

Can we have an overall osteoarthritis severity score for the patellofemoral joint using magnetic resonance imaging? Reliability and validity

Sarah Kobayashi¹ · Anthony Peduto² · Milena Simic¹ · Marlene Fransen¹ · Kathryn Refshauge¹ · Jean Mah³ · Evangelos Pappas¹

Received: 9 October 2017 / Accepted: 23 October 2017 / Published online: 6 November 2017
© International League of Associations for Rheumatology (ILAR) 2017

Abstract This work aimed to assess inter-rater reliability and agreement of a magnetic resonance imaging (MRI)-based Kellgren and Lawrence (K&L) grading for patellofemoral joint osteoarthritis (OA) and to validate it against the MRI Osteoarthritis Knee Score (MOAKS). MRI scans from people aged 45 to 75 years with chronic knee pain participating in a randomised clinical trial evaluating dietary supplements were utilised. Fifty participants were randomly selected and scored using the MRI-based K&L grading using axial and sagittal MRI scans. Raters conducted inter-rater reliability, blinded to clinical information, radiology reports and other rater results. Intra- and inter-rater reliability and agreement were evaluated using the intra-class correlation coefficient (ICC) and Cohen's weighted kappa. There was a 2-week interval between the first and second readings for intra-rater reliability. Validity was assessed using the MOAKS and evaluated using Spearman's correlation coefficient. Intra-rater reliability of the K&L system was excellent: ICC 0.91 (95% CI 0.82–0.95); weighted kappa ($\kappa = 0.69$). Inter-rater reliability was high (ICC 0.88; 95% CI 0.79–0.93), while agreement between raters was moderate ($\kappa = 0.49$ –0.57). Validity analysis

demonstrated a strong correlation between the total MOAKS features score and the K&L grading system ($\rho = 0.62$ –0.67) but weak correlations when compared with individual MOAKS features ($\rho = 0.19$ –0.61). The high reliability and good agreement show consistency in grading the severity of patellofemoral OA with the MRI-based K&L score. Our validity results suggest that the scale may be useful, particularly in the clinical environment. Future research should validate this method against clinical findings.

Keywords Bone · Cartilage · Diagnostic imaging · Knee · Magnetic resonance imaging · Osteoarthritis

Osteoarthritis (OA) is a chronic musculoskeletal condition frequently resulting in varying degrees of pain, stiffness and activity limitations [1]. The weight-bearing joint most commonly affected by OA is the knee, with approximately 24% of adults (aged 15 to 99 years) with knee OA [2]. Research on knee OA has predominantly focused on the tibiofemoral joint or considered the knee as one joint [3], despite studies showing that OA in the patellofemoral compartment is just as, if not more, prevalent than in the medial or lateral tibiofemoral compartments [4–8]. In fact, a systematic review of population-based studies that recruited at least 215 people through “random sampling or convenience sampling from the community” demonstrates around 25% of people aged 20–99 years have patellofemoral OA [9]. Furthermore, a study reported that of those with radiographic knee OA and knee pain, 31% had isolated patellofemoral OA, 24% had isolated medial tibiofemoral OA and 20% had combined patellofemoral and tibiofemoral OA [4]. These findings emphasise the need for a greater focus on the patellofemoral joint in OA research.

More recently, magnetic resonance imaging (MRI) has been used in OA research to comprehensively quantify

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10067-017-3888-y>) contains supplementary material, which is available to authorized users.

✉ Sarah Kobayashi
skob9241@uni.sydney.edu.au

¹ Faculty of Health Sciences, Discipline of Physiotherapy, University of Sydney, Sydney, Australia

² Department of Radiology Westmead Hospital and Western Clinical School, Discipline of Medical Imaging, Sydney Medical School, University of Sydney, Sydney, Australia

³ Department of Imaging, Westmead Hospital, Westmead, Australia

structural joint changes that may be difficult to visualise in radiographic imaging. Although MRI is expensive and not advised for OA diagnosis [10], it allows better visualisation of joint structural pathologies than radiography [11]. Bone marrow lesions, subchondral cysts, sclerosis, synovitis and effusion are a few of the features that are believed to be contributors to, or associated with, the pathophysiology and symptoms of knee OA [12–15]. Scoring systems, such as the Boston-Leeds Osteoarthritis Knee Score, Whole Organ MRI Score and MRI Osteoarthritis Knee Score (MOAKS), have been established to standardise the grading and reporting of OA features visible on MRI scans in all compartments of the knee [12, 16, 17]. In particular, the MOAKS is a comprehensive assessment tool for knee OA MRI evaluation. It was developed from the strengths and weaknesses of previous MRI-based assessment tools for OA evaluation [12, 16, 18]; it is reliable [12], and it has been widely used in research [11, 15]. However, the MOAKS is complex and requires specialised training, therefore limiting its practical use in the research environment.

As a potential alternative to these more complex MRI-based OA assessment tools, this paper aims to evaluate an MRI-based Kellgren & Lawrence (K&L) grading tool for the patellofemoral joint. The K&L assessment tool is widely used for radiographic OA assessment. It evaluates the presence of osteophytes, joint space narrowing (JSN), sclerosis and bony deformity on radiograph [19]. Although Kellgren and Lawrence [19] only describes its use for the tibiofemoral joint, it has been previously used for the assessment of the patellofemoral joint in lateral and skyline/axial radiographic views [6, 20–23]. Riddle et al. [24] developed an MRI-based K&L grading for the patellofemoral joint within a cohort study, evaluating the appropriateness of joint replacement surgery, as they did not have any axial radiographic views [24]. The MRI-based K&L grading assesses the patellofemoral joint mostly using two OA features: osteophytes and cartilage loss. The intra-rater reliability appeared excellent [24]; however, inter-rater reliability and agreement were not evaluated. The primary aim of this study was to assess inter-rater reliability and agreement of an MRI-based K&L grading of the patellofemoral joint. The secondary aim was to validate the MRI-based K&L grading by comparing it with the reliable and validated MOAKS.

Methods

Overview

The present study is a secondary analysis of MRI scans undertaken at baseline during a double-blind randomised placebo-controlled trial, which was conducted over a 2-year period (2007–2009). The trial investigated the effect of glucosamine sulphate, chondroitin sulphate or the combination of

both on disease progression in people aged 45–75 years with chronic knee pain [25].

Participants

Participants were recruited through general media advertising and general practices in New South Wales, Australia. Eligible participants aged 45 years and over had joint space narrowing in the medial tibiofemoral joint of a symptomatic knee. Exclusion criteria included rheumatoid arthritis or other inflammatory joint diseases, lower limb surgery within the last 6 months, bilateral knee replacements or plans for knee replacements during the study period. Participants gave informed written consent. The study is in accordance with the Declaration of Helsinki and it was approved by the local human research ethics committee.

Baseline demographics

Age, height, weight, analgesia and knee pain duration (years) were collected. Participants completed the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC); pain (0–20) and physical function (0–68) subscale scores were extrapolated, with higher scores representing greater knee pain and activity limitations.

MRI technique

The participants' knees were imaged using a dedicated knee coil in a 3-T magnet (GE Signa HDx). Each examination consisted of axial proton density-weighted turbo spin echo images (with repetition time (TR) of 3900; echo time (TE) of 40; echo train length (ETL) 8; 3 mm slice thickness; 0.3 mm intersection gap; 13 cm field of view (FOV); 384 × 320 matrix), sagittal proton density-weighted fat suppressed turbo spin echo images (TR 3400; TE 40; ETL 7; 3 mm slice thickness; 0.3 mm gap; 14 cm FOV; 384 × 320 matrix) and sagittal T2-weighted turbo spin echo images (TR 1060; TE 6.5; ETL 1; 3 mm slice thickness; no gap; 16 cm FOV; 320 × 224 matrix). Total acquisition time (including the initial survey sequence) was 30 min.

MRI-based K&L grading

The MRI-based K&L grading is a surrogate for the radiographic K&L scale. The MRI-based K&L grading ranges from 0 to 4:

- 0: Normal
- 1: No definite osteophytes or joint space narrowing, but there may be minimal cartilage, bone or periarticular changes

- 2: Definite osteophyte with focal cartilage loss but no extensive cartilage involvement/no joint space narrowing
- 3: Osteophytes with significant cartilage loss at either the medial or lateral patellar and/or trochlear surfaces
- 4: Osteophytes with complete cartilage loss involving more than 50% of the medial and/or lateral patellofemoral joint

The axial and sagittal views of the patellofemoral joint were assessed.

MRI osteoarthritis knee score

All participant MRI scans were also graded using the MOAKS, a reliable and validated scoring system that was developed for the specific use of knee OA assessment using MRI scans [12]. There are 12 subscales in the MOAKS, including size of osteophyte, percentage of any cartilage loss (partial and full-thickness loss), percentage of full-thickness cartilage loss, volume of bone marrow lesions and effusion-synovitis. Raters used this assessment tool to evaluate individual OA features from normal (0) to severe [3]. In order to validate the MRI-based K&L grading for the patellofemoral joint, cartilage and osteophyte MOAKS scores in the patellofemoral joint (medial patellar, lateral patellar, medial trochlear and lateral trochlear regions) were compared with the scores attained from the MRI-based K&L grading. Other subscales of the MOAKS, such as bone marrow lesions, cysts and synovitis-effusion, were excluded for this study as they were features that were not assessed in the MRI-based K&L grading. The cartilage and osteophyte MOAKS scores were then averaged in order to attain an overall patellofemoral joint score and compared with the MRI-based K&L grading scores.

Procedure

A senior radiologist (A.P), who was experienced with using the MRI-based K&L grading, conducted the training, which consisted of evaluating images from 10 randomly selected participants with each rater according to the MRI-based K&L grading. The raters were then given another set of MRI scans from 20 randomly selected participants to assess independently. Disagreements and inconsistencies among the readers were addressed and discussed until a consensus was reached. An atlas was created from the 30 sets of MRI scans for training to visually demonstrate each grade (Appendix).

Of the 304 participants with available knee MRI scans, 50 MRI scans were randomly selected. The 30 MRI scans that were used for training were excluded from the selection. The images were graded by three raters: the primary investigator (S.K) and two radiologists (A.P, J.M). All raters were blinded to clinical information and radiologic reports. The primary investigator was a health researcher and a novice reader with

no formal radiology training. RadiAnt DICOM viewer [26] was used to view the MRI scans for assessment. Sagittal and axial MR images were used together to attain a total view of the patellofemoral joint. When the raters gave discordant grades to the views of the same participant, the worst grade was used. There was a 2-week interval between the first and second readings for the assessment of intra-rater reliability. The primary investigator also assessed the MRI scans using the MOAKS approximately 1 year after the MRI-based K&L grading assessments.

Statistical analysis

Intra-class correlation coefficient (ICC) was used to assess intra- and inter-rater reliability, using SPSS (SPSS Inc., Chicago, IL). ICC assesses intra- and inter-rater reliability by measuring the variance of scores between the raters [27]. Model 3 was used to calculate the ICC as the three raters assessed each participant, and the raters were fixed. An average measure between the scores of the three raters was taken to determine the ICC for the test. Reliability is considered to be poor when the ICC < 0.40, fair when the ICC is 0.41–0.59, good when the ICC 0.60–0.74 and excellent when the ICC is greater than 0.75 [28].

Intra-rater (S.K) and inter-rater agreement were measured using Cohen’s weighted kappa in Excel. A predefined table of weights (Table 1) was used to measure the degree of disagreement between the two raters (linear weighted kappa). The observed frequencies of scores were tabulated into a 2 × 2 contingency table. A resource package provided by real-statistics.com [29] was then used to calculate the weighted kappa, standard error and 95% confidence intervals (CI), using the two constructed tables. A weighted kappa of less than 0.20 indicates poor agreement, 0.21–0.40 indicates fair agreement, 0.41–0.60 indicates moderate agreement, 0.61–0.80 indicates good agreement and a weighted kappa greater than 0.80 is interpreted as very good agreement [30]. Spearman’s correlation coefficient for ordinal scales was used to statistically compare the MRI-based K&L grading and the MOAKS for validity. A strength of a correlation is considered to be small when $\rho = 0.10$, medium when $\rho = 0.30$ and large when $\rho = 0.50$ [31].

Table 1 Predefined table of weights to calculate Cohen’s weighted kappa

	0	1	2	3	4
0	0	1	2	3	4
1	1	0	1	2	3
2	2	1	0	1	2
3	3	2	1	0	1
4	4	3	2	1	0

Results

Study sample

The mean age for the sample (26 females) was 61.1 years (SD 8.4), and the mean BMI was 27.4 kg/m² (SD 4.3). On average, participants had mild symptoms at baseline, with mean WOMAC pain score of 6.0 (SD 3.9) and physical function score of 15.6 (SD 12.1) (Table 2).

Intra-rater reliability

Two-week intra-rater reliability was 0.91 (95%CI: 0.82–0.95), indicating excellent reliability. Intra-rater agreement was good ($\kappa = 0.69$).

Inter-rater reliability

Inter-rater reliability among all three raters was excellent (ICC = 0.88; Table 3). When pairs of readers were compared, the ICC remained above 0.75 for each pair of raters (Table 3). However, Cohen's weighted kappa showed that inter-rater agreement among raters was moderate, ranging between 0.49 and 0.57 (Table 3).

Validity

Our results reveal that the MRI-based K&L grading was correlated with MOAKS osteophytes in the superior patellar, inferior patellar, medial trochlear and lateral trochlear regions (Table 4). The percentage of cartilage loss (both partial and full thickness) and the percentage of full-thickness cartilage loss in all regions of the patellofemoral joint were also correlated with MRI-based K&L grading scores (Table 4). Although these correlations were statistically significant, the coefficients were not large and the strength of the correlations were medium ($\rho = 0.37$ – 0.58) (Table 4).

When the average of the MOAKS scores were calculated and compared with the MRI-based K&L grading scores, the correlation coefficients were strong and remained statistically significant (Table 4). The strongest correlation occurred when the average percentage of full-thickness cartilage loss in all of the patellofemoral joint regions was calculated ($\rho = 0.65$, $p < 0.001$) (Table 4). The correlation coefficient decreased when osteophytes were included in the analysis ($\rho = 0.55$, $p < 0.01$) (Table 4).

Discussion

We evaluated intra- and inter-rater reliability and agreement of a recently described MRI-based K&L grading [24]. The

Table 2 Demographics ($n = 50$): mean (standard deviation)

	Females ($n = 26$)	Males ($n = 24$)	Total sample ($n = 50$)
Age (years)	60.7 (8.6)	61.5 (8.3)	61.1 (8.4)
Height (cm)	164.8 (6.5)	177.7 (7.5)	171.0 (9.5)
Weight (kg)	74.4 (12.0)	86.9 (17.9)	80.4 (16.2)
Mean BMI (kg/m ²)	27.4 (4.1)	27.4 (4.6)	27.4 (4.3)
Analgesia (%)	19.2%	8.3%	14.0%
WOMAC			
Pain (0–20)	6.2 (3.5)	5.7 (4.3)	6.0 (3.9)
Physical function (0–68)	15.6 (11.8)	15.6 (12.7)	15.6 (12.1)
Duration of symptoms			
< 3 years	31%	16.7%	24.0%
3–5 years	19%	50.0%	34.0%
6–10 years	35%	20.8%	28.0%
> 10 years	15.4%	8.3%	12.0%
Not applicable	0.0%	4.2%	2.0%
Patellofemoral joint grade (graded by primary investigator (SK))			
0	0.0%	0.0%	0.0%
1	7.7%	29.2%	18.0%
2	38.5%	50.0%	44.0%
3	34.6%	12.5%	24.0%
4	19.2%	8.3%	14.0%

Table 3 Agreement for MRI-based K&L grading between three raters using weighted kappa and ICC, as well as standard errors and 95% CIs

Raters	Weighted kappa	Standard error	95% CI	ICC	95% CI
Between all three raters	–	–	–	0.88	0.79–0.93
Between SK (inexperienced) and AP (experienced)	0.57	0.08	0.42–0.72	0.84	0.62–0.92
Between AP (experienced) and JM (experienced)	0.49	0.09	0.32–0.67	0.79	0.63–0.88
Between SK (inexperienced) and JM (experienced)	0.54	0.07	0.39–0.69	0.84	0.64–0.92

grading from MRI scans demonstrated good intra-rater reliability and agreement even when performed by an inexperienced reader. We found that while inter-rater agreement was

Table 4 Spearman’s correlation coefficients between MRI-based K&L grading for the patellofemoral joint scores and MOAKS features (individual and combined)

MOAKS feature	MOAKS region	Correlation coefficient	Sig (two-tailed)
Osteophytes	Superior patella	0.369	0.008
	Inferior patella	0.573	< 0.001
	Medial patella	0.158	0.274
	Lateral patella	0.190	0.185
	Medial trochlea	0.453	0.001
	Lateral trochlea	0.498	< 0.001
	<i>Average of osteophytes scores</i>	0.554	< 0.001
Percentage of any cartilage loss (partial and full thickness)	Medial patella	0.582	< 0.001
	Lateral patella	0.563	< 0.001
	Medial trochlea	0.563	< 0.001
	Lateral trochlea	0.470	< 0.001
	<i>Average of any cartilage loss scores</i>	0.629	< 0.001
Percentage of full-thickness cartilage loss	Medial patella	0.611	< 0.001
	Lateral patella	0.559	< 0.001
	Medial trochlea	0.462	0.001
	Lateral trochlea	0.393	0.005
	<i>Average of full thickness cartilage loss scores</i>	0.649	< 0.001
Total MOAKS score (osteophytes + %cartilage loss + %full-thickness loss)	<i>Average of total score</i>	0.621	< 0.001

moderate, there was excellent inter-rater reliability between the three raters. We also demonstrated validity of the scale with a strong correlation between the total MOAKS scores and the MRI-based K&L score. These findings indicate that the MRI-based K&L grading for the patellofemoral joint examined in this study could be a useful tool for researchers and clinicians to assess and monitor patellofemoral OA.

The MRI-based K&L grading was originally developed so that the patellofemoral joint could be assessed using MRI, in conjunction with the radiographic K&L assessment of the tibiofemoral joint [24]. Unlike the original study, where an experienced radiologist was employed, a novice reader performed the intra-rater reliability assessments in the present study. Despite having less experience with reading MRIs, we obtained good agreement (weighted $\kappa = 0.69$). However, our intra-rater reliability was not as high as reported in the original study when an experienced radiologist assessed the MRI scans (weighted $\kappa = 0.80$) [24]. This finding suggests that when a novice rater receives extensive training as in our study, acceptable consistency in the grading of patellofemoral OA severity by one rater can be achieved.

In contrast to the results from the intra-rater reliability and agreement, the inter-rater reliability and agreement results appeared very different from each other, which may be due to the conceptual differences between agreement and reliability. Reliability (using ICC) is an assessment of the variability of the selected study objects (i.e. participant MRI scans) [32]. Agreement (using weighted kappa) assesses how much the raters (measurement error) agree on the same measures [32]. The good ICC and moderate weighted kappa results reveal that, although the raters are able to differentiate between the different severities consistently, the raters are not always agreeing. Perhaps more training is required for more novice readers so that they can identify more subtle features that could result in more severe OA grades, and therefore, be more consistent with expert readers. However, only moderate agreement was also seen between the expert readers (Table 3), suggesting that disagreement between the raters could be reflective of the limitations of the MRI-based K&L grading, rather than the raters’ experience. In order to overcome the limitations of the MRI-based K&L grading, perhaps the assessment tool needs to be more refined to improve the agreement

between raters. For now, any disagreements need to be discussed further to reinforce the distinguishing features of each grade. Furthermore, in clinics, it would be preferable if the same reader evaluates the MRI scans, when following their patients' structural disease progression.

The validity findings suggest that the MRI-based K&L grading is not a good alternative to the MOAKS if the assessor is evaluating individual structural pathologies of OA in the patellofemoral joint. However, when the MOAKS scores were combined, the correlation coefficient between the total MOAKS and the MRI-based K&L grading was stronger. Since the MRI-based K&L grading evaluates osteophytes and cartilage in combination to yield an overall patellofemoral joint score, it would be appropriate to compare it with a similar outcome (that is, combining MOAKS osteophytes and cartilage loss subscale scores). However, from the authors' knowledge, the MOAKS scores have not been previously combined, and the clinical significance of a total MOAKS score is unknown. Furthermore, a total patellofemoral joint score may not be ideal for MRI-based OA assessment as severe OA changes in some regions may outweigh milder changes in other regions. That is, a total patellofemoral joint score may be reported as more severe than specific areas of the joint (e.g. medial and lateral patellar surfaces). Yet, changes in a small region of the joint may not be clinically meaningful. With the large number of subregions and subscales in the MOAKS and other MRI-based OA assessment tools, it may be difficult to yield a sound conclusion for the patient that could be clinically meaningful. It is up to the clinician or researcher to decide whether they present to their patients an overall patellofemoral joint score or all the findings in each region and each subscale. The latter could be potentially detrimental to the patient, as medical terms that insinuate presence of disease could result in less understanding and more fear [33], and perhaps poorer prognosis. A previous study has shown that "degenerative terms" on radiological reports lead to poorer perceived prognosis among people with low back pain [34]. Reports using complex MRI-based OA assessment tools may lead to the patient catastrophizing their OA, provoking fear and potentially leading to poorer prognosis. The MRI-based K&L grading provides a simplified score, which may be clinically useful and more meaningful to patients.

The strengths of this study include a training protocol for all raters in the study, the development of an atlas and detailed statistical consideration. The most experienced radiologist trained the other raters, ensuring consistency, as the three raters had various degrees of training and experience with assessing MRI scans. All raters were blinded to any clinical data and radiological reports during the assessments. The raters had no prior knowledge of the participants'

patellofemoral and/or tibiofemoral joint status, eliminating bias from the assessments. Furthermore, participants' MRIs were randomly selected. For the statistical analyses, we employed two statistical tests to ensure evaluation of both reliability and agreement. Additionally, we assessed the validity of the MRI-based K&L grading by comparing it to a reliable MRI-based assessment tool.

Study limitations

Limitations of this study include the exclusion of the tibiofemoral joint and limiting evaluation to the patellofemoral joint of the knee, as well as the inclusion of only participants with knee pain. The inclusion of tibiofemoral joint assessment and a "normal" participant cohort (no knee pain) would strengthen the usefulness of the MRI-based K&L grading for knee OA assessment using MRI scans. Future studies should consider assessing the MRI-based K&L grading for the tibiofemoral joint to allow a whole joint assessment to be conducted. Furthermore, future studies should also consider comparing the MRI-based K&L grading with radiographic K&L evaluation of the patellofemoral joint to determine comparability. Since MRI could be more sensitive than radiographic evaluations, raters may identify more people with mild knee OA using MRI, than with radiography. It would be also valuable to evaluate the MRI-based K&L grading in the clinical setting and to evaluate patients' responses to standardised terminology or definitions of OA as seen on radiological reports. It would be interesting to see if the terminology used in the MRI-based K&L grading and other MRI-based OA assessment tools elicit fear and potentially lead to perceived poorer prognosis.

Conclusions

The radiographic K&L is a simple and well-recognised assessment, evaluating OA structural disease severity. Therefore, this MRI-based K&L grading can be potentially adopted by those working in radiography and are less familiar with MRI, when radiographs are not available for assessment. Our results demonstrate that researchers and clinicians with different levels of experience can use the grading assessment to assess OA. As it only assesses two MRI features of OA, it is simple and easy to follow and understand. Furthermore, this grading assessment is another option for the assessment of the patellofemoral joint [35]. Despite disease in the patellofemoral joint (combined with tibiofemoral OA) contributing to more pain and functional limitations, compared with isolated tibiofemoral OA [6, 36, 37], fewer MRI studies have been

conducted on the patellofemoral joint, than either the medial or lateral tibiofemoral compartments [11]. Therefore, the MRI-based K&L grading is an important contribution as it provides a less time-consuming score that could be utilised to monitor OA progression in larger cohorts.

Funding information The MRI scans that were used for this manuscript came from the Longitudinal Evaluation of Glucosamine Sulphate (LEGS) trial, which was funded by the National Health and Medical Research Council of Australia (ID 402511), the Department of Health and Ageing (Australia) (ID 402781) and by some supplementary funding (< 15% total) from Sanofi-Aventis Consumer Healthcare Pty Ltd., Australia. The funders had no role in the design of the study; collation, analysis or interpretation of the data or preparation, review or approval of the manuscript.

Compliance with ethical standards The study is in accordance with the Declaration of Helsinki and it was approved by the local human research ethics committee.

Disclosures None.

References

- Litwic A, Edwards MH, Dennison EM, Cooper C (2013) Epidemiology and burden of osteoarthritis. *Br Med Bull* 105(1): 185–199
- Pereira D, Peleteiro B, Araújo J, Branco J, Santos RA, Ramos E (2011) The effect of osteoarthritis definition on prevalence and incidence estimates: a systematic review. *Osteoarthr Cartil* 19(11): 1270–1285
- Crossley KM, Hinman RS (2011) The patellofemoral joint: the forgotten joint in knee osteoarthritis. *Osteoarthr Cartil* 19(7):765–767
- Cooper C, McAlindon T, Snow S, Vines K, Young P, Kirwan J et al (1994) Mechanical and constitutional risk factors for symptomatic knee osteoarthritis—differences between medial tibiofemoral and patellofemoral disease. *J Rheumatol* 21(2):307–313
- Duncan RC, Hay EM, Saklatvala J, Croft PR (2006) Prevalence of radiographic osteoarthritis—it all depends on your point of view. *Rheumatology* 45(6):757–760
- 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(11):1704–1707
- Kumm J, Tamm A, Lintrop M, Tamm A (2012) The prevalence and progression of radiographic knee osteoarthritis over 6 years in a population-based cohort of middle-aged subjects. *Rheumatol Int* 32(11):3545–3550
- Schiphof D, Van Middelkoop M, De Klerk BM, Oei EHG, Hofman A, Koes BW et al (2014) Crepitus is a first indication of patellofemoral osteoarthritis (and not of tibiofemoral osteoarthritis). *Osteoarthr Cartil* 22(5):631–638
- Kobayashi S, Pappas E, Fransen M, Refshauge K, Simic M (2016) The prevalence of patellofemoral osteoarthritis: a systematic review and meta-analysis. *Osteoarthr Cartil* 24(10):1697–1707
- Zhang W, Doherty M, Peat G, Bierma-Zeinstra SMA, Arden NK, Bresnihan B et al (2010) EULAR evidence-based recommendations for the diagnosis of knee osteoarthritis. *Ann Rheum Dis* 69(3):483–489
- Hunter DJ, Zhang W, Conaghan PG, Hirko K, Menashe L, Reichmann WM (2011) Responsiveness and reliability of MRI in knee osteoarthritis: a meta-analysis of published evidence. *Osteoarthr Cartil* 19
- Hunter DJ, Guermazi A, Lo GH, Grainger AJ, Conaghan PG, Boudreau RM et al (2011) Evolution of semi-quantitative whole joint assessment of knee OA: MOAKS (MRI osteoarthritis knee score). *Osteoarthr Cartil* 19(8):990–1002
- Barr AJ, Campbell TM, Hopkinson D, Kingsbury SR, Bowes MA, Conaghan PG A systematic review of the relationship between subchondral bone features, pain and structural pathology in peripheral joint osteoarthritis. *Arthritis Res Therapy* 17:228
- Bastick AN, Runhaar J, Belo JN, Bierma-Zeinstra SM (2015) Prognostic factors for progression of clinical osteoarthritis of the knee: a systematic review of observational studies. *Arthritis Res Therapy* 17:152
- Yusuf E, Kortekaas MC, Watt I, Huizinga TW, Kloppenburg M Do knee abnormalities visualised on MRI explain knee pain in knee osteoarthritis? A systematic review. *Ann Rheum Dis* 70(1):60–67
- Hunter DJ, Lo GH, Gale D, Grainger AJ, Guermazi A, Conaghan PG (2008) The reliability of a new scoring system for knee osteoarthritis MRI and the validity of bone marrow lesion assessment: BLOKS (Boston-Leeds osteoarthritis knee score). *Ann Rheum Dis* 67(2):206–211
- Felson DT, Lynch J, Guermazi A, Roemer FW, Niu J, McAlindon T et al (2010) Comparison of BLOKS and WOMBS scoring systems part II. Longitudinal assessment of knee MRIs for osteoarthritis and suggested approach based on their performance: data from the osteoarthritis initiative. *Osteoarthr Cartil* 18(11):1402–1407
- Guermazi A, Roemer FW, Haugen IK, Crema MD, Hayashi D (2013) MRI-based semiquantitative scoring of joint pathology in osteoarthritis. *Nature reviews. Rheumatology* 9(4):236–251
- Kellgren JH, Lawrence JS (1957) Radiological assessment of osteoarthrosis. *Ann Rheum Dis* 16(4):494–502
- Ersöz M, Ergun S (2003) Relationships between knee range of motion and Kellgren-Lawrence radiographic scores in knee osteoarthritis. *Am J Phys Med Rehabil* 82(2):110–115
- Huang J, Ushiyama T, Inoue K, Kawasaki T, Hukuda S, Vitamin D (2000) Receptor gene polymorphisms and osteoarthritis of the hand, hip, and knee: a case-control study in Japan. *Rheumatology* 39(1):79–84
- Neame R, Zhang W, Deighton C, Doherty M, Doherty S, Lanyon P et al (2004) Distribution of radiographic osteoarthritis between the right and left hands, hips, and knees. *Arthritis & Rheumatism* 50(5): 1487–1494
- Sharif M, Granell R, Johansen J, Clarke S, Elson C, Kirwan JR (2006) Serum cartilage oligomeric matrix protein and other biomarker profiles in tibiofemoral and patellofemoral osteoarthritis of the knee. *Rheumatology* 45(5):522–526
- 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 & Rheumatology* 66(8):2134–2143
- Fransen M, Agalotiis M, Naim L, Votrubec M, Bridgett L, Su S et al (2015) Glucosamine and chondroitin for knee osteoarthritis: a double-blind randomised placebo-controlled clinical trial evaluating single and combination regimens. *Ann Rheum Dis* 74(5):851–858
- Medixant. RadiAnt DICOM Viewer. 2011
- McGraw KO, Wong SP (1996) Forming inferences about some intraclass correlation coefficients. *Psychol Methods* 1(1):30–46

28. Li L, Zeng L, Lin Z-J, Cazzell M, Liu H (2015) Tutorial on use of intraclass correlation coefficients for assessing interest reliability and its application in functional near-infrared spectroscopy-based brain imaging. *J Biomed Opt* 20(5):50801
29. Zaiontz C. Real statistics resource pack. 2013
30. Schorer J, Weiß CA (2007) Weighted kappa coefficient for three observers as a measure for reliability of expert ratings on characteristics in handball throwing patterns. *Meas Phys Educ Exerc Sci* 11(3):177–187
31. Cohen J (2013) *Statistical power analysis for the behavioral sciences*. Burlington: Elsevier Science
32. de Vet HCW, Terwee CB, Knol DL, Bouter LM (2006) When to use agreement versus reliability measures. *J Clin Epidemiol* 59(10): 1033–1039
33. Bossen JKJ, Hageman MGJS, King JD, Ring DC (2013) Does rewording MRI reports improve patient understanding and emotional response to a clinical report? *Clin Orthop Relat Res* 471(11):3637–3644
34. Sloan TJ, Walsh DA (2010) Explanatory and diagnostic labels and perceived prognosis in chronic low back pain. *Spine* 35(21): E1120–E11E5
35. Hunter DJ, Zhang W, Conaghan PG, Hirko K, Menashe L, Reichmann WM et al (2011) Responsiveness and reliability of MRI in knee osteoarthritis: a meta-analysis of published evidence. *Osteoarthr Cartil* 19(5):589–605
36. Farrokhi S, Piva SR, Gil AB, Fitzgerald GK (2012) Severity of patellofemoral joint disease is associated with increased functional limitations and lower limb impairments in patients with coexisting tibiofemoral osteoarthritis. *Osteoarthr Cartil* 20:S15
37. Szebenyi B, Hollander AP, Dieppe P, Quilty B, Duddy J, Clarke S et al (2006) Associations between pain, function, and radiographic features in osteoarthritis of the knee. *Arthritis Rheumatism* 54(1): 230–235