



Published in final edited form as:

*Skeletal Radiol.* 2012 February ; 41(2): 179–186. doi:10.1007/s00256-011-1142-2.

## Different thresholds for detecting osteophytes and joint space narrowing exist between the site investigators and the centralized reader in a multicenter knee osteoarthritis study—data from the Osteoarthritis Initiative

**Ali Guermazi,**

Quantitative Imaging Center, Department of Radiology, Boston University School of Medicine, 820 Harrison Avenue, FGH Building 3rd Floor, Boston, MA 02118, USA

**David J. Hunter,**

Division of Research, New England Baptist Hospital, 125 Parker Hill Avenue, Boston, MA 02120, USA

Northern Clinical School, University of Sydney, Sydney, Australia

**Ling Li,**

Division of Research, New England Baptist Hospital, 125 Parker Hill Avenue, Boston, MA 02120, USA

**Olivier Benichou,**

Eli Lilly & Co, Indianapolis, IN 46285, USA

**Felix Eckstein,**

Paracelsus Medical University, Salzburg, Austria

Chondrometrics GmbH, Ainring, Germany

**C. Kent Kwoh,**

Division of Rheumatology and Clinical Immunology, University of Pittsburgh School of Medicine, S700 Biomedical Science Tower, 3500 Terrace Street, Pittsburgh, PA 15261, USA

**Michael Nevitt, and**

Department of Epidemiology and Biostatistics, University of California, 185 Berry Street, Lobby 5, Suite 5700, San Francisco, CA 94107, USA

**Daichi Hayashi**

Quantitative Imaging Center, Department of Radiology, Boston University School of Medicine, 820 Harrison Avenue, FGH Building 3rd Floor, Boston, MA 02118, USA

Ali Guermazi: guermazi@bu.edu; David J. Hunter: david.hunter@sydney.edu.au; Ling Li: lli7@caregroup.harvard.edu; Olivier Benichou: Benichou\_Olivier@lilly.com; Felix Eckstein: felix.eckstein@pmu.ac.at; C. Kent Kwoh: kentkwoh2@gmail.com; Michael Nevitt: MNeivitt@psg-ucsf.org; Daichi Hayashi: dhayashi@bu.edu

### Abstract

---

© ISS 2011

Correspondence to: Daichi Hayashi, dhayashi@bu.edu.

**Competing interests** Ali Guermazi received grants from General Electric Healthcare and National Institute of Health. He is the President of Boston Imaging Core Lab, LLC; consultant to MerckSerono, Facet Solutions, Novartis, Genzyme, and Stryker. Felix Eckstein is CEO of Chondrometrics, GmbH, and provides consulting services to MerckSerono and Novartis. Olivier Benichou has a full time employment with Lilly & Co, IN. Other authors declared nothing to disclose.

**Objective**—To evaluate how the reading of knee radiographs by site investigators differs from that by an expert musculoskeletal radiologist who trained and validated them in a multicenter knee osteoarthritis (OA) study.

**Materials and methods**—A subset of participants from the Osteoarthritis Initiative progression cohort was studied. Osteophytes and joint space narrowing (JSN) were evaluated using Kellgren-Lawrence (KL) and Osteoarthritis Research Society International (OARSI) grading. Radiographs were read by site investigators, who received training and validation of their competence by an expert musculoskeletal radiologist. Radiographs were re-read by this radiologist, who acted as a central reader. For KL and OARSI grading of osteophytes, discrepancies between two readings were adjudicated by another expert reader.

**Results**—Radiographs from 96 subjects (49 women) and 192 knees (138 KL grade $\geq$ 2) were included. The site reading showed moderate agreement for KL grading overall ( $\kappa=0.52$ ) and for KL $\geq$ 2 (i.e., radiographic diagnosis of “definite OA”;  $\kappa=0.41$ ). For OARSI grading, the site reading showed substantial agreement for lateral and medial JSN ( $\kappa=0.65$  and  $0.71$ ), but only fair agreement for osteophytes ( $\kappa=0.37$ ). For KL grading, the adjudicator’s reading showed substantial agreement with the centralized reading ( $\kappa=0.62$ ), but only slight agreement with the site reading ( $\kappa=0.10$ ).

**Conclusion**—Site investigators over-graded osteophytes compared to the central reader and the adjudicator. Different thresholds for scoring of JSN exist even between experts. Our results suggest that research studies using radiographic grading of OA should use a centralized reader for all grading.

## Keywords

Osteoarthritis; Radiography; X-ray; Knee; Centralized reading

## Introduction

Centralized outcome assessments are widely applied in large scale multicenter clinical trials in various fields of medicine [1, 2]. In rheumatology, centralized independent radiographic assessment is performed in epidemiologic studies or clinical trials of rheumatologic disorders, such as osteoporosis [3], rheumatoid arthritis [4], and osteoarthritis (OA) [5]. This methodology is encouraged by both the Food and Drug Administration (FDA) [6, 7] and European Medicines Agency (EMA) [8] for approval of therapeutic drugs.

The Osteoarthritis Initiative (OAI) is a 4-year, multicenter, longitudinal, prospective observational cohort study, designed to identify biomarkers for the development and/or progression of symptomatic knee OA. In the OAI, radiographs were initially analyzed by clinicians (“site investigators”) at one of the four OAI clinical sites to determine eligibility for the study; results are reported to the public OAI database. From a pool of subjects thus collected, subcohorts are selected specifically for research questions formulated by different teams of researchers and centralized adjudicated readings of the radiographs are then performed. It is on this centralized reading that the results of the published and ongoing studies are based. There is potential for reading disagreement among the different clinic sites. This is an important issue not only because the initial reading affects the eligibility of subjects for a given study, but also because it is the basis for analyzing the relationship of certain radiographic features with the onset of symptoms (incident cohort) or with the structural progression of disease, e.g., MRI-based measurement of cartilage loss. Centralizing the assessments is intended to reduce investigator bias and increase the accuracy of analysis [2, 9, 10].

For non-radiographic studies, reports on the value of centralized assessment have been mixed. Some reported notable discrepancies between site and centralized assessments of myocardial infarction endpoints in the multicenter clinical trials [9, 10], but others have reported no significant impact on clinical outcomes including myocardial infarction, stroke, or other vascular/cardiovascular events [11–13]. In contrast, there is little documentation to assess the value of centralized radiographic assessments in OA studies. However, a recent report showed discrepancies between OAI screening readings by site investigators and the centralized adjudicated reading on baseline knee OA status [14] and highlighted different interpretations of definite osteophytes by different readers.

To ensure site investigators have competence in semiquantitative grading of knee radiographs, they received standardized training and validation before being qualified as a site investigator. Ideally, differences between site readings and the reading by their trainer should be minimal. The aim of this study was to compare the diagnostic performance of an expert musculoskeletal radiologist (who acted as a trainer) and the trainees (site investigators) for reading of knee radiographs using the Kellgren-Lawrence (KL) [15] and the Osteoarthritis Research Society International (OARSI) grading systems [16] in a subsample of the OAI progression cohort.

## Materials and methods

### Study sample

Subjects included in this analysis are a subset of the 4,796 participants in the OAI study, recruited at the four OAI clinical sites. The study protocol, amendments, and informed consent documentation were reviewed and approved by the local institutional review boards, and written informed consent was obtained from all participants. Data used in the present study were obtained from the OAI database available for public access (URL: <http://www.oai.ucsf.edu/>). The specific data sets used are public-use data sets 0.1.1 and 0.E. 1.

All of the participants for the present study are drawn from the progression subcohort (details of the inclusion criteria have been described previously [17]). These criteria were used to select patients for this study: bilateral frequent knee pain, body mass index (BMI)  $>25 \text{ kg/m}^2$ , medial compartment JSN (OARSI grade 1–3) in one knee with no or lower-grade lateral compartment JSN (OARSI grade 0–2) in the same knee, and no or lower-grade JSN (OARSI grade 0–2) in the contralateral knee.

### Acquisition of radiographs

The bilateral posteroanterior view was obtained using a SynaFlexer™ frame (Synarc, San Francisco, CA, USA) to position the subject's feet reproducibly [18]. Body weight is distributed equally between the two legs and the knees and thighs are pressed directly against the wall of the frame, the anterior wall of which was in contact with the Bucky or reclining tabletop of the radiographic unit. This positioning results in a fixed angle of knee flexion of about  $20^\circ$ . A V-shaped angulation support on the base of the frame is used to fix the foot below the index knee in  $10^\circ$  external rotation. The x-ray beam is angled  $10^\circ$  caudally [19].

### Training of site investigators

Baseline radiographs were initially assessed at four clinical sites by site investigators, who were experienced rheumatologists and had received standardized training and validation in the use of KL and OARSI radiographic grading methods by means of teleconference and a web-based program. Each trainee was provided a username and a password to login to the

online training and validation program. In the training mode, site investigators read 30 knee radiographs and graded osteophytes and joint space width, and had the opportunity to give their own answers. They were then able to view the correct answers with annotated images and comments about possible pitfalls. In the validation mode, the investigators read another set of 30 knee radiographs and their results were compared against the gold standard, which was provided by an experienced musculoskeletal radiologist (AG) who was an expert in KL and OARSI grading. This trainer acted as the central reader in the present study. The trainees were allowed to take training sessions as many times as they wished and were required to pass the validation process before being qualified to participate as a site investigator in the OAI. At an early stage of recruitment for the OAI, a sample of site readings was centrally reviewed by the aforementioned musculoskeletal radiologist, and feedback was provided when there were discrepancies [14].

### Radiographic assessment

Site investigators evaluated the KL grade on a 0–4 scale as well as individual radiographic features, i.e., osteophytes and JSN, on a modified OARSI grade [16]. In this modification, the original OARSI grades of 1 and 2 for the assessment of medial and lateral tibiofemoral JSN were collapsed to 1, and the original grade of 3 became 2. Thus, grade 0=no JSN, grade 1=definite JSN present, and grade 2=severe JSN. For osteophytes, the original OARSI grades of 2 and 3 were collapsed to grade 2. Thus, grade 0=no osteophytes, grade 1=definite osteophytes present, and grade 2=large osteophytes. These readings constituted the “site reading” in this study.

For the purpose of this study, the same set of radiographs was read by an expert musculoskeletal radiologist (AG), who is the trainer of site investigators in OAI and has 8 years of experience in radiographic assessment of knee OA with KL and OARSI grading systems using the OARSI atlas [20]. The modified OARSI grading was also used instead of the original OARSI grading to enable direct comparison of results with the site reading. This reader was blinded to the results of the site readings. These readings constituted the “centralized reading” in this study and were used as a reference standard.

As an additional assessment, an adjudicator [a rheumatologist (DJH) who is also an expert with 12 years of experience in radiographic assessment of knee OA, but was not involved in OAI as a site investigator or as a trainer] reviewed cases in which disagreements between the site and the centralized readings existed with regard to diagnosis of radiographic OA (i.e., detection of the definite osteophytes using KL and OARSI grading). The adjudicator made the final decision in an independent and blinded manner, i.e., neither site investigators nor the central reader were present during the adjudicator’s reading session, and the adjudicator did not know the results of either of the readings. The results of the adjudication reading served as an alternative reference standard for comparing the site readings and the centralized readings. Radiographic readings were performed using digital imaging software (eFilm Workstation, version 2.0.0, Merge Healthcare, Milwaukee, WI).

### Statistical analysis

Agreement between the site reading and the centralized reading was compared. For the OARSI grading, agreement [kappa and 95% confidence interval (CI)] was calculated for evaluation of each OA feature, i.e., lateral JSN, medial JSN, and osteophytes. For the KL grading, overall agreement was evaluated for the two readings.

Because there is no definitive gold standard against which to compare the diagnostic performance of all readers, it was not possible to calculate the sensitivity or specificity to judge the diagnostic performance of the site investigators, the central reader, or the

adjudicator. In this paper, therefore, we will compare diagnostic performance of the site investigators, the central reader, and the adjudicator relative to one another.

Lastly, for the cases where the site reading and the central reading disagreed, the adjudicator's reading using KL grading was used as an alternative reference to calculate agreement (kappa) with each reading. This was done to assess if the adjudicator agreed more with the site reading or the centralized reading for overall KL grading. We used a weighted least-squares approach to test the significance of differences between kappa values for the centralized and the site readings.

All statistical calculations were performed using SAS<sup>®</sup> software (Version 9.1 for Windows, SAS Institute, Cary, NC).

## Results

### Demographics

The radiographs from 96 subjects (192 knees) were read in the present study. Their mean age was 61.1 (standard deviation $\pm$ 8.8) years, mean BMI was 30.7 $\pm$ 4.1 kg/m<sup>2</sup>, 49 (51.0%) were women, and 138 (71.9%) knees were KL grade $\geq$ 2 according to the centralized radiographic reading.

### Agreement with the centralized reading

The number of knees in which the two readings agreed on all four features (i.e., KL grade and osteophytes, medial and lateral JSN on the OARSI grading) was 62 (32%). In one case, the site reading completely disagreed with the centralized reading on all features.

Table 1 summarizes the agreement of the KL grades and each feature of the OARSI grading between the site reading and the centralized. For KL grades, grading of KL 0–4 overall showed moderate agreement (kappa=0.52), which was a higher level of agreement than for grading of KL $\geq$ 2 (i.e., radiographic diagnosis of “definite OA”; kappa=0.41). For the OARSI grading, the site reading showed substantial agreement for grading of lateral and medial JSN (kappa= 0.65 and 0.71), but only fair agreement for osteophytes (kappa=0.37). Site investigators over-graded osteophytes in 88 cases (46%) relative to the central reader (Table 2).

### Comparison for the Kellgren-Lawrence grade

There were 47 KL grade 0 knees, 7 grade 1 knees, 30 grade 2 knees, 83 grade 3 knees, and 25 grade 4 knees, and 138 knees had radiographic OA according to the centralized reading. Notably, the site investigators over-graded 33 KL grade 0 knees (by the centralized reading) as grade 1 or 2 (21 and 12 cases, respectively, Fig. 1), and 5 KL grade 1 knees as grade 2 or 3 (4 and 1 cases, respectively, Fig. 2).

### Comparison with the adjudicated reading

The site reading and the centralized reading disagreed in 93 knees in regard to KL grading and OARSI grading of osteophytes. For KL grading, the adjudicated reading agreed with the site reading in 19 cases and disagreed in 74 cases, whereas it agreed with the centralized reading in 58 knees and disagreed in 35 knees (Table 3). The overall agreement between the site reading and the adjudicated reading was kappa=0.10 (−0.01 to −0.21), and that for the centralized reading was kappa=0.62 (0.51–0.74). The adjudicated reading showed statistically higher agreement with the centralized reading than the site reading ( $s < 0.0001$ , by weighted least-squares approach). For OARSI grading, the site reading showed virtually no agreement, while the centralized reading showed almost complete agreement (Table 4).

## Discussion

In the present study, agreement between the site reading and the centralized reading for grading of KL $\geq$ 2 (i.e., radiographic diagnosis of “definite OA”) was moderate. For the OARSI grading, the site reading showed substantial agreement for grading of lateral and medial JSN, but only fair agreement for osteophytes. In 93 cases in which two readings had discrepant KL grading, an independent adjudicator’s reading showed substantial agreement with the centralized reading, but only slight agreement with the site reading.

A potential limitation of this study was that the adjudicator only assessed cases where the site and the centralized readings disagreed. Thus, there was a risk that the adjudicator would favor the judgment of the central reader. We attempted to minimize this bias by performing the adjudication process independently and blinded. Also, since there was no “definitive” gold standard, we cannot completely rule out the possibility that both the central reader and the adjudicator were under-grading osteophytes, relative to site investigators. Thus, our study cannot prove if the centralized and adjudicated readings are better than the site reading, and it is only possible to state how different they are from one another.

Radiographic diagnosis of “definite” knee OA is made when a knee has a “definite” osteophyte (i.e., classified as KL grades $\geq$ 2). Detection of “definite” osteophytes is often used as the threshold for inclusion criteria in knee OA studies and thus diagnostic accuracy for osteophyte detection needs to be high. Recently, discrepancies between OAI screening readings by site investigators and the centralized reading using KL grading were reported [14]. There was a tendency for site investigators to over-grade osteophytes compared to central readers, i.e., 25% of knees with definite osteophytes by site reading had KL grades  $<$ 2 by the centralized reading. It was suggested that there were different interpretations of “definite” osteophytes.

In the present study, site investigators also had a tendency to over-grade osteophytes relative to the central reader based on OARSI grading (Tables 2, 3, and Fig. 1). The site investigators over-graded osteophytes in 46% (88/192) of cases relative to the central reader. Of those, 47 cases were graded as grade 0 (i.e., no osteophyte) by the central reader. Likewise, the site investigators over-graded osteophytes in 94% (87/93) of cases relative to the adjudicator. The adjudicated reading almost completely disagreed with the site reading, which is in stark contrast with almost complete agreement between the centralized and the adjudicated readings (Table 3).

Often, readers are faced with lesions that may be considered normal by some and abnormal by others. For example, femur or tibia may have a double-contour, and it is possible that site investigators were more likely to consider it to be an osteophyte, and the central reader/adjudicator tended to think it was not an osteophyte. This seemed to occur frequently in the present study, as site investigators over-graded knees without osteophytes (as determined by the central reader and the adjudicator) as having grade 1 or 2 osteophytes in a number of cases (Fig. 1).

Regarding over-grading of OARSI grade 1 osteophytes as grade 2, one plausible explanation might be that site investigators were given images of the smallest possible “definite” osteophyte as an example of grade 1 lesion in the OARSI atlas [14]. Thus, any lesion that was only slightly larger than the example could have been classified as OARSI grade 2, even if that lesion would not be large enough to qualify for a grade 2 lesion in the eyes of the central reader.

A possible explanation for these differences in the thresholds for grading osteophytes between site investigators and the central reader/adjudicator may be that the site

investigators were motivated to recruit as many subjects as possible to meet the recruitment goal of the research study within a certain time limit. It is plausible that they wished to include doubtful/borderline cases, rather than exclude them, which would not bring them closer to meeting their recruitment goal. Such external pressure did not exist for the central reader or the adjudicator in this study. Another possible reason might be that the site investigators had biases to over-grade OA based on their clinical assessment of the patient whom they felt had OA. Although these remain speculations, these possibilities may highlight the importance of the centralized reading, which is unlikely to be influenced by such external factors or availability of clinical information. It would thus be suggested that all radiographs should be sent to the centralized reader for consideration for study inclusion rather than use the site investigators for initial selection.

The site reading and the centralized reading demonstrated substantial agreement for OARSI grading of both lateral and medial JSN. For the lateral compartment, 185 of 192 knees (96%) had no JSN by the centralized reading and there were only 5 cases in which two readings differed by 1 grade. For the medial compartment, the site investigators under-graded 23 grade 1 JSN as grade 0, and over-graded 12 grade 0 cases as grade 1, relative to the central reader, and the two readings disagreed in 39 cases. The adjudicator disagreed with the site reading in 23 cases, and also disagreed with the centralized reading in 12 cases. Thus, there seemed to be a variable threshold for JSN among different readers, even between two expert readers (Fig. 3), but we did not detect any trends towards either over-grading or under-grading by the site investigators.

Finally, a literature review yielded several publications dealing with observer variation in radiographic grading of OA. Bellamy et al. [21] assessed the interrater reliability of two radiologists experienced in reporting hand, hip, and knee radiographs. Interrater agreement (kappa) was 0.58 for JSN and 0.82 for osteophytes in the knee. However, in this study, two readers were both expert radiologists, and KL or OARSI grading methods were not used. A study by Cooper et al. [22] compared diagnostic performance of five observers, one of whom was an experienced consultant radiologist who trained four other readers. Inter-observer agreement (kappa) was about 0.5 for JSN and less than 0.6 for osteophytes (NB: exact values not presented in the publication). In this study, interobserver agreement was calculated for all five observers. Gunther and Yi [23] demonstrated that grading of JSN is highly reader-dependent and interobserver reliability (intraclass correlation coefficient, ICC) ranged from 0.24 to 0.73 among one experienced orthopaedic consultant and two orthopaedic residents. Unlike our findings, they found interobserver reliability for grading of osteophytes (ICC 0.65–0.81) was higher than that for JSN. Moreover, interobserver reliability for KL grading was 0.76–0.84 among three readers. Thus, this study showed two relatively inexperienced readers may achieve a high degree of agreement relative to the experienced reader for the grading of osteophytes and overall assessment of knee radiographs using KL grading, but not for grading of JSN. Lastly, Vilalta et al. [24] used the modified KL grade to evaluate osteophytes and JSN. Three observers (specialist orthopaedic surgeon, orthopaedic resident, and a newly qualified physician) received training before the reading, but showed interobserver agreement (kappa) of 0.31–0.55 for osteophytes and 0.31–0.49 for JSN. However, the last two studies did not involve any radiologists, and it is unknown if the diagnostic performance of the orthopaedic consultant is comparable to an experienced musculoskeletal radiologist. Inclusion of a radiologist could potentially have led to different results.

To sum up, our findings may not agree with previously published reports, but it can be said that all previous reports have used different analytic approaches with various grading methods and readers with different qualifications and variable degrees of experience in radiographic assessment [25]. Thus, one needs to be cautious when interpreting results of

these studies; it is difficult to directly compare our results with those from the previous reports. However, overall, these reports show that observer variability may exist even among expert readers or even after provision of training for non-expert readers. In this respect, our findings are in line with previous publications.

## Conclusion

Our study has shown that the site reading showed substantial agreement with the centralized reading regarding OARSI grading of JSN, but site investigators had a tendency to over-grade osteophytes, compared with the central reader and the adjudicator. Even after receiving standardized training and being validated to become a site investigator by the trainer of OAI, i.e., the central reader of this study, different thresholds for grading osteophytes and JSN still existed. These findings may have an impact when a given study is based on eligibility criteria that take into account the KL grade or the presence/absence of osteophytes, and would suggest that research studies using radiographic grading of OA should use a centralized reader for all grading.

## Abbreviations

<b>OA</b>	Osteoarthritis
<b>OAI</b>	Osteoarthritis Initiative
<b>KL</b>	Kellgren-Lawrence
<b>OARSI</b>	Osteoarthritis Research Society International
<b>JSN</b>	Joint space narrowing
<b>MRI</b>	Magnetic resonance imaging
<b>FDA</b>	Food and Drug Administration
<b>EMA</b>	European Medicines Agency

## Acknowledgments

**Funding sources** The Osteoarthritis Initiative (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 (NIH), 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. The central radiographic readings and adjudications (i.e., data collection) were funded by Eli Lilly & Company. The sponsors did not have any role in study design, analysis, and interpretation of data; writing of the manuscript; or the decision to submit the manuscript for publication.

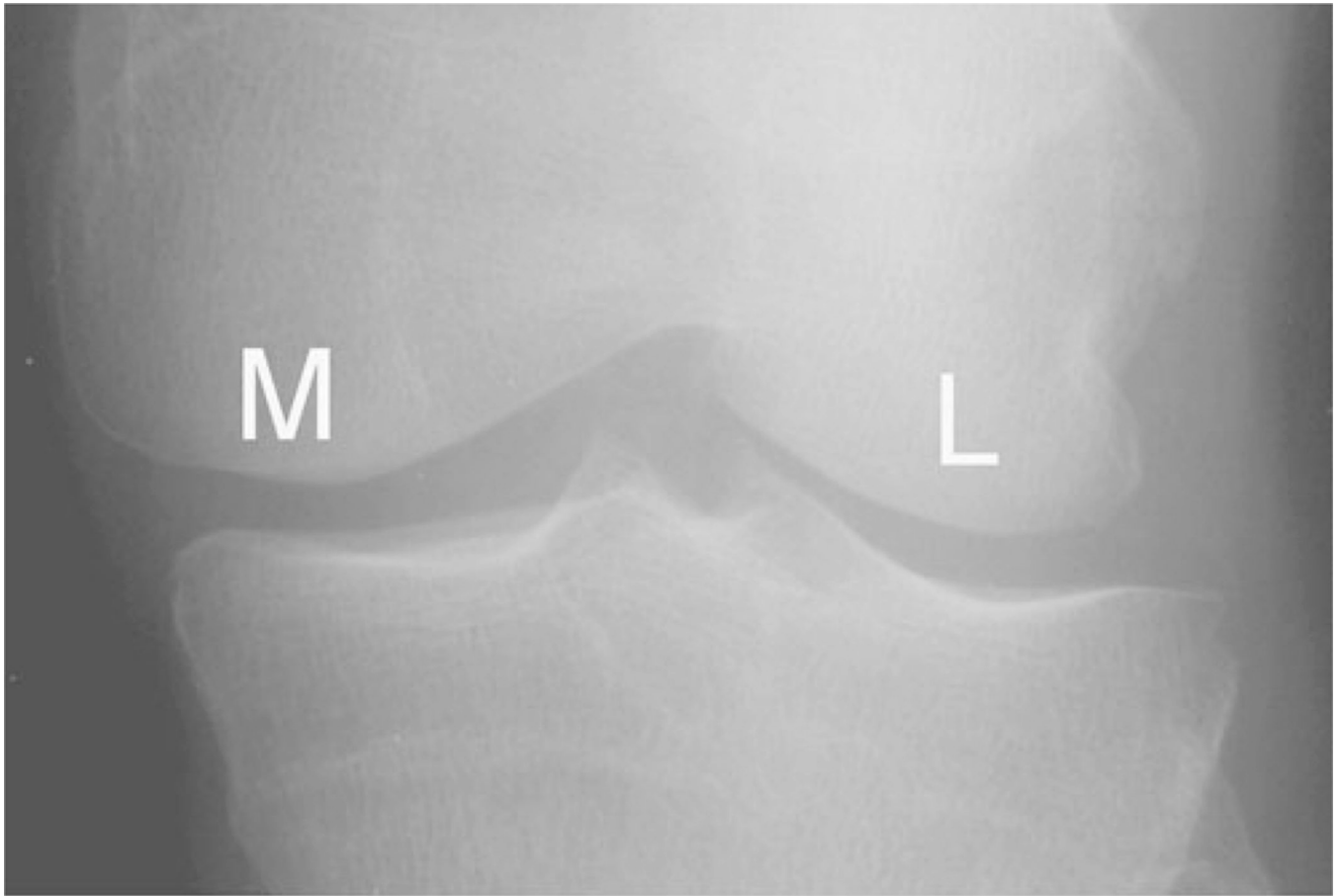
## References

1. Agnelli G, Gussoni G, Bianchini C, Verso M, Mandala M, Cavanna L, et al. Nadroparin for the prevention of thromboembolic events in ambulatory patients with metastatic or locally advanced solid cancer receiving chemotherapy: a randomised, placebo-controlled, double-blind study. *Lancet Oncol.* 2009; 10:943–949. [PubMed: 19726226]
2. Walter SD, Cook DJ, Guyatt GH, King D, Troyan S. Outcome assessment for clinical trials: how many adjudicators do we need? *Control Clin Trials.* 1997; 18:27–42. [PubMed: 9055050]
3. Saag KG, Zanchetta JR, Devogelaer JP, et al. Effects of teriparatide versus alendronate for treating glucocorticoid-induced osteoporosis: thirty-six-month results of a randomized, double-blind, controlled trial. *Arthritis Rheum.* 2009; 60:3346–3355. [PubMed: 19877063]



4. Soubrier M, Puechal X, Sibilia J, et al. Evaluation of two strategies (initial methotrexate monotherapy vs. its combination with adalimumab) in management of early active rheumatoid arthritis: data from the GUEPARD trial. *Rheumatology*. 2009; 48:1429–1434. [PubMed: 19741011]
5. Kwoh CK, Roemer FW, Hannon MJ, Moore CE, Jakicic JM, Guermazi A, et al. The Joints On Glucosamine (JOG) Study: a randomized, double-blind, placebo-controlled trial to assess the structural benefit of glucosamine in knee osteoarthritis based on 3 T MRI. *Arthritis Rheum*. 2009; 60:S725–S726.
6. U.S. Department of Health and Human Services, Food and Drug Administration. [Accessed Feb. 18, 2010] Clinical development programs for drugs, devices, and biological products intended for the treatment of osteoarthritis (OA). Draft guidance. 1999 July. <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM071577.pdf>
7. [Accessed Feb. 18, 2010] OARSI FDA OA initiative. [http://www.oarsi.org/index2.cfm?section=OARSI\\_Initiatives&content=FDA\\_OA\\_Initiative](http://www.oarsi.org/index2.cfm?section=OARSI_Initiatives&content=FDA_OA_Initiative)
8. European Medicines Agency Committee for Medicinal Products for Human Use (CHMP). [Accessed Feb. 18, 2010] Guidelines on clinical investigation of medicinal products used in the treatment of osteoarthritis. 2010 Jan. (Doc. Ref. CPMP/EWP/784/97 Rev. 1) <http://www.ema.europa.eu/pdfs/human/ewp/078497en.pdf>
9. Mahaffey KW, Harrington RA, Akkerhuis M, Kleiman NS, Berdan LG, Crenshaw BS, et al. Disagreements between central clinical events committee and site investigator assessments of myocardial infarction end-points in an international clinical trial: review of the PURSUIT study. *Curr Control Trials Cardiovasc*. 2001; 2:187–194.
10. Mahaffey KW, Roe MT, Dyke CK, Newby LK, Kleiman NS, Connolly P, et al. Misreporting of myocardial infarction end points: results of adjudication by a central clinical events committee in the PARAGON-B trial. *Am Heart J*. 2002; 143:242–248. [PubMed: 11835026]
11. Ninomiya T, Donnan G, Anderson N, Bladin C, Chambers B, Gordon G, et al. Effects of the end point adjudication process on the results of the perindopril protection against recurrent stroke study (PROGRESS). *Stroke*. 2009; 40:2111–2115. [PubMed: 19359647]
12. Granger CB, Vogel V, Cummings SR, Held P, Fiedorek F, Lawrence M, et al. Do we need to adjudicate major clinical events? *Clin Trials*. 2008; 5:56–60. [PubMed: 18283081]
13. Pogue J, Walter SD, Yusuf S. Evaluating the benefit of event adjudication of cardiovascular outcomes in large simple RCTs. *Clin Trials*. 2009; 6:239–251. [PubMed: 19528133]
14. Nevitt MC, Felson DT, Maeda JS, Aliabadi PA, McAlindon T, Lynch JA, et al. Central vs. clinic reading of knee radiographs for baseline OA in the Osteoarthritis Initiative Progression Cohort: implications for public data users. *Osteoarth Cartil*. 2009; 17 Suppl 1:S229.
15. Kellgren JH, Lawrence JS. Radiological assessment of osteoarthritis. *Ann Rheum Dis*. 1957; 16:494–502. [PubMed: 13498604]
16. Guermazi A, Hunter DJ, Roemer FW. Plain radiography and magnetic resonance imaging diagnostics in osteoarthritis: validated staging and scoring. *J Bone Joint Surg Am*. 2009; 91:54–62. [PubMed: 19182026]
17. Hunter DJ, Niu J, Zhang Y, Totterman S, Tamez J, Dabrowski C, et al. Change in cartilage morphometry: a sample of the progression cohort of the Osteoarthritis Initiative. *Ann Rheum Dis*. 2009; 68:349–356. [PubMed: 18408248]
18. Kothari M, Guermazi A, von Ingersleben G, Miaux Y, Sieffert M, Block JE, et al. Fixed-flexion radiography of the knee provides reproducible joint space width measurements in osteoarthritis. *Eur Radiol*. 2004; 14(9):1568–1573. [PubMed: 15150666]
19. Peterfy C, Li J, Zaim S, Duryea J, Lynch J, Mizux Y, et al. Comparison of fixed-flexion positioning with fluoroscopic semi-flexed positioning for quantifying radiographic joint-space width in the knee: test-retest reproducibility. *Skeletal Radiol*. 2003; 32:128–132. [PubMed: 12605275]
20. Altman RD, Hochberg M, Murphy WAJ, Wolfe F, Lequesne M. Atlas of individual radiographic features in osteoarthritis. *Osteoarthritis Cartilage*. 1995; 3:3–70. [PubMed: 8581752]

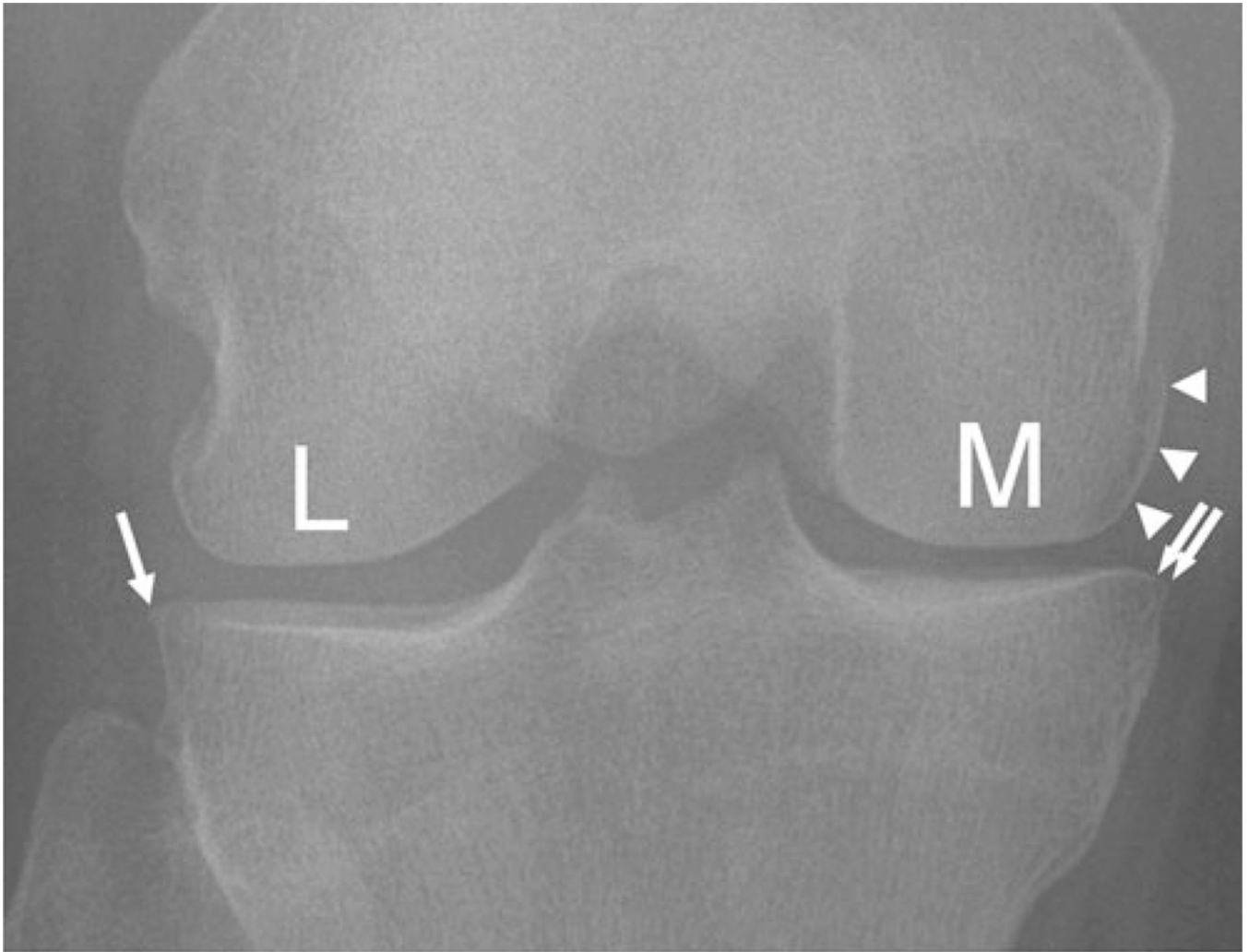
21. Bellamy N, Tesar P, Walker D, Klestov A, Muirden K, Kuhnert P, et al. Perceptual variation in grading hand, hip and knee radiographs: observations based on an Australian twin registry study of osteoarthritis. *Ann Rheum Dis.* 1999; 58(12):766–769. [PubMed: 10577964]
22. Cooper C, Cushnaghan J, Kirwan JR, Dieppe PA, Rogers J, McAlindon T, et al. Radiographic assessment of the knee joint in osteoarthritis. *Ann Rheum Dis.* 1992; 51(1):80–82. [PubMed: 1540044]
23. Gunther KP, Sun Y. Reliability of radiographic assessment in hip and knee osteoarthritis. *Osteoarthritis Cartilage.* 1999; 7(2):239–246. [PubMed: 10222223]
24. Vilalta C, Nunez M, Segur JM, Domingo A, Carbonell JA, Macule F. Knee osteoarthritis: interpretation variability of radiological signs. *Clin Rheumatol.* 2004; 23(6):501–504. [PubMed: 15278759]
25. Sun Y, Gunther KP, Brenner H. Reliability of radiographic grading of osteoarthritis of the hip and knee. *Scand J Rheumatol.* 1997; 26(3):155–165. [PubMed: 9225869]



**Fig. 1.** Example of a left knee with OARSI grade 0 osteophytes (i.e., no osteophytes) and KL grade 0 by the centralized reading. Site investigators classified this knee as KL grade 2 and recorded OARSI grade 2 for osteophytes. The adjudicator also classified this knee to be KL grade 0 with OARSI grade 0 for osteophyte, agreeing with the central reader. All three readings agreed that there was no medial (*M*) or lateral (*L*) JSN



**Fig. 2.** Example of a right knee with OARSIS grade 0 JSN [i.e., no medial (*M*) or lateral (*L*) JSN] and KL grade 1 [i.e., lateral tibial osteophytic lipping (*arrow*), but no definite osteophyte] by the centralized reading. Site investigators classified this knee as KL grade 3 and recorded OARSIS grade 1 for medial JSN and grade 2 for osteophyte. The adjudicator completely agreed with the central reader and assigned OARSIS grade 0 for JSN and KL grade 1



**Fig. 3.**

Example of a right knee with OARSI grade 1 medial (*M*) JSN, grade 2 medial femoral osteophyte (*arrowheads*), and KL grade 3 by the centralized reading. There is also lateral (*arrow*) and medial (*double arrows*) tibial osteophytic lipping. The site investigator assigned OARSI grade 0 for both medial (*M*) and lateral (*L*) JSN, grade 1 for osteophyte, and KL grade 1. The adjudicator also assigned OARSI grade 0 for JSN agreeing with the site reading, but grade 2 for medial femoral osteophyte agreeing with the centralized reading, and consequently classified this knee as KL grade 2. This is an example to demonstrate that different thresholds can exist for detection of radiographic OA features even between expert readers

**Table 1**

Agreement of site reading with the centralized reading (the reference standard)

Grading methods	Kappa (95% CI)	
KL		
All grades	0.52	(0.43–0.60)
KL $\geq$ 2	0.41	(0.27–0.54)
OARSI		
JSN lateral	0.65	(0.37–0.94)
JSN medial	0.71	(0.62–0.80)
Osteophyte	0.37	(0.28–0.45)

CI Confidence interval, KL Kellgren-Lawrence grade, OARSI Osteoarthritis Research International grade, JSN joint space narrowing

**Table 2**

Summary of the site reading vs. the centralized reading using the modified OARSI grading

<i>(n=192)</i>	Site		
	0	1	2
Centralized			
0	18	30	17
1	0	18	41
2	0	5	63

The number in each cell denotes the number of knees classified into each grade

**Table 3**

Summary of the site reading vs. the adjudicated reading and the centralized reading using the modified OARSI grading for osteophytes

<i>(n=93)</i>	Site		<i>(n=93)</i>		Centralized		
	0	1	2	Adjudicated	0	1	2
0	0	29	17	0	45	1	0
1	0	2	41	1	2	40	1
2	0	4	0	2	0	0	4

The number in each cell denotes the number of knees classified into each grade



**Table 4**

Summary of the site reading and the centralized reading using KL grading relative to the adjudicated reading for the 93 knees in which there were discrepancies between the site and the centralized readings

<i>(n=93)</i>	Site reading					Centralized reading				
	0	1	2	3	4	0	1	2	3	4
Adjudicated	0	1	2	3	4	0	1	2	3	4
0	2	24	13	0	0	32	1	0	6	0
1	0	9	6	1	0	1	4	5	6	0
2	0	6	6	2	0	0	2	12	0	0
3	0	14	6	2	1	0	1	0	19	3
4	0	0	0	1	0	0	0	0	0	1

The number in each cell denotes the number of knees classified into each grade by the KL system