Clinical Rheumatology (2019) 38:1469–1476
https://doi.org/10.1007/s10067-018-04414-z

ORIGINAL ARTICLE

Magnetic resonance imaging of patellofemoral osteoarthritis: intertester reliability and associations with knee pain and function

Daniel L. Riddle[1] · Josephina A. Vossen[2] · Kevin B. Hoover[2]

Received: 8 November 2018 / Revised: 12 December 2018 / Accepted: 25 December 2018 / Published online: 7 January 2019

Abstract

Objectives We examined the intertester reliability of patellofemoral compartment (PFC) osteoarthritis (OA) severity using magnetic resonance images (MRI) and a modified Kellgren and Lawrence (K&L) system. Second, we determined if these grades were associated with clinical tests of PFC involvement or self-reported pain/difficulty with stair climbing. Third, we assessed the association between PFC OA severity and knee pain or disability, after accounting for potential confounders including tibiofemoral OA severity.

Method We examined the 9-year Osteoarthritis Initiative data from 114 subjects in the year prior to undergoing knee arthroplasty. The weighted kappa (κ_w) was used to determine intertester reliability, and the Pearson chi-square was used to assess associations among PFC OA scores and clinical tests. Multiple regressions were used to determine independent associations between self-reported pain/function and PFC OA.

Results Reliability was substantial ($\kappa_w = 0.73$ (SE = 0.05)). Chi-square associations between PFC OA severity and clinical tests were not significant ($p > 0.05$). Multiple regression models between PFC OA and self-reported pain or function scores were not significant ($p > 0.05$).

Conclusions MRI-based measures of PFC OA were highly reliable indicating that musculoskeletal radiologists can reliably grade the PFCs of subjects using MRI. The extent of PFC OA is not associated with either clinical tests of PFC involvement or activities associated with PFC pain in persons with moderate to severe symptomatic tibiofemoral and PFC OA.

Keywords Function · MRI · Osteoarthritis · Pain · Patellofemoral

Knee osteoarthritis is a debilitating disease affecting approximately one third of older adults [1]. The structure of the knee joint with two tibiofemoral compartments, medial and lateral, and the patellofemoral compartment (PFC) can result in compartment-specific osteoarthritis. The PFC either in isolation or in combination with the medial tibiofemoral compartment is frequently symptomatic, particularly in middle-aged and older women with osteoarthritis [2, 3].

Radiographic evaluation of the PFC is optimally performed using multiple views, including the lateral and skyline, or tangential, views [4]. Radiographic views specifically designed for the PFC are not acquired in one of the larger longitudinal clinical and imaging osteoarthritis studies—the Osteoarthritis Initiative (OAI) [4]. However, axial and sagittal MRI sequences are available to evaluate the PFC in this and other longitudinal studies of knee osteoarthritis. Recently, MRI sequences and a modified Kellgren and Lawrence (K&L) grading scale (to accommodate MRI imaging) for PFC osteoarthritis have been used to generate reliable estimates of PFC OA [5, 6].

Kobayashi and colleagues [6] recently suggested that a simple-to-obtain measure of PFC OA severity was needed versus more time-consuming and complex magnetic resonance imaging (MRI) measures such as the MRI Osteoarthritis Knee Score (MOAKS) [7]. The investigators examined, in part, the reliability of PFC OA severity using MRIs and a modified K&L grading system originally reported

✉ Daniel L. Riddle
dlriddle@vcu.edu

[1] Departments of Physical Therapy, Orthopaedic Surgery and Rheumatology, Basement, West Hospital, Room B-100, Virginia Commonwealth University, Richmond, VA 23298-0224, USA

[2] Department of Radiology, Virginia Commonwealth University Health System, Richmond, VA 23298, USA

by Riddle et al. [5]. The weighted kappa, a chance-corrected reliability coefficient of $\kappa = 0.49$, was reported for intertester reliability and corresponds to a moderate level of agreement [8]. Kobayashi et al. concluded that MRI-derived K&L grades for the PFC showed promise, but that additional work was needed to determine the extent to which these measures associate with clinical findings in patients with knee pain.

In the current study, we were interested in building on our prior work and that of Kobayashi and colleagues [6]. Our study had three purposes. First, we determined the extent of intertester reliability for PFC osteoarthritis grades using MRI and a modified K&L system [5] applied by two experienced musculoskeletal radiologists. Kobayashi and colleagues reported moderate agreement between experienced radiologists ($\kappa = 0.49$), and we hypothesized that among experienced musculoskeletal radiologists, reliability would be higher. Second, to better understand the clinical relevance of MRI-derived PFC K&L grades, we determined if these grades were significantly associated with results obtained from self-reported pain or difficulty with stair climbing or clinical tests of PFC involvement, the patellar grind test and the patellar crepitus test. Both clinical tests provide reliable data when conducted by trained examiners [9, 10]. Stair climbing is an activity that has been associated with PFC inflammation and pain [11]. Given that PFC compressive forces increase with stair climbing, as compared to level walking and other less strenuous activities [12], we hypothesized that worse PFC OA would be associated with greater pain and greater difficulty with stair climbing. Third, we determined if measures of PFC OA severity were associated with knee pain or disability, as measured with the WOMAC scale, after accounting for potential confounders including tibiofemoral OA severity. Given the importance of the PFC in daily function [2], we hypothesized that greater PFC OA would be associated with worse (higher) WOMAC pain and function scores after adjustment for potential confounding.

Materials and methods

Source of data: the Osteoarthritis Initiative

The Osteoarthritis Initiative (OAI) is a National Institutes of Health and privately funded prospective community-based 9-year longitudinal study of subjects with radiographic knee osteoarthritis (OA) or at high risk for knee OA [4]. Knee OA risk was confirmed if the subjects were overweight or obese, reported prior knee injury or surgery, had knee symptoms, or had a family history of knee replacement surgery. Subjects were recruited from the communities of four study sites. The study was approved by the Institutional Review Boards at the following sites: (1) University of Pittsburgh in Pittsburgh, Pennsylvania; (2) University of Maryland in Baltimore, Maryland; (3) Ohio State University in Columbus, Ohio; and (4) Memorial Hospital of Rhode Island, in Pawtucket, Rhode Island. All subjects signed an IRB-approved consent form prior to participation.

A total of 17,457 women and men between the ages of 45 and 79 years were screened, and 4796 were enrolled. The most common exclusionary criteria were magnetic resonance imaging height and weight restrictions ($n = 2328$), recruitment thresholds for age and gender ($n = 2954$), and dropouts prior to the enrollment visit ($n = 4381$). Enrollment began in 2004, and publically available data are available with 9 years of yearly follow-up.

Subjects

Over the 9-year study period, a total of 427 subjects underwent knee arthroplasty (KA) on at least one knee. A subset of these subjects with preoperative magnetic resonance images ($N = 114$) were selected for the current study because of the need for modified K&L grades of index PFCs of presurgical knees, to be used in a future study of a KA appropriateness classification system [13]. Additionally, persons who undergo KA over the next year are likely to have more substantial tibiofemoral and PFC OA and greater pain than persons with milder symptoms. Because of our focus on the PFC, we wanted to include persons with the full spectrum of PFC OA severity, from mild to severe, and a sample of persons who had KA surgery over the next year were likely to have the full spectrum of PFC involvement. Subjects had an approximately equal percentage of preoperative radiographic tibiofemoral K&L grades of 1 or 2 ($n = 22$, 20%), or 4 ($n = 23$, 20%). A total of 60% of the sample had tibiofemoral K&L grades of 3 ($n = 69$). The yearly distribution of KA surgeries over the study period is illustrated in Fig. 1. Because our KA sample had surgery at different time points (i.e., the surgeries were time varying), and some subjects had multiple arthroplasty surgeries over the study period, we selected the knee with the earliest knee arthroplasty for each subject.

Radiographic tibiofemoral osteoarthritis severity assessment

Standing anterior knee radiographs were used in the current study for tibiofemoral OA grading and were obtained yearly (except for years 5 and 7 because OAI did not include in-person visits during these years) over the study period on all subjects using a highly standardized, reliable, and valid approach that positioned the standing subject's knees in 20° of flexion. Subjects were instructed to evenly distribute body weight during radiographic acquisition [14]. An extensive adjudication process was used for K&L grades for all knees over all time periods. Two central site readers and a third adjudicator, all either a rheumatologist or a musculoskeletal radiologist

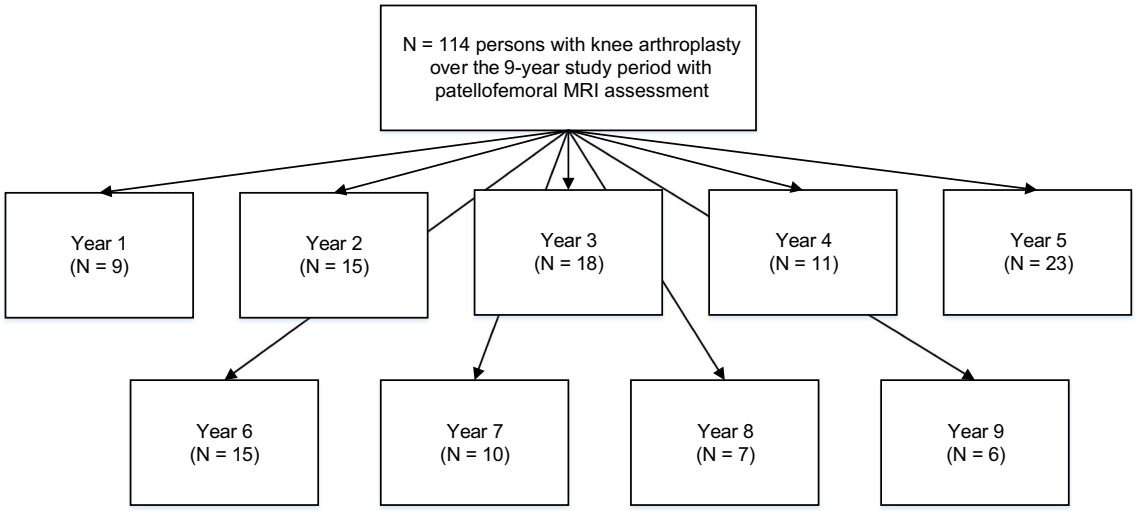

Fig. 1 Summary of the year of knee arthroplasty for the subjects in the sample

with extensive training, served as radiographic readers, and all K&L grades have been obtained on the OAI website. Radiographic images of lateral or sunrise views of the PFC were not obtained in the OAI study. For subjects who had surgery in years 6 and 8 ($n = 22$), we used the MRI and radiographic data from years 4 and 6, respectively.

MRI measures of patellofemoral compartment osteoarthritis

MRIs were used in the current study to grade the extent of PFC OA. The images were acquired using identical 3.0-Tesla Siemens Trio systems at each site. Sequences evaluated were sagittal 3D DESS (TR/TE 16.32/4.70, flip angle 25°, slice thickness 0.7 mm); sagittal proton density–weighted turbo spine echo T2-weighted (TR/TE 3000/32.0, slice thickness 3.0 mm); and axial multiplanar reformats based on sagittal 3D DESS. Two experienced sub-specialty trained musculoskeletal radiologists, one with 5 years (JAV) and 11 years (KBH) of clinical experience, underwent pre-study calibration and training by applying an operationally defined modified K&L osteoarthritis grading system we developed in a prior study [5] to a set of 10 knee MRIs. These knees were randomly selected from subjects who had bilateral knee MRIs available. Only the contralateral (i.e., non-surgical) knees were included in the calibration sample, and these knee images were not used in the full study. During the pre-study phase, the two radiologists independently graded the PFC of 10 MRIs using the grading system and then together discussed discrepancies. All subjects in the OAI underwent MRI imaging yearly during the first 4 years of study and also at year 6 and year 8. The preoperative MR images used in this study were taken within 1 year of the patients' knee replacement surgeries.

Kellgren and Lawrence originally described radiographic grading on a 0 to 4 scale based only on radiographic data. The original description did not include application to the patellofemoral joint but it has since been extensively applied [15–17]. Because we used MRIs, both in this and a prior study [5], we slightly modified the K&L system to apply to PFC MRI data. For the MRIs, PFCs judged as normal were scored a 0 and a joint with no definite osteophyte or joint space narrowing but with limited cartilage or periarticular changes was graded as 1. When a definite osteophyte was present without joint space narrowing, a grade of 2 was given. A grade of 3 indicated significant cartilage loss on at least one facet or trochlear surface with joint space narrowing. A grade of 4 indicated complete cartilage loss (i.e., complete loss of joint space) of either the medial or lateral PFC of > 50%. These grades are, in our view, consistent with the original OA grades proposed by Kellgren and Lawrence [18] but are based on MRI rather than radiographic data. The two musculoskeletal radiologists each viewed the 114 series of MRIs and rated PFC OA severity using the modified K&L scale. Both were blinded to each other's ratings.

Clinical tests of the patellofemoral joint

Each year from baseline to year 4, all subjects in the OAI underwent a clinical examination including the patellar grind test and the patellar crepitus test. For the patellar grind test, the examiner applied a gentle downward force on the patella while the patient was positioned supine with the knee extended. The patient was instructed to gradually tighten the quadriceps muscle by pushing the back of the knee into the examination table. A positive test was recorded when an increase in retropatellar pain was reported. For the patellar crepitus test, the patient was positioned supine on the examination table and the examiner passively flexed and extended the knee through

a 90° arc of motion. A positive test was recorded when "a nearly continuous grinding, crackling or crunching sensation was perceived by the examiner during knee extension or flexion." A negative test was recorded either when no crepitus was felt or if only one or two clicks or pops were felt during the test. Both the patellar crepitus test [9, 19] and the patellar grind test [19] have been reported to be highly reliable when conducted by trained personnel. In the OAI, an extensive training was done for all testers. In addition, a quality assurance assessment was conducted after training and monthly quality assessments were conducted at each site.

Other OAI tests used in the current study

Age, during the visit of interest, race (African American or other), sex, and body mass index, measured in kilograms per square meter, were collected to characterize the sample. A validated modified Charlson comorbidity scale with higher scores indicating higher comorbidity [20] was used to characterize the extent of comorbidity. The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) was used to quantify the extent of pain during activity (WOMAC pain scale) and difficulty during activity (WOMAC function scale) [21]. The WOMAC was collected yearly in OAI for the entire study period. Both scales are highly reliable self-reported measures commonly used on persons with knee arthritis [22].

To specifically focus on PFC pain and function, we also used a single item from both the WOMAC pain and function scales. The WOMAC pain item addressed pain with stair climbing, and the WOMAC function item asked about difficulty descending stairs; both rated as none/mild, moderate, or severe/extreme. These items were used to determine associations between PFC OA severity and pain or activity difficulty potentially attributable to the patellofemoral joint. The psychometric quality of single items from the WOMAC scale is unknown.

Data analysis

Descriptive data for the sample are reported in Table 1. To determine the intertester reliability of modified K&L grades, we used the linear weighted kappa [23] which accounts for chance agreement and provides linear weights to account for more serious disagreements. For example, if one rater indicated the PF joint grade was a 1 and the other rated it a 4, this would be a more serious disagreement as compared to a rating of 3 by one rater and a rating of 4 by the second rater. Weighted kappa values range from 0 to 1 with 0 being equivalent to chance agreement and 1 being equivalent to perfect agreement. A weighted kappa of 0.41 to 0.60 is considered moderate agreement while a weighted kappa of 0.81 or higher is considered almost perfect agreement [8].

Table 1 Sample characteristics ($n = 114$)

Preoperative characteristics	Mean or *N*	SD or %
Age (years), mean (SD)	65.2	(8.1)
Sex (female), *N* (%)	73	(64.0)
Body mass index (kg/m^2), mean (SD)	30.0	(4.7)
Race or ethnic group, *N*, (%)		
White	99	(86.8)
African American	15	(13.2)
Modified Charlson comorbidity, mean (SD)[a]	0.5	(0.8)
WOMAC pain scale, mean (SD)[b]	7.2	(4.3)
WOMAC physical function, mean (SD)[c]	23.2	(13.2)
Tibiofemoral osteoarthritis grade[d]		
Grade 1	3	(2.6)
Grade 2	19	(16.7)
Grade 3	69	(60.5)
Grade 4	23	(20.2)
Examiner no. 1 patellofemoral grade[e]		
Grade 1	4	(3.5)
Grade 2	38	(33.3)
Grade 3	44	(38.6)
Grade 4	28	(24.6)
Examiner no. 2 patellofemoral grade[e]		
Grade 1	12	(10.5)
Grade 2	35	(30.7)
Grade 3	33	(28.9)
Grade 4	34	(29.8)
Positive patellar grind test (yes) *N* (%)	32	(47.1)
Positive patellar crepitus test (yes) (%)	54	(75.0)
WOMAC pain item: pain with stairs (yes) (%)		
None/mild	32	(28.1)
Moderate	34	(29.8)
Severe/extreme	45	(40.5)
WOMAC function item: difficulty with stairs (yes) (%)		
None/mild	36	(32.4)
Moderate	46	(41.4)
Severe/extreme	29	(26.1)

WOMAC Western Ontario and McMaster Universities Osteoarthritis Index, *SD* standard deviation

[a] Modified Charlson comorbidity score range is 0 to 45. Higher scores equate to greater comorbidity burden

[b] WOMAC pain scale score range is 0 to 20. Higher scores equate to more function-limiting pain

[c] WOMAC function scale range is 0 to 68. Higher scores equate to more difficulty with functional activities

[d] Tibiofemoral K&L OA grades were based on radiographic data provided by the OAI study

[e] Patellofemoral compartment OA grades were based on modified K&L MRI-based readings by JAV and KBH

We used the Pearson chi-square test to determine the association between the average K&L grade by the two raters and the patellar grind test, the patellar crepitus test, the WOMAC pain stair climbing item, and the WOMAC function stair climbing item. The WOMAC stair climbing items are each scored on a 5-point scale from none to extreme, but because of relatively few scores on the extremes of the scale (i.e., 0 or 5), scores were collapsed into tertiles (i.e., none/mild, moderate, severe/extreme). Because patellar crepitus and grind tests were only obtained during the first 4 years of the OAI (i.e., allowing for inclusion of KA surgeries conducted only during the first 5 years), sample sizes were somewhat smaller for these analyses ($N = 76$). For the remainder of the analyses, relevant subject data over the entire study period were included ($N = 114$). Because the MRI-based PFC OA grades were skewed toward more severe OA, for the chi-square analyses, we collapsed grades 1 and 2 into one category, grade 3 as one category, and grade 4 as one category.

To determine the extent of contribution of PFC OA severity to the overall self-reported pain during activity (WOMAC pain score) and difficulty during activity (WOMAC function score) experience, we used multiple linear regression. The WOMAC score (pain or function, depending on the model) served as the dependent variable and the predictors were added hierarchically with demographic variables of age, sex, race (African American or other), and BMI (kilograms per m^2) forced into the model first, followed by the tibiofemoral K&L grade, and lastly, the average PFC OA grade. Demographic variables of age, sex, race, and BMI as well as tibiofemoral OA severity were included to adjust for potential confounding given that these variables are associated with knee OA symptom severity, incidence, and progression [24–26]. Preoperative comorbidity scores were not included because 70% of the sample scored a 0.

Analyses allowed us to determine if the average PFC OA grade by the two raters contributed to the pain during activity or difficulty during activity experience after accounting for demographic variables and tibiofemoral OA severity. All two-way interactions were tested, and only significant interactions at $p < 0.05$ were retained. For both models, the F statistic from the model analysis of variance was used to determine if the model was statistically significant at $p < 0.05$ and the r^2 was used to judge the extent of explained variance. No a priori sample size determination was made. A p value of ≤ 0.05 was used to judge statistical significance for all tests.

Data availability All data in the current study are publically available at https://ndar.nih.gov/.

Results

The linear weighted kappa for the two examiners was $\kappa_w = 0.73$ (SE = 0.05) 95% CI = 0.64, 0.82. The percentage agreement was 73.7% and there was no clear systematic bias in the data (see Table 2). Raters disagreed with each other in more severe and less severe PFC OA directions. The Pearson chi-square describing the association between averaged modified K&L grades for the two examiners and the PFC clinical tests was $X^2 = 2.09$ ($p = 0.35$) for the patellar grind test and $X^2 = 0$ ($p = 1.0$) for the patellar crepitus test. The Pearson chi-square describing the association between MRI-based modified K&L grades averaged for the two examiners and the WOMAC pain stair climbing item and the WOMAC function stair descent item was $X^2 = 4.4$ ($p = 0.36$) and $X^2 = 8.9$ ($p = 0.06$) respectively.

The multiple linear regression model for the association between WOMAC pain and PFC OA severity, after accounting for demographic and tibiofemoral OA variables, indicated that only BMI was associated with preoperative WOMAC pain scores (see Table 3). For every 1 point increase in BMI, WOMAC pain scores increase by 0.2 points. Neither PFC nor tibiofemoral OA severity was associated with WOMAC pain after adjustment for the other variables in the model. The WOMAC pain model F test was not significant, indicating that the model was not better than chance at predicting WOMAC pain scores from the variables in the model. The model r^2 was 0.07 indicating that the independent variables explained 7% of the variance in WOMAC pain scores.

Findings for WOMAC function were similar in that only BMI associated with WOMAC function in the multivariable model ($p = 0.01$). The model F test was significant at $p = 0.04$, and the model r^2 was 0.12 indicating that the predictor variables explained 12% of the variance in WOMAC functions (see Table 4). No two-way interactions were found for either model.

Table 2 Distribution of modified Kellgren and Lawrence osteoarthritis grades by the two examiners ($n = 114$)

		Examiner no. 1 modified K&L grade			
Examiner no. 2 modified K&L grade		Grade 1	Grade 2	Grade 3	Grade 4
	Grade 1	4	0	0	0
	Grade 2	7	27	3	1
	Grade 3	1	8	28	7
	Grade 4	0	0	2	26

Table 3 Multiple linear regression model for association between WOMAC pain and knee osteoarthritis severity after accounting for demographic variables

Variable	Unstandardized β	t test	Significance
Sex (male)	− 0.81	− 0.96	0.34
Race (African American)	1.18	0.97	0.34
Body mass index (kg/m^2)	0.19	2.08	0.04
Age (years)	0.02	0.36	0.72
Tibiofemoral OA (grades 1 or 2, 3, 4)	0.72	1.02	0.31
Patellofemoral OA (grades 1 or 2, 3, 4)	0.05	0.09	0.91

Discussion

We found that musculoskeletal radiologists are able to use MRIs to reliably determine the extent of OA present in the PFCs of patients with symptomatic knee OA. Intertester reliability was judged to be substantial, based on the recommendations of Landis and Koch [8], and the weighted kappa reliability estimate was 0.73 (95% CI = 0.64, 0.82). The 95% CI was higher than the point estimate of weighted kappa = 0.49 reported by Kobayashi and colleagues [6]. In our view, musculoskeletal radiologists should consider using this OA grading system to efficiently estimate the extent of PFC OA using MRIs. Importantly, training and consensus discussions were used in our study and should be a standard part of musculoskeletal radiologists' adoption of the modified K&L system of rating MRI-based PFC OA.

Unfortunately, MRI-based PFC OA measures did not associate with the two clinical examination results or the stair climbing self-report scores. Our finding of no association between the crepitus test and PFC OA severity contrasts with that of a large sample study of women with no radiographic knee OA but with MRI-based changes indicating pre-OA degeneration [3]. Associations were found between the PFC crepitus test and MRI-based changes to the PFC, with odds ratios ranging from 2.6 to 5.5. These data suggest that early OA-related changes in the absence of radiographic knee OA may be more strongly associated with symptoms on PFC clinical examination than more severe overall knee OA. A possible reason may be that in patients with no radiographic knee OA but with early MRI-based changes, clinical symptoms associate with early PFC changes. In contrast, in the patients with moderate to severe tibiofemoral knee OA evaluated in our study, the contribution of PFC OA severity to clinical symptoms may be difficult or impossible to discern.

Our findings may have been influenced by the fact that our PFC OA grades were skewed toward more severe OA (K&L grades of 3 and 4). Whether our findings would hold for subjects with mild or moderate PFC OA and a milder spectrum of symptoms cannot be determined from our study. Our findings raise concerns about the clinical importance of MRI measures of PFC OA for inferring the potential clinical relevance of PFC OA particularly among persons with more severe OA. These no-association findings were consistent for stair climbing items and PFC clinical examination findings as well as models that assessed associations between WOMAC pain and WOMAC function scores and PFC OA severity after accounting for relevant demographic variables and tibiofemoral OA severity.

It is possible that, because our sample had moderate to severe tibiofemoral OA, the symptomatic tibiofemoral joints made it difficult to discern whether the PFC contributed to the stair climbing or WOMAC pain or difficulty. A study of persons with only PFC OA would need to be conducted to determine the true extent of pain arising from the PFC during stair climbing. To our knowledge, this study has not been done. Our findings also raise doubts about the clinical relevance of MRI-based PFC OA severity measures in persons with more advanced disease and their potential relevance to either pain with daily activity or difficulty with daily activity including stair climbing, a problem frequently associated with PFC pain [11]. The one caveat is that our multiple regression models for WOMAC function and our stair climbing difficulty analysis found a p value of 0.06 for PFC measures. It is possible that, with a larger sample of subjects and a similar degree of

Table 4 Multiple linear regression models for association between WOMAC function and knee osteoarthritis severity after accounting for demographic variables

Variable	Unstandardized β	t test	Significance
Sex (male)	1.44	0.55	0.58
Race (African American)	2.68	0.71	0.48
Body mass index (kg/m^2)	0.69	2.51	0.01
Age (years)	0.05	0.28	0.78
Tibiofemoral OA (grades 1 or 2, 3, 4)	2.08	0.96	0.34
Patellofemoral OA (grades 1 or 2, 3, 4)	3.07	1.9	0.06

association, PFC OA severity measures may be significantly associated with WOMAC function and stair climbing scores, but this would need to be confirmed in a different and larger sample of patients.

In our view, the current study findings do not substantively differ from prior reports of associations between radiographic PFC OA severity and pain severity. Increased severity of radiographically assessed OA of the PFC was associated with increased WOMAC pain scores such that persons with isolated moderate or severe PFC OA had WOMAC pain scores that were, on average, 1.3 points higher as compared to subjects with no or mild PFC OA [27]. Similarly, Szebenyi and colleagues found that the presence of radiographic knee OA, with or without joint space narrowing, was associated with a 10-point higher (worse) pain score measured on a 0 to 100 scale, when both tibiofemoral and PFC OA were present (i.e., a mean visual analog pain scale (VAS) of 42) relative to only PFC OA (i.e., a mean VAS of 32) [28]. While both Szebenyi and colleagues and Duncan et al. found statistically significant differences in pain scores for persons with mild versus moderate PFC OA or persons with isolated PFC OA versus persons with both tibiofemoral and PFC OA, these differences are not likely to be clinically relevant. This is because differences in WOMAC pain scores on the order of 1 point or differences in pain intensity on the order of 10 points on a 0 to 100 scale are generally considered to be too small to be clinically important [29–31].

Our study had some important limitations. First, our study was cross sectional, and therefore we cannot determine the extent to which MRI-derived measures of PFC OA severity associate with future knee pain or function, an important prognostic consideration in clinical practice. Second, the subjects included in our study eventually underwent knee replacement surgery over the coming year, and the results may not apply to subjects with knee pain and either no OA or a milder OA disease spectrum or subjects who are not considering knee replacement surgery as a viable intervention. While the recruitment of persons with a more severe disease spectrum, particularly for tibiofemoral OA, allowed for variation in PFC severity, overall, this sample generally had moderate to severe knee OA. Third, the tibiofemoral and PFC OA grades were treated as continuous measures in the multiple regression models, and this approach may have influenced the findings given that these measures are not on a continuous scale. With this said, our model diagnostics indicated acceptable homoscedasticity, a normal distribution of residuals, and no multicollinearity. Fourth, our sample size was relatively small and may have been underpowered to detect statistically significant differences for the measures of interest. Finally, MRI utilization for knee imaging is likely to vary among countries and may be more commonly used in the USA as compared to other countries.

In conclusion, MRI-based K&L measures of PFC OA provide reliable measures among musculoskeletal radiologists. These measures, therefore, have potential for informing clinicians about the severity of PFC OA in a simple and commonly used way via a modified K&L score, much like that in common use for radiographic assessment of the tibiofemoral joint. PFC measures do not appear to be meaningfully associated with clinical tests commonly used to assess the PFC or self-reported knee pain and functional status in persons with mostly moderate to severe PFC and tibiofemoral OA. However, the extent of PFC OA may be related to clinical tests of PFC involvement and activities associated with PFC pain in patients with either no tibiofemoral OA or mild tibiofemoral OA. Clinicians should recognize that measurements of PFC OA have substantial limitations in persons with more advanced knee OA disease in that they do not appear to associate in a clinically meaningful way with either concurrently obtained PFC clinical examination procedures or self-reported stair climbing, knee pain with activity or knee function.

Role of the funding source The OAI is a public–private partnership comprised of five contracts funded by the National Institutes of Health. Private funding partners include Merck Research Laboratories; Novartis Pharmaceuticals Corporation, GlaxoSmithKline; and Pfizer, Inc. The authors of the current paper are not part of the OAI investigative team. The funding source played no role in the conduct or reporting of this study.

Funding information The OAI is a public-private partnership comprised of five contracts (N01-AR-2-2258; N01-AR-2-2259; N01-AR-2-2260; N01-AR-2-2261; N01-AR-2-2262) funded by the National Institutes of Health, a branch of the Department of Health and Human Services, and conducted by the OAI Study Investigators. Private funding partners include Merck Research Laboratories; Novartis Pharmaceuticals Corporation, GlaxoSmithKline; and Pfizer, Inc. Private sector funding for the OAI is managed by the Foundation for the National Institutes of Health. This manuscript was prepared using an OAI public use data set and does not necessarily reflect the opinions or views of the OAI investigators, the NIH, or the private funding partners.

Compliance with ethical standards

The study was approved by the Institutional Review Boards at the following sites: (1) University of Pittsburgh in Pittsburgh, Pennsylvania; (2) University of Maryland in Baltimore, Maryland; (3) Ohio State University in Columbus, Ohio; and (4) Memorial Hospital of Rhode Island, in Pawtucket, Rhode Island. All subjects signed an IRB-approved consent form prior to participation.

Disclosures None.

References

1. Felson DT, Naimark A, Anderson J, Kazis L, Castelli W, Meenan RF (1987) The prevalence of knee osteoarthritis in the elderly. The Framingham osteoarthritis study. Arthritis Rheum 30:914–918. https://doi.org/10.1002/art.1780300811

2. McAlindon TE, Snow S, Cooper C, Dieppe PA (1992) Radiographic patterns of osteoarthritis of the knee joint in the community: the importance of the patellofemoral joint. Ann Rheum Dis 51:844–849. https://doi.org/10.1136/ard.51.7.844
3. Schiphof D, Van Middelkoop M, De Klerk BM et al (2014) Crepitus is a first indication of patellofemoral osteoarthritis (and not of tibiofemoral osteoarthritis). Osteoarthr Cartil 22:631–638. https://doi.org/10.1016/j.joca.2014.02.008
4. Lester G (2012) The Osteoarthritis Initiative: a NIH public-private partnership. HSS J 8:62–63. https://doi.org/10.1007/s11420-011-9235-y
5. Riddle DL, Jiranek WA, Hayes CW (2014) Use of a validated algorithm to judge the appropriateness of total knee arthroplasty in the United States: a multicenter longitudinal cohort study. Arthritis Rheumatol 66:2134–2143. https://doi.org/10.1002/art.38685
6. Kobayashi S, Peduto A, Simic M, Fransen M, Refshauge K, Mah J, Pappas E (2018) Can we have an overall osteoarthritis severity score for the patellofemoral joint using magnetic resonance imaging? Reliability and validity. Clin Rheumatol 37:1091–1098. https://doi.org/10.1007/s10067-017-3888-y
7. Hunter DJ, Guermazi A, Lo GH, Grainger AJ, Conaghan PG, Boudreau RM, Roemer FW (2011) Evolution of semi-quantitative whole joint assessment of knee OA: MOAKS (MRI Osteoarthritis Knee Score). Osteoarthr Cartil 19:990–1002. https://doi.org/10.1016/j.joca.2011.05.004
8. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33:159–174
9. Maricar N, Callaghan MJ, Parkes MJ, Felson DT, O'Neill TW (2016) Interobserver and intraobserver reliability of clinical assessments in knee osteoarthritis. J Rheumatol 43:2171–2178. https://doi.org/10.3899/jrheum.150835
10. Cibere J, Thorne A, Bellamy N, Greidanus N, Chalmers A, Mahomed N, Shojania K, Kopec J, Esdaile JM (2004) Reliability of the knee examination in osteoarthritis. Arthritis Rheum 50:458–468. https://doi.org/10.1002/art.23310
11. Décary S, Frémont P, Pelletier B, Fallaha M, Belzile S, Martel-Pelletier J, Pelletier JP, Feldman D, Sylvestre MP, Vendittoli PA, Desmeules F (2018) Validity of combining history elements and physical examination tests to diagnose patellofemoral pain. Arch Phys Med Rehabil 99:607–614. https://doi.org/10.1016/j.apmr.2017.10.014
12. Briani RV, Pazzinatto MF, Waiteman MC, de Oliveira Silva D, de Azevedo FM (2018) Association between increase in vertical ground reaction force loading rate and pain level in women with patellofemoral pain after a patellofemoral joint loading protocol. Knee 25:398–405. https://doi.org/10.1016/j.knee.2018.03.009
13. Escobar A, Quintana JM, Arostegui I et al (2003) Development of explicit criteria for total knee replacement. Int J Technol Assess Health Care 19:57–70
14. Kothari M, Guermazi A, von IG et al (2004) Fixed-flexion radiography of the knee provides reproducible joint space width measurements in osteoarthritis. EurRadiol 14:1568–1573
15. Paradowski PT, Lohmander LS, Englund M (2016) Osteoarthritis of the knee after meniscal resection: long term radiographic evaluation of disease progression. Osteoarthr Cartil 24:794–800. https://doi.org/10.1016/j.joca.2015.12.002
16. Parsons C, Fuggle NR, Edwards MH et al (2018) Concordance between clinical and radiographic evaluations of knee osteoarthritis. Aging Clin Exp Res 30:17–25. https://doi.org/10.1007/s40520-017-0847-z
17. Heng HY, Bin Abd Razak HR, Mitra AK (2015) Radiographic grading of the patellofemoral joint is more accurate in skyline compared to lateral views. Ann Transl Med 3:263
18. Kellgren JH, Lawrence JS (1957) Radiological assessment of osteo-arthrosis. Ann Rheum Dis 16:494–502
19. Cibere J, Bellamy N, Thorne A, Esdaile JM, McGorm KJ, Chalmers A, Huang S, Peloso P, Shojania K, Singer J, Wong H, Kopec J (2004) Reliability of the knee examination in osteoarthritis: effect of standardization. Arthritis Rheum 50:458–468
20. Katz JN, Chang LC, Sangha O et al (1996) Can comorbidity be measured by questionnaire rather than medical record review? Med Care 34:73–84
21. Bellamy N (2005) The WOMAC knee and hip osteoarthritis indices: development, validation, globalization and influence on the development of the AUSCAN hand osteoarthritis indices. Clin Exp Rheumatol 23:S148–S153
22. McConnell S, Kolopack P, Davis AM (2001) The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC): a review of its utility and measurement properties. Arthritis Rheum 45:453–461
23. Sim J, Wright CC (2005) The kappa statistic in reliability studies: use, interpretation, and sample size requirements. PhysTher 85:257–268
24. Chapple CM, Nicholson H, Baxter GD, Abbott JH (2011) Patient characteristics that predict progression of knee osteoarthritis: a systematic review of prognostic studies. Arthritis Care Res (Hoboken) 63:1115–1125
25. Neogi T (2013) The epidemiology and impact of pain in osteoarthritis. Osteoarthr Cartil 21:1145–1153
26. Neogi T, Felson D, Niu J, Nevitt M, Lewis CE, Aliabadi P, Sack B, Torner J, Bradley L, Zhang Y (2009) Association between radiographic features of knee osteoarthritis and pain: results from two cohort studies. BMJ 339:b2844
27. Duncan R, Peat G, Thomas E, Wood L, Hay E, Croft P (2008) How do pain and function vary with compartmental distribution and severity of radiographic knee osteoarthritis? Rheumatology 47:1704–1707. https://doi.org/10.1093/rheumatology/ken339
28. Szebenyi B, Hollander AP, Dieppe P, Quilty B, Duddy J, Clarke S, Kirwan JR (2006) Associations between pain, function, and radiographic features in osteoarthritis of the knee. Arthritis Rheum 54:230–235. https://doi.org/10.1002/art.21534
29. Tubach F, Ravaud P, Martin-Mola E, Awada H, Bellamy N, Bombardier C, Felson DT, Hajjaj-Hassouni N, Hochberg M, Logeart I, Matucci-Cerinic M, van de Laar M, van der Heijde D, Dougados M (2012) Minimum clinically important improvement and patient acceptable symptom state in pain and function in rheumatoid arthritis, ankylosing spondylitis, chronic back pain, hand osteoarthritis, and hip and knee osteoarthritis: results from a prospective multina. Arthritis Care Res (Hoboken) 64:1699–1707
30. Tubach F, Ravaud P, Baron G et al (2005) Evaluation of clinically relevant changes in patient reported outcomes in knee and hip osteoarthritis: the minimal clinically important improvement. Ann Rheum Dis 64:29–33
31. Danoff JR, Goel R, Sutton R, Maltenfort MG, Austin MS (2018) How much pain is significant? Defining the minimal clinically important difference for the visual analog scale for pain after total joint arthroplasty. J Arthroplast 33:S71–S75. https://doi.org/10.1016/j.arth.2018.02.029