

Do Surgeon Expectations Predict Clinically Important Improvements in WOMAC Scores After THA and TKA?

Hassan M. K. Ghomrawi PhD, MPH, Carol A. Mancuso MD,
Allison Dunning MS, Alejandro Gonzalez Della Valle MD, Michael Alexiades MD,
Charles Cornell MD, Thomas Sculco MD, Matthias Bostrom MD,
David Mayman MD, Robert G. Marx MD, Geoffrey Westrich MD,
Michael O'Dell MD, Alvin I. Mushlin MD

Received: 16 August 2016 / Accepted: 21 March 2017
© The Association of Bone and Joint Surgeons® 2017

Abstract

Background Failure of THA or TKA to meet a patient's expectations may result in patient disappointment and litigation. However, there is little evidence to suggest that surgeons can consistently anticipate which patients will benefit from those interventions.

The institution of one or more of the authors (HMKG) has received, during the study period, funding from the National Institutes of Health (NICHD grant # R00HD060686) to support this research.

Each author certifies that neither he or she, nor any member of his or her immediate family, have 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.

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.

Each author certifies that his or her institution approved the human protocol for this investigation, that all investigations were conducted in conformity with ethical principles of research, and that informed consent for participation in the study was obtained.

This work was performed at the Feinberg School of Medicine, Chicago, IL, USA, and Weill Cornell Medical College and Hospital for Special Surgery, New York, NY, USA.

Electronic supplementary material The online version of this article (doi:10.1007/s11999-017-5331-8) contains supplementary material, which is available to authorized users.

H. M. K. Ghomrawi (✉)
Department of Surgery, Feinberg School of Medicine,
Northwestern University, 20th Floor, 633 N Saint Claire,
Chicago, IL 60611, USA
e-mail: hassan.ghomrawi1@northwestern.edu

H. M. K. Ghomrawi
Department of Pediatrics, Feinberg School of Medicine,
Northwestern University, Chicago, IL, USA

Questions/purposes To determine the ability of surgeons to identify, in advance of surgery, patients who will benefit from THA or TKA and those who will not, where 'benefit' is defined as a clinically important improvement in a validated patient-reported outcomes score.

Methods In this prospective study, eight high-volume orthopaedic surgeons completed validated THA and TKA expectations questionnaires (score 0–100, 100 being the highest expectation) as part of preoperative assessment of all their patients scheduled for a THA or TKA and enrolled in the Hospital for Special Surgery institutional registry. Enrolled patients completed the WOMAC preoperatively and at 2 years. Successful outcomes were defined as achieving the minimum clinically important difference

H. M. K. Ghomrawi
Center for Healthcare Studies, Feinberg School of Medicine,
Northwestern University, Chicago, IL, USA

C. A. Mancuso
Department of Medicine, Hospital for Special Surgery, New
York, NY, USA

A. Dunning
Duke University, Durham, NC, USA

A. Gonzalez Della Valle, M. Alexiades, C. Cornell, T. Sculco,
M. Bostrom, D. Mayman, R. G. Marx, G. Westrich
Department of Orthopedics, Hospital for Special Surgery, New
York, NY, USA

M. O'Dell
Department of Rehabilitation Medicine, Weill Cornell Medical
College, New York, NY, USA

A. I. Mushlin
Departments of Healthcare Policy and Research and Medicine,
Weill Cornell Medical College, New York, NY, USA

(MCID) in WOMAC pain and function subscales. Sensitivity, specificity, and receiver operating characteristic (ROC) curves were used to evaluate the ability of surgeons' expectation scores to identify patients likely to achieve the MCID on the WOMAC scale. Analyses were run separately for patients having THA and TKA. We enrolled 259 patients undergoing THA and 247 undergoing TKA, of whom 77% ($n = 200$) and 77% ($n = 191$) completed followup surveys 2 years after their procedures, respectively.

Results Surgeons' expectation scores effectively anticipated patients who would improve after THA, but they were no better than chance in identifying patients who would achieve the MCID on the WOMAC score 2 years after TKA. For patients having THA, the areas under the ROC curve were 0.67 (95% CI, 0.53–0.82; $p = 0.02$) and 0.74 (95% CI, 0.63–0.85; $p < 0.01$) for WOMAC function and pain outcomes, respectively, indicating good accuracy. Sensitivity and specificity were maximized on WOMAC pain and function scores (sensitivity = 0.69, specificity = 0.72, both for pain and function) at an expectations score of 83 or greater of 100. Surgeons' expectations were more accurate for patients who were men, who had a BMI less than 30 kg/m², who had more than one comorbidity, and who were older than 65 years. For patients having TKA, surgeons' expectation scores were not better than chance for identifying those who would experience a clinically important improvement on the WOMAC scale (area under ROC curve: Function = 0.51, [95% CI, 0.42–0.61], $p = 0.78$; Pain = 0.51, [95% CI, 0.40–0.61], $p = 0.92$).

Conclusions Most patients having THA and TKA achieved the MCID improvement after surgery. However, the inability of surgeons' expectation scores to discriminate accurately between patients who benefit and those who do not among patients scheduled for THA who are young, with no comorbidities, and with elevated BMIs, and among all patients scheduled for TKA, calls for surgeons to spend more time with these patients to fully understand and address their needs and expectations. Using standardized assessment tools to compare surgeons' expectations and those of their patients may help focus the surgeon-patient discussion further, and address patients' expectations more effectively.

Level of Evidence Level II, therapeutic study.

Introduction

Although THA and TKA have proven clinically effective and cost-effective [4, 9], their elective nature highlights the importance of addressing patients' expectations, because helping patients meet those expectations is the ultimate goal of elective surgery. Failing to do so leaves patients

disappointed, unsatisfied, and perhaps litigious [19, 21, 22]. Evidence from a randomized trial showed that expectations can be modified through preoperative educational interventions [20]. It therefore is reasonable to expect that adjusting patients' expectations to be closer to reality will favorably affect their postoperative course, leading to improved outcomes and greater satisfaction. However, benchmarks representing realistic goals for patients do not currently exist.

Surgeons' expectations have the potential to serve as benchmarks for patients to use in setting realistic expectations for themselves. Surgeons attempt to realistically set patients' expectations based on their clinical knowledge, expertise, and experience. However, despite such efforts to provide genuine informed consent, some patients may still have elective surgery with higher (or sometimes lower) expectations than do their surgeons. In previous studies, it was shown that patients and surgeons differ in their preoperative expectations for THA and TKA and that a preoperative educational program helps to decrease this difference [10, 11]. Studies are needed to empirically validate surgeons' expectations as a potential prognostic tool by establishing that they are associated with actual patient outcomes after THA and TKA. We found only one study that examined this relationship; it showed no association between surgeons' preoperative expectations and the Knee Society functional rating score [14] at 1 year after surgery; however, this no-difference finding may have been the result of that study's sample size (53 patients) being insufficient to detect a difference that might have been present [25].

In the current study, we aimed to determine the ability of surgeons to identify, in advance of surgery, patients who will benefit from THA or TKA and those who will not, where 'benefit' is defined as a clinically important improvement in a validated patient-reported outcomes score.

Patients and Methods

This was a prospective study nested in the institutional THA and TKA registries at the Hospital for Special Surgery. Of all patients undergoing THA or TKA at the Hospital for Special Surgery, approximately 84% consented to registry participation. Of those, 81% returned a baseline survey [17, 18]. For the purposes of this study, patients were consecutively recruited from the practices of eight orthopaedic surgeons between 2010 and 2012 after obtaining written informed consent. We included patients 18 years or older scheduled to undergo unilateral THA or TKA and excluded patients only if they had cognitive deficits that prevented them from completing their surveys.

Using the surgeon versions of the validated Hospital for Special Surgery 18-item THA expectations survey and the Hospital for Special Surgery 19-item TKA expectations survey, surgeons recorded their expectations for each patient after the preoperative evaluation [20, 23]. This questionnaire was developed to evaluate expectations of different aspects of recovery including pain relief; walking; the ability to perform personal, recreational, and social activities of daily living; and psychological well-being [20]. Validation included face and content validity, construct validity, and test–retest reliability [20, 23]. The improvement expected on each item was rated on a four-point Likert scale (4 = back to normal or complete improvement; 3 = not back to normal but a lot of improvement; 2 = not back to normal but a moderate amount of improvement; 1 = not back to normal but little improvement) and each item also has a no-expectation option (0 = I do not have this expectation or this expectation does not apply to me). An overall score is calculated by summing the scores of all the questions and converting it to a 0 to 100 scale with 100 being the highest expectation of returning to normal in all aspects and 0 being the most pessimistic (meaning that there are no expectations of improvement after surgery in any aspect) [20].

To recruit patients for the registry, patients scheduled for surgery were approached when they came to the hospital for their required preoperative screening, which primarily involved blood tests and medical examinations. Patients who agreed to participate were enrolled at that time and completed a series of baseline questionnaires including demographics, the SF-36 [32–34], and the Hip Injury Osteoarthritis Outcomes Score (HOOS) for patients undergoing THA [26] and the Knee Injury Osteoarthritis Outcomes Score (KOOS) for patients undergoing TKA [28, 29]. The HOOS and the KOOS include all of the WOMAC questions [3]. At the 2-year followup, patients undergoing THA completed the HOOS survey and those undergoing TKA completed the KOOS survey. For the purposes of this study, the HOOS and KOOS surveys were used solely to derive the WOMAC three subscale scores of pain (five items; score range, 0–20; 20 = highest level of pain), stiffness (two items; score range, 0–8; 8 = highest level of stiffness), and function (17 items; score range, 0–68; 68 = lowest level of function).

Preoperatively, we enrolled 259 patients undergoing THA and 247 undergoing TKA between 2009 and 2012. The 2-year followup rate was 77% ($n = 200$) and 77% ($n = 191$) for THA and TKA, respectively. Patients undergoing THA with no followup had worse baseline WOMAC and SF-36 Mental Component Summary scores and were more likely to be obese and have no college education. Patients undergoing TKA lacking followup were more likely to have worse American Society of

Anesthesiologists and WOMAC scores. Patients undergoing TKA were older, heavier, and included more females than those undergoing THA. The mean surgeon expectation scores were 85.6 ± 14.0 for THA and 79.1 ± 13.8 for TKA, respectively (Table 1). The distributions of expectations scores varied widely among surgeons; however, the distributions were similar for patients undergoing THA and those undergoing TKA when examining the scores for each surgeon separately.

A high proportion of patients improved as shown by the change in their WOMAC scores (Table 1). Ninety-percent of patients undergoing THA and 79% of those undergoing TKA achieved the minimum clinically important difference (MCID) on the pain subscale, whereas 90% of patients undergoing THA and 65% of those undergoing TKA achieved the MCID on the function subscale.

Analytic Plan

First, we applied univariate statistics to describe the study cohort. Means and SDs were calculated for continuous variables and frequencies for categorical variables. Second, we used statistical methods commonly used for assessing diagnostic tests (sensitivity, specificity, and receiver operating characteristic [ROC] curve) to evaluate the accuracy of surgeons' predictions. First, we classified patients in terms of whether they achieved the MCID in the WOMAC pain and function scores (1 = achieved MCID, 0 = did not achieve MCID) and used this as the gold standard (for purposes of testing sensitivity and specificity) when we evaluated the ability of surgeons' expectation scores to discriminate between patients who would benefit from THA or TKA and those who would not. The MCID represents the minimum change in the WOMAC subscale score that is a clinically significant improvement associated with these procedures. We used baseline-adjusted MCIDs, as described by Escobar et al. [8] for TKA and Quintana et al. [27] for THA, to account for the fact that patients starting with higher (better) baseline WOMAC scores have less room for improvement than those starting worse off. Escobar et al. [8] estimated baseline-adjusted MCIDs for TKA were 45 for the worst tertile, 28 for the medium tertile, and 16 for the best tertile on the WOMAC pain scale and 45, 33, and 17 for the three tertiles on the WOMAC function scale respectively. Similarly, Quintana et al. [27] estimated that baseline-adjusted MCIDs for THA were 36 for the worst tertile, 23 for the medium tertile, and 15 for the best tertile on the WOMAC pain scale and 31, 22, and 9 for the three tertiles on the WOMAC function scale respectively. Second, using whether patients achieved a MCID as the outcome, we estimated the areas under the ROC curves for the surgeons' expectations scores. We used

Table 1. Baseline characteristics and outcome scores of the THA and TKA cohorts

Parameter	THA		TKA	
	All patients recruited at baseline (n = 259)	Patients with 2-year followup (n = 200)	All patients recruited at baseline (n = 247)	Patients with 2-year followup (n = 191)
Baseline characteristics				
Age (years), mean (SD)	62 (12)	86 (43%)	68 (10)	123 (64%)
Female gender, number (%)	138 (53%)	111 (56%)	161 (65%)	124 (65%)
BMI (> 30 kg/m ²), number (%)	80 (31%)	54 (27%)	106 (43%)	77 (40%)
Charlson Comorbidity Index (> 0), number (%)	63 (24%)	49 (25%)	82 (33%)	63 (33%)
Living alone, number (%)	55 (21%)	45 (23%)	58 (24%)	43 (23%)
College education or above, number (%)	171 (66%)	141 (71%)	155 (63%)	124 (65%)
ASA score, number (%)				
P1	20 (8%)	19 (10%)	7 (3%)	6 (3%)
P2	197 (76%)	151 (76%)	197 (80%)	159 (83%)
P3	42 (16%)	30 (15%)	43 (17%)	26 (14%)
SF-36 Physical Component Summary, mean (SD)	33 (9)	33 (9)	33 (8)	34 (8)
SF-36 Mental Component Summary, mean (SD)	49 (12)	50 (12)	52 (12)	52 (12)
Outcome variables				
WOMAC Pain				
Baseline, mean (SD)	50 (18)	52 (18)	52 (17)	54 (16)
2 years, mean (SD)		92 (14)		87 (14)
% achieving MCID		90%		79%
WOMAC Function				
Baseline, mean (SD)	47 (19)	48 (19)	53 (17)	54 (16)
2 years, mean (SD)		90 (14)		86 (16)
% achieving MCID		90%		65%

ASA = American Society of Anesthesiologists; MCID = minimal clinically important difference.

adjusted scores that are derived from generalized estimating equations regression analysis, rather than raw scores, to calculate the area under the ROC curve. This approach takes into consideration potential similarities that exist among patients of the same surgeon [12]. An area of 0.5 indicates that surgeons' expectations are not better than chance in predicting the MCID and an area of 1 indicates that surgeons' expectations perfectly predict whether patients achieve MCIDs. Generally, an area under the ROC curve of 0.70 to 0.80 indicates acceptable discrimination, and areas above 0.80 indicate excellent discrimination [13]. We identified cut points for an expectations score as the threshold for predicted treatment success that maximizes the sensitivity and specificity. Post hoc subgroup analyses also were conducted. Regression-based areas under the ROC curve were calculated separately for men and women, older versus younger than 65 years, patients with a BMI of 30 kg/m² or greater versus those with a BMI less than

30 kg/m², and for patients with any comorbidities versus those with no comorbidities.

We conducted two sensitivity analyses to test the robustness of our results. First, we did an exploratory orthogonal factor analysis with varimax rotation to exclude items (item subsets) on the surgeons' expectations rating that may not be good predictors of outcomes because the original expectations survey was derived from patient interviews. Exploratory factor analysis is a statistical method used to group items that covary together into factors [1]. Once these factors were generated, we calculated surgeons' expectations scores, using the same method, and estimated regression-based areas under the ROC curves for each factor. In the second sensitivity analysis, we reran all generalized estimating equations models adjusting for the SF-36 Mental Component Summary score and generated areas under the ROC curve. Adjusting for this score controls for the patient's

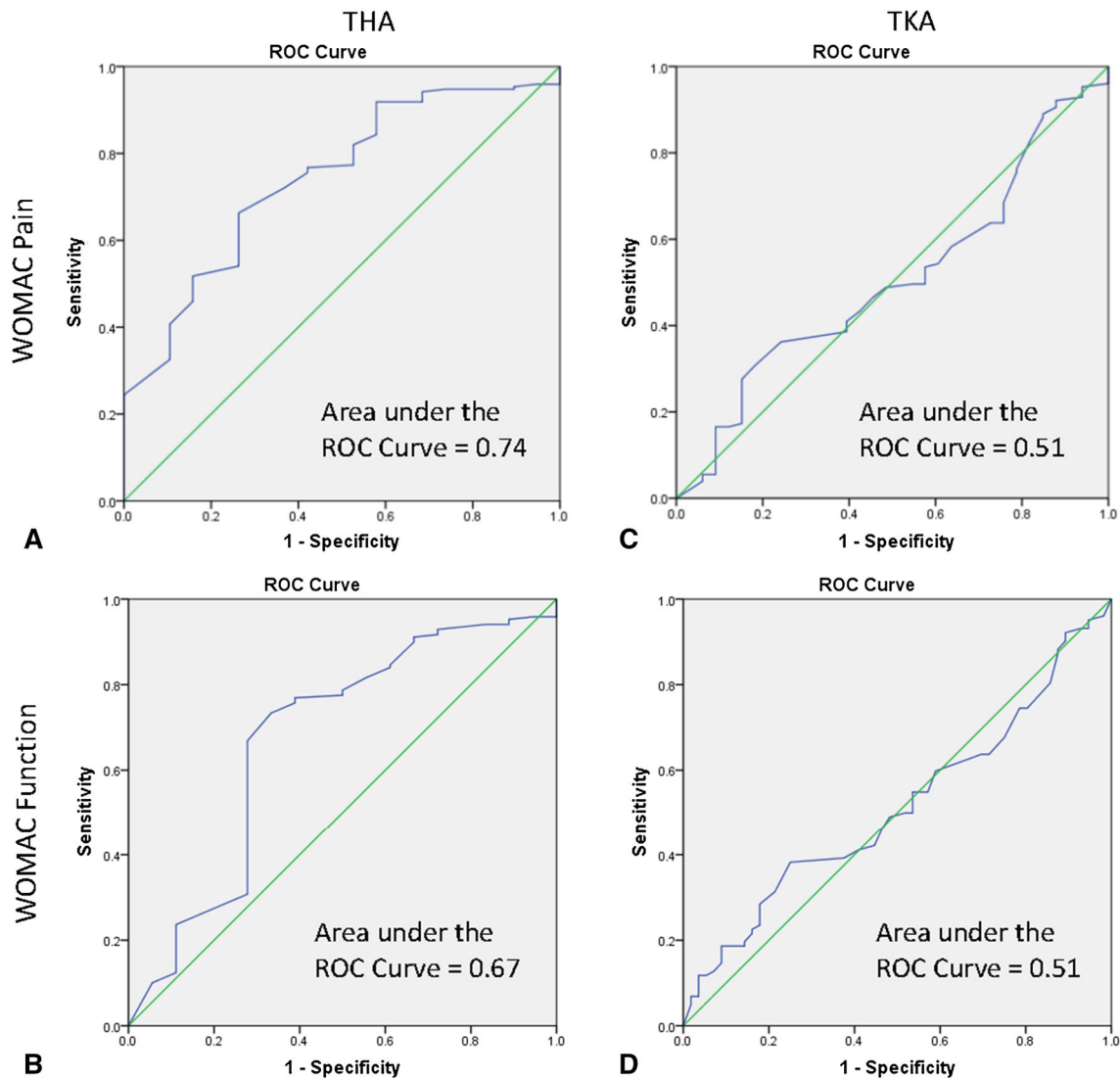


Fig. 1A–D The ROC curves show how accurately surgeons' expectations scores predict achieving the minimum clinically important difference in improvement in (A) pain relief and (B) function in patients undergoing THA and (C) pain and (D) function in patients undergoing TKA. The diagonal line indicates an area under the curve

of 0.5 (or no better than chance). The area under the curved line represents the area under the ROC curve and is significantly larger than 0.5 in patients undergoing THA but not in patients undergoing TKA.

psychologic well-being, which is known to affect outcomes, yet may not be as apparent to surgeons as other patient characteristics such as functional disability and thus not fully factored in the surgeons' expectations [5]. The SF-36 Mental Component Summary score has good specificity and sensitivity in detecting anxiety and depression [24], two psychologic problems known to be underdiagnosed in clinical practice [15, 30]. Our study was approved by the institutional review board of the Hospital for Special Surgery. All statistical analyses were conducted using SPSS Version 22.0 (IBM Corp, Armonk, NY, USA) using a significance level of 0.05.

Results

Surgeons' expectation scores effectively anticipated patients who would improve after THA, but they were no better than chance (that is, there was no clear trend regarding whether they overestimated or underestimated improvement) in identifying patients who would achieve the MCID on the WOMAC score 2 years after TKA. For THA, the area under the ROC curve for surgeons' expectations scores as a predictor of outcomes was 0.74 (95% CI, 0.63–0.85; $p < 0.01$) for WOMAC pain and 0.67 (95% CI, 0.53–0.82; $p = 0.02$) for WOMAC function, but not

Table 2. Ability of surgeons' expectations scores to discriminate good from bad outcomes

Surgeons' expectations	Percent achieving MCID, area under the ROC curve for achieving MCID	
	WOMAC function	WOMAC pain
Entire sample (n = 200)	90%, 0.67	90%, 0.74
BMI \geq 30 kg/m ² (n = 54)	80%, 0.55	83%, 0.65
BMI < 30 kg/m ² (n = 146)	86%, 0.72	87%, 0.78
Age \geq 65 years (n = 86)	84%, 0.78	81%, 0.62
Age < 65 years (n = 114)	85%, 0.62	89%, 0.79
Gender (female) (n = 111)	78%, 0.62	82%, 0.70
Gender (male) (n = 89)	93%, 0.94	91%, 0.85
Charlson Comorbidity Index = 0 (n = 151)	87%, 0.65	88%, 0.82
Charlson Comorbidity Index > 0 (n = 49)	78%, 0.74	80%, 0.51

MCID = minimum clinically important difference; ROC = receiver operating characteristic.

different from 0.5 for patients undergoing TKA (area under the ROC curve: Function = 0.51, [95% CI 0.42–0.61], $p = 0.78$; Pain = 0.51, [95% CI, 0.40–0.61], $p = 0.92$) (Fig. 1). Factor analysis revealed that the 18 items in the THA expectations survey grouped into three factors, explaining 78% of the variance (Appendix 1. Supplemental material is available with the online version of *CORR*®). Two items (improving ability to tie shoelaces and improving ability to cut toenails) grouped into one factor that had areas under the ROC curve that were not different from 0.5 (meaning they were no better than chance). When we excluded these items and recalculated the scores, the areas under the ROC curve were marginally improved (0.75 for WOMAC pain and 0.68 for WOMAC function). Four factors were generated for the patients undergoing TKA, and they explained 76% of the variance (Appendix 2. Supplemental material is available with the online version of *CORR*®). No factors had ROC curve areas that were significantly different from 0.5. These results did not change after adjusting for the SF-36 Mental Component Summary scores.

The higher the surgeons' preoperative expectation scores were, the more likely they were to be associated with a patient achieving the MCID on the WOMAC scale, and the surgeons' expectations scores were more predictive in patients who were men with a BMI less than 30 kg/m². The discriminating ability of surgeons' expectations in patients undergoing THA was maximized (sensitivity = 0.69, specificity = 0.74) at an expectations score of 82.6 or greater for WOMAC pain and maximized (sensitivity = 0.69, specificity = 0.72) at an expectations score of 82.6 or greater for WOMAC function. Subgroup analysis was conducted for THA only, since there was no predictive value to the expectations score for patients undergoing TKA. The area under the ROC curve was above 0.8, that is, indicating excellent discrimination, for functional improvement and pain relief in patients

undergoing THA who were men, and between 0.7 and 0.8, indicating acceptable discrimination, in those with a BMI less than 30 kg/m² (Table 2). We did not evaluate race because the number of nonwhite participants was low.

Discussion

Educating and informing patients about their likely outcomes using realistic benchmarks is important to help them reach their best-achievable outcome. Surgeons' expectations are thought of as realistic because of their training, knowledge, and clinical experience; however, this assumption has not been empirically validated. We showed that surgeons' expectations were reasonably good in discriminating between patients who did not do well versus those who did well after THA, that is, they generally had good sensitivity and specificity. Subgroup analyses revealed that they generally were more accurate for patients with a BMI less than 30 kg/m² and male gender, but also in more-challenging groups such as those with comorbidities and older patients. However, surgeons' expectation scores did not discriminate between good and poor outcomes in patients undergoing TKA.

This study had some limitations. First, our study was limited to evaluating the association between surgeons' expectations and patient-reported outcomes. Some outcomes judged as important by surgeons, such as ROM and stiffness, were not captured in our study and may be more predictable by surgeons, especially in patients undergoing TKA where these problems typically are more prevalent. Second, although the MCID values used in this study were robustly developed and baseline-adjusted, they may still fall short of representing patient-centered MCID values. We ran sensitivity analyses with the WOMAC pain and function scores as continuous rather than dichotomous variables and found similar results (association in the case

of THA and no association in the case of TKA). In addition, the MCID values used in this study represent minimal clinically important improvement, and therefore our conclusions are based on conservative estimates of improvement. Third, our TKA sample had a large proportion of women and patients with a BMI greater than 30 kg/m²; however, because our subgroup analysis showed that surgeons' expectation scores are equally not capable of predicting improvement in all TKA groups, we do not speculate that this may have contributed to the no-difference finding we observed in patients undergoing TKA. Fourth, sample-size limitations and the need to collect additional information prohibited us from conducting subgroup analysis by race and socioeconomic status. Fifth, the area under the ROC curve is independent of the prevalence of the outcome (ie, achieving MCID) [35], and therefore is not affected by the low number of patients who did not achieve the MCID, especially in patients undergoing THA, in our study. Sixth, we did not exclude patients with complications; however, the occurrence of complications is low (approximately 1%) [6], and is unlikely to change the results of our study. Finally, our study was limited to one specialized orthopaedic center (Hospital for Special Surgery) and our participants generally were high-volume surgeons and relatively well-educated patients. As such, our findings may not generalize well to settings where those qualities are not present. Prior work found the registry data from the Hospital for Special Surgery to be similar to the nationally representative Function and Outcomes Research for Comparative Effectiveness in Total Joint Replacement and Quality Improvement (FORCE-TJR) registry (www.force-tjr.org), and also may be affected by the loss to followup and the differences between patients with and without followup. Given that this subgroup includes more patients with obesity, our current ROC estimates may be overestimated.

Compared with the study by Meijerink et al. [25], which examined the association of the preoperative assessment of difficulty of the procedure and immediate postoperative satisfaction with a 1-year Knee Society clinical rating score in 51 patients, our study included a much larger sample of patients, we had 2-year followup, and we used a more-robust approach to studying expectations by anchoring analysis to MCIDs rather than unvalidated "satisfaction" scores. Our results suggest that surgeons' expectation scores may have some utility in predicting THA outcomes; however, this utility varied considerably among patient subgroups. Surgeons' scores appear better able to predict MCID improvement in function in patients undergoing THA who have a BMI less than 30 kg/m², are older than 65 years, and have one or more comorbidities; they were less predictive in patient subgroups where this surgery has only more recently become prevalent (younger, patients

with obesity, and those with no comorbidities). These findings highlight the importance of understanding and addressing the needs of patients as THA indications have expanded to include these other groups. In addition, the important gender differences in the accuracy of predicting outcomes must be noted. Gender differences have been documented in physician referral to orthopaedic surgeons and in recommendations for surgery, but we are not aware of prior studies that have shown gender differences in surgeons' expectations of the outcomes. These findings deserve further investigation.

The inability of surgeons' expectations scores to identify which patients would achieve a clinically important improvement after TKA using all or subsets of the items derived from factor analysis, and in all subgroups of patients, underscores the greater difficulty in predicting its outcomes compared with THA. Patients undergoing THA are more likely than patients undergoing TKA to report having a "forgotten joint" implant, an implant that so resembles the natural joint that the patient forgets his or her joint was replaced [2]. It also is more challenging to accurately predict certain complications such as stiffness, which occurs more often in patients undergoing TKA and substantially affects patients' progress and their likelihood of achieving a good outcome. Based on our findings, surgeon's expectations appear to be no better than chance in predicting that a patient will achieve the minimal clinically important improvement after a TKA, and therefore, it behooves surgeons to more realistically set patients' expectations and explain our findings to potential patients before surgery. While avoiding surgery may be one option for the patient in light of these findings, more than 2/3 of the patients undergoing TKA in our study actually achieved the MCID. In addition, TKA remains the most-effective treatment and a cost-effective treatment for advanced knee osteoarthritis, and research shows that continuing to live with advanced osteoarthritis is not only associated with disability, but also may increase the risk of cardiovascular disease [16, 31]. Further research is needed to identify other prognostic measures than surgeons' expectations for improvement for patients undergoing TKA.

The similar expectations that surgeons list for patients undergoing TKA and those undergoing THA are echoed by the results of a national survey of 358 surgeons performing THA and TKA that evaluated their expectations for four hypothetical patient vignettes (two THA and two TKA vignettes) and found similar patterns [7]. These results may indicate that surgeons still treat THA and TKA generally as one category and, therefore, expect similar outcomes. Informing the orthopaedic community about the similarities in surgeons' expectations for these two different procedures may make surgeons more conscious of the distinctive outcomes of these procedures, encourage them

to make their expectations after TKA more customized to TKA, and thus have better benchmarks for this procedure in the future.

The results of this study should encourage surgeons to carefully evaluate their expectations and those of their patients preoperatively. Comparing patients' and surgeons' expectations may provide a simple-yet-effective approach that can help surgeons improve on the approaches they use to counsel patients. Prior research has shown that a seven-point difference between a patient's score and surgeon's score was a clinically valid indicator of a meaningful difference in perceiving outcomes [11, 19]. For the "typical" patient undergoing THA, surgeons may use their expectations scores as benchmarks to adjust the expectations of their patients. For the other patients undergoing THA, and perhaps for patients undergoing TKA, additional discussions based on comparing the two expectations (patient's and surgeon's) should inform and improve the shared decision-making process regarding surgery. This study highlights the importance of surgeons' expectations in informing the discussions between patients and surgeons and does not address the surgeon's approach during these discussions, nor does it limit the efforts of addressing patients' expectations to only education by the surgeon and his or her staff.

Surgeons play an important role by informing and shaping their patients' expectations for outcomes after surgery. With expanding indications for surgery, this study shows that surgeons' expectations are reasonably predictive of improvement for patients undergoing THA who are 65 years or older, have one or more comorbidities, and who are male, but surgeons' expectations do not appear to anticipate results as accurately in other THA groups, nor in patients undergoing TKA. Therefore, although most patients improve after surgery, surgeons should spend adequate time with their patients to better understand their expectations and more realistically address them.

Acknowledgments Funding was provided by National Institute of Arthritis and Musculoskeletal and Skin Diseases (Grant No. R00 HD060686).

References

1. Bartholomew DJ. *Analysis of Multivariate Social Science Data*. 2nd ed. Boca Raton, FL: CRC Press; 2008.
2. Behrend H, Giesinger K, Giesinger JM, Kuster MS. The "forgotten joint" as the ultimate goal in joint arthroplasty: validation of a new patient-reported outcome measure. *J Arthroplasty*. 2012;27:430-436.e1.
3. Bellamy N. *WOMAC Osteoarthritis Index: A User's Guide*. London, Ontario: University of Western Ontario; 1995.
4. Chang RW, Pellisier JM, Hazen GB. A cost-effectiveness analysis of total hip arthroplasty for osteoarthritis of the hip. *JAMA*. 1996;275:858-865.
5. Duivenvoorden T, Vissers MM, Verhaar JA, Busschbach JJ, Gosens T, Bloem RM, Bierma-Zeinstra SM, Reijman M. Anxiety and depressive symptoms before and after total hip and knee arthroplasty: a prospective multicentre study. *Osteoarthritis Cartilage*. 2013;21:1834-1840.
6. Dushey CH, Bornstein LJ, Alexiades MM, Westrich GH. Short-term coagulation complications following total knee arthroplasty: a comparison of patient-reported and surgeon-verified complication rates. *J Arthroplasty*. 2011;26:1338-1342.
7. Dy CJ, Gonzalez Della Valle A, York S, Rodriguez JA, Sculco TP, Ghomrawi HM. Variations in surgeons' recovery expectations for patients undergoing total joint arthroplasty: a survey of the AAHKS membership. *J Arthroplasty*. 2013;28:401-405.
8. Escobar A, Garcia Perez L, Herrera-Espineira C, Aizpuru F, Sarasqueta C, Gonzalez Saenz de Tejada M, Quintana JM, Bilbao A. Total knee replacement; minimal clinically important differences and responders. *Osteoarthritis Cartilage*. 2013;21:2006-2012.
9. Ethgen O, Bruyere O, Richy F, Dardennes C, Reginster JY. Health-related quality of life in total hip and total knee arthroplasty: a qualitative and systematic review of the literature. *J Bone Joint Surg Am*. 2004;86:963-974.
10. Ghomrawi HM, Franco Ferrando N, Mandl LA, Do H, Noor N, Gonzalez Della Valle A. How often are patient and surgeon recovery expectations for total joint arthroplasty aligned? Results of a pilot study. *HSS J*. 2011;7:229-234.
11. Ghomrawi HM, Mancuso CA, Westrich GH, Marx RG, Mushlin AI; Expectations Discordance Study Group. Discordance in TKA expectations between patients and surgeons. *Clin Orthop Relat Res*. 2013;471:175-180.
12. Hanley JA, Negassa A, Edwards MD, Forrester JE. Statistical analysis of correlated data using generalized estimating equations: an orientation. *Am J Epidemiol*. 2003;157:364-375.
13. Hosmer DW, Lemeshow S. *Applied Logistic Regression*. New York, NY: John Wiley; 1989.
14. Insall JN, Dorr LD, Scott RD, Scott WN. Rationale of the Knee Society clinical rating system. *Clin Orthop Relat Res*. 1989;248:13-14.
15. Kasper S. Anxiety disorders: under-diagnosed and insufficiently treated. *Int J Psychiatry Clin Pract*. 2006;10(suppl 1):3-9.
16. Kim HS, Shin JS, Lee J, Lee YJ, Kim MR, Bae YH, Park KB, Lee EJ, Kim JH, Ha IH. Association between knee osteoarthritis, cardiovascular risk factors, and the Framingham Risk Score in South Koreans: a cross-sectional study. *PLoS One*. 2016;11:e0165325.
17. Lyman S, Lee YY, 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.
18. Lyman S, Lee YY, 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.
19. Mahomed NN, Liang MH, Cook EF, Daltroy LH, Fortin PR, Fossel AH, Katz JN. The importance of patient expectations in predicting functional outcomes after total joint arthroplasty. *J Rheumatol*. 2002;29:1273-1279.
20. Mancuso CA, Graziano S, Briskie LM, Peterson MG, Pellicci PM, Salvati EA, Sculco TP. Randomized trials to modify patients' preoperative expectations of hip and knee arthroplasties. *Clin Orthop Relat Res*. 2008;466:424-431.
21. Mancuso CA, Salvati EA, Johanson NA, Peterson MG, Charlson ME. Patients' expectations and satisfaction with total hip arthroplasty. *J Arthroplasty*. 1997;12:387-396.
22. Mancuso CA, Sculco TP, Salvati EA. Patients with poor preoperative functional status have high expectations of total hip arthroplasty. *J Arthroplasty*. 2003;18:872-878.

23. Mancuso CA, Sculco TP, Wickiewicz TL, Jones EC, Robbins L, Warren RF, Williams-Russo P. Patients' expectations of knee surgery. *J Bone Joint Surg Am*. 2001;83:1005-1012.
24. Matcham F, Norton S, Steer S, Hotopf M. Usefulness of the SF-36 Health Survey in screening for depressive and anxiety disorders in rheumatoid arthritis. *BMC Musculoskelet Disord*. 2016;17:224.
25. Meijerink HJ, Brokelman RB, van Loon CJ, van Kampen A, de Waal Malefijt MC. Surgeon's expectations do not predict the outcome of a total knee arthroplasty. *Arch Orthop Trauma Surg*. 2009;129:1361-1365.
26. Nilsson AK, Lohmander LS, Klassbo M, Roos EM. Hip disability and osteoarthritis outcome score (HOOS): validity and responsiveness in total hip replacement. *BMC Musculoskelet Disord*. 2003;4:10.
27. Quintana JM, Aguirre U, Barrio I, Orive M, Garcia S, Escobar A. Outcomes after total hip replacement based on patients' baseline status: what results can be expected? *Arthritis Care Res (Hoboken)*. 2012;64:563-572.
28. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynon BD. Knee Injury and Osteoarthritis Outcome Score (KOOS): development of a self-administered outcome measure. *J Orthop Sports Phys Ther*. 1998;28:88-96.
29. Roos EM, Toksvig-Larsen S. Knee injury and Osteoarthritis Outcome Score (KOOS): validation and comparison to the WOMAC in total knee replacement. *Health Qual Life Outcomes*. 2003;1:17.
30. Sheehan DV. Depression: underdiagnosed, undertreated, underappreciated. *Manag Care*. 2004;13(6 suppl Depression):6-8.
31. Wang H, Bai J, He B, Hu X, Liu D. Osteoarthritis and the risk of cardiovascular disease: a meta-analysis of observational studies. *Sci Rep*. 2016;6:39672.
32. Ware JE Jr, Gandek B, Kosinski M, Aaronson NK, Apolone G, Brazier J, Bullinger M, Kaasa S, Leplege A, Prieto L, Sullivan M, Thunedborg K. The equivalence of SF-36 summary health scores estimated using standard and country-specific algorithms in 10 countries: results from the IQOLA Project. International Quality of Life Assessment. *J Clin Epidemiol*. 1998;51:1167-1170.
33. Ware JE Jr, Kosinski M, Dewey JE, Gandek B. *SF-36 Health Survey: Manual and Interpretation Guide*. Lincoln, RI: Quality Metric Inc; 2000.
34. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. *Med Care*. 1992;30:473-483.
35. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem*. 1993;39:561-577.