OR exp "knee arthroplasty"/ OR exp "Knee Prosthesis"/ OR "Knee Prosthesis".ti,ab OR "Knee Replacement".ti,ab OR "Knee Arthroplasty".ti,ab OR "Prosthetic Knee".ti,ab OR "Prosthetic Knees".ti,ab OR ((*"Arthroplasty"/ OR *"Joint Prosthesis"/ OR "Joint Arthroplasty".ti OR "Joint Replacement".ti OR "Joint Prosthesis".ti OR "Prosthetic Joint".ti OR "Prosthetic Joints".ti) AND (exp *"Hip"/ OR "hip".ti OR "hips".ti OR exp *"Knee"/ OR "knee".ti OR "knees".ti))) AND ("revision rate".mp OR "revision rates".mp OR (("revision".mp OR "revisions".mp) ADJ5 ("rate".mp OR "rates".mp OR "median".mp OR "mean".mp OR "percentage".mp OR "percent".mp OR percent*.mp)) OR "revision".mp OR "Repeat Surgery".mp OR "Joint Revision".mp OR "Reoperation"/ OR "Reoperation".mp OR Reoperat*.mp OR "Re-operation".mp OR "Re-operat*".mp OR "Length of Stay"/ OR "length of stay".mp OR "lengths of stay".mp OR "length of stays".mp OR "lengths of stays".mp OR "stay length".mp OR "stay lengths".mp OR "Hospital Readmission"/ OR "patient readmission".mp OR "hospital readmission".mp OR "patient re-admission".mp OR "hospital re-admission".mp OR "readmission".mp OR "re-admission".mp OR readmit*.mp OR "re-admit*".mp OR exp "Mortality"/ OR "mortality".mp OR mortalit*.mp OR "death".mp OR "deaths".mp OR "Cause of Death".mp OR "fatality rate".mp OR "fatality rates".mp OR "Fatal Outcome".mp OR "Fatal Outcomes".mp OR "Survival Rate"/ OR *"complication"/ OR "Postoperative Complication"/ OR exp "Prosthesis Complication"/ OR "Surgical Infection"/ OR "infectious complication"/

OR "Prosthesis Failure".mp OR "Prosthesis-Related Infections".mp OR "Prosthesis-Related Infection".mp OR "Prosthesis Infections".mp OR "Prosthesis Infection".mp OR "Surgical Wound Infection".mp OR "Surgical Infection".mp OR "Surgical Wound Infections".mp OR "Surgical Infections".mp OR "Infection"/ OR "infection".mp OR "infections".mp OR "infected".mp OR "surgical injury".mp OR "surgical injuries".mp OR "complication".mp OR "complications".mp OR "Joint Dislocation"/ OR "Dislocations".mp OR "Dislocation".mp OR dislocat*.mp OR "subluxation"/ OR Subluxat*.mp OR "prosthesis loosening"/ OR "loosening".mp OR "malalignment".mp OR "malalignments".mp OR "malaligned".mp OR exp "Joint Instability"/ OR "Instability".mp OR "Instabilities".mp OR "Patella Dislocation"/ OR "Patellar Dislocation".mp OR "Patella Dislocation".mp OR "Patellar Dislocations".mp OR "Patella Dislocations".mp) AND ((("variation".ti OR "variations".ti OR "difference".ti OR "differences".ti) AND (exp *"Hospital"/ OR "hospital".ti OR "hospitals".ti)) OR "hospital characteristics".ti OR "hospital outcome".ti OR "hospital outcomes".ti OR "international variation".ti OR "international variations".ti OR "ranking hospitals".ti OR "hospital rank".ti OR "hospital ranking".ti OR "hospital rankings".ti OR "hospital ranks".ti)) AND (2009 OR 201* OR 202*).yr

6. Academic Search Premier (http://search.ebscohost.com/login.aspx?authtype=ip,uid&profile=lumc&defaultdb=aph)

((TI("tha" OR "total hip" OR "Hip Replacement" OR "hip arthroplasty" OR "Hip Prosthesis" OR "Hip Prosthesis" OR "Hip Replacement" OR "Hip Arthroplasty" OR "Prosthetic Hip" OR "Prosthetic Hip" **OR** "tka" OR "total knee" OR "Knee Replacement" OR "knee arthroplasty" OR "Knee Prosthesis" OR "Knee Prosthesis" OR "Knee Replacement" OR "Knee Arthroplasty" OR "Prosthetic Knee" OR "Prosthetic Knees" OR (("Arthroplasty" OR "Joint Prosthesis" OR "Joint Arthroplasty" OR "Joint Replacement" OR "Joint Prosthesis" OR "Prosthetic Joint" OR "Prosthetic Joints") AND ("Hip" OR "hip" OR "hips" OR "Knee" OR "knee" OR "knees"))) AND TI("revision rate" OR "revision rates" OR (("revision" OR "revisions") AND ("rate" OR "rates" OR "median" OR "mean" OR "percentage" OR "percent" OR percent*)) OR "revision" OR "Repeat Surgery" OR "Joint Revision" OR "Reoperation" OR "Reoperation" OR Reoperat* OR "Re-operation" OR "Re-operat" OR "Length of Stay" OR "length of stay" OR "lengths of stay" OR "length of stays" OR "lengths of stays" OR "stay length" OR "stay lengths" OR "Hospital Readmission" OR "patient readmission" OR "hospital readmission" OR "patient re-admission" OR "hospital re-admission" OR "readmission" OR "re-admission" OR readmit* OR "re-admit" OR "Mortality" OR "mortality" OR mortalit* OR "death" OR "deaths" OR "Cause of Death" OR "fatality rate" OR "fatality rates" OR "Fatal Outcome" OR "Fatal Outcomes" OR "Survival Rate" OR "complication" OR "Postoperative Complication" OR "Prosthesis Complication" OR "Surgical Infection" OR "infectious complication"

OR "Prosthesis Failure" OR "Prosthesis-Related Infections" OR "Prosthesis-Related Infection" OR "Prosthesis Infections" OR "Prosthesis Infection" OR "Surgical Wound Infection" OR "Surgical Infection" OR "Surgical Wound Infections" OR "Surgical Infections" OR "Infection" OR "infection" OR "infections" OR "infected" OR "surgical injury" OR "surgical injuries" OR "complication" OR "complications" OR "Joint Dislocation" OR "Dislocations" OR "Dislocation" OR dislocat* OR "subluxation" OR Subluxat* OR "prosthesis loosening" OR "loosening" OR "malalignment" OR "malalignments" OR "malaligned" OR "Joint Instability" OR "Instability" OR "Instabilities" OR "Patella Dislocation" OR "Patellar Dislocation" OR "Patella Dislocation" OR "Patellar Dislocations" OR "Patella Dislocations") AND SU((("variation" OR "variations" OR "difference" OR "differences") NEAR5 ("hospital" OR "hospitals")) OR (("variation" OR "variations" OR "difference" OR "differences") AND "Hospital") OR "hospital characteristics" OR "hospital outcome" OR "hospital outcomes" OR "international variation" OR "international variations" OR "ranking hospitals" OR "hospital rank" OR "hospital ranking" OR "hospital rankings" OR "hospital ranks")) **OR** (SU("tha" OR "total hip" OR "Hip Replacement" OR "hip arthroplasty" OR "Hip Prosthesis" OR "Hip Prosthesis" OR "Hip Replacement" OR "Hip Arthroplasty" OR "Prosthetic Hip" OR "Prosthetic Hip" **OR** "tka" OR "total knee" OR "Knee Replacement" OR "knee arthroplasty" OR "Knee Prosthesis" OR "Knee Prosthesis" OR "Knee Replacement" OR "Knee Arthroplasty" OR "Prosthetic Knee" OR "Prosthetic Knees" OR (("Arthroplasty" OR "Joint Prosthesis" OR "Joint Arthroplasty" OR "Joint Replacement" OR "Joint Prosthesis" OR "Prosthetic Joint" OR "Prosthetic Joints") AND ("Hip" OR "hip" OR "hips" OR "Knee" OR "knee" OR "knees"))) AND SU("revision rate" OR "revision rates" OR (("revision" OR "revisions") NEAR/5 ("rate" OR "rates" OR "median" OR "mean" OR "percentage" OR "percent" OR percent*)) OR "revision" OR "Repeat Surgery" OR "Joint Revision" OR "Reoperation" OR "Reoperation" OR Reoperat* OR "Re-operation" OR "Re-operat" OR "Length of Stay" OR "length of stay" OR "lengths of stay" OR "length of stays" OR "lengths of stays" OR "stay length" OR "stay lengths" OR "Hospital Readmission" OR "patient readmission" OR "hospital readmission" OR "patient re-admission" OR "hospital re-admission" OR "readmission" OR "re-admission" OR readmit* OR "re-admit" OR "Mortality" OR "mortality" OR mortalit* OR "death" OR "deaths" OR "Cause of Death" OR "fatality rate" OR "fatality rates" OR "Fatal Outcome" OR "Fatal Outcomes" OR "Survival Rate" OR "complication" OR "Postoperative Complication" OR "Prosthesis Complication" OR "Surgical Infection" OR "infectious complication" OR "Prosthesis Failure" OR "Prosthesis-Related Infections" OR "Prosthesis-Related Infection" OR "Prosthesis Infections" OR "Prosthesis Infection" OR "Surgical Wound Infection" OR "Surgical Infection" OR "Surgical Wound Infections" OR "Surgical Infections" OR "Infection" OR "infection" OR "infections" OR "infected" OR "surgical injury" OR

"surgical injuries" OR "complication" OR "complications" OR "Joint Dislocation" OR "Dislocations" OR "Dislocation" OR dislocat* OR "subluxation" OR Subluxat* OR "prosthesis loosening" OR "loosening" OR "malalignment" OR "malalignments" OR "malaligned" OR "Joint Instability" OR "Instability" OR "Instabilities" OR "Patella Dislocation" OR "Patellar Dislocation" OR "Patella Dislocation" OR "Patellar Dislocations" OR "Patella Dislocations") AND TI((("variation" OR "variations" OR "difference" OR "differences") AND ("Hospital" OR "hospital" OR "hospitals")) OR "hospital characteristics" OR "hospital outcome" OR "hospital outcomes" OR "international variation" OR "international variations" OR "ranking hospitals" OR "hospital rank" OR "hospital ranking" OR "hospital rankings" OR "hospital ranks")))

((TI("tha" OR "total hip" OR "Hip Replacement" OR "hip arthroplasty" OR "Hip Prosthesis" OR "Hip Prosthesis" OR "Hip Replacement" OR "Hip Arthroplasty" OR "Prosthetic Hip" OR "Prosthetic Hip" **OR** "tka" OR "total knee" OR "Knee Replacement" OR "knee arthroplasty" OR "Knee Prosthesis" OR "Knee Prosthesis" OR "Knee Replacement" OR "Knee Arthroplasty" OR "Prosthetic Knee" OR "Prosthetic Knees" OR (("Arthroplasty" OR "Joint Prosthesis" OR "Joint Arthroplasty" OR "Joint Replacement" OR "Joint Prosthesis" OR "Prosthetic Joint" OR "Prosthetic Joints") AND ("Hip" OR "hip" OR "hips" OR "Knee" OR "knee" OR "knees"))) AND TI("revision rate" OR "revision rates" OR (("revision" OR "revisions") AND ("rate" OR "rates" OR "median" OR "mean" OR "percentage" OR "percent" OR percent*)) OR "revision" OR "Repeat Surgery" OR "Joint Revision" OR "Reoperation" OR "Reoperation" OR Reoperat* OR "Re-operation" OR "Re-operat" OR "Length of Stay" OR "length of stay" OR "lengths of stay" OR "length of stays" OR "lengths of stays" OR "stay length" OR "stay lengths" OR "Hospital Readmission" OR "patient readmission" OR "hospital readmission" OR "patient re-admission" OR "hospital re-admission" OR "readmission" OR "re-admission" OR readmit* OR "re-admit" OR "Mortality" OR "mortality" OR mortalit* OR "death" OR "deaths" OR "Cause of Death" OR "fatality rate" OR "fatality rates" OR "Fatal Outcome" OR "Fatal Outcomes" OR "Survival Rate" OR "complication" OR "Postoperative Complication" OR "Prosthesis Complication" OR "Surgical Infection" OR "infectious complication" OR "Prosthesis Failure" OR "Prosthesis-Related Infections" OR "Prosthesis-Related Infection" OR "Prosthesis Infections" OR "Prosthesis Infection" OR "Surgical Wound Infection" OR "Surgical Infection" OR "Surgical Wound Infections" OR "Surgical Infections" OR "Infection" OR "infection" OR "infections" OR "infected" OR "surgical injury" OR "surgical injuries" OR "complication" OR "complications" OR "Joint Dislocation" OR "Dislocations" OR "Dislocation" OR dislocat* OR "subluxation" OR Subluxat* OR "prosthesis loosening" OR "loosening" OR "malalignment" OR "malalignments" OR "malaligned" OR "Joint Instability" OR "Instability" OR "Instabilities" OR "Patella Dislocation" OR "Patellar Dislocation"

OR "Patella Dislocation" OR "Patellar Dislocations" OR "Patella Dislocations") AND AB((("variation" OR "variations" OR "difference" OR "differences") NEAR5 ("hospital" OR "hospitals")) OR (("variation" OR "variations" OR "difference" OR "differences") AND "Hospital") OR "hospital characteristics" OR "hospital outcome" OR "hospital outcomes" OR "international variation" OR "international variations" OR "ranking hospitals" OR "hospital rank" OR "hospital ranking" OR "hospital rankings" OR "hospital ranks")) **OR** (AB("tha" OR "total hip" OR "Hip Replacement" OR "hip arthroplasty" OR "Hip Prosthesis" OR "Hip Prosthesis" OR "Hip Replacement" OR "Hip Arthroplasty" OR "Prosthetic Hip" OR "Prosthetic Hip" **OR** "tka" OR "total knee" OR "Knee Replacement" OR "knee arthroplasty" OR "Knee Prosthesis" OR "Knee Prosthesis" OR "Knee Replacement" OR "Knee Arthroplasty" OR "Prosthetic Knee" OR "Prosthetic Knees" OR (("Arthroplasty" OR "Joint Prosthesis" OR "Joint Arthroplasty" OR "Joint Replacement" OR "Joint Prosthesis" OR "Prosthetic Joint" OR "Prosthetic Joints") AND ("Hip" OR "hip" OR "hips" OR "Knee" OR "knee" OR "knees"))) AND AB("revision rate" OR "revision rates" OR (("revision" OR "revisions") NEAR/5 ("rate" OR "rates" OR "median" OR "mean" OR "percentage" OR "percent" OR percent*)) OR "revision" OR "Repeat Surgery" OR "Joint Revision" OR "Reoperation" OR "Reoperation" OR Reoperat* OR "Re-operation" OR "Re-operat" OR "Length of Stay" OR "length of stay" OR "lengths of stay" OR "length of stays" OR "lengths of stays" OR "stay length" OR "stay lengths" OR "Hospital Readmission" OR "patient readmission" OR "hospital readmission" OR "patient re-admission" OR "hospital re-admission" OR "readmission" OR "re-admission" OR readmit* OR "re-admit" OR "Mortality" OR "mortality" OR mortalit* OR "death" OR "deaths" OR "Cause of Death" OR "fatality rate" OR "fatality rates" OR "Fatal Outcome" OR "Fatal Outcomes" OR "Survival Rate" OR "complication" OR "Postoperative Complication" OR "Prosthesis Complication" OR "Surgical Infection" OR "infectious complication" OR "Prosthesis Failure" OR "Prosthesis-Related Infections" OR "Prosthesis-Related Infection" OR "Prosthesis Infections" OR "Prosthesis Infection" OR "Surgical Wound Infection" OR "Surgical Infection" OR "Surgical Wound Infections" OR "Surgical Infections" OR "Infection" OR "infection" OR "infections" OR "infected" OR "surgical injury" OR "surgical injuries" OR "complication" OR "complications" OR "Joint Dislocation" OR "Dislocations" OR "Dislocation" OR dislocat* OR "subluxation" OR Subluxat* OR "prosthesis loosening" OR "loosening" OR "malalignment" OR "malalignments" OR "malaligned" OR "Joint Instability" OR "Instability" OR "Instabilities" OR "Patella Dislocation" OR "Patellar Dislocation" OR "Patella Dislocation" OR "Patellar Dislocations" OR "Patella Dislocations") AND TI((("variation" OR "variations" OR "difference" OR "differences") AND ("Hospital" OR "hospital" OR "hospitals")) OR "hospital characteristics" OR "hospital outcome" OR "hospital outcomes" OR

"international variation" OR "international variations" OR "ranking hospitals" OR "hospital rank" OR "hospital ranking" OR "hospital rankings" OR "hospital ranks"))**)**

((TI("tha" OR "total hip" OR "Hip Replacement" OR "hip arthroplasty" OR "Hip Prosthesis" OR "Hip Prosthesis" OR "Hip Replacement" OR "Hip Arthroplasty" OR "Prosthetic Hip" OR "Prosthetic Hip" **OR** "tka" OR "total knee" OR "Knee Replacement" OR "knee arthroplasty" OR "Knee Prosthesis" OR "Knee Prosthesis" OR "Knee Replacement" OR "Knee Arthroplasty" OR "Prosthetic Knee" OR "Prosthetic Knees" OR (("Arthroplasty" OR "Joint Prosthesis" OR "Joint Arthroplasty" OR "Joint Replacement" OR "Joint Prosthesis" OR "Prosthetic Joint" OR "Prosthetic Joints") AND ("Hip" OR "hip" OR "hips" OR "Knee" OR "knee" OR "knees"))) AND TI("revision rate" OR "revision rates" OR (("revision" OR "revisions") AND ("rate" OR "rates" OR "median" OR "mean" OR "percentage" OR "percent" OR percent*)) OR "revision" OR "Repeat Surgery" OR "Joint Revision" OR "Reoperation" OR "Reoperation" OR Reoperat* OR "Re-operation" OR "Re-operat" OR "Length of Stay" OR "length of stay" OR "lengths of stay" OR "length of stays" OR "lengths of stays" OR "stay length" OR "stay lengths" OR "Hospital Readmission" OR "patient readmission" OR "hospital readmission" OR "patient re-admission" OR "hospital re-admission" OR "readmission" OR "re-admission" OR readmit* OR "re-admit" OR "Mortality" OR "mortality" OR mortalit* OR "death" OR "deaths" OR "Cause of Death" OR "fatality rate" OR "fatality rates" OR "Fatal Outcome" OR "Fatal Outcomes" OR "Survival Rate" OR "complication" OR "Postoperative Complication" OR "Prosthesis Complication" OR "Surgical Infection" OR "infectious complication" OR "Prosthesis Failure" OR "Prosthesis-Related Infections" OR "Prosthesis-Related Infection" OR "Prosthesis Infections" OR "Prosthesis Infection" OR "Surgical Wound Infection" OR "Surgical Infection" OR "Surgical Wound Infections" OR "Surgical Infections" OR "Infection" OR "infection" OR "infections" OR "infected" OR "surgical injury" OR "surgical injuries" OR "complication" OR "complications" OR "Joint Dislocation" OR "Dislocations" OR "Dislocation" OR dislocat* OR "subluxation" OR Subluxat* OR "prosthesis loosening" OR "loosening" OR "malalignment" OR "malalignments" OR "malaligned" OR "Joint Instability" OR "Instability" OR "Instabilities" OR "Patella Dislocation" OR "Patellar Dislocation" OR "Patella Dislocation" OR "Patellar Dislocations" OR "Patella Dislocations") AND KW((("variation" OR "variations" OR "difference" OR "differences") NEAR5 ("hospital" OR "hospitals")) OR (("variation" OR "variations" OR "difference" OR "differences") AND "Hospital") OR "hospital characteristics" OR "hospital outcome" OR "hospital outcomes" OR "international variation" OR "international variations" OR "ranking hospitals" OR "hospital rank" OR "hospital ranking" OR "hospital rankings" OR "hospital ranks")) **OR** (KW("tha" OR "total hip" OR "Hip Replacement" OR "hip arthroplasty" OR "Hip Prosthesis" OR "Hip Prosthesis" OR

Chapter 2

"Hip Replacement" OR "Hip Arthroplasty" OR "Prosthetic Hip" OR "Prosthetic Hip" **OR** "tka" OR "total knee" OR "Knee Replacement" OR "knee arthroplasty" OR "Knee Prosthesis" OR "Knee Prosthesis" OR "Knee Replacement" OR "Knee Arthroplasty" OR "Prosthetic Knee" OR "Prosthetic Knees" OR (("Arthroplasty" OR "Joint Prosthesis" OR "Joint Arthroplasty" OR "Joint Replacement" OR "Joint Prosthesis" OR "Prosthetic Joint" OR "Prosthetic Joints") AND ("Hip" OR "hip" OR "hips" OR "Knee" OR "knee" OR "knees"))) AND KW("revision rate" OR "revision rates" OR (("revision" OR "revisions") NEAR/5 ("rate" OR "rates" OR "median" OR "mean" OR "percentage" OR "percent" OR percent*)) OR "revision" OR "Repeat Surgery" OR "Joint Revision" OR "Reoperation" OR "Reoperation" OR Reoperat* OR "Re-operation" OR "Re-operat" OR "Length of Stay" OR "length of stay" OR "lengths of stay" OR "length of stays" OR "lengths of stays" OR "stay length" OR "stay lengths" OR "Hospital Readmission" OR "patient readmission" OR "hospital readmission" OR "patient re-admission" OR "hospital re-admission" OR "readmission" OR "re-admission" OR readmit* OR "re-admit" OR "Mortality" OR "mortality" OR mortalit* OR "death" OR "deaths" OR "Cause of Death" OR "fatality rate" OR "fatality rates" OR "Fatal Outcome" OR "Fatal Outcomes" OR "Survival Rate" OR "complication" OR "Postoperative Complication" OR "Prosthesis Complication" OR "Surgical Infection" OR "infectious complication" OR "Prosthesis Failure" OR "Prosthesis-Related Infections" OR "Prosthesis-Related Infection" OR "Prosthesis Infections" OR "Prosthesis Infection" OR "Surgical Wound Infection" OR "Surgical Infection" OR "Surgical Wound Infections" OR "Surgical Infections" OR "Infection" OR "infection" OR "infections" OR "infected" OR "surgical injury" OR "surgical injuries" OR "complication" OR "complications" OR "Joint Dislocation" OR "Dislocations" OR "Dislocation" OR dislocat* OR "subluxation" OR Subluxat* OR "prosthesis loosening" OR "loosening" OR "malalignment" OR "malalignments" OR "malaligned" OR "Joint Instability" OR "Instability" OR "Instabilities" OR "Patella Dislocation" OR "Patellar Dislocation" OR "Patella Dislocation" OR "Patellar Dislocations" OR "Patella Dislocations") AND TI((("variation" OR "variations" OR "difference" OR "differences") AND ("Hospital" OR "hospital" OR "hospitals")) OR "hospital characteristics" OR "hospital outcome" OR "hospital outcomes" OR "international variation" OR "international variations" OR "ranking hospitals" OR "hospital rank" OR "hospital ranking" OR "hospital rankings" OR "hospital ranks")))

((TI("tha" OR "total hip" OR "Hip Replacement" OR "hip arthroplasty" OR "Hip Prosthesis" OR "Hip Prosthesis" OR "Hip Replacement" OR "Hip Arthroplasty" OR "Prosthetic Hip" OR "Prosthetic Hip" **OR** "tka" OR "total knee" OR "Knee Replacement" OR "knee arthroplasty" OR "Knee Prosthesis" OR "Knee Prosthesis" OR "Knee Replacement" OR "Knee Arthroplasty" OR "Prosthetic Knee" OR "Prosthetic Knees" OR (("Arthroplasty" OR "Joint Prosthesis" OR "Joint Arthroplasty" OR "Joint

Replacement" OR "Joint Prosthesis" OR "Prosthetic Joint" OR "Prosthetic Joints") AND ("Hip" OR "hip" OR "hips" OR "Knee" OR "knee" OR "knees"))) AND TI("revision rate" OR "revision rates" OR (("revision" OR "revisions") AND ("rate" OR "rates" OR "median" OR "mean" OR "percentage" OR "percent" OR percent*)) OR "revision" OR "Repeat Surgery" OR "Joint Revision" OR "Reoperation" OR "Reoperation" OR Reoperat* OR "Re-operation" OR "Re-operat" OR "Length of Stay" OR "length of stay" OR "lengths of stay" OR "length of stays" OR "lengths of stays" OR "stay length" OR "stay lengths" OR "Hospital Readmission" OR "patient readmission" OR "hospital readmission" OR "patient re-admission" OR "hospital re-admission" OR "readmission" OR "re-admission" OR readmit* OR "re-admit" OR "Mortality" OR "mortality" OR mortalit* OR "death" OR "deaths" OR "Cause of Death" OR "fatality rate" OR "fatality rates" OR "Fatal Outcome" OR "Fatal Outcomes" OR "Survival Rate" OR "complication" OR "Postoperative Complication" OR "Prosthesis Complication" OR "Surgical Infection" OR "infectious complication" OR "Prosthesis Failure" OR "Prosthesis-Related Infections" OR "Prosthesis-Related Infection" OR "Prosthesis Infections" OR "Prosthesis Infection" OR "Surgical Wound Infection" OR "Surgical Infection" OR "Surgical Wound Infections" OR "Surgical Infections" OR "Infection" OR "infection" OR "infections" OR "infected" OR "surgical injury" OR "surgical injuries" OR "complication" OR "complications" OR "Joint Dislocation" OR "Dislocations" OR "Dislocation" OR dislocat* OR "subluxation" OR Subluxat* OR "prosthesis loosening" OR "loosening" OR "malalignment" OR "malalignments" OR "malaligned" OR "Joint Instability" OR "Instability" OR "Instabilities" OR "Patella Dislocation" OR "Patellar Dislocation" OR "Patella Dislocation" OR "Patellar Dislocations" OR "Patella Dislocations") AND TI((("variation" OR "variations" OR "difference" OR "differences") NEAR5 ("hospital" OR "hospitals")) OR (("variation" OR "variations" OR "difference" OR "differences") AND "Hospital") OR "hospital characteristics" OR "hospital outcome" OR "hospital outcomes" OR "international variation" OR "international variations" OR "ranking hospitals" OR "hospital rank" OR "hospital ranking" OR "hospital rankings" OR "hospital ranks")) **OR** (TI("tha" OR "total hip" OR "Hip Replacement" OR "hip arthroplasty" OR "Hip Prosthesis" OR "Hip Prosthesis" OR "Hip Replacement" OR "Hip Arthroplasty" OR "Prosthetic Hip" OR "Prosthetic Hip" **OR** "tka" OR "total knee" OR "Knee Replacement" OR "knee arthroplasty" OR "Knee Prosthesis" OR "Knee Prosthesis" OR "Knee Replacement" OR "Knee Arthroplasty" OR "Prosthetic Knee" OR "Prosthetic Knees" OR (("Arthroplasty" OR "Joint Prosthesis" OR "Joint Arthroplasty" OR "Joint Replacement" OR "Joint Prosthesis" OR "Prosthetic Joint" OR "Prosthetic Joints") AND ("Hip" OR "hip" OR "hips" OR "Knee" OR "knee" OR "knees"))) AND TI("revision rate" OR "revision rates" OR (("revision" OR "revisions") NEAR/5 ("rate" OR "rates" OR "median" OR "mean" OR "percentage" OR "percent" OR percent*)) OR "revision"

OR "Repeat Surgery" OR "Joint Revision" OR "Reoperation" OR "Reoperation" OR Reoperat* OR "Re-operation" OR "Re-operat" OR "Length of Stay" OR "length of stay" OR "lengths of stay" OR "length of stays" OR "lengths of stays" OR "stay length" OR "stay lengths" OR "Hospital Readmission" OR "patient readmission" OR "hospital readmission" OR "patient re-admission" OR "hospital re-admission" OR "readmission" OR "re-admission" OR readmit* OR "re-admit" OR "Mortality" OR "mortality" OR mortalit* OR "death" OR "deaths" OR "Cause of Death" OR "fatality rate" OR "fatality rates" OR "Fatal Outcome" OR "Fatal Outcomes" OR "Survival Rate" OR "complication" OR "Postoperative Complication" OR "Prosthesis Complication" OR "Surgical Infection" OR "infectious complication" OR "Prosthesis Failure" OR "Prosthesis-Related Infections" OR "Prosthesis-Related Infection" OR "Prosthesis Infections" OR "Prosthesis Infection" OR "Surgical Wound Infection" OR "Surgical Infection" OR "Surgical Wound Infections" OR "Surgical Infections" OR "Infection" OR "infection" OR "infections" OR "infected" OR "surgical injury" OR "surgical injuries" OR "complication" OR "complications" OR "Joint Dislocation" OR "Dislocations" OR "Dislocation" OR dislocat* OR "subluxation" OR Subluxat* OR "prosthesis loosening" OR "loosening" OR "malalignment" OR "malalignments" OR "malaligned" OR "Joint Instability" OR "Instability" OR "Instabilities" OR "Patella Dislocation" OR "Patellar Dislocation" OR "Patella Dislocation" OR "Patellar Dislocations" OR "Patella Dislocations") AND TI((("variation" OR "variations" OR "difference" OR "differences") AND ("Hospital" OR "hospital" OR "hospitals")) OR "hospital characteristics" OR "hospital outcome" OR "hospital outcomes" OR "international variation" OR "international variations" OR "ranking hospitals" OR "hospital rank" OR "hospital ranking" OR "hospital rankings" OR "hospital ranks")))

AND py=(2009 OR 2010 OR 2011 OR 2012 OR 2013 OR 2014 OR 2015 OR 2016 OR 2017 OR 2018 OR 2019 OR 2020)

Appendix IIA Study details

Authors	Arthroplasty	Number of patients	Number of hospitals	Time period	Data category	Data collection method	Data completeness	Indicator(s) reported
Bozic(14) *USA, 2014*	THA&TKA *Combined*	THA: 251199 TKA: 626781 THA + TKA: 118	3479	2008 to 2010	Administrative	Routinely	---	Complications
Thirukumaran(15) *USA, 2020*	THA + TKA *Combined*	---	2326	2009 to 2012 & 2013 to 2016	Administrative	Routinely	---	Readmission Complications
Courtney(39) *USA, 2018*	THA + TKA *Combined*	458259	2702	April 1, 2012 to March 31, 2015 & July 1, 2012 to June 30, 2015	Administrative	Routinely	>75%	Readmission Complications
van Schie(30) *The Netherlands, 2020*	THA&TKA *Separately*	THA: 86468 TKA: 73077	THA:97 TKA:98	2014 to 2016	Clinical	Clinician-reported	THA&TKA: 98% Revisions: 96%	Revision
Graham(21) *USA, 2019*	THA&TKA *Separately*	THA: 45833 TKA: 24338	THA: 96 TKA: 95	October 1, 2007 to September 30, 2014	Clinical	Clinician-reported	THA&TKA: 83%	Readmission Complications
Sheetz(9) *USA, 2019*	THA	143174	106	2005 to 2014	Administrative	Routinely	---	Readmission Complications Mortality
Bottle(36) *The United Kingdom, 2018*	THA&TKA *Separately*	THA: 259980 TKA: 311033	---	April 2010 to March 2015	Administrative	Routinely	---	Readmission
Padegimas(34) *USA, 2018*	TKA	430	2	2014	Administrative	Routinely	---	Revision Readmission Complications LOS
Marang-vd Mheen(50) *The Netherlands, 2017*	THA + TKA *Combined*	41397	22	July 2007 to December 2010	Administrative	Routinely	---	Complications LOS

Appendix IIA Continued

Authors	Arthroplasty	Number of patients	Number of hospitals	Time period	Data category	Data collection method	Data completeness	Indicator(s) reported
Hollis(40) *USA, 2017*	THA + TKA *Combined*	THA: 22797 TKA: 41927	84	October 1, 2007 to September 30, 2014	Clinical	Clinician-reported	---	Readmission
Qian(42) *USA, 2013*	THA	54405	77	June 2006 to September 2010	Administrative	Routinely	---	Complications
Voorn(10) *The Netherlands, 2017*	THA&TKA *Separately*	THA: 1163 TKA: 986	THA: 23 TKA: 23	May 2013 to October 2013	Clinical	Routinely	---	Complications LOS
Cram(43) *USA, 2012*	THA	202773	---	2007 to 2008	Administrative	Routinely	---	Complications
Cai(49) *USA, 2012*	TKA	610285	3101	July 1, 2002 to June 30, 2005	Administrative	Routinely	---	Complications
Padegimas(31) *USA, 2017*	THA	658	2	2014	Administrative	Routinely	---	Revision Readmission LOS
Chen(52) *USA, 2017*	THA&TKA *Separately*	THA: 122 TKA: 84	THA: 2 TKA: 2	January 2010 to August 2014	Clinical	Routinely	---	LOS
Courtney(16) *USA, 2017*	THA + TKA *Combined*	458259	2702	April 1, 2012 to March 31, 2015 & July 1, 2012 to June 30, 2015	Administrative	Routinely	>75%	Readmission Complications
Husni(54) *USA, 2010*	TKA	9157	295	2000	Administrative	Routinely	---	LOS
Hofstede(41) *The Netherlands, 2018*	THA + TKA *Combined*	120106	61	2007 to 2012	Administrative	Routinely	2007: 88% 2012: 76%	Readmission LOS
Pross(27) *Germany, 2017*	THA	---	551 to 1117	2006 to 2014	Clinical	Clinician-reported	35%	Revision Complications
Calderwood(44) *USA, 2013*	THA	524892	3296	2005 to 2007	Administrative	Routinely	≥95%	Complications
Kurtz(38) *USA, 2016*	TKA	952593	3848	2010 to 2013	Administrative	Routinely	---	Readmission

Appendix IIA Continued

Authors	Arthroplasty	Number of patients	Number of hospitals	Time period	Data category	Data collection method	Data completeness	Indicator(s) reported
Kurtz(37) *USA, 2016*	THA	442333	3730	2010 to 2013	Administrative	Routinely	---	Readmission
Jergesen(48) *USA, 2016*	THA&TKA *Separately*	THA: 272 TKA: 261	THA: 2 TKA: 2	2008 to 2012	Administrative	Routinely	---	Complications
Makela(32) *USA, 2011*	THA&TKA *Separately*	Total: 12187	THA: 60 TKA: 60	2005 to 2007	Administrative	Clinician-reported	---	Revision
Mittal(28) *USA, 2018*	THA + TKA *Combined*	---	2461	2012 to 2015	Administrative	Routinely	---	Readmission
Skufca(45) *Finland, 2017*	THA&TKA *Separately*	THA: 73227 TKA: 56860	THA: 17 TKA: 18	1999 to 2014	Clinical	Clinician-reported	---	Complications
Asaid(51) *Australia, 2013*	THA + TKA *Combined*	THA: 938 TKA: 658	2	June 28, 2002 to September 3, 2007	Clinical	Routinely	---	Complications
Dailey(46) *Australia, 2009*	THA&TKA *Separately*	THA: 4131 TKA: 4858	THA: 10 TKA: 9	July 1, 2005 to June 30, 2007	Clinical	Clinician-reported	THA: 83% TKA: 78%	Complications
Martino(53) *USA, 2018*	THA&TKA *Separately*	THA: 495 TKA: 796	THA: 2 TKA: 2	June 2013 to December 2014	Clinical	Routinely	---	LOS
Singh(33) *United Kingdom, 2017*	THA + TKA *Combined*	THA: 4915 TKA: 5928	2	April 2003 to October 2014	Clinical	Clinician-reported	---	Revision Complications
Husted(29) *Denmark, 2010*	THA&TKA *Separately*	---	THA: 48 TKA: 43	2004	Clinical	Clinician-reported	---	LOS
Lopez-Contreras(47) *Spain, 2012*	THA&TKA *Separately*	THA: 7804 TKA: 16781	THA: 49 TKA: 51	2007 to 2009	Clinical	Clinician-reported	---	Complications

Appendix IIB Registry reports details

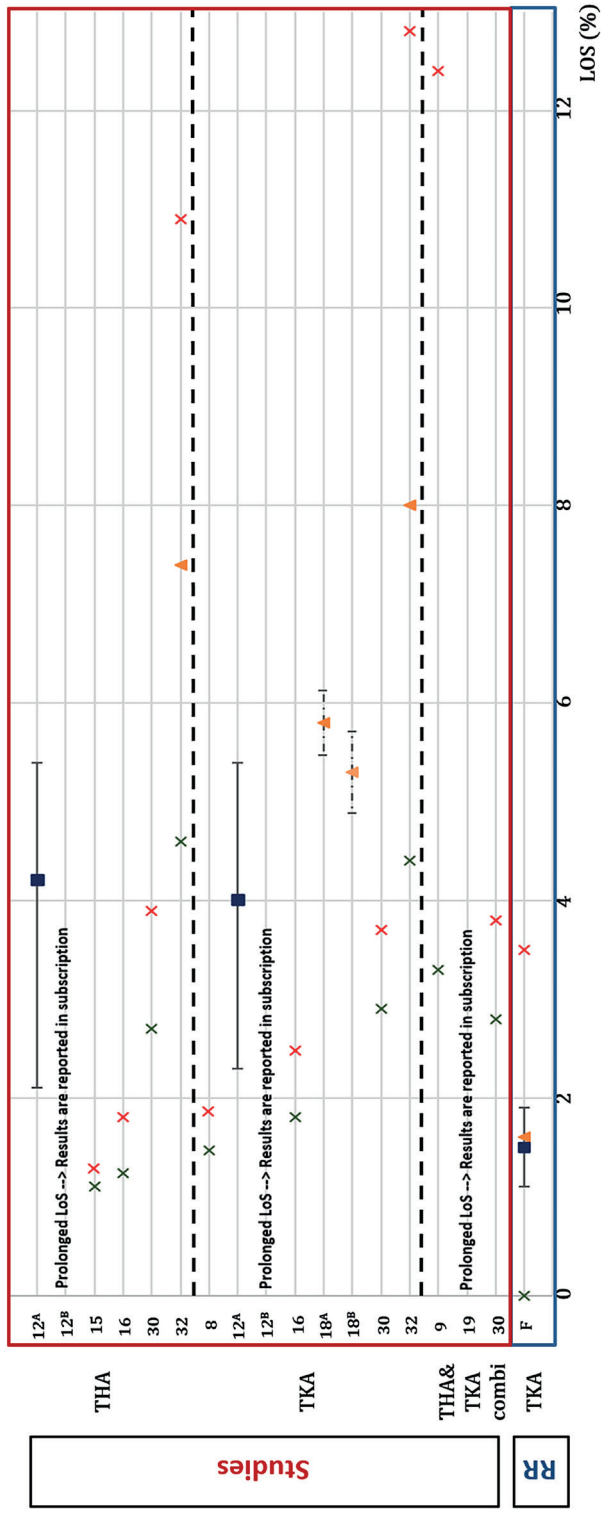
Register reports	Arthroplasty	Number of patients	Number of hospitals	Completeness of data		Indicator(s) reported
				Overall	Individual hospitals	
Norwegian Arthroplasty Registry(1) *Norway, 2020* *Founded in 1987*	THA & TKA	THA (2019): 9879 TKA (2019): 7161	THA: 57 TKA: 58	THA (2017-2018): 97.5% THA revision: 93.1% TKA (2017-2018): 97.6% TKA revision: 93.2%	Yes	Revision
Dutch Arthroplasty Registry(2) *The Netherlands, 2020* *Founded in 2007*	THA & TKA	THA (2019): 33248 TKA (2019): 25859	THA: 95 TKA: 91	THA (2018): 99% THA revision (2018): 97% TKA (2018): 99% TKA revision (2018): 97%	No	Revision
Swedish Knee Arthroplasty Registry(3) *Sweden, 2020* *Founded in 1975*	TKA	2019: 14977	72	2018: 97.1%	Yes	Revision Complications Mortality
Danish Hip Arthroplasty Registry(4) *Denmark, 2020* *Founded in 1995*	THA	2019: 11193	47	2019: 96.5% Revision (2019): 86.2%	Yes	Revision Readmission Complications
Swedish Hip Arthroplasty Registry(5) *Sweden, 2019* *Founded in 1979*	THA	2018: 18629	81	2017: 98%	Yes	Revision Complications Mortality
Danish Knee Arthroplasty Registry(6) *Denmark, 2020* *Founded in 1997*	TKA	2019: 11124	39	2019: 95.7% Revision (2019): 89.7%.	Yes	Revision Readmission LOS

Appendix IIB Continued

Register reports	Arthroplasty	Number of patients	Number of hospitals	Completeness of data		Indicator(s) reported
				Overall	Individual hospitals	
Finnish Arthroplasty Registry(7) *Finland, 2020* *Founded in 1980*	THA & TKA	THA (2019): 10495 TKA (2019): 13460	THA: 51 TKA: 51	THA (2019): 94.9% TKA (2019): 97.1%	Yes	Revision
Swiss Arthroplasty Registry(8) *Switzerland, 2020* *Founded in 2012*	THA & TKA	THA (2019): 19897 TKA (2019): 15378	THA: 152 TKA: 148	THA and TKA combined (2019): >95%	Yes	Revision

LOS=length-of-stay; THA=total hip arthroplasty; TKA=total knee arthroplasty; USA=United States of America.

Appendix III

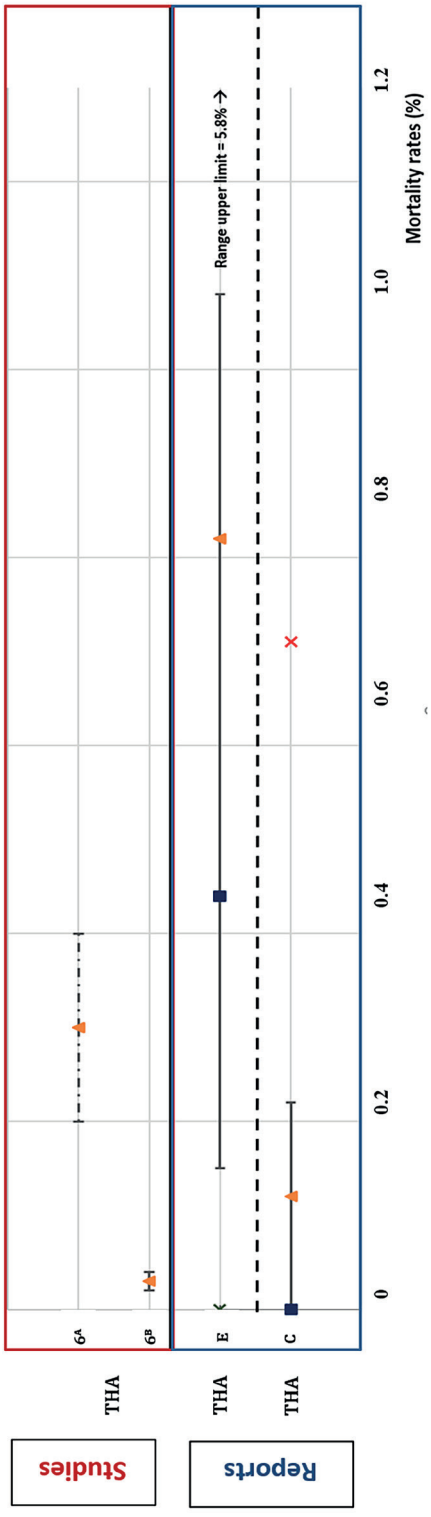


Appendix IV Measurements to report between-hospital variation for length-of-stay

Length-of-stay	Studies (*n*=9)			Registry reports (*n*=1)
	THA (*n*=6) [12[A],12[B],15, 16,30,32]	**TKA (*n*=8)** [8,12[A],12[B],16,18[A], 18[B],30,32]	**THA&TKA (*n*=3)** [9,19,30]	**TKA (*n*=1)** [F]
1a) Outcome definition I				
Nights spent in hospital	[15,30]	[8,30]	[30]	---
Post-operative stay	[12[A],12[B]]	[12[A],12[B],18[A],18[B]]	---	[F]
Total LOS	[16]	[16]	---	---
+LOS rehabilitation centre	[32]	[32]	---	---
Unclear	---	---	[9,19]	---
1b) Outcome definition II				
LOS (days)	[12[A],15,16,30,32]	[8,12[A],16-32]	[9,30]	[F]
Extended LOS (%)	[12[B]]	[12[B]]	[19]	---
2) Case-mix adjusted				
Yes	[12[B]]	[12[B]]	---	---
No	[12[A],15,16,30,32]	[8,12[A],16-32]	[9-30]	[F]
3a) Type of patient selected				
Age selection(s)	[12[A],12[B],30]	[12[A],12[B],30]	[30]	---
Osteoarthritis	---	---	[19]	---
No trauma patients	[15]	[8]	---	---
Medicare patients	---	[18[A],18[B]]	---	---
Elective surgery	[12[A],12[B]]	[12[A],12[B]]	---	---
Matching of patient groups	[15,16]	[8,16]	---	---
No selections	32	[32]	[9]	[F]
3b) Type of hospitals selected				
Academic hospitals	---	---	[9]	---
Non-academic hospitals	[12[A],12[B]]	[12[A],12[B]]	---	---
Critical pathway hospitals	---	[18[A]]	---	---
Non-critical pathway hospitals	---	[18[B]]	---	---
Unclear	[32]	[32]	---	---
No selections	[15,16,30]	[8,16,30]	[19,30]	[F]

The definitions for length-of-stay were defined for 5 domains. The numbers in brackets correspond to the study numbers from Table 1A and the letters in brackets correspond to the report codes from Table 1B.
LOS=length-of-stay ; THA=total hip arthroplasty; TKA=total knee arthroplasty.

Appendix V


Studies
Reports
THA
THA
THA
6A
6B
E
C
Range upper limit = 5.8% →
0
0.2
0.4
0.6
0.8
1.0
1.2
Mortality rates (%)

Appendix VI Measurements to report between-hospital variation for mortality

	Studies (*n*=1)	Registry reports (*n*=2)	
	THA (*n*=2)	THA (*n*=1)	TKA (*n*=1)
Mortality	[6[A],6[B]]	[E]	[C]
1) Outcome definition			
Post-operative	[6[A],6[B]]	[E]	[C]
2) Follow-up			
Within 30 days	[6[A],6[B]]	---	---
Within 90 days	---	[E]	[C]
3) Case-mix adjusted			
Yes	[6[A],6[B]]	---	---
No	---	[E]	[C]
4a) Type of patient selected			
Age	[6[A],6[B]]	---	---
Medicare patients	[6[A],6[B]]	---	---
Multiple selections were made	---	[E]	---
4b) Type of hospitals selected			
Honor roll hospitals	[6[A]]	---	---
Affiliated honor roll hospitals	[6[B]]	---	---

The definitions for mortality were defined for 5 domains. The numbers in brackets correspond to the study numbers from Table 1A, and the letters in brackets correspond to the report codes from Table 1B.

Chapter 3

Between-Hospital Variation in Revision Rates After Total Hip and Knee Arthroplasty in the Netherlands

Directing Quality-Improvement Initiatives

Peter van Schie[1,2], Liza N. van Steenbergen[3], Leti van Bodegom-Vos[2], Rob G.H.H. Nelissen[1], Perla J. Marang- van de Mheen[2]

1. Department of Orthopaedics, Leiden University Medical Centre, Leiden, Netherlands;
2. Department of Biomedical Data Sciences, Medical Decision Making, Leiden University Medical Centre, Leiden, Netherlands;
3. Dutch Arthroplasty Register (LROI), 's-Hertogenbosch, Netherlands.

J Bone Joint Surg Am. 2020 Feb 19;102(4):315-324

Abstract

Background

Variation in 1-year revision rates between Dutch hospitals after primary total hip and knee arthroplasty (THA and TKA) may direct quality-improvement initiatives if this variation accurately reflects true hospital differences. The aim of the present study was to assess the extent of variation, both overall and for specific indications, as well as the statistical reliability of ranking hospitals.

Methods

All primary THAs and TKAs that were performed between January 2014 and December 2016 were included. Observed/expected (O/E) ratios regarding 1-year revision rates were depicted in a funnel plot with 95% control limits to identify outliers based on 1 or 3 years of data, both overall and by specific indication for revision. The expected number was calculated on the basis of patient mix with use of logistic regression models. The statistical reliability of ranking hospitals (rankability) on these outcomes indicates the percentage of total variation that is explained by "true" hospital differences rather than chance. Rankability was evaluated using fixed and random effects models, for overall revisions and specific indications for revision, including 1 versus 3 years of data.

Results

The present study included 86,468 THAs and 73,077 TKAs from 97 and 98 hospitals, respectively. Thirteen hospitals performing THAs were identified as negative outliers (median O/E ratio, 1.9; interquartile range [IQR], 1.5-2.5), with 5 hospitals as outliers in multiple years. Eight negative outliers were identified for periprosthetic joint infection; 4, for dislocation; and 2, for prosthesis loosening. Seven hospitals performing TKAs were identified as negative outliers (median O/E-ratio, 2.3; IQR, 2.2-2.8), with 2 hospitals as outliers in multiple years. Two negative outlier hospitals were identified for periprosthetic joint infection and 1 was identified for technical failures. The rankability for overall revisions was 62% (moderate) for THA and 46% (low) for TKA.

Conclusion

There was large between-hospital variation in 1-year revision rates after primary THA and TKA. For most outlier hospitals, a specific indication for revision could be identified as contributing to worse performance, particularly for THA; these findings are starting points for quality-improvement initiatives.

Introduction

Recent studies showed variation in hospital performance after total hip arthroplasty (THA) and total knee arthroplasty (TKA) in terms of outcomes such as length of stay, readmission, complications, allogeneic blood transfusion, and mortality as well as costs, suggesting that improvement is possible for some hospitals(1-10). For instance, a fivefold-higher postoperative complication rate was found in some United States hospitals compared with others, which may be due to patient characteristics, true hospital differences, and random variation(2,9). The statistical reliability of ranking (rankability) indicates the percentage of variation that is due to "true" hospital differences as opposed to random (chance) variation(11-16). Random variation is particularly likely if the number of events is small(17). Therefore, rankability adds crucial information to calculated between-hospital variation in clinical outcomes. Conclusions about the quality of care delivered are only reliable if an outcome is ranked reliably.

Most arthroplasty registries publish annual reports with the aim of monitoring hospital performance and providing hospitals with feedback. In orthopaedics, the 1-year revision rate is an important clinical outcome that is often used to monitor the quality of care delivered. The consequences of revision—both for the patient and in terms of associated costs—are considerable. Because of low revision rates, several years of outcomes are usually combined to have more events(18-20). As a result, it may take longer before deteriorating hospital performance is noticed. However, if the rankability for single years of data analysis is acceptable, a reliable earlier signal can be given when performance deteriorates. Furthermore, variation in 1-year revision rates for specific indications (infection, dislocation, etc.) may direct quality-improvement initiatives for worse-performing hospitals. The variation in 1-year revision rates among hospitals is currently unknown, as are the variations for specific indications and the statistical reliability of ranking on revisions.

The aims of the present investigation were (1) to assess the extent of variation in 1-year revision rates between Dutch hospitals after primary THA and TKA procedures in the period 2014-2016 and (2) to estimate rankability to determine the extent to which our findings represent true hospital differences. In both cases, we evaluated both 3 years of procedures as well as single years for both overall revision and specific indications for revision.

Materials and Methods

Study design and setting

This observational study used routinely collected data from the nationwide Dutch Arthroplasty Register (LROI). The LROI was established in 2007, and by 2012 all Dutch hospitals were participating. For every arthroplasty procedure, the product number of the implanted component is registered to identify the prosthesis, as well as the encrypted social security number of each patient so that it is possible to connect a revision procedure to the hospital where the primary procedure took place. Surgeons provide information about the procedure, patient characteristics, and any possible revisions (including the type of, and indication for, revision). The vital status of patients (dead or alive) is obtained from Vektis, the Dutch insurance health-care database. The LROI uses the opt-out system for informed consent, whereby patients must actively object in order not to be included. The completeness of data on primary THA and TKA procedures is checked against hospital electronic health records and currently exceeds 98% for primary procedures and 96% for revisions, meaning that >98% of primary procedures performed are included in the register(21,22). THAs were performed at 97 hospitals, and TKAs were performed at 98.

Patients and outcomes

Anonymous data on all Dutch patients undergoing a primary THA or TKA procedure between January 2014 and December 2016 were included. The following patient characteristics that may influence the need for revision were recorded: age, sex, body mass index (BMI), smoking status (yes or no), American Society of Anaesthesiologists (ASA) classification (I, II, III-IV), Charnley score (A, B1, B2, C, not applicable), and diagnosis (osteoarthritis or non-osteoarthritis)(23). Revision within 1 year (yes or no) was the primary outcome measure (defined as exchange, removal, or addition of any component). When a revision was performed, surgeons registered 1 of the available indications for revision:

- THA: wear (cup and/or insert), dislocation, prosthesis removal, prosthesis loosening (femur and/or acetabulum), periarticular ossification, symptomatic metal-on-metal bearing;

- TKA: patellar pain, wear (modular tibial polyethylene insert), patellar dislocation, malalignment, instability, prosthesis loosening (femur, tibia, and/or patella), progressive patellofemoral osteoarthritis, arthrofibrosis;

- THA and TKA: periprosthetic joint infection, periprosthetic fracture, “other” indication.

To direct quality-improvement initiatives if a hospital performed more revisions than expected given its patient mix, the indications for revision were categorized into the following groups:

- THA: infection, loosening (acetabular and/or femur), and dislocation;
- TKA: infection, loosening (femur, tibia, and/or patella), and technical failure (malalignment, instability, and/or patellar dislocation).

Statistical analysis

First, we estimated the between-hospital variation in 1-year revision rates after primary THA and TKA procedures, with adjustment for differences in patient mix (using the same method as the LROI). For each patient, the expected revision risk was calculated with use of logistic regression analysis, with all of the patient characteristics listed above as independent variables and 1-year revision the dependent variable. Missing patient characteristic values (<10% for all variables) were imputed with the mean for numeric variables or the mode for categorical variables (so that the most frequently occurring category was imputed). All expected revision risks at the patient level were summed within a hospital to obtain the aggregated expected number (E) of revisions per hospital. The observed number (O) of revisions per hospital was then divided by the expected number to calculate an O/E ratio for each hospital.

The O/E ratios were depicted in a funnel plot with 95% control limits, including 3 years of procedures (2014-2016) and single years(24). Hospitals outside these limits either have significantly lower revision rates than expected and thus perform better (positive outlier, represented by green dots on the funnel plot) or have significantly higher revision rates and perform worse (negative outlier, represented by red dots on the funnel plot). Hospitals within these limits do not perform different than expected. Feedback based on a single year of performance data may be given sooner, but 3-year data may have better power. The analyses were repeated for specific indications for revision as these indications might point to starting points for quality improvement.

Statistical reliability of ranking (Rankability)

Rankability was introduced in previous research and refers to the statistical reliability of ranking hospitals(11-13). Rankability reflects a signal-to-noise ratio and is expressed as the percentage of the hospital variation being due to 'true' hospital differences rather than chance variation with use of the following equation(11-14):

$$\text{Rankability} = \frac{\text{Between-hospital variation}(\tau^2)}{\text{Between-hospital variation}(\tau^2) + \text{Within-hospital variation}(\sigma^2)} X\ 100\%$$

Between-hospital variation was estimated with use of a random effects logistic regression model to adjust for clustering of patients within hospitals, with the hospital variable as a random factor and patient characteristics as fixed factors. Within-hospital variation was estimated with use of a fixed effect logistic regression model, with all patient characteristics as fixed factors to adjust for patient mix and hospital as a categorical variable. The median squared standard error (SE) of the hospital estimate was taken to represent the within-hospital variance(15). The between and within-hospital variation from these models are included in the equation above to evaluate the extent to which hospital variation can be attributed to "true" hospital differences. If the within-hospital variation is relatively large, for example, as a result of low-frequency outcomes, it will become harder to detect any between-hospital differences. Rankability was classified as low (<50%), moderate (50% to 75%), or high (>75%), as previously suggested(15). Rankability was determined for THA and TKA separately for both 3 years and single years of data as well as by the indication for revision. If hospitals are ranked reliably over a shorter period of time (e.g., single years), deteriorating hospital performance is identified sooner and hospitals do not have to wait several years for data, potentially preventing more failures from occurring(16).

The analyses on hospital variations were performed with use of SPSS (version 23; IBM). Analyses on rankability were performed with use of STATA (version 14; StataCorp). The LUMC Medical Ethical Committee waived the need for ethical approval under Dutch law.

Results

In total, 86,468 primary THA procedures from 97 hospitals and 73,077 primary TKA procedures from 98 hospitals were included. Less than 4% of patient mix variables were missing, except for smoking, which was below 10%. The average 1-year revision rate across all patients was 1.8% for THA and 1.2% for TKA, with a median rate per hospital of 1.6% (interquartile range [IQR], 1.0%-2.3%) for THA and 1.1% (IQR, 0.7%-1.6%) for TKA (Table I).

Table I Distribution of Patient Characteristics and 1-Year Revisions in Dutch Hospitals in the Period 2014-2016.

	THA (*n*=97 hospitals)		TKA (*n*=98 hospitals)	
	Median (IQR)	**Range**	**Median (IQR)**	**Range**
Procedures (*n*)	759 (526-1173)	2-2502	699 (463-938)	9-1998
Mean age (years)	69.3 (67.8-70.1)	50.6-71.8	68.8 (67.4-69.7)	56.5-72.2
Gender, female (%)	66.1 (63.3-68.0)	0.0-74.1	65.2 (61.9-67.8)	8.3-100.0
Mean BMI (kg/m^2)	27.3 (27.0-27.8)	25.9-28.6	29.8 (29.3-30.4)	20.5-31.0
Smoking (%)	13.2 (10.7-15.2)	0.0-27.9	9.8 (8.4-11.8)	1.0-20.5
ASA classification (%)				
• ASA I	17.4 (14.2-21.4)	3.3-100	11.8 (9.8-16.0)	3.8-54.5
• ASA II	65.0 (59.8-70.4)	0.0-96.7	68.7 (63.7-73.6)	42.5-91.6
• ASA III-IV	15.6 (11.5-20.4)	0.0-40.1	16.6 (10.8-21.8)	0.0-50.6
Charnley score (%)				
• A	49.3 (43.7-53.9)	23.7-78.2	45.3 (35.6-52.4)	13.1-100.0
• B1	27.8 (22.9-33.4)	3.6-50.7	33.0 (27.3-40.3)	0.0- 57.8
• B2	20.1 (18.1-22.9)	4.7-28.3	19.4 (16.2-21.5)	0.0-28.0
• C	1.9 (1.0-3.3)	0.0-12.2	2.3 (1.1-4.2)	0.0-17.4
Diagnosis (%)				
• OA	87.1 (83.5-90.8)	42.2-100.0	96.6 (95.5-97.9)	58.6-100.0
• Non-OA	12.9 (9.3-16.5)	0.0-57.8	3.4 (2.1-4.5)	0.0-41.4
1-year revision (%)	1.6 (1.0-2.3)	0.0-7.0	1.1 (0.7-1.6)	0.0-16.7
• Infection*	0.3 (0.1-0.8)	0.0-4.5	0.3 (0.1-0.6)	0.0-8.3
• Loosening**	0.3 (0.1-0.4)	0.0-1.5	0.1 (0.0-0.2)	0.0-1.0
• Dislocation***	0.5 (0.2-0.8)	0.0-2.5		
• Technical failure****			0.3 (0.1-0.5)	0.0-8.3

The values under ''Median (IQR)" indicate the mean or the percentage of the median hospital. The value under ''Range" indicate the highest or lowest mean or percentage of the hospitals. ASA = American Society of Anaesthesiologists, BMI = body mass index, IQR = Interquartile range, OA = osteoarthritis, THA = total hip arthroplasty and TKA = total knee arthroplasty.

*Revision within 1 year because of infection. **Revision within 1 year because of prosthesis loosening; following THA (acetabulum and/or femur or TKA (femur, tibia and/or patella). ***Revision within 1 year because of dislocation following THA only. ****Revision within 1 year because of technical failure (malalignment, instability and/or patella dislocation) following TKA only.

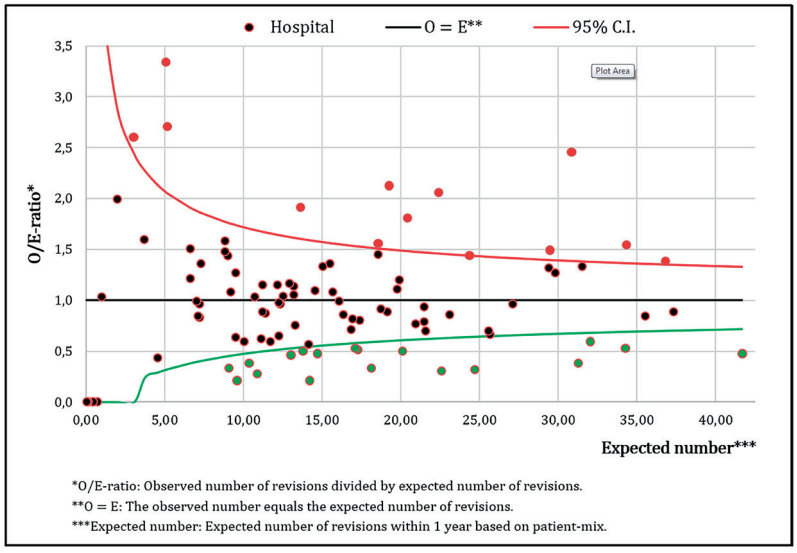


Figure 1 Funnel-plot of Hospital Variation in 1-year Revisions after THA during 2014-2016.

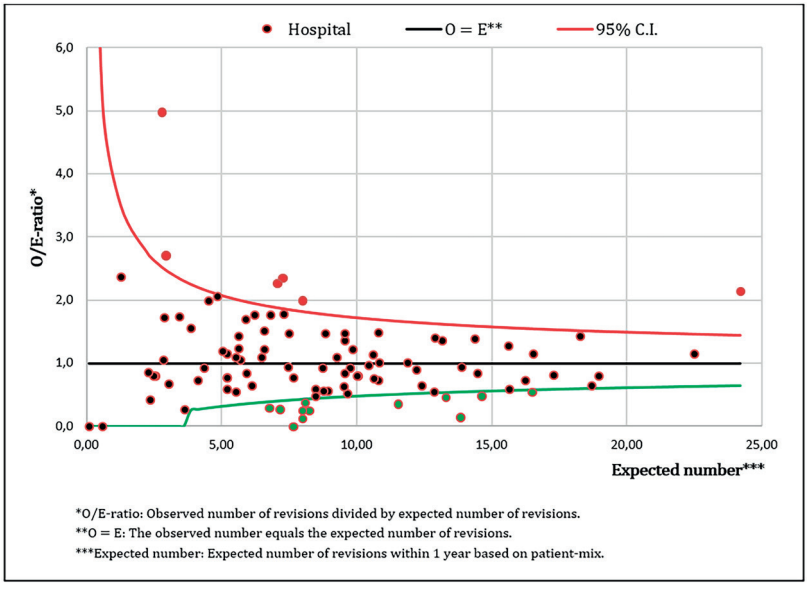


Figure 2 Funnel-plot of Hospital Variation in 1-year Revisions after TKA during 2014-2016.

Hospital variation

Thirteen hospitals performing THA were negative outliers (median O/E ratio, 1.9; IQR, 1.5-2.5), and 18 hospitals were positive outliers (median O/E ratio, 0.4; IQR, 0.3-0.5) (Fig. 1). Seven hospitals performing TKA were negative outliers (median O/E ratio, 2.3; IQR, 2.2-2.8), and 14 hospitals were positive outliers (median O/E ratio, 0.3; IQR, 0.2-0.5) (Fig. 2). Of the 13 negative outliers for THA, 1 hospital was an outlier in all 3 years, 4 were outliers in 2 years, 6 were outliers in 1 year, and 2 were not an outlier in any year (Table II). Of the 7 negative outliers for TKA, 2 hospitals were an outlier in 2 years, 4 hospitals were an outlier in 1 year, and 1 hospital was not an outlier in any single year (Table III). Some outliers in the 3-year period were not an outlier in any single year because only slightly more revisions were performed than expected, causing the difference to become significant only when a lager sample size was available (e.g., hospital 4) (Table II) or because the low volume of procedures at that hospital resulted in wider funnel-plot control limits reflecting the smaller sample size (e.g., hospital 90) (Table II).

Table II Outlier Hospitals with Significantly More Revisions Than Expected Within 1 Year After THA During 2014-2016 According to Year.

	THA outliers (*n*=13 hospitals)			
Hospital	**2014-2016 O/E**	**2014 (O/E)**	**2015 (O/E)**	**2016 (O/E)**
4	1.4			
6	1.5		2.1	1.8
9	2.5	2.2	2.6	2.6
13	1.5	1.8		
14	1.4		2.3	
21	2.1	3.1	1.9	
28	1.8	2.3		1.8
33	2.1		2.1	2.6
37	1.6		2.3	
52	1.9	2.5		
87	2.7			4.9
88	3.3		5.2	
90	2.6			
Median (IQR) for negative outliers	1.9 (1.5-2.5)	2.3 (2.2-2.5)	2.3 (2.1-2.5)	2.6 (1.8-2.6)
Median (IQR) for non-negative outliers	0.9 (0.5-1.1)	0.8 (0.5-1.2)	0.8 (0.5-1.2)	0.8 (0.4-1.4)

*An O/E-ratio is only provided for negative outliers during the 3-year and 1-year periods. IQR = Interquartile range, O/E = observed number of revisions within 1 year divided by the expected number of revisions within 1 year based on patient-mix, and THA = total hip arthroplasty.

Table III Outlier Hospitals with Significantly More Revisions Than Expected Within 1 Year After TKA During 2014-2016 According to Year.

	TKA outliers (*n*=7 hospitals)			
Hospital	**2014-2016**	**2014**	**2015**	**2016**
	O/E	**(O/E)**	**(O/E)**	**(O/E)**
9	2.2		2.4	2.4
35	2.3	3.3		
39	2.0			
41	2.3			4.0
87	2.8		8.9	3.8
89	2.7			4.7
95	13.3	43.9		
Median (IQR) for negative outliers	2.3 (2.2-2.8)	23.6 (13.5-33.8)	5.7 (4.0-7.3)	3.9 (3.5-4.2)
Median (IQR) for non-negative outliers	0.8 (0.6-1.2)	0.9(0.4-1.6)	0.9 (0.3-1.3)	0.9 (0.5-1.2)

An O/E-ratio is only provided for negative outliers during the 3-year and 1-year periods. IQR = Interquartile range, O/E = observed number of revisions within 1 year divided by the expected number of revisions within 1 year based on patient-mix, and TKA = total knee arthroplasty.

Of the 13 negative outliers for THA, 8 hospitals had more revisions for infection (with the cup and/or stem being replaced in 13% of cases), 4 had more revisions for dislocation, and 2 had more revisions for prosthesis loosening. Two hospitals had more revisions for both infection and dislocation. For 1 negative outlier, no specific indication was found. Four hospitals had more revisions for infection in multiple years, and 1 hospital had more revisions for dislocation in multiple years (Table IV). Of the 7 negative outliers for TKA, 2 hospitals had more revisions for infection (with the tibial and/or femoral component being replaced in 13% of cases), with 1 hospital having more infections in all single years. Furthermore, 1 hospital had more technical failures (Table V).

Statistical reliability of ranking (Rankability)

Rankability for overall revision during 2014 to 2016 was 62% (moderate) for THAs and 46% (low) for TKAs, indicating that 62% of the observed variation for THAs and 46% for TKAs reflect "true" hospital differences rather than random variation. For THA, the 3-year indication-specific rankabilities were 61% (moderate) for infection, 39% (low) for dislocation, and 32% (low) for loosening. Rankabilities in single years were low (Table VI). For TKA, these values were 43% (low) for infection, 14% (low) for technical failures, and 11% (low) for loosening. Rankabilities for single years were low (Table VI).

Table IV Outlier Hospitals with Significantly More Revisions Than Expected Within 1 Year After THA During 2014-2016 According to Reason for Revision.

	THA outliers (*n*=13 hospitals)												
Hospital	**All revisions 2014-2016 O/E**	**Revisions for infection (O/E)**				**Revisions for loosening (O/E)**				**Revisions for dislocation (O/E)**			
		2014-2016	**2014**	**2015**	**2016**	**2014-2016**	**2014**	**2015**	**2016**	**2014-2016**	**2014**	**2015**	**2016**
4	1.4					2.4	3.1						
6	1.5	2.5		3.2	3.1								
9	2.5	3.0		3.6	3.7					2.5	3.6		
13	1.5									2.0		2.7	
14	1.4	2.2		3.7									
21	2.1	3.2	4.6	3.3	2.3					2.3	4.6		
28	1.8												
33	2.1	4.5	4.3	3.9	5.1								
37	1.6					4.0		6.7					
52	1.9									3.5	3.9	3.9	
87	2.7	1.6			5.3								
88	3.3	7.3		14.2									
90	2.6	5.1											
Median (IQR) for negative outlier	1.9 (1.5-2.5)	3.1 (2.4-4.7)	4.5 (4.4-4.5)	3.7 (3.4-3.9)	3.7 (3.1-5.1)	3.2 (2.8-3.6)	N/A	N/A	N/A	2.4 (2.2-2.8)	3.9 (3.8-4.3)	3.3 (3.0-3.6)	N/A
Median (IQR) for non-negative outlier	0.9 (0.5-1.1)	0.5 (0.2-1.1)	0.0 (0.0-1.0	0.5 (0.0-1.1)	0.6 (0.0-1.4)	0.8 (0.3-1.2)	0.7 (0.0-1.6)	0.7 (0.0-1.4)	0.5 (0.0-1.6)	0.8 (0.4-1.3)	0.6 (0.0-1.2)	0.8 (0.0-1.3)	0.8 (0.0-1.5)

An O/E-ratios is provided if a hospital was a negative outlier for 1 of the reasons for revision in the period 2014-2016. IQR = Interquartile range, O/E = Observed number of revisions within 1 year divided by the expected number of revisions within 1 year based on patient-mix, THA = total hip arthroplasty, and N/A = not applicable.

Table V Outlier Hospitals with Significantly More Revisions Than Expected Within 1 Year After TKA During 2014-2016 According to Reason for Revision.

TKA outliers (*n*=7 hospitals)													
Hospital	**All revisions 2014-2016 O/E**	**Revisions for infection (O/E)**				**Revisions for loosening (O/E)**				**Revisions for technical failure (O/E)**			
		2014-2016	**2014**	**2015**	**2016**	**2014-2016**	**2014**	**2015**	**2016**	**2014-2016**	**2014**	**2015**	**2016**
9	2.2	3.2	3.5	2.9	3.4								
35	2.3												
39	2.0												
41	2.3	4.4			7.8								
87	2.8									8.5		17.0	
89	2.7												
95	13.3												
Median (IQR) for negative outliers	2.3 (2.3-2.8)	3.8 (3.5-4.1)	N/A	N/A	5.6 (4.5-6.7)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Median (IQR) for non-negative outliers	0.8 (0.6-1.2)	0.8 (0.3-1.4)	0.0 (0.0-1.6)	0.7 (0.0-1.6)	0.6 (0.0-1.3)	0.7 (0.0-1.5)	0.0 (0.0-1.4)	0.0 (0.0-0.0)	0.0 (0.0-1.6)	0.9 (0.4-1.4)	0.0 (0.0-1.6)	0.6 (0.0-1.5)	0.8 (0.0-1.6)

An O/E-ratios is provided if a hospital was a negative outlier for 1 of the reasons for revision in the period 2014-2016. IQR = Interquartile range, O/E = Observed number of revisions within 1 year divided by the expected number of revisions within 1 year based on patient-mix, TKA = total knee arthroplasty, and N/A = not applicable.

Table VI Rankabilities of 1-Year Revision Rates per Reason for Revision during 2014-2016 According to Year.

	1-year reason for revisions	2014-2016	2014	2015	2016
	Infection	61%	34%	36%	31%
THA	Prosthesis loosening	32%	23%	28%	23%
	Dislocation	39%	29%	18%	15%
	Infection	43%	28%	19%	39%
TKA	Prosthesis loosening	11%	15%	25%	8%
	Technical failure	14%	6%	10%	13%

THA = total hip arthroplasty, TKA = total knee arthroplasty.

Discussion

The present study demonstrated large variation in overall 1-year revision rates as well as indication-specific 1-year revision rates between Dutch hospitals after primary THA and TKA. For THA, 13 hospitals performed significantly worse than expected and 18 hospitals performed significantly better than expected during 2014-2016. For TKA, these values were 7 and 14 hospitals, respectively. Eleven of the 13 negative outliers for THA and 6 of the 7 outliers for TKA were an outlier for 1 or more single years. When 3 years of data on specific indications were included, worse performance was identified for 12 of the 13 outlier hospitals for THA and 3 of the 7 outlier hospitals for TKA, with the specific indications consisting mainly of infection (after both THA and TKA) as well as dislocation (after THA). These findings are starting points for further in-depth hospital-based investigations to improve quality of care. When 3 years of data were used, rankability was moderate for THA and low for TKA. For single years of data, rankabilities were all low. When 3 years of data on specific indications were used, rankabilities were low for both THA and TKA, with the exception of infection after THA, for which rankability was moderate. Therefore, we would recommend using 3 years of data to reliably rank hospitals on their revision rates and identify areas for improvement.

Comparison with the literature

Neither hospital variation nor the statistical reliability of ranking hospitals on 1-year revision rates after primary THA and TKA have been described previously, to our knowledge. However, hospital variation for other outcomes has been described. U.S. studies have shown large variation in terms of the rates of complications (range, 1.8% to 9.0%) and blood transfusions (median, 15.9% [IQR, 5.4% to 26.2%] for THA and 11.0% [IQR, 3.5% to 18.5%] for TKA)(2,9). A Dutch study found large variation in allogeneic blood transfusions (O/E-ratio range, 0.0-4.4) but also showed that the reliability of ranking was only 34% for THA and 21% for TKA(6).

Hofstede et al. showed large between-hospital variation in the number of patients with a prolonged length of hospital stay and the number of acute readmissions after THA and TKA, with high and low rankability for these outcomes, respectively (16). For other outcomes and diseases, rankability was mostly moderate at best. Lingsma et al. found a rankability of 55% for a "poor" outcome after acute brain ischemia, and Henneman et al. found a rankability of 35% for mortality after colorectal cancer surgery(11,14). Only for surgical site infection after colectomy was a high rankability previously found(16). The present study demonstrated moderate rankability for revision after THA, which seems in line with previous findings.

Strengths and limitations

A strength of the present study is that a large population-based registry with completeness of >98% was used(21,22). However, low 1-year revision rates make detection of hospital differences difficult. In addition, low-volume hospitals are more difficult to monitor compared with high-volume hospitals because they can escape the outlier status in the funnel plot as a result of the higher uncertainty embedded in their revision estimates. This is reflected in a large range between upper and lower control limits for low-volume hospitals (on the left side of the funnel plot), which means that a higher O/E-ratio is needed to become an outlier compared with high-volume hospitals (on the right side of funnel plot). However, in this cohort, only 6 hospitals performed <25 THAs per year and 3 hospitals performed <25 TKAs per year.

The limitations of the present study are that limited patient variables are collected in registries, causality cannot be proven due to its observational nature, and there may be underreporting of revisions. However, given the high completeness of the LROI database for primary and revision procedures (currently exceeding 98% and 96%, respectively), underreporting is unlikely to affect the results(21,22). Only if a revision were performed abroad would it be missed in the registry. Another limitation may be underreporting of periprosthetic joint infections as incision and drainage is not reported when no prosthesis component is exchanged or removed(25,26). In the Netherlands, the exchange of loose prosthetic components during acute periprosthetic joint infection became standard care in 2015, which likely resulted in some underreporting before 2015(27). Limited data on patient characteristics were available, thereby limiting the possibility of case-mix correction. This factor might have resulted in overestimation of the differences now attributed to quality of care. For example, diabetes mellitus or the use of immunosuppressants influence the revision risk, but data on these factors were not available(28).

Implementation and further research

Determining hospital variation and detecting outliers seem to be a simple and efficient way to get insight into hospital performance, provided that this variation reflects "true" hospital differences with minimal random variation(12,15,29,30). A categorization for rankability is still arbitrary but is suggested to be good when >75%(13,14,31,32). Our results showed moderate rankability for THA for both the overall revision rate (62%) and infection (61%) on the basis of 3 years of data. For single years, the rankabilities were all low, making these outcomes less useful for reliably assessing performance in practice.

In 2017, the Dutch Orthopaedic Association started a procedure to detect possible negative outlier hospitals and to discuss activities to improve care, resulting in a customized plan of improvement(33). The present study used the same outlier procedure, while adding indications for revision as a method to direct quality-improvement initiatives. Most outliers for overall revision were also outliers for a specific indication for revision, so adding these analyses seems a useful addition to direct improvement activities. Although part of the variation may be surgeon-related (e.g., surgical approach), the current approach is likely to provide a broader overview to improve the quality of care given (e.g., individual surgeons are dependent on their team to achieve the best outcomes). Furthermore, having more revisions for infection is likely not surgeon-related but rather is due to antibiotic prophylactic protocol and implementation, wound-care policy, preoperative preparation of the patient with chlorhexidine, and so on.

In the future, other outcome measures associated with quality of care after THA and TKA should be used, such as length of hospital stay, readmissions, and patient-reported outcome measures, provided that these outcomes have sufficient rankability. The use of moving periods of 3-year data over time may be useful for follow-up research because this method reliably examines the outcome over several years. Furthermore, another option might be to combine outcomes in an ordinal composite measure, which would improve rankability and thereby the ability to detect "true" hospital differences while also providing a broader quality-of-care perspective than 1 single indicator(16). In addition, other methods such as statistical process control techniques may enable detection of worsening performance sooner and thereby also improve the quality of care.

Conclusion

Large variation in 1-year revision rates after THA and TKA between Dutch hospitals was identified. The majority of outliers could be linked to a specific indication for revision, which gives clear starting points for quality-improvement initiatives. Earlier detection of worse performance by using a single year of data had low reliability.

References

1. Siciliani L, Sivey P, Street A. Differences in length of stay for hip replacement between public hospitals, specialised treatment centres and private providers: selection or efficiency? Health Econ. 2013;22(2):234-42.
2. Menendez ME, Lu N, Huybrechts KF, Ring D, Barnes CL, Ladha K, et al. Variation in Use of Blood Transfusion in Primary Total Hip and Knee Arthroplasties. J Arthroplasty. 2016;31(12):2757-63.e2.
3. Solomon DH, Losina E, Baron JA, Fossel AH, Guadagnoli E, Lingard EA, et al. Contribution of hospital characteristics to the volume-outcome relationship: dislocation and infection following total hip replacement surgery. Arthritis Rheum. 2002;46(9):2436-44.
4. So-Osman C, Nelissen R, Te Slaa R, Coene L, Brand R, Brand A. A randomized comparison of transfusion triggers in elective orthopaedic surgery using leucocyte-depleted red blood cells. Vox Sang. 2010;98(1):56-64.
5. Voorn VM, Marang-van de Mheen PJ, Wentink MM, So-Osman C, Vliet Vlieland TP, Koopman-van Gemert AW, et al. Frequent use of blood-saving measures in elective orthopaedic surgery: a 2012 Dutch blood management survey.
6. Voorn VMA, Marang-van de Mheen PJ, van der Hout A, So-Osman C, van den Akker-van Marle ME, Koopman-van Gemert A, et al. Hospital variation in allogeneic transfusion and extended length of stay in primary elective hip and knee arthroplasty: a cross-sectional study. BMJ Open. 2017;7(7):e014143.
7. Fry DE, Pine M, Nedza SM, Locke DG, Reband AM, Pine G. Risk-Adjusted Hospital Outcomes in Medicare Total Joint Replacement Surgical Procedures. J Bone Joint Surg Am. 2017;99(1):10-8.
8. Weeks WB, Kotzbauer GR, Weinstein JN. Using Publicly Available Data to Construct a Transparent Measure of Health Care Value: A Method and Initial Results. Milbank Q. 2016;94(2):314-33.
9. Bozic KJ, Grosso LM, Lin Z, Parzynski CS, Suter LG, Krumholz HM, et al. Variation in hospital-level risk-standardized complication rates following elective primary total hip and knee arthroplasty. J Bone Joint Surg Am. 2014;96(8):640-7.
10. Cram P, Vaughan-Sarrazin MS, Wolf B, Katz JN, Rosenthal GE. A comparison of total hip and knee replacement in specialty and general hospitals. J Bone Joint Surg Am. 2007;89(8):1675-84.
11. Henneman D, van Bommel AC, Snijders A, Snijders HS, Tollenaar RA, Wouters MW, et al. Ranking and rankability of hospital postoperative mortality rates in colorectal cancer surgery. Ann Surg. 2014;259(5):844-9.
12. van Dishoeck AM, Koek MB, Steyerberg EW, van Benthem BH, Vos MC, Lingsma HF. Use of surgical-site infection rates to rank hospital performance across several types of surgery. Br J Surg. 2013;100(5):628-36; discussion 37.
13. Lingsma HF, Eijkemans MJ, Steyerberg EW. Incorporating natural variation into IVF clinic league tables: The Expected Rank. BMC Med Res Methodol. 2009;9:53.
14. Lingsma HF, Steyerberg EW, Eijkemans MJ, Dippel DW, Scholte Op Reimer WJ, Van Houwelingen HC. Comparing and ranking hospitals based on outcome: results from The Netherlands Stroke Survey. QJM. 2010;103(2):99-108.
15. van Dishoeck AM, Lingsma HF, Mackenbach JP, Steyerberg EW. Random variation and rankability of hospitals using outcome indicators. BMJ quality & safety. 2011;20(10):869-74.
16. Hofstede SN, Ceyisakar IE, Lingsma HF, Kringos DS, Marang-van de Mheen PJ. Ranking hospitals: do we gain reliability by using composite rather than individual indicators? BMJ quality & safety. 2018.

17. Dimick JB, Welch HG, Birkmeyer JD. Surgical mortality as an indicator of hospital quality: the problem with small sample size. JAMA. 2004;292(7):847-51.
18. Dutch Arthroplasty Register (LROI). Annual report 2018. www.lroi-report.nl.
19. Swedisch Hip Arthroplasty Register. Annual Report 2017. https://registercentrum.blob.core.windows.net/shpr/r/Eng_Arsrapport_2017_Hoftprotes_final-Syx2fJPhMN.pdf.
20. Swedisch Knee Arthroplasty Register. Annual Report 2018. http://www.myknee.se/en/.
21. van Steenbergen LN, Denissen GA, Spooren A, van Rooden SM, van Oosterhout FJ, Morrenhof JW, et al. More than 95% completeness of reported procedures in the population-based Dutch Arthroplasty Register. Acta Orthop. 2015;86(4):498-505.
22. *LROI website. Completeness of Registering Hospitals and Completeness of Registered Arthroplasties in the LROI Based on the Hospital Information System in 2016,* http://www.lroi-rapportage.nl/data-quality-coverage-and-completeness, accessed febr 2019.
23. Jasper LL, Jones, C. A., Mollins, J., Pohar, S. L., Beaupre, L. A.. Risk factors for revision of total knee arthroplasty: a scoping review. BMC Musculoskelet Disord. 2016;17:182.
24. Rothman KJ, Boice JD. Epidemologic analysis with a programmable calculator. National Institutes of Health. 1979 31-2.
25. Gundtoft PH, Overgaard S, Schonheyder HC, Moller JK, Kjaersgaard-Andersen P, Pedersen AB. The "true" incidence of surgically treated deep prosthetic joint infection after 32,896 primary total hip arthroplasties: a prospective cohort study. Acta Orthop. 2015;86(3):326-34.
26. Veltman ES, Moojen DJF, Nelissen RG, Poolman RW. Antibiotic Prophylaxis and DAIR Treatment in Primary Total Hip and Knee Arthroplasty, A National Survey in The Netherlands. J Bone Jt Infect. 2018;3(1):5-9.
27. Dutch Orthopaedic Association. Protocol: Aanbeveling werkwijze behandeling Prothese Infecties Orthopedie, 2015. www.orthopeden.org.
28. Marchant MH, Jr., Viens NA, Cook C, Vail TP, Bolognesi MP. The impact of glycemic control and diabetes mellitus on perioperative outcomes after total joint arthroplasty. J Bone Joint Surg Am. 2009;91(7):1621-9.
29. Adab P, Rouse AM, Mohammed MA, Marshall T. Performance league tables: the NHS deserves better. BMJ. 2002;324(7329):95-8.
30. Cram P, Cai X, Lu X, Vaughan-Sarrazin MS, Miller BJ. Total knee arthroplasty outcomes in top-ranked and non-top-ranked orthopedic hospitals: an analysis of Medicare administrative data. Mayo Clin Proc. 2012;87(4):341-8.
31. Huedo-Medina TB, Sanchez-Meca J, Marin-Martinez F, Botella J. Assessing heterogeneity in meta-analysis: Q statistic or I2 index? Psychol Methods. 2006;11(2):193-206.
32. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003;327(7414):557-60.
33. CommisionQuality. Commision Quality (Collaboration between Dutch Orthopaedic Association and Dutch Arthroplasty Registry). Protocol: Quality procedure, 2017. www.orthopeden.org.

Chapter 4

Monitoring Hospital Performance with Statistical Process Control After Total Hip and Knee Arthroplasty: A Study to Determine How Much Earlier Worsening Performance Can Be Detected

Peter van Schie[1,2], Leti van Bodegom-Vos[2], Liza N. van Steenbergen[3] Rob G.H.H. Nelissen[1], Perla J. Marang- van de Mheen[2]

1. Department of Orthopaedics, Leiden University Medical Centre, Leiden, Netherlands;
2. Department of Biomedical Data Sciences, Medical Decision Making, Leiden University Medical Centre, Leiden, Netherlands;
3. Dutch Arthroplasty Register (LROI), 's-Hertogenbosch, Netherlands.

J Bone Joint Surg Am. 2020 Dec 2;102(23):2087-2094

Abstract

Background

Given the low early revision rate after total hip arthroplasty (THA) and total knee arthroplasty (TKA), hospital performance is typically compared using 3 years of data. The purpose of this study was to assess how much earlier worsening hospital performance in 1-year revision rates after THA and TKA can be detected.

Methods

All 86,468 THA and 73,077 TKA procedures performed from 2014 to 2016 and recorded in the Dutch Arthroplasty Register were included. Negative outlier hospitals were identified by significantly higher O/E (observed divided by expected) 1-year revision rates in a funnel plot. Monthly Shewhart p-charts (with 2 and 3-sigma control limits) and cumulative sum (CUSUM) charts (with 3.5 and 5 control limits) were constructed to detect a doubling of revisions (odds ratio of 2), generating a signal when the control limit was reached. The median number of months until generation of a first signal for negative outliers and the number of false signals for non-negative outliers were calculated. Sensitivity, specificity and accuracy were calculated for all charts and control limit settings, using outlier status in the funnel plot as the golden standard.

Results

The funnel plot showed that 13 of 97 hospitals had significantly higher O/E 1-year revision rates and were negative outliers for THA and 7 of 98 hospitals had significantly higher O/E 1-year revision rates and were negative outliers for TKA. The Shewhart p-chart with the 3-sigma control limit generated 68 signals (34 false-positive) for THA and 85 signals (63 false-positive) for TKA. The sensitivity for THA and TKA was 92% and 100% respectively; the specificity was 69% and 51%, respectively; and the accuracy was 72% and 54%, respectively. The CUSUM chart with a 5 control limit generated 18 signals (1 false-positive) for THA and 7 (1 false-positive) for TKA. The sensitivity was 85% and 71% for THA and TKA, respectively; the specificity was 99% for both; and accuracy was 97% for both. The Shewhart p-chart with a 3-sigma control limit generated the first signal for negative outliers after a median of 10 months [Interquartile range (IQR):2 to 18] for THA and 13 months [IQR:5 to 18] for TKA. The CUSUM charts with a 5 control limit generated the first signal after a median of 18 months [IQR:7 to 22] for THA and 21 months [IQR:9 to 25] for TKA.

Conclusion

Monthly monitoring using CUSUM charts with a 5 control limit enables earlier detection of worsening 1-year revision rates with accuracy so that initiatives to improve care can start earlier.

Introduction

Most arthroplasty registries publish annual reports including funnel plots for binary clinical outcomes, with the purpose of monitor hospital performance and providing feedback. Funnel plots are graphical tools to compare outcomes with those of other hospitals and detect hospitals performing significantly better or worse in terms of these outcomes. In orthopaedics, the 1-year revision rate is an important performance indicator to monitor quality of hospital care. Consequences of a revision are dramatic for patients and entail considerable costs. However, due to low event rates for 1-year revision as well as for many orthopaedic performance outcomes, multiple years of outcomes are usually combined in funnel plots to obtain detectable and reliable hospital differences.(1-6) Because arthroplasty registries typically combine 3 years of data, it may take a long time before deteriorating performance is noticed, resulting in late action plans to improve care.(3) Thus, more frequent monitoring of clinical endpoints such as 1-year revision rates is needed, as are reliable and earlier signals if outcomes deteriorate.

Statistical Process Control (SPC) charts such as Shewhart p-charts and Cumulative SUM (CUSUM) charts may offer additional information because the performance is plotted more frequently over time (for example, monthly). Several good clinical studies and the focus to improve the quality of care, led to growing interest in these charts.(7-16) SPC-charts with their control limits can distinguish between an "in-control" process, showing only chance variation within control limits, and an "out-of-control" process showing systematic (special-cause) variation and generating a signal (alert) when the control limit is reached.(17) However, with SPC charts there is a trade-off between the number of false positive and the number of false negative signals, determined by the level at which control limits are set. In practice, minimization of the number of false-positive signals in particular is recommended because they may result in alert and improvement fatigue by clinicians.(18,19)

Various SPC charts are available, but there is uncertainty about which chart and control limit to choose.(20,21) In the present study we opted for Shewhart p-charts and CUSUM charts. The Shewhart p-chart is considered to be accessible, especially with regard to implementation and easy interpretation.(22) However, the CUSUM chart has superior performance in detecting small (<10%) and large (>10%) increases in event rates.(13,22-24) These two SPC charts thus seemed logical candidates to test. The authors of a previous orthopaedic study already described CUSUM charts implementation, but did not address how much earlier a signal was generated or its reliability compared with the more commonly used funnel plot, which seems crucial for these techniques to be accepted in routine clinical practice.(25)

The aim of this study was to assess the extent to which Shewhart p-charts and CUSUM charts enable monitoring such that worsening 1-year revision total hip arthroplasty (THA) or total knee arthroplasty (TKA) rates in Dutch hospitals are detected earlier within a timeframe of 3 years, with good sensitivity, specificity and accuracy, compared with the current method of arthroplasty registries using funnel plots.

Methods

Study design

This observational study used routinely collected data from the nationwide Dutch Arthroplasty Register (Landelijke Registratie Orthopedische Implantaten (LROI)).(6) Data completeness in this register is checked against in-hospital patient records and currently exceeds 98% for primary arthroplasties and 96% for revisions.(26,27)

Study population

All Dutch patients who underwent a primary THA or TKA procedure from January 2014 to December 2016 as recorded in the LROI were included. The following patient characteristics were available: age, sex, body mass index (BMI, kg/m^2), smoking (yes or no), American Society of Anaesthesiologists (ASA) classification (I,II,III-IV), Charnley score (A,B1,B2,C, and not applicable) and diagnosis (osteoarthritis or non-osteoarthritis).(28) Revision within one year (yes or no) was the primary outcome measure (defined as replacement, removal or addition of any component).

Statistical analysis

The between-hospital variation in 1-year revision rates after primary THA and TKA during 2014-2016 was estimated, applying the same method as used by the LROI. For each patient the expected revision risk was calculated using logistic regression analysis, including all patient characteristics described above as independent variables and 1-year revision as the dependent variable. Missing patient characteristic values (<10% for all variables) were imputed with the mean for numeric variables or the mode for categorical variables (meaning that the most frequently occurring category was imputed). All expected revision risks were then summed within a hospital to obtain the aggregated expected number (E) of revisions per hospital. The observed numbers (O) divided by expected numbers were depicted in a funnel plot with 95% control limits. Negative outlier hospitals are those outside the upper limit, meaning that they had significantly higher revision rates than expected given their patient-mix. Positive outlier hospitals are those outside the lower limit, meaning that they had significantly lower revision rates.

Second, the extent to which SPC-methods can generate an earlier signal for deteriorating performance within a 3-year time frame was estimated. Risk-adjusted monthly Shewhart p-charts (with 2 and 3-sigma control limits) and risk-adjusted log-likelihood CUSUM charts (with 3.5 and 5 control limit) for 1-year revisions were constructed to detect an odds ratio of 2, for each hospital across 3 years.(22) Figure 1 shows an example of a Shewhart-p-chart, in which the center line indicates the mean hospital performance and the area between both control limits is where variation is considered random (by chance). A value outside control limits is considered a systematic variation and generates a signal. Usually 2 and 3-sigma control limits are used, with the 2-sigma control limit having a higher likelihood of type a 1 error (false-positive signal) and the 3-sigma control limit having a higher likelihood of a type 2 error (false-negative signal). Figure 2 shows an example of a CUSUM chart with 3.5 and 5 control limits. This chart shows the cumulative performance across patients over a period of time. When the chart-statistic reaches the control limit, a signal is generated and the chart resets to zero. Similarly, the control limits are chosen to balance the likelihood of false-positive and false-negative signals, with 3.5 and 5 most commonly used in practice.(22,25) Appendix I gives a more detailed description of the Shewhart p-chart and CUSUM chart.

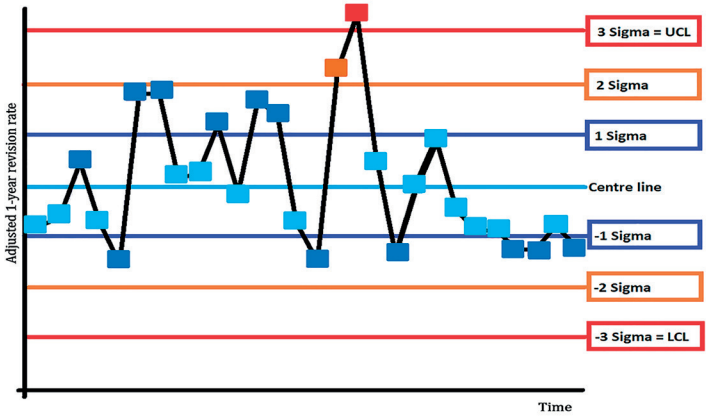


Figure 1 Example of a Shewhart p-chart
See text for explanation of chart.
UCL = upper control limit and LCL = lower control limit.

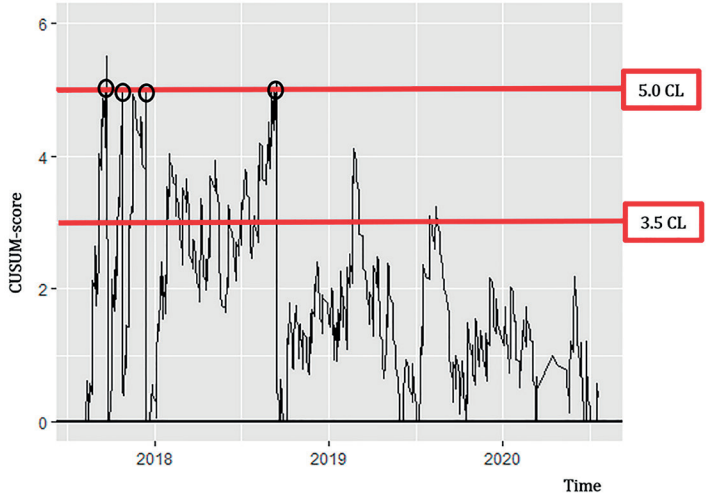


Figure 2 Example of a CUSUM chart
A hypothetical CUSUM chart with 3.5 and 5 control limit (CL). The chart resets to 0 after the 5 control limit is reached. In this example, 4 signals are generated, and the hospitals shows an improvement for this outcome over time. See text for further explanation of chart.

For both charts and control limit settings, we calculated the median number of months needed to generate the first signal for negative outlier hospitals and the number of false signals for other hospitals. Furthermore, we calculated the signals missed for negative outlier hospitals. Additionally, sensitivity, specificity and accuracy for both charts and control limit settings were calculated within the 3-year time frame using the negative outlier status of a hospital in the funnel plots as the ''golden" standard. The accuracy for correctly classifying a hospital was defined as:

$$\frac{\text{(Number of true positive classified hospitals+Number of true negative classified hospitals)}}{\text{(Total number of hospitals)}} \text{ X } 100\%.$$

Analyses were performed using SPSS version 25 (IBM). The LUMC Medical Ethical Committee considered the study exempt for ethical approval under Dutch law (CME, G18.140).

Results

The study included 86,468 primary THA procedures from 97 hospitals and 73,077 primary TKA procedures from 98 hospitals. The rate of missing data was <4% for

all variables, except for smoking (<10%). On the patient-level, the average 1-year revision rate was 1.8% for THA and 1.2% for TKA. On hospital-level, the median revision rate was 1.6% (interquartile range (IQR):1.0 to 2.3) for THA and 1.1% (IQR:0.7 to 1.6) for TKA (Table 1).

Table 1 Distribution of patient characteristics and outcomes in participating hospitals

	THA (*n*=97 hospitals)		TKA (*n*=98 hospitals)	
	Median (IQR)	**Range**	**Median (IQR)**	**Range**
Procedures (*n*)	759 (526-1173)	2-2502	699 (463-938)	9-1998
Mean age (years)	69.3 (67.8-70.1)	50.6-71.8	68.8 (67.4-69.7)	56.5-72.2
Gender, female (%)	66.1 (63.3-68.0)	0.0-74.1	65.2 (61.9-67.8)	8.3-100.0
Mean BMI (kg/m^2)	27.3 (27.0-27.8)	25.9-28.6	29.8 (29.3-30.4)	20.5-31.0
Smoking (%)	13.2 (10.7-15.2)	0.0-27.9	9.8 (8.4-11.8)	1.0-20.5
ASA classification (%)				
• ASA I	17.4 (14.2-21.4)	3.3-100	11.8 (9.8-16.0)	3.8-54.5
• ASA II	65.0 (59.8-70.4)	0.0-96.7	68.7 (63.7-73.6)	42.5-91.6
• ASA III-IV	15.6 (11.5-20.4)	0.0-40.1	16.6 (10.8-21.8)	0.0-50.6
Charnley score* (%)				
• A	49.3 (43.7-53.9)	23.7-78.2	45.3 (35.6-52.4)	13.1-100.0
• B1	27.8 (22.9-33.4)	3.6-50.7	33.0 (27.3-40.3)	0.0- 57.8
• B2	20.1 (18.1-22.9)	4.7-28.3	19.4 (16.2-21.5)	0.0-28.0
• C	1.9 (1.0-3.3)	0.0-12.2	2.3 (1.1-4.2)	0.0-17.4
Diagnosis (%)				
• OA	87.1 (83.5-90.8)	42.2-100.0	96.6 (95.5-97.9)	58.6-100.0
• Non-OA**	12.9 (9.3-16.5)	0.0-57.8	3.4 (2.1-4.5)	0.0-41.4
1-year revisions (%)	1.6 (1.0-2.3)	0.0-7.0	1.1 (0.7-1.6)	0.0-16.7

The values under ''Median (IQR)" indicate the mean or the percentage of the median hospital. The values under ''Range" indicate the highest and lowest means or percentages among the hospitals. *The Charnley score was used to evaluate comorbidity in relation to levels of activity. **All diagnoses except osteoarthritis (fracture, osteonecrosis, rheumatoid arthritis, inflammatory arthritis, etc).

Outlier hospitals

Based on 3-year funnel plots, 13 hospitals performing THA were negative outliers with a median O/E (observed divided by expected) ratio of 1.9 (IQR:1.5 to 2.5) compared with 0.9 (IQR:0.5 to 1.1) for the other hospitals. For TKA, there were 7 negative outliers with a median O/E ratio of 2.3 (IQR:2.3 to 2.8) compared with 0.8 (IQR:0.6 to 1.2) for the other hospitals (Table 2 and Appendices II and III; red dots). Two hospitals were negative outliers for both THA and TKA. Eighteen hospitals were positive outliers for THA with a median O/E ratio of 0.4 (IQR:0.3 to 0.5) and 14 hospitals were positive outliers for TKA with a median O/E ratio of 0.3 (IQR:0.2 to 0.5) (Appendices II and III; green dots).

Table 2 Outlier hospitals with significantly more revisions than expected during 2014-2016

Hospital	Negative outliers	
	THA (*n*=13 hospitals)	TKA (*n*=7 hospitals)
	2014-2016 O/E	2014-2016 O/E
4	1.4	
6	1.5	
9	2.5	2.2
13	1.5	
14	1.4	
21	2.1	
28	1.8	
33	2.1	
37	1.6	
35		2.3
39		2.0
41		2.3
52	1.9	
87	2.7	2.8
88	3.3	
89		2.7
90	2.6	
95		13.3
Median (IQR) negative outliers	1.9 (1.5-2.5)	2.3 (2.3-2.8)
Median (IQR) all other Dutch hospitals	0.9 (0.5-1.1)	0.8 (0.6-1.2)

An O/E ratios is provided only for negative outlier during the 3-year period.

Earlier signals compared with false signals using two SPC methods

I. Shewhart p-chart

For THA, 195 signals of worsening performance were generated for 70 hospitals at the 2-sigma (similar to 2 standard deviation in hypothesis testing) control limit with all 13 negative outlier hospitals alerted, but also 57 hospitals incorrectly alerted (sensitivity 100%, specificity 32%, accuracy 41%). At the 3-sigma control limit, 68 signals were generated for 38 hospitals, with 12 negative outlier hospitals alerted (sensitivity 92%, specificity 69%, accuracy 72%). At 3-sigma, the first signal for negative outliers was generated after a median of 10 months (IQR:2 to 18), which should be considered against 34 false-positive signals for other hospitals. For 1 negative outlier hospital, no signal was generated. More than 1 signal was generated for 9 negative outliers and 7 other hospitals (table 3).

For TKA, 214 signals were generated for 85 hospitals at the 2-sigma control limit, with all 7 negative outlier hospitals alerted (sensitivity 100%, specificity

14%, accuracy 20%) and 85 signals were generated for 52 hospitals at 3-sigma (sensitivity 100%, specificity 51%, accuracy 54%). At 3-sigma, the first signal for negative outliers was generated after a median of 13 months (IQR:5 to 18), which should be considered against 63 false-positive signals. All negative outlier hospitals were alerted. More than 1 signal was generated for 6 negative outliers and 14 other hospitals (table 3).

II. CUSUM chart

For THA, 33 signals were generated for 16 hospitals at 3.5 control limit (sensitivity 85%, specificity 94%, accuracy 93%) and 18 signals were generated for 12 hospitals at 5 control limit, correctly alerting 11 of 13 negative outliers (sensitivity 85%, specificity 99%, accuracy 97%). At the 5 control limit, the first signal for negative outliers was generated after a median of 18 months (IQR:7 to 22), which should be considered against one false-positive signal for other hospitals. Two negative outlier hospitals were not alerted. More than 1 signal was generated for 4 negative outliers and none for other hospitals (table 3).

For TKA, 16 signals were generated for 12 hospitals at the 3.5 control limit (sensitivity 71%, specificity 92%, accuracy 91%) and 7 signals were generated for 6 hospitals at 5 control limit with 5 of the 7 outliers correctly alerted (sensitivity 71%, specificity 99%, accuracy 97%). At the 5 control limit, the first signal for negative outliers was generated after a median of 21 months (IQR:9-25) which should be considered against one false-positive signal. Two negative outliers were not alerted. More than 1 signal was generated for 1 negative outlier and none for the other hospitals (table 3).

Table 3 Characteristics of statistical process control charts

	THA (97 hospitals; 13 outliers)				TKA (98 hospitals; 7 outliers)			
	Shewhart p-chart		CUSUM chart		Shewhart p-chart		CUSUM chart	
	2-sigma	**3-sigma**	**3.5 C.L.**	**5 C.L.**	**2-sigma**	**3-sigma**	**3.5 C.L.**	**5 C.L.**
Total number of signals	195	68	33	18	214	85	16	7
• Number of good signals* (%)	76 (39%)	34 (50%)	27 (82%)	17 (94%)	36 (17%)	22 (26%)	9 (56%)	6 (86%)
• Number of false signals** (%)	119 (61%)	34 (50%)	6 (18%)	1 (6%)	178 (83%)	63 (74%)	7 (44%)	1 (14%)
Total number of hospitals with signal	70	38	16	12	85	52	12	6
• Number of good signals⁺ (%)	13 (19%)	12 (32%)	11 (69%)	11 (92%)	7 (8%)	7 (13%)	5 (42%)	5 (83%)
• Number of false signals⁺⁺ (%)	57 (81%)	26 (68%)	5 (31%)	1 (8%)	78 (92%)	45 (87%)	7 (58%	1 (17%)
Total number of hospitals with >1 signal	43	16	9	4	54	20	2	1
• Negative outliers with >1 signal (%)	13 (30%)	9 (56%)	8 (89%)	4 (100%)	7 (13%)	6 (30%)	2 (100%)	1 (100%)
• Other hospitals with >1 signal (%)	30 (70%)	7 (44%)	1 (11%)	0 (0%)	47 (87%)	14 (70%)	0 (0%)	0 (0%)
First signal for outliers (months + IQR)	5 [2-10]	10 [2-18]	16 [4-18]	18 [7-22]	5 [3-13]	13 [5-18]	15 [7-22]	21 [9-25]
Sensitivity	100%	92%	85%	85%	100%	100%	71%	71%
Specificity	32%	69%	94%	99%	14%	51%	92%	99%
Accuracy^	41%	72%	93%	97%	20%	54%	91%	97%

*Number of signals generated for negative outliers. **Number of signals generated for other hospitals. ⁺Number of negative outliers that received a signal. ⁺⁺Number of other hospitals that received a signal. ^Accuracy for correctly classifying a hospital (number of true-positive classified hospitals + number of true-negative classified hospitals) / (total number of hospitals) x 100%.

Discussion

Most arthroplasty registers report revision rates after THA and TKA, as well as differences between hospitals using funnel plots to detect hospitals with significantly worse performance than others (negative outlier hospitals).(1-6) Because of the low event rate, this is typically done by combining multiple years of data. The present study shows that monthly monitoring of THA and TKA revision rates using CUSUM charts with the 5 control limit detected worsening performance earlier than did the funnel plots, with good accuracy within a 3-year time frame; the first signal for negative outliers was generated at a median of 18 months for THA and 21 months for TKA. Using CUSUM charts to monitor deteriorating patterns for revision rates thus makes it possible to initiate improvement initiatives earlier rather than waiting for the results to appear in the funnel plot after 3 years.

Some limitations of this study should be noted. First, given the LROI privacy protocol, we could not confirm that the the negative outlier hospitals were actually being audited for worse performance by the Dutch Orthopaedic Association. However, since we and the Dutch Orthopaedic Association used both the same data source and the same statistical code to generate the outlier status in a funnel plot, it seems highly unlikely that our identification of negative outliers would have differed. Second, the number of months that the signal generation by the CUSUM chart was earlier than the signal generation by the funnel plot may not be directly generalizable to other countries, but it is likely that the differences in favor of the CUSUM chart are generalizable, particularly because the benefits have been shown previously.(25,29) Third, there is a possibility of insufficient adjustment for differences in patient-mix between hospitals because we could control only for those patient characteristics that were collected. However, this limitation would be expected to be similar for both the funnel plot and SPC charts, so it seems unlikely that it affected our conclusions regarding which method is best to detect changing performance. Fourth, registry data are self-reported by orthopaedic surgeons who may not register all revisions, but given the completeness of the Dutch register we do not believe that this affected our results considerably.(26,27) Fifth, surgeons may postpone revisions, resulting in hospitals having low 1-year revision rates but higher revision rates beyond one year. Therefore, using registries to monitor performance reflects daily practice as well as physician's behaviour. We recommend monitoring long-term revision rates (such as at 2 to 5 years) as a balancing measure to check for such occurrences.

There are few examples in orthopaedics of using SPC-methods for quality improvement.(25,29) The Scottish Arthroplasty Project reported using CUSUM chart with the 5 control limit to identify hospital variation in complications.(25) When a signal was

generated by exceeding the control limit, surgeons had to submit a review of their complications for assessment by the Scottish Orthopaedic Association. A reduction in complication rates was observed over the last years since the introduction of this quality improvement strategy. However, due to lack of a control group a causal relationship between CUSUM chart implementation and reduction in complications could not be demonstrated, as a general time trend due to other factors could have been responsible for this reduction. To our knowledge, no empirical studies have been performed to investigate how much earlier worsening performance could be detected using SPC methods before that worsening appeared as an outlier in funnel plots. These empirical data from daily practice are what the present study adds to the simulations in previous studies that already pointed to more rapid detection of small changes in performance with CUSUM charts. This is relevant for (for example) registries and scientific associations deciding whether to implement such SPC charts in their hospital feedback to initiate quality improvement.(22) By examining patient outcomes over time, SPC charts were able to detect deviating performance even when performance had been "in control" in the past, which may be difficult for a funnel plot to detect, because it uses the average outcome over a 3 year period. In addition, the CUSUM chart can be employed to examine the effect of quality improvement initiatives. Using SPC charts thus seems to add relevant information to act upon in daily practice and improve quality of care.

Similar to our study, another study showed the possibility of earlier detection of surgical site infections (SSI) outbreaks using SPC charts.(30) The Shewhart p-charts and exponential weighted moving average (EWMA) charts (another SPC chart) in that study both detected 8 out of 10 SSI outbreaks (including all 4 orthopaedic related outbreaks). In each case, a signal was generated prior to signal generation by the traditional detection methods, with a specificity of 70% and 90% for the Shewhart-p-chart and EWMA chart, respectively.

The English hospital mortality surveillance system generates CUSUM charts, on monthly-collected hospital administrative data.(7,8) After implementation of CUSUM charts, the average risk of death fell by 61% in the 9 months following a signal and reached the level of expected risk within 18 months.(7) It could be that signals were triggered by random variation and subsequent reductions occurred due to regression to the mean (a phenomenon in which extreme outcomes are likely to be followed by a fall in subsequent outcomes).(31) This may overestimate the effect of a signal. However, findings could also be explained by hospitals monitoring their own performance and taking action before a signal is generated.(7)

In contrast, one study showed no improvement in incidence rates of ward-acquired methicillin-resistant Staphylococcus aureus (MRSA) after implementation of monthly SPC feedback (with or without diagnostic tools).(14)

In 2017, the Dutch Orthopaedic Association, in collaboration with the LROI started to identify negative outlier hospitals using funnel plots including 3 years of data, with the aim of providing insight into their clinical practice compared with other Dutch hospitals.(32) This study showed that SPC charts should be included as additional hospital feedback information to provide earlier alerts if performance deteriorates and to provide hospitals with the opportunity to introduce quality improvement initiatives earlier to improve patient care. Further research must be performed to determine whether using SPC charts in daily practice will in fact initiate more quality improvement initiatives, which is the focus of an ongoing randomised controlled trial.(33) Crucial for the effectiveness is that professionals can trust the signals from the SPC chart to be reliable, as was demonstrated by data in this study, and therefore known that they warrant subsequent actions to be taken. Using SPC charts allows initiatives to be introduced earlier than is possible if hospitals wait to become an outlier in a funnel plot.

References

1. Swedisch Hip Arthroplasty Register. Annual Report 2017. https://registercentrum.blob.core.windows.net/shpr/r/Eng_Arsrapport_2017_Hoftprotes_final-Syx2fJPhMN.pdf.
2. Swedisch Knee Arthroplasty Register. Annual Report 2018. http://www.myknee.se/en/.
3. Dutch Arthroplasty Register (LROI). Annual report 2018. www.lroi-report.nl.
4. Danish Hip Arthroplasty Register. Annual Report 2018. http://danskhoftealloplastikregister.dk/en/dhr/.
5. Australian Orthopaedic Association National Joint Replacement Registry. Annual report 2018. https://aoanjrr.sahmri.com/annual-reports-2018.
6. van Schie P, van Steenbergen LN, van Bodegom-Vos L, Nelissen R, Marang-van de Mheen PJ. Between-Hospital Variation in Revision Rates After Total Hip and Knee Arthroplasty in the Netherlands: Directing Quality-Improvement Initiatives. J Bone Joint Surg Am. 2019.
7. Cecil E, Bottle, A., Esmail, A., Wilkinson, S., Vincent, C., Aylin, P. P.. Investigating the association of alerts from a national mortality surveillance system with subsequent hospital mortality in England: an interrupted time series analysis. BMJ quality & safety. 2018;27(12):965-73.
8. Cecil E, Wilkinson, S., Bottle, A., Esmail, A., Vincent, C., Aylin, P. P.. National hospital mortality surveillance system: a descriptive analysis. BMJ quality & safety. 2018;27(12):974-81.
9. Dyrkorn OA, Kristoffersen, M., Walberg, M.. Reducing post-caesarean surgical wound infection rate: an improvement project in a Norwegian maternity clinic. BMJ quality & safety. 2012;21(3):206-10.
10. Benning A, Ghaleb, M., Suokas, A., Dixon-Woods, M., Dawson, J., Barber, N., Franklin, B. D., Girling, A., Hemming, K., Carmalt, M., Rudge, G., Naicker, T., Nwulu, U., Choudhury, S., Lilford, R.. Large scale organisational intervention to improve patient safety in four UK hospitals: mixed method evaluation. BMJ. 2011;342:d195.
11. Nicolay CR, Purkayastha, S., Greenhalgh, A., Benn, J., Chaturvedi, S., Phillips, N., Darzi, A.. Systematic review of the application of quality improvement methodologies from the manufacturing industry to surgical healthcare. Br J Surg. 2012;99(3):324-35.
12. Woodall WH, Fogel, S.L., Steiner, S.H.. The monitoring and improvement of surgical-outcome quality. J Qual Technology. 2015(47):383-99.
13. Grigg O, Farewell, V.. An overview of risk-adjusted charts. J R Stat Soc Ser A Stat Soc. 2004;167:523-39.
14. Curran E, Harper, P., Loveday, H., Gilmour, H., Jones, S., Benneyan, J., Hood, J., Pratt, R.. Results of a multicentre randomised controlled trial of statistical process control charts and structured diagnostic tools to reduce ward-acquired meticillin-resistant Staphylococcus aureus: the CHART Project. J Hosp Infect. 2008;70(2):127-35.
15. Cohen ME, Liu, Y., Ko, C. Y., Hall, B. L.. Improved Surgical Outcomes for ACS NSQIP Hospitals Over Time: Evaluation of Hospital Cohorts With up to 8 Years of Participation. Ann Surg. 2016;263(2):267-73.
16. Hollesen RVB, Johansen, R. L. R., Rorbye, C., Munk, L., Barker, P., Kjaerbye-Thygesen, A.. Successfully reducing newborn asphyxia in the labour unit in a large academic medical centre: a quality improvement project using statistical process control. BMJ quality & safety. 2018;27(8):633-42.
17. Woodall WH. The use of control charts in health-care and public health surveillance. J Qual Technology. 2006;26:89-104.

18. Anhoj J, Hellesoe, A.B.. The problem with red, amber, green: the need to avoid distraction by random variation in organisational performance measures. BMJ quality & safety. 2017;26(1):81-4.
19. Schmidtke KA, Watson, D. G., Vlaev, I.. The use of control charts by laypeople and hospital decision-makers for guiding decision making. Q J Exp Psychol (Hove). 2017;70(7):1114-28.
20. Schmidtke KA, Poots, A. J., Carpio, J., Vlaev, I., Kandala, N. B., Lilford, R. J.. Considering chance in quality and safety performance measures: an analysis of performance reports by boards in English NHS trusts. BMJ quality & safety. 2017;26(1):61-9.
21. Thor J, Lundberg, J., Ask, J., Olsson, J., Carli, C., Harenstam, K. P., Brommels, M.. Application of statistical process control in healthcare improvement: systematic review. Qual Saf Health Care. 2007;16(5):387-99.
22. Neuburger J, Walker, K., Sherlaw-Johnson, C., van der Meulen, J., Cromwell, D. A.. Comparison of control charts for monitoring clinical performance using binary data. BMJ quality & safety. 2017;26(11):919-28.
23. Montgomery DC. Introduction to Statistical Quality Control. 6th ed New York: Wiley & Sons, 2009.
24. Spiegelhalter D, Sherlaw-Johnson C, Bardsley M. Statistical methods for healthcare regulation: rating, screening and surveillance. J R Stat Soc Ser A Stat Soc. 2012;175:1-47.
25. Macpherson GJ, Brenkel, I. J., Smith, R., Howie, C. R.. Outlier analysis in orthopaedics: use of CUSUM: the Scottish Arthroplasty Project: shouldering the burden of improvement. J Bone Joint Surg Am. 2011;93 Suppl 3:81-8.
26. van Steenbergen LN, Denissen GA, Spooren A, van Rooden SM, van Oosterhout FJ, Morrenhof JW, et al. More than 95% completeness of reported procedures in the population-based Dutch Arthroplasty Register. Acta Orthop. 2015;86(4):498-505.
27. *LROI website. Completeness of Registering Hospitals and Completeness of Registered Arthroplasties in the LROI Based on the Hospital Information System in 2016,* http://www.lroi-rapportage.nl/data-quality-coverage-and-completeness, accessed febr 2019.
28. Jasper LL, Jones, C. A., Mollins, J., Pohar, S. L., Beaupre, L. A.. Risk factors for revision of total knee arthroplasty: a scoping review. BMC Musculoskelet Disord. 2016;17:182.
29. Biau DJ, Milet A, Thevenin F, Anract P, Porcher R. Monitoring surgical performance: an application to total hip replacement. J Eval Clin Pract. 2009;15(3):420-4.
30. Baker AW, Haridy, S., Salem, J., Ilies, I., Ergai, A.O., Samareh, A., Andrianas, N., Benneyan, J.C., Sexton, D.J., Anderson, D.J.. Performance of statistical process control methods for regional surgical site infection surveillance: a 10-year multicentre pilot study. BMJ quality & safety. 2018;27(8):600-10.
31. Marang-van de Mheen PJ, Abel, G. A., Shojania, K. G.. Mortality alerts, actions taken and declining mortality: true effect or regression to the mean? BMJ quality & safety. 2018.
32. CommisionQuality. Commision Quality (Collaboration between Dutch Orthopaedic Association and Dutch Arthroplasty Registry). Protocol: Quality procedure, 2017. www.orthopeden.org.
33. ClinicalTrial.gov. https://clinicaltrials.gov/ct2/show/NCT04055103?term=Arthroplasty&cntry=NL&city=Leiden&draw=2&rank=2.

Supplemental data

Appendix I Description of the Shewhart-p-chart and CUSUM-chart.

Introduction and theory

In recent years, Statistical Process Control (SPC)-methods have gained growing interest in healthcare as a method to monitor quality of care and evaluate quality improvement initiatives.[1-3] In this study we opted for Shewhart-p-charts and CUSUM-charts, but other types of SPC-charts exist e.g. the exponentially weighted moving average (EMWA)-chart, and the g-chart. The general theory behind SPC-charts is that random variation is inherent in all processes, caused by common causes. A process is in-control when there is only random variation (common cause variation). However, situations may arise that cause a process to become out-of-control, due to the particular causes of this situation (special cause variation). SPC-charts with a control limit intend to distinguish between common cause variation and special cause variation, with the intention to investigate for possible causes when special cause variation is detected. The advantage of a SPC-chart over, for example, the funnel-plot where data of multiple years are taken together, is that the time variable is added by plotting the outcomes over time, showing the possible effect of changes in practice nearly real-time rather than that these remain hidden in the pooled data over a longer period.

Shewhart-p-chart

The Shewhart-p-chart generally uses a standard format, as shown in Figure 1 in the manuscript. The x-axis indicates time, e.g. weeks, months or quarters. Because it is a p-chart, the y-axis displays a proportion of a certain outcome (e.g. revision rate). The chart thus presents e.g. the weekly proportion of patients with a certain outcome over time. Three horizontal lines are depicted: the center line (CL), the upper control limit (UCL) and the lower control limit (LCL). The center line represents the average or median level of performance over a certain period. Given the random variation, an outcome will usually vary across this central tendency line and remain within the control limits, assuming that the long-term rate of that outcome does not change and will only present some random variation over time. Usually 2 and 3-sigma control limits are used, with a 2-sigma control limit having higher likelihood of type 1 error (false positive signal) and a 3-sigma control limit a higher likelihood of type 2 error (false negative signal). Control limits are computed statistically based on probability distributions such as the Gaussian ('normal' distribution), similar to hypothesis testing. In general, 95% of data will fall within ±2 standard deviations (SD) or 2

sigma and 99,7% within ±3 SD or 3 sigma. Values that fall outside the chosen upper and lower control limits exceed that range of most values, making it unlikely that this is due to random variation but rather reflects a true difference, in this study indicating that the revision rate has doubled.

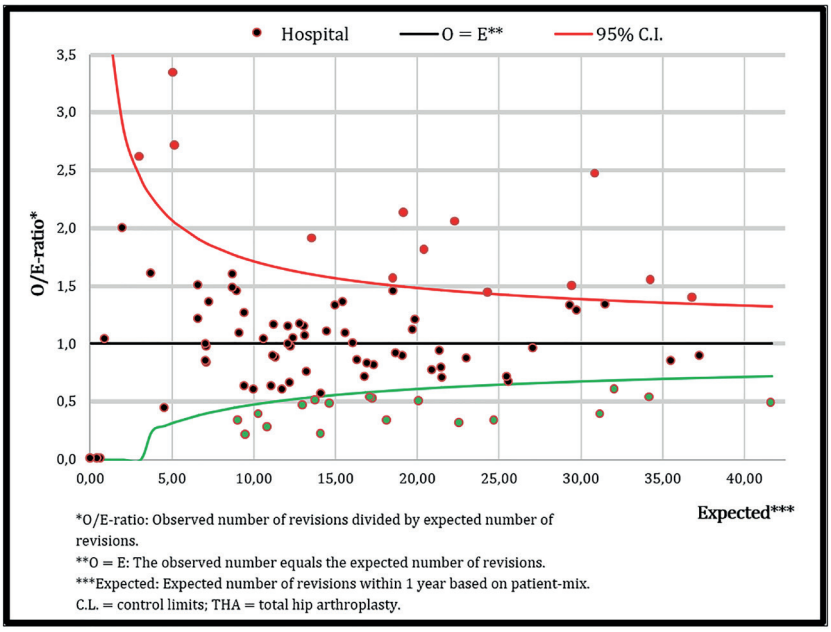
CUSUM-chart

Where the Shewhart-p-chart works with aggregated data over weeks, months of quarters, the CUSUM-chart uses every patient to plot the graph chronologically. For each patient undergoing an operation the expected chance on e.g. a revision is calculated based on certain patient characteristics and compared with the observed outcome, whether this patient has a revision or not. The line in the CUSUM-chart declines when ''good" outcomes occur (e.g. no revisions) representing better performance than expected and increases when 'unfavorable' outcomes occur (e.g. revisions) representing worse performance than expected (Figure 2 in manuscript). When performance is in balance, an increase in the line in the CUSUM-chart because of an ''unfavorable" outcome is counteracted by many small decreases in the line in the CUSUM-chart resulting from ''good" outcomes. Regardless of the use of the CUSUM-chart for detecting a better or worse outcome, the baseline always indicates that a surgeon or hospital is performing as expected. The more the CUSUM-chart line drifts away from the baseline, the more this proves that a surgeon or hospital is performing better or worse than expected. A signal for better or worse performance is generated when the control limit is exceeded, in this case to detect a doubling of the revision rate. Similar to the Shewhart-p-chart, control limit setting of CUSUM-charts allow us to balance the risk of false positive and false negative signals. The control limits in CUSUM-charts are most commonly set at 3.5 or 5, with the 3.5 having higher likelihood of false-positive signals but the 5 having higher likelihood of false negative signals.[4,5] The CUSUM-chart is reset to zero when the control limit is reached. For a detailed description of the Shewhart-p-chart and CUSUM-chart formulas, we refer to Neuburger et al and Benneyan.[4,6]

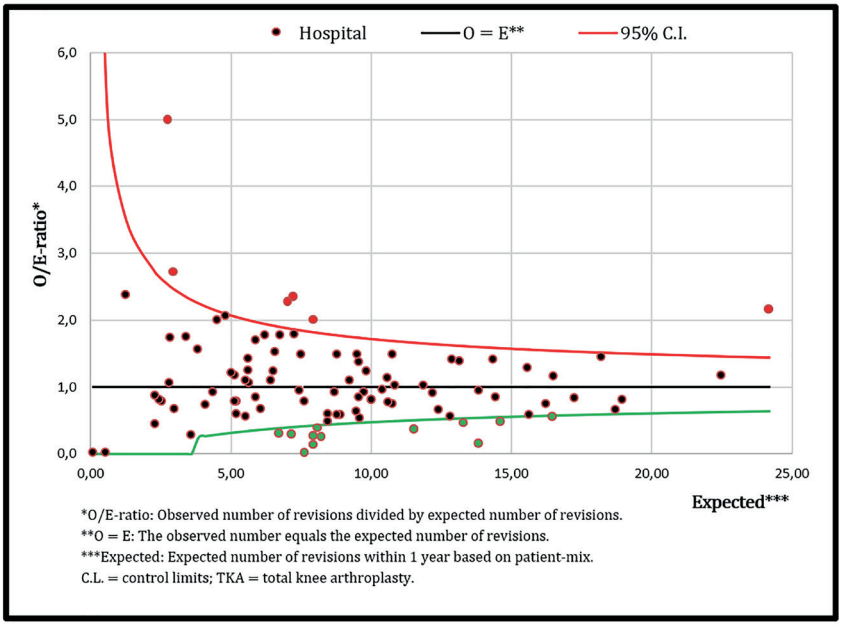
Literature

1. Benning A, Ghaleb, M., Suokas, A., Dixon-Woods, M., Dawson, J., Barber, N., Franklin, B. D., Girling, A., Hemming, K., Carmalt, M., Rudge, G., Naicker, T., Nwulu, U., Choudhury, S., Lilford, R.. Large scale organisational intervention to improve patient safety in four UK hospitals: mixed method evaluation. *BMJ.* 2011;342:d195.
2. Nicolay CR, Purkayastha, S., Greenhalgh, A., Benn, J., Chaturvedi, S., Phillips, N., Darzi, A.. Systematic review of the application of quality improvement methodologies from the manufacturing industry to surgical healthcare. *Br J Surg.* 2012;99(3):324-335.
3. Woodall WH, Fogel, S.L., Steiner, S.H.. The monitoring and improvement of surgical-outcome quality. *J Qual Technology.* 2015(47):383-399.
4. Neuburger J, Walker, K., Sherlaw-Johnson, C., van der Meulen, J., Cromwell, D. A.. Comparison of control charts for monitoring clinical performance using binary data. *BMJ quality & safety.* 2017;26(11):919-928.
5. Macpherson GJ, Brenkel, I. J., Smith, R., Howie, C. R.. Outlier analysis in orthopaedics: use of CUSUM: the Scottish Arthroplasty Project: shouldering the burden of improvement. *J Bone Joint Surg Am.* 2011;93 Suppl 3:81-88.
6. Benneyan JC, Lloyd RC, Plsek PE. Statistical process control as a tool for research and healthcare improvement. *Qual Saf Health Care.* 2003;12(6):458-464.

Appendix II. Funnel-plot of between-hospital variation in 1-year revisions after THA during 2014-2016



Appendix III. Funnel-plot of between-hospital variation in 1-year revisions after TKA during 2014-2016



Chapter 5

A more comprehensive evaluation of quality of care after total hip and knee arthroplasty: combining 4 indicators in an ordered composite outcome

Peter van Schie[1,2], Leti van Bodegom-Vos[2], Liza N. van Steenbergen[3], Rob G.H.H. Nelissen[1], Perla J. Marang-van de Mheen[2], for the IQ Joint Study group

1. Department of Orthopedics, Leiden University Medical Centre, Leiden;
2. Department of Biomedical Data Sciences, Medical Decision Making, Leiden University Medical Centre, Leiden;
3. Dutch Arthroplasty Register (LROI), 's-Hertogenbosch, The Netherlands.

Acta Orthop. 2022 Jan 3;93:138-145.

Abstract

Background and purpose

Most arthroplasty registers give hospital-specific feedback on revision rates after total hip and knee arthroplasties (THA/TKA). However, due to the low number of events per hospital, multiple years of data are required to reliably detect worsening performance, and any single indicator provides only part of the quality of care delivered. Therefore, we developed an ordered composite outcome including revision, readmission, complications, and long length-of-stay (LOS) for a more comprehensive view on quality of care and assessed the ability to reliably differentiate between hospitals in their performance (rankability) with fewer years of data.

Methods

All THA and TKA performed between 2017 and 2019 in 20 Dutch hospitals were included. All combinations of the 4 indicators were ranked from best to worst to create the ordinal composite outcome for THA and TKA separately. Between-hospital variation for the composite outcome was compared with individual indicators standardized for case-mix differences, and we calculated the statistical rankability using fixed and random effects models.

Results

22,908 THA and 20,423 TKA were included. Between-hospital variation for the THA and TKA composite outcomes was larger when compared with revision, readmission and complications, and similar to long LOS. Rankabilities for the composite outcomes were above 80% even with 1 year of data, meaning that largely true hospital differences were detected rather than random variation.

Interpretation

The ordinal composite outcome gives a more comprehensive overview of quality of delivered care and can reliably differentiate between hospitals in their performance using 1 year of data, thereby allowing earlier introduction of quality improvement initiatives.

Introduction

Traditionally, arthroplasty registries monitor and compare implant survival, with the 1-year revision rate as an indicator to detect any problems with implants at an early stage. In recent years, however, these registry data is increasingly also used to provide feedback to hospitals on their outcomes after implant surgery and as an indicator for the quality of care compared with other hospitals (1). As quality of care covers different domains such as effectiveness, safety, and efficiency, these are measured with additional indicators (2). This is acknowledged in a recent Dutch study showing that orthopedic surgeons would like to receive feedback not only on revision, but also regarding readmission, complications, and length of stay (LOS) for hospital comparisons and to monitor the quality of care delivered (3). The rationale is that benchmarking and feedback may spur quality improvement initiatives in case of suboptimal performance.

Arthroplasty registries primarily provide feedback on single indicators, such as revision surgery or mortality, but any single indicator provides an incomplete overview of the quality of care (4). Furthermore, comparing hospital performance on multiple individual indicators is difficult, because a hospital may have a high score on one indicator but a low score on another. Because of these limitations, there is growing interest in composite measures, in which multiple relevant indicators are combined to provide a more comprehensive overview of delivered quality of care for patients when choosing a hospital for treatment, and also increase the number of events to make it better suitable for benchmarking hospitals (5-10). The higher number of events for composite outcomes increases the accuracy by which hospital performance is estimated (lower statistical uncertainty). A previous study showed that 3 years of data were needed to reliably differentiate between hospitals for 1-year revisions due to the low numbers of events per hospital (4). Therefore, a long time is needed before worsening performance is detected reliably, resulting in late action plans to improve quality of care. Combining multiple indicators into a composite outcome could help to increase the number of events, so that a shorter time period is needed to reliably differentiate between hospitals in their performance (6-8,11).

Existing composite outcomes often represent an all-or-none concept, like the proportion of patients with all desired indicators realized, also known as Textbook Outcome (TO). For orthopedics, 8 related indicators for total hip and knee arthroplasty (THA/TKA) were recently combined into such a binomial outcome (5). However, all-or-none measures are less informative, as outcome frequencies may vary considerably between indicators and frequently occurring outcomes will dominate the composite outcome results, a well-known disadvantage from the trial literature (12).

These measures are also less useful for quality improvement, as they give equal weight to all outcomes and do not provide feedback on where and how to improve, i.e., in which of the (combination of) outcomes, nor will they be very sensitive to monitor the effect of targeted initiatives to improve a single outcome (13). An ordinal composite measure with all combinations of indicators ranked from best to worst would be able to provide such feedback, taking into account possible interrelationships between individual indicators and pointing more specifically at where to improve care (7,11).

We therefore developed an ordered composite outcome including 1-year revision, 30-day readmission, 30-day complications, and upper-quartile LOS, separately for THA and TKA. In addition, we compared the statistical reliability of ranking hospitals between the composite and individual outcomes, both when including 3 years and 1 year of data, to assess when hospital differences in performance could be reliably detected.

Methods

Data collection

Anonymous data of all patients undergoing a primary THA or TKA between January 1, 2017 and December 31, 2019 were included from 20 Dutch hospitals (2 university, 5 teaching, 7 general, and 6 private hospitals, which reflects the national distribution). These hospitals are participating in a randomized controlled trial to test whether an intervention consisting of monthly feedback, interactive education, and a toolbox with suggested quality improvement initiatives, is effective to result in more initiatives undertaken and better patient outcomes (ClinicalTrial.gov). (14) Routinely submitted data to the Dutch Arthroplasty Register (LROI) were used to generate feedback, supplemented with hospital data on readmissions, complications, and LOS for each patient. The LROI-data collection methods and completeness have been described previously (4). In summary, data completeness is checked against Hospital Electronic Health Records and currently exceeds 98% for primary procedures and 96% for revisions (15,16) (LROI website). The hospitals have been given a clear definition for each indicator as described below to avoid measurement variability. Less than 9% of readmission, complications, and LOS data were missing for both THA and TKA, so a composite outcome could not be calculated for 8.7% of THA and 7.9% of TKA patients.

Hospital performance indicators

The 1-year revision was calculated based on the primary surgery and revision dates, routinely collected in the LROI. Other indicators were calculated based on the index

hospitalization when the primary THA or TKA was performed. The indicators were defined as:

- Revision: Exchange, removal, or addition of any component within 1 year after surgery;

- Readmission: An admission within 30 days after discharge of the index hospitalization;

- Complication: An adverse event other than revision and death during the index hospitalization or within 30 days after discharge. The most commonly registered complications were postoperative bladder retention (13%), hip dislocation (10%), and surgical site infection (7%) for THA and postoperative bladder retention (17%), wound leakage (8%), and surgical site infection (7%) for TKA;

- Long LOS: LOS of the index hospitalization longer than the 75th percentile, based on all patients treated, included to also take into account possible hospital differences in sensitivity of reporting complications.

All indicators were case-mix adjusted for fair hospital comparison. The following patient characteristics are available in the LROI: age, sex, BMI, current smokers (yes/no), ASA classification (I,II,III-IV), Charnley score (A,B1,B2,C,n/a) and diagnosis (osteoarthritis/non-osteoarthritis).

Ordinal composite outcome

To order the individual indicators, an anonymous internet-based questionnaire was sent during June-July 2020 using Qualtrics (QualtricsXM, Provo, UT, USA). All 135 orthopedic surgeons performing THA and/or TKA in the 20 hospitals were asked to rank the indicators with the patient's perspective in mind, from 1 (least severe outcome) to 4 (most severe outcome). Reminders were sent 1 and 2 weeks after the first invitation, resulting in a response rate of 39%. The final ordering was based on the mean number of points assigned per indicator across respondents: 1) long LOS (1.1 points); 2) complications (2.5 points); 3) readmission (2.6 points) and 4) revision (3.9 points). This ordering seems to be supported by previous studies, showing that complications during admission (resulting in long LOS) did not affect patient's quality of care evaluation, while complications after discharge (resulting in readmission) did, suggesting that patients consider readmissions to be worse than long LOS (11,17,18).

All possible combinations of indicators were then ranked from best to worst using the above ordering. Patients with a revision were combined into one group to avoid

subgroups with few events and because we considered the impact of a revision to be higher (19-23). This resulted in the following 9 combinations:

- No revision, no readmission, no complications, no long LOS (TO);
- No revision, no readmission, no complications, long LOS;
- No revision, no readmission, complications, no long LOS;
- No revision, no readmission, complications, long LOS;
- No revision, readmission, no complications, no long LOS;
- No revision, readmission, no complications, long LOS;
- No revision, readmission, complications, no long LOS;
- No revision, readmission, complications, long LOS;
- Revision.

Statistics

Patient characteristics were missing in less than 5% of patients. These were considered to be missing at random and imputed using multiple imputations for 10 rounds with predictive mean matching as the underlying model. All variables were used as predictors, including the outcome variables, but only patient characteristics were imputed.

1st, the standardized ordered composite outcome for each hospital was calculated using ordinal logistic regression with all patient characteristics and hospital as fixed effect independent variables. The coefficient of each hospital was compared to the average across all hospitals, and the difference exponentiated to give a proportional odds ratio higher or lower than the average, similar to the standardized individual indicators.

2nd, the standardized rates for the individual indicators revision, readmission, complications, and long LOS were calculated. For each indicator, the expected risk for each patient was calculated using logistic regression analysis with all patient characteristics as independent variables (excluding hospital) and the indicator (yes/no) as dependent variable. Summing all patients' expected probabilities treated in a hospital, resulted in the expected number of patients having the indicator for that

hospital. The observed number of patients for that indicator was divided by the expected number to calculate the standardized indicator (observed/expected) for each hospital.

The between-hospital variation for standardized individual indicators and the composite outcome were described using the median and interquartile range (IQR). Hospital-level correlations between standardized individual indicators were calculated using Pearson correlation coefficients, to indicate to what extent hospital performance on individual indicators would point in the same direction or not, and thereby the added value of capturing more information in the composite. The strength of correlations was defined as: ≤0.35 weak >0.35-0.67 moderate and >0.67 strong (24).

Statistical reliability of ranking

We examined the reliability of ranking (rankability) hospitals to assess whether the composite outcome would more reliably differentiate between hospitals in their performance than individual indicators. The rankability is the percentage of between-hospital variation (in terms of the indicator) that is due to "true" hospital differences as opposed to natural/random (chance) variation due to unexplained factors (4,7,11,25-27) and was calculated as previously described (4). In short, the between-hospital variation from random effect logistic regression models, was divided by the sum of between-hospital and within-hospital variation from fixed-effect logistic regression models, both adjusted for case-mix. Rankabilities were calculated for all 10 imputed datasets and the mean and range were given across datasets. Rankability was classified as low (<50%), moderate (50%-75%), or high (>75%) (27). Rankability was calculated for single years and 3 years of data to assess whether hospitals can be reliably ranked with less data.

All analyses were performed using SPSS version 25 (IBM Corp, Armonk, NY, USA), except for rankability analyses for which STATA version 14.2 (StataCorp, College Station, TX, USA) was used.

Ethics, funding, and potential conflict of interest

The LUMC Medical Ethical Committee waived the need for ethical approval under Dutch law (CME, G18.140). PvS received a grant from the Van Rens Foundation (VRF2018-001) to perform this study. The authors declare no conflicts of interest.

Results

22,908 THA and 20,423 TKA procedures were included. Overall patient-level revision, readmission, complication, and long LOS rates were lower for TKA than THA (Table 1). LOS was not normally distributed making it difficult to create equal quartiles, so the closest integer value was chosen resulting in above 4 days defined as long LOS for both THA and TKA. This explains the percentage of patients with long LOS being considerably smaller than 25%. The mean LOS was 3.3 (sd 2.9) days for THA and 3.0 (sd 2.1) for TKA. At hospital level, the number of procedures performed varied considerably with a median of 1,188 for THA and 848 for TKA (Table 2). The overall patient-level and hospital-level revision rates (Tables 1 and 2) were comparable to those observed in all Dutch hospitals of patients operated on between January 2014 and December 2016 (4).

Table 1 Baseline patient characteristics and indicators after THA and TKA in the period 2017-2019 in 20 Dutch hospitals.

	THA (*n*=22908)	**TKA (*n*=20423)**
Patient characteristics		
Mean age in years (SD)	69 (10)	68 (8.8)
Sex, female (%)	14707 (64)	12606 (62)
BMI (SD)	27 (4.5)	29 (4.8)
Current Smokers (%)	2395 (11)	1667 (8.4)
ASA classification (%)		
• ASA I	4113 (18)	2736 (13)
• ASA II	14533 (63)	13759 (68)
• ASA III-IV	4259 (19)	3924 (19)
Charnley score (%)		
• A	9205 (42)	7529 (37)
• B1	7082 (32)	7598 (37)
• B2	4984 (23)	4470 (22)
• C	711 (3)	722 (4)
Diagnosis (%)		
• Osteoarthritis	20214 (88)	19723 (97)
• Non-Osteoarthritis	2669 (12)	697 (3.4)
Indicators		
1-year revision (%)	410 (1.8)*	250 (1.2)**
30-day readmission (%)	829 (3.9)	633 (3.4)
30-day complication (%)	1027 (4.5)	620 (3.3)
Long LOS, upper quartile (%)	2794 (13.3)	2123 (11.4)

*The 1-year revision percentage for THA was 1.8% in the Netherlands during 2014-2016.(4)
**The 1-year revision percentage for TKA was 1.2% in the Netherlands during 2014-2016.(4)
ASA = American Society of Anesthesiologists; BMI = body mass index; LOS = length of stay; SD = standard deviation; THA = total hip arthroplasty; TKA = total knee arthroplasty.

Table 2 Baseline hospital-level characteristics and indicators for 20 Dutch hospitals performing THA and TKA.

	THA (*n*=20 hospitals)	TKA (*n*=20 hospitals)
	Median (IQR)	Median (IQR)
Procedures (*n*)	1188 (623-1630)	848 (593-1552)
Patient characteristics		
Mean age (years)	69 (65-70)	69 (66-70)
Sex, female (%)	64 (62-65)	63 (59-64)
Mean BMI	27 (27-27)	30 (29-30)
Current smokers (%)	11 (9.2-13)	9 (7.3-10)
ASA classification (%)		
• ASA I	14 (9.7-23)	9.8 (7.5-19)
• ASA II	64 (59-70)	67 (61-72)
• ASA III-IV	23 (14-28)	23 (14-31)
Charnley score (%)		
• A	43 (37-48)	39 (28-42)
• B1	31 (28-34)	34 (32-42)
• B2	22 (19-24)	22 (19-25)
• C	3.0 (1.2-5.3)	2.5 (1.5-5.6)
Diagnosis (%)		
• Osteoarthritis	89 (83-93)	97 (95-98)
• Non-Osteoarthritis	11 (7.3-17)	3.0 (2.3-5.3)
Indicators		
1-year revision (%)	1.7 (0.8-2.7)*	1.3 (0.7-1.7)**
• ***Standardized***	*0.9 (0.6-1.6)*	*1.0 (0.7-1.4)*
30-day readmission (%)	4.2 (1.8-6.0)	3.8 (1.7-5.5)
• ***Standardized***	*0.9 (0.4-1.3)*	*1.0 (0.5-1.4)*
30-day complication (%)	3.8 (2.3-5.5)	2.3 (1.0-4.3)
• ***Standardized***	*0.7 (0.5-1.1)*	*0.7 (0.3-1.2)*
Long LOS, upper quartile (%)	11 (2.2-23)	11 (2.6-21)
• ***Standardized***	*0.7 (0.3-1.4)*	*0.9 (0.2-1.4)*

The value under ''Median (IQR)" indicate the mean or the percentage of the median hospital. All standardized indicators were adjusted for: age, gender, BMI, current smokers, ASA classification, Charnley score and diagnosis.
*The median percentage on hospital-level for THA was 1.6% (IQR:1.0-2.3) in the Netherlands during 2014-2016;(4)
**The median percentage on hospital-level for TKA was 1.1% (IQR:0.7-1.6) in the Netherlands during 2014-2016;(4)
ASA = American Society of Anesthesiologists; BMI = body mass index; IQR = Interquartile range; THA = total hip arthroplasty; TKA = total knee arthroplasty.

Between-hospital variation of individual indicators

Hospitals differed considerably in their case-mix (especially ASA-classification and diagnosis for THA) and crude indicator outcomes (Table 2). Largest variation was found for long LOS (THA: IQR [2.2%-23.4%] TKA: IQR [2.6%-20.9%]) and smallest for revision (THA: IQR [0.8%-2.7%] TKA: IQR [0.7%-1.7%]). After adjustment for case-mix, considerable variation remained with largest variation for long LOS (THA: IQR [0.3-1.4] TKA: IQR [0.2-1.4]) and smallest for complications after THA (IQR [0.5-1.1]) (Table 2).

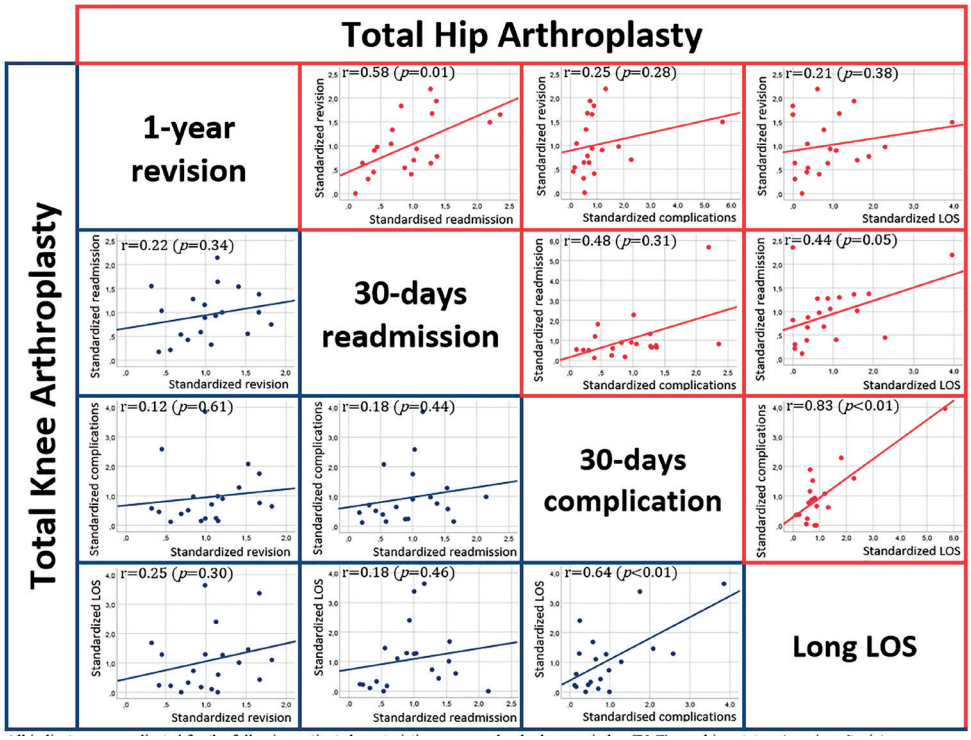


Figure 1 Correlation between standardized rates of individual indicators at hospital-level.All indicators were adjusted for the following patient characteristics: age, gender, body mass index (BMI), current smokers, American Society Anesthesiologists (ASA) classification, Charnley score and diagnosis.
LOS = length of stay.

Relations between individual indicators

Most individual hospital-level indicators were not related, meaning that hospitals with a good performance on one indicator do not necessarily have a good performance on another indicator (Figure 1). For THA, only revision rates are moderately correlated with readmission rates (r=0.58, p=0.01) and complications with long LOS have a strong correlation (r=0.83, p<0.01). For TKA, only complications with long LOS are moderately correlated (r=0.64, p<0.01). The ordinal composite outcomes will capture these relationships but also add the information captured by unrelated indicators.

Ordinal composite outcome

Figure 2 shows the hospital variation in the composite outcome. The median hospital had 18% (IQR [8.4%-28%]) patients without TO for THA and 21% (IQR [7.9%-25%]) for TKA, both increasing the number of events and between-hospital variation compared with the median revision rates of 1.8% (IQR [1.0%-2.8%]) and 1.3% (IQR [0.7%-1.7%]) respectively. Among patients with a revision after THA, 50%

were readmitted after the index hospitalization, 38% had complication(s), 27% had a long LOS, but 25% of the patients had no other indicators (7% missing data). Estimates for TKA were 41%, 28%, 17% and 40% respectively (9% missing data). The between-hospital variation in the standardized ordinal composite outcome, expressed as proportional odds ratios (THA: median 1.0 (IQR:[0.5-1.7]) and TKA: 1.3 (IQR:[0.4-1.6]) were larger than for revisions, readmissions, complications, and similar to long LOS (Table 2 and Figure 2).

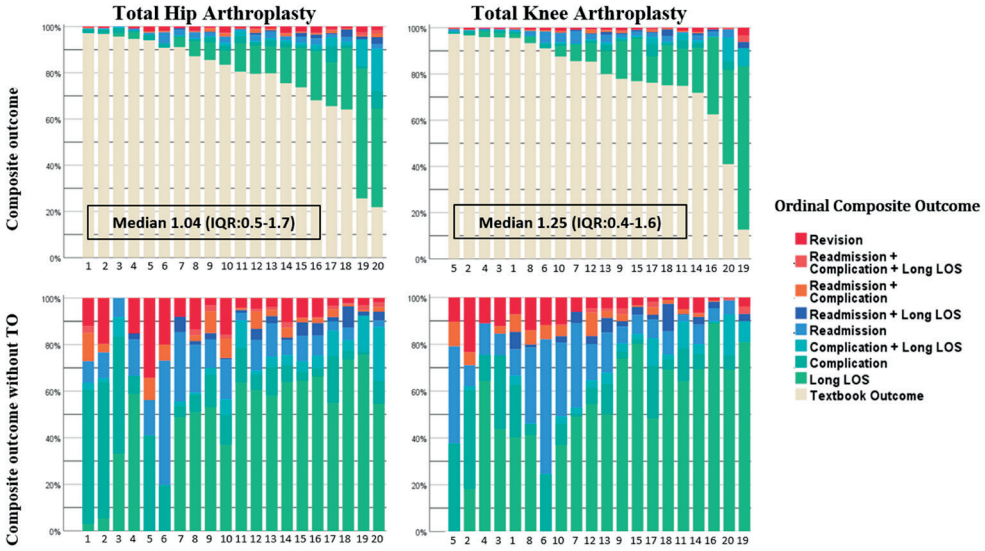


Figure 2 Crude ordinal composite outcome distribution per hospital and standardized effect of the hospitals on the composite outcome (median and IQR).
This graphs show the crude outcome distribution per hospital (*n*=20). The hospitals are numbered on the x-axis. The hospitals for TKA were labelled according to their rank for TO in THA's The standardised odds of the hospital effect (median and IQR) were adjusted for the following patient characteristics: age, gender, body mass index (BMI), current smokers, American Society Anaesthesiologists (ASA) classification, Charnley score and diagnosis.
IQR = interquartile ranges; TO = Textbook Outcome.

Reliability of ranking hospitals

Using 3 years of data, hospitals can be reliably ranked as rankabilities were high for most individual indicators and the composite outcome (Figure 3), except for the moderate rankability for readmission (THA and TKA) and revision (THA) and low rankability for revision (TKA). Using single years, rankability was low for revision, low to moderate for readmission, moderate to high for complications, and for long LOS but consistently high for the composite outcomes.

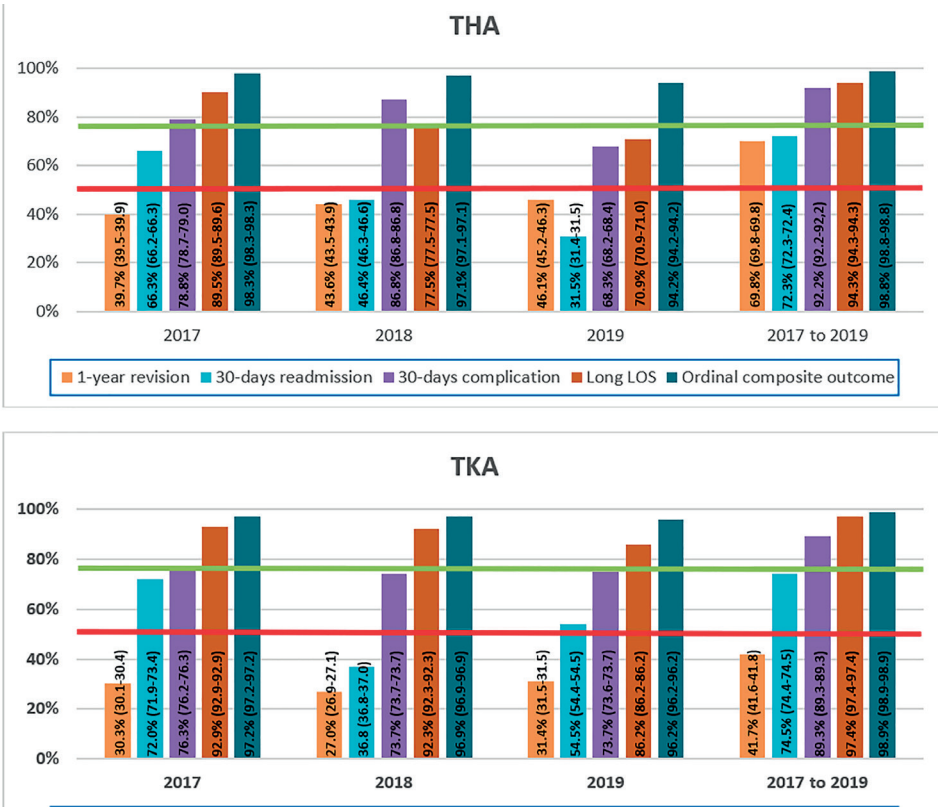


Figure 3 Rankabilities of individual indicators and ordinal composite outcomes.
The mean (range) of the rankabilities are described within the graph. The rankability is high when the bar is above the green line, moderate when between the red and green line and low when below the red line.
LOS = length of stay; THA = total hip arthroplasty; TKA = total knee arthroplasty.

Discussion

We developed an ordered composite outcome including all combinations of 4 relevant quality of care indicators to give a more comprehensive overview i.e., not only whether patients had a revision, but also whether they were readmitted, experienced complications, or had long LOS. Using this composite outcome, quality improvement initiatives can be tailored to specific patient groups based on the combination of indicators. The between-hospital variation in the composite outcomes was larger than for the individual outcomes revision, readmission and complications, and similar for long LOS. Statistically, this contributed to a higher rankability (i.e., a higher percentage of the variation being due to "true" differences rather than chance). The composite outcome was able to reliably differentiate between hospitals in their

performance when using only 1 year of data, thereby allowing earlier introduction of quality improvement initiatives. The added value of the composite was also supported by the lack of hospital-level correlation between many individual indicators, meaning that hospital performance may be quite different depending on which indicator is being examined whereas these are all included in the composite outcome. It thus gives a more comprehensive view on quality of delivered care and is better able to differentiate between hospitals in their performance.

Comparison with literature

Compared with 2 previously developed orthopedic composite measures, our measure includes revision rather than only short-term indicators, which adds relevant information as revision is generally considered a serious adverse event for patients and a quality indicator used in arthroplasty registries (5,7). Furthermore, previously developed binary measures miss the underlying relations among individual indicators, making it unclear on which outcome a hospital needs to improve if performance is worse than in other hospitals. The reliability of ranking hospitals on revision using 3 years of data among these 20 hospitals was similar to previous estimates among all Dutch hospitals, where rankabilities of 62% for THA and 42% for TKA were reported in 2014-2016, versus 70% and 42% in the present study (4). Similarly, another study also reported higher rankabilities for LOS than for readmission using single and 3 years of data (7). A higher rankability by combining individual indicators into an (ordered) composite outcome was also seen in other studies, meaning that most variation reflects true hospital differences rather than merely chance (7,11,28,29). A recent simulation study showed that the rankability of an ordinal composite outcome depends on the rankability of the more prevalent individual indicator, and the extent to which individual indicators making up the composite are correlated within hospitals (30). If individual indicators are completely independent, the rankability of the composite will often be less than at least 1 individual indicator, whereas it will be higher if the within-hospital correlation is at least 0.5. As indicated in Figure 1, the within-hospital correlation for several indicators was around or above 0.5 in our study, for which the simulations showed higher rankabilities for the composite than the individual indicators in 50% of the scenarios.

Strengths and limitations

Strengths of this study are the limited risk of selection bias and using case-mix adjusted rates, because the LROI data includes over 98% of patients and patient characteristics have only less than 5% missing values (15,16). Data supplemented by the hospitals (i.e., readmission, complications, and LOS) was missing in less than 9%. In addition, our approach can be readily applied in other arthroplasty registries that include data

on these indicators, or use data linkage with administrative data sources as done in the present study (1).

This study also has some limitations. The generalizability to other countries may be limited due to differences in e.g., discharge policies, and availability of resources for supporting patients at home, which may result in different estimates and hospital variation for readmissions and long LOS (31). However, it seems likely that combining indicators in a composite will similarly improve rankability unless indicators would have completely different interrelations.(30) 2nd, only 20 of 102 Dutch hospitals were included, since hospital readmission, complications, and LOS are not routinely collected by the LROI. However, both the average patient-level revision rate and the hospital-level variation in revisions were similar to that shown for all hospitals (Tables 1 and 2) (4) suggesting the sample to be fairly representative although we do not have data on the other indicators. 3rd, when readmissions and complications occurred in another hospital, these would be missed and result in these rates being underestimated. However, as early complications fall within the diagnosis-related group (DRG) paid to the hospital performing the primary arthroplasty, it is very likely that patients go back to the same hospital. And even if it would occur, it will only influence the relationship between indicators if systematically and frequently in some hospitals while not in others, which does not seem likely. 4th, all complications were included regardless of severity, although this is partly reflected in whether they occur in combination with a readmission or merely prolonged LOS. Future research could refine the composite outcome including this distinction by severity, but this would also increase the number of combinations, potentially making it less useful as feedback.

Implementation

Individual indicators measure one aspect of quality of care, but lack the ability to measure the entire chain of delivered quality. One hospital may perform well on one indicator (e.g., 1-year revision), while at the same time performing worse on another (e.g., 30-day readmission). The composite outcome includes this and may thus help patients, for example, if they want to know how often the procedure is going as planned for a specific hospital to look at the TO which is still visible. For healthcare providers, it provides insight how often combinations of indicators of adverse outcome occur (as each patient can only be classified into one of the predetermined categories). Furthermore, this also guides which medical records have to be reviewed (characterized by the specific combinations of outcomes) to investigate whether care can be improved for these patients. For example, hospital 5 and 6 for THA in Figure 2 had zero long LOS patients, but a relatively high number of readmissions, which may indicate that patients were discharged too early. Rather than reviewing the records of all readmitted patients in case of a relatively high readmission rate, hospitals can

now more selectively review records of readmitted patients with a normal length of stay to investigate more specifically whether e.g. information around discharge can be improved to ensure adequate patient management at home and avoid readmissions because patients can be monitored at the outpatient clinic. Hospital 16 for TKA had a relatively high number of patients with a long LOS, but a low number of patients with other adverse outcomes, suggesting there may be a delay in transfers or that this is caused by other logistical issues that can be addressed. The hospitals 5, 6, 7, 8 and 10 had many readmissions within 30 days without a complication within 30 days recorded, providing insight into whether to improve on reporting completeness or if there was no complication, discussing whether the readmission was needed or could have been adequately treated at the outpatient clinic, which may improve care. A final advantage of the composite outcome is that it prevents "gaming" of the individual indicator, e.g., when hospitals receive incentives or penalties when individual indicators are too high, because reducing one indicator may increase another if they are related. For implementation in registries, they work towards more frequent data submissions, preferably monthly rather than annually which currently is often the norm. This is needed to allow for near real-time monitoring of indicator outcomes, and so that any subsequent improvement actions can be undertaken without delay.

Conclusion

The newly developed ordinal composite outcome provides a more comprehensive overview of the quality of care delivered, as it has ordered all combinations of revision, readmission, complications and long LOS. This composite outcome more reliably differentiates between hospitals in their performance than individual indicators using only 1 year of data, thereby allowing earlier introduction of quality improvement initiatives targeted to more specific patient groups.

Acknowledgement

We gratefully acknowledge the input of the members of the 20 hospitals who provided their data to complete this study, as part of the IQ Joint study group (in alphabetic order): Antonius Hospital, Sneek (S.T. Hokwerda); Bergman Clinics, Arnhem (P. van Kampen); Bergman Clinics, Breda (J. Schrier); Bergman Clinics, Delft (F. de Graaff); Bergman Clinics, Naarden (H. Bouma); Bergman Clinics, Rijswijk (R. van Hierden); Bergman Clinics, Rotterdam (M. Vischjager); Catharina Hospital, Eindhoven (R.W.T.M. van Kempen); Dijklander Hospital, Hoorn (W.C. Neve); Elisabeth-TweeSteden Hospital (T. Gosens); Gelderse Vallei Hospital, Ede (W. Beijneveld);

Leiden University Medical Centre, Leiden (H.M.J van der Linden); Maxima Medical Centre, Eindhoven (M. van den Besselaar); Medical Spectrum Twente, Enschede (W. Verra); Onze Lieve Vrouwe Gasthuis, Amsterdam (R.W. Poolman); Sint Anna Hospital, Geldrop (W. van der Weegen); Sint Franciscus Hospital, Schiedam (A. Polak); Tjongerschans Hospital, Heerenveen (M. Mulder); University Medical Centre Groningen, Groningen (M. Stevens); Zuyderland Hospital, Sittard (B. Boonen).

References

1. Lubbeke A, Silman AJ, Barea C, Prieto-Alhambra D, Carr AJ. Mapping existing hip and knee replacement registries in Europe. Health Policy. 2018;122(5):548-57.
2. WorldHealthOrganisation. World Health Organisation (WHO) - Quality of Care. https://www.who.int/health-topics/quality-of-care#tab=tab_1.
3. Van Schie P, Van Bodegom-Vos L, Zijdeman TM, Nelissen R, Marang-Van De Mheen PJ. Awareness of performance on outcomes after total hip and knee arthroplasty among Dutch orthopedic surgeons: how to improve feedback from arthroplasty registries. Acta Orthop. 2020:1-8.
4. van Schie P, van Steenbergen LN, van Bodegom-Vos L, Nelissen R, Marang-van de Mheen PJ. Between-Hospital Variation in Revision Rates After Total Hip and Knee Arthroplasty in the Netherlands: Directing Quality-Improvement Initiatives. J Bone Joint Surg Am. 2019.
5. Hollenbeck B, Hoffman MA, Tromanhauser SG. High-Volume Arthroplasty Centers Demonstrate Higher Composite Quality Scores and Enhanced Value: Perspective on Higher-Volume Hospitals Performing Arthroplasty from 2001 to 2011. J Bone Joint Surg Am. 2020;102(5):362-7.
6. Karthaus EG, Lijftogt N, Busweiler LAD, Elsman BHP, Wouters M, Vahl AC, et al. Textbook Outcome: A Composite Measure for Quality of Elective Aneurysm Surgery. Ann Surg. 2017;266(5):898-904.
7. Hofstede SN, Ceyisakar IE, Lingsma HF, Kringos DS, Marang-van de Mheen PJ. Ranking hospitals: do we gain reliability by using composite rather than individual indicators? BMJ quality & safety. 2019;28(2):94-102.
8. Dimick JB, Birkmeyer NJ, Finks JF, Share DA, English WJ, Carlin AM, et al. Composite measures for profiling hospitals on bariatric surgery performance. JAMA Surg. 2014;149(1):10-6.
9. Marang-van de Mheen PJ, Dijs-Elsinga J, Otten W, Versluijs M, Smeets HJ, Vree R, et al. The relative importance of quality of care information when choosing a hospital for surgical treatment: a hospital choice experiment. Med Decis Making. 2011;31(6):816-27.
10. Dijs-Elsinga J, Otten W, Versluijs MM, Smeets HJ, Kievit J, Vree R, et al. Choosing a hospital for surgery: the importance of information on quality of care. Med Decis Making. 2010;30(5):544-55.
11. Lingsma HF, Bottle A, Middleton S, Kievit J, Steyerberg EW, Marang-van de Mheen PJ. Evaluation of hospital outcomes: the relation between length-of-stay, readmission, and mortality in a large international administrative database. BMC Health Serv Res. 2018;18(1):116.
12. Montori VM, Permanyer-Miralda G, Ferreira-González I, Busse JW, Pacheco-Huergo V, Bryant D, et al. Validity of composite end points in clinical trials. BMJ. 2005;330(7491):594-6.
13. Barclay M, Dixon-Woods M, Lyratzopoulos G. The problem with composite indicators. BMJ quality & safety. 2019;28(4):338-44.
14. ClinicalTrial.gov. https://clinicaltrials.gov/ct2/show/NCT04055103?term=Arthroplasty&cntry=NL&city=Leiden&draw=2&rank=2.
15. van Steenbergen LN, Denissen GA, Spooren A, van Rooden SM, van Oosterhout FJ, Morrenhof JW, et al. More than 95% completeness of reported procedures in the population-based Dutch Arthroplasty Register. Acta Orthop. 2015;86(4):498-505.
16. *LROI website. Completeness of Registering Hospitals and Completeness of Registered Arthroplasties in the LROI Based on the Hospital Information System in 2016,* http://www.lroi-rapportage.nl/data-quality-coverage-and-completeness, accessed febr 2019.
17. Marang-van de Mheen PJ, van Duijn-Bakker N, Kievit J. Surgical adverse outcomes and patients' evaluation of quality of care: inherent risk or reduced quality of care? Qual Saf Health Care. 2007;16(6):428-33.

18. Poelemeijer YQM, Marang-van de Mheen PJ, Wouters M, Nienhuijs SW, Liem RSL. Textbook Outcome: an Ordered Composite Measure for Quality of Bariatric Surgery. Obes Surg. 2019;29(4):1287-94.
19. Jaffer AK, Barsoum WK, Krebs V, Hurbanek JG, Morra N, Brotman DJ. Duration of anesthesia and venous thromboembolism after hip and knee arthroplasty. Mayo Clin Proc. 2005;80(6):732-8.
20. Khatod M, Barber T, Paxton E, Namba R, Fithian D. An analysis of the risk of hip dislocation with a contemporary total joint registry. Clin Orthop Relat Res. 2006;447:19-23.
21. Mahomed NN, Barrett JA, Katz JN, Phillips CB, Losina E, Lew RA, et al. Rates and outcomes of primary and revision total hip replacement in the United States medicare population. J Bone Joint Surg Am. 2003;85(1):27-32.
22. Peersman G, Laskin R, Davis J, Peterson M. Infection in total knee replacement: a retrospective review of 6489 total knee replacements. Clin Orthop Relat Res. 2001(392):15-23.
23. Pulido L, Parvizi J, Macgibeny M, Sharkey PF, Purtill JJ, Rothman RH, et al. In hospital complications after total joint arthroplasty. J Arthroplasty. 2008;23(6 Suppl 1):139-45.
24. Taylor R. Interpretation of the Correlation Coefficient: A Basic Review. JDSM. 1990;6(6):35-9.
25. Henneman D, van Bommel AC, Snijders A, Snijders HS, Tollenaar RA, Wouters MW, et al. Ranking and rankability of hospital postoperative mortality rates in colorectal cancer surgery. Ann Surg. 2014;259(5):844-9.
26. van Dishoeck AM, Koek MB, Steyerberg EW, van Benthem BH, Vos MC, Lingsma HF. Use of surgical-site infection rates to rank hospital performance across several types of surgery. Br J Surg. 2013;100(5):628-36; discussion 37.
27. van Dishoeck AM, Lingsma HF, Mackenbach JP, Steyerberg EW. Random variation and rankability of hospitals using outcome indicators. BMJ quality & safety. 2011;20(10):869-74.
28. Dimick JB, Welch HG. The zero mortality paradox in surgery. J Am Coll Surg. 2008;206(1):13-6.
29. Drye EE, Chen J. Evaluating quality in small-volume hospitals. Arch Intern Med. 2008;168(12):1249-51.
30. Austin PC, Ceyisakar IE, Steyerberg EW, Lingsma HF, Marang-van de Mheen PJ. Ranking hospital performance based on individual indicators: can we increase reliability by creating composite indicators? BMC Med Res Methodol. 2019;19(1):131.
31. Bottle A, Middleton S, Kalkman CJ, Livingston EH, Aylin P. Global comparators project: international comparison of hospital outcomes using administrative data. Health Serv Res. 2013;48(6 Pt 1):2081-100.

Chapter 6

Is postoperative improvement in PROM scores after THA and TKA overestimated due to selective non-response? Linking PROM scores to adverse event data provides insights

Peter van Schie MD1,2, Leti van Bodegom-Vos PhD1, Tristan M Zijdeman1, Taco Gosens, MD PhD3, Rob G.H.H. Nelissen MD PhD2, Perla J. Marang-van de Mheen PhD1,4 for the IQ Joint Study group

1. Department of Biomedical Data Sciences, Medical Decision Making, Leiden University Medical Centre, Leiden, The Netherlands;
2. Department of Orthopaedics, Leiden University Medical Centre, Leiden, The Netherlands;
3. Department of Orthopaedics, ETZ, Tilburg, The Netherlands;
4. Department of Safety & Security Science, Faculty of Technology, Policy & Management, Delft University of Technology, Delft, The Netherlands.

Submitted

Abstract

Background and purpose: Improvement in Patient Reported Outcome Measure (PROM) scores may be under or overestimated after total hip and knee arthroplasty (THA and TKA) as questionnaires are filled in by a selection of patients (PROM respondents). Linking PROM scores to adverse event data may provide insight in the direction of bias as these are known for all patients and likely associated with improvement in PROM scores. We, therefore, compared PROM respondents and non-respondents on their adverse event rates, examined whether patients experiencing adverse events had different improvement rates in PROM scores and whether hospitals with better adverse event rates showed different PROM response rates and improvement in PROM scores.

Methods: All primary THAs and TKAs performed in 19 Dutch hospitals between January 2017 and December 2019 were included. The HOOS and KOOS were used to assess the physical function after THA and TKA, respectively. Adverse events included were 1-year revision, 30-day readmission, 30-day complications, and long (i.e., >75th percentile) length-of-stay (LOS). A clinically relevant improvement was defined as at least a 10point decrease in HOOS and 9 points in KOOS scores. Patient-level associations were assessed using binary logistic regression models adjusted for patient characteristics, baseline PROM score, and clustering of patients within hospitals. Hospital-level correlations were assessed using Pearson correlations.

Results: 20,338 THA and 18,082 TKA procedures were included. HOOS and KOOS respondents mostly had more favourable adverse event rates than non-respondents. THA patients experiencing revision, complications, or long LOS were less likely to experience clinically relevant HOOS improvements (Odds ratios 0.11 [0.06-0.20], 0.44 [0.30-0.63] and 0.66 [0.50-0.88] respectively). TKA patients experiencing revision or long LOS were less likely to experience clinically relevant KOOS improvements (Odds ratios 0.26 [0.12-0.55] and 0.63 [0.50-0.80], respectively). Hospital performance on adverse events, and PROM response rates were not associated. Hospitals with better adverse event rates also had higher percentages of THA patients achieving clinically relevant HOOS improvements, while no associations were found for TKA patients.

Interpretation: Clinically relevant improvement in PROM scores are likely overestimated as PROM non-respondents had less favourable adverse event rates, which were associated with a lower likelihood to achieve a clinically relevant PROM improvement. Hospital differences are unlikely to be affected as hospital PROM response rates were not associated with adverse event rates.

Introduction

Clinical performance outcomes such as revision, readmission, complications, and length-of-stay (LOS) are unintended adverse events for patients and generally occur with low frequency after total hip and total knee arthroplasty (THA and TKA). (Van Schie et al. 2022) However, up to 10% and 20% of patients following THA and TKA, respectively, are dissatisfied with results, mainly related to continued pain and disability.(Rolfson et al. 2011, Dunbar et al. 2013) Patient Reported Outcome Measures (PROMs) measure the intended outcomes such as pain reduction, functionality improvement, and health-related quality of life gain and thereby complement the information provided by adverse events by identifying additional areas for improvement.(FDA 2009, Black 2013, Franklin et al. 2014, Rolfson et al. 2016, Johnston et al. 2019, Docter et al. 2021, Makhni 2021)

Similar to the need for high data completeness regarding adverse events to ensure no selection bias is at play, we also need high response rates of patients completing both pre-and postoperative questionnaires to calculate the improvement in PROMs. As for other national and regional arthroplasty registries, Dutch PROM response rates are low, with approximately 60% of patients completing the preoperative questionnaire for THA and TKA patients. In the absence of better response rates, we should at least gain insight into how those who complete questionnaires are a selection of all patients and in what direction this may cause bias (i.e., under- or overestimation of PROM improvement). Non-response bias is challenging to assess because, by definition, non-respondent data are not available and these non-respondents may differ systematically from respondents which would introduce bias. Previous studies have shown differences in patient characteristics, such as patients completing questionnaires being healthier, more likely to be white, having higher literacy rates, and lower rates of cognitive impairment, including dementia.(Norquist et al. 2000, Hutchings et al. 2013, Jahagirdar et al. 2013, Gibbons 2016, Kaur et al. 2023) However, these may still provide only a partial view by representing baseline patient characteristics rather than outcomes. Unintended adverse events are likely associated with improvement in PROM scores, and given the 97% completeness in registries for both THA and TKA, these seem well-suited to provide further insights into the relation between adverse event rates and postoperative PROM scores.(FDA 2009, Black 2013, Rolfson et al. 2016, Johnston et al. 2019, Bohm et al. 2021, Makhni 2021)

This study, therefore, aimed to provide insight into how improvement in PROM scores may be under-or overestimated relative to all patients who received a THA or TKA by 1) comparing PROM questionnaire respondents and non-respondents on their adverse event rates (i.e., revision, readmission, complications, and long LOS)

for THA and TKA, 2) examining whether patients experiencing adverse events had different improvement in PROM scores and 3) whether hospitals with better adverse event rates showed different PROM response rates and improvement in PROM scores.

Patients and Methods

Study design and setting

This observational study was performed in 19 hospitals (two university, four teaching, seven general, and six private clinics), reflecting the distribution across the Netherlands, using routinely collected data (i.e., data on revision and PROMs as well as patient characteristics) from the Dutch Arthroplasty Register (LROI). (van Steenbergen et al. 2015, Van Steenbergen et al. 2021, Van Schie et al. 2022) These were linked to hospital data on readmissions, complications, and LOS. These hospitals participated in a cluster randomised controlled trial to assess the effectiveness of a prospective multifaceted quality improvement intervention on patient outcomes after THA and TKA.(van Schie et al. 2023) The LUMC Medical Ethical Committee waived the need for ethical approval under Dutch law (CME, G18.140). This study was funded by the Van Rens Foundation (VRF-2018-001).

Participants

Anonymous data of all patients undergoing a primary THA or TKA between January 1, 2017, and December 31, 2019, were included from 20 Dutch hospitals. One general hospital did not provide PROMs data to the LROI and was therefore excluded, leaving data from 19 hospitals eligible for this study. Participating hospitals were comparable to all other Dutch hospitals in the distribution of median revision rates (1.7% vs 1.7% for THA, p=1.00 and 1.4% versus 0.9% for TKA, p=0.62).(van Schie et al. 2023)

Data source

Routinely submitted LROI data regarding patient characteristics, revision, and PROMs were used, which were linked for each patient to hospital data on readmission, complications, and LOS. The following data were provided by the LROI for each patient: (1) patient characteristics; age at surgery, gender, body mass index (BMI; kg/m2), current smoking status (yes/no), American Society of Anaesthesiologists (ASA) classification (I-IV), Charnley score (A/B1/B2/C/n/a), and indication for surgery (osteoarthritis/non-osteoarthritis); (2) whether a revision had taken place within 1 year after surgery; and (3) preoperative, 3-months postoperative (for THA), 6-months postoperative (for TKA) and 12-months postoperative PROM outcomes. The LROI data completeness is checked against Hospital Electronic Health Records and currently

exceeds 99% for primary procedures and 97% for revisions. Completeness is lower for PROMs data, currently 63% for preoperative PROMs for THA and 58% for TKA and lower for postoperative PROM questionnaires. LROI data were linked to hospital data by an IT specialist from each hospital. A clear definition for each adverse event was provided below to avoid measurement variability.

Adverse events

The one-year revision was calculated using LROI data based on primary surgery and revision dates. Other adverse events were calculated using the index hospitalisation during which the primary THA or TKA was performed. The outcomes were defined as:

- Revision: Any change, removal, or addition of any component within one year after primary surgery;
- Readmission: An admission within 30 days after discharge of the index hospitalisation;
- Complication: An adverse event other than revision during the index hospitalisation or within 30 days after discharge;
- Long LOS: LOS of the index hospitalisation is longer than the 75th percentile, based on all patients in the 19 hospitals, which was included because of possible hospital differences in sensitivity to report complications.

Patient-Reported Outcome Measures

The LROI routinely collects the HOOS-Physical function Short form (HOOS) and KOOS-Physical function Short form (KOOS), which are joint-specific PROMs and the most frequently collected PROMs in arthroplasty registries.(Bohm et al. 2021) The PROMs were collected preoperatively at the time of indication for surgery (with a maximum of 182 days before surgery), and 3 months (for THA), 6 months (for TKA), and 12 months postoperatively. The LROI does not compute an overall score when one or more questions are incomplete. The HOOS and KOOS contain 5 and 7 questions, respectively, to measure physical function.(Davis et al. 2008, Perruccio et al. 2008) Despite their brevity, these questionnaires have sufficient internal consistency and reliability and have been included in the standard set of outcome measures for hip and knee osteoarthritis of the International Consortium for Health Outcomes Measurement (ICHOM).(Braaksma et al. 2020) The scores range from 0 to 100, with higher scores reflecting more effort to perform activities (and thus worse function). Since patients are unlikely to notice a small improvement in PROM scores, a 10 points difference with the baseline PROM score was taken as a clinically relevant

improvement or worsening for the 3 months or 12 months postoperative HOOS score and a 9 points difference for the KOOS, as approximately half a standard deviation (sd) has been shown to reflect the minimally clinical relevant improvement in health-related quality of life for chronic diseases.(Norman et al. 2003)

Statistical analysis

Because both the pre-and postoperative PROM scores are needed to calculate an improvement in PROM scores, respondents on preoperative, 3 months (THA), 6 months (TKA), and 12 months (THA and TKA) postoperative PROM questionnaires were compared with non-respondents on their adverse event rates and patient characteristics using t-tests for continuous data and chi-square tests for categorical data.

Before examining patient-level and hospital-level correlations, missing data for patient characteristics (occurring in less than 5% of patients) were imputed. Missing data were considered missing at random and imputed using multiple imputations for ten rounds with predictive mean matching as the underlying model. All patient characteristics (i.e., age, gender, BMI, smoking, ASA score, Charnley score, and diagnosis), adverse events, and preoperative PROM scores were used as predictors, but only patient characteristics were imputed.

Patient-level associations between adverse events, and clinically relevant improvement in PROM scores were assessed using binary logistic regression models, separately for THA and TKA. The models included clinically relevant PROM improvement (yes/no) as the dependent variable and the adverse events as independent variables. For 30-day readmission, 30-day complications, and long LOS, we used improvement at the first postoperative PROM measurement (i.e., 3 months postoperative for THA and 6 months for TKA) as the dependent variable as this time point is more likely to reflect the impact of surgery. For revision, the 1 year postoperative PROM measurement was used because the exact revision date was unknown to us as researchers, which could occur before or after the first postoperative PROM measurement. All models were adjusted for the preoperative PROM score and all patient characteristics, as these have been shown to predict postoperative PROM scores.(Hofstede et al. 2016, Rolfson et al. 2016) Hospital was included as a random effect to account for the clustering of patients within hospitals.

Hospital-level correlations between standardised adverse event rates and standardised rates of patients with clinically relevant PROM improvements were estimated using Pearson correlation. We also examined correlations between standardised adverse event rates and PROM response rates. As mentioned above, the first postoperative PROM measurement was used for readmission, complications, and long LOS, and

the 12 months postoperative measurement for revision. Standardised rates were used to adjust for the possible differences between hospitals in their patient-mix, which would influence their risk on adverse events. Therefore, the observed (O) number of events in a hospital was divided by the expected (E) number based on the hospital's patient-mix to give an O/E ratio per hospital for each adverse event. For each patient, the expected adverse event risk was calculated using logistic regression analysis among all patients from all hospitals, including the patient characteristics described above as independent variables and the adverse event as the dependent variable.(Berliner et al. 2016, Hofstede et al. 2016, Rolfson et al. 2016, van Schie et al. 2020) The expected number for a hospital was then obtained by summing all patients' expected probabilities. The expected number of patients with a clinically relevant improvement was calculated similarly with the baseline PROM score additionally included as an independent variable.

All p-values were two-sided, and values below 0.05 were considered statistically significant in all analyses. Analyses were performed using SPSS (version 25; IBM) and STATA (version 14; StataCorp).

Results

20,338 primary THA procedures and 18,082 primary TKA procedures from 19 hospitals were included. Less than 10% of THA and TKA patients had missing data on readmission, complications, and LOS. Revision, readmission, complication, and long LOS rates were lower for TKA than THA (Table 1). LOS data were not normally distributed, making it challenging to create equal quartiles, so the closest integer value was chosen, which resulted in above 4 days being defined as long LOS for both THA and TKA. The mean LOS was 3.1 days (sd 2.5) for THA and 3.2 days (sd 1.9) for TKA. Revision rates were comparable to those observed among all Dutch hospitals. (van Schie et al. 2023)

Table 1 Clinical outcomes and patient characteristics after THA and TKA in 19 Dutch hospitals during 2017-2019.

	THA (*n*=20,338)	TKA (*n*=18,082)
Clinical outcomes		
1-year revision (%)	376 (1.8)	237 (1.3)
30-day readmission (%)	724 (3.9)	551 (3.4)
30-day complication (%)	735 (3.9)	417 (2.5)
Long LOS (%)	2,205 (11.8)	1,778 (10.9)
Patient characteristics		
Mean age in years (SD)	68.36 (10.3)	68.10 (8.8)
Sex, female (%)	13,029 (64.1)	11,199 (61.9)
BMI (SD)	26.94 (4.4)	29.29 (4.9)
Current Smokers (%)	2,122 (10.4)	1,524 (8.4)
ASA classification (%)		
• ASA I	3,853 (18.9)	2,507 (13.9)
• ASA II	12,622 (62.1)	11,997 (66.4)
• ASA III-IV	3,860 (19.0)	3,575 (19.8)
Charnley score (%)		
• A	8,158 (41.8)	6,587 (36.6)
• B1	6,241 (32.0)	6,743 (37.5)
• B2	4,502 (23.1)	3,994 (22.2)
• C	630 (3.2)	663 (3.7)
Diagnosis (%)		
• Osteoarthritis	18,019 (88.6)	17,510 (96.9)
• Non-Osteoarthritis	2,315 (11.4)	569 (3.1)

Less than 10% of the values for clinical outcomes were missing, and less than 5% for patient characteristics.
ASA=American Society of Anaesthesiologists; BMI=body mass index; LOS=length of stay; SD=standard deviation.

The mean HOOS and KOOS scores significantly improved postoperatively regardless of whether adverse events occurred (Figure 1). However, patients undergoing revision had significantly worse postoperative PROM scores than patients without revision. Comparable results were found for readmission, complications, and long LOS. For THA patients, 86% had clinically relevant improvements in the HOOS, and 2% had worsened scores at 3 months postoperatively. At 12 months, it was 90% and 2%, respectively. For TKA patients, 73% had clinically relevant improvement in the KOOS, and 3% had worsened scores at 6 months postoperative. At 12 months, it was 78% and 3%, respectively.

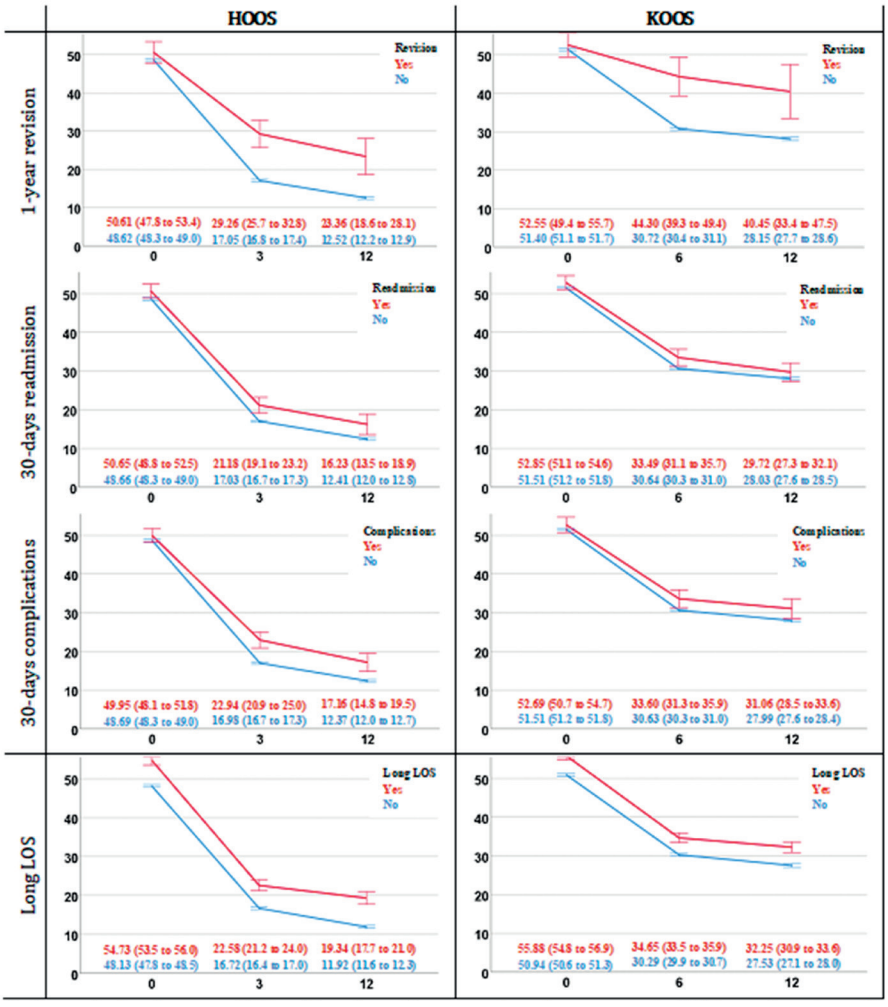


Figure 1 Mean PROM scores over time for clinical outcomes.
The line graphs show the mean and 95% confidence interval for preoperative and two postoperative PROM scores for patients with and without revisions, readmission, complications and long LOS.
X-axis: 0=preoperative; 3=3-months postoperative; 6=6-months postoperative; 12=12-months postoperative.
Y-axis: HOOS or KOOS scores.

Table 2 Respondents versus non-respondents of the HOOS.

	Preoperative AND postoperative HOOS					
Respondent preoperative and 3 months postoperative (*n*=7,731; 38.0%)	**Non-respondent preoperative and 3 months postoperative (*n*=12,607; 62.0%)**	***p*-value**	**Respondent preoperative and 12 months postoperative (*n*=5,382; 26.5%)**	**Non-respondent preoperative and 12 months postoperative (*n*=14,956; 73.5%)**	***p*-value**	
Clinical outcomes						
1-year revision (%)	93 (1.2)	283 (2.2)	<0.01	66 (1.2)	310 (2.1)	<0.01
30-day readmission (%)	252 (3.3)	472 (3.7)	<0.01	166 (3.1)	558 (3.7)	<0.01
30-day complication (%)	239 (3.1)	496 (3.9)	<0.01	170 (3.2)	565 (3.8)	<0.01
Long LOS (%)	566 (7.3)	1639 (13.0)	<0.01	403 (7.5)	1802 (12.0)	<0.01
Patient characteristics						
Mean age in years (SD)	68.98 (9.7)	67.98 (10.7)	<0.01	69.02 (9.9)	68.12 (10.5)	<0.01
Gender, female (%)	4,876 (63.1)	8,153 (64.7)	0.02	3,379 (62.8)	9,650 (64.5)	0.02
Mean BMI in kg/m^2 (SD)	27.07 (4.4)	26.85 (4.5)	<0.01	27.05 (4.4)	26.89 (4.4)	0.02
Current Smokers (%)	719 (9.3)	1403 (11.1)	<0.01	499 (9.3)	1623 (10.9)	<0.01
ASA classification (%)						
• ASA I	1,376 (17.8)	2,477 (19.6)		944 (17.5)	2,909 (19.5)	
• ASA II	4,983 (64.5)	7,639 (60.6)	<0.01	3,556 (66.1)	9,066 (60.6)	<0.01
• ASA III-IV	1,370 (17.7)	2,490 (19.8)		880 (16.4)	2,980 (19.9)	
Charnley score (%)						
• A	3,078 (39.8)	5,080 (40.3)		2,161 (40.2)	5,997 (40.1)	
• B1	2,694 (34.8)	3,547 (28,1)	<0.01	1,828 (34.0)	4,413 (29.5)	<0.01
• B2	1,644 (21.3)	2,858 (22.7)		1,161 (21.6)	3,341 (22.3)	
• C	217 (2.8)	413 (3.3)		158 (2.9)	472 (3.2)	
Indication (%)						
Osteoarthritis	7,278 (94.1)	10,741 (85,2)	<0.01	5,061 (94.0)	12,958 (86.6)	<0.01
Non-Osteoarthritis	450 (5.9)	1865 (14.8)		318 (6.0)	1197 (13.4)	

Percentages might not sum to 100 because of rounding.

ASA=American Society of Anaesthesiologists; BMI=body mass index; LOS=length of stay; N/A=not applicable; SD=standard deviation.

Table 3 Respondents versus non-respondents of the KOOS.

	Preoperative AND postoperative KOOS					
Respondent preoperative and 6 months postoperative (*n*=5,519; 30.5%)	**Non-respondent preoperative and 6 months postoperative (*n*=12,563; 69.5%)**	***p*-value**	**Respondent preoperative and 12 months postoperative (*n*=4,319; 23.9%)**	**Non-respondent preoperative and 12 months postoperative (*n*=13,763; 76.1%)**	***p*-value**	
Clinical outcomes						
1-year revision (%)	51 (0.9)	61 (1.4)	0.03	35 (0.8)	202 (1.5)	<0.01
30-day readmission (%)	155 (2.8)	169 (3.8)	<0.01	115 (2.7)	436 (3.2)	<0.01
30-day complication (%)	139 (2.6)	107 (2.4)	<0.01	108 (2.6)	309 (2.2)	<0.01
Long LOS (%)	577 (10.5)	574 (12.9)	<0.01	485 (11.2)	1,293 (9.4)	<0.01
Patient characteristics						
Mean age in years (SD)	68.50 (8.5)	67.93 (8.9)	<0.01	68.58 (8.5)	67.95 (8.8)	<0.01
Gender, female (%)	3,393 (61.5)	7,806 (62.1)	0.40	2,641 (61.1)	8,558 (62.2)	0.22
Mean BMI in kg/m^2 (SD)	29.43 (4.8)	29.22 (4.8)	<0.01	29.38 (4.8)	29.26 (4.8)	0.15
Current Smokers (%)	451 (8.2)	1,073 (8.5)	0.14	352 (8.2)	1,172 (8.5)	0.35
ASA classification (%)						
• ASA I	666 (12.1)	1,841 (14.7)		525 (12.2)	1982 (14.4)	
• ASA II	3,718 (67.4)	8,279 (65.9)	<0.01	2,937 (68.0)	9,060 (65.8)	<0.01
• ASA III-IV	1,135 (20.6)	2,440 (19.4)		857 (19.8)	2,718 (19.7)	
Charnley score (%)						
• A	2,103 (38.1)	4,484 (35.7)		1,627 (37.7)	4,960 (36.0)	
• B1	2,024 (36.7)	4,719 (37.6)	<0.01	1,586 (36.7)	5,157 (37.5)	<0.01
• B2	1,147 (20.8)	2,847 (22.7)		913 (21.1)	3,081 (22.4)	
• C	227 (4.1)	436 (3.5)		175 (4.1)	488 (3.5)	
Indication (%)						
• Osteoarthritis	5,374 (97.4)	12,136 (96.6)	0.02	4,205 (97.4)	1,3305 (96.7)	0.04
• Non-Osteoarthritis	145 (2.6)	424 (3.4)		114 (2.6)	455 (3.3)	

Percentages might not sum to 100 because of rounding.

ASA=American Society of Anaesthesiologists; BMI=body mass index; LOS=length of stay; N/A=not applicable; SD=standard deviation.

Pre- and 3 months postoperative HOOS questionnaires were completed by 7,731 (38%) THA patients, and 5382 (27%) completed both pre-and 12 months postoperative questionnaires (Table 2). Patients who completed the pre-and postoperative HOOS questionnaire differed from those not completing it in having more favourable adverse event rates. Considering the 3206 patients with at least one adverse event, 2212 (69%) did not return both the preoperative and 3 months postoperative questionnaires compared with 59% of patients experiencing no adverse event, and thus were more likely to be a non-respondent. For not returning the preoperative and 12 months questionnaires, these figures were 80% and 71% respectively. Although absolute differences between respondents and non-respondents for most patient characteristics were small, they were nevertheless significant likely due to the large sample size. For the KOOS, pre- and 6 months postoperative questionnaires were completed by 5,519 (31%) TKA patients, and 4319 (24%) completed both pre-and 12 months postoperative questionnaires (Table 3). Patients who completed the pre- and postoperative KOOS questionnaires differed from those not completing it in having more favourable revision and readmission rates, but comparable complication rates and less favourable long LOS rates. Of 2549 patients with at least one adverse event, 1988 (78%) did not return both preoperative and 6 months postoperative questionnaires compared with 67% of patients experiencing no adverse event. For not returning the preoperative and 12 months postoperative questionnaires, these figures were and 82% respectively 74%. Significant differences were found for most patient characteristics, except gender, smoking and BMI.

Patients experiencing revision, complications, or long LOS were less likely to achieve a clinically relevant improvement in the HOOS (89%, 56%, and 34% less likely, respectively), with the association for readmission going in the same direction but non-significant (Table 4). Patients experiencing revision or long LOS were less likely to achieve a clinically relevant improvement in the KOOS (74% and 37% less likely, respectively), with associations for readmission and complications in the same direction but non-significant.

Table 4 Patient-level associations between clinical outcomes and clinically relevant improvement in PROM scores.

	HOOS	KOOS
	OR (95% CI)	OR (95% CI)
1-year revision	0.11 (0.06 – 0.20)	0.26 (0.12 – 0.55)
30-days readmission	0.71 (0.48 – 1.06)	0.70 (0.46 – 1.05)
30-days complications	0.44 (0.30 – 0.63)	0.79 (0.51 – 1.23)
Long LOS	0.66 (0.50 – 0.88)	0.63 (0.50 – 0.80)

A difference of 10 points was taken as a clinically relevant improvement for the HOOS and 9 points for the KOOS. LOS=length of stay.

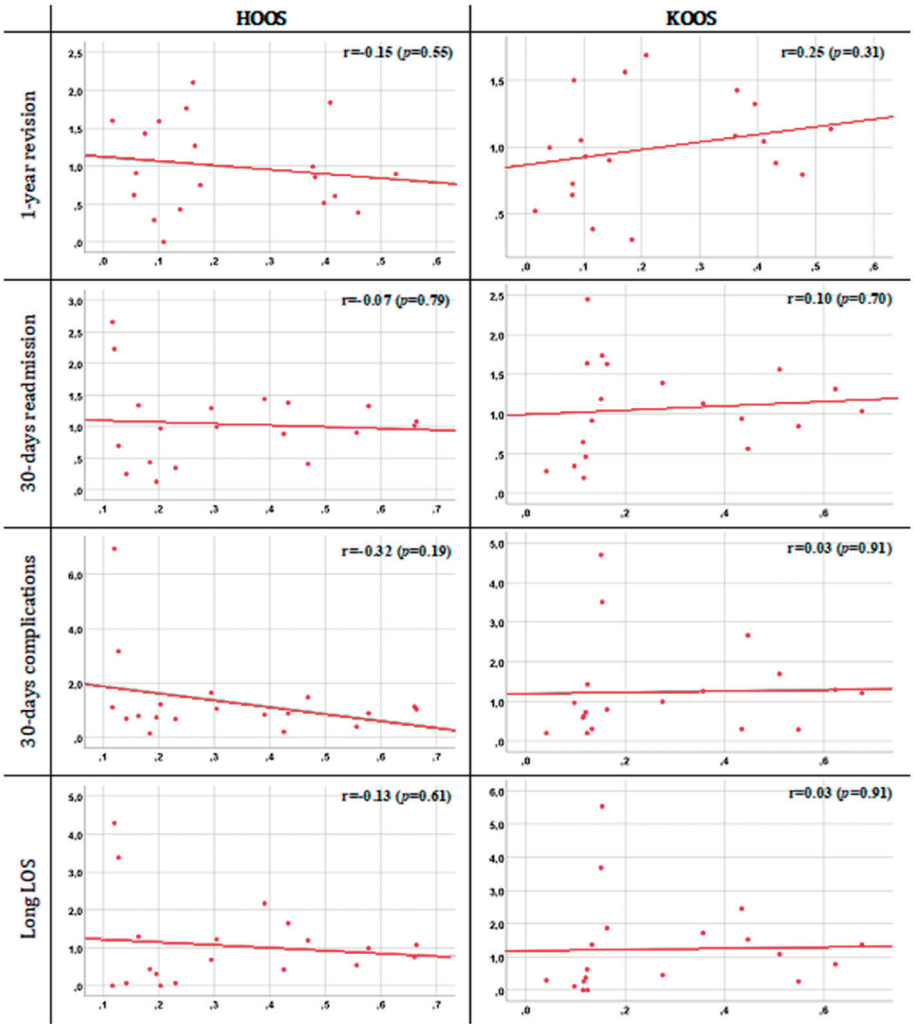


Figure 2 Hospital-level correlations between clinical outcomes and PROM response rates.
The y-axis shows the standardised (O/E ratio) clinical outcomes, with a higher value representing hospitals with more events than expected based on patient characteristics (worse performance).
The x-axis shows the PROM response rates for the hospitals.
The strength of correlations is defined as: ≤0.35 weak >0.35-0.67 moderate and >0.67 strong (47).
LOS=Length of stay; O/E=Observed divided by expected; r=correlation coefficient.

Considerable between-hospital variation as shown by the interquartile range was found in (standardised) adverse event rates, patient characteristics, response rates on PROM questionnaires, mean PROM scores, and to a smaller extent in the standardised percentage of patients achieving clinically relevant improvement in PROM scores (Table 5). There was no association between hospital performance on standardised adverse event rates and PROM response rates (Figure 2). Hospitals with lower (i.e.,

better) standardised readmission, complications, and long LOS rates also had higher standardised percentages of patients with clinically relevant improvements in the HOOS (r=-0.58 p<0.01, r=-0.82 p<0.01 and r=-0.65 p<0.01 respectively), with correlations for revision in the same direction but non-significant (r=-0.19, p=0.44) (Figure 3). For TKA, no significant correlations were shown with improvement in the KOOS.

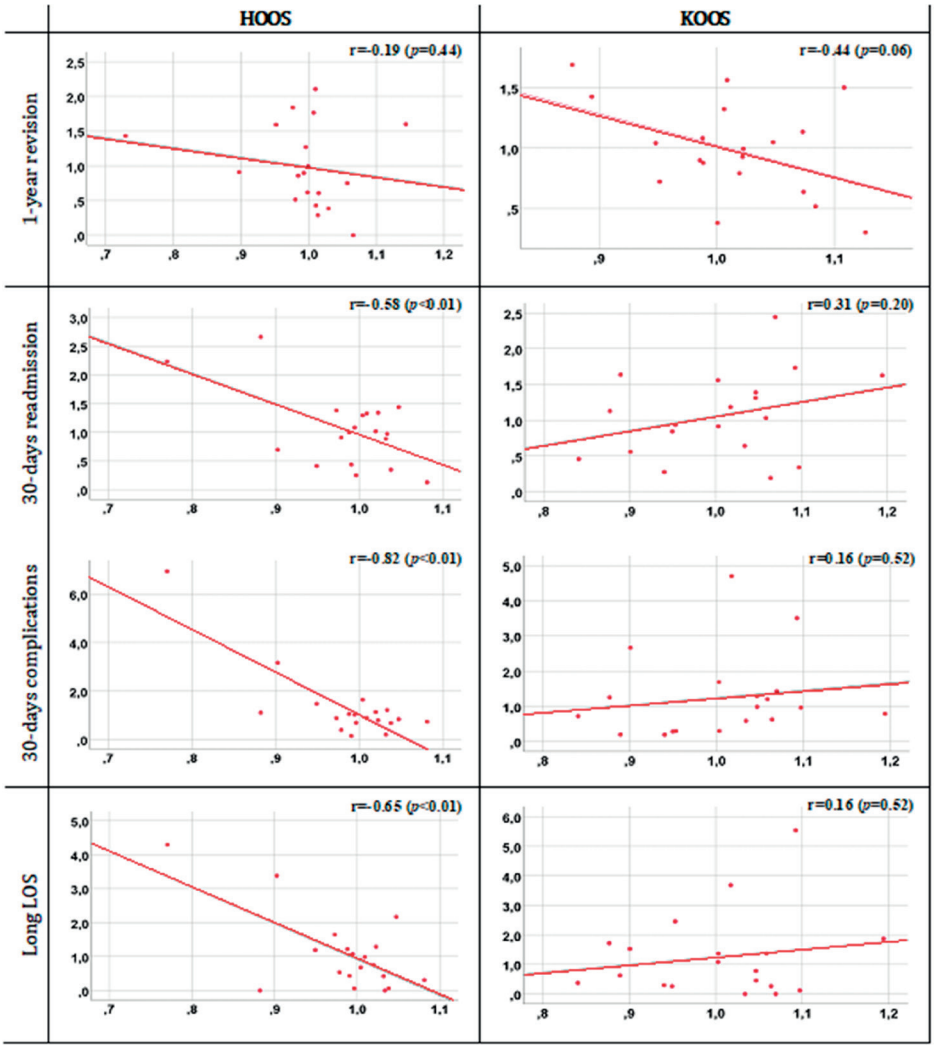


Figure 3 Hospital-level correlations between clinical outcomes and clinically relevant improvement.The y-axis shows the standardised (O/E ratio) clinical outcomes, with a higher value representing hospitals with more events than expected based on patient characteristics (worse performance).
The x-axis shows the standardised (O/E ratio) percentages of patients with clinically relevant improvement in PROM scores, with a higher value representing hospitals with more patients with clinically relevant improvement than expected based on patient characteristics and preoperative PROM scores (better performance).
The strength of correlations is defined as: ≤0.35 weak >0.35-0.67 moderate and >0.67 strong (47).
LOS=Length of stay; O/E=Observed divided by expected; r=correlation coefficient.

Table 5 Hospital-level variation between 19 Dutch hospitals performing THA and TKA during 2017-2019.

	THA (*n*=19 hospitals)	TKA (*n*=19 hospitals)
	Median (IQR)	Median (IQR)
Procedures (*n*)	1,144 (615-1,388)	838 (580-1,378)
Clinical outcomes		
1-year revision (%)	1.7 (0.9-2.7)*	1.3 (0.9-1.7)**
Standardised	*0.9 (0.6-1.5)*	*1.0 (0.8-1.2)*
30-day readmission (%)	4.0 (2.0-6.0)	3.9 (1.8-5.5)
Standardised	*1.0 (0.6-1.3)*	*1.0 (0.6-1.5)*
30-day complication (%)	3.7 (2.3-4.5)	2.0 (1.0-3.5)
Standardised	*0.89 (0.4-1.3)*	*1.0 (0.4-1.4)*
Long LOS (%)	9.0 (2.6-19.8)	9.4 (2.5-20.9)
Standardised	0.8 *(0.4-1.3)*	*0.8 (0.3-1.6)*
Patient characteristics		
Mean age (years)	69.1 (64.3-70.0)	68.9 (66.3-69.6)
Sex, female (%)	63.7 (62.1-65.1)	62.7 (59.9-64.0)
Mean BMI	27.1 (26.7-27.2)	29.5 (28.6-30.0)
Current smokers (%)	10.7 (9.2-12.2)	9.1 (7.5-10.3)
ASA classification (%)		
• ASA I	14.6 (9.8-19.9)	9.8 (7.7-14.3)
• ASA II	62.7 (58.7-66.8)	67.3 (60.9-71.0)
• ASA III-IV	22.3 (7.1-27.2)	22.7 (13.1-30.2)
Charnley score (%)		
• A	43.1 (37.8-47.1)	38.7 (28.2-41.5)
• B1	30.4 (28.1-33.0)	32.6 (31.6-42.1)
• B2	22.6 (19.4-24.1)	22.5 (19.0-24.4)
• C	2.9 (1.2-5.3)	2.2 (1.2-5.2)
Indications (%)		
• Osteoarthritis	89.4 (83.6-91.3)	97.1 (95.8-97.8)
• Non-Osteoarthritis	10.6 (7.7-16.8)	2.9 (2.3-4.4)
HOOS/KOOS response rate		
Pre- and 3-months postop (%)	29.5 (17.4-45.9)	N/A
Pre- and 6-months postop (%)	N/A	15.4 (12.2-44.4)
Pre- and 12-months postop (%)	16.2 (9.6-39.3)	17.1 (8.9-38.7)
Mean HOOS/KOOS scores		
3-months postoperative	17.5 (15.8-18.9)	N/A
6-months postoperative	N/A	31.0 (29.5-34.1)
12-months postoperative	12.8 (9.7-14.4)	28.0 (24.9-29.7)
Standardised percentage of patients with clinically relevant improvement		
Pre- and 3-months postop	1.00 (0.98-1.03)	N/A
Pre- and 6-months postop	N/A	1.02 (0.95-1.06)
Pre- and 12-months postop	1.00 (0.98-1.01)	1.01 (0.99-1.06)

The value under "Median (IQR)" indicates the median hospital's mean or percentage. Standardised clinical outcomes were adjusted for patient characteristics (age, sex, BMI, current smokers, ASA classification, Charnley score, and diagnosis). Standardised percentages of patients with clinically relevant improvements were adjusted for patient characteristics and preoperative PROM scores.

*Dutch 1-year revision percentage for THA was 1.7%.(20)

**Dutch 1-year revision percentage for TKA was 0.9%.(20)

ASA=American Society of Anaesthesiologists; BMI=body mass index; IQR=interquartile range.

Discussion

Interpretation of the results

To investigate whether missing PROM data for THA and TKA may result in under- or overestimation of PROM improvement scores, we used adverse event rates to examine how these differed between respondents and non-respondents and their association with PROM improvement scores. We found that PROM respondents had more favourable adverse event rates than non-respondents for the HOOS, and also for revision and readmission for the KOOS. Patients experiencing revision, complications, or long LOS after THA were less likely to have a clinically relevant improvement in the HOOS, with results for readmission in the same direction but non-significant. The strongest association was found for revision, suggesting that quality improvement initiatives should focus most on reducing revision rates to benefit patient care. Comparable results were found for the KOOS. Since patients completing PROM questionnaires generally have more favourable adverse event rates than non-respondents, this means that part of the patients who were less likely to achieve a clinically relevant improvement would be missed, and therefore that PROM improvement scores are likely overestimated. This means, for example, that less than our estimated 90% of THA and 78% of TKA patients had a clinically relevant improvement at 12 months postoperatively in the HOOS and KOOS, respectively.

Hospital-level results for THA and TKA patients showed no association between adverse event rates and PROMS response rates, and for THA that better adverse event rates were associated with higher percentages of patients achieving clinically relevant improvement in the HOOS. Despite the large between-hospital variation in PROM response rates, which in general tends to remove those with less favourable adverse events, there was no association between hospital adverse events and their PROM response rates. In other words, hospitals with different PROM response rates do not systematically have different adverse event rates, suggesting that hospital differences in PROM improvement scores are likely unaffected.

Strengths and limitations

This study is the first to report associations between a set of commonly used adverse events and the likelihood of achieving a clinically relevant improvement in physical functioning after THA and TKA. Given the observed associations, it seems likely that initiatives to improve the quality of care by reducing revision, readmission, complications, and long LOS rates will be accompanied by increased percentages of patients achieving clinically relevant improvement in physical functioning, but also that improvement after THA and TKA is likely overestimated. However, we found no evidence that this has affected the estimated between-hospital differences

in PROM improvement rates. Some limitations should be noted. First, data were obtained from 19 hospitals rather than all Dutch hospitals performing THA and TKA in the Netherlands. However, these hospitals reflected the national distribution of hospital types (i.e., university, teaching, general, and private clinics) and had comparable revision rates to national data, so it seems unlikely that the selection of hospitals would affect our results.(van Schie et al. 2023) Second, this study can only provide indirect evidence due to the lack of information about changes in PROM scores among non-respondents. Third, non-response is known to be affected by patient characteristics. So, if non-respondents would systematically have more favourable patient characteristics (e.g., younger and with better health status) known to be related to more improvement in PROMs, then this could be the case for the non-respondents without adverse events. However, Tables 2 and 3 did not indicate such systematic differences, as respondents were older, but less often smoked and ASA 3+. Potential risk factors such as mental health and emotional health status were not included to adjust the associations between adverse events and improvement in PROM scores as these variables are not collected by the LROI.(Ayers et al. 2013, Giesinger et al. 2013) Although the available patient characteristics are likely the most relevant, some residual confounding may remain.(Hofstede et al. 2016, Rolfson et al. 2016) We could not think of other reasons why non-respondents would be expected to experience a greater or lesser improvement in PROM scores after an adverse event. Fourth, patients undergoing THA or TKA may not improve as much in their PROM scores if another joint is also affected. The latter will influence associations when the prevalence of such patients is unevenly distributed among patients with or without adverse events.

Comparison to the literature

This study showed that 1 year after surgery, 90% of THA patients achieved a clinically relevant improvement and 78% of TKA patients. However, 2% of THA and 3% of TKA patients reported a clinically relevant worsening, with respectively 8% and 19% showing no relevant change. The lower percentages of improvement in PROM scores for TKA than THA are consistent with earlier studies from our group and others in Sweden and the United States.(Rolfson et al. 2011, Keurentjes et al. 2013, Keurentjes et al. 2013, Keurentjes et al. 2014, Lyman et al. 2018) It should be noted that while patients with worse preoperative PROM scores may improve more, they do not achieve the same postoperative level as patients with better preoperative function scores.(Hofstede et al. 2016) Additionally, previous studies have reported higher satisfaction rates for THA than TKA, which would seem consistent with more patients achieving clinically relevant improvement in PROM scores.(Bourne et al. 2010, Dunbar et al. 2013, Haanstra et al. 2015, Tilbury et al. 2016) Satisfaction rates may improve further by addressing preoperative expectations, a significant predictor

for dissatisfaction following TKA.(Dunbar et al. 2013)In the Dutch registry, 63% and 58% of patients completed the preoperative HOOS and KOOS questionnaire, respectively, which are low compared with the Scandinavian registries but higher than the Italian Register of the Orthopaedic Prosthetic Implants and the Michigan Arthroplasty Register.(Rolfson et al. 2016) This would suggest that the extent of overestimation in PROM improvement is likely smaller for countries with better response rates, provided that adverse event rates are similar. In line with our results, one study including THA patients found that non-response during follow-up was not at random; non-respondents had significantly lower PROM scores at the previous time point than respondents, so indicating that patients reporting good outcomes were overrepresented.(Imam et al. 2014) In another study, TKA patients respondents reported a higher mean Knee Society Score (KSS), mean function score and lower mean pain score than non-respondents.(Kim et al. 2004) Comparable results were reported in another study, including patients after shoulder replacement, and another study identified a trend of worse outcomes for non-respondents (Norquist et al. 2000, Polk et al. 2013).

The PROMs Working Group of the ISAR stated that a response rate above 80% is recommended for reliable outcome assessment but proposes a 60% threshold for an acceptable response rate.(Rolfson et al. 2016) Only six of the 16 arthroplasty registries collecting PROMs capture >80% of their preoperative and postoperative PROMs; the remaining registries reported response rates less than 60%.(FDA 2009, Black 2013, Rolfson et al. 2016, Johnston et al. 2019, Bohm et al. 2021, Makhni 2021) Another study stated that a 100% response rate is needed to adequately evaluate PROM difference scores because of a change in the distribution of predictors when a selection of patients is analysed, resulting in unreliable outcomes.(Pronk et al. 2020) This seems only feasible if PROM collection is mandatory and becomes part of the doctor-patient conversations on THA and TKA care goals.(Murthy et al. 2022)

Conclusions

Respondents to PROM questionnaires less often experienced adverse events. This likely results in an overestimation of the clinically relevant improvement in PROMs as adverse events were associated with a lower likelihood to achieve a clinically relevant PROM improvement. Hospital differences in PROM response rates were not associated with differences in adverse event rates, suggesting that estimated between-hospital differences in PROM improvement are likely unaffected. Ultimately, patients undergo THA and TKA to improve their function, and our results suggest that this may be further improved by continued efforts to reduce adverse event rates.

Author contributions

PvS: Conceptualisation and design of manuscript; Methodology; Formal analysis; Data curation; Writing (original draft); Project administration. LvB: Conceptualisation and manuscript design; Writing (review & editing); Supervision. TZ: Conceptualisation and manuscript design; Data curation; Writing (original draft). TG: Writing (review & editing). RN: Conceptualisation and design of manuscript; Writing (review & editing); Supervision; Funding acquisition. PM: Conceptualisation and design of manuscript; Conceptualisation of the manuscript; Methodology; Formal analysis; Writing (review & editing); Supervision; Project administration; Funding acquisition.

Acknowledgments

In addition to the authors, the IQ Joint study group consists of (in alphabetic order): Antonius Hospital, Sneek (S.T. Hokwerda, MD); Bergman Clinics (P.M. van Kampen, PhD); Bergman Clinics, Arnhem (I. Buchholz, MD); Bergman Clinics, Breda (J. Schrier, MD); Bergman Clinics, Delft (F. de Graaff, MD); Bergman Clinics, Naarden (H. Bouma, MD); Bergman Clinics, Rijswijk (T. Hogervorst, MD, PhD and J. Wolkenfelt, MD); Bergman Clinics, Rotterdam (M. Vischjager, MD, PhD); Catharina Hospital, Eindhoven (R.W.T.M. van Kempen, MD); Dijklander Hospital, Hoorn (G.C. Huitema, MD); Dijklander Hospital, Hoorn (L. de Vries, PhD); Elisabeth-TweeSteden Hospital (T. Gosens, MD, PhD); Gelderse Vallei Hospital, Ede (W. Beijneveld, MD); Maxima Medical Centre, Eindhoven (M. van den Besselaar, MD); Medical Spectrum Twente, Enschede (W. Verra, MD, PhD); OLVG, Amsterdam (R.W. Poolman, MD, PhD); OLVG, Amsterdam (V.A. Scholtes, PhD); Sint Anna Hospital, Geldrop (W. van der Weegen, PhD); Sint Franciscus Hospital, Schiedam (A. Polak, MD); Tjongerschans Hospital, Heerenveen (M. Mulder, PhD); University of Groningen University Medical Center Groningen, Groningen (M. Stevens, PhD).

We thank the Dutch Arthroplasty Register for providing data for this study (L.N. van Steenbergen, PhD)

References

American Academy of Orthopaedic Surgeons (AAOS) Joint Replacement Registry. https://www.aaos.org/registries/registry-program/american-joint-replacement-registry/.

ClinicalTrial.gov. https://clinicaltrials.gov/ct2/show/NCT04055103?term=Arthroplasty&cntry=NL&city=Leiden&draw=2&rank=2.

Danish Hip Arthroplasty Register (DHAR). Annual report 2020. http://danskhoftealloplastikregister.dk/wp-content/uploads/2020/11/dhr-aarsrapport-2020_til_offentliggoerelse-1.pdf.

Dutch Arthroplasty Register (LROI). Data completeness. https://www.lroi-report.nl/data-quality/completeness/.

Dutch Arthroplasty Register (LROI). Online Annual Report. https://www.lroi-report.nl/.

Dutch PROM response rates for THA. https://www.lroi-report.nl/hip/proms/response/.

Dutch PROM response rates for TKA. https://www.lroi-report.nl/knee/proms/response/.

International Consortium for Health Outcomes Measurement. Hip and Knee Osteoarthritis Standard Set. https://connect.ichom.org/standard-sets/hip-knee-osteoarthritis/.

National Joint Registry (NJR). Annual report 2021. https://reports.njrcentre.org.uk/Portals/0/PDF-downloads/NJR%2018th%20Annual%20Report%202021.pdf.

Swedish Hip Arthroplasty Register (SHAR). Annual Report 2018. https://registercentrum.blob.core.windows.net/shpr/r/Arsrapport_2018_Hoftprotes_ENG_26mars_Final-rJepCXNsLI.pdf.

Ayers DC, Franklin PD, Ring DC. The role of emotional health in functional outcomes after orthopaedic surgery: extending the biopsychosocial model to orthopaedics: AOA critical issues. J Bone Joint Surg Am. 2013;95(21):e165.

Berliner JL, Brodke DJ, Chan V, SooHoo NF, Bozic KJ. John Charnley Award: Preoperative Patient-reported Outcome Measures Predict Clinically Meaningful Improvement in Function After THA. Clin Orthop Relat Res. 2016;474(2):321-9.

Black N. Patient reported outcome measures could help transform healthcare. BMJ. 2013;346:f167.

Bohm ER, Kirby S, Trepman E, Hallstrom BR, Rolfson O, Wilkinson JM, Sayers A, Overgaard S, Lyman S, Franklin PD, Dunn J, Denissen G, A WD, Ingelsrud LH, Navarro RA. Collection and Reporting of Patient-reported Outcome Measures in Arthroplasty Registries: Multinational Survey and Recommendations. Clin Orthop Relat Res. 2021;479(10):2151-66.

Bourne RB, Chesworth B, Davis A, Mahomed N, Charron K. Comparing patient outcomes after THA and TKA: is there a difference? Clin Orthop Relat Res. 2010;468(2):542-6.

Braaksma C, Wolterbeek N, Veen MR, Prinsen CAC, Ostelo R. Systematic review and meta-analysis of measurement properties of the Hip disability and Osteoarthritis Outcome Score - Physical Function Shortform (HOOS-PS) and the Knee Injury and Osteoarthritis Outcome Score - Physical Function Shortform (KOOS-PS). Osteoarthritis Cartilage. 2020;28(12):1525-38.

Cabitza F, Dui LG. Collecting Patient Reported Outcomes in the Wild: Opportunities and Challenges. Stud Health Technol Inform. 2018;247:36-40.

Davis AM, Perruccio AV, Canizares M, Tennant A, Hawker GA, Conaghan PG, Roos EM, Jordan JM, Maillefert JF, Dougados M, Lohmander LS. The development of a short measure of physical function for hip OA HOOS-Physical Function Shortform (HOOS-PS): an OARSI/OMERACT initiative. Osteoarthritis Cartilage. 2008;16(5):551-9.

Docter S, Fathalla Z, Lukacs MJ, Khan MCM, Jennings M, Liu SH, Dong S, Getgood A, Bryant DM. Interpreting Patient-Reported Outcome Measures in Orthopaedic Surgery: A Systematic Review. J Bone Joint Surg Am. 2021;103(2):185-90.

Dunbar MJ, Richardson G, Robertsson O. I can't get no satisfaction after my total knee replacement: rhymes and reasons. Bone Joint J. 2013;95-b(11 Suppl A):148-52.

FDA. Guidance for Indusctry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. 2009.

Franklin PD, Lewallen D, Bozic K, Hallstrom B, Jiranek W, Ayers DC. Implementation of patient-reported outcome measures in U.S. Total joint replacement registries: rationale, status, and plans. J Bone Joint Surg Am. 2014;96 Suppl 1(Suppl 1):104-9.

Gibbons E. Patient-reported outcome measures and the evaluation of services. Health Serv Deliv Res. 2016;4:55-68.

Giesinger JM, Kuster MS, Behrend H, Giesinger K. Association of psychological status and patient-reported physical outcome measures in joint arthroplasty: a lack of divergent validity. Health Qual Life Outcomes. 2013;11:64.

Haanstra TM, Tilbury C, Kamper SJ, Tordoir RL, Vliet Vlieland TP, Nelissen RG, Cuijpers P, de Vet HC, Dekker J, Knol DL, Ostelo RW. Can Optimism, Pessimism, Hope, Treatment Credibility and Treatment Expectancy Be Distinguished in Patients Undergoing Total Hip and Total Knee Arthroplasty? PLoS One. 2015;10(7):e0133730.

Hofstede SN, Gademan MG, Vliet Vlieland TP, Nelissen RG, Marang-van de Mheen PJ. Preoperative predictors for outcomes after total hip replacement in patients with osteoarthritis: a systematic review. BMC Musculoskelet Disord. 2016;17:212.

Hutchings A, Grosse Frie K, Neuburger J, van der Meulen J, Black N. Late response to patient-reported outcome questionnaires after surgery was associated with worse outcome. J Clin Epidemiol. 2013;66(2):218-25.

Imam MA, Barke S, Stafford GH, Parkin D, Field RE. Loss to follow-up after total hip replacement: a source of bias in patient reported outcome measures and registry datasets? Hip Int. 2014;24(5):465-72.

Jahagirdar D, Kroll T, Ritchie K, Wyke S. Patient-reported outcome measures for chronic obstructive pulmonary disease : the exclusion of people with low literacy skills and learning disabilities. Patient. 2013;6(1):11-21.

Johnston B, Patrick D, Devji T, al e. Chapter 18: Patient_reported outcomes. Cochrane Handbook for Systematic Reviews of Interventions. 2019.

Kaur S, Alhaug OK, Dolatowski FC, Solberg TK, Lønne G. Characteristics and outcomes of patients who did not respond to a national spine surgery registry. BMC Musculoskelet Disord. 2023;24(1):164.

Keurentjes JC, Blane D, Bartley M, Keurentjes JJ, Fiocco M, Nelissen RG. Socio-economic position has no effect on improvement in health-related quality of life and patient satisfaction in total hip and knee replacement: a cohort study. PLoS One. 2013;8(3):e56785.

Keurentjes JC, Fiocco M, Nelissen RG. Willingness to undergo surgery again validated clinically important differences in health-related quality of life after total hip replacement or total knee replacement surgery. J Clin Epidemiol. 2014;67(1):114-20.

Keurentjes JC, Fiocco M, So-Osman C, Onstenk R, Koopman-Van Gemert AW, Pöll RG, Kroon HM, Vliet Vlieland TP, Nelissen RG. Patients with severe radiographic osteoarthritis have a better prognosis in physical functioning after hip and knee replacement: a cohort-study. PLoS One. 2013;8(4):e59500.

Kim J, Lonner JH, Nelson CL, Lotke PA. Response bias: effect on outcomes evaluation by mail surveys after total knee arthroplasty. J Bone Joint Surg Am. 2004;86(1):15-21.

Lyman S, Lee YY, McLawhorn AS, Islam W, MacLean CH. What Are the Minimal and Substantial Improvements in the HOOS and KOOS and JR Versions After Total Joint Replacement? Clin Orthop Relat Res. 2018;476(12):2432-41.

Makhni EC. Meaningful Clinical Applications of Patient-Reported Outcome Measures in Orthopaedics. J Bone Joint Surg Am. 2021;103(1):84-91.

Murthy S, Clapp JT, Burson RC, Fleisher LA, Neuman MD. Physicians' perspectives of prognosis and goals of care discussions after hip fracture. J Am Geriatr Soc. 2022.

Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. Med Care. 2003;41(5):582-92.

Norquist BM, Goldberg BA, Matsen FA, 3rd. Challenges in evaluating patients lost to follow-up in clinical studies of rotator cuff tears. J Bone Joint Surg Am. 2000;82(6):838-42.

Perruccio AV, Stefan Lohmander L, Canizares M, Tennant A, Hawker GA, Conaghan PG, Roos EM, Jordan JM, Maillefert JF, Dougados M, Davis AM. The development of a short measure of physical function for knee OA KOOS-Physical Function Shortform (KOOS-PS) - an OARSI/OMERACT initiative. Osteoarthritis Cartilage. 2008;16(5):542-50.

Polk A, Rasmussen JV, Brorson S, Olsen BS. Reliability of patient-reported functional outcome in a joint replacement registry. A comparison of primary responders and non-responders in the Danish Shoulder Arthroplasty Registry. Acta Orthop. 2013;84(1):12-7.

Pronk Y, van der Weegen W, Vos R, Brinkman JM, van Heerwaarden RJ, Pilot P. What is the minimum response rate on patient-reported outcome measures needed to adequately evaluate total hip arthroplasties? Health Qual Life Outcomes. 2020;18(1):379.

Rolfson O, Bohm E, Franklin P, Lyman S, Denissen G, Dawson J, Dunn J, Eresian Chenok K, Dunbar M, Overgaard S, Garellick G, Lubbeke A. Patient-reported outcome measures in arthroplasty registries Report of the Patient-Reported Outcome Measures Working Group of the International Society of Arthroplasty Registries Part II. Recommendations for selection, administration, and analysis. Acta Orthop. 2016;87 Suppl 1:9-23.

Rolfson O, Eresian Chenok K, Bohm E, Lübbeke A, Denissen G, Dunn J, Lyman S, Franklin P, Dunbar M, Overgaard S, Garellick G, Dawson J. Patient-reported outcome measures in arthroplasty registries. Acta Orthop. 2016;87 Suppl 1(Suppl 1):3-8.

Rolfson O, Kärrholm J, Dahlberg LE, Garellick G. Patient-reported outcomes in the Swedish Hip Arthroplasty Register: results of a nationwide prospective observational study. J Bone Joint Surg Br. 2011;93(7):867-75.

Taylor R. Interpretation of the Correlation Coefficient: A Basic Review. JDSM. 1990;6(6):35-9.

Tilbury C, Haanstra TM, Leichtenberg CS, Verdegaal SH, Ostelo RW, de Vet HC, Nelissen RG, Vliet Vlieland TP. Unfulfilled Expectations After Total Hip and Knee Arthroplasty Surgery: There Is a Need for Better Preoperative Patient Information and Education. J Arthroplasty. 2016;31(10):2139-45.

Van Schie P, Van Bodegom-Vos L, Van Steenbergen LN, Nelissen R, Marang-van de Mheen PJ. A more comprehensive evaluation of quality of care after total hip and knee arthroplasty: combining 4 indicators in an ordered composite outcome. Acta Orthop. 2022;93:138-45.

van Schie P, van Bodegom-Vos L, Zijdeman TM, Nelissen R, Marang-van de Mheen PJ. Effectiveness of a multifaceted quality improvement intervention to improve patient outcomes after total hip and knee arthroplasty: a registry nested cluster randomised controlled trial. BMJ quality & safety. 2023;32(1):34-46.

van Schie P, van Steenbergen LN, van Bodegom-Vos L, Nelissen R, Marang-van de Mheen PJ. Between-Hospital Variation in Revision Rates After Total Hip and Knee Arthroplasty in the Netherlands: Directing Quality-Improvement Initiatives. J Bone Joint Surg Am. 2020;102(4):315-24.

van Steenbergen LN, Denissen GA, Spooren A, van Rooden SM, van Oosterhout FJ, Morrenhof JW, Nelissen RGHH. More than 95% completeness of reported procedures in the population-based Dutch Arthroplasty Register. Acta Orthop. 2015;86(4):498-505.

Van Steenbergen LN, Mäkelä KT, Kärrholm J, Rolfson O, Overgaard S, Furnes O, Pedersen AB, Eskelinen A, Hallan G, Schreurs BW, Nelissen R. Total hip arthroplasties in the Dutch Arthroplasty Register

(LROI) and the Nordic Arthroplasty Register Association (NARA): comparison of patient and procedure characteristics in 475,685 cases. Acta Orthop. 2021;92(1):15-22.

Chapter 7

Awareness of performance on outcomes after total hip and knee arthroplasty among Dutch orthopaedic surgeons: how to improve feedback from arthroplasty registries

Peter van Schie[1,2], Leti van Bodegom-Vos[2], Tristan M. Zijdeman[2], Rob G.H.H. Nelissen[1], Perla J. Marang- van de Mheen[2]

1. Department of Orthopaedics, Leiden University Medical Centre, Leiden, Netherlands;
2. Department of Biomedical Data Sciences, Medical Decision Making, Leiden University Medical Centre, Leiden, Netherlands;

Acta Orthop. 2021 Feb;92(1):54-61

Abstract

Background and purpose

The Netherlands Registry of Orthopedic Implants (LROI) uses audit & feedback (A&F) as the strategy to improve performance outcomes after total hip and knee arthroplasty (THA/TKA). Effectiveness of A&F depends on awareness of below average performance to initiate improvement activities. We explored the awareness of Dutch orthopedic surgeons regarding their performance on outcomes after THA/TKA and factors associated with this awareness.

Methods

An anonymous questionnaire was sent to all 445 eligible Dutch orthopedic surgeons performing THA/TKA. To assess awareness on own surgeon group performance, they were asked whether their 1-year THA/TKA revision rates over the past 2 years were below average (negative outlier), average (non-outlier), above average (positive-outlier) in the funnel plot on the LROI dashboard or did not know. Associations were determined with 1) dashboard login at least once a year (yes/no); 2) correct funnel plots interpretation (yes/no) and; 3) recall of their 1-year THA/TKA revision rate (yes/no).

Results

44% respondents started the questionnaire, 158 THA and 156 TKA surgeons. 55% of THA surgeons and 55% of TKA surgeons were aware of their performance. Surgeons aware of their performance more often logged in on the LROI dashboard, more often interpreted funnel plots correctly and more often recalled their revision rate. 38% of THA and 26% of TKA surgeons scored good on all 3 outcomes.

Interpretation

Only half of orthopedic surgeons were aware of their performance status regarding outcomes after THA/TKA. This suggests that to increase awareness, orthopedic surgeons need to be actively motivated to look at the dashboard more frequently and educated on interpretation of funnel plots for audit and feedback to be effective.

Introduction

Several studies have shown large between-hospital variation in performance outcomes after total hip and knee arthroplasty (THA/TKA) including revision rates, suggesting opportunities to improve care (1-6). Audit and feedback (A&F) is a frequently used approach to reduce between-hospital variation, and defined as provision of clinical performance summaries to healthcare providers or organizations intended to initiate activities to improve performance (7,8). Worldwide, A&F from arthroplasty registries is provided in different ways. In the Netherlands, performance indicators such as revision rates, Patient Reported Outcome Measures (PROMs) and patient characteristics are shown on surgeon-group-level in a real-time password protected web-based dashboard and the extent of variation is shown in an anonymized version in annual reports.

Following a Cochrane review of 140 studies from multiple fields, A&F is effective with a median absolute improvement of 4% of the desired outcome, but with the effect size varying from a 9% decrease to a 70% increase (9). Part of the reason for this large variation in effectiveness may be the varying degree to which A&F leads to an increased awareness on own performance. For example, A&F is not received, information including graphs (e.g. funnel plots) and/or tables is not interpreted correctly, or the reported performance outcomes are not considered interesting (10). Sufficient awareness on own performance relative to others in combination with motivation to improve is more likely to result in targeted quality improvement initiatives (11-13).

Due to a lack of awareness on own performance, it is often overestimated (10). This can limit quality improvement initiatives, because it is assumed that performance is good even though there may be room for improvement. Furthermore, it is important that performance indicators give sufficient direction where to improve care, so that professionals are able to select focused interventions to improve care. A recent study showed that for most surgeon groups with significantly higher revision rates, the direction of improvement could be pointed out by looking at the reason for revision (e.g. infection, prosthesis loosening, dislocation etc.) (6). By looking at a more specific outcome, professionals can figure out in which part of the care process improvements are possible, e.g. timing of antibiotic prophylaxis (infection), cementation techniques (prothesis loosening) or femoral head size (dislocation).

We explored the awareness of orthopedic surgeons regarding their performance on outcomes after THA/TKA and factors associated with this awareness, to gain insight into the ways to increase the effectiveness of A&F provided by the LROI.

Methods

An anonymous internet-based questionnaire study was performed in December 2018 to explore the awareness of orthopedic surgeons on outcomes after THA/TKA provided by the LROI and associated factors.

Netherlands Registry of Orthopedic Implants (LROI)

The LROI was established in 2007 and in 2012 all Dutch surgeon groups participated. In 2015, the LROI dashboard was developed to allow surgeons to better monitor their performance showing information on the number of procedures performed, revision rates, PROMs and patient characteristics on surgeon group-level compared to other surgeon groups, which can be viewed at any time. The completeness for primary THA and TKA procedures is checked against Electronic Health Records and is currently above 98% for primary procedures and 96% for revisions (14,15). 97 surgeon groups performed THA and 98 performed TKA in the study period.

Study population

The questionnaire was sent to all 445 Dutch orthopedic surgeons performing primary THA/TKA, who were members of the hip and knee working groups from the Dutch Orthopedic Association. Reminders were sent by email 4 and 8 weeks after the first invitation. The survey was compiled using NetQ software (version 2014.Q3).

Survey

The information collected with the survey regarding the feedback provided on the LROI dashboard, is divided into 4 parts (Appendix, see Supplementary data).

In the first part, awareness regarding possible deviating performance (outlier status) of their own surgeon-group over the last 2 years was assessed by asking whether their 1-year revision rate was below average (negative-outlier), average (non-outlier), above average (positive-outlier) in the funnel plot on the LROI dashboard, or that they did not know. Second, we searched for 3 potential underlying factors that might be related to the level of awareness. It was assessed whether respondents 1) logged in at least once a year on their LROI-dashboard; 2) were able to interpret funnel plots correctly; 3) could recall the 1-year revision rate of their surgeon group. Respondents answering, they did not know were counted as giving a non-positive answer. By combining these 3 questions, a composite outcome was created. A respondent only scored ''good" when all 3 individual measures were positive, i.e. he/she logged in at least once a year, correctly interpreted the funnel plots and could recall their 1-year revision rate. We also asked about hospital work setting (university-, teaching-, general hospital or private clinic) and number of arthroplasties performed annually (<50,

50-100, >100). Third, respondents were asked about quality improvement initiatives following possible below average performance (negative-outlier) in the past 2 years, and whether the effects of these initiatives were checked using the available feedback information on the LROI dashboard. Finally, there were questions about perceived needs for changes in the current feedback, which current performance indicators were considered important, which indicators should be added to improve healthcare and the preferred frequency (every 1, 3, 6, or 12 months) and way of receiving feedback (tailored for their surgeon group or ability to make selections and explore the data oneself).

Statistics

Analyses were performed separately for THA and TKA surgeons. First, the proportion of respondents who were aware of deviating performance for their own surgeon group in the past 2 years was assessed. To examine the associations between awareness of deviating performance and the pre-defined potentially underlying factors (login to the dashboard, correct interpretation of funnel plots, recall of their own revision rate), univariate logistic regression analysis was performed. All questions answered by respondents regardless of whether they completed the full survey were included in the analyses. If surgeons stopped the survey but answered the previous question, we assumed there was a reason for stopping at that specific question (e.g. because it would be not acceptable to say not logging in) and coded this question as don't know, meaning these were included as non-positive answers. In addition, we examined whether the composite outcome differed across hospital settings and number of THA/TKA performed annually.

Data were analyzed with the statistical software of SPSS version 25. *P*-values <0.05 were considered statistically significant in all analyses.

Ethics, funding and potential conflict of interest

The LUMC Medical Ethical Committee waived the need for ethical approval under Dutch law (CME, G18.140). Author PvS received a grant from the Van Rens Foundation (VRF2018-001) to perform this study. The authors declare that there are no conflicts of interest.

Results

From 445 invited orthopedic surgeons, 194 (44%) started the survey, 158 surgeons performed THA and 156 TKA. 78 answered the questions within 4 weeks, 56 after the first and 60 after the second reminder. 169 (87%) respondents completed the survey (Figure 1). Median time to complete the survey was 6:4 minutes (Interquartile range: 5:3-8:5).

91% of respondents were male and 52% were between 40 and 50 years old. Most respondents (40%) were employed in a general hospital and evenly distributed across volume groups for THA and TKA (Table 1).

Table 1 Characteristics of the respondents.

	Respondents (*n*=194)	
	Frequency	**Percentage (%)**
Gender, male	177	91
Age (years)		
• <40	32	16
• 40-50	101	52
• 51-60	42	22
• >60	19	10
Hospital setting		
• University medical center	20	10
• Teaching hospital	72	37
• General hospital	78	40
• Private clinic	24	13
Type of surgeon*		
• Performing THA	158	81
• Performing TKA	156	80
• Performing THA & TKA	120	62
No. of THA per surgeon/year**		
• <50	34	21
• 50-100	75	48
• >100	46	29
• *No response*	3	2
No. of TKA per surgeon/year***		
• <50	37	24
• 50-100	78	50
• >100	32	20
• *No response*	9	6

* Does the respondent perform only THA, only TKA or both THA and TKA.
** There were 158 THA surgeons.
*** There were 156 TKA surgeons.
No. = Number; THA = Total Hip Arthroplasties; TKA = Total Knee Arthroplasties.

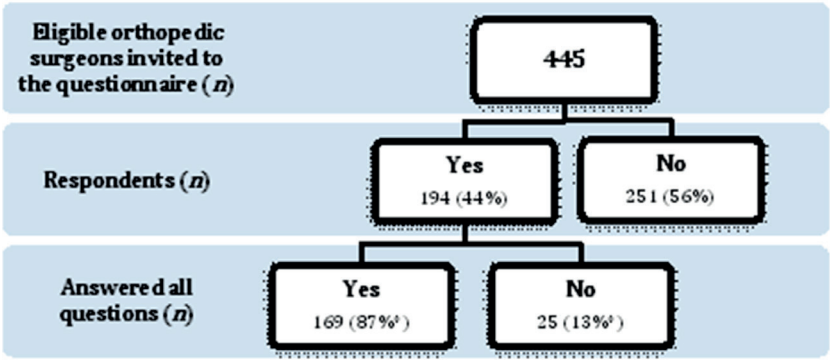


Figure 1 Respondence flowchart.
* Percentage of total number of respondents.

Awareness about performance and underlying factors (Table 2 and 3)

158 THA surgeons answered the questions on logging in, funnel plot interpretation and recalling their revision rate. Only 141 THA surgeons answered the questions on awareness of their surgeon-group performance, with 77 (55%) THA surgeons indicating to be aware of any deviating performance in their surgeon-group over the past 2 years. From the 158 THA surgeons, 105 (67%) logged in on the LROI-dashboard at least once a year, 96 (61%) interpreted the funnel plot correctly and 105 (67%) recalled their 1-year revision rate. THA surgeons who were aware of any deviating performance were 8 times more likely to log in, twice as likely to correctly interpret the funnel plot and 4 times more likely to recall their 1-year revision rate. Overall, 66 (38%) respondents scored good on all these individual items and thus on the composite outcome. THA surgeons who are aware of deviating performance were 5 times more likely to score good on the composite outcome.

156 TKA surgeons answered the questions on logging in, funnel plot interpretation and recalling their revision rate. Only 142 TKA surgeons answered the questions on awareness of own surgeon-group performance, with 78 (55%) TKA surgeons indicating awareness of any deviating performance in their surgeon-group over the past 2 years. Among the 156 TKA surgeons, 103 (66%) logged in to the LROI dashboard at least once a year, 95 (61%) interpreted the funnel plot correctly and 103 (66%) recalled their 1-year revision rate. TKA surgeons who were aware of any deviating performance were 4 times more likely to log in, twice as likely to correctly interpret the funnel plot and 5 times more likely to recall their 1-year revision rate. Overall, 41 (26%) respondents scored good on the composite outcome and TKA surgeons who are aware of deviating performance were 4 times more likely to score good on the composite outcome.

Table 2 Associations between awareness of surgeon-group performance and logging in to dashboard, correct funnel plot interpretation and knowledge about 1-year revision rate.

Yes		Logging in on LROI-dashboard*		Correct funnel plot interpretation**		Knowledge about 1-year revision rate***	
		No	**Yes**	**No**	**Yes**	**No**	
THA Surgeons	**All THA performing respondents** (*n*=158)	105 (67%)	53 (33%)	96 (61%)	62 (39%)	105 (67%)	53 (33%)
		Yes	**OR (CI)**	**Yes**	**OR (CI)**	**Yes**	**OR (CI)**
	Aware of surgeon-group performance						
	Yes (*n*=77)	69 (90%)	7.6 (3.2-18)	59 (77%)	2.4 (1.2-4.9)	66 (86%)	4.4 (2.0-9.8)
	No (*n*=64)	34 (53%)	*reference*	37 (58%)	*reference*	37 (58%)	*reference*

Yes		Logging in on LROI-dashboard*		Correct funnel plot interpretation**		Knowledge about 1-year revision rate***	
		No	**Yes**	**No**	**Yes**	**No**	
TKA Surgeons	**All TKA performing respondents** (*n*=156)	103 (66%)	53 (34%)	95 (61%)	61 (39%)	103 (66%)	53 (34%)
		Yes	**OR (CI)**	**Yes**	**OR (CI)**	**Yes**	**OR (CI)**
	Aware of surgeon-group performance						
	Yes (*n*=78)	65 (83%)	4.1 (1.9-9.0)	56 (72%)	1.6 (0.8-3.3)	66 (85%)	4.9 (2.2-10.7)
	No (*n*=64)	35 (55%)	*reference*	39 (61%)	*reference*	34 (53%)	*reference*

*Logging in at least once every year

**Correctly interpreted both funnel plots

***Know the 1-year revision rate of their healthcare center of the past 2 years.

CI = 95% confidence interval; OR = Odds ratio; THA = Total Hip Arthroplasty; TKA = Total Knee Arthroplasty.

Table 3 Composite outcome stratified by hospital setting and number of arthroplasties performed annually.

Yes (good)		Composite outcome	
		No	
	All THA performing respondents (*n*=158)	60 (38%)	98 (62%)
		Yes	**OR (CI)**
	Aware of surgeon-group performance		
	Yes (*n*=77)	46	5.3
	No (*n*=64)	14	*reference*
	Hospital setting (*n*=158)		
THA Surgeons	University medical center	4	0.4
	Teaching hospital	19	0.6
	General hospital	31	*reference*
	Private clinic	6	0.7
	No. of THA performed per year (*n*=155)*		
	<50	7	0.4
	50-100	35	1.4
	>100	18	*reference*

Yes (good)		Composite outcome	
		No	
	All TKA performing respondents (*n*=156)	41	115
		Yes	**OR (CI)**
	Aware of surgeon-group performance		
	Yes (*n*=78)	31	3.6
	No (*n*=64)	10	*reference*
	Hospital setting (*n*=156)		
TKA Surgeons	University medical center	2	0.3
	Teaching hospital	15	0.6
	General hospital	23	*reference*
	Private clinic	1	0.1
	No. of TKA performed per year (*n*=147)**		
	<50	7	2.4
	50-100	28	1.0
	>100	6	*reference*

*The number of THA performed per year by the respondent.
**The number of TKA performed per year by the respondent.
CI = 95% confidence interval; No. = Number; OR = Odds ratio; THA = Total Hip Arthroplasties; TKA = Total Knee Arthroplasties.

The proportion of surgeons who met the criteria of the composite outcome did not differ by the number of arthroplasties performed annually or across hospital settings, except for a lower proportion for TKA surgeons in private clinics.

Quality improvement initiatives

20 respondents indicated that they were employed in a healthcare center that had a significantly higher 1-year revision rate (negative-outlier) in the past 2 years. 9 of them

did not see this deviating performance coming, because they had never checked the LROI dashboard for performance indicators. 17 indicated that quality improvement initiatives had been introduced and all of them used performance indicators from the LROI dashboard to monitor the effect. A positive effect of these initiatives on the revision rate was reported by 9 respondents and a negative effect by 3 respondents when checking progress in the LROI dashboard. 5 respondents were currently following the effect.

Future feedback

From the current available performance indicators, the number of procedures performed was mostly considered as the most interesting information on the LROI dashboard, followed by 1-year revision rates, PROMs and patient characteristics respectively (Figure 2).

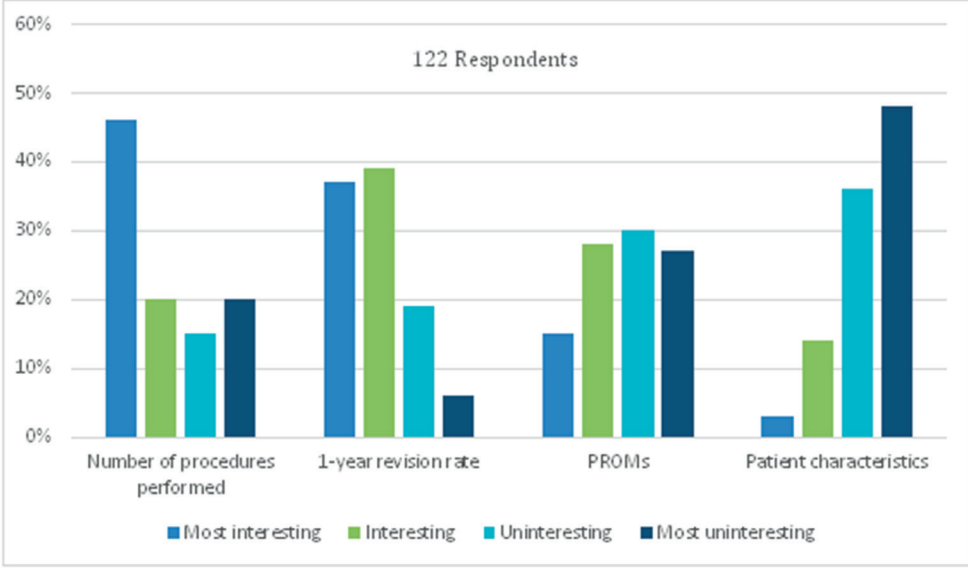


Figure 2 Currently available performance indicators on the secure LROI dashboard ranked from most to least interesting by respondents.
LROI = Dutch arthroplasty register; PROMs = Patient reported outcome measures.

Prosthesis survival and complications are currently not available on the LROI dashboard, but 138 (82%) THA surgeons and 129 (76%) TKA surgeons indicated this information to constitute relevant indicators (Figure 3). 106 (62%) respondents would prefer to receive feedback every 6 months, and a minority every month (*n*=6, 4%), every quarter (*n*=40, 23%) and some respondents having no preference (*n*=18, 11%). 139 (82%) respondents prefer feedback that is tailored for their surgeon-group

without making any selections and 30 respondents (18%) indicated to prefer making their own selections of LROI indicators.

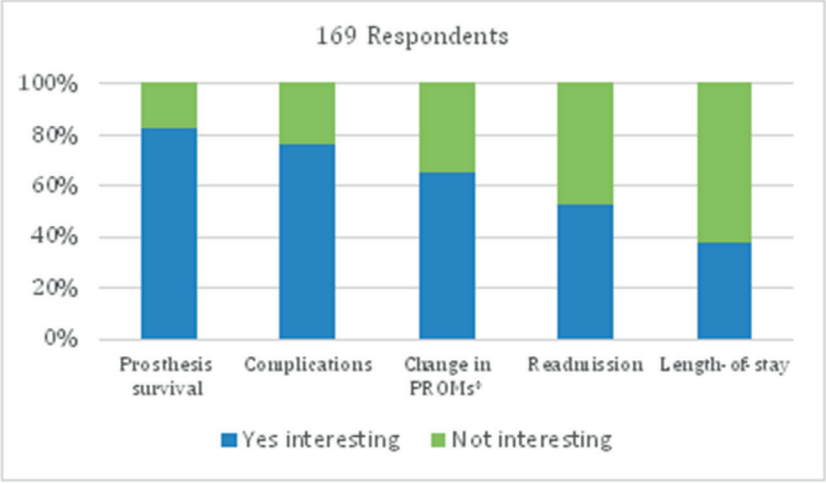


Figure 3 Percentage of orthopedic surgeons interested in additional performance indicators.
*Difference between pre- and post-operative PROMs
LROI = Dutch arthroplasty register; PROMs = Patient reported outcome measures.

Discussion

Although Dutch orthopedic surgeons performing THA/TKA can view their surgeon-group performance on a web-based A&F-dashboard, only half of them are actually aware of their performance over the past 2 years. This lack of awareness on own performance and the associations found in our study, suggests that orthopedic surgeons need to be actively motivated to log in more often, need to be educated on how to interpret funnel plots correctly and must be able to reproduce their revision rate for the A&F to be effective in improving care. To act upon the feedback information all underlying factors must be met, but this was the case in only one third of THA surgeons and one fourth of TKA surgeons, fairly similar across different types of hospitals and annual volume. Yet, it seems important to increase the effectiveness of feedback, given that 9 out of 20 respondents of the negative-outlier surgeon-groups indicated that they did not see their worsening performance coming. Without effective feedback, surgeon-groups would continue to provide care without modification, while 17 out of these 20 respondents indicated that they conducted quality improvement initiatives once identified as showing poor performance.

Differences and similarities between national arthroplasty registries in providing A&F

The way in which A&F is offered varies, from publicly available annual reports including only nationwide averages with sometimes additional surgeon-group specific performance, where others publish their indicators on surgeon-group-level and surgeon-level only in password-protected online dashboards (16-25). The LROI, National Joint Registry, United Kingdom (NJR) and Swedish Hip Arthroplasty Registries (SHAR) use a web-based password-protected A&F dashboard to provide surgeons with peer comparison indicators in visual graphs on surgeon-group-level and in the United Kingdom also on surgeon-level (16,17,25,26). In contrast, the Swedish Knee Arthroplasty Registries (SKAR) and the Danish Hip Arthroplasty Registries (DHAR) make no use of online dashboards, where the SKAR publishes only some indicators (e.g. patient demographics and PROMs) on their publicly accessible website once a year. Some arthroplasty registers may inform participating hospitals once a year about their performance e.g. by emailing performance indicators without this being listed on their website. The feedback generated by the NJR is updated every 6 months, which was also indicated as the preferred frequency to receive feedback by two thirds of respondents in our study (16,18). The Finnish Arthroplasty Registries (FAR), even uses a daily updated publicly accessible website, which includes patient demographics and revision rates at surgeon-group-level (17). What all these different methods of feedback have in common, is that it is passive education, not requiring any action which may be one of the explanations for orthopedic surgeons being unaware about their performance. Public availability of performance indicators may increase the likelihood of action being taken, given that both patients and other stakeholders like insurance companies can review the data and may use them in their decision making.

Comparison with literature

Besides the Cochrane review, there are more studies that found wide variation in the effect of A&F (9). A review, evaluating interactive computer feedback, found a highly variable effect of improvement in quality of care in 3 out of 7 studies (27). Another more recent study found a significant improvement for 4 out of 6 performance indicators, 2,5 years after implementation of online A&F interventions in maternal-new-born hospitals (28). Given the varying effect of A&F, the results of our study can make a relevant contribution to further improve current feedback as provided by arthroplasty registries. We have gained insight into whether A&F reached the target group (i.e. how often do surgeons log in), the ability to interpret the funnel plot and recall of revision rates. In addition, we investigated which performance indicators currently provided by the LROI are considered important by the target group and which indicators should be added. Furthermore, it would be useful to provide feedback on the reasons for revisions, given that this has been shown able to direct quality improvement

initiatives although, we did not specifically ask whether orthopedic surgeons would be interested in this information (6). 2 meta-analysis have shown that a single A&F strategy is one of the less effective interventions showing little to no improvement when examined (29,30). On the other hand, it seems obvious that accessible A&F that is interpreted correctly will ultimately improve the quality of care, as 17 out of 20 orthopedic surgeons indicated that they would conduct quality improvement initiatives as soon as they become aware of poorer performance. It seems likely that more active elements need to be added both to motivate orthopedic surgeons to log in and to ensure correct interpretation of the funnel plot, which is needed to be aware of outlier status regarding their performance.

Trust in A&F data quality is often identified as a barrier to change clinical behavior. This is unlikely to play a major role in the current LROI feedback given the 98% completeness for primary procedures and 96% for revisions, which is similar for the data in above mentioned arthroplasty registries (10-12,14). Another barrier may be that physicians do not consider some indicators as an essential part of quality or deem benchmarks unrealistic (10,31-35). In this study, for instance, it was found that one third of both THA and TKA surgeons do not know their 1-year revision rate, which may suggest that some surgeons do not recognize the importance of this outcome. This is striking because this outcome is already widely used by arthroplasty registries and considered an indicator to reflect the quality of care (17,19,20,22,24). Moreover, A&F does not use absolute benchmarks, but performance indicators are compared with national surgeon-group averages, thereby making it likely that other similar surgeon-groups are able to achieve that level of performance.

Strengths and limitations

A possible limitation of this study is response bias if awareness of performance differs between responders and non-responders and the association with underlying factors were to be different. Given that survey responses were collected anonymously, we were unable to compare whether the characteristics of the non-respondents differed from the respondents to assess whether bias may have occurred. However, considering the overall response rate of 44%, and the fact that non-respondents in general are not as involved as respondents and thus more likely to be not aware of their performance, the associations are likely underestimated. A second limitation, is that some self-reported outcomes (e.g. frequency of logging in or recall of revision rate) were analyzed. It is therefore possible that there were socially desirable answers to certain questions e.g. knowledge about certain indicators. If this affected the results, even fewer orthopedic surgeons may be aware of their performance. However, because this was an anonymous survey, it seems more likely that respondents are surgeons dedicated to good performance and making feedback information more useful rather than giving

socially desirable answers, so that reported rates are likely to reflect actual practice. An exception on the self-reported outcomes was the funnel plot interpretation, where answers given by respondents were compared with the correct answer so that social desirability was not an issue. A third limitation may be the generalization of our results to other countries. Increasingly information becomes publicly available on differences between hospitals in patient outcomes, as we have previously shown for revision rates in the Netherlands and Bozic et al (2014) have shown for complication rates after total hip and knee arthroplasty in the US (1,6). The magnitude of the between-hospital variation in risk-adjusted rates in these studies is surprisingly similar, with both studies showing about 3-4 fold differences between hospitals. Furthermore, although not looking at awareness in performance specifically, a previous international survey study showed only minor differences between orthopaedic surgeons operating in different continents, taking into account their demographics (e.g. sex, age), surgical experience (e.g. number of years in practice, number of arthroplasties performed per year), use of additional diagnostics (e.g. plain radiographs, CT, MRI) and final treatment chosen (e.g. surgical versus non-surgical) (36). So, there is no evidence to suggest that there would be smaller differences between surgeons regarding their performance in other countries, and a difference in awareness has to our knowledge not been described before. Yet, such difference in awareness may be crucial in explaining why hospital differences in performance continue to exist, rather than that public reporting of hospital differences will by itself result in improvement.

Implementation and further research

As alluded to above, more active elements need to be added to improve the A&F design to make it more attractive to log in and result in more awareness on own performance. This could be encouraged by emphasizing the importance of already available indicators (e.g. revision rates) and adding new indicators to the A&F dashboard that are considered relevant and of interest as reported in this study (prosthesis survival, complications, readmissions and length-of-hospital-stay). As a result, more surgeons may be actually reached by the feedback, because the number of orthopedic surgeons who log in as well as the frequency of logging in will then increase. In addition, teaching material must be available on how to interpret funnel plots and be actively promoted by the orthopedic association during meetings, which will also increase awareness and possibly increase the reach of feedback, when more surgeons can interpret the performance indicators. Ultimately, an increased awareness of one's own performance will likely lead to more quality improvement initiatives.

The question arises as to whether voluntary quality control by providing only passive A&F on performance is sufficient in modern orthopedic society. A&F could be more effective when offered in a more active and multifaceted way instead of a single element

(which in this study was only the LROI-dashboard) (9,37). A possible addition to the feedback would be that indicators are also verbally explained by an independent person, with clear targets discussed and action plans created, for instance based on a toolbox (7,9,11,38-44). In addition, setting up committees that will actively approach poorly performing hospitals to create action plans to improve quality of care, may increase interest in one's own performance as orthopedic surgeons want to avoid being under supervision. The Dutch Orthopedic Association has initiated up a quality committee in 2017 with the aim to detect negative outlier hospitals using LROI-data and discuss activities to improve care (45). This new procedure may stimulate logging in to check on performance and in this way increase awareness of own performance in the coming years. After all, orthopedic surgeons have no valid reason not to be interested in their own performance, given that they want the best care for their patients and continuously improving the quality of care is thus inherently linked to that.

This survey is part of the ''Improving Quality based on the Joint registries project" (IQ Joint study). Within this study, what will be tested includes whether more active intervention including monthly feedback on THA/TKA performance indicators, active education on how to use indicators for quality improvement, asking for improvement activities and linking hospitals with better performing hospitals to exchange information and find areas for improvement will result in better outcomes, fewer complications and more quality improvement initiatives compared to the LROI dashboard alone. During this randomized trial, A&F on surgeon-group-level will be provided according to the preferences of the orthopedic surgeons as has been evaluated in this study.

Conclusion

Orthopedic surgeons performing THA/TKA have limited awareness on performance of their surgeon-group. Awareness could be increased by encouraging them to log in more often on their A&F dashboard, teaching them how to interpret funnel plots and emphasizing the importance of performance indicators. Improvement of the effectiveness of feedback is important, because the majority of orthopedic surgeons indicated that quality improvement initiatives were introduced once they learned that their performance was worsening. To provide orthopedic surgeons with better feedback in the future, the feedback information should be extended with the indicators prosthesis survival and complications compared with peers at a national level, tailored to their specific surgeon-group rather than making any selections themselves, with 6-month frequency.

References

1. Bozic KJ, Grosso LM, Lin Z, Parzynski CS, Suter LG, Krumholz HM, et al. Variation in hospital-level risk-standardized complication rates following elective primary total hip and knee arthroplasty. J Bone Joint Surg Am. 2014;96(8):640-7.
2. Fry DE, Pine M, Nedza SM, Locke DG, Reband AM, Pine G. Risk-Adjusted Hospital Outcomes in Medicare Total Joint Replacement Surgical Procedures. J Bone Joint Surg Am. 2017;99(1):10-8.
3. Weeks WB, Kotzbauer GR, Weinstein JN. Using Publicly Available Data to Construct a Transparent Measure of Health Care Value: A Method and Initial Results. Milbank Q. 2016;94(2):314-33.
4. Siciliani L, Sivey P, Street A. Differences in length of stay for hip replacement between public hospitals, specialised treatment centres and private providers: selection or efficiency? Health Econ. 2013;22(2):234-42.
5. Menendez ME, Lu N, Huybrechts KF, Ring D, Barnes CL, Ladha K, et al. Variation in Use of Blood Transfusion in Primary Total Hip and Knee Arthroplasties. J Arthroplasty. 2016;31(12):2757-63.e2.
6. van Schie P, van Steenbergen LN, van Bodegom-Vos L, Nelissen R, Marang-van de Mheen PJ. Between-Hospital Variation in Revision Rates After Total Hip and Knee Arthroplasty in the Netherlands: Directing Quality-Improvement Initiatives. J Bone Joint Surg Am. 2019.
7. Ivers NM, Sales A, Colquhoun H, Michie S, Foy R, Francis JJ, et al. No more 'business as usual' with audit and feedback interventions: towards an agenda for a reinvigorated intervention. Implement Sci. 2014;9:14.
8. Brehaut JC, Eva KW. Building theories of knowledge translation interventions: use the entire menu of constructs. Implement Sci. 2012;7:114.
9. Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, French SD, et al. Audit and feedback: effects on professional practice and healthcare outcomes. Cochrane Database Syst Rev. 2012(6):Cd000259.
10. Gude WT, Roos-Blom MJ, van der Veer SN, Dongelmans DA, de Jonge E, Francis JJ, et al. Health professionals' perceptions about their clinical performance and the influence of audit and feedback on their intentions to improve practice: a theory-based study in Dutch intensive care units. Implement Sci. 2018;13(1):33.
11. van der Veer SN, de Keizer NF, Ravelli AC, Tenkink S, Jager KJ. Improving quality of care. A systematic review on how medical registries provide information feedback to health care providers. Int J Med Inform. 2010;79(5):305-23.
12. de Vos Maartje LG, van der Veer SN, Graafmans WC, de Keizer NF, Jager KJ, Westert GP, et al. Process evaluation of a tailored multifaceted feedback program to improve the quality of intensive care by using quality indicators. BMJ quality & safety. 2013;22(3):233-41.
13. Davis DA, Mazmanian PE, Fordis M, Van Harrison R, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. JAMA. 2006;296(9):1094-102.
14. van Steenbergen LN, Denissen GA, Spooren A, van Rooden SM, van Oosterhout FJ, Morrenhof JW, et al. More than 95% completeness of reported procedures in the population-based Dutch Arthroplasty Register. Acta Orthop. 2015;86(4):498-505.
15. *LROI website. Completeness of Registering Hospitals and Completeness of Registered Arthroplasties in the LROI Based on the Hospital Information System in 2016,* http://www.lroi-rapportage.nl/data-quality-coverage-and-completeness, accessed febr 2019.

16. Porter M, Armstrong R, Howard P, Porteous M, Wilkinson JM. Orthopaedic registries - the UK view (National Joint Registry): impact on practice. EFORT Open Rev. 2019;4(6):377-90.
17. Tabak AY, Bicimoglu A, Celebi L, Muratli HH, Yagmurlu MF, Aktekin CN. Early results of revision total hip replacements that used self locking revision stem (Wagner stem). [Turkish]. Artroplasti Artroskopik Cerrahi. 2002;13(2):65-72.
18. Assi CC, Barakat HB, Caton JH, Najjar EN, Samaha CT, Yammine KF. Mortality Rate and Mid-Term Outcomes of Total Hip Arthroplasty Using Dual Mobility Cups for the Treatment of Femoral Neck Fractures in a Middle Eastern Population. J Arthroplasty. 2019;34(2):333-7.
19. Bonutti PM, Khlopas A, Chughtai M, Cole C, Gwam CU, Harwin SF, et al. Unusually High Rate of Early Failure of Tibial Component in ATTUNE Total Knee Arthroplasty System at Implant-Cement Interface. J Knee Surg. 2017;30(5):435-9.
20. Itonaga I, Sabokbar A, Murray DW, Athanasou NA. Effect of osteoprotegerin and osteoprotegerin ligand on osteoclast formation by arthroplasty membrane derived macrophages. Ann Rheum Dis. 2000;59(1):26-31.
21. Kurcz B, Lyons J, Sayeed Z, Anoushiravani AA, Iorio R. Osteolysis as it Pertains to Total Hip Arthroplasty. Orthop Clin North Am. 2018;49(4):419-35.
22. Li TF, Santavirta S, Virtanen I, Könönen M, Takagi M, Konttinen YT. Increased expression of EMMPRIN in the tissue around loosened hip prostheses. Acta Orthop Scand. 1999;70(5):446-51.
23. Pelt CE, Stagg ML, Van Dine C, Anderson MB, Peters CL, Gililland JM. Early outcomes after revision total hip arthroplasty with a modern modular femoral revision stem in 65 consecutive cases. Arthroplasty today. 2019;5(1):106-12.
24. Talmo CT, Anderson MC, Jia ES, Robbins CE, Rand JD, McKeon BP. High Rate of Early Revision After Custom-Made Unicondylar Knee Arthroplasty. J Arthroplasty. 2018;33(7s):S100-s4.
25. Yoon C, Chang MJ, Chang CB, Choi JH, Lee SA, Kang SB. Does unicompartmental knee arthroplasty have worse outcomes in spontaneous osteonecrosis of the knee than in medial compartment osteoarthritis? A systematic review and meta-analysis. Arch Orthop Trauma Surg. 2019;139(3):393-403.
26. Toomey SD, Hopper RH, Jr., McAuley JP, Engh CA. Modular component exchange for treatment of recurrent dislocation of a total hip replacement in selected patients. J Bone Joint Surg Am. 2001;83(10):1529-33.
27. Tuti T, Nzinga J, Njoroge M, Brown B, Peek N, English M, et al. A systematic review of electronic audit and feedback: intervention effectiveness and use of behaviour change theory. Implement Sci. 2017;12(1):61.
28. Weiss D, Dunn SI, Sprague AE, Fell DB, Grimshaw JM, Darling E, et al. Effect of a population-level performance dashboard intervention on maternal-newborn outcomes: an interrupted time series study. BMJ quality & safety. 2018;27(6):425-36.
29. Shojania KG, Ranji SR, McDonald KM, Grimshaw JM, Sundaram V, Rushakoff RJ, et al. Effects of quality improvement strategies for type 2 diabetes on glycemic control: a meta-regression analysis. JAMA. 2006;296(4):427-40.
30. Tricco AC, Ivers NM, Grimshaw JM, Moher D, Turner L, Galipeau J, et al. Effectiveness of quality improvement strategies on the management of diabetes: a systematic review and meta-analysis. Lancet. 2012;379(9833):2252-61.
31. Gude WT, van Engen-Verheul MM, van der Veer SN, de Keizer NF, Peek N. How does audit and feedback influence intentions of health professionals to improve practice? A laboratory experiment and field study in cardiac rehabilitation. BMJ quality & safety. 2017;26(4):279-87.

32. van der Veer SN, de Vos ML, Jager KJ, van der Voort PH, Peek N, Westert GP, et al. Evaluating the effectiveness of a tailored multifaceted performance feedback intervention to improve the quality of care: protocol for a cluster randomized trial in intensive care. Implement Sci. 2011;6:119.
33. Eva KW, Armson H, Holmboe E, Lockyer J, Loney E, Mann K, et al. Factors influencing responsiveness to feedback: on the interplay between fear, confidence, and reasoning processes. Adv Health Sci Educ Theory Pract. 2012;17(1):15-26.
34. Gude WT, van der Veer SN, de Keizer NF, Coiera E, Peek N. Optimizing Digital Health Informatics Interventions Through Unobtrusive Quantitative Process Evaluations. Stud Health Technol Inform. 2016;228:594-8.
35. Catelas I, Lehoux EA, Ning Z, Figeys D, Baskey SJ, Beaulé PE. Differential proteomic analysis of synovial fluid from hip arthroplasty patients with a pseudotumor vs. Periprosthetic osteolysis J Orthop Res. 2018;36(7):1849-59.
36. Li CS, Karlsson J, Winemaker M, Sancheti P, Bhandari M. Orthopedic surgeons feel that there is a treatment gap in management of early OA: international survey. Knee Surg Sports Traumatol Arthrosc. 2014;22(2):363-78.
37. Soong C, Shojania KG. Education as a low-value improvement intervention: often necessary but rarely sufficient. BMJ quality & safety. 2020;29(5):353-7.
38. Brehaut JC, Colquhoun HL, Eva KW, Carroll K, Sales A, Michie S, et al. Practice Feedback Interventions: 15 Suggestions for Optimizing Effectiveness. Ann Intern Med. 2016;164(6):435-41.
39. Gude WT, Roos-Blom MJ, van der Veer SN, de Jonge E, Peek N, Dongelmans DA, et al. Electronic audit and feedback intervention with action implementation toolbox to improve pain management in intensive care: protocol for a laboratory experiment and cluster randomised trial. Implement Sci. 2017;12(1):68.
40. Bradley EH, Holmboe ES, Mattera JA, Roumanis SA, Radford MJ, Krumholz HM. Data feedback efforts in quality improvement: lessons learned from US hospitals. Qual Saf Health Care. 2004;13(1):26-31.
41. Bradley EH, Curry LA, Webster TR, Mattera JA, Roumanis SA, Radford MJ, et al. Achieving rapid door-to-balloon times: how top hospitals improve complex clinical systems. Circulation. 2006;113(8):1079-85.
42. Brown B, Gude WT, Blakeman T, van der Veer SN, Ivers N, Francis JJ, et al. Clinical Performance Feedback Intervention Theory (CP-FIT): a new theory for designing, implementing, and evaluating feedback in health care based on a systematic review and meta-synthesis of qualitative research. Implement Sci. 2019;14(1):40.
43. de Vos M, Graafmans W, Kooistra M, Meijboom B, Van Der Voort P, Westert G. Using quality indicators to improve hospital care: a review of the literature. Int J Qual Health Care. 2009;21(2):119-29.
44. Roos-Blom MJ, Gude WT, de Jonge E, Spijkstra JJ, van der Veer SN, Peek N, et al. Impact of audit and feedback with action implementation toolbox on improving ICU pain management: cluster-randomised controlled trial. BMJ quality & safety. 2019;28(12):1007-15.
45. CommisionQuality. Commision Quality (Collaboration between Dutch Orthopaedic Association and Dutch Arthroplasty Registry). Protocol: Quality procedure, 2017. www.orthopeden.org.

Supplemental data

Appendix Survey

In order to make the questionnaire run as efficiently as possible, a number of ''loops'' have been incorporated into the questionnaire, so that orthopedic surgeons do not have to answer questions that do not apply to them. These loops are indicated in the questionnaire.

Questions for all respondents: Surgeon-specific questions

1. What is your age in years?
   a. Under 40
   b. 40 to 50
   c. 51 to 60
   d. Above 60

2. What is your gender?
   a. Male
   b. Female

3. In what type of healthcare center do you work for the majority of your time? *(This question involves part 2, see methods section survey in article)*
   a. University hospital
   b. Teaching hospital
   c. General hospital
   d. Private clinic

4. Do you perform primary total hip arthroplasties and/or primary total knee arthroplasties?
   a. Yes, only hip arthroplasties → Loop I (questions 5,6, (skip questions 7,8))
   b. Yes, only knee arthroplasties → Loop II (questions 7,8 (skip questions 5,6))
   c. Yes, both hip- and knee arthroplasties → Loop I &II (questions 5,6,7,8)
   d. No → End of questionnaire

Loop I: Orthopedic surgeons performing Total Hip Arthroplasties

5. How many primary total hip arthroplasties do you perform annually? *(This question involves part 2, see methods section survey in article)*
   a. Less than 50
   b. 50 to 100

c. More than 100

Questions regarding the online LROI-dashboard for total hip arthroplasty

6. What was the overall 1-year revision rate of your department for total hip arthroplasties over the last 2 years? *(This question involves part 2, see methods section survey in article)*
   a. Worse than average
   b. Average
   c. Better than average
   d. I do not know

Loop II: Orthopedic surgeons performing Total Knee Arthroplasties

7. How many primary total knee arthroplasties do you perform annually? *(This question involves part 2, see methods section survey in article)*
   a. Less than 50
   b. 50 to 100
   c. More than 100

Questions regarding the online LROI-dashboard for total knee arthroplasty

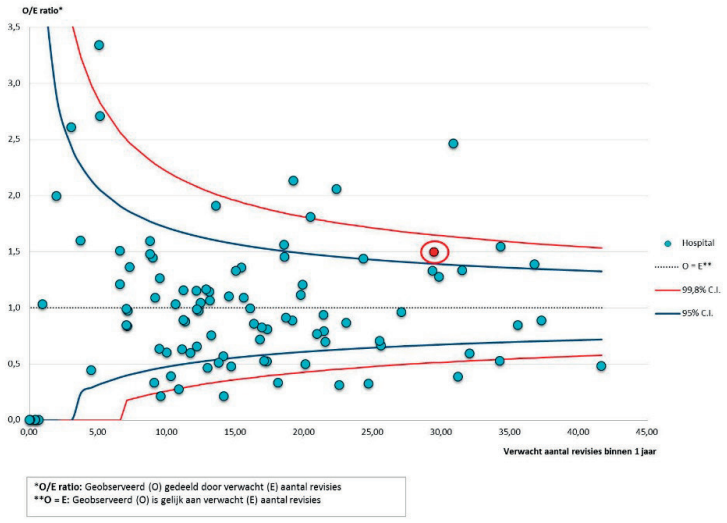
8. What was the overall 1-year revision rate of your department for total knee arthroplasties over the last 2 years? *(This question involves part 2, see methods section survey in article)*
   a. Worse than average
   b. Average
   c. Better than average
   d. I do not know

Questions for all respondents: Frequency of logging in on LROI-dashboard.

9. How often do you log in on the LROI-dashboard? *(This question involves part 2, see methods section survey in article)*
   a. Never
   b. Once a week
   c. Once a month
   d. Once per 6 months
   e. Once a year

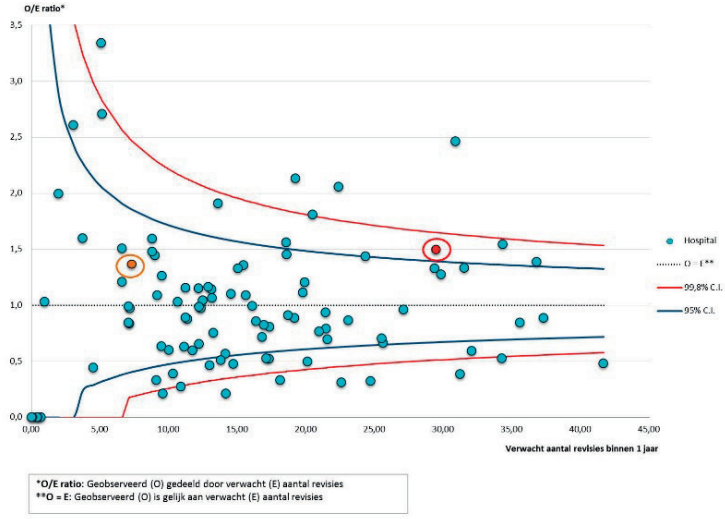
Questions for all respondents: Funnel-plot interpretation.

10. How does the department encircled with red perform? *(This question involves part 2, see methods section survey in article)*



a. Much worse than average (outside 99.8% control limit)
b. Worse than average (outside 95% control limit)
c. Average (within 95% control limit)
d. Better than average (outside 95% control limit)
e. Much better than average (outside 99.8% control limit)
f. I do not know

11. How does the department encircled with orange perform compared to the department encircled with red? *(This question involves part 2, see methods section survey in article)*



a. Better
b. Equal
c. Worse

Questions for all respondents: Awareness about performance of own healthcare center

12. Has your department, according to the funnel-plots displayed on the LROI-dashboard ever performed worse (above the 95% confidence interval) in the past two years? *(This question involves part 1, see methods section survey in article)*
    a. Yes → Loop III (question 13,14,15)
    b. No → Question 16
    c. I do not know → Question 16

Loop III: Worse than average performance based on the LROI-dashboard

13. Did you see in advance that your department performed worse than average? *(This question involves part 3, see methods section survey in article)*
    a. No,
    b. Yes, I felt this coming
    c. Yes, because we collect the same data as the LROI
    d. Yes, because my colleague drew my attention to this

14. Did you undertake quality improvement initiatives to improve? *(This question involves part 3, see methods section survey in article)*
    a. Yes, because we already knew what caused it
    b. Yes, after investigating the cause
    c. No, because the results may be due to coincidence and this is probably an one-off incident
    d. No, because in this period we were treating on a relatively difficult patient population. The results will therefore improve automatically
    e. No, other reason, namely…

15. Did you use the LROI-data to check whether the quality improvement initiative(s) have had effect(s)? *(This question involves part 3, see methods section survey in article)*
    a. No, we have not taken any action
    b. No, we did not check the effect of the intervention
    c. No, we introduced the intervention recently and are monitoring whether an effect is sorting
    d. Yes, the intervention(s) had no effect
    e. Yes, the intervention(s) had a positive effect

Questions for all respondents: Future improvements for feedback

16. Would you prefer to receive a signal earlier if the performance of your department improves or deteriorates compared to the national average? *(This question involves part 4, see methods section survey in article)*
    a. No, I can see that in the funnel-plot in the LROI-dashboard
    b. Yes, I would like an update every 6 months
    c. Yes, I would like an update every 3 months
    d. Yes, I would like an update every month

17. Which tabs on the LROI-dashboard interest you the most? Put the results in order from most interesting (1) to least interesting (4) by dragging the 'blocks". *(This question involves part 4, see methods section survey in article)*
    a. Total number of procedures performed
    b. 1-year revision rate
    c. PROMs
    d. Patient characteristics

18. Are there, in addition to the 1-year revision rates, in comparison with other healthcare centers, more outcomes in which you are interested (Yes/No)? You can check multiple options here. *(This question involves part 4, see methods section survey in article)*

a. Prosthesis survival
b. Improvement in PROMs postoperative compared to preoperative
c. Length-of-hospital-stay
d. Hospital readmission
e. Complications (other than revisions)
f. None

19. How would you like to receive feedback on the outcomes of your department? *(This question involves part 4, see methods section survey in article)*
    a. As the current situation, make selections on the LROI-dashboard
    b. Make selections on a mobile application
    c. Tailored for my surgeon-group, without making selections myself, on the LROI-dashboard
    d. Tailored for my surgeon-group, without making selections myself, on a mobile application
    e. Tailored for my surgeon-group, without making selections myself, send by email.

Chapter 8

Effectiveness of a multifaceted quality improvement intervention to improve patient outcomes after Total Hip and Knee Arthroplasty: A registry nested cluster randomised controlled trial

Peter van Schie, MD[1,2], Leti van Bodegom-Vos, PhD[2], Tristan M. Zijdeman[2], Rob G.H.H. Nelissen, MD, PhD[1], Perla J. Marang-van de Mheen, PhD[2], and the IQ Joint study group

1. Department of Orthopaedics, Leiden University Medical Centre, Leiden, The Netherlands;
2. Department of Biomedical Data Sciences, Medical Decision Making, Leiden University Medical Centre, Leiden, The Netherlands.

BMJ Qual Saf. 2023 Jan;32(1):34-46

Abstract

Objective

To assess the effectiveness of a prospective multifaceted quality improvement intervention on patient outcomes after total hip and knee arthroplasty (THA and TKA).

Design

Cluster randomised controlled trial nested in a national registry. From 1 January 2018 to 31 May 2020 routinely submitted registry data on revision and patient characteristics were used, supplemented with hospital data on readmission, complications, and length of stay (LOS) for all patients.

Setting

20 orthopaedic departments across hospitals performing THA and TKA in The Netherlands.

Participants

32,923 patients underwent THA and TKA, in 10 intervention and 10 control hospitals (usual care).

Intervention

The intervention period lasted 8 months and consisted of the following components: 1) Monthly updated feedback on 1-year revision, 30-day readmission, 30-day complications, long (upper quartile) LOS, and these 4 indicators combined in a composite outcome; 2) Interactive education; 3) An action toolbox including evidence-based quality improvement initiatives (QIIs) to facilitate improvement of above indicators; and 4) Bimonthly surveys to report on QII undertaken.

Main outcome measures

The primary outcome was Textbook Outcome (TO), an all-or-none composite representing the best outcome on all performance indicators (i.e., the absence of revision, readmissions, complications, and long LOS). The individual indicators were analysed as secondary outcomes. Changes in outcomes from pre-intervention to intervention period were compared between intervention versus control hospitals, adjusted for case-mix and clustering of patients within hospitals using random effect binary logistic regression models. The same analyses were conducted for intervention hospitals that did and did not introduce QII.

Results

16,314 patients were analysed in intervention hospitals (12,475 before and 3,839 during intervention) versus 16,609 in control hospitals (12,853 versus 3,756). After the intervention period, the absolute probability to achieve TO increased by 4.32% (95% confidence interval (CI) 4.30-4.34) more in intervention than control hospitals, corresponding to 21.6 (95%CI 21.5-21.8), i.e., 22 patients treated in intervention hospitals to achieve one additional patient with TO. Intervention hospitals had a larger increase in patients achieving TO (ratio of adjusted odds ratios 1.24, 95%CI 1.05-1.48) than control hospitals, a larger reduction in patients with long LOS (0.74, 95%CI 0.61-0.90) but also a larger increase in patients with reported 30-day complications (1.34, 95%CI 1.00-1.78). Intervention hospitals that introduced QII increased more in TO (1.32, 95%CI 1.10-1.57) than control hospitals, with no effect shown for hospitals not introducing QII (0.93, 95%CI 0.67-1.30).

Conclusion

The multifaceted QI intervention including monthly feedback, education, and a toolbox to facilitate QII effectively improved patients achieving TO. The effect size was associated with the introduction of (evidence-based) QII, considered as the causal link to achieve better patient outcomes.

Trial registration number

NCT04055103.

Summary boxes

What is already known on this topic

Given the increasing number of total hip and knee arthroplasties (THA and TKA) performed worldwide, the number of adverse events and revision surgeries are expected to increase as well as societal costs. High-quality care may reduce the risk of adverse events and improve efficiency by avoiding unnecessarily long length-of-stays (LOS).

What this study adds

A multifaceted quality improvement intervention including frequent feedback on performance, interactive education combined with an action implementation toolbox containing evidenced-based quality improvement initiatives (QIIs) was effective to improve patient outcomes after THA and TKA. The absolute probability of patients achieving Textbook outcome (TO)(i.e., absence of revision, readmission, complications, and length-of-stay (LOS)) increased by 4.32% (95% CI 4.30 to 4.34) more in intervention hospitals than control hospitals, with effect size depending on QII introduced. Intervention hospitals particularly improved more in reducing patients with long LOS.

How this study might affect research, practice and/or policy?

These findings support that frequent feedback to surgical teams should be supplemented by interactive education and facilitated by evidence-based improvement initiatives tailored to specific outcomes, to further improve the quality of delivered patient care in arthroplasty surgery.

Introduction

Total hip and knee arthroplasties (THA and TKA) are frequently used cost-effective treatments for symptomatic osteoarthritis and end-stage rheumatoid arthritis to reduce pain and improve patients' functionality.(1-3) Due to the increasing number of procedures, the absolute number of adverse events and costly revision surgeries are likely to increase if the risk remains the same.(4-7) Several studies and arthroplasty reports have shown considerable between-hospital variation in revision, readmission, complications, and length of stay (LOS) for both THA and TKA, indicating huge improvement potential.(7-17)

In recent years, arthroplasty registries have provided surgeons and hospitals with Audit and Feedback (A&F) on their performance, aiming to improve the quality of care delivered.(7,11-18). A Cochrane review showed A&F to be effective with a median absolute improvement of 4.3% (interquartile range (IQR): 0.5% to 16%).(19) Worldwide, arthroplasty registries include different performance indicators in their feedback, with revision most commonly used.(7,11-18). A recent study showed that Dutch orthopaedic surgeons would like to receive feedback not just on revisions but also on readmission, complications, and LOS.(20) For arthroplasty surgery, even a relatively small absolute improvement will have huge impact considering the large annual number of THA and TKA performed worldwide. Studies have shown that A&F maybe more effective when for example, an action implementation toolbox is added to facilitate actions undertaken instead of a "passive" single element (feedback or education alone).(19,21-23) However, a previous study including such an action implementation toolbox only showed improvement in process indicators whereas the ultimate goal is to improve patient outcomes.(23) In addition, A&F seems to be more effective when feedback is delivered by a senior colleague, at least monthly, in both verbal and written format and when explicit goals and specific actions are planned.(19)

We aimed to evaluate the effect of a prospective multifaceted A&F intervention on a composite of clinical outcomes (including 1-year revision, 30-day readmission, 30-day complications, and long LOS) for patients undergoing THA and TKA.

Methods

Study design and participants

A cluster randomised controlled trial (RCT) was nested in the nationwide Dutch Arthroplasty Register (LROI), including 20 hospitals performing THA and TKA across the Netherlands. During the 21 months pre-intervention period (from 1 January 2018 to 30 September 2019) all 20 hospitals established a data linkage between the registry and hospital data, and the research group developed the feedback and action implementation toolbox. Participating orthopaedic departments were then randomised into 10 intervention and 10 control hospitals. The intervention was applied over an 8 months period (1 October 2019 - 31 May 2020) (Figure 1). Control hospitals continued with usual care . We compared the change in patient outcomes from the pre-intervention to the end of the intervention period between intervention and control hospitals as the effect attributable to introduction of the intervention. The trial was pre-registered (ClinicalTrial.gov, NCT04055103) and the LUMC Medical Ethical Committee waived the need for ethical approval under Dutch law (CME, G18.140).(24) The study was announced on the website of the Dutch Orthopaedic Association and the first 20 orthopaedic departments agreeing to participate in the study were included. All THA and TKA procedures performed in the 20 departments were included. No exclusion criteria were used.

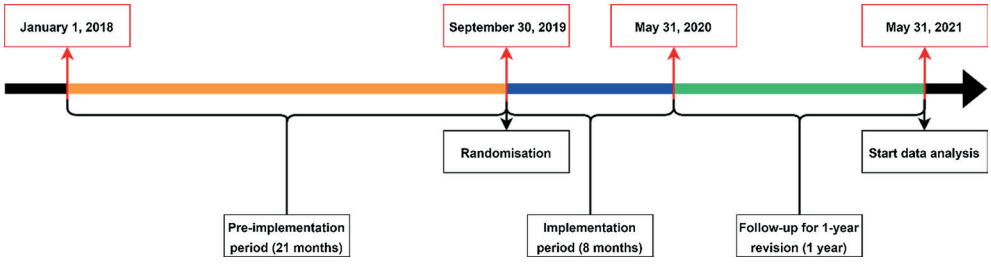


Figure 1 Study period

COVID-19

We originally planned to include a "sustainability phase" after 6 months, where intervention hospitals would no longer be actively supported and the control hospitals would receive the intervention (ClinicalTrial.gov, NCT04055103).(24) This planned sustainability phase was delayed by 2 months so that the intervention period ended May 2020 rather than the planned March 2020 to maintain sufficient statistical power, given the reduction in elective care caused by the COVID-19 outbreak in early March 2020. The number of procedures decreased to 625, 54, and 545 in March, April, and May 2020, respectively, compared with an average 1,215 THA and TKA per month in

2018-2019. In addition, we had planned to match hospitals as part of the intervention to exchange information on best practices and identify areas for improvement, which could not be implemented due to government-imposed COVID-19 restrictions.(24)

Randomisation and masking

Randomisation was stratified by hospital type to achieve an equal distribution of academic, teaching, non-teaching, and private hospitals, as these generally differ in size and are therefore likely to differ in available IT and quality improvement (QI) capacity. Participating hospitals were categorised within one of four groups and then allocated in a 1:1 ratio to the intervention or control group. Due to the nature of the intervention, orthopaedic surgeons of intervention hospitals could not be masked but patients were masked to study group allocation. By liaising with hospital IT specialists to extract hospital data on readmissions, complications, and LOS, we tried to minimize potential bias as they were masked to study group allocation. In all intervention hospitals, the head of the orthopaedic department was appointed as contact person and acted as ''clinical champion".

Intervention

The intervention was designed based on evidence regarding effective feedback(19,21-23) for orthopaedic surgeons(20) and included the following components:

1) Monthly updated feedback was (securely) emailed individually to all orthopaedic surgeons performing THA and TKA in the intervention hospitals. Feedback included case-mix-adjusted indicator outcomes graphically presented in funnel plots and CUSUM charts.(25) The following indicators were reported: 1-year revision (including reasons for revision to align with Quality Improvement Initiatives (QIIs), that is, infection, dislocation (only THA), prosthesis loosening, and technical failure (only TKA)(8)), 30-day readmission, 30-day complications, long (upper quartile) LOS and a composite outcome including all above mentioned indicators. A brief description to interpret the findings for each indicator was provided below each graph tailored to that specific hospital (Appendix I).

2) Education to interpret the feedback was provided by PvS (medical doctor) in the first month of the intervention period, combined with clear targets for improvement of specific indicators. Orthopaedic surgeons learned how to interpret funnel plots and CUSUM charts, and how to use these charts for QI. This was based on a previous survey showing this represented a knowledge gap so that education should be part of the intervention .(20,21) As a reference, a video (Appendix II) and pocket card (Appendix III) were available to summarize the educational meeting.

3) An action implementation toolbox including evidence-based QII for each indicator reported in the feedback, to facilitate taking actions to improve care, based on scientific literature, expert opinion, and guidelines. The plan-do-study-act (PDSA) cycle was added to help surgeons design local QI projects (Appendix IV).

4) A short survey was emailed every 2 months together with the feedback to evaluate adherence to the intervention, encourage reviewing the feedback, verify which QII were introduced, and stimulate trial engagement. Participants could report best practices and experiences to be added to the toolbox and shared with others, also to stimulate trial engagement (Appendix V).

Control hospitals continued with usual care, meaning that no specific intervention was implemented. This means that orthopedic surgeons have access to the password-protected LROI-dashboard where overall between-hospital variation in revision could be viewed in real-time, as well as averages for patient characteristics and patient-reported outcome measures. However, it requires logging in to look up the information, rather than receiving it through email, and gives no comparative information on readmission, complications, and long LOS (or the composite). Since 2015, all surgeons in both control and intervention hospitals have had access to the LROI-dashboard, however, a recent study showed that 39% of the orthopaedic surgeons did not interpret funnel plot correctly, and 34% never logged in.(20)

Graphical displays of performance

Funnel plots are already used in the LROI dashboard as a graphical aid to show between-hospital variation in revisions, adjusted for case-mix. Hospitals plotted between the control limits (2 standard deviations (SD)) perform as expected given their case-mix, while hospitals plotted above or under the control limit perform significantly worse or better, respectively.

A CUSUM chart was added to the monthly feedback since it shows patient-level rather than aggregated performance data during a time period. For every consecutive patient, the observed minus expected probability for an event is plotted. When the score goes up, the observed performance is worse than expected, and vice versa when going down. A signal (alert) was generated when crossing the 5.0 control limit meaning that hospital performance was ''out-of-control" for the quality indicator, after which the chart was reset to zero. When no signal is generated a hospital is "in-control".(26) The level of the control limit determines the trade-off between the number of false-positive and false-negative signals. We showed previously that the CUSUM chart with a 5.0 control limit enabled earlier detection of worsening performance for 1-year revisions

with good accuracy compared with the funnel plot, thereby allowing initiatives to start earlier.(25)

Outcome evaluation

The primary outcome was the Textbook Outcome (TO) composite, with the individual outcomes included in TO analysed as secondary outcomes. The TO composite is an all-or-none concept representing the best outcome on commonly used indicators for THA and TKA (i.e., the absence of 1-year revision, 30-day readmission, 30-day complications, and long LOS). The 1-year revision was calculated based on primary and revision surgery dates, with all patients having at least 1-year follow-up, as routinely collected in the LROI. By including revisions in the composite, a 1-year follow-up was needed after the implementation period to calculate TO (Figure 1). Other outcomes were calculated based on the index hospitalisation in which the primary THA or TKA was performed. Outcomes were defined as follows:

- Revision: Exchange, removal, or addition of any component within one year after the primary surgery.
- Readmission: Any admission within 30 days after discharge of the index hospitalisation to the same hospital.
- Complication: Any complication other than revision during the index hospitalisation or within 30 days after discharge, using the nationwide definition of a complication.(27)
- Long LOS: LOS of the index hospitalisation longer than the 75th percentile (upper quartile), based on all patients treated, to take into account possible between-hospital differences in sensitivity to report complications.(28)

Data collection

Routinely submitted LROI data regarding revisions and patient characteristics were used, supplemented with hospital data on readmission, complications, and LOS for all patients. LROI data were linked to hospital data by an IT specialist from each hospital to ensure anonymous data exchange. LROI data completeness is checked annually against Hospital Electronic Health Records and currently exceeds 99% for primary procedures, and 97% for revisions.(29,30) The LROI uses barcode scanning to enable tracing of prosthetic components so revisions performed in another hospital are included. The following patient-level LROI data were provided: whether a revision had taken place, reason for revision and the patient characteristics age at surgery,

gender, body mass index (kg/m^2), current smoking status (yes/no), American Society of Anaesthesiologists (ASA) classification (I-IV) and diagnosis (osteoarthritis/non-osteoarthritis). Collected data were locked prior to the analyses.

Statistical analysis

At least 18 participating hospitals (9 per arm) were needed to detect a difference in TO of 70% versus 80% with 80% power, α of 0.05, a median of 100 procedures per hospital, and assuming an intra-hospital correlation of 0.02. We included 20 hospitals (10 per arm) in case hospitals would drop out. To assess whether participating hospitals were a representative selection, we compared the median number of procedures and median percentage of revisions for both THA and TKA during the study period in participating centers versus all other Dutch centers, using a Mann-Whitney U test.

Patient characteristics were missing in less than 2% of patients. These were considered missing at random and imputed using multiple imputations for 10 rounds with predictive mean matching as the underlying model. All patient characteristics and outcomes (i.e., TO, revision, readmission, complications, and long LOS) were used as predictors, but only patient characteristics were imputed.

Data were analyzed following an intention to treat approach, classifying hospitals in study groups as randomised. Random effects binary logistic regression models were used to estimate the impact of the intervention on TO as the primary outcome and each secondary outcome, while accounting for patient clustering within hospitals. All models were adjusted for all measured patient characteristics. Outcomes between pre-intervention and intervention period were first compared within intervention and control hospitals, expressed as adjusted odds ratios. The change in outcomes from pre-intervention period was then compared between intervention and control hospitals by including an interaction term between study group and period, and quantified as the ratio of adjusted odds ratios with corresponding 95% confidence intervals (CI). We calculated the number of patients needed to be treated during the intervention period to achieve one additional patient with TO in intervention hospitals as 1 divided by the absolute risk difference. The absolute difference in TO probability was derived from the estimated parameters obtained by the above logistic regression models using a marginal standardisation method.(31) Corresponding 95% CI were computed from non-parametric bootstrapping based on 2000 replications. The same analyses were performed to compare intervention hospitals introducing QII and intervention hospitals not introducing QII, versus control hospitals. These analyses were conducted from the rationale that these QII were the causal link to achieve improved patient outcomes, which would thereby support intervention efficacy. Information on QII introduced was taken from the process evaluation (see below). All

analyses were conducted for all patients, and separately for THA and TKA patients, given the known difference in revisions and complication risks and that a difference in baseline risk may affect the absolute risk reduction achieved.(7-17,32)

All p values were two-sided and values below 0.05 were considered statistically significant in all analyses. Analyses were performed using STATA (version 14; StataCorp).

Process evaluation and intervention fidelity

Surveys were sent by email in November 2019, January 2020, and March 2020 and compiled using Qualtrics^{XM} (Appendix V). As surveys were sent together with the feedback, response also indicated the email was read and feedback received. Questions were asked to evaluate adherence to intervention components and therefore included whether orthopaedic surgeons could interpret the feedback and what other information or tools were needed for further improvement. In addition, we asked which QIIs were undertaken as ultimately the feedback was intended to initiate actions, including whether these QII were based on the toolbox or other evidence. Descriptive statistics were used to explore the number of QII per intervention hospital and the source of the initiatives.

Patient and public involvement

Patients or the public were not involved in the design of the study.

Results

Of the 33,205 patients who underwent THA or TKA in the 20 participating hospitals during the pre-intrevention and intervention period, 282 had missing data for TO, leaving 32,923 (99.2%) patients eligible for analysis. Of these, 16,314 patients were analysed in the intervention hospitals (12,475 before and 3,839 during the intervention period) versus 16,609 in control hospitals (12,853 versus 3,756). Participating hospitals were comparable to all other Dutch hospitals in distribution of type of hospital, median revision rate (1.7% versus 1.7% for THA, p=1.00 and 1.4% versus 0.9% for TKA, p=0.62) and median number of THA surgeries (930 versus 699; p=0.21) but had higher median number of TKA surgeries (700 versus 582; p<0.05) (Figure 2).

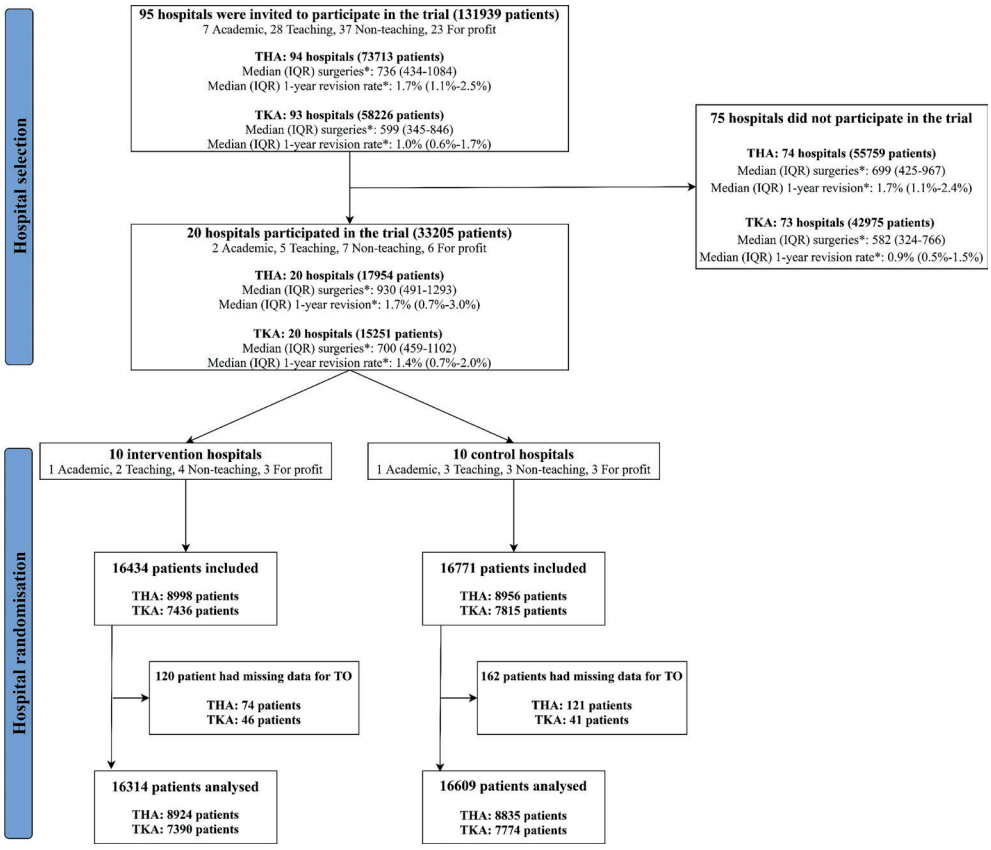


Figure 2 Trial profile

Table 1 shows that hospital and patient characteristics were comparable between intervention and control hospitals, except for slightly more smokers and fewer ASA III-IV patients in intervention hospitals. During the study period, 28,108 patients achieved TO (85.4%), 529 (1.6%) underwent a revision within one year, 1218 (3.7%) had a readmission within 30 days, 1,214 (3.7%) experienced a complication within 30 days, and 3,662 (11.1%) had a long LOS, with considerable between-hospital variation in all outcomes (shown by the interquartile range in table 1) in both intervention and control hospitals. LOS was not normally distributed, making it challenging to create equal quartiles so that the closest integer value was chosen. This resulted in above 4 days defined as long LOS for both THA and TKA, and explains that the median percentage of patients with long LOS is considerably smaller than 25%.

Table 1 Hospital and patient characteristics by hospital group during the study period (1 January 2018 to 31 May 2020)

Hospitals characteristics	**All hospitals (*n*=20)**			**Intervention hospitals (*n*=10)**			**Control hospitals (*n*=10)**		
Geographical region:									
Northwest		3 (15)			1 (10)			1 (10)	
Northeast		4 (20)			2 (20)			2 (20)	
Southwest		6 (30)			3 (30)			2 (20)	
Southeast		5 (25)			3 (30)			3 (30)	
Central area		2 (10)			1 (10)			2 (20)	
Status:									
Academic		2 (10)			1 (10)			1 (10)	
Teaching		5 (25)			2 (20)			3 (30)	
Non-teaching		7 (35)			4 (40)			3 (30)	
For-profit		6 (30)			3 (30)			3 (30)	
Median (IQR) number of surgeons*:		6 (5-7)			6 (5-7)			7 (5-9)	
	All surgeries	**THA**	**TKA**	**All surgeries**	**THA**	**TKA**	**All surgeries**	**THA**	**TKA**
Median (IQR) number of surgeries in study period*:	1472 (1059-2371)	920 (487-1292)	667 (454-1098)	1472 (1350-2152)	920 (723-1133)	612 (388-1041)	1452 (766-2525)	772 (257-1473)	728 (605-1177)
Indicator outcomes:									
Median (IQR) TO (%)*	85.1 (77.0-95.6)	84.4 (75.9-95.6)	84.5 (77.1-95.7)	87.0 (75.0-96.6)	86.9 (76.8-96.9)	84.5 (72.2-95.5)	85.0 (77.2-95.0)	83.6 (71.8-94.1)	85.0 (79.9-96.1)
Median (IQR) revision (%)*	1.4 (0.8-2.6)	1.7 (0.7-3.0)	1.3 (0.7-2.0)	1.6 (0.8-3.0)	1.7 (0.6-3.2)	1.7 (0.8-2.2)	1.3 (0.9-2.3)	1.8 (0.8-3.1)	0.9 (0.7-1.8)
Median (IQR) readmission (%)*	4.4 (1.6-5.0)	4.3 (1.8-5.1)	4.0 (1.6-5.2)	4.6 (1.2-4.9)	4.0 (1.1-5.4)	4.7 (2.5-5.4)	4.2 (1.8-5.1)	4.3 (2.2-5.1)	3.6 (1.4-5.2)
Median (IQR) complications (%)*	3.0 (1.9-4.5)	3.6 (2.1-5.9)	2.5 (1.3-3.7)	2.2 (1.5-3.9)	2.8 (1.6-4.7)	2.3 (1.2-3.9)	3.3 (2.9-6.1)	4.1 (3.2-7.6)	2.8 (1.0-4.6)
Median (IQR) Long LOS (%)*	10.9 (2.1-19.7)	10.3 (2.1-19.3)	9.9 (2.1-19.4)	8.6 (2.0-20.7)	9.3 (2.1-19.0)	9.5 (1.7-23.3)	11.4 (1.5-19.6)	10.9 (1.4-25.0)	11.8 (2.1-17.1)

Table 1 Continued

Hospitals characteristics	All hospitals (*n*=20)			Intervention hospitals (*n*=10)			Control hospitals (*n*=10)		
Patients characteristics	**All surgeries (*n*=32923)**	**THA (*n*=17759)**	**TKA (*n*=15164)**	**All surgeries (*n*=16314)**	**THA (*n*=8924)**	**TKA (*n*=7390)**	**All surgeries (*n*=16609)**	**THA (*n*=8835)**	**TKA (*n*=7774)**
Mean (SD) age (years)	68.5 (9.6)	68.5 (10.3)	68.4 (8.7)	68.1 (9.6)	67.9 (10.3)	68.3 (8.6)	68.9 (9.5)	69.1 (10.2)	68.6 (8.7)
Gender, female	20656 (62.7)	11323 (63.8)	9333 (61.5)	10233 (62.7)	5670 (63.5)	4563 (61.7)	10423 (62.8)	5653 (64.0)	4770 (61.4)
Mean (SD) BMI (kg/m^2)	28.1 (4.8)	27.0 (4.5)	29.4 (4.8)	28.0 (4.8)	26.9 (4.4)	29.3 (4.8)	28.3 (4.8)	27.2 (4.5)	29.5 (4.9)
Smoking	3068 (9.3)	1830 (10.3)	1238 (8.2)	1595 (9.8)	964 (10.8)	631 (8.5)	1473 (8.9)	866 (9.8)	607 (7.8)
ASA classification								1392 (15.8)	959 (12.3)
• ASA I	5158 (15.7)	3204 (18.0)	1954 (12.9)	2807 (17.2)	1812 (20.3)	995 (13.5)	2351 (14.2)	5501 (62.3)	5138 (66.1)
• ASA II	21083 (64.1)	11029 (62.1)	10054 (66.3)	10444 (64.0)	5528 (62.0)	4916 (66.5)	10639 (64.1)	1940 (22.0)	1676 (21.6)
• ASA III-IV	6673 (20.3)	3521 (19.8)	3152 (20.8)	3057 (18.7)	1581 (17.7)	1476 (20.0)	3616 (21.8)		
Diagnosis								7744 (89.4)	7502 (98.4)
• Osteoarthritis	30258 (93.7)	15643 (89.8)	14615 (98.2)	15012 (93.7)	7899 (90.2)	7113 (97.9)	15246 (93.6)	921 (10.6)	119 (1.6)
• Non-osteoarthritis	2051 (6.3)	1780 (10.2)	271 (1.8)	1011 (6.3)	859 (9.8)	152 (2.1)	1040 (6.4)		

Values are numbers (percentages) unless stated otherwise. Data relative to the pre-implementation and implementation period were pooled in intervention and control hospitals. Percentages for ASA classifications might not sum to 100 because of roundings. ASA=American Society of Anaesthesiologists; BMI=body mass index; IQR=Interquartile range; LOS=length-of-stay; SD=standard deviation; THA=total hip arthroplasty; TKA=total knee arthroplasty; TO=Textbook Outcome.

*The value under "Median (IQR)" indicates percentage of the median hospital.

Table 2 Comparison of primary and secondary outcomes by study group (i.e., intervention or control hospitals) between pre-implementation and implementation period

Surgical outcomes	Intervention hospitals			Control hospitals			Intervention versus control hospitals		
	Pre-implementation	Implementation	Implementation versus pre-implementation adjusted odds ratio (95% CI)	Pre-implementation	Implementation	Implementation versus pre-implementation adjusted odds ratio (95% CI)	Ratio of adjusted odds ratios (95% CI)	*p*-value	ICC
All surgeries (*n*=32923)									
Textbook Outcome*	10930 (87.6)	3421 (89.1)	1.39 (1.23 to 1.58)	10615 (82.6)	3142 (83.7)	1.14 (1.02 to 1.27)	1.24 (1.05 to 1.48)	0.011	0.299
1-year revision	199 (1.6)	61 (1.6)	0.93 (0.69 to 1.24)	202 (1.6)	67 (1.8)	1.12 (0.85 to 1.48)	0.82 (0.55 to 1.23)	0.341	0.074
30-days readmission	399 (3.2)	118 (3.1)	0.89 (0.72 to 1.10)	547 (4.3)	154 (4.1)	0.95 (0.79 to 1.14)	0.95 (0.71 to 1.25)	0.697	0.108
30-days complications	303 (2.4)	112 (2.9)	1.12 (0.90 to 1.40)	638 (5.0)	161 (4.3)	0.85 (0.71 to 1.01)	1.34 (1.00 to 1.78)	0.046	0.208
Long LOS	1154 (9.3)	302 (7.9)	0.67 (0.57 to 0.77)	1731 (13.5)	475 (12.6)	0.85 (0.71 to 1.01)	0.74 (0.61 to 0.90)	0.002	0.533
Total hip arthroplasty (*n*=17759)									
Textbook Outcome*	6015 (88.0)	1863 (89.3)	1.39 (1.17 to 1.67)	5533 (80.9)	1611 (80.8)	1.06 (0.92 to 1.23)	1.34 (1.06 to 1.69)	0.013	0.303
1-year revision	112 (1.6)	36 (1.7)	0.98 (0.69 to 1.43)	130 (1.9)	44 (2.2)	1.13 (0.80 to 1.60)	0.86 (0.51 to 1.30)	0.572	0.104
30-days readmission	224 (3.3)	61 (2.9)	0.82 (0.61 to 1.10)	295 (4.3)	104 (5.2)	1.20 (0.95 to 1.51)	0.69 (0.48 to 1.00)	0.051	0.105
30-days complications	194 (2.8)	72 (3.4)	1.12 (0.85 to 1.49)	390 (5.7)	106 (5.3)	0.90 (0.72 to 1.13)	1.25 (0.87 to 1.80)	0.218	0.219
Long LOS	616 (9.0)	164 (7.9)	0.66 (0.54 to 0.82)	1025 (15.0)	296 (14.8)	0.91 (0.77 to 1.07)	0.70 (0.54 to 0.92)	0.010	0.531
Total knee arthroplasty (*n*=15164)									
Textbook Outcome*	4915 (87.2)	1558 (88.9)	1.39 (1.16 to 1.67)	5082 (84.5)	1531 (86.9)	1.25 (1.05 to 1.48)	1.12 (0.87 to 1.44)	0.371	0.333
1-year revision	87 (1.5)	25 (1.4)	0.89 (0.57 to 1.40)	72 (1.2)	23 (1.3)	1.11 (0.69 to 1.79)	0.80 (0.41 to 1.53)	0.492	0.053
30-days readmission	175 (3.1)	57 (3.3)	0.98 (0.72 to 1.33)	252 (4.2)	50 (2.8)	0.65 (0.48 to 0.89)	1.50 (0,97 to 2.32)	0.071	0.095
30-days complications	109 (1.9)	40 (2.3)	1.13 (0.78 to 1.65)	248 (4.1)	55 (3.1)	0.75 (0.57 to 1.01)	1.52 (0.94 to 2.46)	0.085	0.204
Long LOS	538 (9.5)	138 (7.9)	0.66 (0.53 to 0.82)	706 (11.7)	179 (10.2)	0.83 (0.68 to 1.00)	0.79 (0.59 to 1.06)	0.115	0.515

Values are numbers (percentages) unless stated otherwise. Odds ratios were estimated using random effect binary logistic regression models to compare surgical outcomes between pre-implementation and implementation periods in intervention and control hospitals. The ratio of adjusted odds ratios captured the effect of the intervention by comparing the change in outcomes from the pre-implementation to implementation periods between the intervention and control hospitals based on the interaction term. A ratio of adjusted odds ratios value above 1 for the Textbook outcome indicates improvement in the intervention versus control hospitals and a value below 1 deterioration. In contrast, a ratio of adjusted odds ratios value below 1 for 1-year revision, 30-days readmission, 30-days complications, and long LOS indicates improvement and a value below 1 deterioration. Estimates and 95% CI considered clustering of patients at the hospital level. All outcomes were adjusted for gender, age, body mass index, smoking, American Society of Anesthesiologists score, and diagnosis (osteoarthritis versus non-osteoarthritis)

*Composite of an all-or-none concept with the best outcome on all performance indicators (i.e., 1-year revision, 30-day readmission, 30-day complications, and long LOS).

CI=confidence interval; ICC=intraclass correlation; LOS=length-of-stay; PROMs=patient-reported outcome measures.

Outcome evaluation

Table 2 shows changes in clinical outcomes from the pre-intervention to intervention period for both intervention and control hospitals. Intervention hospitals significantly improved in achieving more patients with TO over time for THA/TKA combined (adjusted odds ratio 1.39, 95 % CI 1.23-1.58), as did control hospitals (adjusted odds ratio 1.14, 95% CI 1.02-1.48). Even though intervention hospitals had better pre-intervention TO performance, that is, potentially less room for improvement, they improved significantly more than control hospitals (ratio of adjusted odds ratios 1.24, 95% CI 1.05-1.48). The effect was also significant for THA alone (1.34, 95% CI 1.06-1.69), but not for TKA (1.12, 95% CI 0.87-1.44) although it went in the same direction. For the secondary outcomes, intervention hospitals also showed a significantly higher reduction in the percentage of patients with long LOS than control hospitals for THA/TKA combined and THA. The same trend was observed for 30-day readmission for THA, albeit non-significant. The percentage of patients with reported 30-day complications increased more in intervention than control hospitals for THA/TKA combined but not for THA or TKA separately. No significant effects were found for revisions.

The absolute probability of TO increased by 4.32% (95% CI 4.30-4.34) more in intervention hospitals than control hospitals, corresponding to 21.6 (21.5-21.8), that is, 22 patients treated in intervention hospitals to achieve one additional patient with TO.

Process evaluation and intervention fidelity

The education meetings were scheduled such that all orthopedic surgeons could attend (unless still in surgery). Each meeting ended by discussing which performance indicators would be the focus of improvement initiatives and which specific actions would be taken. Two orthopaedic surgeons were absent during this interactive education session in 3 intervention hospitals and 1 surgeon in 3 hospitals, meaning that 52 of the total of 61 orthopaedic surgeons (85%) attended. From all surgeons, 45 (74%) completed the first survey, 39 (64%) the second survey, and 35 (57%) the third survey. Fifty-five surgeons (90%) completed the survey at least once, meaning that the feedback was reviewed by at least 90% of the surgeons since it was sent together with the survey. Twenty-three (38%) orthopedic surgeons completed the survey at all time points, and at least 1 surgeon for each hospital. In addition, 91% of respondents reported the feedback was clear after receiving the education. In terms of trial engagement, 4 hospitals reported they needed additional educational explanations on funnel plots and CUSUM charts, and 2 hospitals would appreciate more QIIs included in the toolbox. In addition, 7 hospitals requested being linked to a hospital that scored better on a performance indicator to improve further ("learning from the best").

Table 3 Quality improvement initiatives per hospital and the source of the initiatives

Intervention hospital	Quality improvement initiatives	Toolbox	Literature	Expert opinion
1	*LOS:*			
	- Discharge 1 day postoperative if possible.	No	Yes	No
2	*Revision:*			
	- Reduce the number of dislocations for THA by no longer placing a 28mm cup and placing an "Avantage" cup earlier in older patients.	Yes	Yes	Yes
	- Pairing surgeons with more dislocations with surgeons with few dislocations to learn from best practices.	No	Yes	Yes
	LOS:			
	- Start mobilizing earlier after surgery.	No	No	No
	- Improve patient expectation management.	No	Yes	No
	- Earlier consultation of transfer agency.	No	Yes	No
3	*Readmission and complications:*			
	- Reduce wound leakage and surgical site infections by adjusting the wound closure technique, tissue protector for THA, and tranexamic acid during wound closure for TKA.	No	Yes	Yes
4	*LOS:*			
	- Earlier consultation of transfer agency.	No	No	Yes
5	*Revision, readmission, and complications:*			
	- Reduce surgical site infections and prosthetic joint infections by adjusting the wound closure technique (Roerdink et al, 2019).	No	Yes	Yes
	- Covering the sterile surgical field differently.	No	No	Yes
	- Short-term use of the tourniquet for TKA.	No	No	Yes
	- Use of prophylactic antibiotic as suggested in de guidelines of the Netherlands Orthopaedic association.	No	Yes	No
6	*Revision:*			
	- Reduce the number of infections by adopting pre-operative, intra-operative and post-operative interventions from the toolbox and the literature (not defined).	Yes	Yes	No
	LOS:			
	- Earlier consultation of transfer agency.	Yes	Yes	No
7	*LOS:*			
	- Mobilizing on the day of surgery.	Yes	No	No
	- Inform the patient before surgery about the expected LOS.	Yes	No	No
8	*Revision:*			
	- Introduction of a new type of prosthesis.	No	Yes	Yes
	- Introduction of an infection discussion in which improvement initiatives are evaluated.	No	Yes	No
	LOS:			
	- Prevent wound leakage by keeping the compression bandage in place longer in patients who have had surgery late in the day.	No	No	Yes
	- Closing the fascia with polydioxanone suture.	No	No	Yes
	- Close the subcutis in 2 layers.	No	No	Yes
	- Improve patient flow to the care hotel.	No	No	Yes

These are the quality improvement initiatives as reported in the bi-monthly surveys by the orthopaedic surgeons in the intervention hospitals. The initiatives are described under the indicator that the hospital aimed to improve with the initiative. However, the quality initiatives mentioned could affect other indicators, both positively and negatively. Two hospitals did not introduce any initiatives and are not included in the table.
LOS=length-of-stay; THA=total hip arthroplasty; TKA=total knee arthroplasty.

Table 3 shows descriptions of QIIs introduced in each hospital, intended to improve patient outcomes, including whether these were taken from the toolbox or based on other evidence. The median number of performance indicators for which QII were undertaken per hospital was 2 (IQR 1-2). Two hospitals did not introduce any QII, and of the remaining 8 hospitals most introduced QII to improve LOS.

Intervention hospitals that introduced QIIs improved significantly more in TO than control hospitals (1.32, 95% CI 1.10-1.57), whereas intervention hospitals not introducing any QII showed similar changes as control hospitals (0.93, 95% CI 0.67-1.30) (Figure 3). Of note, pre-intervention TO on average was lower for intervention hospitals that introduced QII compared with hospitals not introducing QII (85.2% versus 94.5%, $p<0.01$) with control hospitals at 82.6%. For the secondary outcomes, intervention hospitals that introduced QII to reduce long LOS improved significantly more than control hospitals. For complications, no difference was found for intervention hospitals that introduced QII targeting complications but hospitals not introducing these QII increased more in complications than control hospitals. No significant differences were found for hospitals introducing QII to reduce revisions or readmissions.

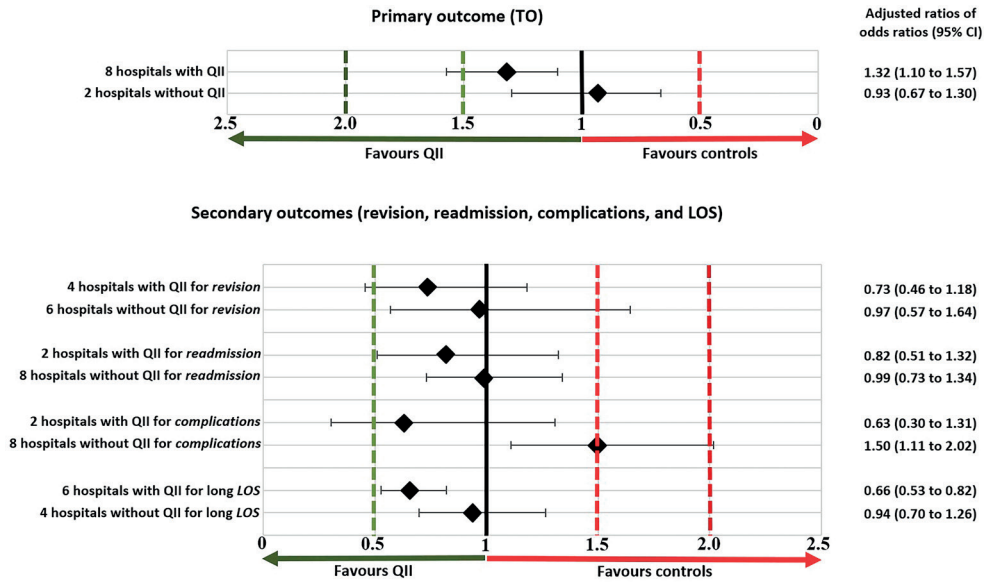


Figure 3 Primary and secondary outcomes by implementation of quality improvement initiatives
LOS=length of stay; TO=textbook outcome.

Discussion

The present study has shown that the multifaceted QI intervention was effective to increase the percentage of patients achieving TO more in intervention than control hospitals. Intervention hospitals that introduced QII improved significantly more in performance on TO than control hospitals, whereas intervention hospitals not introducing any QII showed comparable changes as control hospitals (but had higher pre-implementation TO). For the secondary outcomes, a higher reduction in patients with long LOS was found for intervention than control hospitals, and hospitals introducing QII to reduce LOS improved significantly more than control hospitals whereas hospitals not introducing these QIIs showed similar changes. Effects for readmission and revision seemed to go in the same direction, but were non-significant. However, intervention hospitals also showed a higher increase in the percentage of patients with reported complications than control hospitals. This seemed to be due to hospitals not introducing QII targeting complications, as those hospitals showed a higher increase in reported complications than control hospitals whereas hospitals introducing QIIs targeting complications showed similar changes. In addition, it may reflect increased sensitivity in reporting complications associated with the intervention. Taken together, these findings suggest that our multifaceted QI intervention was effective to improve TO for THA/TKA patients, most likely through the introduction of targeted QII particularly reducing long LOS.

Although a significant effect was only found for THA, the direction of the effect and some secondary outcomes (e.g., long LOS) were similar for TKA. This may be due to the smaller volume, as the median number of TKA surgeries was considerably smaller than for THA in intervention hospitals (Table 1), which may explain why the effect went in the same direction but was not significant. In addition, the baseline risks for revisions and complications were lower for TKA, mostly associated with smaller absolute risk reductions.

Comparison to the literature

The present study showed an absolute larger improvement of 4.32% in intervention versus control hospitals, similar to the median improvement shown for A&F interventions in a Cochrane review including 140 studies.(19) This suggests that about 50% of included studies in that review had smaller effects than the present study. A comparable cluster RCT using control charts and regular feedback resulted in an absolute reduction of major adverse events of 0.9%, or 114 patients needed to treat in intervention hospitals to prevent 1 adverse event.(33) However, such comparisons need to be done cautiously as included studies involve different populations being targeted and different control groups. For interventions like statins and aspirin it

is known that both the absolute reduction and the number needed to treat (NNT) depend on the baseline risk.(32) This is equally relevant in our study, as the baseline risk for particularly revision and complications are already low, meaning that absolute risk reductions tend to be lower. This likely explains why the overall effect is driven by LOS and readmissions, with higher baseline risk.

Other QI initiatives have been described within orthopedics, such as the Continuous Quality Improvement Program for hip and knee replacement surgical care Canada.(34) A standardised care pathway was developed guided by the Triple Aim framework and six quality dimensions derived from the Institute of Medicine, using key performance indicators and benchmarked to give feedback twice a year to individual physicians, hospital administrators, and quality review teams on how they compare against a set threshold of good quality. Currently, 83% of orthopaedic surgeons participate in the program, representing 95% of the total volume of THA and TKA. In another QI project performed in the UK, a reduction in LOS was achieved from 3.6 to 2.4 days in one hospital for THA and TKA and 3.6 to 2.0 days in another, both by the introduction of PDSA cycles to improve on postoperative analgesia, physiotherapy and local policy.(35)

A previous study targeting quality of pain management in intensive care units showed an improvement in pain management when an action implementation toolbox was added to feedback compared with feedback alone, but only in process indicators and not in clinical outcomes.(23) The present study therefore adds that a comparable intervention where the toolbox included evidence-based measures targeting outcomes rather than merely process measures such as having a protocol in place, was effective in improving patient outcomes.

Strengths and limitations of this study

The strengths of this study include the robust randomised trial design, limited selection bias given that LROI data include more than 99% of all primary THA and TKA performed, and the required sample size of the power calculation achieved. [29,30] In addition, the intervention was developed guided by evidence and following the latest theory and recommendations.(22,23,36) The risk of contamination among control hospitals seems unlikely because control hospitals were not aware of the start of the implementation period and received the intervention at a later point in time, and feedback was tailored to a specific hospital. If contamination did occur, this would have diluted the intervention effect so that the true effect would potentially be larger. In addition, given the lower pre-implementation TO in control hospitals than in intervention hospitals, one would expect a larger change in control hospitals due to

more potential for improvement and regression to the mean, making the opposite effect all the more notable. (19,37)

However, some limitations remain. First, the potential influence of a Hawthorne effect on study findings was largely compensated by control hospitals, as performance improved in both hospital groups. Second, information bias may occur if coding accuracy changed within hospitals between periods and differently for intervention than control hospitals. This seems unlikely, including that it would occur to such an extent that it would explain our results. Third, since outcome frequencies vary considerably between performance indicators, TO is dominated by long LOS, a well-known disadvantage of binary all-or-none composite outcomes.(38,39) However, estimates for the individual outcomes largely went in the same direction, albeit non-significant. Fourth, implementing the intervention in a specific country and for a specific type of surgery limits the generalisability of the results. Thus, the feasibility and impact of the intervention in a different context requires further study. Finally, patients were not involved in the design of this study which could have resulted in different outcomes being targeted by QI initiatives.

Implications and future research

Even though the multifaced QI intervention in the present study was shown to improve the quality of delivered care, the question is what is needed for hospitals to sustain these effects and potentially continue improving further. Only few studies describe how QI interventions became adopted in everyday practice.(40-43) Implementing a bundle of common QI interventions (e.g., staff education, A&F, alerts) to "quick fix" poor hospital performance may provide a temporary solution, but is not sustainable.(40,41) It appears from the emerging literature that sustainable interventions must provide solutions for the underlying problem. Only through understanding the problem, both an effective and sustainable intervention can be created that becomes part of everyday practice in the long term.(44) This may require that first an effective intervention needs to be found, to then solve how it can be adapted for everyday practice to be sustainable, using resources that remain available after the QI intervention ends.(45) In the present study's design, it was taken into account that intervention components would remain available for the LROI to apply them in other hospitals if the intervention proved to be effective. The CUSUM charts developed for the intervention are currently being implemented by the LROI.(25) In addition, the educational video and pocket card remain available, as well as the toolbox which can be kept up-to-date and further expanded with new effective QII appearing in the literature. Also, an annual educational session or workshop may keep hospitals both engaged to continue improving their care and act as further education, where participants in the current study may act as champions to share what worked

and what not. Further engagement can also be supported by the toolbox being continuously updated as participants share and use each other's best practices and experiences, rather than a static list that may become outdated when new evidence appears.

Conclusions

The effect of QI interventions is known to vary, but an optimal design will reasonably improve patient care. The multifaceted intervention in the present study has shown that monthly updated feedback, education, an evidence-based implementation toolbox with suggested QII, and bimonthly surveys achieve a statistically significant larger improvement in outcomes for patients undergoing THA and TKA. The intervention effect was most likely achieved by introducing targeted QII particularly reducing long LOS.

Collaborators

We gratefully acknowledge the Dutch Arthroplasty Register (L.N. van Steenbergen, PhD) and the 20 hospitals who provided their data to complete this study, as part of the IQ Joint study group (in alphabetic order): Antonius Hospital, Sneek (S.T. Hokwerda, MD); Bergman Clinics (P.M. van Kampen, PhD); Bergman Clinics, Arnhem (I. Buchholz, MD); Bergman Clinics, Breda (J. Schrier, MD); Bergman Clinics, Delft (F. de Graaff, MD); Bergman Clinics, Naarden (H. Bouma, MD); Bergman Clinics, Rijswijk (T. Hogervorst, MD, PhD and J. Wolkenfelt, MD); Bergman Clinics, Rotterdam (M. Vischjager, MD, PhD); Catharina Hospital, Eindhoven (R.W.T.M. van Kempen, MD); Dijklander Hospital, Hoorn (G.C. Huitema, MD); Dijklander Hospital, Hoorn (L. de Vries, PhD); Elisabeth-TweeSteden Hospital (T. Gosens, MD, PhD); Gelderse Vallei Hospital, Ede (W. Beijneveld, MD); Maxima Medical Centre, Eindhoven (M. van den Besselaar, MD); Medical Spectrum Twente, Enschede (W. Verra, MD, PhD); OLVG, Amsterdam (R.W. Poolman, MD, PhD); OLVG, Amsterdam (V.A. Scholtes, PhD); Sint Anna Hospital, Geldrop (W. van der Weegen, PhD); Sint Franciscus Hospital, Schiedam (A. Polak, MD); Tjongerschans Hospital, Heerenveen (M. Mulder, PhD); University of Groningen, University Medical Center Groningen, Groningen (M. Stevens, PhD); Zuyderland Hospital, Sittard (B. Boonen, MD, PhD).

References

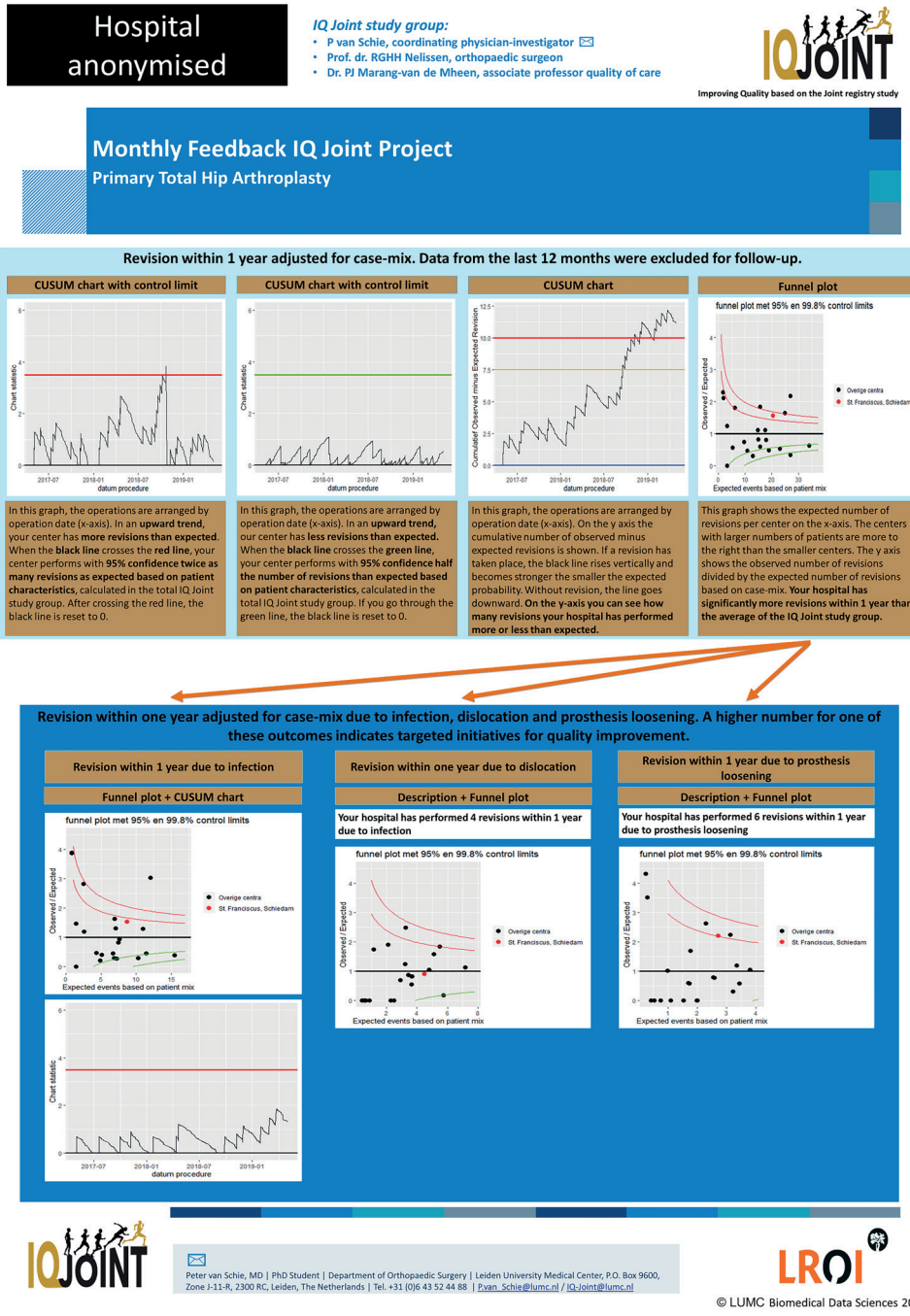
1. Nelson AE, Allen KD, Golightly YM, Goode AP, Jordan JM. A systematic review of recommendations and guidelines for the management of osteoarthritis: The chronic osteoarthritis management initiative of the U.S. bone and joint initiative. Semin Arthritis Rheum. 2014;43(6):701-12.
2. Bachmeier CJ, March LM, Cross MJ, Lapsley HM, Tribe KL, Courtenay BG, et al. A comparison of outcomes in osteoarthritis patients undergoing total hip and knee replacement surgery. Osteoarthritis Cartilage. 2001;9(2):137-46.
3. Wilson RA, Gwynne-Jones DP, Sullivan TA, Abbott JH. Total Hip and Knee Arthroplasties Are Highly Cost-Effective Procedures: The Importance of Duration of Follow-Up. J Arthroplasty. 2021.
4. Kurtz SM, Ong KL, Lau E, Bozic KJ. Impact of the economic downturn on total joint replacement demand in the United States: updated projections to 2021. J Bone Joint Surg Am. 2014;96(8):624-30.
5. Khan M, Osman K, Green G, Haddad FS. The epidemiology of failure in total knee arthroplasty: avoiding your next revision. Bone Joint J. 2016;98-b(1 Suppl A):105-12.
6. Kurtz SM, Lau E, Watson H, Schmier JK, Parvizi J. Economic burden of periprosthetic joint infection in the United States. J Arthroplasty. 2012;27(8 Suppl):61-5.e1.
7. Dutch Arthroplasty Register (LROI). Online Annual Report. https://www.lroi-report.nl/.
8. van Schie P, van Steenbergen LN, van Bodegom-Vos L, Nelissen R, Marang-van de Mheen PJ. Between-Hospital Variation in Revision Rates After Total Hip and Knee Arthroplasty in the Netherlands: Directing Quality-Improvement Initiatives. J Bone Joint Surg Am. 2019.
9. Bozic KJ, Grosso LM, Lin Z, Parzynski CS, Suter LG, Krumholz HM, et al. Variation in hospital-level risk-standardized complication rates following elective primary total hip and knee arthroplasty. J Bone Joint Surg Am. 2014;96(8):640-7.
10. Fry DE, Pine M, Nedza SM, Locke DG, Reband AM, Pine G. Risk-Adjusted Hospital Outcomes in Medicare Total Joint Replacement Surgical Procedures. J Bone Joint Surg Am. 2017;99(1):10-8.
11. Norwegian Arthroplasty Register (NAR). Annual Report 2020. http://nrlweb.ihelse.net/eng/Rapporter/Report2020_english.pdf.
12. Swedish Knee Arthroplasty Register (SKAR). Annual Report 2020. http://myknee.se/pdf/SVK_2020_Eng_1.0.pdf.
13. Danish Hip Arthroplasty Register (DHAR). Annual report 2020. http://danskhoftealloplastikregister.dk/wp-content/uploads/2020/11/dhr-aarsrapport-2020_til_offentliggoerelse-1.pdf.
14. Swedish Hip Arthroplasty Register (SHAR). Annual Report 2018. https://registercentrum.blob.core.windows.net/shpr/r/Arsrapport_2018_Hoftprotes_ENG_26mars_Final-rJepCXNsLI.pdf.
15. Danish Knee Arthroplasty Register (DKAR). Annual Report 2020. https://www.sundhed.dk/content/cms/99/4699_dkr-arsrapport-2020_offentliggorelse.pdf.
16. Finnish Arthroplasty Register (FAR). Online Annual Report. https://www.thl.fi/far/#index.
17. Swiss Arthroplasty Register (SAR). Annual Report 2020. https://www.siris-implant.ch/de/Downloads&category=16.
18. National Joint Registry (NJR). Annual report 2021. https://reports.njrcentre.org.uk/Portals/0/PDFdownloads/NJR%2018th%20Annual%20Report%202021.pdf.
19. Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, French SD, et al. Audit and feedback: effects on professional practice and healthcare outcomes. Cochrane Database Syst Rev. 2012(6):Cd000259.

20. Van Schie P, Van Bodegom-Vos L, Zijdeman TM, Nelissen R, Marang-Van De Mheen PJ. Awareness of performance on outcomes after total hip and knee arthroplasty among Dutch orthopedic surgeons: how to improve feedback from arthroplasty registries. Acta Orthop. 2020:1-8.
21. Soong C, Shojania KG. Education as a low-value improvement intervention: often necessary but rarely sufficient. BMJ quality & safety. 2020;29(5):353-7.
22. Gude WT, Roos-Blom MJ, van der Veer SN, de Jonge E, Peek N, Dongelmans DA, et al. Electronic audit and feedback intervention with action implementation toolbox to improve pain management in intensive care: protocol for a laboratory experiment and cluster randomised trial. Implement Sci. 2017;12(1):68.
23. Roos-Blom MJ, Gude WT, de Jonge E, Spijkstra JJ, van der Veer SN, Peek N, et al. Impact of audit and feedback with action implementation toolbox on improving ICU pain management: cluster-randomised controlled trial. BMJ quality & safety. 2019;28(12):1007-15.
24. ClinicalTrial.gov. https://clinicaltrials.gov/ct2/show/NCT04055103?term=Arthroplasty&cntry=NL&city=Leiden&draw=2&rank=2.
25. van Schie P, van Bodegom-Vos L, van Steenbergen LN, Nelissen R, Marang-van de Mheen PJ. Monitoring Hospital Performance with Statistical Process Control After Total Hip and Knee Arthroplasty: A Study to Determine How Much Earlier Worsening Performance Can Be Detected. J Bone Joint Surg Am. 2020;102(23):2087-94.
26. Woodall WH, Fogel, S.L., Steiner, S.H.. The monitoring and improvement of surgical-outcome quality. J Qual Technology. 2015(47):383-99.
27. Marang-van de Mheen PJ, van Hanegem N, Kievit J. Effectiveness of routine reporting to identify minor and serious adverse outcomes in surgical patients. Qual Saf Health Care. 2005;14(5):378-82.
28. El Bitar YF, Illingworth KD, Scaife SL, Horberg JV, Saleh KJ. Hospital Length of Stay following Primary Total Knee Arthroplasty: Data from the Nationwide Inpatient Sample Database. J Arthroplasty. 2015;30(10):1710-5.
29. van Steenbergen LN, Denissen GA, Spooren A, van Rooden SM, van Oosterhout FJ, Morrenhof JW, et al. More than 95% completeness of reported procedures in the population-based Dutch Arthroplasty Register. Acta Orthop. 2015;86(4):498-505.
30. Dutch Arthroplasty Register (LROI). Data completeness. https://www.lroi-report.nl/data-quality/completeness/.
31. Austin PC. Absolute risk reductions, relative risks, relative risk reductions, and numbers needed to treat can be obtained from a logistic regression model. J Clin Epidemiol. 2010;63(1):2-6.
32. McAlister FA. The "number needed to treat" turns 20--and continues to be used and misused. CMAJ. 2008;179(6):549-53.
33. Duclos A, Chollet F, Pascal L, Ormando H, Carty MJ, Polazzi S, et al. Effect of monitoring surgical outcomes using control charts to reduce major adverse events in patients: cluster randomised trial. BMJ. 2020;371:m3840.
34. Marshall DA, Christiansen T, Smith C, Squire Howden J, Werle J, Faris P, et al. Continuous quality improvement program for hip and knee replacement. Am J Med Qual. 2015;30(5):425-31.
35. Parkes RJ, Ayeko O, Brunton L, Griffiths-Jones W, Ungvari Z, Goss H. Revolutionising rapid recovery: a quality improvement project in hip and knee replacement. BMJ Open Qual. 2021;10(2).
36. Brehaut JC, Colquhoun HL, Eva KW, Carroll K, Sales A, Michie S, et al. Practice Feedback Interventions: 15 Suggestions for Optimizing Effectiveness. Ann Intern Med. 2016;164(6):435-41.
37. Marang-van de Mheen PJ, Abel, G. A., Shojania, K. G.. Mortality alerts, actions taken and declining mortality: true effect or regression to the mean? BMJ quality & safety. 2018.

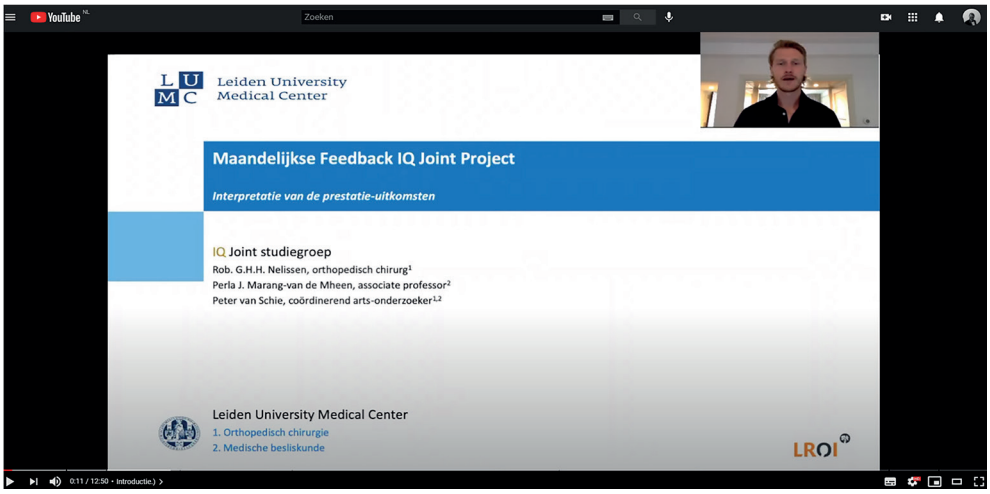
38. Montori VM, Permanyer-Miralda G, Ferreira-González I, Busse JW, Pacheco-Huergo V, Bryant D, et al. Validity of composite end points in clinical trials. BMJ. 2005;330(7491):594-6.
39. Barclay M, Dixon-Woods M, Lyratzopoulos G. The problem with composite indicators. BMJ quality & safety. 2019;28(4):338-44.
40. Glasgow JM, Davies ML, Kaboli PJ. Findings from a national improvement collaborative: are improvements sustained? BMJ quality & safety. 2012;21(8):663-9.
41. Bridges J, May C, Fuller A, Griffiths P, Wigley W, Gould L, et al. Optimising impact and sustainability: a qualitative process evaluation of a complex intervention targeted at compassionate care. BMJ quality & safety. 2017;26(12):970-7.
42. Lipitz-Snyderman A, Needham DM, Colantuoni E, Goeschel CA, Marsteller JA, Thompson DA, et al. The ability of intensive care units to maintain zero central line-associated bloodstream infections. Arch Intern Med. 2011;171(9):856-8.
43. Pronovost PJ, Watson SR, Goeschel CA, Hyzy RC, Berenholtz SM. Sustaining Reductions in Central Line-Associated Bloodstream Infections in Michigan Intensive Care Units: A 10-Year Analysis. Am J Med Qual. 2016;31(3):197-202.
44. Trbovich P, Shojania KG. Root-cause analysis: swatting at mosquitoes versus draining the swamp. BMJ quality & safety. 2017;26(5):350-3.
45. Burke RE, Marang-van de Mheen PJ. Sustaining quality improvement efforts: emerging principles and practice. BMJ quality & safety. 2021.

Supplemental data

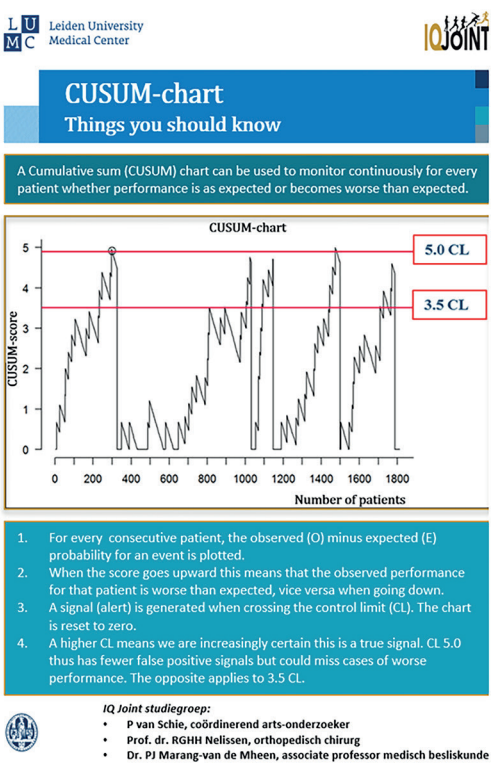
Appendix I

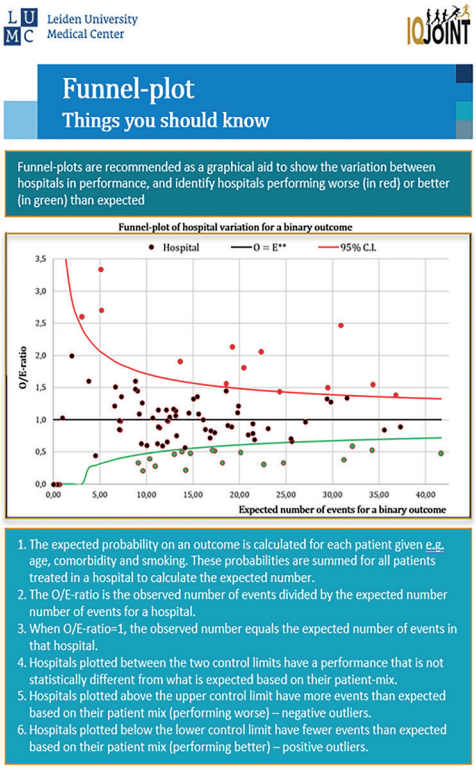


Appendix II



Appendix III





Appendix IV

Leiden University Medical Center

IQ Joint study
Toolbox

Content of Toolbox

In this toolbox we offer starting points for quality improvement initiatives based on the currently available literature. The chapters are ordered according to the performance outcomes as offered in the monthly feedback. It is noted that this is not an exhaustive list.

We advise to implement quality improvement initiatives according to the Plan-Do-Check-Act cycle.

Plan-Do-Check-Act cycle 2

Performance outcomes
- 1-year revision rate due to infection (THA &TKA) 3
- 1-year revision rate due to prosthesis loosening (THA&TKA) 5
- 1-year revision rate due to dislocation (THA) 7
- 1-year revision rate due to technical failure (TKA) 9
- Length-of-stay in hospital 11
- Readmissions 13
- Textbook Outcome & Ordinal Composite Outcome Measure 15

IQ Joint studiegroep:
- P van Schie, coördinerend arts-onderzoeker
- Prof. dr. RGHH Nelissen, orthopedisch chirurg
- Dr. PJ Marang-van de Mheen, assistent professor medisch besliskunde

Leiden University Medical Center

IQJOINT

IQ Joint study
Toolbox

Plan - Do - Check – Act cycle[1]

A iterative four-step management method used for the control and continuous improvement of processes, service and care delivery.

Plan

1. For which outcome an improvement is possible (outlier).
   - *Example: Infection.*
2. Identify possible reasons for the problem (define the process).
   - *Example: Check the electronic patient files to see if the last 100 patient have all received antibiotic prophylaxis. No, 4 patient did not.*
3. Plan Improvement/change.
   - *Example: Ask the anaesthetist for each patient during the "Time Out Procedure" whether the patient had received antibiotic prophylaxis and do this for 6 months.*

↓

Do

1. Pilot → Carry out the improvement/change.
   - *Example: As described under Plan → 3.*

↓

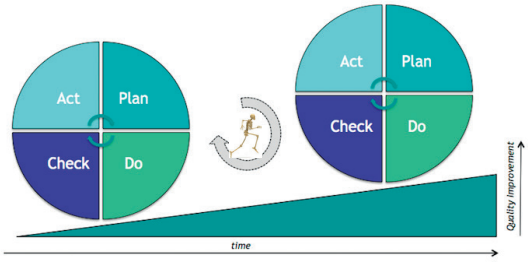
Check

1. What is the effect and is it as desired.
   - *Example: Check the electronic patient files to see if the last 100 patients have all received antibiotic prophylaxis. No, 1 patient did not.*

↓

Act

1. Adopt the improvement/change or abandon it.
   - *Example: Adopt the change because it has led to a substantial improvement.*
2. Run through the cycle again.



Plan-Do-Check-Act cycle - Tague, Nancy R 2005

IQ Joint studiegroep:
- P van Schie, coördinerend arts-onderzoeker
- Prof. dr. RGHH Nelissen, orthopedisch chirurg
- Dr. PJ Marang-van de Mheen, assistent professor medisch besliskunde

Chapter 8

Leiden University Medical Center

IQ JOINT

IQ Joint study
Toolbox

Outcome: Infection (THA & TKA)

Topics are described where quality improvement initiatives could be considered.

Pre-operative

Patient-specific factor optimization
- Poor nutritional status: Aim for Albumine blood levels >34g/L (healthy range: 34-54 g/L).[1,2,25]
- Overweight: Aim for a BMI <30 kg/m². Every BMI-point decrease in obese patients reduces the chance on postoperative infection.[3-6]
- Smoking: Convince patients to participate in smoking cessation programs. Smoking cessation for at least 4 weeks before surgery reduced infections.[7-9]
- Immunocompromising diseases / Immunosuppressive drugs: Choose the most suitable moment to perform the operation. Consult other physicians if needed.
- Glycaemic blood level control. Different glucose target levels were specified.[24]

MRSA screening & decolonisation
- Screening & decolonisation with mupirocin ointment and chlorhexidine show minimal reduction for infections.[10-15,24] Not recommended in NOV-guidelines (NOV guidelines - preoperative decolonisation)

Skin disinfection
- Consider to apply chlorhexidine around the operating area the night before and the morning of surgery.[16,17]

Intra-operative

Antibiotic prophylaxis
- As recommended in NOV-guidelines (NOV guideline - systemic antibiotic prophylaxis).
- Consider vancomycin for MRSA-colonized patients and institutions with high prevalence of MRSA-infections.[18]

Lavage
- Consider 3 minutes lavage with dilute anitsepticum (betadine/chlorhexidine).[19] Avoid lavage with surfactants or antibiotics.[23] Use a low-pressure delivery system for a <2L volume of solution.[23]

Prevent transfusions
- Check pre-operative hemoglobin level and correct if necessary preoperatively.
- Tranexamic acid might help minimize blood loss and wound infection.[20,21]

Cement loaded with antibiotics
- As recommended in NOV-guidelines (NOV guidelines - Antibiotica-laden cement).

Surgical approach
- Lateral surgical approach results in more infections compared to posterior approach.[6] However, aach of the approaches has their own set of complications and benefits.

Bearing surface
- Ceramic-on-ceramic and ceramic-on-polyethylene surfaces are associated with lower risk of revisions for infection after 12 and 24 months respectively compared to metal-on-polyethylene.[6]

Post-operative

Antibiotic prophylaxis
- As recommended in NOV-guidelines (NOV guideline - systemic antibiotic prophylaxis).

Wound leakage
- Is a wound leakage protocol available and is it followed sufficient?

Patient-specific factor optimization
- Blood glucose levels: Fasting blood glucose value <200mg/dl is suggested.[22]

IQ Joint studiegroep:
•P van Schie, coördinerend arts-onderzoeker
•Prof. dr. RGHH Nelissen, orthopedisch chirurg
•Dr. PJ Marang-van de Mheen, assistent professor medisch besliskunde

Leiden University Medical Center

IQJOINT

IQ Joint study

References

1. Mainous, M.R. and E.A. Deitch, Nutrition and infection. Surg Clin North Am, 1994. 74(3): p. 659-76.
2. Devoto, G., et al., Prealbumin serum concentrations as a useful tool in the assessment of malnutrition in hospitalized patients. Clin Chem, 2006. 52(12): p. 2281-5.
3. Dowsey, M.M. and P.F. Choong, Obese diabetic patients are at substantial risk for deep infection after primary TKA. Clin Orthop Relat Res, 2009. 467(6): p. 1577-81.
4. Dowsey, M.M. and P.F. Choong, Early outcomes and complications following joint arthroplasty in obese patients: a review of the published reports. ANZ J Surg, 2008. 78(6): p. 439-44.
5. D'Apuzzo, M.R., W.M. Novicoff, and J.A. Browne, The John Insall Award: Morbid obesity independently impacts complications, mortality, and resource use after TKA. Clin Orthop Relat Res, 2015. 473(1): p. 57-63.
6. Lenguerrand, E., et al., Risk factors associated with revision for prosthetic joint infection after hip replacement: a prospective observational cohort study. Lancet Infect Dis, 2018. 18(9): p. 1004-1014.
7. Kapadia, B.H., et al., Tobacco use may be associated with increased revision and complication rates following total hip arthroplasty. J Arthroplasty, 2014. 29(4): p. 777-80.
8. Kapadia, B.H., et al., Increased revision rates after total knee arthroplasty in patients who smoke. J Arthroplasty, 2012. 27(9): p. 1690-1695.e1.
9. Sorensen, L.T., Wound healing and infection in surgery. The clinical impact of smoking and smoking cessation: a systematic review and meta-analysis. Arch Surg, 2012. 147(4): p. 373-83.
10. Kapadia, B.H., et al., Pre-admission cutaneous chlorhexidine preparation reduces surgical site infections in total hip arthroplasty. J Arthroplasty, 2013. 28(3): p. 490-3.
11. Baratz, M.D., et al., Twenty Percent of Patients May Remain Colonized With Methicillin-resistant Staphylococcus aureus Despite a Decolonization Protocol in Patients Undergoing Elective Total Joint Arthroplasty. Clin Orthop Relat Res, 2015. 473(7): p. 2283-2290.
12. Rao, N., et al., Preoperative screening/decolonization for Staphylococcus aureus to prevent orthopedic surgical site infection: prospective cohort study with 2-year follow-up. J Arthroplasty, 2011. 26(8): p. 1501-7.
13. Sporer, S.M., T. Rogers, and L. Abella, Methicillin-Resistant and Methicillin-Sensitive Staphylococcus aureus Screening and Decolonization to Reduce Surgical Site Infection in Elective Total Joint Arthroplasty. J Arthroplasty, 2016. 31(9 Suppl): p. 144-7.
14. Schweizer, M.L., et al., Association of a bundled intervention with surgical site infections among patients undergoing cardiac, hip, or knee surgery. Jama, 2015. 313(21): p. 2162-71.
15. Stambough, J.B., et al., Corrigendum to "Decreased Hospital Costs and Surgical Site Infection Incidence With a Universal Decolonization Protocol in Primary Total Joint Arthroplasty" [Journal of Arthroplasty 32 (2017) 728-734]. J Arthroplasty, 2017. 32(8): p. 2625.
16. Kapadia, B.H., et al., Pre-admission cutaneous chlorhexidine preparation reduces surgical site infections in total hip arthroplasty. J Arthroplasty, 2013. 28(3): p. 490-3.
17. Johnson, A.J., et al., Chlorhexidine reduces infections in knee arthroplasty. J Knee Surg, 2013. 26(3): p. 213-8.
18. Bratzler, D.W., et al., Clinical practice guidelines for antimicrobial prophylaxis in surgery. Surg Infect (Larchmt), 2013. 14(1): p. 73-156.
19. Brown, N.M., et al., Dilute betadine lavage before closure for the prevention of acute postoperative deep periprosthetic joint infection. J Arthroplasty, 2012. 27(1): p. 27-30.
20. Sukeik, M., et al., Systematic review and meta-analysis of the use of tranexamic acid in total hip replacement. J Bone Joint Surg Br, 2011. 93(1): p. 39-46.
21. Sukeik, M., et al., The effect of tranexamic acid on wound complications in primary total Hip Arthroplasty: A meta-analysis. Surgeon, 2019.
22. Mraovic, B., et al., Perioperative hyperglycemia and postoperative infection after lower limb arthroplasty. J Diabetes Sci Technol, 2011. 5(2): p. 412-8.
23. Kavolus, J., et al., Irrigation fluids used for the prevention and treatment of orthopaedic infections. J Bone Joint Surg Am, 2020;102:76-84.
24. Fields, A.C., J.C. Pradarelli, and K.M.F. Itani, Preventing Surgical Site Infections: Looking Beyond the Current Guidelines. Jama, 2020.
25. Gu A, et al., Preoperative malnutrition negatively correlates with postoperative wound complications and infection after total join arthroplasty: a systematic review and meta-analysis. J Arthroplasty. 2019 May;34(5):1013-24.

IQ Joint studiegroep:

•P van Schie, coördinerend arts-onderzoeker
•Prof. dr. RGHH Nelissen, orthopedisch chirurg
•Dr. PJ Marang-van de Mheen, assistent professor medisch besliskunde

Leiden University Medical Center

IQ JOINT

IQ Joint study
Toolbox

Outcome: Prosthesis loosening (THA & TKA)

Topics are described where quality improvement initiatives could be considered.

Aseptic loosening

Patient risk factors

Patient-specific factor optimization
- Smoking: Convince patients to participate in smoking cessation programs. Smoking significantly increases risk of aseptic loosening.[1]
- Age: Postpone the operation if possible. Lower age has a higher chance on aseptic loosening in the future.[2]

Prosthesis factors

THA
- Advise low-impact activities such as walking, swimming and cycling. Patient undertaking intermediate to intense activity are four times more likely than less active people to develop acetabular prosthesis loosening due to more wear.[3-6]
- Use cross-linked polyethylene liners instead of conventional liners to reduce wear and revisions.[7-9]

TKA
- Roentgen stereophotogrammetric analyses (RSA)-studies showed favorable outcomes on prosthesis loosening in the first two year for cemented implants, but unfavorable outcomes after two years when compared to uncemented implants.[10]

THA &TKA
- Has a new prothesis been implemented recently? Has sufficient training taken place? Schedule a meeting where experiences can be shared.

Surgical factors

Cementation techniques
- Distal and proximal prosthesis centralization
- Adequate canal preparation with pulsatile lavage to increase cement penetration and interdigitation.
- Is there profit to be gained within one of the phases: mixing, waiting, working or setting?
- Check the most recent manual for use of the cement.

Septic loosening

Take a look at the toolbox for infection

IQ Joint studiegroep:
•P van Schie, coördinerend arts-onderzoeker
•Prof. dr. RGHH Nelissen, orthopedisch chirurg
•Dr. PJ Marang-van de Mheen, assistent professor medisch besliskunde

Leiden University Medical Center

IQ JOINT

IQ Joint study

References

1. Teng, S., et al., Smoking and risk of prosthesis-related complications after total hip arthroplasty: a meta-analysis of cohort studies. PLoS One, 2015. 10(4): p. e0125294.
2. Bordini, B., et al., Factors affecting aseptic loosening of 4750 total hip arthroplasties: multivariate survival analysis. BMC Musculoskelet Disord, 2007. 8: p. 69.
3. Flugsrud, G.B., et al., The effect of middle-age body weight and physical activity on the risk of early revision hip arthroplasty: a cohort study of 1,535 individuals. Acta Orthop, 2007. 78(1): p. 99-107.
4. Lubbeke, A., et al., Influence of patient activity on femoral osteolysis at five and ten years following hybrid total hip replacement. J Bone Joint Surg Br, 2011. 93(4): p. 456-63.
5. Schmalzried, T.P., et al., The John Charnley Award. Wear is a function of use, not time. Clin Orthop Relat Res, 2000(381): p. 36-46.
6. Emms, N.W., et al., Long-term outcome of a cementless, hemispherical, press-fit acetabular component: survivorship analysis and dose-response relationship to linear polyethylene wear. J Bone Joint Surg Br, 2010. 92(6): p. 856-61.
7. Australian Orthopaedic Association National Joint Registry, Annual report. 2017
8. Paxton, E.W., et al., Metal-on-conventional polyethylene total hip arthroplasty bearing surfaces have a higher risk of revision than metal-on-highly crosslinked polyethylene: results from a US registry. Clin Orthop Relat Res, 2015. 473(3): p. 1011-21.
9. Hanna, S.A., et al., Highly cross-linked polyethylene decreases the rate of revision of total hip arthroplasty compared with conventional polyethylene at 13 years' follow-up. Bone Joint J, 2016. 98-b(1): p. 28-32.
10. Nakama, G.Y., et al., Cemented, cementless or hybrid fixation options in total knee arthroplasty for osteoarthritis and other non-traumatic diseases. Cochrane Database Syst Rev, 2012. 10: p. CD006193.

IQ Joint studiegroep:
•P van Schie, coördinerend arts-onderzoeker
•Prof. dr. RGHH Nelissen, orthopedisch chirurg
•Dr. PJ Marang-van de Mheen, assistent professor medisch besliskunde

Leiden University Medical Center

IQ JOINT

IQ Joint study
Toolbox

Outcome: Dislocation (THA)

Topics are described where quality improvement initiatives could be considered.

Pre-operative

Patient-specific factor optimization

- Overweight: Aim for a BMI<30 kg/m^2. Dislocation after THA occurs more often in obese patients.[1-4]

Intra-operative

Femoral head size

- In most patients a femoral head size of 32 mm is recommended following the NOV-guidelines (NOV guideline - femoral head diameter).
- Larger femoral head size decreases dislocation rates due to greater jumping distance and a greater range of motion.[1,5-7] However, heads above 32-mm lead to more friction and more wear with prosthesis loosening as a possible consequence.[8,9]
- If a head larger than 32 mm is indicated, it seems best to use a ceramic-on-ceramic prosthesis because this combination shows lowest wear (NOV guideline - femoral head diameter).[8]
- For the posterolateral approach 36-mm head can safely further reduce the risk of revision for dislocation (without an increased number of revisions for all other reasons within 6 years).[10]

Surgical approach

- Each of the approaches has their own set of complications and benefits. Both the posterior, lateral and anterior approach can be used (NOV guideline - surgical approach).
- Some registries report increased dislocation rates for posterior approach when compared to anterior and direct lateral at 6-year follow-up.[10,11] However, the revision rate for all other revisions was higher with anterior approach and lowest with posterior approach.[10]
- If the posterior approach is chosen, surgeons should reconstruct the posterior capsule and the external rotators to prevent dislocations (NOV guideline - surgical approach).

Dual mobility cup

- Dual mobility articulations are a viable alternative in cases with increased risk of instability or dislocation, however, evidence is limited (NOV guideline - dual mobility cup). Following patient groups have an increased risk of dislocations and may benefit from a dual mobility cup: spinal injury, poliomyelitis, cerebral palsy, femoral neck fracture, acetabular dysplasia, muscular dystrophy and intellectual impairment.[12-16]
- The 5-year cup revision rates are comparable to that of traditional unipolar cups.[17]

Stability Assessment

- Minimize impingement by removing osteophytes, thickened capsule or increase offset.
- A lipped liner can offer stability in extremes of movement.[18]

New prosthesis

- Has sufficient training taken place?
- Schedule a meeting where experiences can be shared.

Post-operative

Hip dislocation precaution

- Early dislocation rates do not decrease with hip dislocation precaution.[16,19] Evidence is limited and included only studies with anterolateral and posterolateral approaches. Further, abandoning mobilization restrictions increases patient satisfaction through earlier return of daily activities to preoperative levels.[20,21]

IQ Joint studiegroep:
•P van Schie, coördinerend arts-onderzoeker
•Prof. dr. RGHH Nelissen, orthopedisch chirurg
•Dr. PJ Marang-van de Mheen, assistent professor medisch besliskunde

Leiden University Medical Center

IQ Joint study

References

1. Howie, D.W., O.T. Holubowycz, and R. Middleton, Large femoral heads decrease the incidence of dislocation after total hip arthroplasty: a randomized controlled trial. J Bone Joint Surg Am, 2012. 94(12): p. 1095-102.
2. Haverkamp, D., et al., Obesity in total hip arthroplasty--does it really matter? A meta-analysis. Acta Orthop, 2011. 82(4): p. 417-22.
3. Wagner, E.R., et al., Effect of Body Mass Index on Complications and Reoperations After Total Hip Arthroplasty. J Bone Joint Surg Am, 2016. 98(3): p. 169-79.
4. Li, W., et al., Functional Gain and Pain Relief After Total Joint Replacement According to Obesity Status. J Bone Joint Surg Am, 2017. 99(14): p. 1183-1189.
5. Stroh, D.A., et al., Reduced dislocation rates and excellent functional outcomes with large-diameter femoral heads. J Arthroplasty, 2013. 28(8): p. 1415-20.
6. Sariali, E., et al., Mathematical evaluation of jumping distance in total hip arthroplasty: influence of abduction angle, femoral head offset, and head diameter. Acta Orthop, 2009. 80(3): p. 277-82.
7. Kostensalo, I., et al., Effect of femoral head size on risk of revision for dislocation after total hip arthroplasty: a population-based analysis of 42,379 primary procedures from the Finnish Arthroplasty Register. Acta Orthop, 2013. 84(4): p. 342-7.
8. Tsikandylakis, G., et al., Head size in primary total hip arthroplasty. EFORT Open Rev, 2018. 3(5): p. 225-231.
9. Charnley, J., A. Kamangar, and M.D. Longfield, The optimum size of prosthetic heads in relation to the wear of plastic sockets in total replacement of the hip. Med Biol Eng, 1969. 7(1): p. 31-9.
10. Zijlstra, W.P., et al., Effect of femoral head size and surgical approach on risk of revision for dislocation after total hip arthroplasty. Acta Orthop, 2017. 88(4): p. 395-401.
11. Sheth, D., et al., Anterior and Anterolateral Approaches for THA Are Associated With Lower Dislocation Risk Without Higher Revision Risk. Clin Orthop Relat Res, 2015. 473(11): p. 3401-3408.
12. Hernigou, P., et al., Dual-mobility or Constrained Liners Are More Effective Than Preoperative Bariatric Surgery in Prevention of THA Dislocation. Clin Orthop Relat Res, 2016. 474(10): p. 2202-10.
13. Queally, J.M., A. Abdulkarim, and K.J. Mulhall, Total hip replacement in patients with neurological conditions. J Bone Joint Surg Br, 2009. 91(10): p. 1267-73.
14. Houdek, M.T., et al., Total Hip Arthroplasty in Patients with Cerebral Palsy: A Cohort Study Matched to Patients with Osteoarthritis. J Bone Joint Surg Am, 2017. 99(6): p. 488-493.
15. Sanders, R.J., B.A. Swierstra, and J.H. Goosen, The use of a dual-mobility concept in total hip arthroplasty patients with spastic disorders: no dislocations in a series of ten cases at midterm follow-up. Arch Orthop Trauma Surg, 2013. 133(7): p. 1011-6.
16. Barnsley, L., L. Barnsley, and R. Page, Are Hip Precautions Necessary Post Total Hip Arthroplasty? A Systematic Review. Geriatr Orthop Surg Rehabil, 2015. 6(3): p. 230-5.
17. Bloemheuvel, E.M., L.N. van Steenbergen, and B.A. Swierstra, Dual mobility cups in primary total hip arthroplasties: trend over time in use, patient characteristics, and mid-term revision in 3,038 cases in the Dutch Arthroplasty Register (2007-2016). Acta Orthop, 2019. 90(1): p. 11-14.
18. Sculco, P.K., et al., Avoiding Hip Instability and Limb Length Discrepancy After Total Hip Arthroplasty. Orthop Clin North Am, 2016. 47(2): p. 327-34.
19. Gromov, K., et al., Removal of restrictions following primary THA with posterolateral approach does not increase the risk of early dislocation. Acta Orthop, 2015. 86(4): p. 463-8.
20. Ververeli, P.A., et al., Evaluation of reducing postoperative hip precautions in total hip replacement: a randomized prospective study. Orthopedics, 2009. 32(12): p. 889.
21. Peak, E.L., et al., The role of patient restrictions in reducing the prevalence of early dislocation following total hip arthroplasty. A randomized, prospective study. J Bone Joint Surg Am, 2005. 87(2): p. 247-53.

IQ Joint studiegroep:
•P van Schie, coördinerend arts-onderzoeker
•Prof. dr. RGHH Nelissen, orthopedisch chirurg
•Dr. PJ Marang-van de Mheen, assistent professor medisch besliskunde

Leiden University Medical Center

IQJOINT

IQ Joint study
Toolbox

Outcome: Technical failure (TKA)

Topics are described where quality improvement initiatives could be considered.

Malalignment

Coronal plane (varus/valgus)

- Former studies suggest that mechanical malalignment with an angle >3° results in higher revision rates.[1,2] However, more recent studies suggest anatomical alignment is more important and showed no association between mechanical malalignment (>3°) and revision rates.[3-7] Optimal alignment seems an anatomical tibiofemoral angle of 7 to 9° valgus with the mechanical axis through the medial third of the knee.[8-10]

Sagittal and axial plane

- Sagittal and axial malalignment is associated with increases revision rates.[9]

Rotational alignment

- A positive correlation was found between external rotation of the tibial and the femoral component and the Knee Society Score.[19]

New prosthesis

- Has sufficient training taken place?
- Schedule a meeting where experiences can be shared.

Instability

Etiology

- Early (within one year) postoperative instability may be required for various reasons, including malalignment of components, implant loosening, improper balance of the flexion-extension space, rupture or laxity of the posterior cruciate ligament or medial collateral ligament and patellar tendon rupture or patella fracture.[11-16]
- Some patient are prone to instability. Those who have rheumatoid arthritis, connective tissue disease, severe osteoporosis, neuromuscular pathology, gross deformities who need severe correction with ligament release, foot deformities and quadriceps/medial thrust hip abductor weakness.[15,17]

Patient-specific factor optimization

- Overweight: Aim for a BMI<30 kg/m^2. Obesity is a risk factor because it complicates surgical exposure, jeopardizes the collateral ligament.[15]

Pre-operative / intra-operative

- Evaluation the state of the lateral and medial collateral ligament and posterior cruciate ligament (PCL) with physical examination in order to select the right implant for each patient.[16] Instability can be prevented in most cases with appropriate prosthesis selection and good surgical technique (e.g. prevent soft tissue damage, correct implantation of components in every plane).[12,14] Posterior stabilized implants should be utilized in those patients with PCL insufficiency and in those with increase risk of posterior instability (e.g. rheumatoid arthritis, need to resect the PCL, flexion contracture or previous tibial osteotomy). If the choice is made to preserve the PCL, it is important to take special care in maintaining its integrity when the tibial cut is made. In case of doubt, it is preferable to convert the arthroplasty to a posterior stabilized design. In some patients with marked instability (medial or lateral collateral loss, massive bone loss including the femoral condyles, complete or insufficiency of the PCL, poliomyelitis , or Charcot arthropathy), a primary constrained or linked hinge implant may me indicated.[16]

Patellar dislocation

Intra-operative

- Patella maltracking or dislocation is closely related to malalignment. In most patients, functional patellar tracking is achieved by a good prosthesis positioning by checking the femoral implant rotation, femoral implant flexion, femoral implans varus/valgus positioning, femoral implant mediolateral or medialization, tibial implant rotation. Excessive internal rotation of the tibial component or femur component promotes external rotation during walking, thereby increasing the risk of patellar dislocation. The more externally rotated the implant, the less risk there is for lateral patellar maltracking. However, , this must not be at the expense of tibiofemoral alignment and stability.[11] Postoperative patella alta and non-medialized implantation of a patellar prosthesis are also risk factor.[18]

Leiden University Medical Center

IQ JOINT

IQ Joint study

References

1. Jeffery, R.S., R.W. Morris, and R.A. Denham, Coronal alignment after total knee replacement. J Bone Joint Surg Br, 1991. 73(5): p. 709-14.
2. Kim, Y.H., et al., Computer-assisted surgical navigation does not improve the alignment and orientation of the components in total knee arthroplasty. J Bone Joint Surg Am, 2009. 91(1): p. 14-9.
3. Hadi, M., et al., Does malalignment affect revision rate in total knee replacements: a systematic review of the literature. Springerplus, 2015. 4: p. 835.
4. Parratte, S., et al., Effect of postoperative mechanical axis alignment on the fifteen-year survival of modern, cemented total knee replacements. J Bone Joint Surg Am, 2010. 92(12): p. 2143-9.
5. Bonner, T.J., et al., The effect of post-operative mechanical axis alignment on the survival of primary total knee replacements after a follow-up of 15 years. J Bone Joint Surg Br, 2011. 93(9): p. 1217-22.
6. Magnussen, R.A., et al., Residual varus alignment does not compromise results of TKAs in patients with preoperative varus. Clin Orthop Relat Res, 2011. 469(12): p. 3443-50.
7. Matziolis, G., J. Adam, and C. Perka, Varus malalignment has no influence on clinical outcome in midterm follow-up after total knee replacement. Arch Orthop Trauma Surg, 2010. 130(12): p. 1487-91.
8. Ritter, M.A., et al., The effect of alignment and BMI on failure of total knee replacement. J Bone Joint Surg Am, 2011. 93(17): p. 1588-96.
9. Kim, Y.H., et al., The relationship between the survival of total knee arthroplasty and postoperative coronal, sagittal and rotational alignment of knee prosthesis. Int Orthop, 2014. 38(2): p. 379-85.
10. Fang, D.M., M.A. Ritter, and K.E. Davis, Coronal alignment in total knee arthroplasty: just how important is it? J Arthroplasty, 2009. 24(6 Suppl): p. 39-43.
11. Bozic, K.J., et al., The epidemiology of revision total knee arthroplasty in the United States. Clin Orthop Relat Res, 2010. 468(1): p. 45-51.
12. Sharkey, P.F., et al., Insall Award paper. Why are total knee arthroplasties failing today? Clin Orthop Relat Res, 2002(404): p. 7-13.
13. Gioe, T.J., et al., Why are total knee replacements revised?: analysis of early revision in a community knee implant registry. Clin Orthop Relat Res, 2004(428): p. 100-6.
14. Chang, M.J., et al., Diagnosis, causes and treatments of instability following total knee arthroplasty. Knee Surg Relat Res, 2014. 26(2): p. 61-7.
15. Vince, K.G., A. Abdeen, and T. Sugimori, The unstable total knee arthroplasty: causes and cures. J Arthroplasty, 2006. 21(4 Suppl 1): p. 44-9.
16. Rodriguez-Merchan, E.C., Instability following total knee arthroplasty. Hss j, 2011. 7(3): p. 273-8.
17. Meding, J.B., et al., Genu recurvatum in total knee replacement. Clin Orthop Relat Res, 2003(416): p. 64-7.
18. Berger, Crossett, Jacobs, Rubash. 1998.
19. Valkering K.P., et al. Effect of rotational alignment on outcome of total knee arthroplasty; A systematic review of the literature and correlation analysis. Acta Orthopaedica, 86:4, 432-439.

IQ Joint studiegroep:

•P van Schie, coördinerend arts-onderzoeker
•Prof. dr. RGHH Nelissen, orthopedisch chirurg
•Dr. PJ Marang-van de Mheen, assistent professor medisch besliskunde

Leiden University Medical Center

IQ JOINT

IQ Joint study
Toolbox

Length of hospital stay

Below, topics are described where quality improvement initiatives can be considered.

Pre-operative

Fast-track program

- Implement a fast track surgery program in THA surgery is associated with a reduction in post-operative length of stay (LOS), shorter convalescence and rapid functional recovery, without increased morbidity and mortality.[1,2] (NOV guidelines - fast-track program)
- Mention the expected date of discharge prior to operation so that patients know what is to be expected.

Patient-specific factor optimization

- Optimize glucose levels in patient with type I diabetes, since it gives an elevated risk of complications after THA / TKA surgery, thus a prolonged LOS.[3]
- Pre-operative iron-deficiency anaemia is associated with increased risk of LOS>5 days after adjustment for pre-operative patient-related risk factors. It should be detected in pre-operative evaluation and treated before surgery to ensure maximum rehabilitation.[4]
- Malnutrition; levels of albumin, total lymphocytes and transferrin should be monitored and be restored if not within the normal ranges (34-54 g/L, 3900-10000 cells/µL and 170-370 mg/dl respectively), because malnourished patients are at higher risk for surgical complications and thus a higher probability of prolonged LOS.[5]
- Pre-operative opioid use is a risk factor for post-operative pain at rest and during walking which impairs fast-track recovery among TKA patients and leads to increased opioid consumption post-operative. It should be detected in pre-operative evaluation and the patient need to be persuaded to keep opioid use to a minimum.[6,7]
- Social support; inadequate social support e.g. living alone, is associated with a longer LOS. Optimizing the organizational part of patient pathway and optimizing social support before admission for surgery avoids delayed discharge.[8,9]

Intra-operative

Medication

- A protocol of scheduled oral narcotics, cyclooxygenase-2 inhibitors, a local anesthetic for wound infiltration and no intrathecal narcotics (TKA: add femoral nerve catheter) shows significant improvements regarding LOS and post-operative pain-scores.[10] However, another study showed only a significant improvement in pain-scores and opioid requirements, but showed an effect on LOS although not significant.[11]
- A single dose of 125 mg methylprednisolone given pre-operatively, reduces pain in THA patient in the first post-operative 24 hours, thus enlarging the chance of satisfactory day-of-surgery mobilization and early discharge.[12,13]

Surgical technique

- Direct anterior approach (DAA) shows an advantage regarding mean hospital stay compared with posterolateral (PL) approach in THA surgery.[14]

Post-operative

Delay of discharge

- Medical interventions; delay of discharge due to e.g. waiting for blood transfusion, start of physiotherapy or post-operative radiographic examination, should be avoided through multidisciplinary organization and planning.[15]
- Repeating / mentioning of the expected date and time of delay when there are no complications.

Oral pain treatment

- Oral treatment should be a combination of a NSAID, paracetamol and short acting-opioid.

Mobilization on day of surgery

- Mobilization on the day of surgery significantly increases the probability of early discharge.[16]

IQ Joint studiegroep:
•P van Schie, coördinerend arts-onderzoeker
•Prof. dr. RGHH Nelissen, orthopedisch chirurg
•Dr. PJ Marang-van de Mheen, assistent professor medisch besliskunde

Leiden University Medical Center

IQ JOINT

IQ Joint studie

Referenties

1. Ansari, D., et al. (2013). "Fast-track surgery: procedure-specific aspects and future direction." Langenbecks Arch Surg 398(1): 29-37.
2. Hansen, TB. Fast track in hip arthroplasty. 2017
3. Viens, N. A., et al. (2012). "Role of diabetes type in perioperative outcomes after hip and knee arthroplasty in the United States." J Surg Orthop Adv 21(4): 253-260.
4. Jans, O., et al. (2014). "Role of preoperative anemia for risk of transfusion and postoperative morbidity in fast-track hip and knee arthroplasty." Transfusion 54(3): 717-726.
5. Nicholson, J. A., et al. (2012). "Nutritional status and short-term outcome of hip arthroplasty." J Orthop Surg (Hong Kong) 20(3): 331-335.
6. Aasvang, E. K., et al. (2016). "Chronic pre-operative opioid use and acute pain after fast-track total knee arthroplasty." Acta Anaesthesiol Scand 60(4): 529-536.
7. Goesling, J., et al. (2016). "Trends and predictors of opioid use after total knee and total hip arthroplasty." Pain 157(6): 1259-1265.
8. Napier, R. J., et al. (2013). "Modifiable factors delaying early discharge following primary joint arthroplasty." Eur J Orthop Surg Traumatol 23(6): 665-669.
9. den Hartog, Y. M., et al. (2015). "Which patient characteristics influence length of hospital stay after primary total hip arthroplasty in a 'fast-track' setting?" Bone Joint J 97-B(1): 19-23.
10. Peters, C. L., et al. (2006). "The effect of a new multimodal perioperative anesthetic regimen on postoperative pain, side effects, rehabilitation, and length of hospital stay after total joint arthroplasty." J Arthroplasty 21(6 Suppl 2): 132-138.
11. Andersen, L. O. and H. Kehlet (2014). "Analgesic efficacy of local infiltration analgesia in hip and knee arthroplasty: a systematic review." Br J Anaesth 113(3): 360-374.
12. Husted, H., et al. (2010). "Care principles at four fast-track arthroplasty departments in Denmark." Dan Med Bull 57(7): A4166.
13. Lunn, T. H., et al. (2013). "Effect of high-dose preoperative methylprednisolone on recovery after total hip arthroplasty: a randomized, double-blind, placebo-controlled trial." Br J Anaesth 110(1): 66-73.
14. Godoy-Monzon, D., et al. (2019). "Comparative study of radiological and functional outcomes following a direct anterior approach versus to a posterolateral approach to the hip." Rev Esp Cir Ortop Traumatol 63(5): 370-375.
15. Husted, H., et al. (2011). "Why still in hospital after fast-track hip and knee arthroplasty?" Acta Orthop 82(6): 679-684.
16. Okamoto, T., et al. (2016). "Day-of-Surgery Mobilization Reduces the Length of Stay After Elective Hip Arthroplasty." J Arthroplasty 31(10): 2227-2230.

IQ Joint studiegroep:

•P van Schie, coördinerend arts-onderzoeker

•Prof. dr. RGHH Nelissen, orthopedisch chirurg

•Dr. PJ Marang-van de Mheen, assistent professor medisch besliskunde

Leiden University Medical Center

IQ JOINT

IQ Joint study
Toolbox

Outcome: Number of hospital readmission within 30 days

Topics are described where quality improvement initiatives could be considered.

General information

Readmissions

- Primary diagnoses at readmission that were identified to be directly attributable to surgery comprised 38% readmissions at 0-30 days, 24% at 31-60 days and 16% at 60-90 days. Proportion attributable to surgery decreases significantly over the 90-day period after index surgery.[1]
- The most frequent readmission diagnosis after TKA is surgical site infection.[2]
- An increased length of stay, discharge disposition, blood transfusion and general anaesthesia are associated with readmission.[3]

Length of hospital stay (LOS) and readmissions

- Decreasing the LOS does not increase the risk of readmissions.[4-7] No difference in 90-day-readmission odds between patients with a 1-midnight LOS and those with a 2-midnight LOS for primary TKA was identified.[8,9]

Patient-specific factor optimization

- Smoking: Convince patients to participate in smoking cessation programs. Smoking increases the risk of 90-days readmission.[8,10]

What kind of readmissions are involved

An example of a plan of approach for file investigation:

1. Take a closer look at 50 file. Exclude patients with relatively high mortality risk. These are the patients with probably a high disease burden and therefore relatively little chance of finding points for improvement.
2. How soon after discharge did the readmission take place (within a week or later)? Selection for early readmissions gives maximum chance to find improvement regarding potential too early discharge or incorrect information transfer. Selection for late readmissions often indicates complications after discharge.
3. Make a distinction between re-admissions in the same diagnosis group as the index admission versus in another diagnosis group. If the re-admission diagnosis group is the same as the index admission, this may be an indication that the patient was discharged too soon. If the re-admission concerns a different diagnosis group than the index admission, then there may be a re-admission with a complication. Of course it is possible that there is no relationship with the surgical procedure earlier.

IQ Joint studiegroep:
•P van Schie, coördinerend arts-onderzoeker
•Prof. dr. RGHH Nelissen, orthopedisch chirurg
•Dr. PJ Marang-van de Mheen, assistent professor medisch besliskunde

Leiden University Medical Center

IQ JOINT

IQ Joint study

Referenties

1. Braud, J., Ramanathan, D., Klika, A., Navale, S., Higuera, C., Molloy, R. Temporal Relations of Unplanned Readmissions Following Total Knee Arthroplasty: A Study of Large State Inpatient Databases. J Arthroplasty. 2017; 32(9S): p. S119-S123.
2. WHO Regional Office for Europe. Do current discharge arrangements from inpatient hospital care for the elderly reduce readmission rates, the length of inpatient stay or mortality, or improve health status?. Copenhagen, Denmar: WHO Regional Office for Europe; 2005.
3. Mesko, N.W., Bachmann, K.R., Kovacevic, D., LoGrasso, M.E., O'Rourke, C., Froimson, M.I. Thirty-day readmission following total hip and knee arthroplasty - a preliminary single institution predictive model. J Arthroplasty. 2014; 29(8): p. 1532-8.
4. Sutton, J.C., 3rd, Antoniou, J., Epure, L.M., Huk, O.L., Zukor, D.J., Bergeron, S.G. Hospital Discharge within 2 Days Following Total Hip or Knee Arthroplasty Does Not Increase Major-Complication and Readmission Rates. J Bone Joint Surg Am. 2016; 98(17): p. 1419-28.
5. Bini, S.A., Inacio, M.C., Cafri, G. Two-Day Length of Stay is Not Inferior to 3 Days in Total Knee Arthroplasty with Regards to 30-Day Readmissions. J Arthroplasty. 2015; 30(5): p. 733-8.
6. Berg, U., BuLow, E., Sundberg, M., Rolfson, O. No increase in readmissions or adverse events after implementation of fast-track program in total hip and knee replacement at 8 Swedish hospitals: An observational before-and-after study of 14,148 total joint replacements 2011-2015. Acta Orthopaedica. 2018; 89(5): p. 522-527.
7. Zhu, S., Qian, W., Jiang, C., Ye, C., Chen, X. Enhanced recovery after surgery for hip and knee arthroplasty: a systematic review and meta-analysis. Postgrad Med J. 2017; 93(1106): p. 736-742.
8. Charpentier, P.M., Srivastava, A.K., Zheng, H., Ostrander, J.D., Hughes, R.E. Readmission Rates for One Versus Two-Midnight Length of Stay for Primary Total Knee Arthroplasty: Analysis of the Michigan Arthroplasty Registry Collaborative Quality Initiative (MARCQI) Database. J Bone Joint Surg Am. 2018; 100(20): p. 1757-1764.
9. Sibia, U.S., Waite, K.A., Callanan, M.A., Park, A.E., King, P.J., MacDonald, J.H. Do shorter lengths of stay increase readmissions after total joint replacements? Arthroplast Today. 2017; 3(1): p. 51-55.
10. Sahota, S., Lovecchio, F., Harold, R.E., Beal, M.D., Manning, D.W. The Effect of Smoking on Thirty-Day Postoperative Complications After Total Joint Arthroplasty: A Propensity Score-Matched Analysis. J Arthroplasty. 2018; 33(1): p. 30-35.

IQ Joint studiegroep:

•P van Schie, coördinerend arts-onderzoeker
•Prof. dr. RGHH Nelissen, orthopedisch chirurg
•Dr. PJ Marang-van de Mheen, assistent professor medisch besliskunde

Leiden University Medical Center

IQ JOINT

IQ Joint study
Toolbox

Outcome: Textbook Outcome & Ordinal Composite Outcome Measure

Topics are described where quality improvement initiatives could be considered.

Separately looking at the 3 indicators revision within 1 year, readmission within 30 days and prolonged length-of-stay (a length of stay in the upper tertile) have disadvantages. When conducting file investigations, there is a chance that the same file will be requested 3 times and that many must be investigated to find opportunities for quality improvement. Furthermore, single outcomes do not provide insight for professionals and patients into which part of the patients everything went well. For the above 3 indicators, a TO would mean that a patient did not undergo a revision within 1 year, had no readmission within 30 days and had a normal length-of-stay.

However, hospitals with a TO that differs significantly form the average gives little information about which outcome was specifically worse/good. Therefore the Ordinal Composite Outcome measure (Textbook Outcome Plus; TOP) has been developed. This is an extension of the TO with the additional element that the different combinations of the results are arranged (instead of all in 1 non-TO group), so that it could be seen in which group the hospital deviates from the average. The order is from the best to the worst outcome as follows:

- No revision within one year, no readmission within 30 days, no prolonged length-of-stay (**Textbook Outcome**)
- No revision within one year, no readmission within 30 days, prolonged length-of-stay
- No revision within one year, readmission within 30 days, no prolonged length-of-stay
- No revision within one year, readmission within 30 days, prolonged length-of-stay
- Revision within one year, no readmission within 30 days, no prolonged length-of-stay
- Revision within one year, no readmission within 30 days, prolonged length-of-stay
- Revision within one year, readmission within 30 days, no prolonged length-of-stay
- Revision within one year, readmission within 30 days, prolonged length-of-stay

This ordered outcome measure can also be corrected for patient-mix by using different funnel-plots where group 1 is compared with the rest, group 2 versus the rest, etc.. With this method, it can be indicated in which group your hospital differs significantly. These funnel- plots can be supplied on request.

If a hospital deviates from one of these groups, specific file investigation on these patients could be performed. For possible quality improvement initiatives, I refer to the Toolboxes for revision, readmission and prolonged length of hospital stay.

IQ Joint studiegroep:
•P van Schie, coördinerend arts-onderzoeker
•Prof. dr. RGHH Nelissen, orthopedisch chirurg
•Dr. PJ Marang-van de Mheen, assistent professor medisch besliskunde

Appendix V Surveys

The survey was emailed every two months together with the feedback to evaluate adherence to the intervention, encourage reviewing the feedback, introduce QII if necessary, and stimulate trial engagement. Participants could report best practices and experiences to be added to the toolbox and shared with others, also to stimulate trial engagement. The questionnaire has been translated from Dutch.

1. Have you read and interpreted the monthly feedback containing the performance outcomes of primary total hip and/or knee arthroplasties (THP/TKP)?
   a. Yes, the feedback was clear
   b. Yes, but the feedback was not clear because ……
   c. No, I did not get around to this

2. Has the feedback been discussed within your department?
   a. Yes, we saw potential to improve at least one of the performance outcomes
   b. Yes, but we did not see any potential to improve one of the performance outcomes
   c. No, we did not get around to this

3. Are there any performance outcomes for which, according to the feedback, your center underperforms or has performed worse than the average for the IQ Joint study group? You may select multiple outcomes.
   a. Revision within one year
   b. Readmission within 30 days
   c. Complications within 30 days
   d. Long length of stay
   e. None

4. For which performance outcome(s) have improvement initiatives been undertaken since the start of the study (October 2019)? Enter what you/your healthcare institution have done in the free text field below the relevant performance outcome. You may select multiple outcomes.
   a. Revision within one year
      i. ……..
   b. Readmission within 30 days
      i. ……..
   c. Complications within 30 days
      i. ……..

d. Long length of stay
   i. ……..
e. None

5. If applicable, indicate for each improvement initiative undertaken to improve a performance outcome, how your center chose this improvement initiative.

	Toolbox	Literature	Expert advice	Other (free text field)	No improvement initiative undertaken for this outcome
Revision	Yes/No	Yes/No	Yes/No	…..	Yes/No
Readmission	Yes/No	Yes/No	Yes/No	…..	Yes/No
Complications	Yes/No	Yes/No	Yes/No	…..	Yes/No
Long length of stay	Yes/No	Yes/No	Yes/No	…..	Yes/No

6. What would you or your center need to improve further? You may select multiple answers.
   a. Further explanation of the interpretation of performance outcomes in the feedback
   b. Link to another center that scores better for an outcome for which we score worse
   c. More items in the toolbox, namely ……

7. This is the last question. You can still click back to make adjustments. After this question, the questionnaire is sent immediately. There is still room for questions and/or suggestions in the free text field.

……..…….

Thank you for participating in this survey. In order to be able to learn from each other and to give each other new ideas, the improvement initiatives of all participating centers will be included in the Toolbox.

Chapter 9

Summary, general discussion, and future perspectives

The objective of this thesis was to study how arthroplasty registries can improve their feedback to orthopaedic surgeons in order to give direction to quality improvement initiatives (QII) that improve care for total hip and knee (THA and TKA). A second aim was to evaluate the effectiveness of such improved feedback on patient outcomes.

Giving feedback on performance indicators is a frequently used approach to improve the quality of care delivered. In this context, feedback is defined as the provision of clinical outcome summaries to healthcare providers or organizations intended to initiate activities to improve the performance of delivered care(1,2). Internationally, feedback from arthroplasty registries is provided in various ways. In the Netherlands, clinical outcomes are shown at the hospital-level in a real-time secured web-based dashboard from the LROI. The extent of between-hospital variation is shown in an anonymized version in annual reports.(3,4) The effect of feedback varies (i.e., from a 9% decrease to a 70% increase), but an optimal design will reasonably improve patient care.(5) This thesis provided an overview of national and international between-hospital variation in clinical outcomes to investigate whether improvement is achievable. In order to optimize the content of the feedback, methodological studies have been performed to investigate whether outlier hospitals can be detected earlier, whether reasons for higher revision rates can be identified, and a composite outcome measure is developed and tested. As awareness of performance by surgeons in combination with motivation to improve is more likely to result in targeted QII improving quality, associations with such awareness were assessed to increase feedback effectiveness.(6-8) The knowledge obtained in combination with up-to-date theory for providing effective feedback was incorporated in a multifaceted quality improvement intervention and tested on its effectiveness in a cluster randomised controlled trial.(5,9-11)

This chapter starts with a summary of the main findings, including the practical implications of the previous chapters. Subsequently, these findings and relevant methodological issues are discussed in the context of available literature, and finally, recommendations for future practice and research are given.

Summary and practical implications

Arthroplasty registries were initially established to compare implant survival and monitor the safety of different orthopaedic implants, like total hip and knee arthroplasties (THA and TKA). In recent years, however, registries have also been as quality systems across the healthcare system to show the variation between hospitals for numerous clinical outcome measures, thus providing feedback to hospitals on their performance. The latter is usually compared with a reference standard (i.e., the

benchmark).(12-19) Most registries provide feedback through annual reports intended to encourage QII in low-performing hospitals.(12-19) However, this information needs to be viewed by an action of the orthopaedic surgeon (i.e., log into the secured website to view the data within a secured site). Furthermore, also top-performing hospitals may be interested in comparing their performance with hospitals from other countries to stimulate further improvement within specific domains. Fair international hospital comparison is only achievable when consistent outcome definitions are used, as these will determine the frequency of occurrence.(20-23) Consistency in outcome measure definitions makes it also possible to merge international data providing better opportunities to detect rare safety issues earlier (e.g., the metal-on-metal hip arthroplasty disaster, modular femoral neck corrosion etc.), which will prevent thousands of patients from being exposed to poor performing implant designs, thus decreasing unnecessary suffering in future.(24,25) **Chapter 2** showed that among registry reports and arthroplasty cohort studies: revision, readmission, and complications are the most frequently reported clinical outcomes, with considerable differences in their outcome rates between hospitals, indicating a vast improvement potential for at least some hospitals. However, part of the variation may be explained by the significant heterogeneity in the following domains: 1) outcome definitions, including what is a revision, readmission, or complications, 2) duration of follow-up and starting point of follow-up, 3) characteristics included in patient-mix adjustment, and 4) type of patients- and hospital included. This thesis showed that revision of the implant within five years, readmission within 30 days, and complications up to 2 years postoperative were the most commonly used outcome measures in arthroplasty reports. However, none of these definitions had a perfect agreement with the other domains for THA, TKA, and THA&TKA combined. The least consensus was found on whether or not to adjust for patient characteristics, let alone which characteristics should be included in the adjustment. Although the latter as well as the other domains investigated in this study are essential for fair hospital comparison.(23) In the future, partnerships of arthroplasty registries such as the International Society of Arthroplasty Registries can play a leading role, not only in international collaboration but also striving for more uniformity in the definitions and methods used.(26)

Reporting the between-hospital variation in clinical outcomes and identifying positive and negative outlier hospitals is a simple and effective way to get insight into hospital performance, provided that the rankability (i.e., the percentage of total variation that is explained by "true" hospital differences rather than chance variation) is acceptable.(27-30) However, assessing how to pursue improvement for a given clinical outcome can be challenging, particularly for summary outcomes such as all-cause revision, but this can be facilitated by examining specific indications for revision that may be the reason for the identified worse performance on all-cause revision. **Chapter 3**

showed large variation in 1-year all-cause revision rates between Dutch hospitals with moderate rankability (61%) for THA and low rankability (46%) for TKA within a 3-year time frame, indicating huge improvement potential for a considerable part of the Dutch hospitals. Earlier detection of poor performance using a 1-year time frame has the advantage that QII can be introduced earlier; however, this resulted in low rankabilities and is therefore not recommended. Underlying reasons for worse performance on all-cause revision were found for 12 of the 13 negative outlier hospitals for THA and 3 of the 7 for TKA, mainly consisting of infection (both for THA and TKA) and dislocations (only THA). Implant loosening and technical failure (only TKA) were less likely to be the underlying reason for the worse performance. Rankabilities for the specific indications for revision were all low within a 3-year time frame for THA and TKA, except for infection for THA, for which the rankability was moderate (i.e., 61%). As rankabilities within a 1-year timeframe were all low, it is recommended to use a 3-year time frame to identify underlying reasons for worse hospital performance on all-cause revision.

Where **chapter 3** showed that earlier detection of poorer performance could not be done reliably using funnel plots within a 1-year time frame, the monthly monitoring of revision rates using CUSUM charts with 5 control limits shown in **chapter 4** was able to detect worsening performance earlier than the conventional funnel plots. The first signal for negative outliers was generated at a median of 18 months for THA and 21 months for TKA within a 3-year time frame. CUSUM charts thereby enable detection of deteriorating patterns earlier, making it possible to introduce QII earlier than waiting for the results to appear in the funnel plot after 3 years. This thesis adds to the existing literature how much earlier a signal was generated and with what accuracy (i.e., 97% both for THA and TKA) compared with the traditional funnel plot with a 3-year time frame. These results are highly relevant for registries and scientific associations deciding whether to implement CUSUM charts in their organisation to improve quality.(31) The results on accuracy will contribute to professionals' confidence in CUSUM charts. In response to these findings, the Dutch Arthroplasty Register (LROI) has added CUSUM charts to their routine dashboard reporting on clinical outcomes to provide hospitals with a tool for earlier detection and thereby the opportunity to introduce QII earlier to improve patient care.(32)

By increasing the number of events, composite outcomes may also enable that differences between hospitals in their performance to be detected sooner.(33-36) In **chapter 5**, an ordered composite outcome with all combinations of clinical outcomes (i.e., revision, readmission, complications, and upper-quartile LOS) ranked from best to worst according to the patient's perspective was developed and tested on its ability to differentiate between hospitals. The newly developed composite showed higher

rankability than individual clinical outcomes due to the larger variation between hospitals when more information is included. The composite could reliably differentiate between hospitals in their performance using a 1-year time frame, rather than requiring the usual 3-year time frame, allowing the introduction of QII earlier. An additional advantage is that the composite can measure more aspects of delivered quality of care as hospitals may perform well on one outcome while at the same time performing worse on another. The composite overcomes this issue and shows whether a patient had a revision but also whether they were readmitted, experienced complications, or had a prolonged LOS indicating more specifically where improvement is possible (e.g., in patients with a normal LOS, without complications who were readmitted). The new composite is widely applicable as it may help healthcare providers to select for which patient groups' medical records have to be reviewed to investigate whether and how care can be improved. For instance, rather than reviewing all records for patients who were readmitted, it allows to selectively review only those with a normal LOS without any registered complications, to understand e.g., whether information at discharge needs to be improved to avoid readmission or discuss whether the readmission was needed or could have been treated at the outpatient clinic. Another application is that it is visible in one outcome whether a focus to improve, for example, length of stay, does not come at the expense of another outcome (e.g., readmission). Finally, patients can simply check how often a procedure goes as planned (i.e., without any clinical outcome occurring).

Clinical outcomes such as revision, readmission, complications, and LOS measure unintended adverse events and generally occur with low frequency for THA and TKA.(37) However, up to 10% and 20% of patients following THA and TKA, respectively, are dissatisfied with results, mainly related to persistent pain and disability.(38,39) Patient-Reported Outcome Measures (PROMs), on the other hand, measure the intended outcomes such as pain reduction, functionality improvement, and health-related quality of life, and would therefore complement these clinical outcomes by identifying potential additional areas for improvement in these intended outcomes.(40-44) Similar to the need for a high level of data completeness regarding clinical outcomes to ensure there is no selection bias, we also need high response rates of patients completing both pre- and postoperative questionnaires to allow calculating the improvement in PROMs, or if that is not feasible at least gain insight into how those who complete questionnaires are a selection of all patients. Dutch response rates, as in other national and regional arthroplasty registries, were low (i.e., less than 61% of patients completed the preoperative PROM questionnaire, and only about 40% of patients completed preoperative and postoperative PROM questionnaires) compared with the above 98% completeness of revision, surgical procedure, implant, and patient characteristics data in registries for both THA and TKA.(40-45) Previous

studies have already shown differences in various patient characteristics, such as patients completing questionnaires, in general, being healthier, more likely to be white, with higher literacy rates, and lower rates of cognitive impairment, including dementia.(46-48) To better understand whether missing PROM data for THA and TKA may result in under- or overestimation of PROMS improvement scores, **chapter 6** used clinical outcome rates (i.e., revision, readmission, complications, and upper-quartile LOS) to examine whether these differed between respondents and non-respondent, as well as their association with PROM improvement scores. Chapter 6 showed that respondents to PROM questionnaires less often experienced adverse events. This likely results in an overestimation of the clinically relevant improvement in PROMs as adverse events were associated with a lower likelihood to achieve a clinically relevant PROM improvement. Given the observed patient-level associations in **chapter 6**, it is likely that initiatives to improve the quality of care by reducing readmission, complications, and long LOS for both THA and TKA patients will lead to more patients achieving clinically relevant improvement in HOOS-PS and KOOS-PS scores. Hospital differences in PROM response rates were not associated with differences in adverse event rates, suggesting that estimated between-hospital differences in PROM improvement are likely unaffected.

Feedback may be methodologically sound, but it is only effective if it is viewed and interpreted by practitioners. **Chapter 7** showed that only half (i.e., 55%) of Dutch orthopaedic surgeons performing THA and TKA were aware of their outlier performance status regarding revision rates. Awareness was higher among surgeons that more often logged in on the LROI dashboard, more often interpreted funnel plots correctly, and more often could recall the 1-year revision rates of their surgeon group. Thirty-eight percent of THA and 26% of TKA surgeons met all three conditions necessary to act upon the feedback information, i.e., logging in, correct interpretation of funnel plot, and could recall their 1-year revision rates. Forty-five percent of surgeons in a hospital identified as a negative outlier reported not seeing their worsening performance coming, meaning they continued to provide care without modifications. Thus, a focus on making feedback more effective is very important, as 85% of surgeons indicated that they did start QII once being identified as having worse performance. Logging in on the LROI dashboard should be made more attractive and encouraged, for example, by emphasizing the importance of already reported clinical outcomes (e.g., revision rates) and adding new outcomes such as prosthesis survival, complications, readmissions, and length-of-hospital-stay as these are considered relevant by a large part of surgeons. Second, teaching material should become available to improve interpretation skills of statistical presentation of data like funnel plots, CUSUM charts etc., or provide explanatory text with these funnel plots on the LROI dashboard to help surgeons understand what the data in the figure

represent. Third, feedback should be sent to become easily accessible, readily read (e.g., infographics), and tailored to single hospitals rather than expecting surgeons to make any selections and investigate themselves.

The knowledge obtained from the previous chapters with contemporary theory for providing effective feedback was incorporated into a multifaceted quality improvement intervention and tested on its effectiveness in **chapter 8**.(5,9-11) The intervention was applied over eight months and included monthly feedback, education on the interpretation of the feedback, and an action implementation toolbox including evidence-based QII. Hospitals that received the intervention improved 4.3% more on the Textbook Outcome (i.e., the absence of revision, readmission, complications, and long LOS, the best possible outcome in the composite developed in **chapter 5**) compared with control hospitals. It was found that the effect size was larger for intervention hospitals that introduced QII, suggesting that these QII were likely the reason for the better outcomes. The median number of TKA surgeries performed was considerably smaller than for THA in intervention hospitals, which could explain why the effect for TKA was not significant when outcomes for THA and TKA were analysed separately, even though the effect size pointed in the same direction of improvement. In addition, the difference could be explained by the lower baseline risks for revision and complications for TKA, associated with smaller absolute risk reduction. These findings in **chapter** 8 support that frequent feedback to surgical teams should be supplemented by interactive education and facilitated by evidence-based QII tailored to specific outcomes to improve the quality of care regarding THA and TKA effectively.

General discussion

Components of the quality intervention

Quality improvement interventions are a common strategy to improve patient outcomes, but with highly variable effects across studies.(5,49,50) Two meta-analyses show that quality interventions using only one single intervention component are less effective, with little to no improvement.(51,52) Thus, quality interventions should be designed in a multifaceted way, including components addressing, for example, the gap in knowledge and (surgical) skills, but may also address other components such as audit and feedback that will allow to evaluate the impact of changes made and spur further improvement actions. Even though education on quality parameters is needed in most quality improvement efforts, serious limitations are present when they are used without proper context and a predetermined goal.(9) Education can only be effective if it solves a knowledge problem which is a barrier for quality

improvement. Nevertheless, designing proper multifaceted interventions is less than easy and straightforward. The quality intervention reported in **chapter 8** serves as a model for future intervention efforts by combining theory and previous evidence, thus increasing the likelihood of effectiveness in daily clinical practice.

The first intervention component consisted of monthly updated feedback sent by email to all individual orthopaedic surgeons. The monthly time interval was chosen based on previous evidence that feedback on performance is more effective when given repeatedly and not only once. The reason might be that recipients are more likely to perceive such repeated feedback as more relevant and accurate than only once a year since monthly data feedback is closer to current performance and can thus be easier related to current clinical practice.(5,53) Furthermore, these monthly feedbacks also allow for timely evaluation of the introduced QII, and thus any subsequent improvement actions without delay can be done. Arthroplasty registries can support this as data are routinely collected that could easily be used for near real-time monitoring of clinical outcomes. Sending the feedback by email was based on orthopaedic surgeons' preference of receiving feedback by email and our finding that 33% of orthopaedic surgeons never logged in on the LROI dashboard (**chapter 7**). This adds to the literature regarding the most effective mode of delivery (e.g., electronic, paper, face-to-face) and frequency (e.g., monthly, quarterly, or yearly) which have not been well assessed to date. Moreover, our intervention tapped in on previous knowledge that feedback is more effective when it is also discussed orally in a group, preferably by a senior colleague, rather than just presenting written data to individuals.(5) As for the former, the researcher visited all intervention sites to explain the feedback orally, although this intervention might have been improved if a senior colleague would visit the hospitals.(54,55) Another important aspect we considered was the sender of the feedback. It is known that when feedback comes from a regulatory body, recipients may be more likely to activate affective processes (e.g., distress), distracting attention from the specific task requiring change. In contrast, the current intervention was nested in the registry from which they "trusted" data and procedures for how data were collected and processed.(1)

As for the feedback, the comparator hospitals are essential in helping clinicians assess their performance and identify differences between current and desirable performance.(56) However, the choice of comparators may have critical implications for what message is conveyed by the feedback and how recipients will react.(57) For this reason, we performed stratified randomisation to ensure that, e.g., academic hospitals treating complex cases would not only be compared with private hospitals treating only the healthiest patients (which can never be captured entirely by case-mix adjustment). We included funnel plots in the feedback, as this may already be

a familiar visual presentation because it was available in the LROI dashboard and annual reports, although we realized this required some education based on our findings that 39% of the orthopaedic surgeons did not interpret the funnel plot correctly (**chapter 7**). The funnel plot allowed participants to compare performance with other hospitals. However, the funnel plot will only give an average estimate of performance across a period e.g., 3 years, so we added a CUSUM chart to more clearly show trends over time e.g., indicating whether hospitals were moving in the right direction after implementing a new QII. This trend of performance change (as shown in CUSUM charts) is more motivating to introduce new QII than the distance between performance and best performers (as shown in funnel plots).(58,59) In addition, trends increase the credibility of feedback and enable the introduction of QII according to the *Plan-Do-Study-Act (PDSA)* cycle, in which recipients continuously self-assess their performance and the effect of QII when deciding whether or not to take action. Trends, therefore, add meaningful information and should be added by default to performance feedback. In addition, **chapter 4** showed that changes in hospital performance were detected 18 months and 15 months earlier for THA and TKA, respectively, than the conventional funnel plots using a 3-year time frame.

With regard to comparing hospital performances, a recently published review suggests that comparison with high-performing peers is preferred over benchmarking hospitals to a national average, as it shows that top performance could be achieved. (60) In addition, psychological theories suggest that clinicians are less likely to accept an "externally imposed" performance goal (e.g., 1-year revision rate of less than 1% imposed by an outside party) and that recipients in such cases are more likely to reject feedback recommendations and pursue self-conceived performance levels.(56,58,61). The funnel plot fits well with this understanding, as the performance compared to other hospitals and performance outcomes can be adjusted for differences in case-mix. Due to case-mix adjustment, hospitals that mainly treat patients without comorbidities and therefore expected to have lower frequencies of adverse events could be fairly compared with hospitals that mainly treat patients with multiple comorbidities.(20-22) However, when large differences in hospital performances consist, low performers may experience the feedback as unfeasible and reject the feedback. Tailoring the feedback to individual hospitals could avoid feedback rejection, for example, by comparing low performers with the top 50% and average performers with the top 10%.

The second intervention component consisted of education to interpret the feedback on performance, which was provided during an on-site visit in the first month of the intervention. The latter was attended by a majority of the orthopaedic surgeons within a hospital (**chapter 8**). An online educational video and pocket card containing a summary of the educational meetings were available as a reference. Education was

needed as **Chapter 7** showed that 39% of orthopaedic surgeons could not interpret funnel plots correctly. Even more, they often overestimated their performance if unaware of their performance data.(56) Even more important, "unawareness" will limit the undertaking of necessary QII, because it is assumed that performance is good enough even though there may be room for improvement. It seems evident that correctly interpreted feedback will improve the quality of care, as 17 out of 20 orthopaedic surgeons indicated that they would conduct QII when becoming aware of worse performance compared to the national benchmark. At the end of the educational meeting, explicit goals and specific actions for improvement were discussed, as this will improve implementation and intervention effectiveness.(5,49) This also aligns with theories that goals aimed for can make feedback more tangible for clinicians and thus help to facilitate better-focused action plans, which facilitate steps needed to achieve predefined goals.(62,63) The improvement process is more effective when goals are considered specific, measurable, achievable, relevant, and time-bound, allowing the recipients' attention to be more productive on the task.(53,58,64) We encouraged surgeon groups to set their own goals and create their own action plans as goals may otherwise not be acceptable for a subset of clinicians, even if a credible orthopaedic authority set them (e.g., a national association).(56,58,61) Not embracing the set goals will improve the chance of feedback rejection and impede intentions to improve care, thus likely diluting the effects of the quality interventions.(61) In this way, we also aimed for an established engagement with goals and action plans. To further stimulate engagement, bimonthly questionnaires were sent to all orthopaedic surgeons allocated to the intervention group to verify compliance with self-set goals and action plans. Furthermore, monitoring of progress in achievement of these plans was done.

The final intervention component was an action implementation toolbox including evidence-based QII for each clinical outcome reported in the feedback, which was added to overcome the barrier of translating feedback into what needs to be improved in clinical practice. Clinicians have often been shown to lack the skills or knowledge to interpret statistical feedback and formulate what QII is necessary to improve.(5,7,65,66) The toolbox bridges this gap and lowers the barrier to implementing evidence-based quality improvement initiatives. Adding a toolbox to a quality intervention has shown to be an improvement compared to feedback alone, but only in process indicators and not clinical outcomes.(11) As shown in **chapter 8**, our intervention showed an improvement in clinical outcomes, which could be due to the fact that we included evidence-based QII in the toolbox that targeted the outcomes. In contrast, the toolbox in the study of Roos-Blom et al. mostly targeted process measures, such as the availability of a protocol rather than the outcomes.(11)

In this thesis, we worked on a model for the design of a quality improvement intervention, where insight was gained into potential barriers of the target group in **chapter 7,** and feedback was methodologically improved in **chapters 3, 4, and 5**. Gained knowledge on these topics was supplemented with theory to provide effective feedback (i.e., education, discussing goals and actions, and a toolbox). However, there were also indications for intervention improvement as four of the ten intervention hospitals indicated that they needed additional information on the interpretation of funnel plots and CUSUM chats despite the offered educational session. Two hospitals indicated that they would appreciate more QIIs in the toolbox, and seven hospitals indicated that they would like to be matched with hospitals to exchange information on best practices and identify areas for further improvement. Unfortunately, the latter was initially planned but was not executed due to government restrictions related to the COVID-19 pandemic.(67) Finally, the intervention period could have lasted longer than eight months, as the bimonthly surveys showed that some intervention hospitals started implementing QII after several months. Therefore, it is possible that the end effect of a QII has not yet been achieved at the time the intervention was evaluated after eight months. The intervention, as reported in this thesis, will probably not meet the target group's needs in the future as it is possible that new components would fit better with barriers at that time. This makes designing an appropriate quality intervention, like quality improvement, a continuous improvement process.

However, similar to other multifaceted quality interventions that were tested, it is unclear to what extent each single component of the intervention (i.e., feedback, education, and an action implementation toolbox) contributed to the 4.3% (95% confidence interval 4.30% to 4.34%) absolute improvement in the intervention compared to control hospitals.(5,49,66) **Chapter 8** showed that the intervention effect was most likely achieved through the introduction of targeted QII, making this likely the causal link to the improved patient outcomes, demonstrating that if surgeons are sufficiently engaged to introduce QII, it will improve patient care.(68)

Sustainability

Even if a quality improvement intervention positively affects the quality of care delivered, maintaining access to resources available during the intervention is likely needed to sustain the improvement or even continue to achieve further gains. Little is known about why a quality intervention is sustainable, as most empirical data demonstrate a lack of sustainability, and only a few studies report on sustainability and adoption in everyday practice after the initial improvement initiative ended. (69-72) Implementing a package of common quality interventions (e.g., feedback, education, alerts) as a quick fix to resolve poor hospital performance may then provide a temporary solution but is generally unsustainable.(69,70,73) In the end, it is not

the number of implemented quality interventions by a hospital that is a measure of success but rather the ability to sustain the interventions in the long term.(74-76) To achieve that goal, an effective and sustainable intervention must offer a solution for the underlying problem as the first step, but it also needs to be adapted to the environment and use resources that will continue to be available after the intervention ends to be sustainable and become part of everyday practice.(73,77)

During intervention design, the application of the intervention components by the Dutch Arthroplasty Registry (LROI) was considered, so that intervention components and resources would remain available as much as possible if the quality intervention proved effective. For example, the CUSUM charts shown in **chapter 4** to enable earlier detection of worsening performance, have already been implemented and are currently part of the routine LROI dashboard as well as promoted through communication by the Dutch Orthopaedic association to reach other hospitals that did not take part in the IQ Joint study.(78) In addition, following the IQ Joint study, it is now also possible to register complications in the LROI database in order to further improve the quality of care and safety. The education video to explain the statistical feedback information and pocket card to be used as a reminder in clinical practice remained available, as well as the action implementation toolbox used during the study. The latter must be kept up-to-date and potentially further expanded if new evidence-based QII effectively improves the targeted clinical outcomes. Furthermore, to promote continued engagement with improving the quality of care as an integrated part of orthopaedic patient care, an annual educational session or workshop may act as a stimulus and platform for exchanging best practices to motivate the hospitals to improve their care continually. Finally, the IQ joint study group provided a positive and safe improvement climate, where the created collaboration of hospitals may continue to improve the quality of care in future projects.

Future perspectives

Relevant future directions of research to increase the effectiveness of quality improvement initiatives and prevent adverse effects are mentioned in the following section. The focus is on further developing quality interventions for a specific problem and determining the effect of the prosthesis on the quality of care

Tailoring quality interventions to a specific problem in practice

Even though there is evidence that some bundled interventions are more effective in improving care, the complicated matter of how quality interventions exercise their effect needs to be further unraveled, as engagement with and impact of interventions

is variable.(5,79) The lack of systematic, coordinated research in this field and gaps in studies reporting on the likely mechanism of the intervention effect perpetuates this problem. Only reporting outcome effects without concurrent publication of the process evaluation leaves the reader to guess at the reason for the main findings because they cannot learn from the underlying processes or barriers that may or may not have affected the outcome. Process evaluation is therefore needed to contextualize and understand the effect of the intervention.(80) Process evaluations can focus on the uptake of intervention components and are often based on a mix of interviews, focus groups, and field notes, which can substantiate the intervention's fidelity. For example, as was reported in **chapter 8**, the number of study participants interpreting the feedback, attending educational sessions, and introducing QII were reported. More knowledge on factors hindering or facilitating effective interventions is needed.

Just as care is tailored to the individual patient, quality interventions should also be tailored to the daily barriers and problems clinicians face. Initiatives that truly help clinicians achieve their goals are likely to be well-received, unlike many quality interventions leading to unenthusiastic engagement and unsuccessful outcomes due to little clinical relevance.(49,50) For that matter, some studies state that more than a third of participants who sign up for a quality intervention are not actively participating in the intervention.(81) These numbers are unlikely to increase as long as a one-size-fits-all approach is maintained that focuses on the outcome from a theoretical perspective rather than a clinically relevant perspective, understanding the resources and skills required to achieve better performances. To address this problem, the participating hospitals defined which clinical outcome they wanted to improve (i.e., revision, readmission, complications, and long LOS). Nevertheless, we could still assess the intervention's effect by focusing on a composite outcome as the primary outcome, including all clinical outcomes. In addition, prior to the implementation of the intervention, insight was obtained into the desired resources and skill level of clinicians. The latter requires investments in improving methods to provide information clinicians need to improve on but also ensures a better understanding of the clinical physicians' attention we are trying to capture and the behaviour we are trying to change. Who are these clinicians, what matters to them, and how do their goals align with intervention goals? Answering these questions will facilitate more tailored interventions clinicians want to participate in. Systemically categorising clinicians' preferences, knowledge, skills, and goals is the first step toward achieving a targeted understanding of what needs to be addressed to ensure the quality intervention is designed to drive practice changes. Other topics relevant for future research include the culture in which quality interventions take place, which is part of the contextual factors that can influence the effect enormously. The culture within a clinic strongly influences how feedback is being responded to, even if it is highly credible (i.e., trusting), from a trustworthy source, and constructive (i.e.,

usefulness and actionability), which will generally strengthen the effectiveness of quality interventions.(82) In addition, the culture significantly influences how clinicians work together, engagement and motivation to work on quality improvement.(68) Finally, a hospital culture with room to implement QII is needed in which clinicians feel supported even when the effect is disappointing. Truly exercising PDSA cycles to improve care also means abandoning interventions if there are no or negative effects.(83) Future research should focus on quality intervention trials incorporating up-to-date evidence- and theory-based best practices and address knowledge gaps. In addition, there should be a shift towards tailored audit and feedback studies rather than one-size-fits-all two-arm studies where multiple quality interventions are tested simultaneously.(84) Then the most effective intervention is implemented in clinical practice or taken as a starting point for future research.(1)

Implant choice

Rather than focusing on the quality of care delivered, another possibility to improve care is to focus on choosing the most optimal implant for a particular patient. **Chapter 3** showed that most negative outlier hospitals for overall revision were also a negative outlier for a specific indication for revision (e.g., infection, dislocation, or implant loosening), allowing hospitals to introduce targeted QII to improve the quality of care delivered, thereby lowering their revision rates. However, the specific implant was not included in the analysis, while the implant type can have significant effects on the overall revision rate as well as on specific indications for revision.(85-88) For example, metal-on-metal, large-head, uncemented, and resurfacing implants increased the risk of revision surgery after primary THA compared with metal-on-polyethylene, 32mm diameter heads, and cemented implants, respectively.(85,88-90) Also, introducing a new implant can be a reason for an increase in revisions since the instrumentation is slightly different as well as details on implant placement. The latter affects the surgical team and the surgeon. The Orthopaedic Data Evaluation Panel (ODEP) aims to guide surgeons' implant choices by classifying implants based on whether they have revision rates at an acceptable predefined level (i.e., an established external benchmark based on revision rates). Data are based on both observational single cohort studies as well as registry data with sufficient follow-up and sample size.(91) However, specific patients may benefit more from one type of implant while others may benefit more from a different type depending on patient and implant characteristics. Therefore, a prediction model to guide the most optimal implant choice for a specific patient (group) by estimating the lowest revision risk tailored to the individual patient (group) based on items like sex, age, femoral head size, comorbidity score, and activity level could, in theory, improve performance of arthroplasty surgery for patients.(92-94) Such a prediction model could represent the next step forward in improving the quality of care for patients after THA and TKA.

References

1. Ivers NM, Sales A, Colquhoun H, Michie S, Foy R, Francis JJ, et al. No more 'business as usual' with audit and feedback interventions: towards an agenda for a reinvigorated intervention. Implement Sci. 2014;9:14.
2. Brehaut JC, Eva KW. Building theories of knowledge translation interventions: use the entire menu of constructs. Implement Sci. 2012;7:114.
3. Website orf the Dutch Arthroplasty register (LROI). https://www.lroi.nl/. Checked on June 14, 2021.
4. Dutch Arthroplasty Register (LROI). Annual report 2019. www.lroi-report.nl.
5. Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, French SD, et al. Audit and feedback: effects on professional practice and healthcare outcomes. Cochrane Database Syst Rev. 2012(6):Cd000259.
6. van der Veer SN, de Keizer NF, Ravelli AC, Tenkink S, Jager KJ. Improving quality of care. A systematic review on how medical registries provide information feedback to health care providers. Int J Med Inform. 2010;79(5):305-23.
7. de Vos Maartje LG, van der Veer SN, Graafmans WC, de Keizer NF, Jager KJ, Westert GP, et al. Process evaluation of a tailored multifaceted feedback program to improve the quality of intensive care by using quality indicators. BMJ quality & safety. 2013;22(3):233-41.
8. Davis DA, Mazmanian PE, Fordis M, Van Harrison R, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. JAMA. 2006;296(9):1094-102.
9. Soong C, Shojania KG. Education as a low-value improvement intervention: often necessary but rarely sufficient. BMJ quality & safety. 2020;29(5):353-7.
10. Gude WT, Roos-Blom MJ, van der Veer SN, de Jonge E, Peek N, Dongelmans DA, et al. Electronic audit and feedback intervention with action implementation toolbox to improve pain management in intensive care: protocol for a laboratory experiment and cluster randomised trial. Implement Sci. 2017;12(1):68.
11. Roos-Blom MJ, Gude WT, de Jonge E, Spijkstra JJ, van der Veer SN, Peek N, et al. Impact of audit and feedback with action implementation toolbox on improving ICU pain management: cluster-randomised controlled trial. BMJ quality & safety. 2019;28(12):1007-15.
12. Norwegian Arthroplasty Register (NAR). Annual Report 2020. http://nrlweb.ihelse.net/eng/Rapporter/Report2020_english.pdf.
13. Dutch Arthroplasty Register (LROI). Online Annual Report. https://www.lroi-report.nl/.
14. Swedish Knee Arthroplasty Register (SKAR). Annual Report 2020. http://myknee.se/pdf/SVK_2020_Eng_1.0.pdf.
15. Danish Hip Arthroplasty Register (DHAR). Annual report 2020. http://danskhoftealloplastikregister.dk/wp-content/uploads/2020/11/dhr-aarsrapport-2020_til_offentliggoerelse-1.pdf.
16. Swedish Hip Arthroplasty Register (SHAR). Annual Report 2018. https://registercentrum.blob.core.windows.net/shpr/r/Arsrapport_2018_Hoftprotes_ENG_26mars_Final-rJepCXNsLI.pdf.
17. Danish Knee Arthroplasty Register (DKAR). Annual Report 2020. https://www.sundhed.dk/content/cms/99/4699_dkr-arsrapport-2020_offentliggorelse.pdf.
18. Finnish Arthroplasty Register (FAR). Online Annual Report. https://www.thl.fi/far/#index.
19. Swiss Arthroplasty Register (SAR). Annual Report 2020. https://www.siris-implant.ch/de/Downloads&category=16.

20. Lenguerrand E, Whitehouse MR, Beswick AD, Kunutsor SK, Burston B, Porter M, et al. Risk factors associated with revision for prosthetic joint infection after hip replacement: a prospective observational cohort study. Lancet Infect Dis. 2018;18(9):1004-14.
21. Prokopetz JJ, Losina E, Bliss RL, Wright J, Baron JA, Katz JN. Risk factors for revision of primary total hip arthroplasty: a systematic review. BMC Musculoskelet Disord. 2012;13:251.
22. Jasper LL, Jones, C. A., Mollins, J., Pohar, S. L., Beaupre, L. A.. Risk factors for revision of total knee arthroplasty: a scoping review. BMC Musculoskelet Disord. 2016;17:182.
23. Signorini DF, Weir NU. Any variability in outcome comparisons adjusted for case mix must be accounted for. BMJ. 1999;318(7176):128.
24. Drummond J, Tran P, Fary C. Metal-on-Metal Hip Arthroplasty: A Review of Adverse Reactions and Patient Management. Journal of functional biomaterials. 2015;6(3):486-99.
25. Pfuntner A, Wier LM, Stocks C. Most Frequent Procedures Performed in U.S. Hospitals, 2011: Statistical Brief #165. Healthcare Cost and Utilization Project (HCUP) Statistical Briefs. Rockville (MD): Agency for Healthcare Research and Quality (US); 2006.
26. International Society of Arthroplasty Registries (ISAR). https://www.isarhome.org/home.
27. van Dishoeck AM, Koek MB, Steyerberg EW, van Benthem BH, Vos MC, Lingsma HF. Use of surgical-site infection rates to rank hospital performance across several types of surgery. Br J Surg. 2013;100(5):628-36; discussion 37.
28. van Dishoeck AM, Lingsma HF, Mackenbach JP, Steyerberg EW. Random variation and rankability of hospitals using outcome indicators. BMJ quality & safety. 2011;20(10):869-74.
29. Adab P, Rouse AM, Mohammed MA, Marshall T. Performance league tables: the NHS deserves better. BMJ. 2002;324(7329):95-8.
30. Cram P, Cai X, Lu X, Vaughan-Sarrazin MS, Miller BJ. Total knee arthroplasty outcomes in top-ranked and non-top-ranked orthopedic hospitals: an analysis of Medicare administrative data. Mayo Clin Proc. 2012;87(4):341-8.
31. Neuburger J, Walker, K., Sherlaw-Johnson, C., van der Meulen, J., Cromwell, D. A.. Comparison of control charts for monitoring clinical performance using binary data. BMJ quality & safety. 2017;26(11):919-28.
32. Landelijke Registratie Orthopedische Interventies (LROI); CUSUM chart implementation. https://www.lroi.nl/nieuws/dashboard-uitgebreid-survival-cusum-en-indicatoren.
33. Hofstede SN, Ceyisakar IE, Lingsma HF, Kringos DS, Marang-van de Mheen PJ. Ranking hospitals: do we gain reliability by using composite rather than individual indicators? BMJ quality & safety. 2019;28(2):94-102.
34. Karthaus EG, Lijftogt N, Busweiler LAD, Elsman BHP, Wouters M, Vahl AC, et al. Textbook Outcome: A Composite Measure for Quality of Elective Aneurysm Surgery. Ann Surg. 2017;266(5):898-904.
35. Dimick JB, Birkmeyer NJ, Finks JF, Share DA, English WJ, Carlin AM, et al. Composite measures for profiling hospitals on bariatric surgery performance. JAMA Surg. 2014;149(1):10-6.
36. Lingsma HF, Bottle A, Middleton S, Kievit J, Steyerberg EW, Marang-van de Mheen PJ. Evaluation of hospital outcomes: the relation between length-of-stay, readmission, and mortality in a large international administrative database. BMC Health Serv Res. 2018;18(1):116.
37. Van Schie P, Van Bodegom-Vos L, Van Steenbergen LN, Nelissen R, Marang-van de Mheen PJ. A more comprehensive evaluation of quality of care after total hip and knee arthroplasty: combining 4 indicators in an ordered composite outcome. Acta Orthop. 2022;93:138-45.
38. Rolfson O, Kärrholm J, Dahlberg LE, Garellick G. Patient-reported outcomes in the Swedish Hip Arthroplasty Register: results of a nationwide prospective observational study. J Bone Joint Surg Br. 2011;93(7):867-75.

39. Dunbar MJ, Richardson G, Robertsson O. I can't get no satisfaction after my total knee replacement: rhymes and reasons. Bone Joint J. 2013;95-b(11 Suppl A):148-52.
40. Johnston B, Patrick D, Devji T, al e. Chapter 18: Patient_reported outcomes. Cochrane Handbook for Systematic Reviews of Interventions. 2019.
41. Makhni EC. Meaningful Clinical Applications of Patient-Reported Outcome Measures in Orthopaedics. J Bone Joint Surg Am. 2021;103(1):84-91.
42. FDA. Guidance for Indusctry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. 2009.
43. Black N. Patient reported outcome measures could help transform healthcare. BMJ. 2013;346:f167.
44. Rolfson O, Eresian Chenok K, Bohm E, Lübbeke A, Denissen G, Dunn J, et al. Patient-reported outcome measures in arthroplasty registries. Acta Orthop. 2016;87 Suppl 1(Suppl 1):3-8.
45. Bohm ER, Kirby S, Trepman E, Hallstrom BR, Rolfson O, Wilkinson JM, et al. Collection and Reporting of Patient-reported Outcome Measures in Arthroplasty Registries: Multinational Survey and Recommendations. Clin Orthop Relat Res. 2021;479(10):2151-66.
46. Hutchings A, Grosse Frie K, Neuburger J, van der Meulen J, Black N. Late response to patient-reported outcome questionnaires after surgery was associated with worse outcome. J Clin Epidemiol. 2013;66(2):218-25.
47. Jahagirdar D, Kroll T, Ritchie K, Wyke S. Patient-reported outcome measures for chronic obstructive pulmonary disease : the exclusion of people with low literacy skills and learning disabilities. Patient. 2013;6(1):11-21.
48. Gibbons E. Patient-reported outcome measures and the evaluation of services. Health Serv Deliv Res. 2016;4:55-68.
49. Tuti T, Nzinga J, Njoroge M, Brown B, Peek N, English M, et al. A systematic review of electronic audit and feedback: intervention effectiveness and use of behaviour change theory. Implement Sci. 2017;12(1):61.
50. Weiss D, Dunn SI, Sprague AE, Fell DB, Grimshaw JM, Darling E, et al. Effect of a population-level performance dashboard intervention on maternal-newborn outcomes: an interrupted time series study. BMJ quality & safety. 2018;27(6):425-36.
51. Shojania KG, Ranji SR, McDonald KM, Grimshaw JM, Sundaram V, Rushakoff RJ, et al. Effects of quality improvement strategies for type 2 diabetes on glycemic control: a meta-regression analysis. JAMA. 2006;296(4):427-40.
52. Tricco AC, Ivers NM, Grimshaw JM, Moher D, Turner L, Galipeau J, et al. Effectiveness of quality improvement strategies on the management of diabetes: a systematic review and meta-analysis. Lancet. 2012;379(9833):2252-61.
53. DR I, CD F, MS T. Consequences of individual feedback on behaviour in organisations. J Appl Psychol. 1979:349-71.
54. Hysong SJ, Best RG, Pugh JA. Audit and feedback and clinical practice guideline adherence: making feedback actionable. Implement Sci. 2006;1:9.
55. Kluger AN, Van Dijk D. Feedback, the various tasks of the doctor, and the feedforward alternative. Med Educ. 2010;44(12):1166-74.
56. Gude WT, Roos-Blom MJ, van der Veer SN, Dongelmans DA, de Jonge E, Francis JJ, et al. Health professionals' perceptions about their clinical performance and the influence of audit and feedback on their intentions to improve practice: a theory-based study in Dutch intensive care units. Implement Sci. 2018;13(1):33.

57. Landis-Lewis Z, Douglas GP, Hochheiser H, Kam M, Gadabu O, Bwanali M, et al. Computer-Supported Feedback Message Tailoring for Healthcare Providers in Malawi: Proof-of-Concept. AMIA Annu Symp Proc. 2015;2015:814-23.
58. AN K, A D. The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory. Psychol Bull. 1996(119):254-84.
59. RE J, M H, C-H C. The importance of velocity, or why speed may matter more than distance. Organ Psychol Rev. 2013(3):62-85.
60. Gude WT, Brown B, van der Veer SN, Colquhoun HL, Ivers NM, Brehaut JC, et al. Clinical performance comparators in audit and feedback: a review of theory and evidence. Implement Sci. 2019;14(1):39.
61. Gude WT, van Engen-Verheul MM, van der Veer SN, de Keizer NF, Peek N. How does audit and feedback influence intentions of health professionals to improve practice? A laboratory experiment and field study in cardiac rehabilitation. BMJ quality & safety. 2017;26(4):279-87.
62. Locke EA, Latham GP. Building a practically useful theory of goal setting and task motivation. A 35-year odyssey. Am Psychol. 2002;57(9):705-17.
63. Carver CS, Scheier MF. Control theory: a useful conceptual framework for personality-social, clinical, and health psychology. Psychol Bull. 1982;92(1):111-35.
64. GT D. There's a S.M.A.R.T. way to write management's goals and objectives. Manage Rev. 1981:35-6.
65. de Vos ML, van der Veer SN, Wouterse B, Graafmans WC, Peek N, de Keizer NF, et al. A multifaceted feedback strategy alone does not improve the adherence to organizational guideline-based standards: a cluster randomized trial in intensive care. Implement Sci. 2015;10:95.
66. Ivers NM, Grimshaw JM, Jamtvedt G, Flottorp S, O'Brien MA, French SD, et al. Growing literature, stagnant science? Systematic review, meta-regression and cumulative analysis of audit and feedback interventions in health care. J Gen Intern Med. 2014;29(11):1534-41.
67. ClinicalTrial.gov. https://clinicaltrials.gov/ct2/show/NCT04055103?term=Arthroplasty&cntry=NL&city=Leiden&draw=2&rank=2.
68. Wolfstadt JI, Cohen-Rosenblum A. 'You can't do quality between surgical cases and tea time': barriers to surgeon engagement in quality improvement. BMJ quality & safety. 2022:bmjqs-2022-015083.
69. Glasgow JM, Davies ML, Kaboli PJ. Findings from a national improvement collaborative: are improvements sustained? BMJ quality & safety. 2012;21(8):663-9.
70. Bridges J, May C, Fuller A, Griffiths P, Wigley W, Gould L, et al. Optimising impact and sustainability: a qualitative process evaluation of a complex intervention targeted at compassionate care. BMJ quality & safety. 2017;26(12):970-7.
71. Lipitz-Snyderman A, Needham DM, Colantuoni E, Goeschel CA, Marsteller JA, Thompson DA, et al. The ability of intensive care units to maintain zero central line-associated bloodstream infections. Arch Intern Med. 2011;171(9):856-8.
72. Pronovost PJ, Watson SR, Goeschel CA, Hyzy RC, Berenholtz SM. Sustaining Reductions in Central Line-Associated Bloodstream Infections in Michigan Intensive Care Units: A 10-Year Analysis. Am J Med Qual. 2016;31(3):197-202.
73. Burke RE, Marang-van de Mheen PJ. Sustaining quality improvement efforts: emerging principles and practice. BMJ quality & safety. 2021.
74. Greene SM, Reid RJ, Larson EB. Implementing the learning health system: from concept to action. Ann Intern Med. 2012;157(3):207-10.
75. Kilbourne AM, Glasgow RE, Chambers DA. What Can Implementation Science Do for You? Key Success Stories from the Field. J Gen Intern Med. 2020;35(Suppl 2):783-7.

76. Jackson GL, Cutrona SL, White BS, Reardon CM, Orvek E, Nevedal AL, et al. Merging Implementation Practice and Science to Scale Up Promising Practices: The Veterans Health Administration (VHA) Diffusion of Excellence (DoE) Program. Jt Comm J Qual Patient Saf. 2021;47(4):217-27.
77. Trbovich P, Shojania KG. Root-cause analysis: swatting at mosquitoes versus draining the swamp. BMJ quality & safety. 2017;26(5):350-3.
78. van Schie P, van Bodegom-Vos L, van Steenbergen LN, Nelissen R, Marang-van de Mheen PJ. Monitoring Hospital Performance with Statistical Process Control After Total Hip and Knee Arthroplasty: A Study to Determine How Much Earlier Worsening Performance Can Be Detected. J Bone Joint Surg Am. 2020;102(23):2087-94.
79. Sales A. Reporting on implementation trials with null findings: the need for concurrent process evaluation reporting. BMJ quality & safety. 2022.
80. Eccles MP, Foy R, Sales A, Wensing M, Mittman B. Implementation Science six years on--our evolving scope and common reasons for rejection without review. Implement Sci. 2012;7:71.
81. Desveaux L, Ivers NM, Devotta K, Ramji N, Weyman K, Kiran T. Unpacking the intention to action gap: a qualitative study understanding how physicians engage with audit and feedback. Implement Sci. 2021;16(1):19.
82. Watling C, Driessen E, van der Vleuten CP, Vanstone M, Lingard L. Beyond individualism: professional culture and its influence on feedback. Med Educ. 2013;47(6):585-94.
83. Leis JA, Shojania KG. A primer on PDSA: executing plan-do-study-act cycles in practice, not just in name. BMJ quality & safety. 2017;26(7):572-7.
84. Boggan J. Tailoring audit and feedback - a precision intervention but with added complexity? BMJ quality & safety. 2022.
85. López-López JA, Humphriss RL, Beswick AD, Thom HHZ, Hunt LP, Burston A, et al. Choice of implant combinations in total hip replacement: systematic review and network meta-analysis. BMJ. 2017;359:j4651.
86. Deere KC, Whitehouse MR, Porter M, Blom AW, Sayers A. Assessing the non-inferiority of prosthesis constructs used in hip replacement using data from the National Joint Registry of England, Wales, Northern Ireland and the Isle of Man: a benchmarking study. BMJ Open. 2019;9(4):e026685.
87. Wyatt M, Frampton C, Whitehouse M, Deere K, Sayers A, Kieser D. Benchmarking total hip replacement constructs using noninferiority analysis: the New Zealand joint registry study. BMC Musculoskelet Disord. 2021;22(1):719.
88. Smith AJ, Dieppe P, Vernon K, Porter M, Blom AW. Failure rates of stemmed metal-on-metal hip replacements: analysis of data from the National Joint Registry of England and Wales. Lancet. 2012;379(9822):1199-204.
89. Clarke A, Pulikottil-Jacob R, Grove A, Freeman K, Mistry H, Tsertsvadze A, et al. Total hip replacement and surface replacement for the treatment of pain and disability resulting from end-stage arthritis of the hip (review of technology appraisal guidance 2 and 44): systematic review and economic evaluation. Health Technol Assess. 2015;19(10):1-668, vii-viii.
90. Smith AJ, Dieppe P, Howard PW, Blom AW. Failure rates of metal-on-metal hip resurfacings: analysis of data from the National Joint Registry for England and Wales. Lancet. 2012;380(9855):1759-66.
91. Orthopaedic Data Evaluation Panel (ODEP). ODEP Benchmarks. Available at: http://www.odep.org.uk/Benchmarks/ODEPBenchmarks.aspx.
92. Tsikandylakis G, Mohaddes M, Cnudde P, Eskelinen A, Karrholm J, Rolfson O. Head size in primary total hip arthroplasty. EFORT Open Rev. 2018;3(5):225-31.

93. Paxton E, Cafri G, Havelin L, Stea S, Pallisó F, Graves S, et al. Risk of revision following total hip arthroplasty: metal-on-conventional polyethylene compared with metal-on-highly cross-linked polyethylene bearing surfaces: international results from six registries. J Bone Joint Surg Am. 2014;96 Suppl 1(Suppl 1):19-24.
94. Berry DJ, Harmsen WS, Cabanela ME, Morrey BF. Twenty-five-year survivorship of two thousand consecutive primary Charnley total hip replacements: factors affecting survivorship of acetabular and femoral components. J Bone Joint Surg Am. 2002;84(2):171-7.

Chapter 10

Appendices

Dutch summary (Nederlanse samenvatting)

Het doel van dit proefschrift was om te onderzoeken hoe feedback met betrekking tot totale heup- en knieprotheses (THP en TKP) waarbij gebruik wordt gemaakt van registers kan worden verbeterd. Daarnaast werd de effectiviteit van de verbeterde feedback getest door patiëntenuitkomsten te evalueren.

Protheseregisters werden opgericht om de overleving van implantaten te vergelijken en de veiligheid van verschillende orthopedische implantaten, zoals THP en TKP, te bewaken. De laatste jaren worden registers echter ook gebruikt als kwaliteitssystemen om de variatie tussen ziekenhuizen voor tal van klinische uitkomstmaten inzichtelijk te maken waardoor ziekenhuizen feedback krijgen over hun eigen prestaties. Het laatste wordt meestal gedaan in vergelijking met een referentiestandaard (d.w.z., de benchmark). Het geven van feedback over prestatie-uitkomsten is een veelgebruikte aanpak om de kwaliteit van zorg te verbeteren. In deze context wordt feedback gedefinieerd als het verstrekken van samenvattingen van klinische uitkomsten aan zorgverleners of organisaties met als doel verbeterinitiatieven te initiëren. Internationaal wordt feedback vanuit protheseregisters op verschillende manieren aangeboden. In Nederland worden klinische uitkomsten op ziekenhuisniveau getoond in een actueel dashboard van de Landelijke Registratie Orthopedische Interventies (LROI). Daarnaast wordt de mate van variatie in uitkomsten tussen ziekenhuizen getoond in een geanonimiseerde vorm in jaarlijkse rapporten. Uit de literatuur is gebleken dat de effectiviteit van feedback varieert (d.w.z., van een verslechtering van 9% tot een verbetering van 70%). Een optimale inrichting van de feedback zal dus redelijkerwijs leiden tot een verbetering van de zorg voor de patiënt. Dit proefschrift gaf een overzicht van de nationale en internationale variatie tussen ziekenhuizen in klinische uitkomsten om te onderzoeken of verbetering haalbaar is. Om de feedback inhoudelijk te verbeteren, zijn methodologische studies uitgevoerd om te onderzoeken of zogenaamde uitschieter-ziekenhuizen (zowel positief als negatief) eerder kunnen worden gedetecteerd, of redenen voor hogere revisiepercentages kunnen worden geïdentificeerd, en een samengestelde uitkomstmaat werd ontwikkeld en getest. De verkregen kennis uit deze studies in combinatie met hedendaagse theorieën voor het geven van effectieve feedback werden opgenomen in een veelzijdige kwaliteitsverbeterinterventie waarbij de effectiviteit werd onderzocht in een cluster gerandomiseerde en gecontroleerde trial.

Deze Nederlandse samenvatting beschrijft de belangrijkste bevindingen van dit proefschrift inclusief de praktische implicaties.

Samenvatting en praktische implicaties

Registers geven feedback bedoeld om kwaliteitverbeterinitiatieven in slechter presterende ziekenhuizen aan te jagen. Echter kunnen de best presterende ziekenhuizen ook geïnteresseerd zijn in het vergelijken van hun uitkomsten met ziekenhuizen uit andere landen om verder te verbeteren binnen specifieke domeinen. Eerlijke internationale vergelijking van ziekenhuizen is echter alleen haalbaar wanneer consistente uitkomstdefinities worden gebruikt, omdat deze in grote mate de frequentie bepalen. Consistentie in de definitie van uitkomstmaten maakt het ook mogelijk om internationale gegevens samen te voegen, waardoor betere mogelijkheden ontstaan om zeldzame veiligheidsproblemen eerder te detecteren zoals bijvoorbeeld de problematiek rondom metalen-op-metalen heupprothesen of modulaire femurhalscorrosie, wat duizenden patiënten zal beschermen tegen slecht presterende implantaten, waardoor onnodig lijden in de toekomst wordt verminderd. **Hoofdstuk 2** toonde aan dat revisie, heropname en complicaties de meest gerapporteerde klinische uitkomsten zijn in cohortstudies en rapporten van protheseregisters. Er zijn aanzienlijke verschillen in uitkomsten tussen ziekenhuizen, wat wijst op een groot verbeterpotentieel. Een deel van deze variatie kan echter worden verklaard door de aanzienlijke heterogeniteit in de volgende domeinen: 1) uitkomstdefinities, inclusief wat een revisie, heropname of complicatie inhoudt, 2) duur van follow-up en startpunt van follow-up, 3) kenmerken die zijn opgenomen in de correctie voor de patiëntenmix, en 4) type patiënten en ziekenhuizen die zijn meegenomen in de rapportages. Dit proefschrift toonde aan dat revisie van het implantaat binnen vijf jaar, heropname binnen 30 dagen en complicaties tot 2 jaar na de operatie de meest gebruikte uitkomstmaten waren. Geen van deze definities kwam echter perfect overeen met de andere domeinen voor THA, TKA en THA&TKA gecombineerd. De minste consensus werd gevonden voor welke patiëntkenmerken moet worden gecorrigeerd. Laatstgenoemde evenals de andere onderzochte domeinen zijn essentieel voor een eerlijke vergelijking tussen ziekenhuizen. In de toekomst kunnen samenwerkingen tussen protheseregisters zoals geïnitieerd door de International Society of Arthroplasty Registries, een leidende rol spelen in het streven naar meer uniformiteit in de gebruikte definities en methoden.

Het rapporteren van variatie tussen ziekenhuizen en het identificeren van positieve en negatieve uitschieter-ziekenhuizen wordt gedaan met behulp van funnelplots en is een eenvoudige en effectieve manier om inzicht te krijgen in de prestaties van ziekenhuizen, op voorwaarde dat deze analyses betrouwbaar zijn. De betrouwbaarheid kan worden uitgedrukt als het percentage van de totale variatie dat wordt verklaard door "echte" ziekenhuisverschillen versus toevalsvariatie. **Hoofdstuk 3** toonde een grote variatie in het 1-jaars revisiepercentage tussen Nederlandse ziekenhuizen, met een acceptabele betrouwbaarheid (61%) voor THP en een lage betrouwbaarheid (46%)

voor TKP in een periode van 3 jaar. Dit geeft aan dat er veel verbeteringspotentieel is voor een aanzienlijk aantal Nederlandse ziekenhuizen. Het vroegtijdig opsporen van slechte prestaties met behulp van een tijdspanne van 1 jaar heeft het voordeel dat kwaliteitverbeterinitiatieven eerder kunnen worden geïntroduceerd, maar dit leidde tot een lage betrouwbaarheid en wordt daarom niet aanbevolen. Het beoordelen hoe verbetering voor weinig gespecificeerde uitkomsten zoals 1-jaarsrevisie kan echter uitdagend zijn. Dit kan worden vergemakkelijkt door specifieke redenen voor revisie te onderzoeken die de onderliggende oorzaak kunnen zijn voor slechtere prestaties voor revisie. **Hoofdstuk** 3 toonde dat onderliggende redenen werden gevonden voor 12 van de 13 negatieve uitschieter-ziekenhuizen voor THP en 3 van de 7 voor TKP. De voornaamste onderliggende redenen voor slechtere prestaties waren infectie (zowel voor THP als TKP) en prothesedislocaties (alleen THP). Implant loslating (zowel voor THP als TKP) en technisch falen (alleen TKP) waren minder waarschijnlijk. De betrouwbaarheid voor de specifieke redenen voor revisie waren allemaal laag binnen een tijdspanne van 3 jaar voor THP en TKP, behalve voor infectie voor THP, waarvoor de betrouwbaarheid acceptabel was (d.w.z., 61%). Aangezien de betrouwbaarheid binnen een tijdspanne van 1 jaar allemaal laag waren, wordt aanbevolen om een periode van 3 jaar te gebruiken om de onderliggende redenen voor slechtere ziekenhuisprestaties op het gebied van revisie te identificeren.

In **hoofdstuk 3** werd aangetoond dat het gebruik van funnelplots voor het vroegtijdig opsporen van slechtere prestaties niet betrouwbaar was binnen een tijdspanne van 1 jaar, daarom werd een tijdspannen van 3 jaar aanbevolen. **Hoofdstuk 4** toonde echter aan dat maandelijks monitoren van 1-jaar revisiepercentages met behulp van CUSUM-grafieken met 5 controlelimieten eerder een betrouwbare achteruitgang of verbetering kan detecteren dan de conventionele funnelplots. Het eerste signaal voor negatieve afwijkingen werd gegenereerd na een mediane tijd van 18 maanden voor THP en 21 maanden voor TKP binnen een tijdspanne van 3 jaar. CUSUM-grafieken maken het daardoor mogelijk om eerder een verslechtering te detecteren, waardoor het mogelijk wordt om kwaliteitsverbeterinitiatieven eerder te introduceren dan te wachten op de resultaten in de funnelplot na 3 jaar. De nauwkeurigheid van deze CUSUM-grafieken in vergelijking met de traditionele funnelplot met een tijdspanne van 3 jaar was 97%. Deze resultaten zijn relevant voor protheseregisters en wetenschappelijke verenigingen in de besluitvorming of zij CUSUM-grafieken willen implementeren in hun organisatie. De resultaten over de nauwkeurigheid zullen bijdragen aan het vertrouwen van professionals in CUSUM-grafieken. Naar aanleiding van deze bevindingen heeft de LROI CUSUM-grafieken toegevoegd aan hun routinematige dashboardrapportage voor klinische uitkomsten.

Een andere manier om afwijkende prestaties eerder te kunnen detecteren is het vergrote van het aantal gebeurtenissen zoals bij samengestelde uitkomsten wordt gedaan. In **hoofdstuk 5** werd een geordende samengestelde uitkomst van klinische relevante uitkomsten (d.w.z., revisie, heropname, complicaties en verlengde opnameduur) voor THP en TKP ontwikkeld en getest op de mogelijkheid om nauwkeuriger en eerder tussen ziekenhuizen te differentiëren. Deze nieuwe uitkomst toonde een hogere mate van betrouwbaarheid dan de individuele uitkomsten vanwege de grotere variatie tussen ziekenhuizen wanneer de vier uitkomsten worden samengevoegd. De samengestelde uitkomst kon betrouwbaar verschillen tussen ziekenhuizen in hun prestaties differentiëren met een tijdspanne van 1 jaar, in plaats van de gebruikelijke 3 jaar, waardoor de introductie van kwaliteitverbeterinitiatieven eerder mogelijk wordt. Een bijkomend voordeel is dat de samengestelde uitkomst meerdere aspecten van de geleverde kwaliteit van zorg kan meten, omdat ziekenhuizen het op één uitkomst goed kunnen doen terwijl ze tegelijkertijd slechter kunnen presteren op een andere. De samengestelde uitkomst overwint dit probleem en laat zien of een patiënt een revisie heeft gehad, maar ook of ze opnieuw zijn opgenomen, complicaties hebben gehad of een verlengd opnameduur hebben gehad. Deze nieuwe samengestelde uitkomst kan helpen bij het selecteren van specifieke patiëntengroepen voor medische dossieronderzoek. In plaats van alle dossiers te bekijken voor patiënten die opnieuw zijn opgenomen, maakt de samengestelde uitkomst het mogelijk om selectief bijvoorbeeld alleen diegenen te bekijken die naast een heropname een normale opnameduur en geen complicatie hadden. Zo kan eenvoudiger worden bepaald of bijvoorbeeld informatie bij ontslag moet worden verbeterd om een heropname te voorkomen of dat een heropname mogelijk niet nodig zou zijn geweest indien een patiënt poliklinisch nauwlettender zou zijn vervolgd. Daarnaast is in één uitkomst zichtbaar of een focus om bijvoorbeeld de verblijfsduur te verkorten, niet ten koste gaat van een andere uitkomst zoals bijvoorbeeld heropname. Tot slot kunnen patiënten eenvoudig controleren hoe vaak een procedure verloopt zoals gepland (d.w.z., zonder dat een van de klinische uitkomsten zich voordoet).

Klinische uitkomsten zoals revisie, heropname, complicaties en opnameduur meten onbedoelde gebeurtenissen en komen bij THP en TKP over het algemeen voor met een lage frequentie. Echter 10% van de TKP- en 20% van de THP-patiënten zijn op de lange termijn ontevreden over het resultaat, voornamelijk vanwege aanhoudende pijn en functionele beperkingen. In tegenstelling tot klinische uitkomsten meten Patient-Reported Outcome Measures (PROMs) de bedoelde uitkomsten zoals pijnvermindering, verbetering van functionaliteit en gezondheid gerelateerde kwaliteit van leven en kunnen daarom leiden tot aanvullende verbeterpunten. Net als bij klinische uitkomsten is het van belang dat PROMs data een hoge mate van compleetheid hebben ter voorkomen van selectiebias. In Nederland en andere nationale en regionale

protheseregisters zijn responspercentages echter laag. Dit wil zeggen dat in Nederland 61% van de patiënten de preoperatieve PROM-vragenlijst heeft ingevuld en slechts 40% van de patiënten zowel de pre- als postoperatieve vragenlijst. In tegenstelling tot de compleetheid in registers voor klinische uitkomsten die boven de 98% bedraagt. Eerdere studies hebben al aangetoond dat patiënten die de vragenlijsten invullen over het algemeen gezonder zijn, vaker een witte huidskleur, minder vaak analfabeet en minder vaak cognitieve stoornissen hebben (inclusief dementie). Om beter te begrijpen of ontbrekende PROM-data leidt tot een onder- of overschatting van de postoperatieve PROM-verbeterscores werd in **hoofdstuk 6** het verschil in klinische uitkomsten (d.w.z. revisie, heropname, complicaties en opnameduur) voor respondenten en niet-respondenten beschreven. Ook werd de associatie met PROM-verbeterscores bepaald. PROM-respondenten hadden gunstigere klinische uitkomsten dan niet-respondenten voor THP en voor revisie en heropname voor TKP. THP-patiënten hadden minder kans op een klinisch relevante PROM-verbetering wanneer zij een revisie, complicaties of een verlengde opnameduur hadden met resultaten voor heropname in dezelfde richting maar niet significant. Vergelijkbare resultaten werden gevonden voor TKP. Doordat patiënten die PROM-vragenlijsten invulden over het algemeen gunstigere klinische uitkomsten hadden dan niet-respondenten werd een deel van de patiënten die een kleinere kans hadden op een klinisch relevante verbetering voor de PROMs gemist. PROM-verbeterscores worden daarom waarschijnlijk overschat. Gezien de beschreven associaties in **hoofdstuk 6**, is het waarschijnlijk dat initiatieven om de kwaliteit van zorg te verbeteren door heropname, complicaties en opnameduur voor zowel THP en TKP te verminderen, zullen leiden tot meer patiënten die een klinisch relevante PROM-verbeteringen. Het is onwaarschijnlijk dat ziekenhuisverschillen worden beïnvloed, aangezien de PROM-responspercentages van ziekenhuizen niet geassocieerd waren met klinische uitkomsten.

Feedback is alleen effectief wanneer deze wordt bekeken en correct wordt geïnterpreteerd door de doelgroep. **Hoofdstuk 7** laat zien dat slechts de helft (55%) van de orthopedisch chirurgen in Nederland die THP en TKP plaatsen weet of zij werkzaam zijn in een ziekenhuis die normaal presteert of een positieve of negatieve uitschieter. Het bewustzijn was hoger onder chirurgen die vaker inlogden op het LROI-dashboard, vaker funnelplots correct interpreteerden en vaker de 1-jaars revisiepercentages van hun vakgroep konden herinneren. Achtendertig procent van de THP- en 26% van de TKP-chirurgen voldeed aan deze drie voorwaarden die nodig zijn om actie te ondernemen op basis van de feedbackinformatie. Vijfenveertig procent van de chirurgen in een ziekenhuis dat als negatieve uitbijter was geïdentificeerd, meldde dat ze de verslechterende prestaties van hun ziekenhuis niet zagen aankomen. Dit betekent dat ze doorgingen met het verlenen van zorg zonder wijzigingen aan te brengen. Dit onderbouwt dat het van belang is om te focussen

op het effectiever maken van feedback, mede omdat 85% van de chirurgen aangaf dat er in hun ziekenhuis kwaliteitsverberterinitiatieven werden geïmplementeerd zodra ze werden geïdentificeerd als een slechter presterend ziekenhuis. De chirurgen gaven aan dat inloggen op het LROI-dashboard kan worden aangemoedigd door de inhoud aantrekkelijker te maken, bijvoorbeeld door niet alleen de nadruk te leggen op revisiepercentages, maar nieuwe uitkomsten toe te voegen zoals protheseoverleving, complicaties, heropnames en opnameduur. Deze nieuwe uitkomsten worden namelijk als relevant beschouwd door een groot deel van de chirurgen. Ook zou er onderwijsmateriaal beschikbaar moeten worden gesteld om interpretatievaardigheden van statistische grafieken zoals bijvoorbeeld de funnelplots en CUSUM beter te kunnen interpreteren. Een uitleggende tekst onder statistische grafieken zou ook kunnen helpen om chirurgen de helpen bij het interpreteren van de grafieken. Tot slot zou feedback in een kant en klaar model voor individuele ziekenhuizen (bijvoorbeeld in de vorm van een infographic) moeten worden toegestuurd via de email zodat chirurgen niet zelf selecties moeten maken in een online dashboard en feedback laagdrempelig toegankelijk is.

In **hoofdstuk 8** werden de bevindingen uit de voorgaande hoofdstukken gecombineerd met hedendaagse theorieën over effectieve feedback om een kwaliteitverbeterinterventie te ontwikkelen en de effectiviteit te testen in een cluster gerandomiseerde en gecontroleerde trial. De interventie werd gedurende acht maanden toegepast en omvatte maandelijkse feedback, interactieve educatie over de interpretatie van de feedback en een toolbox met daarin op bewijs gebaseerde kwaliteitverbeterinitiatieven voor alle uitkomsten die in de feedback werden gerapporteerd. Ziekenhuizen die de interventie ontvingen, verbeterden met 4,3% meer op de Textbook Outcome (dat wil zeggen het ontbreken van revisie, heropname, complicaties en een lange opnameduur) in vergelijking met de controlegroep. Het effect was groter bij interventieziekenhuizen die kwaliteitsverbeterinitiatieven introduceerde dan ziekenhuizen die dit niet deden, wat suggereert dat deze kwaliteitsverbeterinitiatieven die werden ondernomen waarschijnlijk de reden waren voor de betere resultaten. Deze bevindingen ondersteunen dat frequente feedback aan chirurgische teams moet worden aangevuld met interactieve educatie en toolbox met daarin kwaliteitverbeterinitiatieven die zijn afgestemd op specifieke uitkomsten om de kwaliteit van zorg met betrekking tot THP en TKP effectief te verbeteren.

Acknowledgment (Dankwoord)

Dit proefschrift is tot stand gekomen dankzij de (in)directe hulp van veel inspirerende mensen. Een aantal wil ik persoonlijk bedanken.

P.J. Marang-van de Mheen (copromotor).
Beste Perla, Jouw kritische blik, doelgerichtheid en oog voor detail zonder het overzicht te verliezen zijn eigenschappen die ik erg bewonder en hebben dit proefschrift tot een vele malen hoger niveau getild. Onze wekelijkse besprekingen kenmerkte zich door een leerzame omgeving vol positiviteit, waar we elkaar ook op persoonlijk vlak goed hebben leren kennen; bedankt voor de fantastische tijd.

R.G.H.H. Nelissen (promotor).
Beste Rob, Het is een talent hoe jij buiten de kaders kan denken en zo ogenschijnlijk gemakkelijk nieuwe onderzoekvragen bedenkt. Het was een genoegen om jouw enthousiasme voor de wetenschap en natuurlijk alles dat met Leiden te maken heeft van dichtbij mee te mogen maken.

Leti van Bodegom-Vos, dank voor de tijd en moeite die je hebt genomen voor het kritisch lezen van mijn artikelen en daarnaast natuurlijk de gezelligheid.

De Landelijke Registratie Orthopedische Interventies (LROI) en in het bijzonder Liza van Steenbergen wil ik bedanken voor de samenwerking in de periode dat wij maandelijks feedback gaven aan de verschillende ziekenhuizen. Ik besef mij maar al te goed dat ook jullie hier veel werk aan hebben gehad.

Collega's van de C2, we hebben een mooie tijd achter de rug! Tijdens de dagelijkse lunch werd een inhoudelijke discussie al snel onderbroken voor luchtig vertier. Wat ik erg waardeerde aan deze groep was dat eenieder altijd bereid was om een ander te helpen wanneer deze voor een lastige uitdaging stond. De legendarische borrels, (sport) activiteiten en als kers op de taart de skivakanties hebben daarnaast bijgedragen aan nieuwe vriendschappen. Geert en Saho wij hebben een bijzondere band opgebouwd en ik hoop dat we in de komende jaren nog veel mogen samenwerken.

Sandro (paranimf), Dit proefschrift bevat te weinig woorden om onze avonturen samen te vatten. Tranen in onze ogen van het lachen en openhartige gesprekken kenmerken onze vriendschap. Ik heb het gevoel dat je altijd naast me staat ook als je er niet bent. We kennen elkaar nu meer dan 20 jaar en daar zal het zeker niet bij blijven.

David (paranimf), Onze dagen begonnen vroeg in de bieb als studiegenoten en eindigde laat in een kroeg als jaargenoten. We zien elkaar niet wekelijks meer, maar de vriendschap houdt stand als een rots in de branding. Geweldig dat je mij wilt steunen tijdens de verdediging van dit proefschrift.

Lieve ouders, Jullie zijn mijn voorbeeld voor het creëren van een warm gezin en hoe je samen gelukkig kunt zijn. In ieder van mijn levensfasen dragen jullie met veel optimistische energie een steen(tje) bij. Jullie staan altijd voor ons klaar. Ik ben minstens zo trots op jullie als jullie wel eens zeggen dat jullie op mij zijn. Ik hou van jullie.

Andrea, de band die wij hebben is uniek en ik ben trots dat jij mijn lieve zus(je) bent. Je bewandelt altijd vol overgaven je eigen pad. Daar kan ik nog iets van leren. En bedankt voor het toevoegen van Tim en Ralf aan de familie:)

Where elsers, Er bestaat voor mij geen leukere dag dan een dag die ik van 's ochtends vroeg tot 's avonds laat met jullie kan doorbrengen. Ik kijk nu alweer uit naar oud en nieuw, de bootfeestjes, koningsdag, pride en natuurlijk ons jaarlijkse weekje in de bergen. Met eenieder vrienden voor het leven.

Kern 08, Roeivrienden? Dat is alweer lang geleden. Vijf van de acht mag ik tot mijn beste vrienden rekenen. Ik zorg altijd dat ik uitgerust ben als ik jullie tref, want iedere activiteit wordt tot een competitief evenement omgetoverd. Nu alweer zin in ons volgende avontuur.

Roos, mijn liefde en geluk. Dolblij ben ik dat jij 13 jaar geleden in mijn leven bent gekomen. Je maakt me gelukkig, haalt het beste in mij naar boven en geeft mij de ruimte om te ontwikkelen. Bedankt voor je oprechtheid, attentheid, reislust en grenzeloze liefde. Nu zijn we getrouwd en hebben we een geweldig zoon Roemer. Ik kan niet wachten op de tijd die ons nog te wachten staat. Daarnaast heb ik door jou er twee lieve schoonouders bij. Jacco en Ilona, bedank voor jullie gastvrijheid, rust, hulp en liefde.

Bibliography

Publications outside this thesis

S. Hasan, **P. van Schie**, B.L. Kaptein, J.W. Schoones, P.J. Marang-van de Mheen, R.G.H.H. Nelissen. *Biomarkers to discriminate between aseptic loosened and stable total hip and knee arthropasties: A systematic review.* Submitted.

T.J.N. van der Lelij & L.M. Gerritsen, E.R.A. van Arkel, R. Munnik-Hagewoud, R.G. Zuurmond, S. Keereweer, **P. van Schie**, P.B.A.A. van Driel. *The role of patient characteristics and the effects of angiogenic therapies on the microvasculature of the meniscus: A systematic review.* The Knee, 2022.

L.A. Hoogervorst, **P. van Schie**, J. Nagels, R.G.H.H. Nelissen, P.J. Marang-van de Mheen. *The reliability of revision rates following primary shoulder arthroplasty as quality indicator to rank hospital performance: a national registry analysis including 13,104 shoulders and 87 hospitals.* J Shoulder Elbow Surg, 2022.

Z. Popal, T. Schepers, **P van Schie**, G. Giannakopoulos, J. Halm. *The use of routine laboratory testing in acute trauma care: a retrospective analysis.* Tur J of Trauma and Emerg Surg, 2022.

L.M. Gerritsen & T.J.N. van der Lelij, **P. van Schie**, M. Fiocco, S. Keereweer, E.R.A. van Arkel, H.M. Hazelbag, P.B.A.A. van Driel. *Higher healing rate after meniscal repair with concomitant ACL reconstruction for tears located in vascular zone 1 compared to zone 2: a systematic review and meta-analysis.* Knee Surg Sport Traumatol Arthrosc, 2022.

P. van Schie, T.J.N. van der Lelij, M. Gerritsen, R.P.J. Meijer, E.R.A. van Arkel, M. Fiocco, J.W. Swen, A.L. Vahrmeijer, H.M. Hazelbag, S. Keereweer & P.B.A.A. van Driel. *Intra-operative assessment of the vascularisation of a cross section of the meniscus using near-infrared fluorescence imaging.* Knee Surg Sport Traumatol Arthrosc, 2021.

P. van Schie, K.E.M. Benders, M.P.J. van den Bekerom. *Letter to the Editor "Comparison Between Corticosteroids and Lidocaine Injection in the Treatment of Tennis Elbow: A Double-Blinded, RCT*. Am J Phys Med Rehabil, 2018.

P. Hoogervorst, **P. van Schie**, M.P.J. van den Bekerom. *Midshaft clavicle fractures: Current concepts*. EFORT Open Rev, 2018.

P. van Schie, W.M. Bakker. *Brace versus geen brace bij thoracolumbale wervelfracturen (AO type A) zonder neurologische verschijnselen*. Nederlandse Gipsverbandmeesters Magazine, 2018.

Presentations

P. van Schie. Interpretatie en toepassing: Gebruik van CUSUMs in de LROI. Nederlandse Orthopedische Vereniging (NOV) congress, Utrecht, 2023 **(oral presentation, invited speaker)**.

P. van Schie, L. van Bodegom-Vos, T. Zijdeman, R.G.H.H. Nelissen, P.J. Marang-van de Mheen, for the IQ Joint study group. *Effectiveness of a multifaceted quality improvement intervention to improve patient outcomes after total hip and knee arthroplasty: A registry nested cluster RCT.* Nordic Orthopaedic Federation, Vilnius, 2022 **(oral presentation, supported by the NOV because the study was rated as the best Dutch abstract of recent years)**

P. van Schie, L. van Bodegom-Vos, T. Zijdeman, R.G.H.H. Nelissen, P.J. Marang-van de Mheen, for the IQ Joint study group. *Effectiveness of a multifaceted quality improvement intervention to improve patient outcomes after total hip and knee arthroplasty: A registry nested cluster RCT.* European Federation of National Associations of Orthopaedics and Tramatology (EFORT) congress, Lissabon, 2021 **(oral presentation)**.

P. van Schie, L. van Bodegom-Vos, T. Zijdeman, R.G.H.H. Nelissen, P.J. Marang-van de Mheen, for the IQ Joint study group. *Does a quality improvement intervention result in better outcomes for total hip and knee arthroplasty? A registry nested cluster randomized controlled trial.* International Congress of Arthroplasty Registries (ISAR) congress, Copenhagen, 2021 **(oral presentation, selected top 5 abstracts)**.

P. van Schie, L. van Bodegom-Vos, L.N. van Steenbergen, R.G.H.H. Nelissen, P.J. Marang-van de Mheen. *A more comprehensive evaluation of quality of care after total hip and knee arthroplasty: Combining indicators in an ordered composite outcome.* International Congress of Arthroplasty Registries (ISAR) congress, Copenhagen, 2021 **(oral presentation)**.

P. van Schie, L. van Bodegom-Vos, L. van Steenbergen, T. Zijdeman, R.G.H.H. Nelissen, P.J. Marang-van de Mheen, for the IQ Joint study group. *Leidt een kwaliteitsinterventie tot betere uitkomsten en gerichte verbeterinitiatieven voor THP en TKP?* Nederlandse Orthopedische Vereniging (NOV) congress, Den Bosch, 2021 **(oral presentation, rewarded with the Van Rens Prijs)**.

P. van Schie, L. van Bodegom-Vos, L.N. van Steenbergen, R.G.H.H. Nelissen, P.J. Marang-van de Mheen. *Evaluation of Performance on Outcomes after Total Hip and Knee: A more comprehensive reflection of delivered care by combining indicators in a composite outcome.* World Arthroplasty Congress (WAC), Munich, 2021 **(oral presentation)**.

P. van Schie, T.J.N. van der Lelij, L.M. Gerritsen, M. Fiocco, E.R.A. van Arkel, J.W. Swen, H.M. Hazelbag, R.P.J. Meijer, A.L. Vahrmeijer, S. Keereweer, P.B.A.A. van Driel. *Intraoperative assessment of the vascularisation of a cross-section of the meniscus using near-infrared fluorescence imaging.* European Society of Sports Traumatology (ESSKA) congress, Milan, 2021 **(oral presentation)**.

P. van Schie, L. van Bodegom-Vos, R.G.H.H. Nelissen, P.J. Marang-van de Mheen. *Komt feedback van het LROI-dashboard over Totale Heup Prothesen aan bij de orthopedisch chirurg?* Nederlandse Orthopedische Vereniging (NOV) congress, Den Bosch, 2020 **(oral presentation)**.

P. van Schie, L. van Bodegom-Vos, L.N. van Steenbergen, R.G.H.H. Nelissen, P.J. Marang-van de Mheen. *Variation in revision rate after Total Hip Arthroplasty in the Netherlands: Can we direct quality improvement initiatives?* European Federation of National Associations of Orthopaedics and Traumatology (EFORT) congress, Lisbon, 2019 **(oral presentation)**.

P. van Schie, L. van Bodegom-Vos, L.N. van Steenbergen, R.G.H.H. Nelissen, P.J. Marang-van de Mheen. *Monitoring Implant Safety and Hospital Quality: Can we detect worse performance sooner using Statistical Process Control?* European Federation of National Associations of Orthopaedics and Traumatology (EFORT) congress, Lisbon, 2019 **(poster presentation, rewarded with the best poster presentation)**.

P. van Schie, K. Treskes, V.M. de Jong, J.S.K. Luitse, J.C. Goslings, T. Schepers. *Beperkte bijdrage routinematig laboratoriumonderzoek bij screening traumakamerpatienten.* Traumadagen congress, Amsterdam, 2015 **(oral presentation)**.

Curriculum vitae

Peter van Schie was born on March 3rd, 1988 in Leiden, The Netherlands and a son of Karel van Schie and Maartje Bakker. In 2006 he graduated from the Rijnlands Lyceum in Oegstgeest. Initially, he wanted to become an astronaut, but due to his height (6ft 7.1), this was not possible. He changed plans and started his studies in Biomedical Sciences at the University of Amsterdam (UvA). At the same time, he worked as a coordinator for organ and tissue donation at the Nederlandse Transplantatie Stichting (NTS). While at university, he lived in Amsterdam, where he met the love of his life, Roos Groeneveld. After obtaining his propaedeutic certificate, a dream came true, as he was admitted to medical school at the Vrije University (VU) of Amsterdam. In 2011, he obtained his Bachelor of Science (BSc) in medicine, and in 2016 his Master of Science (MSc). Peter completed his final year of medicine with internships in surgery and orthopaedic surgery at the Onze Lieve Vrouw Gasthuis (OLVG) in Amsterdam. During his studies, he rowed at the Amsterdamse Studenten Roeivereniging Nereus (ASR Nereus) and later in the Dutch national team. In addition to several national titles and podiums at world cups, he ended his rowing career with a fifth place at the 2016 Olympic games in Rio de Janeiro.
After completing his studies, he started as a non-training resident in orthopaedic surgery at the Onze Lieve Vrouw Gasthuis (OLVG) in Amsterdam (supervisor; Prof. dr. R.W. Poolman). After that, he was employed for another five months as a non-training resident at the Leiden University Medical Center (LUMC) orthopaedics department in Leiden (supervisor; prof. dr. R.G.H.H. Nelissen). In 2018 he started a PhD research project at the Department of orthopaedics and biomedical data sciences at the LUMC under the supervision of dr P.J. Marang-van de Mheen and prof. dr. R.G.H.H. Nelissen. The project team received a grant from the Van Rens Foundation to perform their studies. The results of this research project are reported in this thesis. During his PhD, research results were presented at several (inter)national conferences, and four prizes were won. In addition, several statistical and epidemiological courses were completed, and education was given to medical students in both their bachelor's and master's phases. In 2019 he married Roos, and in 2021 they had a son named Roemer. During the corona epidemic, he worked in the COVID-19 cohort department of the LUMC. He has been a member of the Dutch Orthopaedic Association since 2021.

After being admitted to the orthopaedic surgery residency in May 2021, he started as a clinical resident at the Alrijne in Leiderdorp surgery department in January 2022 (supervisor; dr. A.M. Zeillemaker). This is the first part of the six-year trajectory to becoming an orthopaedic surgeon. In November 2023, his training will continue at the Reinier Haga Orthopaedic Center (RHOC), Haaglanden Medical Center (HMC)

and LUMC. In the future, he will continue to work towards his goal of becoming an outstanding orthopaedic surgeon.