Clin Orthop Relat Res (2023) 00:1-11
DOI 10.1097/CORR.0000000000002606

Clinical Orthopaedics and Related Research®
A Publication of The Association of Bone and Joint Surgeons®

Clinical Research

How Much Perioperative Pain and Dysfunction Underlie the HOOS JR and KOOS JR?

Mark E. Cowen MD, SM[1], Huiyong Zheng PhD[2], Richard E. Hughes PhD[3], Patricia D. Franklin MD, MBA, MPH[4], Michael A. Masini MD[1], Brian R. Hallstrom MD[3]

Received: 9 September 2022 / Revised: 9 December 2022 / Accepted: 30 January 2023 / Published online: 14 March 2023

Abstract

Background The Hip Disability and Osteoarthritis Outcome Score Joint Replacement (HOOS JR) and Knee Injury and Osteoarthritis Outcome Score Joint Replacement (KOOS JR) scores represent pain and dysfunction as a single number ranging from 0 (extreme pain and dysfunction) to 100 (no pain or functional limitations). However, scores between 0 and 100 lack a simple interpretation because they reflect varying combinations of pain levels and dysfunction. Given that most adverse events and improvement occur within the first 90 days after surgery, a deeper understanding of the level of pain and dysfunction may reveal missed opportunities for patient care.

Questions/purposes (1) What does a given preoperative or postoperative HOOS JR and KOOS JR score indicate about pain and ability to perform daily activities? (2) How much of a change in score (that is, delta) is needed to indicate significant improvement in pain control and daily functioning?

Methods The Michigan Arthroplasty Registry Collaborative Quality Initiative contains more than 95% of THAs and TKAs performed in Michigan. Between January 2017 and March 2019, 84,175 people in the registry underwent primary THA or TKA and were potentially eligible for this retrospective, comparative study of the first 90 postoperative days.

Support for the Michigan Arthroplasty Registry Collaborative Quality Initiative is provided by Blue Cross and Blue Shield of Michigan and Blue Care Network as part of the BCBSM Value Partnerships program. Although Blue Cross Blue Shield of Michigan and the Michigan Arthroplasty Registry Collaborative Quality Initiative work collaboratively, the opinions, beliefs, and viewpoints expressed by the authors do not necessarily reflect the opinions, beliefs, and viewpoints of BCBSM or any of its employees.
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.
Ethical approval for this study was obtained from the University of Michigan Medical School, Ann Arbor, MI, USA (Study eResearch ID: HUM00159685).
This work was performed at the University of Michigan, Ann Arbor, MI, USA.

[1]St. Joseph Mercy Health System, Ypsilanti, MI, USA

[2]Michigan Arthroplasty Registry Collaborative Quality Initiative, Ann Arbor, MI, USA

[3]University of Michigan, Ann Arbor, MI, USA

[4]Northwestern University Feinberg School of Medicine, Chicago, IL, USA

M. E. Cowen ✉, St. Joseph Mercy Health System, 5301 McAuley Drive, Ypsilanti, MI 48197, USA, Email: cowenme@comcast.net

Wolters Kluwer

Eighty-four percent (70,608 of 84,175) were excluded because their surgeons did not attain a target survey collection proportion of 70% and another 6% (5042) were missing covariate information or surveys, leaving 10% (8525) for analysis. The mean age and percentage of women were 65 ± 11 years and 55% (2060 of 3716), respectively, for patients undergoing THA and 67 ± 9 years and 61% (2936 of 4809), respectively, for those undergoing TKA. There were no clinically meaningful differences between patients who were analyzed and those who were excluded except for lower representation of non-White patients in the analyzed group. For interpretation, patient responses to Question 7 (pain) and Question 6 (function) from the Patient-Reported Outcomes Measurement Information System global items (PROMIS-10) were dichotomized into "much pain" (rating of pain 4 to 10 of 10) versus "less pain" (rating of ≤ 3) and "good function" (able to perform most activities) versus "poor function" (not able to perform most activities) and combined into four pain-function categories. We examined the mean preoperative and postoperative HOOS JR and KOOS JR scores for each pain-function category, adjusted for patient characteristics. We calculated the size of the delta associated with an increase to a more favorable category postoperatively (versus staying in the same or worse category) via multivariable logistic regression that controlled for patient characteristics.

Results Patients in the least favorable "much pain, poor function" category preoperatively had adjusted mean scores of 40 (95% confidence interval 39 to 41) for both the HOOS JR and KOOS JR. Those with mixed levels of pain and function had mean scores between 46 and 55. Those in the most favorable "less pain, good function" category had means of 60 (95% CI 58 to 62) and 59 (95% CI 58 to 61) for the HOOS JR and KOOS JR, respectively. The adjusted delta to achieve a pain level of ≤ 3 or the ability to perform most activities was 30 (95% CI 26 to 36) on the HOOS JR and 27 (95% CI 22 to 29) on the KOOS JR scales.

Conclusion These adjusted means of the HOOS JR and KOOS JR provide context for understanding the levels of pain and dysfunction for individuals as well for patients reported in other studies. Potential quality improvement efforts could include tracking the proportion of patients with THA or TKA who achieved a sufficient delta to attain pain levels of ≤ 3 or the ability to perform most activities. Future studies are needed to understand pain and function represented by the HOOS JR and KOOS JR at 1 to 2 years, how these may differ by patient subgroups, and whether scores can be improved through quality improvement efforts.

Level of Evidence Level III, therapeutic study.

Introduction

The Hip Disability and Osteoarthritis Outcome Score Joint Replacement (HOOS JR) and Knee Injury and Osteoarthritis Outcome Score Joint Replacement (KOOS JR) scores are merely numbers without an understanding of how much pain and dysfunction they represent [22, 23]. For example, the HOOS JR and KOOS JR are scaled from 0 (indicating extreme pain and dysfunction) to 100 (no pain or functional limitations). A Centers for Medicare and Medicaid Services program incentivizes achievement of a 1-year postoperative KOOS JR score of ≥ 71 derived from patient satisfaction [2, 16]. The extent of pain and dysfunction at this level is unclear.

Given that most adverse events and improvement occur within the first 90 days postoperatively, a deeper understanding of the symptoms behind the scores may uncover unrecognized opportunities for improved patient care [34]. Early efforts focused on the clinically important difference or change in score (delta) that represents a change in clinical condition that a patient can feel, usually by comparing scores among groups categorized by an external reference, such as satisfaction or perceived improvement [7, 20, 24]. More recently, attention has focused on defining the score at which patients feel they can live with their level of symptoms, the patient-acceptable symptom state [19]. Yet it is not known how much pain and dysfunction patients have when they attain an "acceptable" HOOS JR or KOOS JR score. Nor is it known whether a change in score that is large enough for a patient to be satisfied or feel improved is large enough to achieve reasonable pain control or the ability to perform most daily activities.

We therefore asked: (1) What does a given preoperative or postoperative HOOS JR and KOOS JR score indicate about pain and ability to perform daily activities? (2) How much of a change in score (that is, delta) is needed to indicate significant improvement in pain control and daily functioning?

Patients and Methods

Study Design and Setting

The Michigan Arthroplasty Registry Collaborative Quality Initiative, funded by Blue Cross and Blue Shield of Michigan/Blue Care Network, captures more than 95% of hip and knee arthroplasties performed in Michigan, regardless of payer [11, 12]. The collaborative goals include a multi-year evaluation of implants and efforts to reduce adverse events during the 90-day postoperative period [11, 13]. Clinical data abstractors collect preoperative and 90-day event data on all elective primary THAs and TKAs and revision procedures, supplemented by administrative data in the Michigan Inpatient Database, which is managed by the Michigan Health and Hospital Association. The HOOS JR and KOOS JR and the Patient-Reported Outcomes

Measurement Information System global items (PROMIS-10) [9] are collected at scheduled intervals and linked to the registry. We report a retrospective, comparative study of patients 18 years or older, who had an elective primary total joint arthroplasty between January 1, 2017, and March 31, 2019 and who did not have a prior joint arthroplasty in the previous 365 days (which might influence survey responses). In keeping with the quality improvement focus of the registry, the time horizon of our study was the first 90 days after surgery, the interval in which most adverse events occur [34]. We also plan to use the same sample in a future study of the impact of adverse events on scores.

Patients

The first eligible surgery was considered for patients undergoing more than one procedure during the study period. We excluded the small number of those who withdrew from the registry or had bilateral procedures, leaving 84,175 patients in the registry who were potentially eligible (Fig. 1). To ensure that survey respondents represented those in a surgery practice, we focused on surgeons who achieved a 70% survey response rate. This resulted in the exclusion of 84% of the starting procedures (70,608 of 84,175). Because of the high data capture rate in the registry, missing covariate and survey data are more of a problem for this study than loss to follow-up [11]. We required a full set of surveys for each patient, including a preoperative and postoperative PROMIS-10 survey completed on the same date as the preoperative and postoperative HOOS JR and KOOS JR. This resulted in additional exclusion of 6% (5042 of 84,175) of patients owing to either missing covariate information or surveys, leaving 10% (8525 of 84,175) for analysis—3716 patients at 29 sites who underwent hip arthroplasty by 55 surgeons and 4809 patients at 28 sites who underwent knee arthroplasty by 69 surgeons. If patients completed more than preoperative one survey, we retained the survey closest to the surgery and the latest within 120 days after surgery. The postoperative survey was completed at a mean of approximately 60 days (Table 1).

Baseline Data

Of the salient characteristics of the study population, the majority were women, White, nonsmokers, overweight, married, and had Medicare as primary insurance (Table 1). A substantial minority had been taking preoperative narcotics. Preoperatively, almost 75% were limited in their daily activities, and almost 90% reported pain of ≥ 4 on a 10-point scale. Because of the potential for transfer or response bias from the study inclusion and exclusion criteria, we compared patient characteristics between surgeons achieving the 70% response rate ("participating surgeons") versus those of their colleagues who did not (Supplemental Table 1; http://links.lww.com/CORR/B36), and between those in the participating surgeon practices with full survey and covariate information and those without (Supplemental Table 2; http://links.lww.com/CORR/B37). We found no clinically meaningful differences between patients who were analyzed and those excluded except for less participation by Black patients (6.2% in participating surgeon practices versus 9.6% in the other surgeons' practices) and patients categorized as

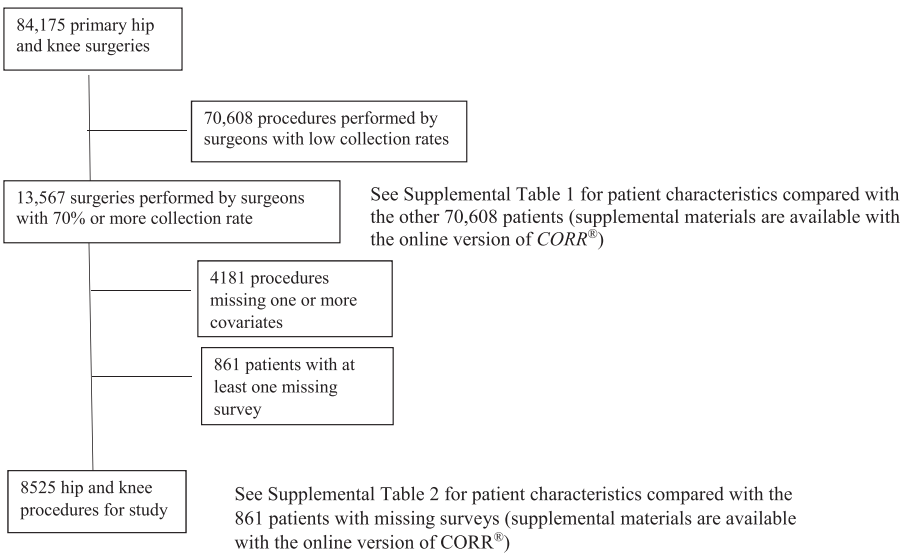

Fig. 1 This flowchart shows eligible patients, exclusions, and study sample.

Table 1. Study sample

Characteristic	Patients who had THA (n = 3716)	Patients who had TKA (n = 4809)
Age in years, mean ± SD	65 ± 11	67 ± 9
Women, % (n)	55 (2060)	61 (2936)
Race, % (n)		
Black	5 (183)	6 (287)
White	93 (3441)	91 (4389)
Other	2 (92)	3 (133)
BMI in kg/m^2, mean ± SD	30 ± 6	33 ± 6
Smoking, % (n)		
Current	10 (386)	8 (369)
Previous	37 (1389)	39 (1894)
Never	52 (1941)	53 (2546)
Married, % (n)	68 (2541)	69 (3322)
Payor, % (n)		
Commercial	28 (1026)	25 (1181)
Medicaid	4 (138)	2 (113)
Medicare	54 (2020)	60 (2873)
Other	14 (532)	13 (642)
Taking preoperative narcotics, % (n)	25 (927)	19 (892)
Number of Elixhauser comorbidities, mean ± SD	1.4 ± 1.2	1.7 ± 1.2
	Preoperative	**Postoperative**	**Preoperative**	**Postoperative**
Days before or after surgery survey was completed, mean ± SD	23 ± 18	63 ± 22	23 ± 18	61 ± 22
HOOS JR score, mean ± SD	47 ± 14	77 ± 15	47 ± 14	67 ± 13
PROMIS-10 Q6: ability to do physical activities				
"Good function" response 4 or 5 (completely or mostly), % (n)	25 (919)	65 (2398)	28 (1357)	59 (2844)
"Poor function" response 1, 2, 3 (not at all, a little, moderately), % (n)	75 (2797)	35 (1318)	72 (3452)	41 (1965)
PROMIS-10 Q7: pain rating, % (n)				
0 to 3 "little pain"	10 (355)	78 (2892)	12 (583)	67 (3241)
4 to 10 "much pain"	90 (3361)	22 (824)	88 (4226)	33 (1568)
Anchor categories, % (n)				
Much pain, poor function	72 (2658)	16 (605)	66 (3191)	23 (1099)
Much pain, good function	19 (703)	6 (219)	22 (1035)	10 (469)
Less pain, poor function	4 (139)	19 (713)	5 (261)	18 (866)
Less pain, good function	6 (216)	59 (2179)	7 (322)	49 (2375)

Percentages may not total 100% owing to rounding.

"other" race (3.1% participating surgeons versus 6.6% of nonparticipating surgeons) (Supplemental Table 1; http://links.lww.com/CORR/B36).

Outcomes Tools and Study Endpoints

The HOOS JR and KOOS JR range from 0 (extreme pain and dysfunction) to 100 (no pain or functional limitations). To answer our first study question regarding the pain and dysfunction represented by the HOOS JR and KOOS JR, we used each patients' responses to PROMIS-10 Question 7 (pain rating for the previous seven days from none to "worst imaginable," 0 to 10) and Question 6 ("To what extent are you able to carry out your everyday physical activities such as walking, climbing stairs, carrying groceries, or moving a chair?"). Among the PROMIS-10 items, these two questions have been shown to have the

strongest correlations with the larger PROMIS health domains of pain impact and physical functioning, and with the EQ-5D, another commonly used quality-of-life survey [9]. We also examined Spearman correlations of the two questions with the individual items and total scores of the HOOS JR and KOOS JR, and found the largest correlations were with the total scores. Based on face validity and ease of interpretation, responses to Question 7 were dichotomized to represent "much pain" (rating of pain between 4 and 10) versus "less pain" (rating of ≤ 3), and similarly for Question 6, "good function" (able to perform tasks completely or mostly) versus "poor function" (moderate or worse ability). We then used the possible permutations to create four categories of pain-function ranging from the least desirable "much pain, poor function" to the most favorable "less pain, good function." To answer the second study question, we defined improvement as attainment of a more favorable pain–function level postoperatively, with either pain levels improved to ≤ 3 on a scale of 10 or the ability to perform at least most daily activities. Otherwise, patients remaining in the same category or declining to a less favorable category postoperatively were considered not to have improved, except for those who began and remained in the most favorable preoperative category of "less pain, good function" who were deemed improved. We examined this assumption with a planned sensitivity analysis and found little impact on the results.

Ethical Approval

Ethical approval for this study was obtained from the University of Michigan Medical School Institutional Review Board (study eResearch ID: HUM00159685).

Statistical Analysis

We used SAS Version 9.4 (SAS Institute) for all analyses. We adjusted the results by age, gender, race, BMI, hemoglobin level, creatinine level, smoking and marital status, primary insurance, comorbidities, use of preoperative narcotics, and time survey was completed. These were chosen on a conceptual basis rather than based on a p value. We report adjusted means and confidence intervals. We used multivariable logistic regression models with binary improvement as the dependent variable and the change in HOOS JR and KOOS JR scores (the delta score) as the explanatory variable, controlling for patient characteristics. To interpret our adjusted results, our reference group was White, women (that is, most of the study population), nonsmoker, married, Medicare primary insurance, with no comorbidities and no use of preoperative narcotics, along with the cohort mean age, BMI, hemoglobin level, creatinine level, and time the survey was completed (Supplemental Table 3; http://links.lww.com/CORR/B38). We determined the threshold for improvement as the delta, providing the highest sensitivity plus specificity for distinguishing patients who achieved a more favorable pain-function category [33, 35]. We did not calculate the optimal delta for patient subgroups because our analytic plan was not designed for this purpose. The strength of the delta-improvement association was measured with the area under the receiver operating curve (AUROC) of the regression models, sometimes referenced as the ROC method, with a value above 0.70 deemed acceptable [31, 32]. The confidence bands around the optimal delta, probability, sensitivity, specificity, and AUROC were calculated via bootstrapping with 5000 replicates (Supplemental Table 4; http://links.lww.com/CORR/B39). Risk-adjusted smoothed spline curves with generalized additive models for the improved versus unimproved groups were plotted on graphs showing risk-adjusted deltas per preoperative scores [8, 21].

Sample Size

Because the second of our planned studies was to examine how adverse events impacted the HOOS JR and KOOS JR, we powered our study sample based on its stricter hypothesis. We found that a sample of 3861 patients who underwent hip arthroplasty and 2965 patients who underwent knee arthroplasty would be adequate for an effect size of 0.4 to detect differences between those who had adverse events and those who did not (Supplemental Digital Content 1; http://links.lww.com/CORR/B40).

Results

Relationship of Preoperative or Postoperative HOOS JR or KOOS JR With Pain and Daily Activities

The adjusted mean score for patients having "much pain, poor function" preoperatively (pain rated as a ≥ 4 on a scale of 10 with the inability to do most of their daily activities) was 40 (95% CI 39 to 41) for both the HOOS JR and KOOS JR. Those with mixed levels of pain and function, that is, either pain ≥ 4 or the inability to perform most daily activities but not both, had adjusted means between 46 and 55. Those in the most favorable "less pain, good function" category (pain ≤ 3 and the ability to perform all or most of their activities) had adjusted means of 60 (95% CI 58 to 62) for the HOOS JR (Supplemental Table 5; http://links.lww.com/CORR/B41) and 59 (95% CI 58 to 61) for the KOOS JR (Supplemental Table 6; http://links.lww.com/CORR/B42). Preoperatively, most patients undergoing either hip and knee surgery were in

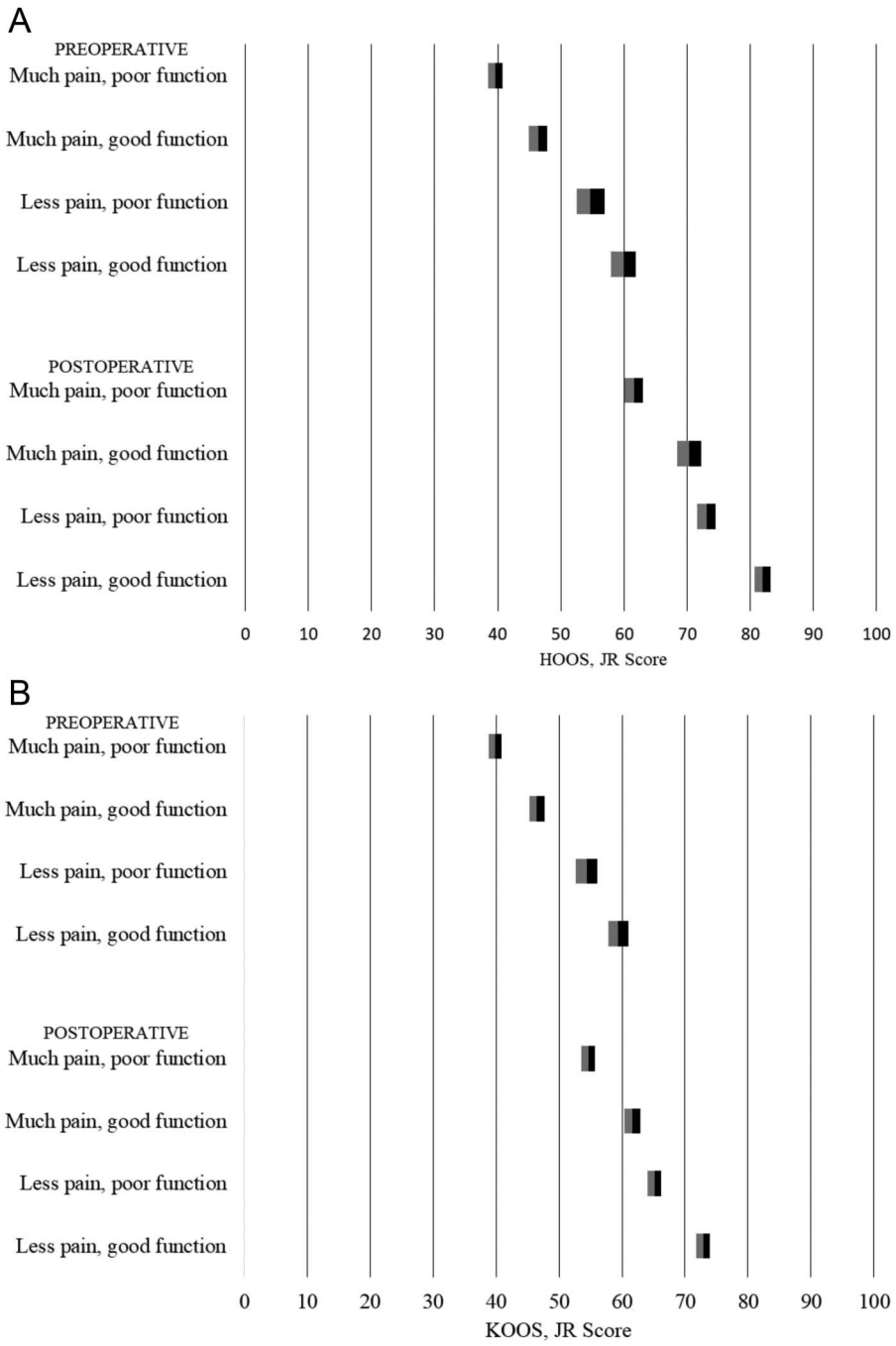

Fig. 2 These figures show the risk-adjusted mean and 95% CIs for the (**A**) HOOS JR and (**B**) KOOS JR per pain and function categories. The vertical axis displays the preoperative and postoperative pain and function categories. The horizontal axis shows the possible range in scores of the preoperative and postoperative HOOS JR and KOOS JR. The range of the 95% CI is displayed in gray and black, with the boundary indicating the mean value.

the "much pain, poor function" category (Table 1). Postoperatively, most hip patients and the plurality of knee patients were in the "less pain, good function" group (Table 1). Although the preoperative and postoperative pain-function categories were defined using the same cutoff points, the postoperative HOOS JR and KOOS JR scores were higher than preoperative scores (Fig. 2A and 2B).

Magnitude of Change in Score to Reflect Improvement in Pain and Function

The number of adjusted points associated with improvement in pain or function was 30 (95% CI 26 to 36) on the HOOS JR and 27 (95% CI 22 to 29) on the KOOS JR scales. Patients attaining a more favorable pain-function category had larger deltas than those who did not, across a wide range of preoperative scores (Figs. 3A and 3B). Patient characteristics affected the probability of improvement differently (Supplemental Table 3; http://links.lww.com/CORR/B38). The adjusted AUROCs were 76 (95% CI 75 to 77) for both THA and TKA (Supplemental Table 4; http://links.lww.com/CORR/B39). The corresponding AUROCs without risk adjustment were smaller: 69 (95% CI 68 to 71) and 70 (95% CI 69 to 71). A total of 79% (2941 of 3716) of the hip group and 68% (3302 of 4809) of the knee group transitioned to a more favorable category, but 24% of patients with TKA who began in the favorable "less pain, good function" category declined to a worse category (Table 2).

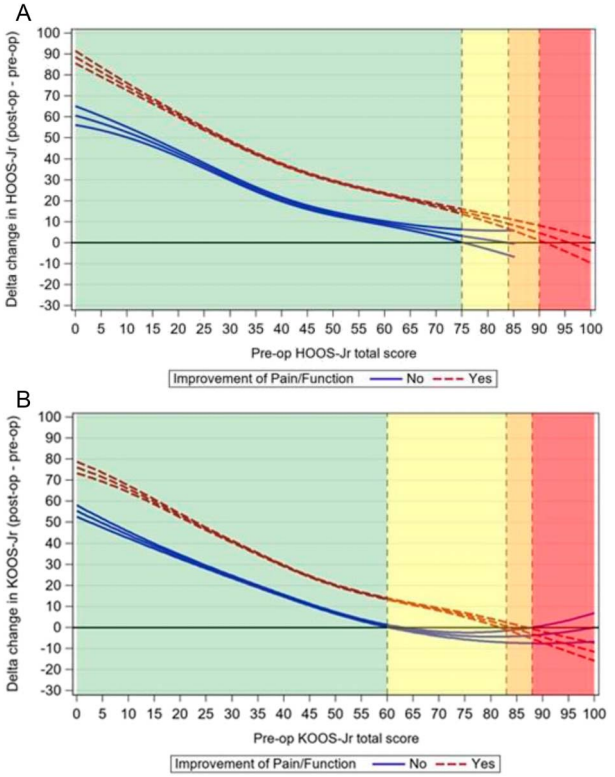

Fig. 3 These risk-adjusted smoothed spline curves display the postoperative change in score (delta, vertical axis) according to the preoperative (**A**) HOOS JR or (**B**) KOOS JR score (horizontal axis). The upper curve represents the 95% confidence boundaries of the risk-adjusted mean delta score of patients who achieved a more favorable pain and function category. The lower curve represents patients who did not achieve either a pain level ≤ 3 or the ability to perform most daily activities. These are mean scores, so the relationships may be different for an individual patient. The green shaded region indicates both improved and unimproved groups had a positive delta; the yellow indicates only the improved group had a positive delta and that the 95% CIs did not overlap; the orange shading indicates the improved group had a nonsignificant delta (0 included within the 95% CI) while the nonimprovement group had a negative delta (entire 95% CI < 0); the red shading indicates both groups had negative deltas (entire 95% CI < 0). The confidence bands are wider at the extremes of the HOOS JR and KOOS JR owing to the small number of patients with scores in these regions.

Discussion

An enhanced appreciation of what the HOOS JR or KOOS JR scores mean may provide additional insights into what patients are experiencing and indicate areas where patient care might be improved. We used two simple pain and function questions from the PROMIS-10 to understand the pain and dysfunction of 8525 patients in the Michigan Arthroplasty Registry Collaborative Quality Initiative according to their HOOS JR and KOOS JR scores. Preoperative scores of approximately 40 represented high levels of pain and problems performing daily activities. Scores between 46 and 55 indicated either high levels of pain or difficulties with daily functioning, but not both. Scores of approximately 60 indicated low levels of pain and the ability to perform most daily activities. Postoperative scores were higher in the postoperative pain-function categories. An increase of 30 HOOS JR or 27 KOOS JR adjusted points after surgery was associated with attaining a more favorable pain-function category.

Limitations

The most important limitation is the possibility of transfer bias given the relatively short follow-up period and small percentage of total joint arthroplasties that were studied. The primary justification for not examining surveys at 1 to 2 years is that our study was in keeping with the Michigan Arthroplasty Registry Collaborative Quality Initiative's clinical quality focus to improve processes and outcomes of care in the 90-day perioperative period, the time during which most adverse events occur [13, 34]. Although the 90-day interval cannot replace a 2-year endpoint, it is difficult to anticipate how interpretations of the scores would be meaningfully different based on our method. Unlike the potential time dependency of a patient-acceptable symptom state based on satisfaction or willingness to have the surgery again, our comparisons of

Table 2. Transition matrix for each preoperative level of pain and function and the percentage of patients in a postoperative category

	Patients undergoing hip arthroplasty				
		Percentage of patients in preoperative category who transitioned to a postoperative category			
	Preoperative category	Postoperative much pain, poor function	Postoperative much pain, good function	Postoperative less pain, poor function	Postoperative less pain, good function
Preoperative much pain, poor function	72 (2658)	20 (534)	6 (161)	23 (601)	51 (1362)
Preoperative much pain, good function	19 (703)	8 (55)	8 (53)	9 (60)	76 (535)
Preoperative less pain, poor function	4 (139)	7 (10)	2 (3)	22 (31)	68 (95)
Preoperative less pain, good function	6 (216)	3 (6)	1 (2)	10 (21)	87 (187)
	Patients undergoing knee arthroplasty				
		Percentage of patients in preoperative category who transitioned to a postoperative category			
	Preoperative category	Postoperative much pain, poor function	Postoperative much pain, good function	Postoperative less pain, poor function	Postoperative less pain, good function
Preoperative much pain, poor function	66 (3191)	29 (924)	9 (273)	20 (654)	42 (1340)
Preoperative much pain, good function	22 (1035)	12 (127)	16 (166)	11 (110)	61 (632)
Preoperative less pain, poor function	5 (261)	10 (25)	5 (12)	25 (66)	61 (158)
Preoperative less pain, good function	7 (322)	7 (23)	6 (18)	11 (36)	76 (245)

Data provided as % (n). We dichotomized the responses to the PROMIS-10 item 6 ("To what extent are you able to carry out your everyday physical activities such as walking, climbing stairs, carrying groceries, or moving a chair?") into categories of "good function" (indicating "mostly" and "completely" able to perform daily activities) and "poor function" ("not at all," "a little," and "moderately"). Similarly, we divided the PROMIS-10 pain rating scale (Question 7) into two categories: "less pain" (rating of 0 to 3) and "much pain" (rating between 4 and 10). The resulting two pain and two function categories were combined into the pain-function composite.

HOOS JR and KOOS JR scores and pain-function category were based on surveys completed on the same day [16]. Still, a future study with longer follow-up is needed. With the goal of quality improvement, important information can be gleaned from the 90-day period, the steep portion of the trajectory in which approximately 80% of improvement in pain and function occurs [5]. In another study, the delta of the WOMAC at 3 months was the strongest predictor of scores at 2 years, surpassing prognostic information from comorbidities and baseline scores [5]. Similarly, postoperative pain levels at 2 months were found to predict pain levels at 6 and 12 months [30]. Theoretically, this timepoint provides an early opportunity for quality improvement projects, but future studies are needed to demonstrate this [5, 30]. The second component of transfer bias is the relatively small percentage of eligible patients in our analysis, primarily because of our focus on surgeons who achieved a robust survey collection rate. Based on comparisons of the patient characteristics of the study sample and a larger Michigan Arthroplasty Registry Collaborative Quality Initiative population, it appears our results are representative. The only potentially meaningful difference was in the smaller proportion of non-White participants. It is difficult to anticipate how the interpretation of pain and dysfunction would be different in a broader population, but it is conceivable [19]. Although tempting, we refrained from providing estimates for patient subgroups because this was not our focus. However, others have begun this work [4, 15, 19].

The next limitation concerns how we defined clinical improvement. Our pain-function categories and improvement defined as attaining a more favorable category were based on face validity and ease of interpretation. Others could define improvement differently. The strength of the association of improvement as we defined with HOOS JR and KOOS JR scores was modest at best, with the CIs of the AUROC for unadjusted models encompassing the threshold for an acceptable anchor of 0.7 [31]. This may reflect a lack of psychometric precision when there is only one question each for pain and function, despite their validity, rather than multiple questions [9, 10]. More importantly, the larger AUROCs with adjusted analyses indicate there are multiple influences on a patient's perception of pain and function over and above a change in HOOS JR and KOOS JR scores. The scores by intent focus on movements and pain in the joint and less on how joint function impacts the whole person with different needs regarding daily activities and sense of comfort. We controlled for available patient characteristics, but there are other influences we did not include; for example, baseline physical function, the nonaffected joint, spine problems, other comorbidities, patient resilience, surgeon effect, or differing lifestyles (Supplemental Table 3; http://links.lww.com/CORR/B38) [1, 3, 28, 29]. This limitation underscores an important take-home point: Patient characteristics must be considered, even informally, to understand how the change in HOOS JR and KOOS JR score reflects a patient's pain and dysfunction [4]. Other analytic approaches could yield different deltas, although the ROC method we used is commonly used [25, 26, 32]. Research provides general support for our approach. Our definition of "less pain" is consistent with a reported patient-acceptable pain level of 2.5 with activities [26]. Pain and function have been included in a composite reference anchor, but with different questions than we used and not applied to the HOOS JR and KOOS JR [6, 27]. Others have reported the relationship between KOOS JR and the PROMIS physical functions scores but without defining categories [3, 17, 18]. For now, readers should consider a range of values for the delta associated with improvement [7, 20].

Relationship of Preoperative or Postoperative HOOS JR or KOOS JR With Pain and Daily Activities

Preoperative scores near 40 represented high levels of pain (≥ 4 on a scale of 10) and difficulties performing most daily activities. Scores between 46 and 55 indicated problems with pain or dysfunction but not both, although pain appeared to have the dominant effect (for example, "much pain or good function" scored lower than "less pain or poor function"). Preoperative scores around 60 generally indicated pain levels ≤ 3 plus the ability to perform most daily activities (Figs. 2A and 2B). This provides context for understanding an individual's scores as well as those in published studies (for example, compare the mean preoperative HOOS JR at a Veterans Health Administration hospital versus that at a specialty hospital, 42 versus 51, respectively [19, 20]). There is also prognostic information. We found that 24% of patients with TKA who started in the most favorable category preoperatively had worse levels of pain or dysfunction at 60 days after surgery (Table 2). This supports ongoing efforts to develop protocols and decision aids for discernment for such patients, perhaps those with a baseline KOOS JR score above 60 [4, 15]. However, as described earlier, the KOOS JR alone has only a modest correlation with pain and function, suggesting a score is insufficient for determining appropriateness. Postoperative scores were higher in the postoperative pain-function categories, presumably reflecting some improvement in HOOS JR and KOOS JR scores but not enough for a patient to improve to a more favorable group (Figs. 3A and 3B). Our results may also be useful for interpreting target scores suggested by others. For example, compare the threshold used by the Centers for Medicare and Medicaid Services for the Merit-based Incentive Payment System Measure of a postoperative KOOS JR score ≥ 71 [2]. Patients at this score or above likely fall into our most favorable postoperative category, a clearly favorable outcome. However, the interpretation is more mixed for another reported patient-acceptable symptom state threshold of 63.7 derived at a specialty hospital [19]. Some of our study group at this threshold could still have substantial problems with pain or dysfunction (Fig. 2B). This also applies, to a lesser extent, for a reported patient-acceptable symptom state threshold of 76.7 for the HOOS JR (Fig. 2A) [19].

Magnitude of Change in Score to Reflect Improvement in Pain and Function

We found that improvement to attain pain levels of ≤ 3 or the ability to perform most activities was associated with an increase of 30 HOOS JR or 27 KOOS JR points, controlling for patient characteristics. Patients with smaller deltas may have persisting pain and dysfunction and may benefit from the surgeon knowing why and what to do, with pain likely to be more of a problem than function (Table 2). Quality improvement efforts at the practice, hospital, or health plan level could be directed to improving the proportion of patients with THA and TKA who achieve a delta, for example, of 26 and 22 (the lower confidence boundary of our optimal deltas), respectively, at 90 days postoperatively. Our deltas provided the highest sum of sensitivity (ability to identify true positive

improvement) plus specificity (true lack of improvement; Supplemental Digital Content 1; http://links.lww.com/CORR/B40). Others have reported smaller deltas using different definitions of improvement. Hung et al. [14] found there were approximately 8 points for patient-perceived improvement. Lyman et al. [24] reported that patients who felt moderately improved scored 18 points higher on the HOOS JR and 14 points higher on the KOOS JR than patients reporting little or no improvement. Patients noting "substantial" improvement improved by 22 points on the HOOS JR and 20 points on the KOOS JR [24]. Kuo et al. [20] found that patients who were somewhat or very satisfied with their results scored 18 unadjusted points higher on the HOOS JR and 21 on the KOOS JR than those who were dissatisfied. These suggest that lower deltas may be sufficient to indicate patient satisfaction and sense of improvement but not high enough to assume adequate pain control and performance of daily activities. Readers can choose the endpoint and associated delta that best serves their clinical improvement activities [7].

Conclusion

The potential value of HOOS JR and KOOS JR scores for patient care requires an interpretation of pain and dysfunction in the context of comorbidities and lifestyle demands, followed by quality improvement efforts. Future studies are needed to understand the pain and function represented by the HOOS JR and KOOS JR at 1 to 2 years, how these may differ by patient subgroups, and how scores can be improved through quality improvement efforts. Without interpretation, HOOS JR and KOOS JR scores are merely numbers that veil the pain and dysfunction that might be addressed earlier.

Acknowledgments We thank Brian R. McCardel MD, who chaired the Michigan Arthroplasty Registry Collaborative Quality Initiative PROS Committee. We also thank the Michigan Arthroplasty Registry Collaborative Quality Initiative Coordinating Center staff who supported the sites collecting data.

References

1. Ayers DC, Li W, Oatis C, Rosal MC, Franklin PD. Patient-reported outcomes after total knee replacement vary on the basis of preoperative coexisting disease in the lumbar spine and other nonoperatively treated joints. *J Bone Joint Surg Am.* 2013;95:1833-1837.
2. Centers for Medicare and Medicaid Services. Overview of CJR quality measures, composite quality score, and pay-for-performance methodology. Available at: https://innovation.cms.gov/Files/x/cjr-qualsup.pdf. Accessed April 14, 2022.
3. Darrith B, Khalil LS, Franovic S, et al.Preoperative Patient-Reported Outcomes Measurement Information System global health scores predict patients achieving the minimal clinically important difference in the early postoperative time period after total knee arthroplasty. *J Am Acad Orthop Surg.* 2021;29:e1417-e1426.
4. Franklin PD, Zheng H, Bond C, Lavallee DC. Translating clinical and patient-reported data to tailored shared decision reports with predictive analytics for knee and hip arthritis. *Qual Life Res.* 2021;30:3171-3178.
5. Gandhi R, Mahomed NN, Cram P, Perruccio AV. Understanding the relationship between 3-month and 2-year pain and function scores after total knee arthroplasty for osteoarthritis. *J Arthroplasty.* 2018;33:1368-1372.
6. Giesinger JM, Hamilton DF, Jost B, Behrend H, Giesinger K. WOMAC, EQ-5D and Knee Society Score thresholds for treatment success after total knee arthroplasty. *J Arthroplasty.* 2015;30:2154-2158.
7. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR; Clinical Significance Consensus Meeting Group. Methods to explain the clinical significance of health status measures. *Mayo Clin Proc.* 2002;77:371-383.
8. Hastie TJ, Tibshirani RJ. *Generalized Additive Models*. Chapman and Hall/CRC; 1990.
9. Hays RD, Bjorner JB, Revicki DA, Spritzer KL, Cella D. Development of physical and mental health summary scores from the Patient-Reported Outcomes Measurement Information System (PROMIS) global items. *Qual Life Res.* 2009;18:873-880.
10. Hays RD, Reise S, Calderón JL. How much is lost in using single items? *J Gen Intern Med.* 2012;27:1402-1403.
11. Hughes RE, Hallstrom BR, Cowen ME, Igrisan RM, Singal BM, Share DA. Michigan Arthroplasty Registry Collaborative Quality Initiative (MARCQI) as a model for regional registries in the United States. *Orthop Res Rev.* 2015;7:47-56.
12. Hughes RE, Hallstrom BR, Zheng H, et al. Michigan Arthroplasty Registry Collaborative Quality Initiative (MARCQI) report: 2012-2016. Available at: https://marcqi.org/marcqi-registry-reports-marcqi-annual-reports/. Accessed October 31, 2022.
13. Hughes RE, Zheng H, Igrisan RM, Cowen ME, Markel DC, Hallstrom BR. The Michigan Arthroplasty Registry Collaborative Quality Initiative experience: improving the quality of care in Michigan. *J Bone Joint Surg Am*. 2018;100:e143.1-7.
14. Hung M, Bounsanga J, Voss MW, Saltzman CL. Establishing minimum clinically important difference values for the Patient-Reported Outcomes Measurement Information System Physical Function, Hip Disability and Osteoarthritis Outcome Score for Joint Reconstruction, and Knee Injury and Osteoarthritis Outcome Score for Joint Reconstruction in orthopaedics. *World J Orthop.* 2018;9:41-49.
15. Jayakumar P, Moore MG, Furlough KA, et al. Comparison of an artificial intelligence-enabled patient decision aid vs educational material on decision quality, shared decision-making, patient experience, and functional outcomes in adults with knee osteoarthritis: a randomized clinical trial. *JAMA Netw Open*. 2021;4:e2037107.
16. Keurentjes JC, Van Tol FR, Fiocco M, et al. Patient acceptable symptom states after total hip or knee replacement at mid-term follow-up: thresholds of the Oxford Hip and Knee Scores. *Bone Joint Res*. 2014;13:7-13.
17. Khalil LS, Darrith B, Franovic S, Davis JJ, Weir RM, Banka TR. Patient-Reported Outcomes Measurement Information System (PROMIS) Global Health short forms demonstrate responsiveness in patients undergoing knee arthroplasty. *J Arthroplasty*. 2020;35:1540-1544.
18. Kortlever JTP, Leyton-Mange A, Keulen MHF, et al. PROMIS Physical Function correlates with KOOS, JR in patients with knee pain. *J Knee Surg*. 2020;33:903-911.
19. Kunze KN, Fontana MA, MacLean CH, Lyman S, McLawhorn AS. Defining the patient acceptable symptom state for the H/KOOS, JR after primary total joint arthroplasty. *J Bone Joint Surg Am*. 2022;104:345-352.

20. Kuo AC, Giori NJ, Bowe TR, et al. Comparing methods to determine the minimal clinically important differences in patient-reported outcome measures for veterans undergoing elective total hip or knee arthroplasty in Veterans Health Administration hospitals. *JAMA Surg*. 2020;155:404-411.
21. Liew J, Forkman J. A guide to generalized additive models in crop science using SAS and R. *Commun Biometry Crop Sci.* 2015;10:41-57.
22. Lyman S, Lee Y-Y, Franklin PD, Li W, Cross MB, Padgett DE. Validation of the KOOS, JR: a short-form knee arthroplasty outcomes survey. *Clin Orthop Relat Res.* 2016;474:1461-1471.
23. Lyman S, Lee Y-Y, Franklin PD, Li W, Mayman DJ, Padgett DE. Validation of the HOOS, JR: a short-form hip replacement survey. *Clin Orthop Relat Res.* 2016;474:1472-1482.
24. Lyman S, Lee Y-Y, McLawhorn AS, Islam W, MacLean CH. What are the minimal and substantial improvements in the HOOS and KOOS and JR versions after total joint replacement? *Clin Orthop Relat Res.* 2018;476:2432-2441.
25. Molino J, Harrington J, Racine-Avila J, Aaron R. Deconstructing the minimum clinically important difference (MCID). *Orthop Res Rev.* 2022;14:35-42.
26. Most J, Hoelen TA, Spekenbrink-Spooren A, Schotanus MGM, Boonen B. Defining clinically meaningful thresholds for patient-reported outcomes in knee arthroplasty. *J Arthroplasty.* 2022;37:837-844.e3.
27. Nguyen UD, Ayers DC, Li W, Harrold LR, Franklin PD. Preoperative pain and function: profiles of patients selected for total knee arthroplasty. *J Arthroplasty.* 2016;31:2402.
28. Nwankwo VC, Jiranek WA, Green CL, Allen KD, George SZ, Bettger JP. Resilience and pain catastrophizing among patients with total knee arthroplasty: a cohort study to examine psychological constructs as predictors of postoperative outcomes. *Health Qual Life Outcomes.* 2021;19:136.
29. Sinclair ST, Klika AK, Jin Y, Piuzzi NS, Cleveland Clinic OME Arthroplasty Group, Higuera CA. The impact of surgeon variability on patient-reported outcomes in total hip arthroplasty. *J Arthroplasty.* 2022;37:S479-S487.e1.
30. Singh JA, Lemay CA, Nobel L, et al. Association of early postoperative pain trajectories with longer-term pain outcome after primary total knee arthroplasty. *JAMA Netw Open*. 2019;2:e1915105
31. Terwee CB, Bot SDM, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol.* 2007;60:34-42.
32. Turner D, Schunemann HJ, Griffith LE, et al. Using the entire cohort in the receiver operating characteristic analysis maximizes precision of the minimal important difference. *J Clin Epidemiol*. 2009;62:374-379.
33. Turner D, Schunemann HJ, Griffith LE, et al. The minimal detectable change cannot reliably replace the minimal important difference. *J Clin Epidemiol.* 2010;63:28-36.
34. Yale New Haven Health Services Corporation/Center for Outcomes Research & Evaluation (YNHHSC/CORE). Hospital-level risk-standardized complication rate following elective primary total hip arthroplasty (THA) and/or total knee arthroplasty (TKA) measure methodology report. Available at: https://www.qualitynet.org/inpatient/measures/complication/methodology. Accessed November 11, 2020.
35. Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950;3:32-35.