The Journal of Arthroplasty 38 (2023) S162–S168

Contents lists available at ScienceDirect

The Journal of Arthroplasty

journal homepage: www.arthroplastyjournal.org

Primary Knee

Do Patient-Reported Outcomes Reflect Objective Measures of Function? Implications for Total Knee Arthroplasty

Brandon G. Hill, MSc [a], Shivesh Shah, BA [b], Wayne E. Moschetti, MD, MS [a, b], Peter L. Schilling, MD, MSc [a, b, *]

[a] Dartmouth Hitchcock Medical Center, One Medical Center Drive, Lebanon, New Hampshire
[b] The Geisel School of Medicine at Dartmouth, Hanover, New Hampshire

A R T I C L E I N F O

Article history:
Received 1 December 2022
Received in revised form
13 April 2023
Accepted 17 April 2023
Available online 25 April 2023

Keywords:
patient reported outcomes
standardized functional tests
functional outcomes
knee osteoarthritis
total knee arthroplasty
osteoarthritis initiative

https://www.aahks.org

A B S T R A C T

Background: Patient-reported outcomes (PROs) are used in research, clinical practice, and by federal reimbursement models to assess outcomes for patients who have knee osteoarthritis (OA) and total knee arthroplasty (TKA). We examined a large cohort of patients to determine if commonly used PROs reflect observed evaluation as measured by standardized functional tests (SFTs).
Methods: We used data from the Osteoarthritis Initiative, a 10-year observational study of knee osteoarthritis patients. Two cohorts were examined: 1) participants who received TKA (n = 281) and 2) participants who have native OA (n = 4,687). The PROs included Western Ontario and McMaster Osteoarthritis Index (WOMAC), Knee Injury and Osteoarthritis Outcome Score (KOOS), 12-Item Short Form Health Survey (SF-12), and Intermittent and Constant Pain Score (ICOAP). The SFTs included 20 m and 400 meter (m) walks and chair stand pace. Repeated measures correlation coefficients were used to determine the relationship between PROs and SFTs.
Results: The PROs and SFTs were not strongly correlated in either cohort. The magnitude of the repeated measures correlation (r_{rm}) between KOOS, WOMAC, SF-12, and ICOAP scores and SFT measurements in native knee OA patients ranged as follows: 400 m walk pace (0.08 to 0.20), chair stand pace (0.05 to 0.12), and 20 m pace (0.02 to 0.21), all with $P < .05$. In the TKA cohort, values ranged as follows: 400 M walk pace (0.00 to 0.29), chair stand time (0.02 to 0.23), and 20 M pace (0.03 to 0.30). Due to the smaller cohort size, the majority, but not all had P values < .05.
Conclusion: There is not a strong association between PROs and SFTs among patients who have knee OA or among patients who received a TKA. Therefore, PROs should not be used as a simple proxy for observed evaluation of physical function. Rather, PROs and SFTs are complementary and should be used in combination for a more nuanced and complete characterization of outcome.

Patient-reported outcome measures (PROs) are routinely used to assess outcomes for patients who have knee osteoarthritis (OA) and following total knee arthroplasty (TKA). Clinical research studies of OA and TKA have used PROs for years, and PROs are now increasingly used to inform clinical decision-making and patient-centered care. As value-based care reforms expand, so too have the potential roles for PROs [1,2]. For example, stakeholders are increasingly interested in not only using PROs to monitor and compare provider performance, but also as a way to adjust provider reimbursement [3]. Many questions remain regarding what PROs should and should not be used for [4]. While PROs are a centerpiece of patient-centered care, it is not clear whether they are also an appropriate measure of provider performance [5]. In fact, the literature suggests that the variation in PROs before and after arthroplasty surgery is heavily influenced by patient-related factors (outside of the control of providers) and that provider performance accounts for relatively little of the variation in PROs [6,7]. There is a perception that PROs should be correlated with objective measures of physical function, such as the ability to walk specified distances, climb stairs, and stand up from a seated position. Yet, the relationship between a patient's appraisal of their function and the observed reality of their function is not well-established among

No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statement refer to https://doi.org/10.1016/j.arth.2023.04.049.

* Address correspondence to: Peter L. Schilling, MD, MSc, Assistant Professor of Orthopaedic Surgery, Dartmouth Hitchcock Medical Center, Lebanon, NH 03766.

https://doi.org/10.1016/j.arth.2023.04.049
0883-5403/

Table 1
Cohort Demographics.

Variable	Native Knee Cohort	Total Knee Arthroplasty Cohort
	(N = 4,687)	(N = 281)
Mean age (range)	61 (45-79)	68 (45-79)
Sex (women), %	58.4	60.9
Race, %	White: 78.9	83.3
	Black: 18.3	12.1
	Asian: 1.0	1.1
	Other: 1.8	3.6
College Graduate, %	59.4	52.0
Income > $50 K, %	58.6	60.1
Mean body mass Index (range)	28.6 (16.9-48.7)	30.1 (21.1-29.9)
Charlson Comorbidity Index	0: 74.4	65.8
	1: 15.1	18.9
	2: 6.1	8.2
	3: 1.9	1.4
	>3: 1.2	1.1
Died before study end	304	13

patients who have knee OA or received a TKA [8]. Relatively few studies have characterized the relationship between PROs and validated objective measures of function as measured by standardized functional tests (SFTs). Moreover, most of the studies that do exist are underpowered and demonstrate mixed results [9−21], with some exceptions [22−24]. As the scope of use for PROs widens, it becomes increasingly important to understand this relationship, particularly as stakeholders consider using PROs to measure provider performance and inform reimbursement decisions. Without proper alignment between PROs and measurement goals, acting based on PROs may be unfounded and result in poor decision-making.

The purpose of this study was to examine the association between commonly used PROs and objective measures of physical function in a large, prospectively collected cohort of patients who have knee OA and a subset who received TKA. We compared PROs to performance-based tests of function selected by the National Institutes of Health sponsored Osteoarthritis Initiative (OAI), including 20-m walk pace, 400-walk pace, and chair stand pace. We will refer to these performance-based tests of function as standardized functional tests (SFTs) throughout the manuscript. We hypothesized that PROs are not strongly associated with SFTs in knee OA or TKA patients.

Material and Methods

Population

The study participants were selected from the National Institutes of Health (NIH) sponsored Osteoarthritis Initiative (OAI). The OAI was a multicentered, 10-year longitudinal, observational study of 4,796 men and women ages 45 to 79 years at enrollment with (29%), or at risk for (68%), symptomatic knee OA. Detailed descriptions of the eligibility criteria and study protocol have been published [25], and the study data can be found on the NIH website (https://nda.nih.gov/oai). The data are available to the public and are not individually identifiable, and therefore analyses would not involve human subjects and was exempt from institutional review board review.

Briefly, participants were excluded if they had rheumatoid or inflammatory arthritis, severe joint space narrowing, or total knee arthroplasty in both knees, or the requirement of ambulatory aids (other than a cane) for the majority of their walking. Annual evaluations of the OAI participants began on February 23, 2004 at 4 study sites (Baltimore, Maryland; Columbus, Ohio; Pittsburgh, Pennsylvania; and Pawtucket, Rhode Island). Missing clinic visit data for the entire OAI sample ranged from 10% at the 1-year clinic follow-up visit to 35% at the 8-year clinic follow-up visit. Sample sizes and missing data are characterized in detail in Tables 1-3.

Table 2
Standardized Functional Tests and Patient Reported Outcomes Collection Years. Shows Which Years Each Measure was Collected During the Osteoarthritis Initiative Study.

Outcome Measure	Visit Year									
	0	1	2	3	4	5	6	7	8	9
400 m Walk	X	-	X	-	X	-	-	-	X	-
20 m Walk	X	X	X	X	X	-	X	-	X	-
Chair Stands	X	X	X	X	X	-	X	-	X	-
KOOS	X	X	X	X	X	X	X	X	X	X
WOMAC	X	X	X	X	X	X	X	X	X	X
ICOAP	-	-	-	-	X	X	X	X	X	-
SF-12	X	X	X	X	X	-	X	-	X	-

From the OAI participants, 2 study cohorts were created. The first cohort was the native knee cohort. It contained the PRO and SFT scores for participants who had no knee or hip arthroplasty. Study participants who had pre-existing hip or knee arthroplasty were excluded from this cohort, leaving 4,687 OAI participants. The data for this cohort include the PRO and SFT scores from each participant's annual visits until a participant received a knee or hip arthroplasty. All subsequent measures for those who received a knee or hip arthroplasty are excluded from this cohort. Thus, the measures in this cohort were only from participants who had their original knees and hips.

The second cohort was the TKA cohort. It followed participants who received their first TKA(s) during the study and had at least 1 follow-up visit. Like the first cohort, all OAI participants who had pre-existing knee or hip arthroplasty were excluded, as were those who received a partial knee arthroplasty or hemiarthroplasty during the study. Also, participants who got bilateral TKAs, but not in the same year, were also dropped as they did not have a clear year to consider as the division for prearthroplasty and postarthroplasty. This cohort was composed of 281 OAI participants.

Measurements

Baseline demographics included age, sex, body mass index (BMI), race, education, and Charlson Comorbidity Index (Table 1), all measured at the start of the study [26]. Kellgren-Lawrence grades, as determined by radiographs, were collected for the majority of participants annually during years 0 through 4 [27]. The SFTs included 20-m walk pace, 400-m walk pace, and chair stand pace, and were chosen based on what was available within the NIH-sponsored OAI dataset. The PROs included the Western Ontario and McMaster Universities Arthritis Index (WOMAC), Knee Injury and Osteoarthritis Outcome Score (KOOS), 12-Item Short Form Survey (SF-12), and Intermittent and Constant Osteoarthritis Pain score (ICOAP). From these 4 surveys, 13 PRO scores and subscores were used: SF-12 Mental and Physical, WOMAC Pain, Stiffness, Daily Living/Activity, and Total, KOOS Pain, Symptoms, Function in Daily Living (ADL), Function in Sport and Recreation (Sports/Rec), and knee-related Quality of Life (QOL), ICOAP Constant, Intermittent, and Total. Collection years for each SFT and PRO can be found in Table 2. The resulting sample sizes can be found in Table 3.

Analyses

We calculated the correlation between PRO scores or subscores and the participants' performance on SFTs for each year. Repeated

Table 3
Standardized Functional Tests and Patient Reported Outcomes Correlations Sample Sizes per Cohort. Sample Sizes are a Function of the Cohort Populations, How Many Times a Measure was Collected, and Individual Patient Participation Over the Course of the Osteoarthritis Initiative Study.

Outcome Measure		Native Knees			TKA		
		400 m Pace	Chair Stand Pace	20 m Pace	400 m Pace	Chair Stand Pace	20 m Pace
SF-12	Physical	13,176	23,686	25,038	854	1,499	1,672
	Mental	13,176	23,686	25,038	854	1,499	1,672
KOOS	Pain	13,338	23,987	25,347	865	1,517	1,692
	Symptoms	13,337	23,987	25,347	865	1,517	1,692
	ADL	13,306	23,924	25,264	861	1,508	1,682
	Sports/Rec	9,751	17,578	18,196	496	855	913
	QOL	13,335	23,984	25,343	865	1,517	1,692
ICOAP	Constant Score	5,390	8,492	8,989	366	556	633
	Intermittent Score	5,387	8,484	8,983	366	556	633
	Total Score	5,386	8,484	8,982	366	556	633

measure correlations were used to account for participants being measured repeatedly across multiple visits, and the bias that can arise from differing patterns between participants versus within participants [28]. We handled missing data with pairwise deletions–an approach that assumes missing data were missing at random or missing completely at random. As such, all cases with data (even those with missing data) were used in the analyses.

The KOOS/WOMAC and ICOAP scores were measured for each knee of the participants. Scatter plots were used to visually evaluate the relationships between all PROs and SFTs. To handle the complexity posed by having both left-sided and right-sided measures for participants, we calculated 3 different sets of correlations: correlations only using the PRO scores and subscores of participants' best scoring knee (the healthiest knee), another only using the results of participants' worst scoring knee, and a last using the mean scores and subscores from both knees in a given visit. The *P* values were adjusted for multiple comparisons using Holm−Bonferroni methods (ie, 13 PRO scores and subscores, each compared to 3 SFT measures) [29].

To rule out age, weight, and health as confounding factors, each cohort was stratified into subgroups and the repeated measure correlations were calculated again. To rule out age, participants were divided by starting ages of 49 to 54, 55 to 64, and 65 to 79 years. The ranges were chosen so each subcohort contained roughly a third of the participants. For weight, initial BMI values were used. The subcohorts were grouped by underweight and normal (BMI < 25), overweight (BMI 25 to 29.9), and obese (BMI ≥ 30). Again, each subcohort contained about a third of the original cohorts. For health, a participant's initial Charlson Comorbidity Index was used [26]. The subcohorts were grouped by none (CCI = 0), mild (CCI = 1 to 2), and moderate/severe (CCI >2). Given the nonlinear nature of this score, the resulting groups were not as even: 75% of participants had no comorbidities, 22% were mild, and 3% were moderate or severe.

Results

Figure 1 shows the change in Kellgren-Lawrence grade (KL grade) distribution over time within the native knee and TKA cohorts, respectively. The KL grades were analyzed for each leg individually. In the native knee cohort, there was a small increase in the mean KL grade over the first 4 years (from 1.3 to 1.4), and only a modest number of participants showed OA progression as measured by KL grade. In contrast, the TKA cohort had higher overall KL grades and more rapid progression over the first 4 years (from 2.4 to 2.5).

Figures 2 and 3 show population trends across time in both PROs (KOOS, ICOAP, SF-12) and SFTs (20 m walk pace, 400 m walk pace, chair stand time) for the native knee and TKA cohorts, respectively. In the native knee cohort, there was very little change in any of these measures over time. This pattern is in sharp contrast to the time trends seen within the TKA cohort: mean PRO scores showed the expected decline before surgery with substantial improvement after TKA. The SFTs in this group showed a similar decline before

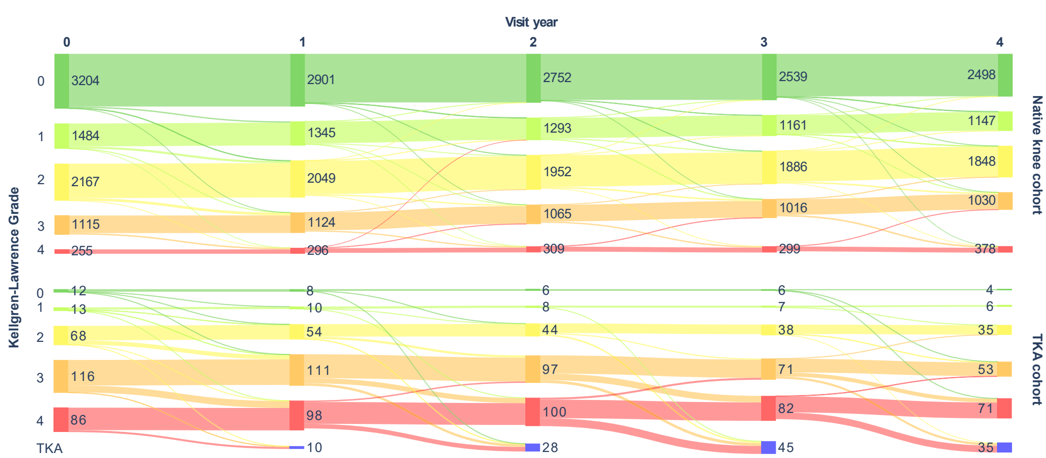
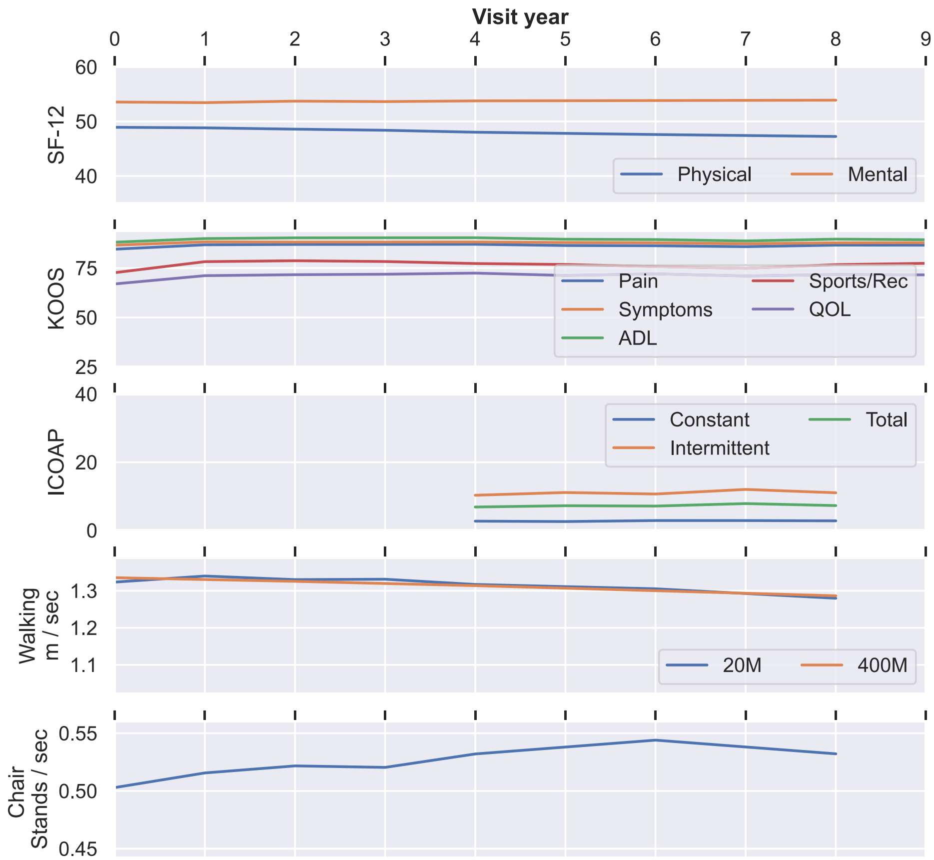


Fig. 1. Kellgren-Lawrence grade progressions. Osteoarthritis Initiative only measured Kellgren-Lawrence grades across all patients during the first 4 years, thus limiting the duration of this plot. For the native cohort, this includes each knee separately. For the total knee arthroplasty cohort, this only included knees that received a total knee arthroplasty during the OAI study (listed after Kellgren-Lawrence grade 4).



Participants

SF-12	4,626	4,162	3,970	3,831	3,775	-	3,369	-	3,148	-
KOOS	9,369	8,684	8,273	8,056	7,904	7,223	6,924	6,768	6,453	5,919
ICOAP	-	-	-	-	7,296	7,228	6,498	6,760	6,146	-
400M	4,461	-	3,456	-	3,048	-	-	-	2,373	-
20M	4,667	4,123	3,832	3,667	3,457	-	2,881	-	2,723	-
Chair Stands	4,455	3,874	3,661	3,431	3,315	-	2,701	-	2,553	-

Fig. 2. Mean standardized functional tests and patient reported outcomes scores for the native knee cohort. Sample sizes are given for each year and measure. Both Knee Injury and Osteoarthritis Outcome Score and Intermittent and Constant Pain Score were measured per knee, giving 2 measures per participant.

surgery with stabilization—as opposed to improvement—after TKA.

Figures 4 and 5 in the Appendix are representative examples of scatter plots for KOOS subscores and SFTs for each participant's least healthy knee in both the native knee and TKA populations, respectively. For the native knee cohort, the scatter plots demonstrate some weakly positive relationships between SFT pace and PRO scores, but in many of the plots there was no discernable relationship on visual inspection (Fig. 4 in the Appendix). Generally speaking, participants whom had better PRO scores demonstrated greater variation in their SFT pace than those who had worse PRO scores. Scatter plots of ADLs and SFTs seem to suggest that those measures have the strongest positive relationship, but the strength of the linear correlation was no greater than 0.15 for any of the 3 comparisons to SFT pace. Overall, no linear correlations between SFT and PRO exceeded 0.2 and most correlations were approximately 0.1. Except for SF-12 Mental, these correlations were all strongly statistically significant to $P < .05$ (Table 4 in Appendix).

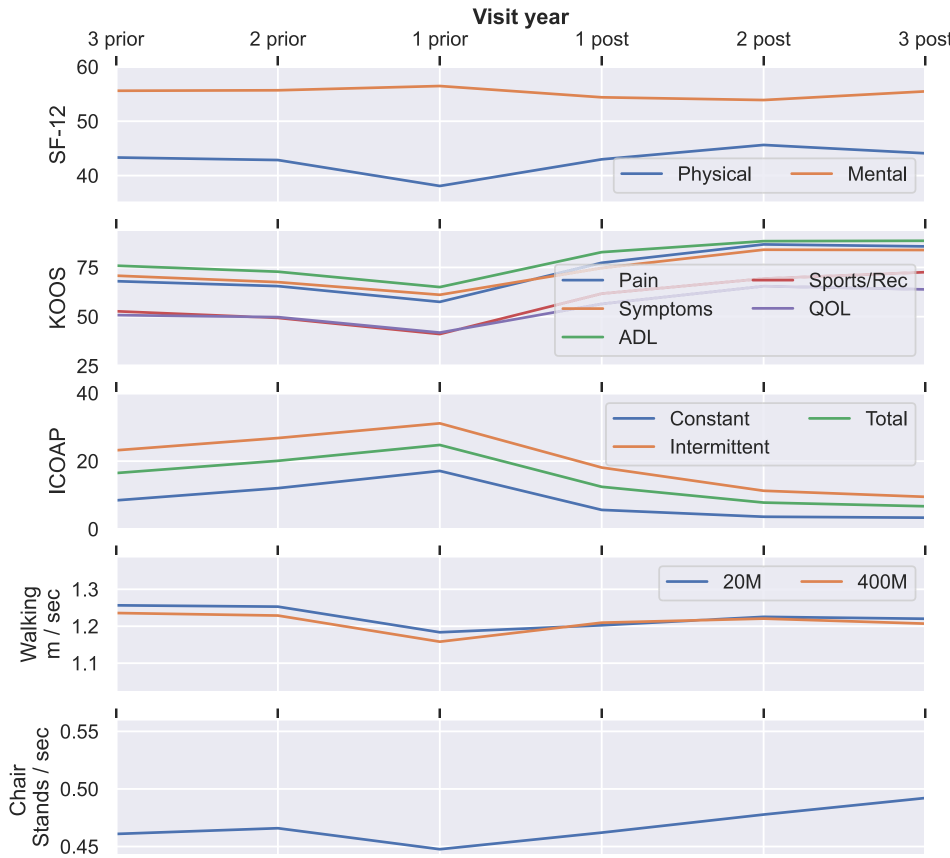
The scatter plots for the TKA cohort demonstrated the same form, direction, and strength of relationships between SFTs and PROs (Fig. 5 in the Appendix). The magnitudes of the correlations in the TKA cohort were like those of the native knee cohort: no correlations exceeded 0.3, and most were approximately 0.2. The majority of the P values for the TKA cohort were strongly statistically significant, even after Holm-Bonferroni corrections (Table 5 in the Appendix).

The data presented here only reflect correlations between each participant's SFT scores and the PROs for their worst-scoring knee. Similar correlations were calculated based on the PRO results of each participant's healthiest knee as well as the mean score between both knees. In both cases, no significant correlations were found, and P values were similarly low, thus ruling out results being driven by a stronger knee or overall knee health.

Cohorts were further subdivided into bands by age, BMI, and Charlson Comorbidity Index. Correlation scores were similarly low, with similar statistical significance.

Discussion

We present the relationships between PROs and measures of objective function in 2 large, prospectively collected cohorts of participants—first, in a cohort of participants at risk for knee



Participants

SF-12	204	188	204	163	133	90
KOOS	273	298	303	318	261	212
ICOAP	110	149	182	185	183	158
400M	95	78	94	50	57	39
20M	197	178	184	143	121	74
Chair Stands	181	161	158	115	108	65

Fig. 3. Mean standardized functional tests and patient-reported outcomes scores for the total knee arthroplasty cohort. Measures' means for the 3 yearly visits before and the 3 yearly visits after receiving a TKA. This only measures Knee Injury and Osteoarthritis Outcome Score and Intermittent and Constant Pain Score from the replaced knee (or both if a bilateral arthroplasty).

osteoarthritis, and second, in a cohort of participants who had knee arthritis who ultimately underwent a TKA. We found that commonly used PROs were not associated with objective measures of functional ability in either cohort, even when the functional tests had been collected in a highly controlled environment. Among participants having undergone TKA, we found expected improvements in both PROs and SFTs from preoperative to postoperative assessments; yet correlations between the classes of measures were universally low (roughly 0.1 to 0.3). In summary, among participants who had knee OA and TKA, we found no correlation between a participant's appraisal of their function and the observed reality of their function. This finding has direct relevance to the role of PROs in value-based payment reforms. Understanding the relationship between PROs and SFTs will facilitate the interpretation of each class of outcome measure so that informed choices can be made about the most appropriate outcome measure for a particular task such as measuring provider performance.

Our findings are consistent with a body of literature suggesting that PROs have limitations and should not be used as surrogates for measures of objective function. While PROs have become increasingly important in the transition to value-based orthopaedics care, SFTs remain the accepted measure of objective physical outcome following TKA [1]. Studies have examined the relationship between various PROs and SFTs, though conclusions have been inconclusive and weak due to limited sample sizes. In 2 separate studies, Bolink et al. investigated the relationship between WOMAC and SFTs, including sit-to-stand transfers and inertial sensor-based gait analyses. Sample sizes in each study were 20 and 36 participants, and moderate associations were noted (Spearman's rho 0.3 to 0.5 and 0.5 to 0.7, respectively) [30]. While some studies have found moderate degrees of association between SFTs and PROs before and following TKA [8–10], sample sizes were limited and there are multiple other studies that show low or only moderate degrees of correlation [11,12,22]. In 2017, Naili et al. conducted a similar study of 28 patients who had undergone TKA. Although TKA patients showed improvements in objective measures of function, self-reported measures of function could not differentiate patients who had improved SFT outcomes from those who did not [12].

Our study has several strengths. To our knowledge, this is the largest prospective study comparing PROs and SFTs in patients who have knee OA and patients who have undergone TKA. This is reflected by the strongly statistically significant estimates of correlation. Also, OAI data comes from a large, NIH-funded study that was conducted with forethought and attention to detail. As such, PROs and SFTs were collected in a highly controlled environment with consistent application of protocols across participants and time periods. Uncontrolled sampling and response bias are known problems with PROs that the original research team thoughtfully addressed with their study protocols.

Our study has potential limitations. Although the sample size was large, it was restricted to patients being treated at only 4 hospitals participating in the OAI study. As such, one could question the study's generalizability. Also, it is valid to question whether the performance-based tests used in this study (so called SFTs) are the right measures to assess physical function in these patients. The measures used in this study are not likely to encompass all attributes of physical function, but were used because they were the ones selected by those who architected the OAI dataset. Future prospective work should look to the set of performance-based tests of physical function recommended by the Osteoarthritis Research Society International (OARSI), a minimum core set of tests chosen by an expert, multidisciplinary advisory group that partially overlap with those available in the OAI (30-second chair stand test, 40M walk test, and stair climb test) [31–33]. In addition, the literature shows significant variance in both PRO use and preference [34]. In the current study, we only measured KOOS, WOMAC, SF-12, and ICOAP. Other validated measurements were not included in our analyses and should be the subjects of subsequent studies. Moreover, our results could be biased if missing data was not missing at random. Furthermore, our analysis was simplistic by design, consisting only of basic correlations between PROs and SFTs across time. Our intention was to empirically evaluate the common perception that PROs ought to correlate with objective physical function.

If objective changes in a patient's physical performance are important, and PROs do not fully capture them, how do we collect this information in patients? Routine administration of SFTs in a clinic or human performance lab is not practical since collecting this data "in vitro" is expensive and tedious. Biomechanical sensors in the form of "wearables" present a potential path forward [35–37]. It may be possible to map raw data from the accelerometers of these devices to existing validated measures of objective function. If successful, wearables have the added benefit of being "in vivo" measures of patients' function in daily life, outside the confines of a clinic or lab. Research has shown the potential for gait kinematics to be associated with PROs, and further research could guide advancements in the use of wearables as predictors of functional status [38].

In conclusion, PROs are not strongly associated with SFTs in knee OA or TKA patients. As such, they should not be used as a proxy for observed physical function. Stated simply, PROs and SFTs are not measuring the same thing. They are, by design, characterizing different domains of pain and function from different vantage points. This should not be controversial. After all, it would be considered crudely reductionist in most fields to contend that any single measure (or class of measures) captures all important attributes of a complex phenomenon. The lack of correlation of these outcome measures does not say that one class of measures is clearly good and one bad. Rather, it emphasizes the importance of using these measures responsibly with an appreciation for each measure's limitations (where it excels and where it has shortcomings). It also suggests that these measures might actually be complementary and combining these measures will give us a much more nuanced and a complete characterization of our patients' outcomes [36,39–41]. This should be the concern of future research, the results of these studies having broad-ranging implications for patient selection, patient expectations, shared decision-making, and comparisons of provider performance and quality.

References

[1] Makhni EC, Baumhauer JF, Ayers D, Bozic KJ. Patient-reported outcome measures: how and why they are collected. Instr Course Lect 2019;68:675–80.

[2] Lan RH, Bell JW, Samuel LT, Kamath AF. Evolving outcome measures in total knee arthroplasty: trends and utilization rates over the past 15 years. J Arthroplasty 2020;35:3375–82.

[3] Starring H, Waddell WH, Steward W, Schexnayder S, McKay J, Leonardi C, et al. Total knee arthroplasty outcomes in patients with medicare, medicare advantage, and commercial insurance. J Knee Surg 2020;33:919–26.

[4] Wang Y, Yin M, Zhu S, Chen X, Zhou H, Qian W. Patient-reported outcome measures used in patients undergoing total knee arthroplasty. Bone Joint Res 2021;10:203–17.

[5] Ehlers AP, Khor S, Cizik AM, Leveque J-CA, Shonnard NS, Oskouian Jr RJ, et al. Use of patient-reported outcomes and satisfaction for quality assessments. Am J Manag Care 2017;23:618–22.

[6] Chen Q, Beal EW, Okunrintemi V, Cerier E, Paredes A, Sun S, et al. The association between patient satisfaction and patient-reported health outcomes. J Patient Exp 2019;6:201–9.

[7] Collins NJ, Roos EM. Patient-reported outcomes for total hip and knee arthroplasty: commonly used instruments and attributes of a "good" measure. Clin Geriatr Med 2012;28:367–94.

[8] Kirschberg J, Goralski S, Layher F, Sander K, Matziolis G. Normalized gait analysis parameters are closely related to patient-reported outcome measures after total knee arthroplasty. Arch Orthop Trauma Surg 2018;138:711–7.

[9] Biggs PR, Whatling GM, Wilson C, Holt CA. Correlations between patient-perceived outcome and objectively-measured biomechanical change following Total Knee Replacement. Gait Posture 2019;70:65–70.

[10] Liebensteiner MC, Herten A, Gstoettner M, Thaler M, Krismer M, Bach CM. Correlation between objective gait parameters and subjective score measurements before and after total knee arthroplasty. Knee 2008;15:461–6.

[11] Luna IE, Kehlet H, Peterson B, Wede HR, Hoevsgaard SJ, Aasvang EK. Early patient-reported outcomes versus objective function after total hip and knee arthroplasty: a prospective cohort study. Bone Joint J 2017;99-B:1167–75.

[12] Naili JE, Iversen MD, Esbjörnsson A-C, Hedström M, Schwartz MH, Häger CK, et al. Deficits in functional performance and gait one year after total knee arthroplasty despite improved self-reported function. Knee Surg Sports Traumatol Arthrosc 2017;25:3378–86.

[13] Mehta SP, Morelli N, Prevatte C, White D, Oliashirazi A. Validation of physical performance tests in individuals with advanced knee osteoarthritis. HSS J 2019;15:261–8.

[14] Graff C, Hohmann E, Bryant AL, Tetsworth K. Subjective and objective outcome measures after total knee replacement: is there a correlation? ANZ J Surg 2016;86:921–5.

[15] Frimpong E, van der Jagt DR, Mokete L, Pietrzak J, Kaoje YS, Smith A, et al. Improvements in objectively measured activity behaviors do not correlate with improvements in patient-reported outcome measures following total knee arthroplasty. J Arthroplasty 2020;35:712–719.e4.

[16] Stratford P, Kennedy D, Clarke H. Confounding pain and function: the WOMAC's failure to accurately predict lower extremity function. Arthroplast Today 2018;4:488–92.

[17] Fransen BL, Pijnappels M, Butter IK, Burger BJ, van Dieën JH, Hoozemans MJM. Patients' perceived walking abilities, daily-life gait behavior and gait quality before and 3 months after total knee arthroplasty. Arch Orthop Trauma Surg 2022;142:1189–96.

[18] Soeno T, Mochizuki T, Tanifuji O, Koga H, Murayama T, Hijikata H, et al. No differences in objective dynamic instability during acceleration of the knee with or without subjective instability post-total knee arthroplasty. PLoS One 2018;13:e0194221.

[19] Fransen BL, Mathijssen NMC, Slot K, de Esch NHH, Verburg H, Temmerman OPP, et al. Gait quality assessed by trunk accelerometry after total knee arthroplasty and its association with patient related outcome measures. Clin Biomech 2019;70:192–6.

[20] Van Onsem S, Verstraete M, Van Eenoo W, Van Der Straeten C, Victor J. Are TKA kinematics during closed kinetic chain exercises associated with patient-reported outcomes? A preliminary analysis. Clin Orthop Relat Res 2020;478:255–63.

[21] Disantis AY, Piva SR, Irrgang JJ. Standardized patient reported outcomes do not capture functional deficits of patients following contemporary total knee replacement: descriptive study. J Exerc Sports Orthop 2018;5. https://doi.org/10.15226/2374-6904/5/1/00167.

[22] Ekanayake CD, DeMik DE, Glass NA, Kotseos C, Callaghan JJ, Ratigan BL. Comparison of patient-reported outcomes and functional assessment using a marker-less image capture system in end-stage knee arthritis. J Arthroplasty 2022;37:2158–63. https://doi.org/10.1016/j.arth.2022.05.039.

[23] Unnanuntana A, Ruangsomboon P, Keesukpunt W. Validity and responsiveness of the two-minute walk test for measuring functional recovery after total knee arthroplasty. J Arthroplasty 2018;33:1737–44.

[24] Van Onsem S, Verstraete M, Dhont S, Zwaenepoel B, Van Der Straeten C, Victor J. Improved walking distance and range of motion predict patient satisfaction after TKA. Knee Surg Sports Traumatol Arthrosc 2018;26:3272–9.

[25] Nevitt MC, Felson DT, Lester G. The Osteoarthritis Initiative: protocol for the cohort study. p. 19-20. http://oai.epi-ucsf.org/datarelease/docs/StudyDesignProtocol.pdf. [Accessed 21 June 2016].

[26] Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40:373–83.

[27] Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. Ann Rheum Dis 1957;16:494–502.

[28] Bakdash JZ, Marusich LR. Repeated measures correlation. Front Psychol 2017;8:456.

[29] Holm S. A simple sequentially rejective multiple test procedure. Scand Stat Theor Appl 1979;6:65–70.

[30] Bolink SAAN, Lenguerrand E, Brunton LR, Wylde V, Gooberman-Hill R, Heyligers IC, et al. Assessment of physical function following total hip arthroplasty: inertial sensor based gait analysis is supplementary to patient-reported outcome measures. Clin Biomech 2016;32:171–9.

[31] Dobson F, Hinman RS, Roos EM, Abbott JH, Stratford P, Davis AM, et al. OARSI recommended performance-based tests to assess physical function in people diagnosed with hip or knee osteoarthritis. Osteoarthritis Cartilage 2013;21:1042–52.

[32] Dobson F, Hinman RS, Hall M, Terwee CB, Roos EM, Bennell KL. Measurement properties of performance-based measures to assess physical function in hip and knee osteoarthritis: a systematic review. Osteoarthritis Cartilage 2012;20:1548–62.

[33] Suwit A, Rungtiwa K, Nipaporn T. Reliability and validity of the osteoarthritis research society international minimal core set of recommended performance-based tests of physical function in knee osteoarthritis in community-dwelling adults. Malays J Med Sci 2020;27:77–89.

[34] Zwiers R, Weel H, Mallee WH, Kerkhoffs GMMJ, van Dijk CN. Ankle Platform Study Collaborative – science of Variation Group. Large variation in use of patient-reported outcome measures: a survey of 188 foot and ankle surgeons. Foot Ankle Surg 2018;24:246–51.

[35] Boekesteijn R, Smolders J, Busch V, Keijsers N, Geurts A, Smulders K. Objective monitoring of functional recovery after total knee and hip arthroplasty using sensor-derived gait measures. PeerJ 2022;10:e14054.

[36] Crizer MP, Kazarian GS, Fleischman AN, Lonner JH, Maltenfort MG, Chen AF. Stepping toward objective outcomes: a prospective analysis of step count after total joint arthroplasty. J Arthroplasty 2017;32:S162–5.

[37] Youn I-H, Leutzinger T, Youn J-H, Zeni JA, Knarr BA. Self-reported and performance-based outcome measures estimation using wearables after unilateral total knee arthroplasty. Front Sports Act Living 2020;2:569932.

[38] Nüesch C, Ismailidis P, Koch D, Pagenstert G, Ilchmann T, Eckardt A, et al. Assessing site specificity of osteoarthritic gait kinematics with wearable sensors and their association with patient reported outcome measures (PROMs): knee versus hip osteoarthritis. Sensors 2021;21:5363. https://doi.org/10.3390/s21165363.

[39] Whatling GM, Biggs PR, Wilson C, Holt CA. Assessing functional recovery following total knee replacement surgery using objective classification of level gait data and patient-reported outcome measures. Clin Biomech 2022;95:105625.

[40] Christensen JC, Blackburn B, Browning B, Wilbur C, Trinity JD, Gililland JM, et al. Patient-reported outcomes measurement information system physical function and knee injury and osteoarthritis outcome score relationship on performance measures in people undergoing total knee arthroplasty. Disabil Rehabil 2022:1–9.

[41] Capin JJ, Bade MJ, Jennings JM, Snyder-Mackler L, Stevens-Lapsley JE. Total knee arthroplasty assessments should include strength and performance-based functional tests to complement range-of-motion and patient-reported outcome measures. Phys Ther 2022;102:pzac033. https://doi.org/10.1093/ptj/pzac033.

Appendix

Table 4
Standardized Functional Tests and Patient Reported Outcomes Repeated Measure Correlations for the Native Knee Cohort.

		20 m Pace			400 m Pace			Chair Stands Pace		
		r_{rm}	Samples	P	r_{rm}	Samples	P	r_{rm}	Samples	P
SF-12	Physical	0.208	25,038	.000	0.197	13,176	.000	0.097	23,686	.000
	Mental	0.020	25,038	.011	0.019	13,176	.079	0.017	23,686	.035
KOOS	Pain	0.117	25,347	.000	0.103	13,338	.000	0.116	23,987	.000
	Symptoms	0.099	25,347	.000	0.096	13,337	.000	0.095	23,987	.000
	ADL	0.142	25,264	.000	0.130	13,306	.000	0.133	23,924	.000
	Sports/Rec	0.101	18,196	.000	0.092	9,751	.000	0.102	17,578	.000
	QOL	0.106	25,343	.000	0.093	13,335	.000	0.117	23,984	.000
ICOAP	Constant	0.094	8,989	.000	0.115	5,390	.000	0.046	8,492	.005
	Intermittent	0.066	8,983	.000	0.081	5,387	.001	0.104	8,484	.000
	Total	0.093	8,982	.000	0.107	5,386	.000	0.107	8,484	.000

Table 5
Standardized Functional Tests and Patient-Reported Outcomes Repeated Measure Correlations for the Total Knee Arthroplasty Cohort.

		20 m Pace			400 m Pace			Chair Stands Pace		
		r_{rm}	Samples	P	r_{rm}	Samples	P	r_{rm}	Samples	P
SF-12	Physical	0.297	1,672	.000	0.288	854	.000	0.176	1,499	.000
	Mental	0.027	1,672	1.000	−0.003	854	1.000	0.017	1,499	1.000
KOOS	Pain	0.151	1,692	.000	0.079	865	.510	0.196	1,517	.000
	Symptoms	0.143	1,692	.000	0.089	865	.333	0.174	1,517	.000
	ADL	0.197	1,682	.000	0.160	861	.002	0.221	1,508	.000
	Sports/Rec	0.214	913	.000	0.193	496	.018	0.206	855	.000
	QOL	0.187	1,692	.000	0.147	865	.005	0.233	1,517	.000
ICOAP	Constant	0.201	633	.001	0.178	366	.432	0.064	556	1.000
	Intermittent	0.127	633	.159	0.110	366	1.000	0.083	556	.965
	Total	0.178	633	.007	0.159	366	.579	0.075	556	1.000

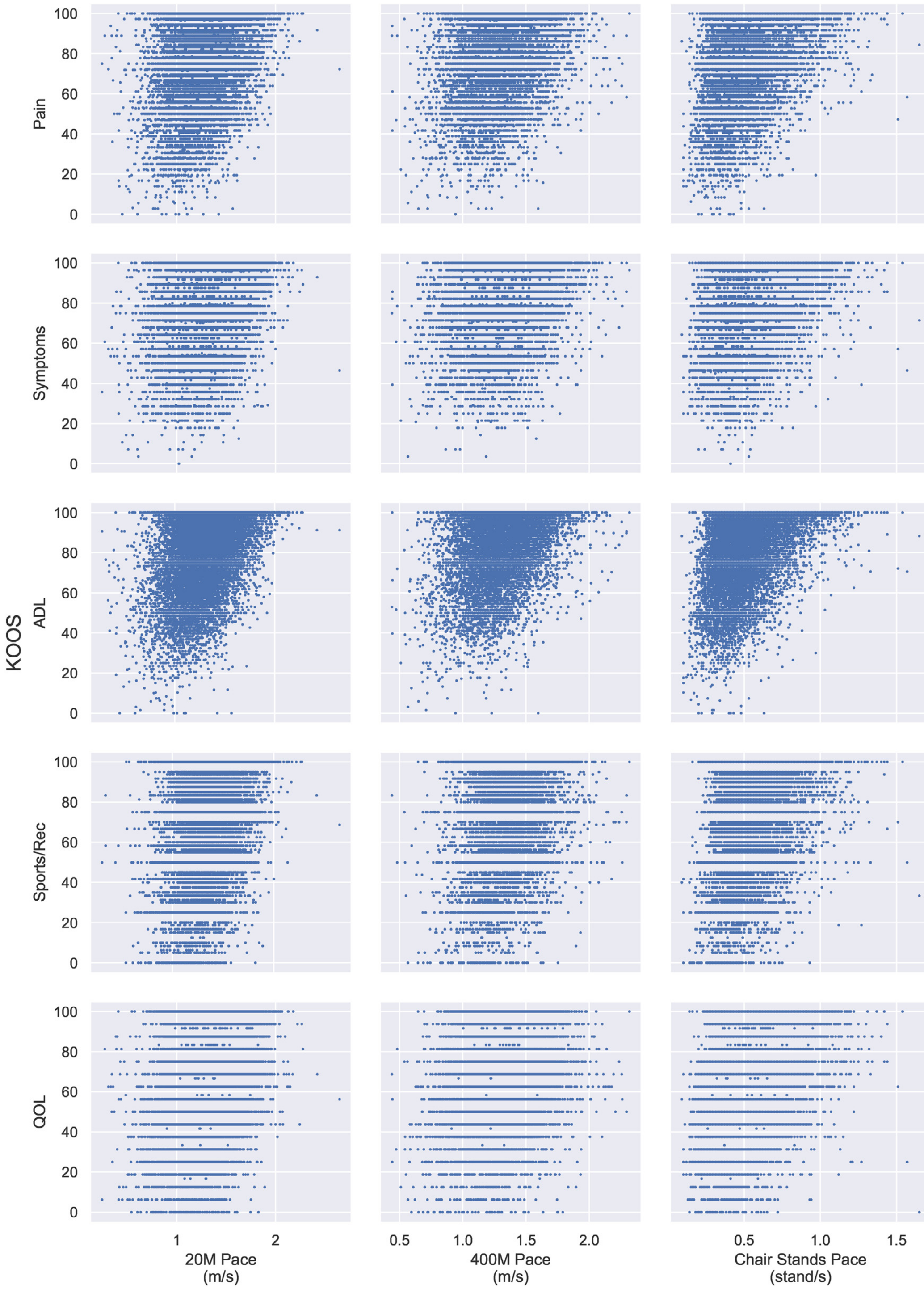


Fig. 4. Scatter plots of patient-reported outcome scores (Knee Injury and Osteoarthritis Outcome Subcores) and standardized functional tests pace for the native knee cohort.

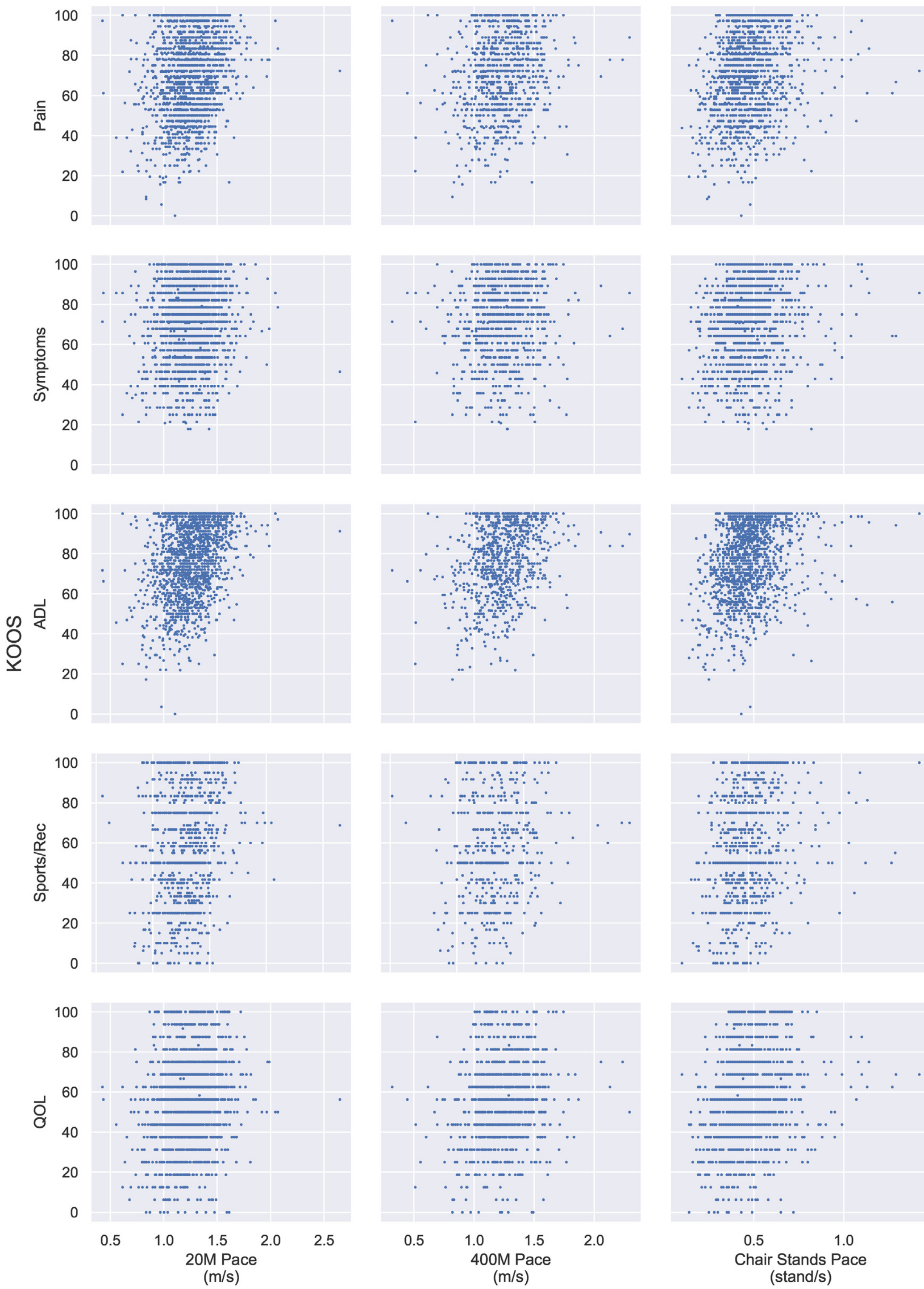


Fig. 5. Scatter plots of patient-reported outcome scores (Knee Injury and Osteoarthritis Outcome Subcores) and standardized functional tests pace for the total knee arthroplasty cohort.