EFORT open reviews

GENERAL ORTHOPAEDICS | 7:4 | 247–263

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] and 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, The Netherlands

Correspondence should be addressed to P van Schie
Email
p.van_schie@lumc.nl

- In order 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 enable early 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.
- PubMed, Embase, Web of Science, Cochrane library, Emcare, and Academic Search Premier were searched from January 2009 to August 2020 for studies reporting on: (i) primary THA/TKA; (ii) revision, readmission, or complications; and (iii) between-hospital variation. Most recent registry reports of Network of Orthopedic Registries of Europe members were also reviewed. Two reviewers independently screened records, extracted data, and assessed the risk of bias using the Integrated quality Criteria for the Review Of Multiple Study designs tool for studies and relevant domains for registries. We assessed agreement for the following domains: (i) outcome definition; (ii) follow-up and starting point; (iii) case-mix adjustment; and (iv) type of patients and hospitals included.
- 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% for THA and TKA, respectively; and for complications, 0–75% and 0–50% for THA and TKA, respectively. Indicator definitions and methodological variables varied considerably across domains.
- 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.

Keywords

- systematic review
- total hip arthroplasty
- total knee arthroplasty
- between-hospital variation
- performance indicators
- revision
- readmission
- complications

EFORT Open Reviews
(2022) **7**, 247–263

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, 2, 3, 4, 5, 6, 7, 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, 2, 3, 4, 5, 6, 7, 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, 10, 11, 12, 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, 15, 16, 17, 18, 19). However, the reliability of hospital rankings has been shown to be affected by for example, 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 the 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).

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: (i) primary THA/TKA; (ii) revision, readmission, complication, length-of-stay (LOS), and mortality; (iii) between-hospital variation (Supplementary data I, see section on supplementary materials given at the end of this article). 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 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 were resolved through discussion. Senior researchers (PM/RN) were available if consensus could not be reached. Inclusion criteria were studies reporting on (i) primary THA and/or TKA; (ii) national or regional between-hospital variation for revision, readmission, complication, LOS, or mortality with at least two 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 was included. The between-hospital variations as reported for the outcomes were collected in the original unit, including mean, S.D.,S.E., 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 was 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 (S.D.), 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: (i) outcome definition (i.e. what constitutes a revision, readmission, or complications); (ii) follow-up and starting point (e.g. post-discharge or post-surgery); (iii) case-mix adjustment (yes/no); and (iv) 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 and 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 S.E. were available, we calculated the 95% CI. If only the S.D. was available, the S.E. was calculated by dividing the S.D. 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 1). 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 2). 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 1643 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 (Fig. 1).

Risk of bias

Thirty-two cohort studies and 1 controlled before-and-after study were included. One study was classified as a HQ

Table 1 Risk of bias (RoB) studies.

Study number	Reference	Year	Design	1A*	2D*	2E*	3B*	3C*	3E	3F	3G*	4C*	5A	5B	F3	7A	7B	7C	7D	7E	ICROMS score
1	Bozic *et al.* (14)†	2014	CS	**2**		**2**			0	2	**2**	**2**		1	2	2	2	2	1	2	22
2	Thirukumaran *et al.* (15)	2020	CBA	**2**	**2**		**2**	**1**	0	2		**0**	2		2	2	2	2	2	0	21
3	Courtney *et al.* (38)	2018	CS	**2**		**1**			0	2	**2**	**2**		1	2	1	2	2	1	2	20
4	van Schie *et al.* (11)	2020	CS	**2**		**1**			0	2	**2**	**0**		1	2	2	2	2	1	2	19
5	Graham *et al.* (21)	2019	CS	**2**		**1**			0	2	**2**	**0**		1	2	2	2	2	1	2	19
6	Sheetz *et al.* (9)	2019	CS	**2**		**1**			0	2	**2**	**0**		1	2	2	2	2	1	2	19
7	Bottle *et al.* (35)	2018	CS	**1**		**1**			0	1	**2**	**2**		1	2	2	2	2	1	2	19
8	Padegimas *et al.* (33)	2018	CS	**2**		**1**			0	2	**2**	**0**		1	2	2	2	2	1	2	19
9	van de Mheen *et al.* (49)	2017	CS	**2**		**1**			0	2	**2**	**0**		1	2	2	2	2	1	2	19
10	Hollis *et al.* (39)	2017	CS	**2**		**1**			0	2	**2**	**0**		1	2	2	2	2	1	2	19
11	Qian *et al.* (41)	2013	CS	**2**		**1**			0	2	**2**	**0**		1	2	2	2	2	1	2	19
12	Voorn *et al.* (10)	2017	CS	**2**		**1**			0	1	**2**	**0**		1	2	2	2	2	1	2	18
13	Cram *et al.* (42)	2012	CS	**2**		**1**			0	2	**2**	**0**		1	1	2	2	2	1	2	18
14	Cai *et al.* (48)	2012	CS	**2**		**1**			0	1	**2**	**0**		1	2	2	2	2	1	2	18
15	Padegimas *et al.* (30)	2017	CS	**2**		**1**			0	2	**2**	**0**		1	2	2	2	2	1	0	17
16	Chen *et al.* (51)	2017	CS	**2**		**1**			0	2	**2**	**0**		1	2	2	2	2	1	0	17
17	Courtney *et al.* (16)	2017	CS	**2**		**0**			0	1	**2**	**1**		1	1	2	2	2	1	2	17
18	Husni *et al.* (53)	2010	CS	**2**		**0**			0	2	**1**	**2**		1	1	2	2	1	1	2	17
19	Hofstede *et al.* (40)	2018	CS	**2**		**1**			0	1	**2**	**0**		1	2	2	2	2	1	0	16
20	Pross *et al.* (27)	2017	CS	**2**		**1**			0	1	**2**	**0**		1	2	2	2	2	1	0	16
21	Calderwood *et al.* (43)	2013	CS	**2**		**1**			0	2	**2**	**2**		1	1	2	0	2	1	0	16
22	Kurtz *et al.* (37)	2016	CS	**2**		**1**			0	1	**1**	**0**		1	2	2	2	2	1	0	15
23	Kurtz *et al.* (36)	2016	CS	**2**		**1**			0	1	**1**	**0**		1	2	2	2	2	1	0	15
24	Jergesen & Yi (47)	2016	CS	**2**		**1**			0	1	**1**	**0**		1	1	2	2	1	1	2	15
25	Makela *et al.* (31)	2011	CS	**2**		**1**			0	1	**2**	**2**		1	2	2	0	1	1	0	15
26	Mittal *et al.* (28)	2018	CS	**2**		**1**			0	1	**1**	**0**		1	1	2	2	2	1	0	14
27	Skufca *et al.* (44)	2017	CS	**2**		**1**			0	1	**2**	**0**		1	0	2	2	2	1	0	14
28	Asaid *et al.* (50)	2013	CS	**2**		**0**			0	2	**1**	**0**		1	1	2	2	2	1	0	14
29	Dailey *et al.* (45)	2009	CS	**2**		**1**			0	2	**2**	**0**		1	1	2	0	2	1	0	14
30	Martino *et al.* (52)	2018	CS	**2**		**1**			0	1	**2**	**0**		1	1	2	0	1	0	2	13
31	Singh *et al.* (32)	2017	CS	**2**		**1**			0	1	**1**	**0**		1	0	2	2	1	1	0	12
32	Husted *et al.* (29)	2010	CS	**2**		**0**			0	2	**1**	**0**		1	1	2	0	2	1	0	12
33	Lopez-Contreras *et al.* (46)	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. †This study has an ICROMS score ≥18 and fulfils the mandatory criteria and was therefore classified as low RoB/high quality. The studies with an ICROMS ≥ 18 points did not fulfil the mandatory criteria and were therefore classified as moderate RoB/moderate quality. The studies with an ICROMS < 18 points did 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 2 Risk of bias (RoB) in 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, that is, 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, did not fulfil the criteria; 1, unclear if criteria are fulfilled; 2, 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.

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 1, domain 2E) and incomplete outcome data addressed (Table 1, 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 2, domain 5)

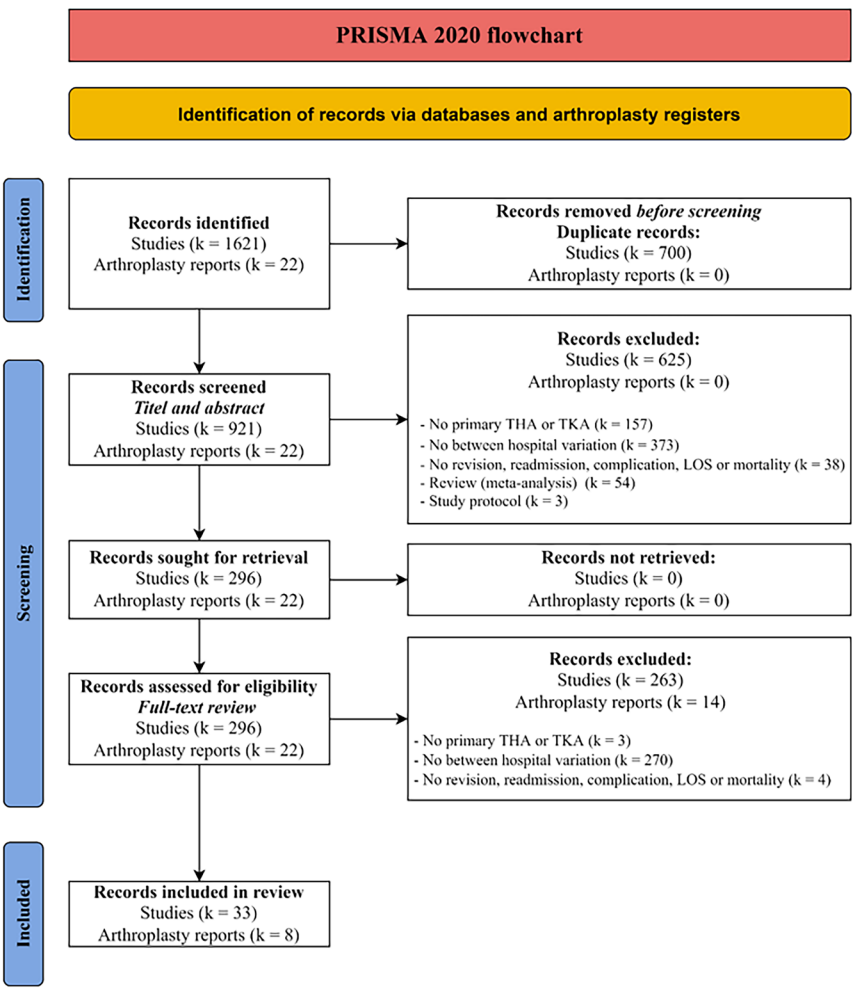

Figure 1
PRISMA 2020 flowchart. THA, total hip arthroplasty; TKA, total knee arthroplasty.

and whether missing values for covariates were imputed (Table 2, domain 7).

Characteristics

Seven of the 33 studies included THA, 4 TKA, 12 both THA and TKA, and 10 studies combined THA and TKA as one group (Supplementary data 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 and TKA combined. Four studies did not report the number of patients included (15, 27, 28, 29). A wide range was found for the number of hospitals included for THA, TKA, and THA and TKA combined (i.e. 2–3479). 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 (Supplementary data IIA).

Two of the eight registry reports included THA, two TKA, and four both THA and TKA. All reports included between-hospital variation for revision, two for readmission, and three for complications. Reports showed smaller variation in the number of patients (i.e. 7161–33248) and hospitals included (i.e. 47–152) 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 seven reports (Supplementary data IIB).

Between-hospitals variation and indicator definitions

Revision

From the six studies reporting revision rates, the between-hospital variation was reported in five studies (1 MQ, 4 LQ) for THA (11, 27, 30, 31, 32) and four studies (2 MQ, 2 LQ) for TKA (11, 31, 32, 33). Notable differences were seen with regard to the extent of between-hospital variation across studies, as shown in Fig. 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 3). Revision within 1 year was mainly reported (33% of studies reporting revision), but even then, the variation remained large (Fig. 2).

Revisions were reported in all eight registry reports, but the between-hospital variation was reported in six reports for both THA (2, 4, 5, 7, 8, 34) and TKA (2, 3, 6, 7, 8, 34). Again, there were notable differences in the between-hospital variation across registries (Fig. 2). As for the included studies, we found a large variety in definitions used, both in what constituted a revision, the follow-up,

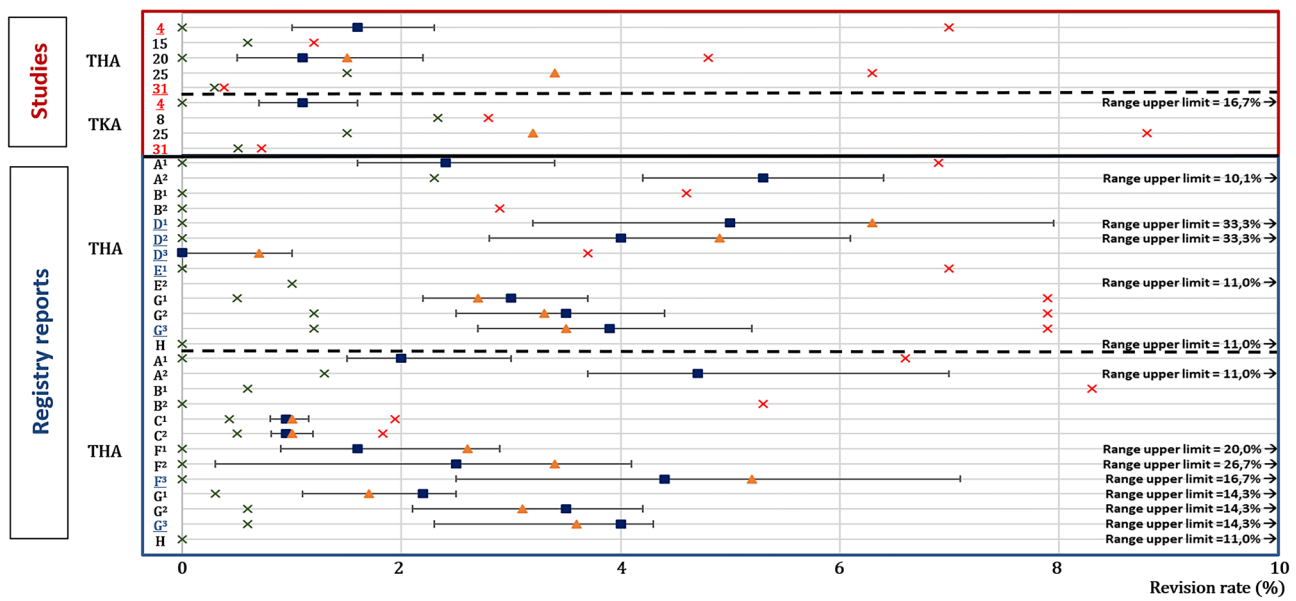

Figure 2

Between-hospital variation for revision. The numbers and the letters on the y-axis correspond with the study numbers from Table 1A and report letters from Table 1B, respectively. A letter in superscript was added to a study number, or a number in superscript was added to a report letter 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 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. THA, total hip arthroplasty; TKA, total knee arthroplasty.

type of patients and hospitals selected, and adjustment for case-mix (Table 3). 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.

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, 30, 35, 36), 4 studies (3 MQ, 1 LQ) for TKA (21, 33, 35, 37), and 6 studies (3 MQ, 3 LQ) for THA and TKA combined (15, 16, 28, 38, 39, 40). 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 4). 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 4). Studies combining THA and 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 two registry reports (25% of reports), but the between-hospital variation was only given in one report with three different patient selections (i.e. all patients, only with osteoarthritis or with a fracture) for THA (4) and in one report for TKA (6) (Fig. 3 and Table 4). 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, 41, 42, 43, 44, 45, 46, 47), 8 studies (4 MQ, 4 LQ) for TKA (10, 21, 33, 44, 45, 46, 47, 48) and 8 studies (1 HQ, 3 MQ, 4 LQ) for THA and TKA combined (14, 15, 16, 32, 38, 47, 49, 50). Eight studies reported the variation more than once with different outcome definition, follow-up, and type of hospitals selected (Table 5). Again, large differences were found in the between-hospital variation, which is (at least) partly explained by the different definitions used (Fig. 4 and Table 5). Two studies used the same dataset and reported comparable in between-hospital variations (16, 38). 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).

Complications were reported in three reports (38% of reports), with between-hospital variation reported in two reports for THA (4, 5) and one report for TKA (3). All reports reported the variation more than once with different outcome definitions, follow-ups, and types of patients selected (Table 5). 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 (Fig. 4 and Table 5).

Length-of-stay and mortality

Nine studies (27% of studies) (10, 29, 30, 33, 40, 49, 51, 52, 53) reported the between-hospital variation for LOS and only one study (3% of studies) (9) represented mortality. Between-hospital variation for LOS was given by one report (13% of reports) (6) and for mortality by two reports (25% of reports) (3, 5) (Supplementary datas III, IV, V and VI).

Perfect agreement

Given the heterogeneity in definitions used across studies and registry reports, none of the outcomes had perfect agreement across all six 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 and TKA combined (Tables 3, 4 and 5).

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 3, 4 and 5). 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, 11), diagnosis (osteoarthritis versus other) (2, 11), BMI (8, 11), Charnley score (8, 11), smoking status (11), use of patellar button (3), previous contralateral arthroplasty (31), bilaterally of the operation (31), heart disease (31), hypertension (31), cancer (31), alcoholism (31), dementia (31), depression (31), Parkinson's disease (31), mental disorders (31), degenerative brain diseases (31), and atherosclerosis (31).

For readmissions, also wide variation in case-mix adjustments: for age (9, 15, 21, 38), 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

Table 3 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 is 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)
One or more of the components are exchanged, removed, or added, including arthrodesis or amputation	–	–	–	(C^1,C^2)
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)
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 six domains. The numbers in brackets correspond to the study number from Table 1 and the letters in brackets correspond to the report codes from Table 2.
No, negative outlier; Po, positive outlier; THA, total hip arthroplasty; TKA, total knee arthroplasty.

Table 4 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)	–
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 honour 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 six domains. The numbers in brackets correspond to the study number from Table 1 and the letters in brackets correspond to the report codes from Table 2.
Fu, follow-up; LOS, length-of-stay; No, negative outlier; Po, positive outlier; THA, total hip arthroplasty; TKA, total knee arthroplasty.

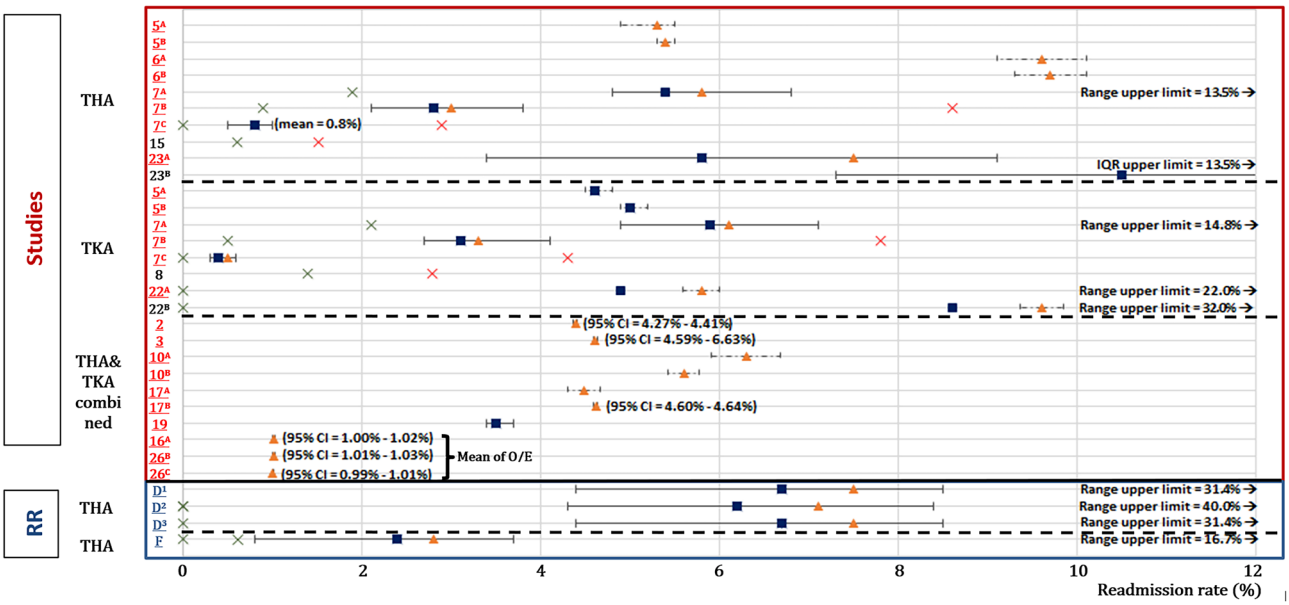

Figure 3

Between-hospital variation for readmission. The numbers and letters on the y-axis correspond to the study numbers from Table 1A and report letters from Table 1B, respectively. A letter in superscript was added to a study number or a number in superscript to a report letter when the readmission rate was reported more than once with different definitions. Study numbers and report letters were underlined and red or 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 represents the mean. The interquartile range is shown in a solid line through the median. The 95% CI is shown with a dashed line through the mean. RR, Registry reports; THA, total hip arthroplasty; TKA, total knee arthroplasty.

mellitus (21), surgical time (21), work relative value unit (21), emergency surgery (21), patient comorbidities (16, 28, 38), Elixhauser comorbidities (9) procedure (THA/TKA) (16), demographics (39), healthcare use (39), comorbidities selected by veteran affair surgical quality improvement programme (VASQIP) nurses (39), and clinical comorbidity (15).

For complications, between-hospital variations were case-mix-adjusted for age (9, 10, 14, 15, 16, 21, 27, 41, 42, 48, 50), gender (9, 10, 14, 15, 21, 27, 41, 42, 48, 50), ethnicity (21, 41, 42), Elixhauser comorbidities (9, 41, 48), patient comorbidities (14, 16, 27, 38, 42), ASA score (10, 21), procedure (THA or TKA) when THA and TKA are combined (14, 16, 50), payer (41), admission status (41), 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, 11, 21, 31, 32, 41, 44, 45, 46), but also to assess variation by structural hospital characteristics (e.g. ownership structure or teaching status) (28, 30, 33, 47, 50, 51, 52), to assess outcome associations between specialisms and hospitals (39), and to assess the impact of coding schemes (42). Regardless of the purpose of the studies, 19 studies (58% of studies) informed individual hospitals about their performance (i.e. 5 for revision (11, 30, 31, 32, 33), 6 for readmission (9, 21, 28, 30, 33, 39), and 12 for complications (9, 10, 21, 32, 33, 41, 42, 44, 45, 46, 47, 50)). The remaining 14 studies (42% of studies) reported the variation in one overall estimate (i.e. mean (S.D.), median (IQR), or range) from which hospitals are unable to infer how they are performing compared with other hospitals (14, 15, 16, 27, 29, 35, 36, 37, 38, 40, 43, 48, 49, 53). All registry reports gave outcomes at the individual hospital level when outcomes were collected (i.e., 8 for revision (2, 3, 4, 5, 6, 7, 8, 34), 2 for readmission (4, 6), 3 for complications (3, 4, 5), 1 for LOS (6), and 2 for mortality (3, 5)).

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

Table 5 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])	–	–	–	–
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)	–	–	–
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])
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])	–	–

(Continued)

Table 5 Continued.

	Studies (n = 20)			Registry reports (n = 3)		
Complications	THA (n = 16) (5,6[A],6[B],11[A],11[B],11[C],1 2,13,20,21,24,27[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])
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 honour 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 six domains. The numbers in brackets correspond to the study number from Table 1 and the letters in brackets correspond to the report codes from Table 2.
[A]National Quality Forum (NQF)-endorsed hospital-level risk-standardized complication rate developed by the Centres for Medicare and Medicaid Services; [B]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.

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 1 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 not only 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 (54, 55). The Australian Orthopaedic Association National Joint Replacement Registry (AOANJRR) identified these implants as having an outlier performance in 2007, more than 3 years before retraction from the market (56). In addition, the mortality risk increased by 8.5% (95% CI: 5.8–11.2%) due to these implants (57). 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 (58). 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 (59). 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 (60, 61). In this context, it is essential to distinguish between

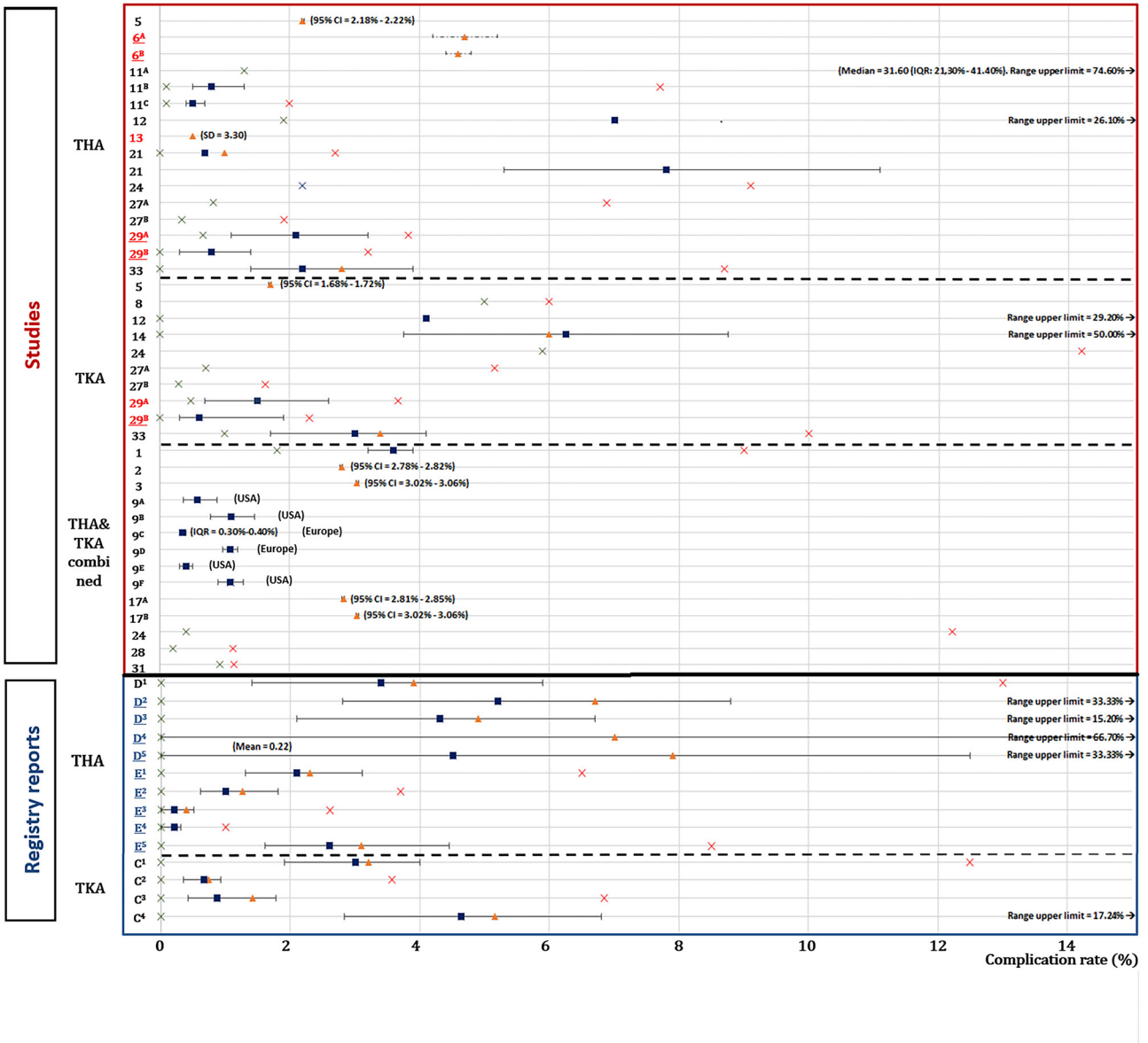

Figure 4

Between-hospital variation for complications. The numbers and letters on the y-axis correspond with the study numbers from Table 1A and report letters from Table 1B, respectively. A letter in superscript was added to a study number or a number in superscript to a report letter when the complication rate was reported more than once with different definitions. Study numbers were underlined and red-colored when complications within 30 days were reported, and report letters were underlined and blue-colored when complications within 2 years were 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% CI is shown with a dashed line through the mean. THA, total hip arthroplasty; TKA, total knee arthroplasty.

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 h imposed by health insurance policies, cooperation agreements with general practitioners, and other financial incentives.

Secondly, to allow for fair hospital comparison between hospitals, it is important to adjust for differences in case-mix (62). 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) (63, 64, 65). 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 three studies and one 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 (66). In addition, a Cochrane review showed a median absolute improvement of 4.3% associated with audit and feedback (IQR:0.5–16.0%) (67). 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) (67, 68, 69, 70). 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 eight studies (24% of studies), making it impossible to assess whether selection bias affected hospital outcomes and thus generalizability of our results (Supplementary data IIIA and 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. Secondly, 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. Thirdly, 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 four domains so that the agreement may have been higher for some domains (Tables 3, 4 and 5).

Conclusion

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

Supplementary materials

This is linked to the online version of the paper at https://doi.org/10.1530/EOR-21-0084.

ICMJE Conflict of Interest Statement

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of this review.

Funding Statement

The authors received a grant from the Van Rens Foundation (VRF2018-001) to perform this study.

References

1. Norwegian Arthroplasty Register (NAR). Annual report 2020. (available at: http://nrlweb.ihelse.net/eng/Rapporter/Report2020_english.pdf)

2. Dutch Arthroplasty Register (LROI). Online annual report. (available at: https://www.lroi-report.nl/)

3. Swedish Knee Arthroplasty Register (SKAR). Annual report 2020. (available at: http://myknee.se/pdf/SVK_2020_Eng_1.0.pdf)

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

5. Swedish Hip Arthroplasty Register (SHAR). Annual report 2018. (available at: 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. (available at: https://www.sundhed.dk/content/cms/99/4699_dkr-arsrapport-2020_offentliggorelse.pdf)

7. Finnish Arthroplasty Register (FAR). Online annual report. (available at: https://www.thl.fi/far/#index)

8. Swiss Arthroplasty Register (SAR). Annual report 2020. (available at: 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 Surgery* 2019 **154** 510–515. (https://doi.org/10.1001/jamasurg.2019.0090)

10. Voorn VMA, Marang-van de Mheen PJ, van der Hout A, So-Osman C, van den Akker-van Marle ME, Koopman-van Gemert AWMM, Dahan A, Vliet Vlieland TPM, Nelissen RGHH, van Bodegom-Vos L, *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** e014143. (https://doi.org/10.1136/bmjopen-2016-014143)

11. van Schie P, van Steenbergen LN, van Bodegom-Vos L, Nelissen RGHH & Marang-van de Mheen PJ. Between-hospital variation in revision rates after total hip and knee arthroplasty in the Netherlands: directing quality-improvement initiatives. *Journal of Bone and Joint Surgery: American Volume* 2020 **102** 315–324. (https://doi.org/10.2106/JBJS.19.00312)

12. van Schie P, van Bodegom-Vos L, van Steenbergen LN, Nelissen RGHH & 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. *Journal of Bone and Joint Surgery: American Volume* 2020 **102** 2087–2094. (https://doi.org/10.2106/JBJS.20.00005)

13. Ranstam J, Wagner P, Robertsson O & Lidgren L. Health-care quality registers: outcome-orientated ranking of hospitals is unreliable. *Journal of Bone and Joint Surgery: British Volume* 2008 **90** 1558–1561. (https://doi.org/10.1302/0301-620X.90B12.21172)

14. Bozic KJ, Grosso LM, Lin Z, Parzynski CS, Suter LG, Krumholz HM, Lieberman JR, Berry DJ, Bucholz R, Han L, *et al.* Variation in hospital-level risk-standardized complication rates following elective primary total hip and knee arthroplasty. *Journal of Bone and Joint Surgery: American Volume* 2014 **96** 640–647. (https://doi.org/10.2106/JBJS.L.01639)

15. Thirukumaran CP, McGarry BE, Glance LG, Ying M, Ricciardi BF, Cai X & Li Y. Impact of hospital readmissions reduction program penalties on hip and knee replacement readmissions: comparison of hospitals at risk of varying penalty amounts. *Journal of Bone and Joint Surgery: American Volume* 2020 **102** 60–67. (https://doi.org/10.2106/JBJS.18.01501)

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** 1888–1894. (https://doi.org/10.2106/JBJS.17.00203)

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 Orthopaedica* 2020 **92** 54–61. (https://doi.org/10.1080/17453674.2020.1827523)

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 and Safety* 2021 **30** 1010–1020. (https://doi.org/10.1136/bmjqs-2020-011247)

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** e112282. (https://doi.org/10.1371/journal.pone.0112282)

20. Veen EJ, Janssen-Heijnen ML, Leenen LP & Roukema JA. The registration of complications in surgery: a learning curve. *World Journal of Surgery* 2005 **29** 402–409. (https://doi.org/10.1007/s00268-004-7358-8)

21. Graham LA, Mull HJ, Wagner TH, Morris MS, Rosen AK, Richman JS, Whittle J, Burns E, Copeland LA, Itani KMF, *et al.* Comparison of a potential hospital quality metric with existing metrics for surgical quality-associated readmission. *JAMA Network Open* 2019 **2** e191313. (https://doi.org/10.1001/jamanetworkopen.2019.1313)

22. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffman TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, *et al.* The PRISMA 2020 statement: an upgrated guideline for reporting systematic reviews. *Systematic Reviews* 2021 **10** 89. (https://doi.org/10.1186/s13643-021-01626-4)

23. Network of orthopaedic registries of Europe (NORE). (available at: https://www.efort.org/about-us/nore/research/). Visited on 31 December 2020.

24. Weir CJ, Butcher I, Assi V, Lewis SC, Murray GD, Langhorne P & Brady MC. Dealing with missing standard deviation and mean values in meta-analysis of continuous outcomes: a systematic review. *BMC Medical Research Methodology* 2018 **18** 25. (https://doi.org/10.1186/s12874-018-0483-0)

25. Cochrane Handbook for Systematic Reviews. Version 6.2, 2021. (available at: https://training.cochrane.org/handbook/current)

26. Zingg W, Castro-Sanchez E, Secci FV, Edwards R, Drumright LN, Sevdalis N & Holmes AH. Innovative tools for quality assessment: integrated quality criteria for review of multiple study designs (ICROMS). *Public Health* 2016 **133** 19–37. (https://doi.org/10.1016/j.puhe.2015.10.012)

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** 842–852. (https://doi.org/10.1016/j.healthpol.2017.06.009)

28. Mittal M, Wang CE, Goben AH & Boyd AD. Proprietary management and higher readmission rates: a correlation. *PLoS ONE* 2018 **13** e0204272. (https://doi.org/10.1371/journal.pone.0204272)

29. Husted H, Hansen HC, Holm G, Bach-Dal C, Rud K, Andersen KL & Kehlet H. What determines length of stay after total hip and knee arthroplasty? A nationwide study in Denmark. *Archives of Orthopaedic and Trauma Surgery* 2010 **130** 263–268. (https://doi.org/10.1007/s00402-009-0940-7)

30. 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. *Journal of Arthroplasty* 2017 **32** 2347–2352. (https://doi.org/10.1016/j.arth.2017.03.032)

31. 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. *Annals of Medicine* 2011 **43** (Supplement 1) S31–S38. (https://doi.org/10.3109/07853890.2011.586362)

32. 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. *European Journal of Orthopaedic Surgery and Traumatology* 2017 **27** 261–265. (https://doi.org/10.1007/s00590-016-1852-1)

33. Padegimas EM, Kreitz TM, Zmistowski B, Teplitsky SL, Namdari S, Purtill JJ, Hozack WJ & Chen AF. Short-term outcomes of total knee arthroplasty performed at an Orthopedic Specialty Hospital. *Orthopedics* 2018 **41** e84–e91. (https://doi.org/10.3928/01477447-20171127-04)

34. The Norwegian National Arthroplasty Register. Annual Report 2020. (available at: http://nrlweb.ihelse.net/eng/Rapporter/Report2020_english.pdf)

35. 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. *Journal of Arthroplasty* 2018 **33** 2014.e2–2019.e2. (https://doi.org/10.1016/j.arth.2018.02.064)

36. 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. *Journal of Arthroplasty* 2016 **31** 2130–2138. (https://doi.org/10.1016/j.arth.2016.03.041)

37. 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? *Journal of Arthroplasty* 2016 **31** 2099–2107. (https://doi.org/10.1016/j.arth.2016.03.045)

38. Courtney PM, Frisch NB, Bohl DD & Della Valle CJ. Improving value in total hip and knee arthroplasty: the role of high volume hospitals. *Journal of Arthroplasty* 2018 **33** 1–5. (https://doi.org/10.1016/j.arth.2017.07.040)

39. Hollis RH, Graham LA, Richman JS, Morris MS, Mull HJ, Wahl TS, Burns E, Copeland LA, Telford GL, Rosen AK, *et al.* Hospital readmissions after surgery: how important are hospital and specialty factors? *Journal of the American College of Surgeons* 2017 **224** 515–523. (https://doi.org/10.1016/j.jamcollsurg.2016.12.034)

40. 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 and Safety* 2019 **28** 94–102. (https://doi.org/10.1136/bmjqs-2017-007669)

41. Qian F, Osler TM, Eaton MP, Dick AW, Hohmann SF, Lustik SJ, Diachun CA, Pasternak R, Wissler RN & Glance LG. Variation of blood transfusion in patients undergoing major noncardiac surgery. *Annals of Surgery* 2013 **257** 266–278. (https://doi.org/10.1097/SLA.0b013e31825ffc37)

42. 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. *Geriatric Orthopaedic Surgery and Rehabilitation* 2012 **3** 17–26. (https://doi.org/10.1177/2151458511435723)

43. Calderwood MS, Kleinman K, Bratzler DW, Ma A, Bruce CB, Kaganov RE, Canning C, Platt R, Huang SS, Centers for Disease Control and Prevention Epicenters Program, ***et al.*** Use of medicare claims to identify US hospitals with a high rate of surgical site infection after hip arthroplasty. *Infection Control and Hospital Epidemiology* 2013 **34** 31–39. (https://doi.org/10.1086/668785)

44. Skufca J, Ollgren J, Virtanen MJ, Huotari K & Lyytikainen O. Interhospital comparison of surgical site infection rates in orthopedic surgery. *Infection Control and Hospital Epidemiology* 2017 **38** 423–429. (https://doi.org/10.1017/ice.2016.333)

45. Dailey L, van Gessel H & Peterson A. Two years of surgical site infection surveillance in Western Australia: analysing variation between hospitals. *Healthcare Infection* 2009 **14** 51–60. (https://doi.org/10.1071/HI09110)

46. 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. *Enfermedades Infecciosas y Microbiologia Clinica* 2012 **30** (Supplement 3) 26–32. (https://doi.org/10.1016/S0213-005X(12)70093-9)

47. Jergesen HE & Yi PH. Early complications in hip and knee arthroplasties in a safety net hospital vs a University Center. *Journal of Arthroplasty* 2016 **31** 754–758. (https://doi.org/10.1016/j.arth.2015.10.031)

48. Cai X, Cram P & Vaughan-Sarrazin M. Are African American patients more likely to receive a total knee arthroplasty in a low-quality hospital? *Clinical Orthopaedics and Related Research* 2012 **470** 1185–1193. (https://doi.org/10.1007/s11999-011-2032-6)

49. Marang-van de Mheen PJ, Bragan Turner E, Liew S, Mutalima N, Tran T, Rasmussen S, Nelissen RGHH & Gordon A. 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 Musculoskeletal Disorders* 2017 **18** 207. (https://doi.org/10.1186/s12891-017-1569-2)

50. 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. *European Journal of Orthopaedic Surgery and Traumatology* 2013 **23** 165–168. (https://doi.org/10.1007/s00590-012-0937-8)

51. 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. *Journal of Arthroplasty* 2017 **32** 2060.e1–2064.e1. (https://doi.org/10.1016/j.arth.2017.02.055)

52. 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. *Journal of Knee Surgery* 2018 **31** 815–821. (https://doi.org/10.1055/s-0037-1615299)

53. 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 Musculoskeletal Disorders* 2010 **11** 160. (https://doi.org/10.1186/1471-2474-11-160)

54. 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** 486–499. (https://doi.org/10.3390/jfb6030486)

55. 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, USA: Agency for Healthcare Research and Quality, 2006.

56. 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. *Journal of Bone and Joint Surgery: American Volume* 2011 **93** 2287–2293. (https://doi.org/10.2106/JBJS.J.01727)

57. Pijls BG, Meessen JM, Schoones JW, Fiocco M, van der Heide HJ, Sedrakyan A & Nelissen RG. 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** e0156051. (https://doi.org/10.1371/journal.pone.0156051)

58. Mäkelä KT, Matilainen M, Pulkkinen P, Fenstad AM, Havelin LI, Engesaeter L, Furnes O, Overgaard S, Pedersen AB, Kärrholm J, *et al.* Countrywise results of total hip replacement. An analysis of 438,733 hips based on the Nordic Arthroplasty Register Association database. *Acta Orthopaedica* 2014 **85** 107–116. (https://doi.org/10.3109/17453674.2014.893498)

59. International Society of Arthroplasty Registries (ISAR). (available at: https://www.isarhome.org/home)

60. 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. *Journal of Bone and Joint Surgery: American Volume* 2014 **96** (Supplement 1) 1–6. (https://doi.org/10.2106/JBJS.N.00739)

61. 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). *Journal of Bone and Joint Surgery: American Volume* 2011 **93** (Supplement 3) 76–80. (https://doi.org/10.2106/JBJS.K.01141)

62. Signorini DF & Weir NU. Any variability in outcome comparisons adjusted for case mix must be accounted for. *BMJ* 1999 **318** 128. (https://doi.org/10.1136/bmj.318.7176.128a)

63. Lenguerrand E, Whitehouse MR, Beswick AD, Kunutsor SK, Burston B, Porter M & Blom AW. Risk factors associated with revision for prosthetic joint infection after hip replacement: a prospective observational cohort study. *Lancet: Infectious Diseases* 2018 **18** 1004–1014. (https://doi.org/10.1016/S1473-3099(18)30345-1)

64. 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 Musculoskeletal Disorders* 2012 **13** 251. (https://doi.org/10.1186/1471-2474-13-251)

65. Jasper LL, Jones CA, Mollins J, Pohar SL & Beaupre LA. Risk factors for revision of total knee arthroplasty: a scoping review. *BMC Musculoskeletal Disorders* 2016 **17** 182. (https://doi.org/10.1186/s12891-016-1025-8)

66. Duclos A, Chollet F, Pascal L, Ormando H, Carty MJ, Polazzi S, Lifante JC & SHEWHART Trial Group. Effect of monitoring surgical outcomes using control charts to reduce major adverse events in patients: cluster randomised trial. *BMJ* 2020 **371** m3840. (https://doi.org/10.1136/bmj.m3840)

67. Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, French SD, O'Brien MA, Johansen M, Grimshaw J & Oxman AD. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database of Systematic Reviews* 2012 **6** CD000259. (https://doi.org/10.1002/14651858.CD000259.pub3)

68. Soong C & Shojania KG. Education as a low-value improvement intervention: often necessary but rarely sufficient. *BMJ Quality and Safety* 2020 **29** 353–357. (https://doi.org/10.1136/bmjqs-2019-010411)

69. Gude WT, Roos-Blom MJ, van der Veer SN, de Jonge E, Peek N, Dongelmans DA & de Keizer NF. 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. *Implementation Science* 2017 **12** 68. (https://doi.org/10.1186/s13012-017-0594-8)

70. Roos-Blom MJ, Gude WT, de Jonge E, Spijkstra JJ, van der Veer SN, Peek N, Dongelmans DA & de Keizer NF. Impact of audit and feedback with action implementation toolbox on improving ICU pain management: cluster-randomised controlled trial. *BMJ Quality and Safety* 2019 **28** 1007–1015. (https://doi.org/10.1136/bmjqs-2019-009588)