

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-VOS², Liza N VAN STEENBERGEN³, Rob G H H NELISSEN¹, and Perla J MARANG-VAN DE MHEEN², 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.

Correspondence: p.van_schie@lumc.nl

Submitted 2021-05-07. Accepted 2021-10-18.

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.

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, this 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 (ClinicalTrials.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-col-

lection 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 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 = textbook outcome).
- 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 with 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 conflicts 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).

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

Table 1. Baseline patient characteristics and indicators after THA and TKA in the period 2017–2019 in 20 Dutch hospitals. Values are count (%) unless otherwise specified

Patient characteristics	THA (n = 22,908)	TKA (n = 20,423)
Mean age (SD)	69 (10)	68 (8.8)
Female sex	14,707 (64)	12,606 (62)
BMI (SD)	27 (4.5)	29 (4.8)
Current smokers	2,395 (11)	1,667 (8.4)
ASA classification		
I	4,113 (18)	2,736 (13)
II	14,533 (63)	13,759 (68)
III–IV	4,259 (19)	3,924 (19)
Charnley score		
A	9,205 (42)	7,529 (37)
B1	7,082 (32)	7,598 (37)
B2	4,984 (23)	4,470 (22)
C	711 (3)	722 (4)
Diagnosis		
Osteoarthritis	20,214 (88)	19,723 (97)
Non-osteoarthritis	2,669 (12)	697 (3.4)
Indicators		
1-year revision	410 (1.8) ^a	250 (1.2) ^b
30-day readmission	829 (3.9)	633 (3.4)
30-day complication	1,027 (4.5)	620 (3.3)
Long LOS, upper quartile	2,794 (13.3)	2,123 (11.4)

^a The 1-year revision percentage for THA was 1.8% in the Netherlands during 2014–2016 (4).
^b The 1-year revision percentage for TKA was 1.2% in the Netherlands during 2014–2016 (4).

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).

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 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

Table 2. Baseline hospital-level characteristics and indicators for 20 Dutch hospitals performing THA and TKA. Values are percentage (IQR) or mean as specified of the median hospital

	THA	TKA
Procedures, n	1,188 (623–1,630)	848 (593–1,552)
Mean age	69 (65–70)	69 (66–70)
Female sex	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		
I	14 (9.7–23)	9.8 (7.5–19)
II	64 (59–70)	67 (61–72)
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) ^a	1.3 (0.7–1.7) ^b
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)

All standardized indicators were adjusted for: age, sex, BMI, current smokers, ASA classification, Charnley score, and diagnosis.

^a The median percentage on hospital-level for THA was 1.6% (IQR: 1.0–2.3) in the Netherlands during 2014–2016 (4).

^b The median percentage on hospital-level for TKA was 1.1% (IQR: 0.7–1.6) in the Netherlands during 2014–2016 (4).

[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).

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.

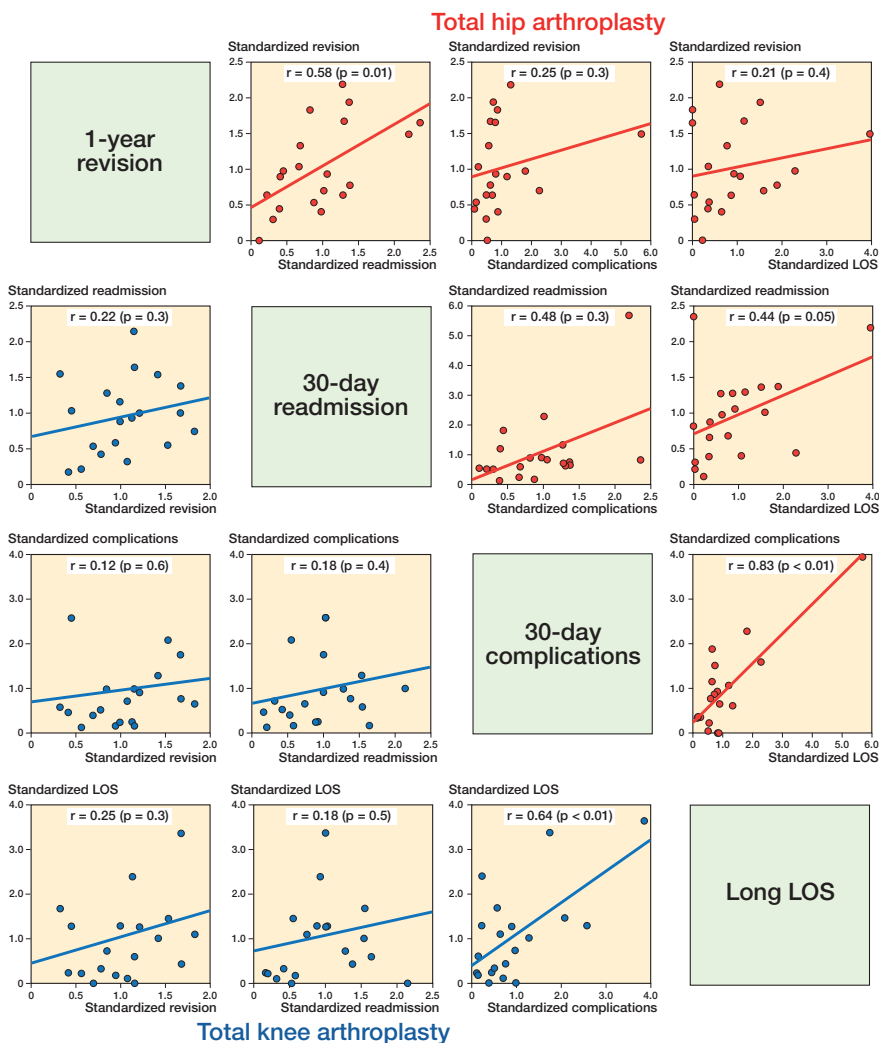


Figure 1. Correlation between standardized rates of individual indicators at hospital level. All indicators were adjusted for the following patient characteristics: age, sex, BMI, current smokers, ASA classification, Charnley score, and diagnosis. LOS = length of stay.

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.

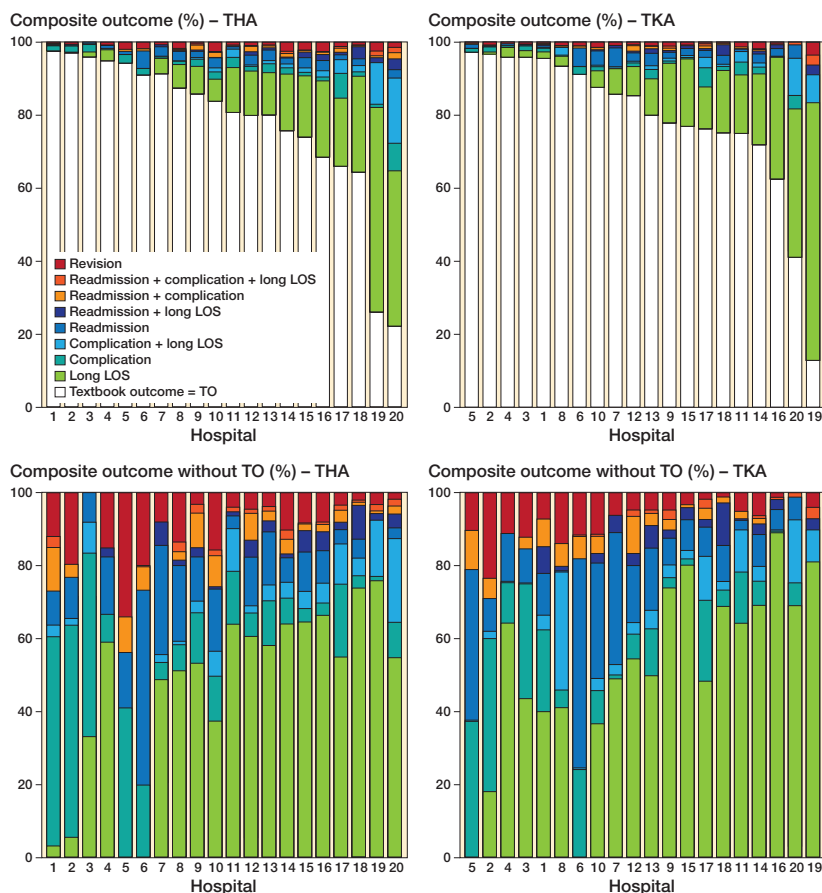


Figure 2. Crude ordinal composite outcome distribution per hospital and standardized effect of the hospitals on the composite outcome (THA median 1.04 [IQR 0.5–1.7] and TKA median 1.25 [IQR 0.4–1.6]). This graphs show the crude outcome distribution per hospital ($n = 20$). The hospitals are numbered on the X-axis. The hospitals for TKA were labelled according to their rank for TO in THAs. The standardized odds of the hospital effect (median and IQR) were adjusted for the following patient characteristics: age, sex, BMI, current smokers, ASA classification, Charnley score, and diagnosis.

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 were to occur, it would only influence the relationship between indicators if systematic and frequent 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 read-

mission). 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 that 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, hospitals 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 the 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

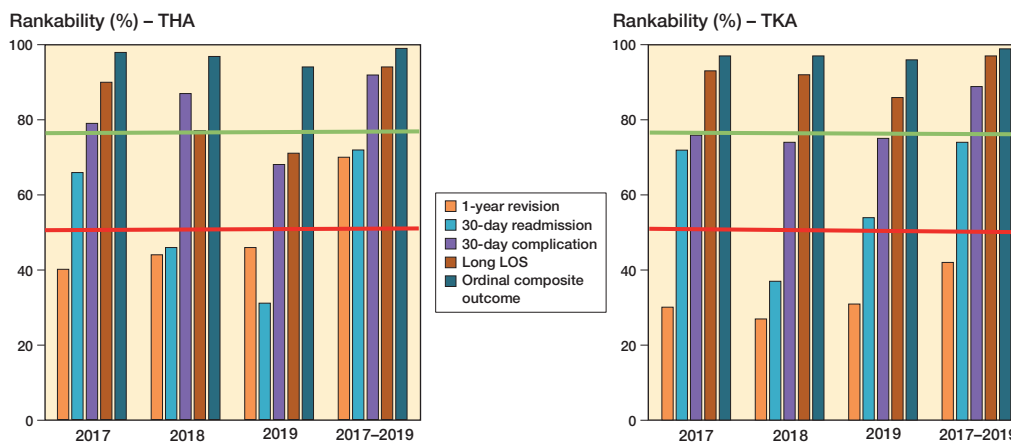


Figure 3. Rankabilities of individual indicators and ordinal composite outcomes. The rankability is high when the bar is above the green line, moderate when between the red and green line, and low when below the red line.

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.

Concept and design: PvS, RN, PM. Collecting data: PvS, LvS, PM. Interpretation of data: PvS, LvB, RN, PM. Writing of the manuscript: PvS. Critical revision of the manuscript: LvB, LvS, RN, PM. Statistical analysis: PvS, PM. Supervision: RN, PM.

The authors 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).

Acta thanks Anne Lübbecke and other anonymous reviewers for help with peer review of this study.

1. **Lubbeke A, Silman A J, Barea C, Prieto-Alhambra D, Carr A J.** Mapping existing hip and knee replacement registries in Europe. *Health Policy* 2018; 122(5): 548-57.
2. **World Health Organization.** World Health Organization (WHO)—Quality of care. https://www.who.int/health-topics/quality-of-care#tab=tab_1.
3. **van Schie P, Van Bodegom-Vos L, Zijdemans T M, Nelissen R, Marang-Van De Mheen P J.** 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 Steenberghe L N, van Bodegom-Vos L, Nelissen R, Marang-van de Mheen P J.** 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.
5. **Hollenbeck B, Hoffman MA, Tromanhauser S G.** 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. **Karthaas E G, Lijftogt N, Busweiler L A D, Elsmann B H P, Wouters M, Vahl A C, et al.** Textbook Outcome: a composite measure for quality of elective aneurysm surgery. *Ann Surg* 2017; 266(5): 898-904.
7. **Hofstede S N, Ceyisakar I E, Lingsma H F, Kringos D S, Marang-van de Mheen P J.** Ranking hospitals: do we gain reliability by using composite rather than individual indicators? *BMJ Qual Saf* 2019; 28(2): 94-102.

8. **Dimick J B, Birkmeyer N J, Finks J F, Share D A, English W J, Carlin A M, et al.** Composite measures for profiling hospitals on bariatric surgery performance. *JAMA Surg* 2014; 149(1): 10-16.
9. **Marang-van de Mheen P J, Dijs-Elsinga J, Otten W, Versluijs M, Smeets H J, 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 M M, Smeets H J, 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 H F, Bottle A, Middleton S, Kievit J, Steyerberg E W, Marang-van de Mheen P J.** 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 V M, Permyer-Miralda G, Ferreira-González I, Busse J W, 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 Qual Saf* 2019; 28(4): 338-44.
14. **ClinicalTrials.gov.** <https://clinicaltrials.gov/ct2/show/NCT04055103?term=Arthroplasty&entry=NL&city=Leiden&draw=2&rank=2>.
15. **van Steenberghe L N, Denissen G A, Spooren A, van Rooden S M, van Oosterhout F J, 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.** 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 February 2019).
17. **Marang-van de Mheen P J, 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 Y Q M, Marang-van de Mheen P J, Wouters M, Nienhuijs S W, Liem R S L.** Textbook Outcome: an ordered composite measure for quality of bariatric surgery. *Obes Surg* 2019; 29(4): 1287-94.
19. **Jaffer A K, Barsoum W K, Krebs V, Hurbanek J G, Morra N, Brotman D J.** 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 N N, Barrett J A, Katz J N, Phillips C B, Losina E, Lew R A, 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 P F, Purtill J J, Rothman R H, 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 A C, Snijders A, Snijders H S, Tollenaar R A, Wouters M W, et al.** Ranking and rankability of hospital postoperative mortality rates in colorectal cancer surgery. *Ann Surg* 2014; 259(5): 844-9.
26. **van Dishoeck A M, Koek M B, Steyerberg E W, van Benthem B H, Vos M C, Lingsma H F.** 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 A M, Lingsma H F, Mackenbach J P, Steyerberg E W.** Random variation and rankability of hospitals using outcome indicators. *BMJ Qual Saf* 2011; 20(10): 869-74.
28. **Dimick J B, Welch H G.** The zero mortality paradox in surgery. *J Am Coll Surg* 2008; 206(1): 13-16.
29. **Drye E E, Chen J.** Evaluating quality in small-volume hospitals. *Arch Intern Med* 2008; 168(12): 1249-51.
30. **Austin P C, Ceyisakar I E, Steyerberg E W, Lingsma H F, Marang-van de Mheen P J.** 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 C J, Livingston E H, Aylin P.** Global comparators project: international comparison of hospital outcomes using administrative data. *Health Serv Res* 2013; 48(6 Pt 1): 2081-100.