The Evaluation of HOOS-JR and KOOS-JR Performance: Implications for Policy and Reimbursement Structure

Rebekah Kleinsmith, Harsh Rajesh Parikh, Michelle Steege, Jessica Jewell, Alexander Kirk Reichard

INTRODUCTION:

Recent policy changes mandate the reporting of KOOS-JR and HOOS-JR at baseline and 1-year following total joint arthroplasty (TJA) in several healthcare settings. The accompanying reimbursement structure is based largely on the consistent reporting of these outcomes and the improvement experienced by patients (i.e. the increase in HOOS-JR or KOOS-JR scores), with the benchmark being a 22-point increase in HOOS or HOOS-JR, and a 20-point increase in KOOS or KOOS-JR. The purpose of this study is to evaluate performance (including correlation and inter-rater agreement) of the JR PROM for both HOOS and KOOS when compared to their full PROM counterpart.

METHODS:

All patients that underwent TJA within a large healthcare system from November 2016 to February 2025 were eligible for inclusion in this retrospective review. Patients were excluded on the basis of incomplete PROs at baseline (preoperatively) or 1-year postoperatively. PROs were collected prospectively via electronic assessment sent to patients by email and/or SMS. Outcomes of interest included baseline (preoperative) and 1-year postoperative HOOS/HOOS-JR for patients that underwent total hip arthroplasty (THA) and KOOS/KOOS-JR for those that underwent total knee arthroplasty (TKA).

Statistical analysis included descriptive statistics including counts and percentages for categorical variables, and means ± standard deviations for scale variables. Two-sample t-tests were utilized to determine differences between the full and JR tools at baseline, 1-year and the 1-year change for both HOOS and KOOS. Pearson's correlation tests were employed to determine the correlation between raw scores and 1-year change in scores for both HOOS (full and JR) and KOOS (full and JR) measures. Resultant correlation parameters were evaluated for visual skew throughout the spectrum of survey responses. Inter-rater agreement (κ) was determined between the full and JR tools for both HOOS and KOOS using the reported CMS cut-offs of 22 for HOOS and 20 for KOOS. Statistical significance was defined as $p \leq 0.05$.

RESULTS:

A total of 5,785 THA patients and 8,084 TKA patients met criteria for inclusion in this study. The average baseline HOOS and HOOS-JR were 36.8 and 42.8, respectively ($p < 0.001$). Baseline KOOS and KOOS-JR scores were 43.9 and 48.1, respectively ($p < 0.001$). There was a statistically significant difference in full PRO and JR scores at both the baseline and 1-year timepoints for HOOS and KOOS ($p < 0.001$ for all) (Table 1).

Pre-operative HOOS vs HOOS-JR calculate a Pearson's correlation (ρ) of 0.83 compared to 0.94 at one-year. Overall, 1-year ΔHOOS correlation between the two tools is 0.90, but suggests evidence of a skewed agreement. Graphical representation of the change in correlation with raw scores and score changes reveals a skewed agreement for HOOS, with lower correlation at lower functionality and lower change. Pre-operative KOOS vs KOOS-JR calculate a Pearson's correlation (ρ) of 0.92 compared to 0.94 at one-year. Overall, 1-year ΔKOOS correlation between the two tools is 0.93.

For the HOOS/HOOS-JR, the CMS-cutoff is a one-year increase in 22 points. When analyzing HOOS and HOOS-JR, the inter-rater correlation coefficient (κ) measured 0.93 [0.92, 0.94], suggesting 93% agreement between the two PROM measures in the attainment of the 22-point cutoff. The inter-rater correlation coefficient (κ) between KOOS and KOOS-JR was 0.85 [0.84, 0.86], suggesting 85% agreement between the two PROM measures when evaluating the 1-year CMS cutoff change of 20 points.

DISCUSSION AND CONCLUSION:

The HOOS-JR and KOOS-JR surveys offer strong overall correlation in the continuous 1-year change in scores; 90% for 1-year ΔHOOS correlation & 93% for 1-year ΔKOOS correlation. Both the HOOS-JR and KOOS-JR surveys provide reliable agreement to their full counterparts in the 1-year change, 93% for hip replacement and 85% for knee replacement, when utilizing CMS-established cutoffs. HOOS and HOOS-JR demonstrate a slight skew in agreement towards the higher end of the functional tool, which is not reflected in the KOOS and KOOS-JR tools. This suggests that the HOOS-JR tool might be overestimating functional ability at lower scores, potentially overshadowing patient functional impairment that is being captured by the full HOOS tool.

These findings have significant implications for health systems and reimbursement as this functional overestimation minimizes the improvement seen by patients, limiting the number of patients that reach the 22-point cutoff as set by CMS when the HOOS-JR is employed. The utilization of target outcome scores (i.e. a pre-defined 1-year target score of X) in tandem with the score improvement cutoff of 22 might better reflect postoperative success in this patient population and result in more just reimbursement for healthcare systems.
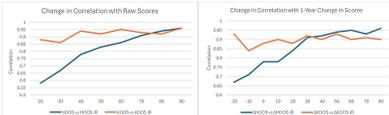
Figure 1. Line graph representing the correlation between the full PRO and the JR based on (a) raw scores and (b) 1-year score changes for both HOOS/HOOS-JR (blue lines) and KOOS/KOOS-JR (orange lines). HOOS /HOOS-JR correlation spectrums demonstrate a slight skew in agreement towards the higher end of the functional tool, which is not reflected in the KOOS and KOOS-JR tools. This indicates that the HOOS-JR tool might be overestimating functional ability at lower scores, potentially overshadowing patient functional impairment that is being captured by the full HOOS tool.

Table 1. Average preoperative, one-year, and change in HOOS, HOOS-JR, KOOS, and KOOS-JR scores.

	HOOS (n = 5,785)	HOOS-JR (n = 5,785)	p-value
Pre-Op	36.8 + 15.9 [36.4, 37.3]	42.8 + 15.0 [42.3, 43.2]	<0.01[1]
1-Year Post-Op	87.8 + 14.2 [87.3, 88.4]	86.3 + 14.9 [85.7, 86.9]	<0.01[1]
Δ1-Year	50.4 + 19.1 [49.7, 51.1]	43.0 + 18.6 [42.3, 43.7]	<0.01[1]
	KOOS (n = 8,084)	**KOOS-JR (n = 8,084)**	
Pre-Op	43.9 + 15.4 [43.5, 44.2]	48.1 + 12.9 [47.7, 48.4]	<0.01[1]
1-Year Post-Op	80.0 + 15.1 [79.5, 80.5]	77.0 + 14.6 [76.5, 77.5]	<0.01[1]
Δ1-Year	35.9 + 18.2 [35.3, 36.6]	28.8 + 16.7 [28.3, 29.4]	<0.01[1]

[1]Resulting p-value from a two-sample t-test.

Δ1-Year indicates improvement in score from baseline to 1-year postoperatively.

Scale variables represented as mean ± standard deviation [95% CI]

<u>Inter-Rater Agreement (κ)</u>
HOOS Jr vs Full HOOS Scoring

	ΔHOOS Full > 22	ΔHOOS Full ≤ 22	
ΔHOOS Jr > 22	2207 (38.2%)	20 (0.3%)	2227 (38.5%)
ΔHOOS Jr ≤ 22	171 (3.0%)	3387 (58.5%)	3558 (61.5%)
	2378 (41.1%)	3407 (58.9%)	5785

KOOS Jr vs Full KOOS Scoring

	ΔKOOS Full > 20	ΔKOOS Full ≤ 20	
ΔKOOS Jr > 20	2384 (29.5%)	50 (0.6%)	2434 (30.1%)
ΔKOOS Jr ≤ 20	481 (6.0%)	5169 (63.9%)	5650 (69.9%)
	2865 (35.4%)	5219 (64.6%)	8084

The blue shading indicates agreement between the two surveys, suggesting that both surveys meet the required CMS-benchmark, while the red shading confirms the opposite