


Clinical Research

Do the Revision Rates of Arthroplasty Surgeons Correlate With Postoperative Patient-reported Outcome Measure Scores? A Study From the Australian Orthopaedic Association National Joint Replacement Registry

Wayne Hoskins MBBS (Hons), FRACS, PhD¹⁻³ , Roger Bingham MBBS, FRACS², Sophia Corfield PhD⁴, Dylan Harries BSc (Hons), PhD⁵, Ian A. Harris MBBS, MMed (Clin Epi), MSc(HDS), PhD, FRACS (Orth), FAHMS^{4,6}, Kelly G. Vince MD, FRCS(C)³

Received: 30 July 2022 / Accepted: 22 May 2023 / Published online: 20 June 2023
Copyright © 2023 by the Association of Bone and Joint Surgeons

Abstract

Background Patient-reported outcome measures (PROMs) are a pragmatic and efficient means to evaluate the functional quality of arthroplasty beyond revision rates, which are used by most joint replacement registries to judge success. The relationship between these two measures of quality—revision rates and PROMs—is unknown, and not every procedure with a poor functional result is revised. It is logical—although still untested—that higher cumulative revision rates correlate inversely with PROMs for individual surgeons; more revisions are associated with lower PROM scores.

Questions/purposes We used data from a large national joint replacement registry to ask: (1) Does a surgeon's early THA cumulative percent revision (CPR) rate and (2) early TKA CPR rate correlate with the postoperative PROMs of patients undergoing primary THA and TKA, respectively, who have not undergone revision?

Methods Elective primary THA and TKA procedures in patients with a primary diagnosis of osteoarthritis that were performed between August 2018 and December 2020 and registered in the Australian Orthopaedic Association

National Joint Replacement Registry PROMs program were eligible. THAs and TKAs were eligible for inclusion in the primary analysis if 6-month postoperative PROMs were available, the operating surgeon was clearly identified, and the surgeon had performed at least 50 primary THAs or TKAs. Based on the inclusion criteria, 17,668 THAs were performed at eligible sites. We excluded 8878 procedures that were not matched to the PROMs program, leaving 8790 procedures. A further 790 were excluded because they were performed by unknown or ineligible surgeons or were revised, leaving 8000 procedures performed by 235 eligible surgeons, including 4256 (53%; 3744 cases of missing data) patients who had postoperative Oxford Hip Scores and 4242 (53%; 3758 cases of missing data) patients who had a postoperative EQ-VAS score recorded. Complete covariate data were available for 3939 procedures for the Oxford Hip Score and for 3941 procedures for the EQ-VAS. A total of 26,624 TKAs were performed at eligible sites. We excluded 12,685 procedures that were not matched to the PROMs program, leaving 13,939 procedures. A further 920 were excluded because

Each author certifies that there are no funding or commercial associations (consultancies, stock ownership, equity interest, patent/licensing arrangements, etc.) that might pose a conflict of interest in connection with the submitted article related to the author or any immediate family members.

All ICMJE Conflict of Interest Forms for authors and *Clinical Orthopaedics and Related Research*® editors and board members are on file with the publication and can be viewed on request.

AOANJRR collection of joint replacement and PROMs data are approved by the Commonwealth of Australia as a federal quality assurance activity (QAA 3/2017) under section 124X of the Health Insurance Act, 1973.

This work was performed at the Australian Orthopaedic Association National Joint Replacement Registry.

they were performed by unknown or ineligible surgeons, or because they were revisions, leaving 13,019 procedures performed by 276 eligible surgeons, including 6730 (52%; 6289 cases of missing data) patients who had had postoperative Oxford Knee Scores and 6728 (52%; 6291 cases of missing data) patients who had a postoperative EQ-VAS score recorded. Complete covariate data were available for 6228 procedures for the Oxford Knee Score and for 6241 procedures for the EQ-VAS. The Spearman correlation between the operating surgeon's 2-year CPR and 6-month postoperative EQ-VAS Health and Oxford Hip or Oxford Knee Score was evaluated for THA and TKA procedures where a revision had not been performed. Associations between postoperative Oxford and EQ-VAS scores and a surgeon's 2-year CPR were estimated based on multivariate Tobit regressions and a cumulative link model with a probit link, adjusting for patient age, gender, ASA score, BMI category, preoperative PROMs, as well as surgical approach for THA. Missing data were accounted for using multiple imputation, with models assuming they were missing at random and a worst-case scenario.

Results Of the eligible THA procedures, postoperative Oxford Hip Score and surgeon 2-year CPR were correlated so weakly as to be clinically irrelevant (Spearman correlation $\rho = -0.09$; $p < 0.001$), and the correlation with postoperative EQ-VAS was close to zero ($\rho = -0.02$; $p = 0.25$). Of the eligible TKA procedures, postoperative Oxford Knee Score and EQ-VAS and surgeon 2-year CPR were correlated so weakly as to be clinically irrelevant ($\rho = -0.04$; $p = 0.004$ and $\rho = 0.03$; $p = 0.006$, respectively). All models accounting for missing data found the same result.

Conclusion A surgeon's 2-year CPR did not exhibit a clinically relevant correlation with PROMs after THA or TKA, and all surgeons had similar postoperative Oxford scores. PROMs, revision rates, or both may be inaccurate or imperfect indicators of successful arthroplasty. Missing data may limit the findings of this study, although the results were consistent under a variety of different missing data scenarios. Innumerable factors contribute to arthroplasty results,

including patient-related variables, differences in implant design, and the technical quality of the procedure. PROMs and revision rates may be analyzing two different facets of function after arthroplasty. Although surgeon variables are associated with revision rates, patient factors may exert a stronger influence on functional outcomes. Future research should identify variables that correlate with functional outcome. Additionally, given the gross level of function that Oxford scores record, outcome measures that can identify clinically meaningful functional differences are required. The use of Oxford scores in national arthroplasty registries may rightfully be questioned.

Level of Evidence Level III, therapeutic study.

Introduction

The original mandate for most joint replacement registries was to identify underperforming prostheses by recording revision surgeries. This was a qualitative, binary, "all or none" outcome measure that successfully flagged some problem implants [17] and improved arthroplasty outcomes. This system is important, but because some problem surgeries never proceed to revision, registries underestimate poor outcomes [26]. Some patients may be unwilling or unfit to undergo further surgery, and some surgeons may decline to perform certain revisions [50]. An indicated revision may not be done; a patient may be advised to live with limited motion or instability after TKA. Some infections go undiagnosed, and some are managed without implant exchange or with chronic antibiotic suppression. A maltracking patella may be dismissed as not worth the risk of revision. Expectations may not be reached.

By contrast, implant revision may not be indicated, despite a poor outcome. Residual pain after a well-performed TKA can originate in the hip, low back, or other source [11]. If knee pathology existed concurrent with other pathology before primary arthroplasty, the patient may perceive partial improvement as a failure. If

¹Faculty of Medicine, Dentistry and Health Sciences, The University of Melbourne, Parkville, Australia


²Traumaplasty Melbourne, East Melbourne, Australia

³Department of Orthopaedics, Whangarei Hospital, Northland District Health Board, Whangarei, New Zealand

⁴Australian Orthopaedic Association National Joint Replacement Registry, Adelaide, Australia

⁵South Australian Health and Medical Research Institute, Adelaide, Australia

⁶Ingham Institute for Applied Medical Research, School of Clinical Medicine, UNSW Sydney, Australia

W. Hoskins , Faculty of Medicine, Dentistry and Health Sciences, Building 181, University of Melbourne, Grattan Street, Melbourne, Victoria 3010, Australia, Email: wayne.hoskins@outlook.com

significant knee pathology did not exist, then surgery may have failed because of misdiagnosis. Technically sound arthroplasties performed for minimal arthritis with modest radiographic findings can result in patient dissatisfaction [13, 31, 34]. Accordingly, some arthroplasties are problematic but should not be revised. Assuming that unrevised arthroplasties are performing well, registry data based on revision rates alone overestimate patient function and the success of surgery.

How can registries identify unrevised yet poorly performing implants? Simple scoring systems were developed to assess the functional quality of arthroplasty, with a focus on the implant [28]. Then, to avoid bias when practitioners collect data directly, patient-reported outcome measures (PROMs) were introduced to allow patients [5] to complete questionnaires themselves [32, 55]. PROMs have largely superseded clinical scoring systems performed by a clinical care team. More recently, extremely basic questionnaires have been sent by mail, asking patients whether they were satisfied with their arthroplasty. These provided sobering insights: Typically, about one of five respondents reported they were either dissatisfied or unsure whether they were satisfied [38]. The factors that contribute to satisfaction are as complex as the questionnaires are simple. Satisfaction assessments have become an unfortunate, misappropriated, and much-repeated trope in many current papers, and it may be difficult to even measure satisfaction meaningfully [36].

PROMs have been studied by numerous national registries [24, 52] to evaluate the functional quality of arthroplasty beyond simple revision rates. PROMs should identify measurable benefits from surgery. The relationship between the two measures of arthroplasty quality, revision rates and PROMs, is unknown. Variables related to the surgeon are recognized as having a substantial influence on revision rates for primary THA and TKA [27, 42]. It is logical, although as yet untested, that higher cumulative revision rates would correlate inversely with PROMs; that is, surgeons with more revisions that result from technical errors or complications would tend to have patients with lower PROM scores for unrevised procedures.

We suggest that if revision rates and PROMs are some legitimate measures of the quality of surgery, then surgeons with higher early revision rates after THA and TKA will also have patients with lower scores. If this thinking is wrong, either PROMs or revision rates alone, or both, may be inaccurate or imperfect indicators of the quality of surgery. Testing this theory should either support the resources allocated to PROMs collection or the validity of PROMs themselves might be questioned, leading registries to consider whether they represent the best expenditure of resources.

We therefore used data from a large national joint replacement registry to ask: (1) Does a surgeon's early THA cumulative percent revision (CPR) rate and (2) early TKA

CPR rate correlate with the postoperative PROMs of patients undergoing primary THA and TKA, respectively, who have not undergone revision?

Patients and Methods

Study Design and Setting

This observational study analyzed the association between the early (2-year) revision rates of THA and TKA surgeons and their patients' postoperative PROMs where revision has not occurred, using data from a large national arthroplasty registry. The Australian Orthopaedic Association National Joint Replacement Registry (AOANJRR) began collecting data on September 1, 1999. Registry data are validated against patient-level data provided by each of the state and territory health departments in Australia with the use of a sequential, multilevel matching process. A matching program is run monthly to search for all primary and revision arthroplasties recorded in the AOANJRR that involved the same side and joint for the same patient, thus enabling each revision to be linked to the primary procedure. Data are also matched to the Australian Institute of Health and Welfare's National Death Index to obtain information on the patient's date of death. After cross-checking data, the registry can obtain data on more than 97% of joint arthroplasties performed in Australia [1].

In 2018, the AOANJRR began a pilot program of electronically recording preoperative and 6-month postoperative PROMs as the sole postoperative timepoint from patients in 44 hospitals before and after elective THA and TKA. A total of 44 hospitals across Australia were included, comprising metropolitan and regional facilities as well as private and public hospitals from all states and one territory [24]. After the pilot study was concluded, recruitment of additional sites commenced with the expansion of the PROMs program nationally. As of December 31, 2020, 47 sites had formally commenced data collection and had data available for analysis in this study. All patients undergoing primary or revision hip or knee arthroplasty at these sites were eligible to participate in the PROMs program. Upon registering and providing consent to participate in the program, patients could provide responses to the preoperative and postoperative surveys when convenient for them via an electronic data collection system. Patients could provide preoperative responses up to 5 months after the index procedure, while postoperative responses could be provided between 5 and 8 months postoperatively. Patients who had not completed their surveys within these periods were considered to have been lost to follow-up for assessing preoperative and postoperative completeness. Procedure and patient details routinely collected by the AOANJRR are subsequently matched to each PROMs registration.

For the study, we considered THA and TKA procedures for a diagnosis of osteoarthritis performed by eligible surgeons at sites eligible for inclusion in the AOANJRR PROMs program. Surgeons were eligible if they had performed at least 50 primary THAs or TKAs for osteoarthritis as of December 31, 2020 that had a 2-year CPR and preoperative and postoperative PROMs recorded.

Patients

Between August 2018 and December 2020, a total of 82,849 primary THA procedures were recorded in the AOANJRR nationally. A total of 17,668 of these THAs were performed at sites that had registered to participate in the PROMs program. The registration rate, defined as the proportion of THA procedures performed at each site that were registered in the PROMs program, ranged from 0% to 95.1%, with an overall registration rate of 49.8%. Procedures that were not registered, accounting for 50% (8878) of procedures at participating sites, were excluded, leaving 8790 procedures. A further 790 of these were excluded because they were performed by unknown or ineligible surgeons or were revised before postoperative PROMs were collected, leaving 8000 procedures performed by 235 eligible surgeons, including 4256 (53%; 3744 cases of missing data) patients who had postoperative scores for the Oxford Hip Score and 4242 (53%; 3758 cases of missing data) patients who had a postoperative EQ-VAS recorded. Complete covariate data, including the preoperative score and patient age, gender, American Society of Anesthesiologists (ASA) score, BMI, and surgical approach, were available for 3939 procedures for the Oxford Hip Score and for 3941 for the EQ-VAS for the final study group (Fig. 1).

For TKA, 131,952 primary procedures with a diagnosis of osteoarthritis were recorded in the AOANJRR between August 2018 and December 2020, of which 26,624 TKAs were performed at sites registered in the PROMs program. As for THA, the registration rate across sites varied widely, from 0% to 91.5%, with an overall registration rate of 52.4%. We again excluded procedures that were not registered, corresponding to 48% (12,685 procedures) of procedures at participating sites, leaving 13,939 procedures. A further 920 were excluded because they were performed by unknown or ineligible surgeons or were revised before postoperative PROMs were collected. This left 13,019 procedures performed by 276 eligible surgeons, including 6730 (52%; 6289 cases of missing data) patients who had had postoperative Oxford Knee Scores and 6728 (52%; 6291 cases of missing data) who had a postoperative EQ-VAS score recorded. Complete covariate data on preoperative score, patient age, gender, ASA score, and BMI were

available for 6228 procedures for the Oxford Knee Score and 6241 procedures for the EQ-VAS for the final study group (Fig. 2).

The proportion of potentially eligible procedures registered in the PROMs program at participating sites and by eligible surgeons varied widely. To assess whether the relationship between postoperative scores and surgeon revision rate differed according to the registration completeness achieved by the center or surgeon, we undertook sensitivity analyses in which only sites or surgeons achieving a registration rate of 60%, corresponding to previously recommended levels of completeness [39], were included.

In addition to differences in registration rates, hospitals and surgeons also differed in the level of completeness of the preoperative and postoperative survey responses collected for their procedures. Among patients with THA procedures registered in the PROMs program who had at least 5 months of follow-up, and hence had either provided their preoperative responses or had been lost to follow-up, the overall rate of preoperative survey completion was 85%, with this rate varying from 57% to 100% across sites. The postoperative completion rate, based on procedures that had reached the end of their postoperative collection period at 8 months postsurgery, was 74%, and this varied from 37% to 100% across sites. For eligible TKA procedures, the corresponding preoperative and postoperative completion rates were 85% and 73%, respectively, with site completeness rates varying from 45% to 100% preoperatively and from 20% to 89% postoperatively. To better address possible biases arising from missing response data for otherwise eligible procedures, multiple imputation was used to examine the relationship between postoperative scores and surgeon revision rates, as described below.

Response rates for completion of preoperative PROMs varied from 50% to 100% of registered procedures across sites, with an overall preoperative completion rate of 85% (19,241 of 22,729) of THA and TKA procedures. Among patients who had at least 8 months of follow-up, 73% (11,718 of 15,962) had complete postoperative data. The response rate for postoperative completion varied between 33% and 88% across sites.

Descriptive Data

Compared with THAs performed since the commencement of PROMs that were either not registered in the PROMs program or performed by an ineligible or unknown surgeon, the 8000 eligible THA procedures tended to have been performed in younger patients and had a higher proportion of procedures using the posterior approach, but they were similar in terms of other demographic factors

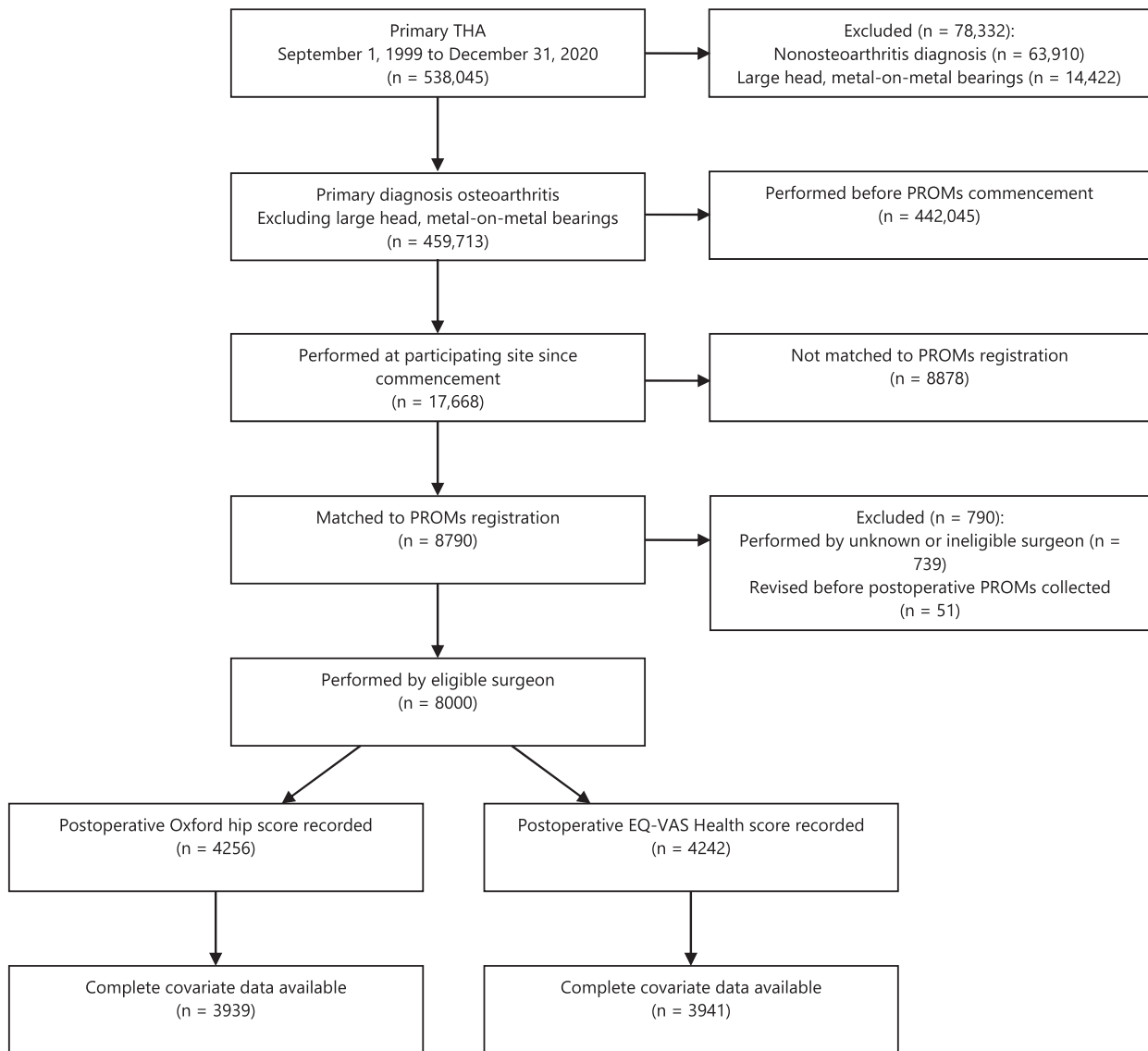


Fig. 1 This flowchart shows a summary of inclusions and exclusions from the analysis of postoperative Oxford Hip Score and EQ-VAS Health in primary conventional THA.

such as gender, ASA class, and BMI (Supplemental Table 1; <http://links.lww.com/CORR/B153>). The mean patient age, age groupings, gender, ASA class, BMI, surgical approach, and preoperative PROMs were similar between surgeons in all quartiles of revision rates for procedures with recorded postoperative Oxford Hip Scores (Table 1) and EQ-VAS scores (Supplemental Table 2; <http://links.lww.com/CORR/B154>).

For TKA, the 13,019 eligible procedures were performed more often in younger patients, but these patients were otherwise similar in demographics to the unregistered or otherwise ineligible patients with TKAs performed since the commencement of the PROMs program (Supplemental

Table 3; <http://links.lww.com/CORR/B155>). The mean patient age, age groupings, gender, ASA class, BMI, and preoperative PROM scores were similar between surgeons in all quartiles of revision rates for procedures with postoperative Oxford Knee Scores (Table 2) and EQ-VAS scores (Supplemental Table 4; <http://links.lww.com/CORR/B156>).

Outcomes

To estimate the THA and TKA revision rate of individual surgeons, we analyzed data for the CPR for all THAs and

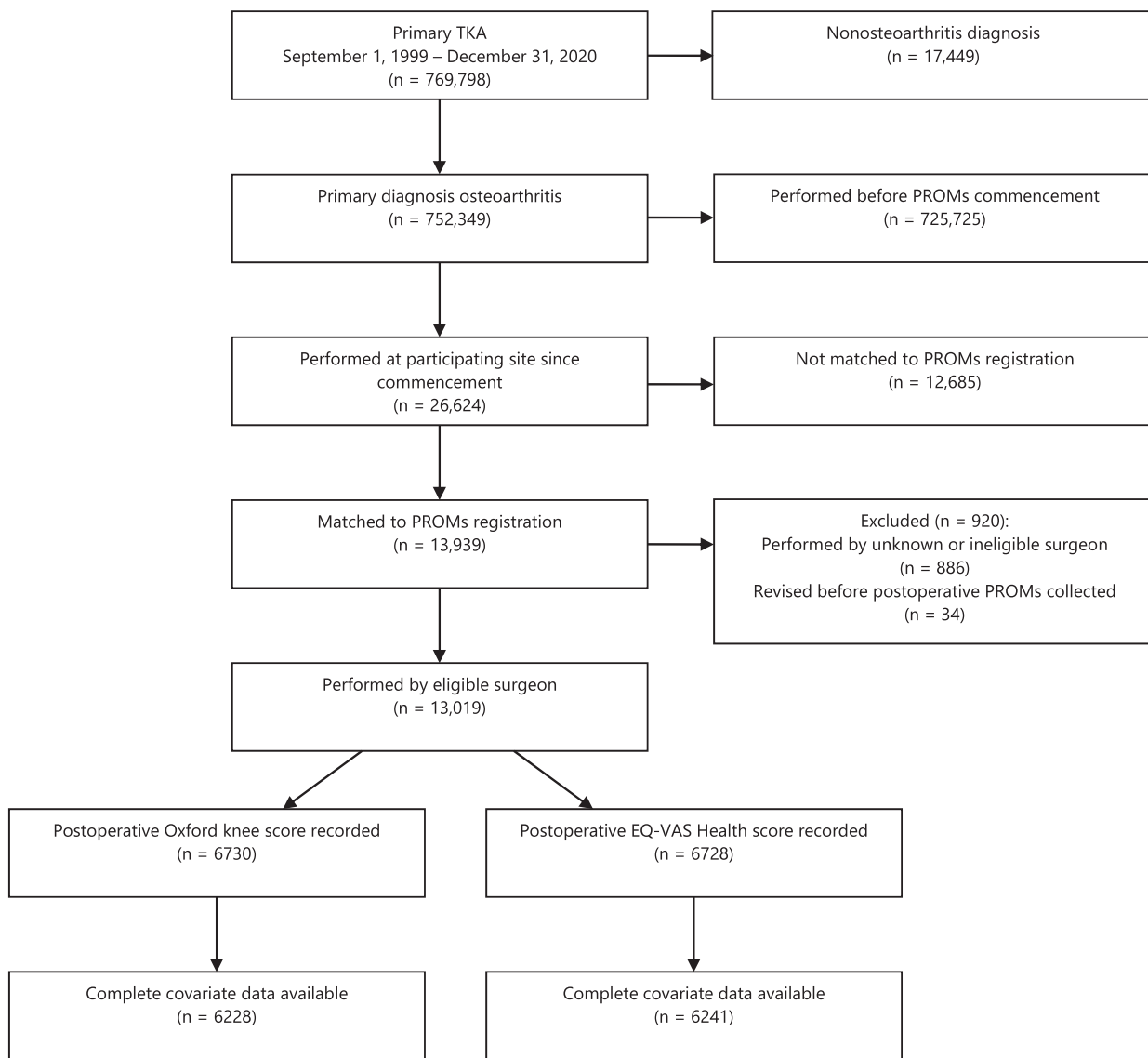


Fig. 2 This flowchart shows a summary of inclusions and exclusions from the analysis of postoperative Oxford Knee Score and EQ-VAS Health in primary TKA.

TKAs performed for osteoarthritis registered in the AOANJRR from September 1, 1999, when collection began, to December 31, 2020. The 2-year CPR for each surgeon was obtained from the Kaplan-Meier survival function estimate based on a surgeon’s total conventional THAs and TKAs with a primary diagnosis of osteoarthritis, excluding THA procedures in which a prosthesis with a large head (head size > 32 mm) and metal-on-metal bearings was used. The 2-year CPR was used as early revision surgery is mainly because of technical factors, surgical complications, and infection [27]. THAs and TKAs were grouped based on the quartiles of 2-year CPR values for

THAs or TKAs for calculating descriptive statistics. Validated PROMs included the Oxford Hip and Knee Scores (range 0 to 48; scored according to worst-best hip or knee function) [15, 16, 41] (the meaningful or minimum important change for both after primary THA and TKA is 5 points [3, 10], although this is typically felt to be population-specific) and EQ-5D-5L and the EQ-VAS Health (range 0 to 100; scored according to worst-best health) [7, 14, 25, 33].

We calculated the correlation between a surgeon’s 2-year CPR for THA and TKA and postoperative PROMs for patients who had procedures with postoperative Oxford Hip

Table 1. Summary of primary conventional THA with postoperative Oxford Hip Score by surgeon 2-year CPR (primary diagnosis osteoarthritis)

Variable	< 25th percentile (n = 993)	25th to < 50th percentile (n = 1220)	50th to < 75th percentile (n = 920)	≥ 75th percentile (n = 1123)
Surgeon 2-year CPR	0.8 (0.5-0.9)	1.2 (1.2-1.4)	2 (1.8-2.1)	2.9 (2.5-3.7)
Follow-up in years	1.5 (1.1-1.9)	1.5 (1.1-1.8)	1.4 (1.1-1.8)	1.5 (1.1-1.9)
Age in years	68 (61-74)	68 (60-74)	67 (59-73)	68 (61-75)
Preoperative Oxford Hip Score ^a	21 (15-27)	21 (15-28)	20 (13-26)	21 (15-28)
Postoperative Oxford Hip Score	45 (40-47)	44 (39-47)	44 (38-47)	43 (39-46)
Age in years				
< 55	9 (85)	13 (164)	13 (119)	10 (114)
55-64	28 (279)	25 (309)	30 (272)	26 (295)
65-74	39 (386)	37 (448)	37 (339)	38 (431)
≥ 75	24 (243)	25 (299)	21 (190)	25 (283)
Women	55 (550)	54 (661)	54 (497)	54 (605)
ASA class ^b				
ASA 1	8 (76)	11 (135)	7 (65)	6 (66)
ASA 2	64 (635)	54 (664)	57 (522)	53 (598)
ASA 3	28 (275)	33 (399)	35 (325)	40 (448)
ASA 4	0.7 (7)	2 (21)	0.9 (8)	0.9 (10)
BMI in kg/m ^{2c}				
Underweight < 18.5	0.5 (5)	0.8 (10)	0.7 (6)	0.5 (6)
Normal 18.5 to 24.9	22 (219)	23 (283)	18 (163)	18 (197)
Preobese 25.0 to 29.9	37 (366)	36 (437)	33 (302)	34 (381)
Obese Class 1 30.0 to 34.9	25 (244)	23 (281)	28 (258)	30 (329)
Obese Class 2 35.0 to 39.9	11 (108)	11 (128)	12 (111)	12 (135)
Obese Class 3 ≥ 40.0	5 (49)	5 (66)	8 (73)	6 (67)
Surgical approach ^d				
Anterior	21 (205)	19 (232)	29 (271)	25 (280)
Lateral	14 (138)	13 (158)	13 (119)	9 (101)
Posterior	65 (650)	68 (827)	58 (530)	66 (742)

Data are presented as median (IQR) or % (n).

^aExcludes 287 procedures with unknown preoperative Oxford Hip Score.

^bExcludes two procedures with unknown ASA score.

^cExcludes 32 procedures with unknown BMI category.

^dExcludes three procedures with unknown surgical approach. CPR = cumulative percent revision; ASA = American Society of Anesthesiologists.

Scores, Oxford Knee Scores, or EQ-VAS Health performed by the included surgeons.

Primary Study Endpoints of Interest

The primary study aim was to identify whether there was a correlation between a surgeon's 2-year CPR for THA and TKA and the postoperative Oxford Hip Score, Oxford Knee Score, and EQ-VAS for patients who had not undergone revision. To achieve this, we identified the 2-year CPR of eligible THA and TKA surgeons and examined the correlation with the respective postoperative PROMs.

Ethical Approval

Ethical approval for this study was obtained from the Australian Government Department of Health. AOANJRR collection of joint replacement and PROMs data is approved by the Commonwealth of Australia as a federal quality assurance activity (QAA 3/2017) under section 124X of the Health Insurance Act, 1973. All AOANJRR studies are conducted in accordance with the ethical principles of research (the Helsinki Declaration II). All patients undergoing joint replacement in Australia provide consent for routine AOANJRR data collection on an opt-out basis.

Table 2. Summary of primary TKA with postoperative Oxford Knee Score by surgeon 2-year CPR (primary diagnosis osteoarthritis)

Variable	< 25th percentile (n = 1558)	25th to < 50th percentile (n = 1495)	50th to < 75th percentile (n = 1969)	≥ 75th percentile (n = 1708)
Surgeon 2-year CPR	0.6 (0.5-0.7)	1.1 (0.9-1.2)	1.8 (1.6-2)	2.8 (2.5-3.5)
Follow-up in years	1.4 (1.1-1.8)	1.4 (1.1-1.8)	1.5 (1.1-1.8)	1.5 (1.1-1.9)
Age in years	67 (62-73)	69 (62-74)	68 (62-73)	67 (60-73)
Preoperative Oxford Knee Score ^a	22 (16-28)	23 (17-29)	22 (16-28)	23 (17-28)
Postoperative Oxford Knee Score	40 (34-44)	40 (34-44)	39 (33-44)	39 (34-43)
Age in years				
< 55	6 (87)	6 (92)	7 (140)	8 (139)
55-64	29 (457)	27 (406)	28 (555)	33 (562)
65-74	44 (689)	43 (650)	44 (873)	40 (679)
≥ 75	21 (325)	23 (347)	20 (401)	19 (328)
Women	55 (862)	60 (900)	57 (1121)	53 (913)
ASA class ^b				
ASA 1	5 (72)	4 (55)	6 (110)	6 (110)
ASA 2	54 (839)	52 (776)	56 (1097)	54 (916)
ASA 3	41 (634)	43 (643)	38 (744)	39 (670)
ASA 4	0.8 (12)	1 (19)	0.8 (16)	0.7 (12)
BMI in kg/m ^{2c}				
Underweight < 18.5	0 (0)	0 (0)	0.1 (1)	0.2 (4)
Normal 18.5 to 24.9	11 (163)	10 (149)	9 (174)	9 (156)
Preobese 25.0 to 29.9	28 (439)	30 (441)	30 (560)	31 (520)
Obese class 1 30.0 to 34.9	31 (476)	30 (450)	34 (622)	30 (518)
Obese class 2 35.0 to 39.9	18 (275)	17 (253)	19 (348)	17 (295)
Obese class 3 ≥ 40.0	13 (194)	12 (184)	13 (250)	12 (206)

Data are presented as median (IQR) or % (n).

^aExcludes 456 procedures with unknown preoperative Oxford Knee Score.

^bExcludes five procedures with unknown ASA score.

^cExcludes 52 procedures with unknown BMI category.

Additional informed consent was obtained from all participants in the PROMs program.

Statistical Analysis

The relationship between the postoperative Oxford and EQ-VAS scores and surgeon 2-year CPR is described graphically using scatter plots of the postoperative score for individual procedures versus the 2-year CPR of the surgeon who performed the procedure. Potential non-linear relationships between postoperative scores and surgeon 2-year CPR were examined using locally weighted regression curves [12], with 95% CIs. No strong evidence for nonlinearity was observed, and further adjusted analyses assumed a linear relationship between postoperative score and surgeon 2-year CPR. The crude association between each postoperative score and the surgeon’s 2-year CPR was further characterized using the Spearman rank correlation, ρ , as a measure of the strength

of the monotonic association between the postoperative PROMs and surgeon CPR. For both primary THA and TKA, we evaluated the Spearman ρ using all eligible procedures for which both the postoperative score and surgeon CPR were available.

We obtained adjusted estimates of the association between postoperative PROMs and surgeon 2-year CPR based on multiple regression analyses of postoperative scores on surgeon 2-year CPR, patient age, gender, ASA class, BMI category, and preoperative PROMs, as well as surgical approach to THA. We analyzed preoperative scores, patient age, and surgeon CPR as continuous variables. Patient BMI was treated as a categorical variable, with procedures assigned BMI categories according to the recommended WHO classification of adult BMI [51]. Because of a small number of procedures with an ASA class of 5 or a BMI of less than 18.5 kg/m², the categories ASA 4 and ASA 5 were combined, as were BMI categories of underweight (< 18.5 kg/m²) and normal weight (18.5 to 24.99 kg/m²). Only main effects were included for all associated factors.

Because of floor and ceiling effects arising from bounded PROMs, we expected the assumptions underlying standard linear regression analyses to be violated, which may have yielded biased effect estimates. To address this, several authors have proposed the use of Tobit models [47] for analyzing bounded measures of health status [2, 43]. In this approach, the observed score is assumed to be a censored observation of an underlying normally distributed latent score. When the underlying latent score is between the lower and upper bound for the scale, the observed score is equal to the latent score. For latent scores outside this range, the observed score is instead fixed to the lower or upper bound, as appropriate. Tobit models were used for both the postoperative Oxford scores and EQ-VAS Health, including the associated factors described above. Censoring of values less than 0 or more than 48 is assumed for the Oxford scores and of values less than 0 or more than 100 for the EQ-VAS Health. To account for non-independence of postoperative scores between procedures performed by the same surgeon, we included a random intercept for each surgeon, with models fitted by the restricted maximum likelihood. In the following, estimated changes in the mean latent score, which are conditional based on the estimated random surgeon effect, are reported for a one percentage point increase in the surgeon's 2-year CPR. We assessed the impact of either neglecting bounds on the PROMs scores or accounting for possible curvilinearity in the scale by additionally fitting linear mixed models and cumulative link mixed models [8], respectively, with the same fixed and random effects. Because the results were similar to those found using the Tobit model, we only report the Tobit model analyses here.

A large proportion of the eligible procedures in our sample had missing data for one or more covariates. For missing PROMs scores, previous studies have found that responders are largely representative of nonresponders in this population [21, 22]. Despite this, missing predictors and outcome values were assumed to be missing at random, and they were accounted for using multiple imputation. Imputation models based on a fully conditional specification that included all associated factors included in the analysis models, as well as the postoperative Oxford score and EQ-VAS Health, were used for drawing imputed values. Preoperative responses to other PROMs collected by the AOANJRR, including the EQ-5D-5L, lower back pain VAS, joint pain VAS, expectations of postoperative mobility, general health and joint pain, and presence or absence of problems with the nonsurgical joint or with walking were included as auxiliary variables. Clustering of procedures by surgeons was accounted for by using multilevel specification for each imputation model, which included the cluster means for each predictor. Imputations were constructed using predictive mean matching as implemented in the mice R package [49]. All procedures performed by eligible surgeons and that were registered in the PROMs

program were included in constructing and analyzing the imputed datasets. For all outcomes, 20 imputations were used. Additionally, a complete case analysis for each outcome was performed to assess possible sensitivity to the imputation model, and, as noted above, the analysis was also repeated restricted to sites and surgeons with a registration rate of at least 60%. Because the results obtained when considering either the complete cases or imputed datasets were very similar, we report the results of the complete case analysis and provide estimates based on multiple imputation (Supplemental Digital Content 1; <http://links.lww.com/CORR/B157>).

The estimated associations between the postoperative scores and surgeon 2-year CPR may be severely biased if postoperative outcomes are not missing at random. For example, this would be the case if patients with poorer postoperative scores were more likely to not provide postoperative responses. If surgeons with a high 2-year CPR had on average lower postoperative scores and hence a greater proportion of missing data, then a complete case analysis would be expected to underestimate the association between the surgeon's CPR and postoperative PROMs. To assess the sensitivity of our results to potential violations of the missing at random assumption, we additionally constructed imputed datasets under a simple scenario in which patients with missing postoperative responses were assumed to have systematically worse postoperative outcomes than those with observed postoperative scores. This was achieved by shifting the imputed postoperative Oxford Hip and Knee Score and EQ-VAS Health scores by fixed constants $\delta_{\text{Oxf}} < 0$ and $\delta_{\text{VAS}} < 0$. All models were then refitted to the resulting imputed data to examine the effect of systematically lower missing postoperative scores (Supplemental Digital Content 1; <http://links.lww.com/CORR/B157>). In general, the estimated associations were shifted to larger negative values for increasing values of the shifts in the unobserved postoperative scores. However, even for very large values of the shift parameters (δ_{Oxf} , δ_{VAS}), the estimated association between the surgeon's 2-year CPR and the mean latent postoperative score remained small, suggesting our conclusions would not be changed under this simple violation of the missing-at-random assumption.

All tests were two-tailed at the 5% level of significance. Statistical analysis was performed using SAS software version 9.4 (SAS Institute Inc) and the R Environment for Statistical Computing, version 4.2.0 (R Core Team 2022), using the packages mice [49], mgcv [53, 54], cenGAM [19] and ordinal [9].

Results

THA

We found no clinically important correlation between postoperative Oxford Hip Score and surgeon 2-year CPR

(Spearman correlation $\rho = -0.09$; $p < 0.001$). Furthermore, we found no correlation between postoperative EQ-VAS and surgeon 2-year CPR (Spearman correlation $\rho = -0.02$; $p = 0.25$). These results were consistent with all statistical modeling.

After adjusting for patient age, gender, ASA class, BMI category, preoperative PROMs, and surgical approach, the association between the postoperative Oxford Hip Score and surgeon 2-year CPR was clinically negligible. A one percentage point increase in surgeon 2-year CPR was estimated to be associated with a 0.48-point decrease in the latent Oxford Hip Score (95% CI -0.74 to -0.23) (Table 3), well below the minimum clinically important difference for this instrument. For the EQ-VAS, the corresponding change in the mean latent score was -0.24 points (95% CI -0.78 to 0.29) (Table 4) and was of similarly limited clinical significance. The results remained if floor and ceiling effects were neglected and when the underlying scale was allowed to be

curvilinear. Similar results were obtained when multiple imputation was used to account for missing data (Supplemental Digital Content 1; <http://links.lww.com/CORR/B157>). When missing postoperative responses were assumed to be systematically lower than the observed scores, the estimated changes in PROMs remained well below the minimum clinically important difference, even for implausibly large shifts in the missing postoperative outcomes.

TKA

We found no clinically important correlation between postoperative Oxford Knee Score and surgeon 2-year CPR (Spearman correlation $\rho = -0.04$; $p = 0.004$). Likewise, for postoperative EQ-VAS, we found no clinically important correlation (Spearman correlation $\rho = 0.03$; $p = 0.006$). These results were consistent with all statistical modeling performed.

Table 3. Change in the mean latent postoperative Oxford Hip Score in primary conventional THA (primary diagnosis osteoarthritis)

Term	Estimate (95% CI)	p value
Age	-0.02 (-0.04 to 0.01)	0.26
Surgical approach ^a		
Anterior	1.04 (0.34 to 1.75)	0.004
Lateral	-1.48 (-2.35 to -0.61)	< 0.001
ASA classification ^b		
ASA 1	2.99 (0.42 to 5.56)	0.023
ASA 2	2.15 (-0.24 to 4.53)	0.08
ASA 3	0.95 (-1.42 to 3.31)	0.43
BMI in kg/m ^{2c}		
BMI underweight < 18.5 or normal 18.5-24.9	3.53 (2.30 to 4.75)	< 0.001
BMI preobese 25.0-29.9	2.65 (1.49 to 3.81)	< 0.001
BMI obese Class 1 30.0-34.9	2.08 (0.92 to 3.23)	< 0.001
BMI obese Class 2 35.0-39.9	0.46 (-0.79 to 1.71)	0.47
Surgeon 2-year CPR	-0.52 (-0.78 to -0.27)	< 0.001
Preoperative Oxford hip score	0.19 (0.16 to 0.22)	< 0.001
Gender		
Men	1.04 (0.53 to 1.55)	< 0.001
Variance components		
Surgeon	1.00 (0.45 to 2.20)	< 0.001

The total number of patients was 3965; residual variance estimate = 59.5.

^aThe reference category for surgical approach is posterior approach.

^bThe reference category for ASA is ASA 4 or 5.

^cThe reference category for BMI is obese Class 3 (≥ 40.0 kg/m²). ASA = American Society of Anesthesiologists; CPR = cumulative percent revision.

Table 4. Change in mean latent postoperative EQ-VAS Health in primary conventional THA (primary diagnosis osteoarthritis)

Term	Estimate (95% CI)	p value
Intercept	60.27 (53.85 to 66.69)	< 0.001
Age	0.00 (-0.06 to 0.05)	0.87
Surgical approach ^a		
Anterior	0.52 (-0.93 to 1.96)	0.48
Lateral	-2.39 (-4.19 to -0.59)	0.01
ASA classification ^b		
ASA 1	6.21 (0.98 to 11.45)	0.02
ASA 2	3.08 (-1.81 to 7.97)	0.21
ASA 3	0.05 (-4.81 to 4.90)	0.99
BMI kg/m ^{2c}		
Underweight < 18.5 or normal 18.5-24.9	5.94 (3.49 to 8.40)	< 0.001
Preobese 25.0 to 29.9	5.08 (2.75 to 7.42)	< 0.001
Obese Class 1 30.0 to 34.9	3.81 (1.47 to 6.15)	0.001
Obese Class 2 35.0 to 39.9	1.92 (-0.60 to 4.45)	0.14
Surgeon 2-year CPR	-0.30 (-0.83 to 0.23)	0.27
Sex		
Male	-0.36 (-1.38 to 0.67)	0.49
Preoperative EQ-VAS Health	0.24 (0.22 to 0.27)	< 0.001
Variance components		
Surgeon	5.35 (2.91 to 9.85)	< 0.001

The total number of patients was 3965; residual variance estimate = 59.5.

^aThe reference category for surgical approach is posterior approach.

^bThe reference category for ASA is ASA 4 or 5.

^cThe reference category for BMI is obese, Class 3 (≥ 40.0 kg/m²). ASA = American Society of Anesthesiologists; CPR = cumulative percent revision.

After adjusting for patient age, gender, ASA class, BMI category, and preoperative PROMs, we estimated that a one percentage point increase in surgeon 2-year CPR was associated with a 0.27-point decrease in Oxford Knee Score (95% CI -0.49 to -0.04) (Table 5). For the EQ-VAS, the estimated increase in the mean latent score was 0.18 points (95% CI -0.24 to 0.60) (Table 6). In both cases, the estimated change in score was too small to be clinically significant. The results remained if floor and ceiling effects were neglected, and when the underlying scale was allowed to be curvilinear. Similar results were obtained when multiple imputation was used to account for missing data and the estimated changes in PROMs remained well below the minimum clinically important difference, even for implausibly large shifts in missing postoperative outcomes.

Discussion

Arthroplasty revision rates are used by most joint replacement registries as one measure of a successful procedure. Surgeon performance, distinct from implant and

patient factors, has a strong association with revision rates, but not every procedure with a poor functional outcome results in revision surgery. PROMs evaluate the functional quality of arthroplasty with quantitative patient-level data, rather than the binary “yes or no” aspect of whether revision was performed. The relationship between PROMs and revision rates as a measure of arthroplasty quality, if any, is unknown. It is logical that higher revision rates should correlate inversely with PROMs for individual surgeons; if more of a surgeon’s procedures result in revision, the functional quality of that surgeon’s procedures might be reduced as well. Contrary to our hypothesis, this study identified no correlation that could be considered clinically relevant between a surgeon’s 2-year CPR for THAs and TKAs with the 6-month postoperative Oxford Hip Score, Oxford Knee Score, and EQ-VAS for their patients, after adjusting for patient age, gender, ASA class, BMI category, preoperative PROMs, and surgical approach to THA. This challenges the assumption that both revision rates and PROMs are valid measures of the quality of surgery. One, or both, may be inaccurate or imperfect. Given the gross level of function recorded by Oxford scores, identification and use of alternate, more-sensitive

Table 5. Change in mean latent postoperative Oxford Knee Score in primary TKA (primary diagnosis osteoarthritis)

Term	Estimate (95% CI)	p value
Intercept	28.85 (26.21 to 31.50)	< 0.001
Age in years	0.00 (-0.02 to 0.03)	0.69
ASA classification ^a		
ASA 1	4.14 (1.92 to 6.36)	< 0.001
ASA 2	2.76 (0.71 to 4.81)	0.01
ASA 3	1.94 (-0.09 to 3.97)	0.06
BMI class in kg/m ^{2b}		
Underweight < 18.5 or normal 18.5-24.9	0.81 (-0.07 to 1.69)	0.07
Preobese 25.0 to 29.9	0.62 (-0.10 to 1.34)	0.09
Obese Class 1 30.0 to 34.9	0.13 (-0.55 to 0.81)	0.70
Obese Class 2 35.0 to 39.9	-0.09 (-0.80 to 0.62)	0.80
Surgeon 2-year CPR	-0.29 (-0.51 to -0.06)	0.01
Preoperative Oxford Knee Score	0.28 (0.25 to 0.30)	< 0.001
Gender		
Men	0.31 (-0.09 to 0.71)	0.13
Variance components		
Surgeon	1.53 (0.95 to 2.45)	< 0.001

There were 6240 patients; residual variance estimate = 57.1.

^aThe reference category for ASA is ASA 4 or 5.

^bThe reference category for BMI is obese Class 3 (≥ 40.0 kg/m²). ASA = American Society of Anesthesiologists; CPR = cumulative percent revision.

Table 6. Change in the mean latent postoperative EQ-VAS Health in primary TKA (primary diagnosis osteoarthritis)

Term	Estimate (95% CI)	p value
Intercept	49.8 (44.5 to 55.2)	< 0.001
Age in years	0.00 (-0.05 to 0.05)	0.99
ASA classification ^a		
ASA 1	11.9 (7.5 to 16.4)	< 0.001
ASA 2	8.6 (4.5 to 12.7)	< 0.001
ASA 3	5.8 (1.7 to 9.9)	0.01
BMI class in kg/m ^{2b}		
Underweight < 18.5 or normal 18.5-24.9	2.1 (0.3 to 3.9)	0.02
Preobese 25.0 to 29.9	1.8 (0.38 to 3.2)	0.01
Obese Class 1 30.0 to 34.9	0.75 (-0.6 to 2.1)	0.28
Obese Class 2 35.0 to 39.9	0.52 (-0.92 to 1.96)	0.48
Surgeon 2-year CPR	0.12 (-0.30 to 0.55)	0.58
Gender		
Men	0.10 (-0.69 to 0.89)	0.81
Preoperative EQ-VAS Health	0.31 (0.29 to 0.33)	< 0.001
Variance components		
Surgeon	4.3 (2.4 to 7.8)	< 0.001

The total number of patients was 6253; residual variance estimate = 230.19.

^aThe reference category for ASA is ASA 4 or 5.

^bThe reference category for BMI is obese Class 3 (≥ 40.0 kg/m²). ASA = American Society of Anesthesiologists; CPR = cumulative percent revision.

outcome measures capable of identifying important performance differences may be required. The ongoing role of current PROMs as an outcome measure in joint replacement registries may be limited given the findings of this research.

Limitations

Missing data, especially when coupled with the pooling of data from surgeons and sites that captured < 5% of data with those that captured > 90%, are a limitation. We addressed this by considering surgeons and sites that achieved acceptable registration rates, as well as by imputing missing response data and examining the impact on our results when missing outcomes were assumed to be systematically worse than observed outcomes. In all cases, the estimated association between the surgeon's 2-year CPR and their postoperative PROMs was small enough to be of little clinical relevance, suggesting our findings hold even when allowing for differential rates of completeness between surgeons. Furthermore, responding and nonresponding patients are similar in this patient population [21, 22]. Another limitation is that the proportion of potentially eligible procedures registered in the PROMs program at participating sites and by eligible surgeons varied widely. Previous studies in this population found that differences in site registration rates are not associated with patient characteristics recorded by the AOANJRR, such that high rates of completeness are not required for a representative sample [22]. Thus, we initially included all eligible procedures from all sites in our analysis. Providers with low registration rates may nevertheless differ meaningfully from those with high registration rates in other respects.

Another limitation is the outcome measures used. Oxford scores record gross levels of function, and a correlation might exist between a surgeon's revision rates and more sensitive PROMs that reveal more subtle functional differences. The timing of outcome data collection may be problematic. Postoperative PROMs were recorded at 6 months, and different results may appear at earlier [6, 35] or later assessment intervals. Recording outcomes at a single timepoint is a limitation of traditional paper-based research. Continuous sensor-based recording of mobility, impact load, asymmetry, and ROM may overcome this limitation, provided clinical significance can be established [4]. Sensor-based recording is likely to be more sensitive than most paper-based PROMs. Technology has advanced rapidly for surgical planning and surgical technique; innovation in outcome measures should follow a similar trajectory.

The number of procedures in this study may be a limitation. Although we identified statistically significant results, our study was powered to detect differences that are not clinically important, so judgement must be applied

when evaluating differences. For example, statistical differences were identified after correlating 2-year CPR and PROMs. These were well below the minimally identifiable clinical difference [4, 11] and "no difference" was reported; more procedures would not change these findings. The inability to explain why a surgeon's 2-year CPR does not correlate with that surgeon's postoperative PROMs is perplexing when surgeon variables are clearly important for CPR [18]. Enhanced patient-, prosthesis-, and surgeon-related data collection could address this question.

Why is There No Correlation Between Revision Rates and Postoperative PROMs?

A negative correlation (higher CPR and lower PROMs) is expected if CPR and PROMs both accurately measure the quality (or success) of surgery. We found, however, no clinical correlation between increasing revision rates at the surgeon level and poorer PROM scores among patients cared for by the same surgeons, after adjusting for many baseline factors.

Our findings are perplexing, given the strength of association between a surgeon's performance and revision rates [27, 44]. Why are two widely accepted measures of the quality of surgery apparently unrelated? First, current PROMs, such as Oxford scores and EQ-VAS, may not detect important differences, especially when very high or very low function are operational—the so-called "floor and ceiling effects." Subtly poor results because of minor pain, occasional stiffness, slight instability, leg length inequality, or ambiguous dissatisfaction may not be discernible with PROMs. This is a contentious issue that new non-PROM scoring systems have attempted to overcome [18, 29, 45]. Apologists for the Oxford Knee Score argue that it does not exhibit ceiling or floor effects. Nonetheless, revision rates may correlate with more sensitive measures of functional outcome.

Second, there is an important threshold phenomenon for some variables. For example, if a surgeon poorly positions acetabular components, revision for instability will only ensue if dislocations recur. All other patients, despite poor-quality positioning of the acetabular component, will enjoy excellent pain relief unless dislocations occur. The poor quality of acetabular component positioning will not reduce PROM scores for most patients, even though more revisions occur. Other surgeons who routinely replace knees with increased laxity in flexion will tend to have patients with above-average scores for flexion, except for those who surpass a threshold and experience flexion instability treated with revision. PROMs record continuous variables that may not indicate a problem until a threshold has been passed.

Third, postoperative PROMs appear to correlate most reliably with preoperative PROMs, suggesting that preoperative patient factors may leave an immutable effect on

surgical results. Most patients benefit from arthroplasty as shown by a general increase in PROMs, but those who start with lower scores seem to have lower scores postsurgery. Patient factors are likely a greater determinant of functional outcome than either surgeon or implant. Numerous patient factors influence the results of surgery. Although satisfaction is a vague measure of the outcome of surgery, it is associated with many patient-related variables [13, 37, 40, 48], including expectations [23]. Despite advances in arthroplasty, the percentage of dissatisfied patients has not decreased appreciably in well-conducted studies over the past two decades. Additionally, dissatisfaction rates of 18% to 23% have been reported for disparate surgical procedures, ranging from unicompartamental to bicompartamental and TKAs [42]. This suggests a patient's disposition or circumstance may exert an overwhelming influence on postoperative satisfaction [20, 30]. Personal experience may influence PROMs scores so heavily that it is difficult to detect other functional differences.

Balanced against the premise that patient factors have the strongest, if not an overwhelming effect on outcome, all surgeons in our study, when divided into quartiles by CPR, cared for a similar cross-section of patients with respect to age, gender, ASA score, and BMI, all of which influence revision rates. These four factors, although significant, are still a small sample of important considerations. Smoking and comorbidities in general affect revision rates, as do the degree of radiologic arthritis, concurrent spine disease, patients' expectations, patient health literacy, socioeconomic status, and mental health status [31, 48]. Patient factors include details of knee pathology that are beyond the scope of this analysis. Some attributes of arthritic knees such as poor preoperative motion, severe deformity, prior incisions, instability, polyarticular involvement, retained metal ware, and extra-articular deformity may independently lead to revision. Variations in the mix of these factors in the practice of an individual surgeon could explain the lack of correlation between PROMs and CPR. Because revision is a relatively uncommon event, even a few procedures could alter a surgeon's CPR independently from general PROMs. On the other hand, an abundance of patients who are at risk of having limited functional results that nonetheless represent improvement and who should not undergo revision could unlink the relationship between PROMs and CPR.

Fourth, could CPR be less valid as a measure of the quality of surgery than we believe? Revision performed, although an incontrovertible expression of failure, is not the whole story. Four permutations of two important variables must be considered: whether revision was performed and whether there was a true problem with the implant and a legitimate indication for revision [50]. Accordingly, we usually construe revision as solving a real problem or being performed for a true indication such as instability. Additionally, revision can be performed for an incorrect

indication (such as pain from the hip or spine in TKA), revision might not have been performed but should have been (such as undiagnosed infection), and revision might not have been performed and should not have been (for example, dissatisfaction related to depression). Because revision rates are low, a few procedures can make a large difference.

Fifth, nonuniform access to revision surgery and variable thresholds for performing revision may create anomalies in the CPR, compromising its reliability as a measure of success. This correlates to situations in which revision is performed although not indicated and in which indicated surgery is not performed. Our PROM data show that primary THA and TKA reliably produce salutary results, with most surgeons achieving similar good outcomes. Whether the original surgeon is confident with revision techniques may influence whether revision will be performed. Conversely, communities with surgeons who are adept at revision surgery may have higher revision rates, quite possibly without there being any worse quality of primary surgery [46]. Accordingly, the success of surgery as measured by PROMs may be relatively uniform, but thresholds for revisions may be inconsistent. Although plausible, this has yet to be tested. Until these concerns have been explained, CPR is probably the most efficient and reliable measure of arthroplasty surgery.

Conclusion

A surgeon's early revision rate, as a surrogate of technical performance, has no clinically important correlation with postoperative Oxford Hip Score, Oxford Knee Score, and EQ-VAS for primary hip and knee arthroplasty performed for osteoarthritis. Missing data may limit the findings of this study, although results were consistent under a variety of different missing data scenarios. When using Oxford scores, no difference in PROMs exists between Australian surgeons. PROMs or revision rates, or both, may be inaccurate or imperfect indicators of successful arthroplasty. Innumerable factors influence arthroplasty results, including patient-related variables, differences in arthroplasty design, and the technical quality of the procedure. PROMs and revision rates may track two different facets of arthroplasty function. Although the surgeon plays an important role in arthroplasty revision rates, the patient may play a more influential role in postoperative PROMs improvement. There are either no real differences in patient function after arthroplasty (except those inherent in the patients themselves), or conventional PROMs cannot detect the differences that exist. This dichotomy can only be investigated with enhanced instruments to measure function. Future research should focus on how detailed surgeon, patient, and prosthesis variables affect postoperative function. Outcome measures that are more sensitive than Oxford scores are probably required. Given measurable differences in CPR

between surgeons and implants, revision rates remain the foundation of registry data collection. Resources currently allocated by registries to PROMs may need to be re-evaluated.

Acknowledgments We thank the AOANJRR as the source of data used in this publication and thank them for the generous assistance provided in the development of the study methodology and data analysis.

References

1. AOANJRR. Hip, knee & shoulder arthroplasty: annual report 2019. Available at <https://aoanjrr.sahmri.com/annual-reports-2019>. Accessed November 14, 2020.
2. Austin PC, Escobar M, Kopec JA. The use of the Tobit model for analyzing measures of health status. *Qual Life Res.* 2000;9:901-910.
3. Beard DJ, Harris K, Dawson J, et al. Meaningful changes for the Oxford hip and knee scores after joint replacement surgery. *J Clin Epidemiol.* 2015;68:73-79.
4. Bolam SM, Batinica B, Yeung TC, et al. Remote patient monitoring with wearable sensors following knee arthroplasty. *Sensors (Basel).* 2021;21:5143.
5. Bowling A. What things are important in people's lives? A survey of the public's judgements to inform scales of health related quality of life. *Soc Sci Med.* 1995;41:1447-1462.
6. Browne JP, Bastaki H, Dawson J. What is the optimal time point to assess patient-reported recovery after hip and knee replacement? A systematic review and analysis of routinely reported outcome data from the English patient-reported outcome measures programme. *Health Qual Life Outcomes.* 2013;11:1-7.
7. Chatterji R, Naylor J, Harris I, et al. An equivalence study: are patient-completed and telephone interview equivalent modes of administration for the EuroQol survey? *Health Qual Life Outcomes.* 2017;15:1-6.
8. Chesher A, Irish M. Residual analysis in the grouped and censored normal linear model. *J Econom.* 1987;34:33-61.
9. Christensen R. ordinal—Regression models for ordinal data. R package version 2022.11-16. 2022. Available at: <https://CRAN.R-project.org/package=ordinal>. Accessed November 22, 2022.
10. Clement N, MacDonald D, Simpson A. The minimal clinically important difference in the Oxford knee score and Short Form 12 score after total knee arthroplasty. *Knee Surg Sports Traumatol Arthrosc.* 2014;22:1933-1939.
11. Clement N, MacDonald D, Simpson A, Burnett R. Total knee replacement in patients with concomitant back pain results in a worse functional outcome and a lower rate of satisfaction. *Bone Joint J.* 2013;95:1632-1639.
12. Cleveland WS, Devlin SJ. Locally weighted regression: an approach to regression analysis by local fitting. *J Am Stat Assoc.* 1988;83:596-610.
13. Connelly JW, Galea VP, Rojanasopondist P, et al. Which preoperative factors are associated with not attaining acceptable levels of pain and function after TKA? Findings from an international multicenter study. *Clin Orthop Relat Res.* 2020;478:1019.
14. Conner-Spady BL, Marshall DA, Bohm E, Dunbar MJ, Noseworthy TW. Comparing the validity and responsiveness of the EQ-5D-5L to the Oxford hip and knee scores and SF-12 in osteoarthritis patients 1 year following total joint replacement. *Qual Life Res.* 2018;27:1311-1322.
15. Dawson J, Fitzpatrick R, Carr A, Murray D. Questionnaire on the perceptions of patients about total hip replacement. *J Bone Joint Surg Br.* 1996;78:185-190.
16. Dawson J, Fitzpatrick R, Murray D, Carr A. Questionnaire on the perceptions of patients about total knee replacement. *J Bone Joint Surg Br.* 1998;80:63-69.
17. de Steiger RN, Hang JR, Miller LN, Graves SE, Davidson DC. Five-year results of the ASR XL acetabular system and the ASR hip resurfacing system: an analysis from the Australian Orthopaedic Association National Joint Replacement Registry. *J Bone Joint Surg Am.* 2011;93:2287-2293.
18. Eckhard L, Munir S, Wood D, et al. The KOOS-12 shortform shows no ceiling effect, good responsiveness and construct validity compared to standard outcome measures after total knee arthroplasty. *Knee Surg Sports Traumatol Arthrosc.* 2021;29:608-615.
19. Fang Z. cenGAM I—Censored regression with smooth terms. R package version 0.5.3. 2017. Available at: <https://CRAN.R-project.org/package=cenGAM>. Accessed November 23, 2022.
20. Gunaratne R, Pratt DN, Banda J, Fick DP, Khan RJ, Robertson BW. Patient dissatisfaction following total knee arthroplasty: a systematic review of the literature. *J Arthroplasty.* 2017;32:3854-3860.
21. Harris IA, Cashman K, Lorimer M, et al. Are responders to patient health surveys representative of those invited to participate? An analysis of the Patient-Reported Outcome Measures Pilot from the Australian Orthopaedic Association National Joint Replacement Registry. *PLoS One.* 2021;16:e0254196.
22. Harris IA, Peng Y, Cashman K, et al. Association between patient factors and hospital completeness of a patient-reported outcome measures program in joint arthroplasty, a cohort study. *J Patient Rep Outcomes.* 2022;6:1-6.
23. Hawker GA, Conner-Spady BL, Bohm E, et al. Patients' preoperative expectations of total knee arthroplasty and satisfaction with outcomes at one year: a prospective cohort study. *Arthritis Rheumatol.* 2021;73:223-231.
24. Heath EL, Ackerman IN, Cashman K, Lorimer M, Graves SE, Harris IA. Patient-reported outcomes after hip and knee arthroplasty: results from a large national registry. *Bone Joint Open.* 2021;2:422-432.
25. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res.* 2011;20:1727-1736.
26. Hermansen LL, Viberg B, Hansen L, Overgaard S. "True" cumulative incidence of and risk factors for hip dislocation within 2 years after primary total hip arthroplasty due to osteoarthritis: a nationwide population-based study from the Danish Hip Arthroplasty Register. *J Bone Joint Surg Am.* 2021;103:295-302.
27. Hoskins W, Rainbird S, Lorimer M, Graves SE, Bingham R. What can we learn from surgeons who perform THA and TKA and have the lowest revision rates? A study from the Australian Orthopaedic Association National Joint Replacement Registry. *Clin Orthop Relat Res.* 2022;480:464-481.
28. Insall JN, Ranawat CS, Aglietti P, Shine J. A comparison of four models of total knee-replacement prostheses. *J Bone Joint Surg Am.* 1976;58:754-765.
29. Jenny J-Y, Louis P, Diesinger Y. High activity arthroplasty score has a lower ceiling effect than standard scores after knee arthroplasty. *J Arthroplasty.* 2014;29:719-721.
30. Kahlenberg CA, Nwachukwu BU, McLawhorn AS, Cross MB, Cornell CN, Padgett DE. Patient satisfaction after total knee replacement: a systematic review. *HSS J.* 2018;14:192-201.
31. Klasan A, Rice DA, Kluger MT, et al. A combination of high preoperative pain and low radiological grade of arthritis is associated with a greater intensity of persistent pain 12 months after total knee arthroplasty. *Bone Joint J.* 2022;104:1202-1208.

32. Novak CB, Mackinnon SE, Stuebe AM. Patient self-reported outcome after ulnar nerve transposition. *Ann Plast Surg.* 2002; 48:274-280.
33. Paulsen A, Roos EM, Pedersen AB, Overgaard S. Minimal clinically important improvement (MCI) and patient-acceptable symptom state (PASS) in total hip arthroplasty (THA) patients 1 year postoperatively: a prospective cohort study of 1,335 patients. *Acta Orthop.* 2014;85:39-48.
34. Polkowski GG, Ruh EL, Barrack TN, Nunley RM, Barrack RL. Is pain and dissatisfaction after TKA related to early-grade preoperative osteoarthritis? *Clin Orthop Relat Res.* 2013;471:162-168.
35. Quah C, Holmes D, Khan T, Cockshott S, Lewis J, Stephen A. The variability in Oxford hip and knee scores in the preoperative period: is there an ideal time to score? *Ann R Coll Surg Engl.* 2017;100:16-20.
36. Ring D, Leopold SS. Measuring satisfaction: can it be done? *Clin Orthop Relat Res.* 2015;473:3071-3073.
37. Rissolito L, Sabatini L, Risitano S, et al. Is it the surgeon, the patient, or the device? A comprehensive clinical and radiological evaluation of factors influencing patient satisfaction in 648 total knee arthroplasties. *J Clin Med.* 2021;10:2599.
38. Robertsson O, Dunbar M, Pehrsson T, Knutson K, Lidgren L. Patient satisfaction after knee arthroplasty: a report on 27,372 knees operated on between 1981 and 1995 in Sweden. *Acta Orthop Scand.* 2000;71:262-267.
39. Rolfson O, Bohm E, Franklin P, et al. Patient-reported outcome measures in arthroplasty registries. *Acta Orthop.* 2016:9-23.
40. Rooks K, Houdek D, Obaid H, Dust W. Primary total knee arthroplasty: correlation between preoperative radiographic severity of arthritis and postoperative patient satisfaction. *J Knee Surg.* 2021;34:1441-1445.
41. Sabah SA, Alvand A, Beard DJ, Price AJ. Minimal important changes and differences were estimated for Oxford hip and knee scores following primary and revision arthroplasty. *J Clin Epidemiol.* 2022;143:159-168.
42. Sabah SA, Knight R, Alvand A, Beard DJ, Price AJ. Early patient-reported outcomes from primary hip and knee arthroplasty have improved over the past seven years: an analysis of the NHS PROMs dataset. *Bone Joint J.* 2022;104:687-695.
43. Sayers A, Whitehouse MR, Judge A, MacGregor AJ, Blom AW, Ben-Shlomo Y. Analysis of change in patient-reported outcome measures with floor and ceiling effects using the multilevel Tobit model: a simulation study and an example from a National Joint Register using body mass index and the Oxford Hip Score. *BMJ Open.* 2020;10:e033646.
44. Siddiqi A, Alamanda V, Barrington J, et al. Effects of hospital and surgeon volume on patient outcomes after total joint arthroplasty: reported from the American Joint Replacement Registry. *J Am Acad Orthop Surg.* 2022;30:e811-e821.
45. Steinhoff AK, Bugbee WD. Knee Injury and Osteoarthritis Outcome Score has higher responsiveness and lower ceiling effect than Knee Society Function Score after total knee arthroplasty. *Knee Surg Sports Traumatol Arthrosc.* 2016;24: 2627-2633.
46. Tay ML, Monk A, Frampton CM, Hooper GJ, Young SW. A comparison of clinical thresholds for revision following total and unicompartmental knee arthroplasty. *Bone Joint J.* 2023;105: 269-276.
47. Tobin J. Estimation of relationships for limited dependent variables. *Econometrica.* 1958:24-36.
48. Tolk JJ, Janssen RP, Haanstra TM, van der Steen M, Bierma-Zeinstra SM, Reijman M. The influence of expectation modification in knee arthroplasty on satisfaction of patients: a randomized controlled trial: the EKSPECT study. *Bone Joint J.* 2021;103:619-626.
49. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *J Stat Softw.* 2011;45: 1-67.
50. Vince K. Rankability and improvability: commentary on an article by Peter van Schie, MD, et al.: "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:e16.
51. WHO Consultation on Obesity. Obesity: preventing and managing the global epidemic: report of a WHO consultation. Available at: <https://apps.who.int/iris/handle/10665/42330>. Accessed January 10, 2023.
52. Wilson I, Bohm E, Lubbeke A, et al. Orthopaedic registries with patient-reported outcome measures. *EFORT Open Rev.* 2019;4: 357-367.
53. Wood SN. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J R Stat Soc Series B Stat Methodol.* 2011;73:3-36.
54. Wood SN, Pya N, Säfken B. Smoothing parameter and model selection for general smooth models. *J Am Stat Assoc.* 2016;111: 1548-1563.
55. Wright RJ, Sledge CB, Poss R, Ewald FC, Walsh ME, Lingard EA. Patient-reported outcome and survivorship after Kinemax total knee arthroplasty. *J Bone Joint Surg Am.* 2004;86: 2464-2470.