


Ranking Surgeon Performance After ACL Reconstruction Using the Knee Injury and Osteoarthritis Outcome Score (KOOS) Subscales

Sarah B. Floyd,^{*†} PhD, MPH, Briggs Ahearn,[‡] MD, Michael J. Kissenberth,[‡] MD, John M. Brooks,^{§†} PhD, and Charles A. Thigpen,^{†||¶} PT, PhD, ATC 
Investigation performed at Center for Effectiveness Research in Orthopaedics, Greenville, South Carolina, USA

Background: Patient-reported outcomes (PROs) are considered the gold standard for evaluating value-based care in orthopaedics. However, there is little evidence to guide implementation of PROs for surgeon performance evaluation.

Purpose: To develop a risk-adjusted surgeon performance measure using the Knee injury and Osteoarthritis Outcome Score (KOOS) for patients undergoing anterior cruciate ligament reconstruction (ACLR).

Study Design: Cross-sectional study; Level of evidence, 3.

Methods: Patients (N = 1248; 662 men; mean age, 30 ± 13 years) who underwent ACLR performed by 40 surgeons between 2010 and 2018 were identified from a large, nationally representative sports medicine clinical data registry. Linear regression was used to predict change scores for each KOOS subscale (Pain, Symptoms, Activities of Daily Living [ADL], Function in Sports and Recreation, and Knee-Related Quality of Life) while adjusting for patient baseline characteristics. A risk-adjusted performance measure was calculated for each KOOS subscale as the difference between the unadjusted and the risk-adjusted predicted change score across all patients treated by a single surgeon. Surgeon-relative quartile ranking was compared across outcome subscale scores.

Results: One-third of the patients (34%) displayed acute cartilage damage, and 56% had a meniscal injury. In the risk adjustment models, older age, presence of diabetes, current smoking status, acute cartilage damage, concurrent cartilage treatment, lower baseline Veterans RAND 12-Item Health Survey mental and physical component scores, and lower baseline Marx and KOOS subscale values all had a significant negative influence on the predicted KOOS subscale change values ($P < .05$ for all). Surgeon performance, ranked in quartile groups, was the same for 10 surgeons but varied by 1 to 2 quartiles for the other 30 surgeons across the different KOOS subscales.

Conclusion: These results showed that surgeon performance varies widely when evaluated using different KOOS subscales for patients undergoing ACLR. Based on the preliminary results and clinical perspective, the authors recommend the ADL and Symptoms subscales as the best options to differentiate surgeon performance for these patients. However, evaluation of surgeon performance may require consideration or use of a set of PROs or the development of a single index PRO that is sensitive to the range of outcome dimensions important to patients.

Keywords: ACL; outcomes; KOOS; PRO; performance measures; knee

Value-based health care, where “value” is defined as the patient health outcomes achieved relative to the resources spent, remains at the forefront of the discussion

surrounding health care system reform in the United States.^{13,18} Initial attempts in orthopaedics to assess quality include the Centers for Medicare and Medicaid Services (CMS) hospital-specific reports of 30-day and 90-day risk standardized readmission rates after elective total hip and knee procedures.^{6,17,23} However, these efforts only reflect process-based measures, which are void of treatment outcomes that are valued by orthopaedic patients.

The Orthopaedic Journal of Sports Medicine, 13(1), 23259671241298993
 DOI: 10.1177/23259671241298993
 © The Author(s) 2025

This open-access article is published and distributed under the Creative Commons Attribution - NonCommercial - No Derivatives License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits the noncommercial use, distribution, and reproduction of the article in any medium, provided the original author and source are credited. You may not alter, transform, or build upon this article without the permission of the Author(s). For article reuse guidelines, please visit SAGE's website at <http://www.sagepub.com/journals-permissions>.

More recent efforts to define and measure value-based care recognize the importance of including patient-reported outcomes (PROs) that capture and track patient expectations, preferences, and distinct dimensions of health status (pain, function, quality of life, etc) over time.^{4,11,14,20} The CMS Quality Payment Program (QPP; <https://qpp.cms.gov>) now includes PROs as high-priority measures that provide a better assessment of clinical performance, quality, and value in orthopaedic medicine compared with process measures alone.⁹ The changes in PROs before and after a treatment are proposed as an approach to compare providers in the form of PRO–performance measures (PRO-PMs).⁹ PRO-PMs can be used to score and compare providers' relative performance. These scores can then be ranked, and cutoff points can be established to inform pay-for-performance compensation bonuses or penalties.²¹ However, the operationalization of this approach is complicated by the multidimensional nature of the outcomes valued by patients.^{4,14}

Recent work has shown that a risk-adjusted PRO-PM approach reveals wide variation in surgeon performance and enables performance comparisons that control for factors outside the surgeon's control.²² This approach is particularly suitable for the assessment of common outpatient surgeries such as anterior cruciate ligament (ACL) reconstruction (ACLR). There are several validated PRO options that could be used for ACLR, including the Knee injury and Osteoarthritis Outcome Score (KOOS), International Knee Documentation Committee (IKDC) subjective score, and Lysholm knee score. The KOOS, which is responsive to changes after ACLR, includes 5 subscales: Pain, Symptoms, Activities of Daily Living (ADL), Function in Sports and Recreation (Sports/Rec), and Knee-Related Quality of Life (QOL).^{3,8} Each of these subscales has been shown to measure distinct outcome dimensions that are valued by patients. However, no study has evaluated if the weighted subscales of pain or function influence surgeon performance ranking. Prior work has shown that the American Shoulder and Elbow Surgeons shoulder score differentiated surgeon performance for rotator cuff repair using risk adjustment of baseline scores and patient factors.²² Thus, when applying this approach to evaluate surgeon performance for ACLR, it is unknown if performance will vary across surgeons or different KOOS subscale dimensions.

To develop risk-adjusted PRO-PMs from KOOS outcome dimensions, it must first be established if a similar risk

adjustment approach differentiates surgeon performance and if surgeon performance varies across KOOS subscales.² This is fundamental if PRO-PMs are to be applied to determine surgeon performance and reimbursement in value-based health care models.^{9,12}

The purpose of this study was to apply a risk-adjusted approach for comparing surgeon performance using the KOOS for patients undergoing ACLR and to evaluate if different KOOS subscale values influence surgeon ranking. We hypothesized that surgeon performance will vary according to KOOS subscale values.

METHODS

Data and Study Sample

This study received institutional review board approval and adhered to the Reporting of Studies Conducted Using Observational Routinely-Collected Data guidelines. Data from 2851 patients who underwent ACLR from 2010 to 2018 were extracted from the Surgical Outcomes System (SOS; Arthrex). The SOS database is a national orthopaedic and sports medicine–certified clinical data registry that was developed for surgeons to easily collect and analyze patient outcomes at baseline and after surgery. Patients were included in this study if they had undergone a primary, unilateral ACLR of a complete ACL tear, had complete patient demographic data, and had complete baseline and 6-month postoperative PRO data. The final analytical sample included data from 85 surgeons and 1248 patients. The full cohort inclusion process is depicted in Figure 1.

Outcome Measures

The KOOS was chosen as the primary outcome of interest because it has been widely used for this study population.^{3,10,15} Additionally, the outcome dimensions of the KOOS have been shown to be responsive to changes after surgical procedures, including ACLR.^{3,8} The KOOS holds 5 separately scored subscales: Pain, Symptoms, ADL, Sports/Rec, and QOL. A normalized score (100 indicating no symptoms and 0 indicating extreme symptoms) is calculated for each subscale. A total score has not been validated

[#]Address correspondence to Charles A. Thigpen, PT, PhD, ATC, ATI Physical Therapy, 200 Patewood Drive, Suite C150, Greenville, SC 29615, USA (email: charles.thigpen@atipt.com).

^{*}Department of Public Health Sciences, Clemson University, Clemson, South Carolina, USA.

[†]Center for Effectiveness Research in Orthopaedics, Greenville, South Carolina, USA.

[‡]Steadman Hawkins Clinic of the Carolinas, Greenville, South Carolina, USA.

[§]University of South Carolina, Department of Health Services Policy and Management, Columbia, South Carolina, USA.

^{||}ATI Physical Therapy, Greenville, South Carolina, USA.

[¶]Institute of Musculoskeletal Advancement, Bolingbrook, Illinois, USA.

Final revision submitted March 27, 2024; accepted May 2, 2024.

One or more of the authors has declared the following potential conflict of interest or source of funding: B.A. has received grant support from Arthrex and DJO and education payments from Smith & Nephew and Peerless Surgical. M.K. has received education payments from Arthrex, consulting fees from Arthrex, nonconsulting fees from Arthrex, and hospitality payments from Exactech. C.A.T. has received consulting fees from Breg and has stock/stock options in Players Health and Trex. AOSSM checks author disclosures against the Open Payments Database (OPD). AOSSM has not conducted an independent investigation on the OPD and disclaims any liability or responsibility relating thereto.

Ethical approval for this study was obtained from Prisma Health (Pro00102962).

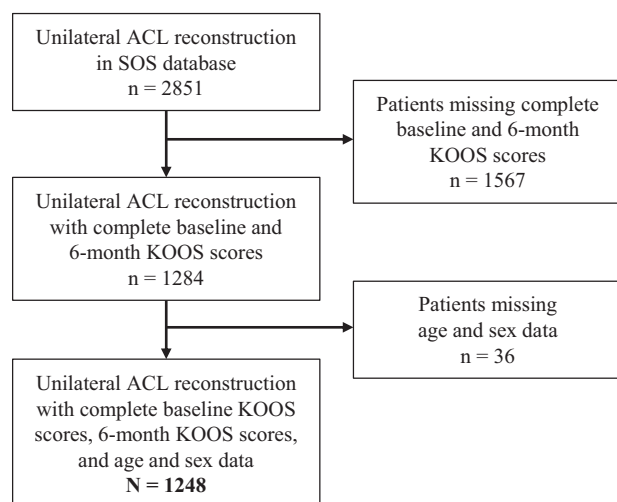


Figure 1. Flowchart of patient inclusion. ACL, anterior cruciate ligament; KOOS, Knee injury and Osteoarthritis Outcome Score; SOS, Surgical Outcomes System.

and is not recommended. For this analysis, we compared the use of different subscale scores to assess how the use of different outcomes affected clinical performance scores and relative surgeon rankings. Postsurgical KOOS subscale values collected at 6 months were used.² The 6-month KOOS was selected because currently the CMS QPP requires annual reporting, and this outcome period would grant physicians a 6-month period to perform surgeries that would be used for annual performance evaluation. The use of this episode-based approach maximizes the treatment period and sample size for each physician under evaluation.

Analytical Approach

For this analysis, we applied a previously published risk-adjusted performance measure approach for rotator cuff repair.²² A summary of each step in that process is presented below, and more details can be found in our previous publication and in the Appendix.²²

Step 1: Select a Disease-Specific PRO. The initial step in developing a performance measure is selecting a disease-specific PRO measure that is sensitive to changes after ACLR. Given that the purpose of this study was to evaluate if different KOOS subscale values influenced surgeon ranking, we chose to model each of the 5 KOOS subscale values.

Step 2: Baseline Risk Adjustment Factors. Patient demographic control measures included in all model specifications were age, sex, diabetes, and smoking status. Clinical control measures included concurrent diagnosis of a multiligament injury, acute cartilage damage (articular or other intra-articular), chronic cartilage damage (osteoarthritis), or meniscal injury. A multiple-ligament injury was defined as a grade 3 injury to the posterior cruciate ligament, medial collateral ligament, lateral collateral ligament, medial patellofemoral ligament, or anterolateral ligament. Concurrent knee procedures were included as control variables and included meniscal repair, meniscectomy,

other ligament repair, or cartilage treatment. Baseline PROs were included in all model specifications and included measures of pain, Veterans RAND 12-Item Health Survey (VR-12) mental and physical component scores, Marx activity scores, and all KOOS subscale values. A P value $<.05$ was considered statistically significant.

Step 3: Model Selection and Assessment. For each KOOS subscale value, we modeled outcome change using a linear regression model estimated using ordinary least squares because the outcome variable was continuous with a large sample size. Although more complex functional forms are available, the linear specification provided the direct relationship between specific factors and subscale scores. The adjusted R^2 and residual standard error were used to assess the explanatory power and accuracy of each model. Coefficients for each risk factor were examined for each factor's impact on the KOOS subscale change value.

Multiple linear predictive models were examined for the effect of each factor on the KOOS subscale change value. Reference class categories were specified as follows: male patient, nonsmoker, no nondiabetes, no concurrent diagnosis of multiligament injury, no acute cartilage damage (articular or other intra-articular), no chronic cartilage damage (osteoarthritis), and no concurrent meniscal repair or meniscectomy surgical procedure performed.

Step 4: Surgeon Performance Scores. The predicted 6-month KOOS subscale change values were estimated for each patient using the risk-adjusted linear models outlined above. The predicted change scores and unadjusted change scores were averaged across all patients treated by a given surgeon. The performance scores for individual surgeons were then calculated as the difference between the risk-adjusted and unadjusted change scores. The resulting "surgeon performance score" represents the number of KOOS points better or worse than the surgeons' panel of patients achieved than expected.

The Efron bootstrap method was used to estimate the distributional characteristics of each performance score and 95% confidence intervals surrounding each surgeon's performance score.⁵ Each surgeon's performance score was estimated from a random sample, with replacements, meaning patients could be redrawn multiple times to create a new sample equal to the original sample size. Surgeons with ≥ 5 complete patient observations were used to estimate performance scores (total of 1165 patients across 40 surgeons). This process was repeated 2000 times to generate a performance score distribution and median score per surgeon. Bootstrapping provides robust control for confounding due to missingness, as it extrapolates the available data to provide robust estimates of the prediction model. All values are presented as means with standard deviations. SAS (Version 9.4; SAS Institute) and R software (Version 1.2.1335; R Foundation for Statistical Computing) were used for building our analytical database and for statistical analysis.

RESULTS

The full study sample included 1248 patients (662 men; mean age, 30 ± 13 years) who underwent primary

TABLE 1
Patient Sample Characteristics (N = 1245)^a

Variable	Value
Age, y	30.2 ± 13
Sex	
Male	662 (53)
Female	583 (47)
Current smoker	14 (1)
Diabetes	15 (1)
Concurrent knee diagnoses	
Multiligament injury	36 (3)
Acute cartilage damage ^b	425 (34)
Chronic cartilage damage (osteoarthritis)	28 (2)
Meniscal injury	702 (56)
Medial	307 (25)
Lateral	395 (32)
Concurrent knee procedures	
Meniscal repair	277 (22)
Meniscectomy	374 (30)
Other ligament repair	42 (3)
Cartilage treatment	309 (25)
Baseline PRO scores	
VAS pain	2.5 ± 2
VR-12 mental component	51.8 ± 11
VR-12 physical component	37.4 ± 9
Marx score	10.7 ± 5
KOOS Pain	65.2 ± 19
KOOS Symptoms	58.7 ± 18
KOOS ADL	72.9 ± 19
KOOS Sports/Rec	34.8 ± 27
KOOS QOL	29.3 ± 19
KOOS 6-mo change scores	
KOOS Pain	18.2 ± 19
KOOS Symptoms	14.9 ± 21
KOOS ADL	17.9 ± 19
KOOS Sports/Rec	30.9 ± 30
KOOS QOL	25.9 ± 23

^aData are presented as mean ± SD or n (%). ADL, Activities of Daily Living; KOOS, Knee injury and Osteoarthritis Outcome Score; PRO, patient-reported outcome; QOL, Knee-Related Quality of Life; Sports/Rec, Function in Sports and Recreation; VAS, visual analog scale; VR-12, Veterans RAND 12-Item Health Survey.

^bArticular or other intra-articular.

unilateral ACLR, with 1% having diabetes and 1% current smokers. One-third of patients (34%) reported having acute cartilage damage, with 25% receiving concurrent cartilage treatment. Similarly, 56% had a meniscal injury and 22% underwent a concurrent meniscal repair. The mean KOOS subscale change values were Pain (18 ± 19), Symptoms (15 ± 21), ADL (18 ± 19), Sports/Rec (31 ± 30), and QOL (26 ± 23). Table 1 shows the complete details of the study sample.

The KOOS ADL model explained the largest amount of the variance ($R^2 = 0.70$), followed by the Pain ($R^2 = 0.55$) and Sports/Rec ($R^2 = 0.53$) subscales. Risk factor degree of influence and direction of effect were relatively consistent across all 5 models with some slight variations. Increasing patient age and smoking had a significant

negative effect on KOOS change values across all subscales (Table 2). Patients undergoing cartilage treatment had significantly lower KOOS change values across all KOOS outcome dimension models. Patients with higher baseline VR-12 mental and physical component scores displayed greater 6-month KOOS change values across all KOOS subscale models except the KOOS ADL model.

Clinical Performance Scores Across KOOS Subscale Dimensions

Risk-adjusted surgeon performance scores ranged widely across surgeons and models (range -25 to 23 points). The KOOS Sports/Rec model displayed the largest score range. However, the Sports/Rec scores were skewed, with the majority of surgeon scores <0. Conversely, the KOOS Pain, ADL, and Symptoms models had the majority of surgeons achieving scores higher than expected (Figure 2). Evaluation of the 95% confidence intervals indicated that the KOOS QOL model was only able to differentiate the performance scores of 8 of the 40 surgeons, suggesting that surgeon performance did not significantly vary across the QOL subscale. However, the KOOS Symptoms, Pain, Sports/Rec, and ADL models were able to significantly differentiate the performance of many more surgeons.

Relative Surgeon Ranking Across KOOS Subscales

Rankings varied greatly for some surgeons and were more consistent for others when comparing individual and quartile group ranks across KOOS subscale dimensions (Figure 3). For example, surgeons 3, 13, 14, 20, 30, and 35 had performance rankings in the top quartile across all 5 KOOS subscale models. Similarly, surgeons 17, 24, and 25 consistently performed in the bottom quartile of surgeons across all 5 subscale models. For these surgeons, performance rankings did not significantly differ when different KOOS outcome dimensions were used.

However, other surgeons saw much more varied rankings across the KOOS dimensions. For example, surgeons 4, 5, 11, 26, and 28 had quartile rankings change by 2 or more quartiles across KOOS dimensions. In total, 10 of 40 surgeons were in the same performance quartile no matter what dimension was used (green in Figure 3), 17 of 40 surgeons changed 1 group position (yellow), and 13 surgeons changed at least quartile 2 groups (red) (Figure 3). The majority of surgeons (30/40) experienced rank changes depending on which KOOS subscale dimension was used for outcome assessment.

DISCUSSION

This is the first study to show that clinical performance measures vary widely across KOOS subscales, which are deemed important to patients undergoing ACLR. The variability in surgeon ranking appears to be influenced by the outcome subscale selected and calls to attention that

TABLE 2
Results of Adjusted Regression Analysis^a

Variable	KOOS QOL	KOOS Sports/Rec	KOOS ADL	KOOS Pain	KOOS Symptoms
Patient characteristics					
Patient age	0.003	-0.27 ^d	-0.12 ^d	-0.10 ^d	-0.06
Female sex	0.46	-0.69	-0.48	-0.86	-2.02 ^c
Current smoker	-0.96	-11.43 ^c	-5.64 ^c	-3.70	-3.60
Diabetes	-10.29 ^c	-13.23 ^c	-9.76 ^d	-14.04 ^d	-12.17 ^d
Concurrent knee diagnoses					
Multiligament injury	-0.24	-4.28	-0.83	-1.00	2.34
Acute cartilage damage	0.39	-1.94	-1.49 ^b	-1.91 ^b	-2.61 ^c
Chronic cartilage damage	-5.81	-1.94	2.72	2.23	6.09 ^c
Meniscal injury	-0.62	-0.14	-0.26	-1.34	-1.08
Concurrent knee procedures					
Meniscal repair	1.20	0.96	0.05	0.57	0.86
Meniscectomy	1.66	3.19	0.08	1.20	2.89 ^c
Other ligament repair	4.49	-0.94	1.21	1.55	-1.64
Cartilage treatment	-6.34 ^d	-4.96 ^c	-2.64 ^d	-3.37 ^d	-10.25 ^d
Baseline PRO scores					
VAS pain	0.46	0.28	-0.41 ^c	-0.06	0.27
VR-12 mental component	0.28 ^d	0.22 ^d	0.14 ^d	0.17 ^d	0.16 ^d
VR-12 physical component	0.19 ^c	0.16 ^b	0.07	0.10 ^b	0.11 ^b
Marx score	0.13	0.36 ^d	0.19 ^d	0.20 ^c	0.15
KOOS Pain	0.10 ^b	0.18 ^d	0.07 ^c	-0.80 ^d	0.10 ^c
KOOS Symptoms	0.04	0.000	-0.01	0.02	-0.78 ^d
KOOS ADL	-0.08	0.05	-0.90 ^d	0.005	-0.04
KOOS Sports/Rec	-0.001	-0.92 ^d	-0.02	-0.04 ^c	-0.03
KOOS QOL	-0.67 ^d	0.12 ^d	0.01	0.04	0.04
No. of observations	1248	1248	1248	1248	1248
Adjusted R^2	0.27	0.53	0.70	0.55	0.46
F statistic	23.17 ^d	67.29 ^d	137.32 ^d	73.55 ^d	50.81 ^d

^aReference class categories were as follows: male sex, nonsmoker, no diabetes, no concurrent diagnosis of multiligament injury, no acute cartilage damage (articular or other intra-articular), no chronic cartilage damage (osteoarthritis), and no concurrent meniscal repair or meniscectomy surgical procedure performed. ADL, Activities of Daily Living; KOOS, Knee injury and Osteoarthritis Outcome Score; PRO, patient-reported outcome; QOL, Knee-Related Quality of Life; Sports/Rec, Function in Sports and Recreation; VAS, visual analog scale; VR-12, Veterans RAND 12-Item Health Survey.

^b $P < .05$.

^c $P < .01$.

^d $P < .001$.

careful consideration must be given to outcome dimension selection when used for surgeon performance rankings. Our results demonstrate that risk adjustment changes surgeon ranking compared with raw KOOS subscale values and support recommendations that provider performance evaluations should report performance across outcome dimensions valued by patients, and the development of value-based payment models should incorporate patient values into payment models. Caution should be used when using subscales for conditions in which improvement in those areas is not expected in the measurement time frame (eg, KOOS Sport/Rec subscale).

While the KOOS subscale models were able to significantly differentiate surgeon rankings, in many cases surgeon rankings varied widely across subscales. In fact, 75% of the surgeons' quartile group ranking changed by >1 quartile based on the different KOOS models selected. For example, surgeon 11 was the fourth-highest ranked surgeon for the KOOS QOL model but was ranked 25th

of the 40 surgeons for the ADL model. This is a critical factor that should be considered as efforts to harmonize outcomes and ongoing conversations surrounding PRO-PM development are occurring.^{2,9} Because our results demonstrated variation in surgeon performance when using different KOOS subscales, it is possible that other PROs in orthopaedics that are a combined single score across subscale dimensions (eg, Western Ontario and McMaster Universities Osteoarthritis Index and IKDC score) would produce similar provider performance variation across outcome dimensions. To improve performance measurement in orthopaedics, the industry must move beyond the initial efforts to reach PRO consensus that have been strictly based on measurement properties (validity, reliability, and responsiveness).^{7,20} Rather, outcome measures that are aligned with or weighted to reflect individual patient values across treatment outcome domains should be selected.

Consideration of the outcomes highly valued by patients with ACLR and thought to be the most responsive to

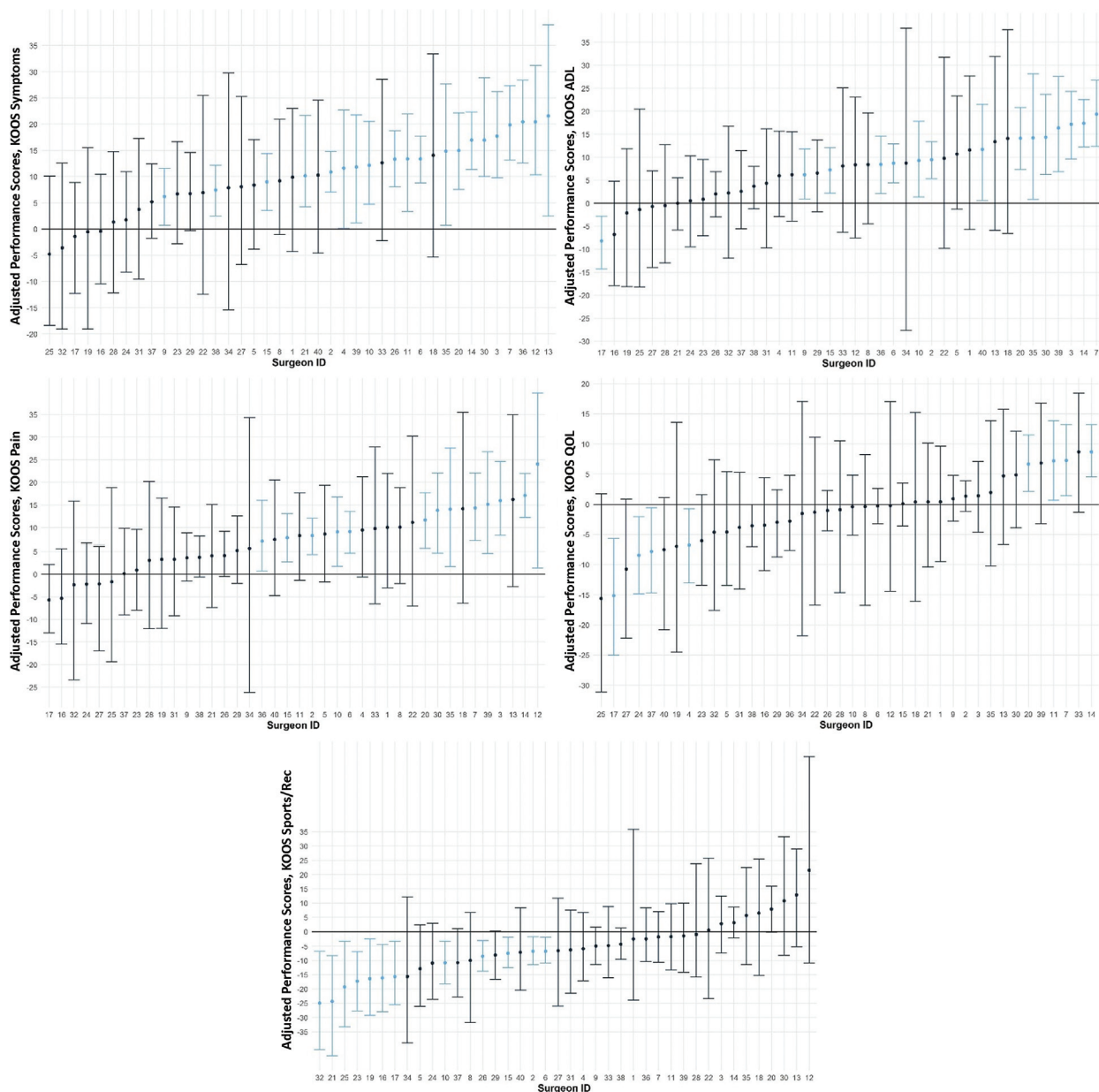


Figure 2. Anterior cruciate ligament reconstruction surgeon performance measure scores and rankings across Knee injury and Osteoarthritis Outcome Score (KOOS) subscales. Error bars represent 95% confidence intervals. ADL, Activities of Daily Living; QOL, Knee-Related Quality of Life; Sports/Rec, Function in Sports and Recreation.

treatment choice should have the most weight in surgeon performance evaluation. Given these considerations and our specific analysis, we would prioritize the ADL and Symptom subscales to measure surgeon performance for ACLR at 6 months with the understanding that these factors are most likely to change in that time frame. Pain subscale change scores suggest that, on average, the vast majority of patients had resolved pain at 6 months and pain is not typically considered a primary indication for ACLR. Similarly, the QOL subscale is not typically a differential subscale before ACLR and differentiated the fewest surgeons compared with other subscale models. We do

not recommend the Sports/Rec subscale, as best-practice recommendations suggest return to sport no earlier than 6 months and most often later than 7 months; thus, using this subscale as a measure would promote substandard clinical practice. In contrast, the ADL and Symptom subscales reflect important dimensions that are associated with the best indications for ACLR and are expected to recover in a substantial manner within the first 6 months after ACLR. These 2 domains likely provide the best measures to differentiate surgeon performance at a 6-month postoperative time point after ACLR for the purpose of value-based payment. However, a longer outcome period

Surgeon ID	ADL Rank Position From Highest	ADL Quartile Group	Pain Rank Position From Highest	Pain Quartile Group	QOL Rank Position From Highest	QOL Quartile Group	Sports/Rec Rank Position From Highest	Sports/Rec Quartile Group	Symptoms Rank Position From Highest	Symptoms Quartile Group
1	11	2	12	2	13	2	15	2	21	3
2	14	2	20	2	11	2	23	3	18	2
3	3	1	4	1	10	1	8	1	5	1
4	26	3	14	2	33	4	19	2	17	2
5	12	2	18	2	30	3	32	4	25	3
6	17	2	16	2	18	2	22	3	12	2
7	1	1	7	1	3	1	13	2	4	1
8	19	2	13	2	19	2	28	3	23	3
9	24	3	29	3	12	2	18	2	31	4
10	15	2	17	2	20	2	30	3	15	2
11	25	3	19	2	4	1	12	2	13	2
12	20	2	1	1	17	2	1	1	3	1
13	9	1	3	1	8	1	2	1	1	1
14	2	1	2	1	1	1	7	1	7	1
15	22	3	21	3	16	2	25	3	22	3
16	39	4	39	4	27	3	35	4	36	4
17	40	4	40	4	39	4	34	4	38	4
18	8	1	6	1	15	2	5	1	10	1
19	38	4	30	3	34	4	36	4	37	4
20	7	1	10	1	6	1	4	1	8	1
21	34	4	27	3	14	2	39	4	20	2
22	13	2	11	2	23	3	9	1	28	3
23	32	4	33	4	32	4	37	4	30	3
24	33	4	36	4	37	4	31	4	34	4
25	37	4	35	4	40	4	38	4	40	4
26	31	4	26	3	22	3	27	3	11	2
27	36	4	37	4	38	4	21	3	26	3
28	35	4	32	4	21	3	10	1	35	4
29	23	3	25	3	26	3	26	3	29	3
30	5	1	9	1	7	1	3	1	6	1
31	27	3	31	4	29	3	20	2	33	4
32	30	3	38	4	31	4	40	4	39	4
33	21	3	15	2	2	1	17	2	14	2
34	16	2	24	3	24	3	33	4	24	3
35	6	1	8	1	9	1	6	1	9	1
36	18	2	23	3	25	3	14	2	2	1
37	29	3	34	4	36	4	29	3	32	4
38	28	3	28	3	28	3	16	2	27	3
39	4	1	5	1	5	1	11	2	16	2
40	10	1	22	3	35	4	24	3	19	2

Figure 3. Surgeon rank score and quartile group score across Knee injury and Osteoarthritis Outcome Score (KOOS) subscales. Red indicates surgeons who changed ≥ 2 quartile rankings across KOOS subscale dimensions ($n = 13$). Yellow indicates surgeons who changed 1 quartile ranking across dimensions ($n = 17$). Green indicates surgeons who did not change quartile rankings across dimensions ($n = 10$). ADL, Activities of Daily Living; QOL, Knee-Related Quality of Life; Sports/Rec, Function in Sports and Recreation.

could be considered for ACLR, with an extended outcome window at 1 or 2 years, and may better differentiate and reflect surgeon performance, keeping in mind that the longer follow-up period will be challenging for current value-based performance models, which are annual assessment periods. This distinction highlights an important consideration when applying these types of performance measures in value-based payment arrangements.

Limitations

Our results should be viewed within the limitations of our study sample and design. We utilized a large, nationally

representative data registry to obtain information and were inherently limited by the extent of information available for each patient and physician. For example, specifics regarding the location, size, and morphology of acute cartilage injuries and meniscal tears as well as the extent of the concomitant cartilage/meniscal procedure performed were not readily available. We acknowledge that cartilage and meniscal injuries and concurrent procedures are important clinical variables and should be considered in future lower extremity sports medicine performance models. As a result, in this study, it is possible that a portion of the performance measurement variation we observe across outcome dimensions could stem from differences in the distributions of injury location, size, and morphology across the patients


seen by each surgeon. However, it is unlikely that these factors varied consistently across surgeons. Thus, the observed differences in risk-adjusted surgeon performance measurement approach previously applied for shoulder patients using a single outcome can be applied to patients undergoing ACLR across KOOS dimensions. Furthermore, we used outcomes at 6 months to generate the surgeon performance score. This time frame may be too short for some treatment outcome domains and may explain why some models failed to differentiate surgeon performance. The expected full recovery for ACLR is not until 9 to 12 months postoperatively. Therefore, outcomes at 1 year may provide the most meaningful differentiator, as this is the “ultimate outcome” of the treatment.^{1,16} However, an episode-based approach and a 6-month outcome period were selected to maximize the treatment period and sample size for each physician under evaluation. This decision was made to harmonize with the current CMS reporting period. We found that physician scores were highly variable using 6-month outcomes and believe that the use of a longer outcome, such as 12 months, would likely have the same conclusion. Therefore, we conclude that careful consideration should be given to the outcome measures selected for performance evaluation.

This approach allows for an iterative process of model refinement, where emerging risk factors can be added to the model and model adjustments can be made to improve overall performance. Finally, complete patient demographic and clinical variables needed for analysis were only available for 43.6% of the ACLR population in this registry, suggesting that implementation barriers to establishing a national path forward for clinical performance measurement should be considered.²⁰ If PRO-PMs are being considered as a method to determine surgeon reimbursement, there must be a widespread commitment to implementing robust data collection practices.

CONCLUSION

Our results showed that surgeon performance varies widely across the KOOS subscales for patients undergoing ACLR. This is the first study to show that the choice of KOOS subscale influences the relative surgeon ranking and thus potentially impacts reimbursement in value-based payment models. Based on our preliminary results and evaluation at 6 months, we recommend the KOOS ADL and Symptoms subscales as the best options to accurately differentiate surgeon performance for patients undergoing ACLR. Our results suggest that a patient-centered evaluation of surgeon performance may require consideration or use of a set of PROs per joint or the development of a single index PRO that is sensitive to the range of outcome dimensions important to patients. These limitations in current PRO-PM should be carefully considered as value-based payment models are developed and deployed.

ORCID iD

Charles A. Thigpen  <https://orcid.org/0000-0003-2487-1399>

REFERENCES

1. American Academy of Orthopaedic Surgeons. Appropriate use criteria for the treatment of anterior cruciate ligament injuries. American Academy of Orthopaedic Surgeons; 2015. Accessed October 25, 2024. <https://www.aaos.org/globalassets/quality-and-practice-re-sources/auc-treatment-acl-injuries.pdf>
2. American Academy of Orthopaedic Surgeons. Principles for musculoskeletal based patient reported outcome-performance measurement development [position statement 1188]. American Academy of Orthopaedic Surgeons; 2018. Accessed October 25, 2024. <https://www.aaos.org/contentassets/1cd7f41417ec4dd4b5c4c48532183b96/1188-principles-for-musculoskeletal-based-patient-reported-outcome-performance-measurement-development.pdf>
3. Antosh LJ, Svoboda SJ, Peck KY, Garcia EJ, Cameron KL. Change in KOOS and WOMAC scores in a young athletic population with and without anterior cruciate ligament injury. *Am J Sports Med*. 2018; 46(7):1606-1616.
4. Brooks JM, Chapman CG, Schroeder MC. Understanding treatment effect estimates when treatment effects are heterogeneous for more than one outcome. *Appl Health Econ Health Policy*. 2018;16(3):381-393.
5. Efron B, Tibshirani R. Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy. *Stat Sci*. 1986;1(1):54-77.
6. Fehring T. AAHS risk adjustment initiative: why is it important? *J Arthroplasty*. 2016;31:1148-1150.
7. Hawkins RJ, Thigpen CA. Selection, implementation, and interpretation of patient-centered shoulder and elbow outcomes. *J Shoulder Elbow Surg*. 2018;27(2):357-362.
8. Ingelsrud LH, Terwee CB, Terluin B, et al. Meaningful change scores in the Knee injury and Osteoarthritis Outcome Score in patients undergoing anterior cruciate ligament reconstruction. *Am J Sports Med*. 2018;46(5):1120-1128.
9. Lowry K, Brox W, Naas P, Tubb C, Muschler G, Dunn W. Musculoskeletal-based patient-reported outcome performance measures, where have we been—where are we going. *J Am Acad Orthop Surg*. 2019;27(13):e589-e595.
10. Meta F, Lizzio VA, Jildeh TR, Makhni EC. Which patient reported outcomes to collect after anterior cruciate ligament reconstruction. *Ann J*. 2017;2(5).
11. MOTION Group. Patient-reported outcomes in orthopaedics. *J Bone Joint Surg Am*. 2018;100(5):436-442.
12. National Quality Forum. Patient-reported outcomes in performance measurement. National Quality Forum; 2012. Accessed October 25, 2024. https://www.qualityforum.org/Publications/2012/12/Patient-Reported_Outcomes_in_Performance_Measurement.aspx
13. Porter M, Larsson S, Lee TH. Standardizing patient outcomes measurement. *N Engl J Med*. 2016;37(4):504-506.
14. Porter ME. What is value in health care? *N Engl J Med*. 2010; 363(26):2477-2481.
15. Roos E, Lohmander L. The Knee injury and Osteoarthritis Outcome Score (KOOS): from joint injury to osteoarthritis. *Health Qual Life Outcomes*. 2003;1:64.
16. Scherer JE, Moen MH, Weir A, Schmikli SL, Tamminga R, van der Hoeven H. Factors associated with a more rapid recovery after anterior cruciate ligament reconstruction using multivariate analysis. *Knee*. 2016;23(1):121-126.
17. Schilling P, Bozic KJ. Development and validation of perioperative risk-adjustment models for hip fracture repair, total hip arthroplasty, and total knee arthroplasty. *J Bone Joint Surg Am*. 2016;98(1):e2.
18. Schwartz AJ, Bozic KJ, Etzioni DA. Value-based total hip and knee arthroplasty: a framework for understanding the literature. *J Am Acad Orthop Surg*. 2019;27(1):1-11.
19. Shea KG, Carey JL, Richmond J, et al. The American Academy of Orthopaedic Surgeons evidence-based guideline on management of anterior cruciate ligament injuries. *J Bone Joint Surg Am*. 2015;97(8):672-674.

20. Siljander MP, McQuivey KS, Fahs AM, Galasso LA, Serdahely KJ, Karadsheh MS. Current trends in patient-reported outcome measures in total joint arthroplasty: a study of 4 major orthopaedic journals. *J Arthroplasty*. 2018;33(11):3416-3421.
21. Sura A, Shah NR. Pay-for-performance initiatives: modest benefits for improving healthcare quality. *Am Health Drug Benefits*. 2010; 3(2):135-142.
22. Thigpen C, Floyd SB, Chapman C, et al. Comparison of surgeon performance of rotator cuff repair risk adjustment toward a more accurate performance measure. *J Bone Joint Surg Am*. 2018; 100(24):2110-2117.
23. Yale New Haven Health Services Corporation—Center for Outcomes Research and Evaluation (YNHHSC/CORE). 2022 procedure-specific complication measure updates and specifications report: elective primary total hip arthroplasty (THA) and/or total knee arthroplasty (TKA)—version 11.0. Centers for Medicare & Medicaid Services; 2022. Accessed October 25, 2024. <https://www.cms.gov/files/document/2022-measure-updates-procedure-specific-complication-measure-updates-and-specifications-report.pdf>

APPENDIX

Risk-Adjusted Performance Measure Approach

To develop a risk-adjusted surgeon-level performance measure, we followed the subsequent steps as outlined in a previous publication.²²

Step 1: Select a Disease-Specific PRO Measure. The initial step in developing a performance measure is selecting a disease-specific PRO measure that is sensitive to changes after orthopaedic treatment. Many PROs exist for each joint, so care should be given to the measure selected.

Step 2: Baseline Risk Adjustment Factors. The selection of clinical control measures was guided by the American Academy of Orthopaedic Surgeons (AAOS)

evidence-based clinical practice guidelines for ACLR,¹⁹ in addition to clinical knowledge of patient characteristic bearing on clinical outcomes. Available SOS variables were mapped to each AAOS variable and included in the risk-adjusted model.

Step 3: Model Selection and Assessment. We established a statistical approach for modeling outcome change using a linear regression model estimated using ordinary least squares because the outcome variable was continuous with a large sample size. Although more complex functional forms are available, the linear specification provided the direct relationship between specific factors and patient outcomes.

Step 4: Surgeon Performance Scores. The predicted 6-month KOOS change values were estimated for each patient using the risk-adjusted linear model outlined above. The predicted change scores and observed change scores were averaged across all patients treated by a given surgeon. The performance scores for individual surgeons were then calculated as the difference between the risk-adjusted and observed KOOS values; the score represents the number of KOOS points better or worse than the surgeons' panel of patients achieved than expected.

The Efron bootstrap method was used to estimate performance scores and confidence intervals surrounding each surgeon's performance score. Performance scores were estimated from a random sample with replacement drawn from each surgeon's sample to create a new sample equal to the original sample size. This process was repeated 2000 times to generate a performance score distribution per surgeon. The median score with 95% confidence intervals from each surgeon-specific distribution was used as the estimate.