Title
Osteoarthritis Classification Scales

Journal
Journal of Bone and Joint Surgery, 96(14)

ISSN
0021-9355

Authors
Wright, Rick W
Wright, Rick W
Ross, James R
et al.

Publication Date
2014-07-16

DOI
10.2106/jbjs.m.00929

Peer reviewed

Osteoarthritis Classification Scales: Interobserver Reliability and Arthroscopic Correlation

The MARS Group*

Background: Osteoarthritis of the knee is commonly diagnosed and monitored with radiography. However, the reliability of radiographic classification systems for osteoarthritis and the correlation of these classifications with the actual degree of confirmed degeneration of the articular cartilage of the tibiofemoral joint have not been adequately studied.

Methods: As the Multicenter ACL (anterior cruciate ligament) Revision Study (MARS) Group, we conducted a multicenter, prospective longitudinal cohort study of patients undergoing revision surgery after anterior cruciate ligament reconstruction. We followed 632 patients who underwent radiographic evaluation of the knee (an anteroposterior weight-bearing radiograph, a posteroanterior weight-bearing radiograph made with the knee in 45° of flexion [Rosenberg radiograph], or both) and arthroscopic evaluation of the articular surfaces. Three blinded examiners independently graded radiographic findings according to six commonly used systems—the Kellgren-Lawrence, International Knee Documentation Committee, Fairbank, Brandt et al., Ahlbäck, and Jäger-Wirth classifications. Interobserver reliability was assessed with use of the intraclass correlation coefficient. The association between radiographic classification and arthroscopic findings of tibiofemoral chondral disease was assessed with use of the Spearman correlation coefficient.

Results: Overall, 45° posteroanterior flexion weight-bearing radiographs had higher interobserver reliability (intraclass correlation coefficient = 0.63; 95% confidence interval, 0.61 to 0.65) compared with anteroposterior radiographs (intraclass
continued

Peer Review: This article was reviewed by the Editor-in-Chief and one Deputy Editor, and it underwent blinded review by two or more outside experts. It was also reviewed by an expert in methodology and statistics. The Deputy Editor reviewed each revision of the article, and it underwent a final review by the Editor-in-Chief prior to publication. Final corrections and clarifications occurred during one or more exchanges between the author(s) and copyeditors.

*Rick W. Wright, MD, James R. Ross, MD, Amanda K. Haas, MA, Laura J. Huston, MS, Elizabeth A. Garofoli, David Harris, BE, Kushal Patel, BA, David Pearson, BA, Jake Schutzman, Majd Tarabichi, David Ying, SB, John P. Albright, MD, Christina R. Allen, MD, Annunziato Amendola, MD, Allen F. Anderson, MD, Jack T. Andrish, MD, Christopher C. Annunziata, MD, Robert A. Arciero, MD, Bernard R. Bach Jr., MD, Champ L. Baker III, MD, Arthur R. Bartolozzi, MD, Keith M. Baumgarten, MD, Jeffery R. Bechler, MD, Jeffrey H. Berg, MD, Geoffrey A. Bernas, MD, Stephen F. Brockmeier, MD, Robert H. Brophy, MD, Charles A. Bush-Joseph, MD, J. Brad Butler V, MD, John D. Campbell, MD, James E. Carpenter, MD, Brian J. Cole, MD, Daniel E. Cooper, MD, Jonathan M. Cooper, DO, Charles L. Cox, MD, R. Alexander Creighton, MD, Diane L. Dahm, MD, Tal S. David, MD, Thomas M. DeBerardino, MD, Warren R. Dunn, MD, MPH, David C. Flanigan, MD, Robert W. Frederick, MD, Theodore J. Ganley, MD, Charles J. Gatt Jr., MD, Steven R. Gecha, MD, James Robert Giffin, MD, Sharon L. Hame, MD, Jo A. Hannafin, MD, PhD, Christopher D. Harner, MD, Norman Lindsay Harris Jr., MD, Keith S. Hechtman, MD, Elliott B. Hershman, MD, Rudolf G. Hoellrich, MD, Timothy M. Hosea, MD, David C. Johnson, MD, Timothy S. Johnson, MD, Morgan H. Jones, MD, Christopher C. Kaeding, MD, Ganesh V. Kamath, MD, Thomas E. Klootwyk, MD, Brett A. Lantz, MD, Bruce A. Levy, MD, C. Benjamin Ma, MD, G. Peter Maiers II, MD, Barton Mann, PhD, Robert G. Marx, MD, Matthew J. Matava, MD, Gregory M. Mathien, MD, David R. McAllister, MD, Eric C. McCarty, MD, Robert G. McCormack, MD, Bruce S. Miller, MD, MS, Carl W. Nissen, MD, Daniel F. O'Neill, MD, EdD, LTC Brett D. Owens, MD, Richard D. Parker, MD, Mark L. Purnell, MD, Arun J. Ramappa, MD, Michael A. Rauh, MD, Arthur Rettig, MD, Jon K. Sekiya, MD, Kevin G. Shea, MD, Orrin H. Sherman, MD, James R. Slauterbeck, MD, Matthew V. Smith, MD, Jeffrey T. Spang, MD, Kurt P. Spindler, MD, Michael J. Stuart, MD, LTC Steven J. Svoboda, MD, Timothy N. Taft, MD, COL Joachim J. Tenuta, MD, Edwin M. Tingstad, MD, Armando F. Vidal, MD, Darius G. Viskontas, MD, Richard A. White, MD, James S. Williams Jr., MD, Michelle L. Wolcott, MD, Brian R. Wolf, MD, James J. York, MD, and James L. Carey, MD, MPH

Disclosure: The lead author did not receive payments or services, either directly or indirectly (i.e., via his institution), from a third party in support of any aspect of this work. He, or his institution, has had a financial relationship, in the thirty-six months prior to submission of this work, with an entity in the biomedical arena that could be perceived to influence or have the potential to influence what is written in this work. The lead author has not had any other relationships, or engaged in any other activities, that could be perceived to influence or have the potential to influence what is written in this work. The complete **Disclosures of Potential Conflicts of Interest** submitted by authors are always provided with the online version of the article.

A commentary by Gregory P. Guyton, MD, is linked to the online version of this article at jbjs.org.

correlation coefficient = 0.55; 95% confidence interval, 0.53 to 0.56). Similarly, the 45° posteroanterior flexion weight-bearing radiographs had higher correlation with arthroscopic findings of chondral disease (Spearman rho = 0.36; 95% confidence interval, 0.32 to 0.39) compared with anteroposterior radiographs (Spearman rho = 0.29; 95% confidence interval, 0.26 to 0.32). With respect to standards for the magnitude of the reliability coefficient and correlation coefficient (Spearman rho), the International Knee Documentation Committee classification demonstrated the best combination of good interobserver reliability and medium correlation with arthroscopic findings.

Conclusions: The overall estimates with the six radiographic classification systems demonstrated moderate (anteroposterior radiographs) to good (45° posteroanterior flexion weight-bearing radiographs) interobserver reliability and medium correlation with arthroscopic findings. The International Knee Documentation Committee classification assessed with use of 45° posteroanterior flexion weight-bearing radiographs had the most favorable combination of reliability and correlation.

Level of Evidence: Diagnostic Level I. See Instructions for Authors for a complete description of levels of evidence.

Osteoarthritis of the knee is a disabling disease resulting in joint discomfort, restricted motion, and diminished function[1-3]. Radiography is the most widely used method to diagnose and monitor the progression of osteoarthritis. Radiographic classification systems are used across many medical disciplines, including rheumatology, orthopaedics, internal medicine, basic science, and clinical research. However, in only a few studies have authors investigated the reliability of classification scales or their correlation with the actual degree of confirmed articular cartilage degeneration within the tibiofemoral compartment of the knee joint[4,5]. To our knowledge, this relationship has not been evaluated by multiple investigators using radiographic analysis of the various classification systems over a wide spectrum of articular changes.

The Multicenter ACL (anterior cruciate ligament) Revision Study (MARS) consortium[6] is a prospective longitudinal cohort that provides a unique opportunity to investigate a wide array of radiographic changes secondary to osteoarthritis with direct arthroscopic assessment of the articular cartilage at the time of revision ACL reconstruction. The spectrum of disease within this cohort ranges from normal articular cartilage to full-thickness chondral disease, which is an appropriate range of disease presentation to adequately assess these classification scales.

The primary aim of this study was twofold: (1) to determine which classification system for tibiofemoral osteoarthritis has the most interobserver reliability, and (2) to determine which tibiofemoral osteoarthritis classification system best correlates with arthroscopic articular cartilage findings. The secondary aim revolved around radiographic issues that also have the potential to change practice: (a) to determine which common radiograph of the knee provides the most reliability, and (b) to determine whether bilateral images (involved and uninvolved knee) are significantly more reliable than unilateral images are. The answers will impact a variety of disciplines as well as the interpretation of all musculoskeletal research relating to osteoarthritis of the knee.

First, we hypothesize that no single radiographic classification system for osteoarthritis is superior in interobserver reliability. However, we believe that a classification scheme that incorporates joint space narrowing will be the most informative. Second, we hypothesize that all radiographic classification systems will correlate poorly with the actual degree of chondral damage detected arthroscopically. Finally, we hypothesize that a posteroanterior weight-bearing radiograph made with the knee in 45° of flexion (Rosenberg radiograph)[7] will be better than an anteroposterior weight-bearing radiograph, and bilateral views will be better than a unilateral view, for predicting intra-articular disease and will have better interobserver reliability.

Materials and Methods

Patients

The MARS consortium is a multicenter cohort consisting of eighty-three surgeons across fifty-two sites, with the aim of identifying modifiable predictors of outcomes for patients undergoing revision ACL reconstruction. It is an American Orthopaedic Society for Sports Medicine-sponsored and National Institutes of Health-funded investigation that began enrolling patients in March 2006[6]. In a descriptive epidemiological study, posttraumatic articular cartilage disease, ranging from mild fibrillation and fraying to full-thickness chondral defects, was noted in >70% of these patients at the time of revision surgery[6]. This cohort provides a unique opportunity to investigate a wide array of radiographic changes secondary to osteoarthritis and allows direct comparison with the status of the articular cartilage at the time of revision ACL reconstruction with use of arthroscopy. Eligible patients were between the ages of twelve and sixty-five years, had an ACL graft deficiency, and underwent a revision ACL reconstruction performed by a participating MARS Group surgeon.

Radiographs

As part of the enrollment in the MARS investigation, each patient had required preoperative weight-bearing anteroposterior and full-extension lateral radiographs made. Additional recommended radiographs included bilateral 45° posteroanterior flexion weight-bearing (Rosenberg) radiographs[7], a patellofemoral radiograph, and full-length weight-bearing alignment (hip, knee, and ankle) radiographs. The anteroposterior radiographs were made with the patient 40 inches (101.6 cm) from the x-ray beam source, and the x-ray beam was centered on the patella and aimed parallel to the tibial condyles[8]. The Rosenberg radiograph was made similarly but posterior to anterior and with the patient's knee flexed 45°, and the x-ray beam was centered at the inferior pole of the patella and directed 10° caudad[7].

Arthroscopic Assessment of Articular Cartilage Pathology

At the time of revision ACL reconstruction, the surgeon completed a study sheet previously shown to be reliable for documenting all intra-articular injuries[9]. Surgeon documentation of the degree of articular cartilage injury was

TABLE I Modified Outerbridge Scale for Grading Cartilage Lesions

Grade	Description
0	Normal articular cartilage
I	Softening of the articular cartilage
II	Fibrillation or superficial fissures of the cartilage
III	Deep fissuring of the cartilage without exposed bone
IV	Exposed bone

recorded according to the modified Outerbridge classification[10] (Table I). The highest grade of articular damage present on a particular surface was assigned to the medial femoral condyle, lateral femoral condyle, medial tibial plateau, and lateral tibial plateau.

Study Procedure and Osteoarthritis Classification Systems

The weight-bearing anteroposterior and Rosenberg radiographs were collected, stripped of identifiers, and read by three experienced orthopaedic surgeons (R.W.W., J.R.R., and J.L.C.). Each surgeon independently read the radiographs without knowledge of the other surgeons' classifications. In the cases of patients who had both anteroposterior and Rosenberg radiographs, the radiographs were read independently so that they could be compared with each other for that particular patient. Reviewers were blinded to the intraoperative arthroscopic findings during radiographic classification. The finding on each radiograph was graded according to six commonly used osteoarthritis classification systems: the Kellgren-Lawrence[11,12], International Knee Documentation Committee (IKDC) radiographic scale[13,14], Fairbank[15,16], Brandt et al.[4], Ahlbäck[17], and Jäger-Wirth[18,19]. Each classification was assigned in six separate reading sessions. Prior to reading the radiographs, the three orthopaedic surgeons reviewed the classification systems to identify discrepancies and inconsistencies with respect to each scoring method and to resolve any conflicts that existed (Table II).

Statistical Analyses

Regarding interobserver reliability, the intraclass correlation coefficient was estimated for each classification system as assessed with anteroposterior and Rosenberg radiographs (Stata Statistical Software, Release 10; StataCorp, College Station, Texas). Combined overall estimates were calculated for unilateral radiographs, bilateral radiographs, and all radiographs.

With respect to the correlation of radiographic grade with arthroscopic findings, Spearman rank correlation coefficient (Spearman rho) was similarly estimated for each classification system as assessed with use of anteroposterior and Rosenberg radiographs. In the statistical analysis, both the maximum and

TABLE II Grading Scales for the Radiographic Osteoarthritis Classification Systems*

Scale	Grade and Characteristics				
Kellgren-Lawrence	0: No JSN or reactive changes	1: Doubtful JSN, possible osteophytic lipping	2: Definite osteophytes, possible JSN	3: Moderate osteophytes, definite JSN, some sclerosis, possible bone-end deformity	4: Large osteophytes, marked JSN, severe sclerosis, definite bone ends deformity
IKDC	A: No JSN	B: >4 mm joint space; small osteophytes, slight sclerosis, or femoral condyle flattening	C: 2-4 mm joint space	D: <2 mm joint space	
Fairbank	0: Normal	1: Squaring of tibial margin	2: Flattening of femoral condyle, squaring and sclerosis of tibial margin	3: JSN, hypertrophic changes, or both	4: All of the characteristics at left, to a more severe degree
Brandt et al.	0: <25% JSN without secondary features (subchondral sclerosis, geodes, and osteophytes)	1: <25% JSN with secondary features or 25%-50% JSN without secondary features	2: 25%-50% JSN with secondary features or 50%-75% JSN without secondary features	3: 50%-75% JSN with secondary features or >75% JSN without secondary features	4: >75% JSN with secondary features
Ahlbäck	0: Normal	1: JSN† (with or without subchondral sclerosis)	2: Obliteration of joint space	3: Bone defect/loss <5 mm	4: Bone defect and/or loss 5-10 mm
Jäger-Wirth	0: No arthrosis	1: Initial arthrosis, small osteophytes, minimal JSN	2: Moderate arthrosis, about 50% JSN	3: Medium-grade arthrosis	4: Heavy arthrosis

*JSN = joint space narrowing, and IKDC = International Knee Documentation Committee. †Joint space narrowing is <3 mm of the joint space or <50% of the other compartment.

TABLE III Interobserver Reliability of the Various Radiographic Classification Systems for Osteoarthritis: Correlation with Arthroscopic Findings*

		Intraclass Correlation Coefficient		
Osteoarthritis Classification System	Radiograph	Estimate	95% Confidence Interval	
Kellgren-Lawrence	AP	0.38	0.33	0.43
	R	0.54	0.48	0.59
IKDC	AP	0.59	0.55	0.63
	R	0.66	0.62	0.71
Fairbank	AP	0.36	0.31	0.41
	R	0.44	0.38	0.50
Brandt et al.	AP	0.47	0.42	0.52
	R	0.57	0.51	0.61
Ahlbäck	AP	0.43	0.38	0.48
	R	0.67	0.62	0.71
Jäger-Wirth	AP	0.53	0.49	0.58
	R	0.51	0.46	0.57
Unilateral radiographs	AP (n = 164)	0.55	0.51	0.58
	R (n = 100)	0.54	0.49	0.58
Bilateral radiographs	AP (n = 428)	0.55	0.52	0.57
	R (n = 315)	0.65	0.63	0.67
Total radiographs	AP (n = 592)	0.55	0.53	0.56
	R (n = 415)	0.63	0.61	0.65

*AP = anteroposterior, R = Rosenberg (posteroanterior weight-bearing radiograph made with the knee in 45° of flexion), and IKDC = International Knee Documentation Committee.

the sum of the arthroscopic grades of articular cartilage degeneration (medial femoral condyle + lateral femoral condyle + medial tibial plateau + lateral tibial plateau) were used to represent the severity of osteoarthritis. Combined overall estimates were calculated for unilateral radiographs, bilateral radiographs, and all radiographs, with use of the Fisher z transformation.

Source of Funding

The MARS investigation was supported by the American Orthopaedic Society for Sports Medicine (AOSSM), Smith & Nephew (Andover, Massachusetts), the National Football League (NFL) Foundation (New York, NY), and the Musculoskeletal Transplant Foundation (Edison, New Jersey). This project was partially funded by grant 5R01-AR060846 from the National Institute of Arthritis and Musculoskeletal and Skin Diseases of the National Institutes of Health.

Results

Patient Population

From March 2006 to May 2010, the radiographs of 690 patients were received and reviewed in the MARS database. Fifty-eight patients were excluded for the following reasons: the radiographs demonstrated bilateral ACL reconstructions (forty-five), the radiographs were of poor quality (eight), or the patient had undergone other additional surgery (five). Of note, no patients were excluded on the basis of incomplete arthroscopic chondral classification, as we had complete chondral classification data for every patient in the study. This resulted in a total of 632 patients with an anteroposterior and/or a Rosenberg radiograph. In this study, 58.2% of the patients were male. The left knee was involved in 50.9% of the cases. The mean age of the patients was twenty-eight years (range, twelve to sixty-three years), and the mean body mass index (BMI) was 26.0 kg/m^2 (range, 18.7 to 47.2 kg/m^2). Five hundred and ninety-four patients had an anteroposterior weight-bearing radiograph (429 had bilateral radiographs), while 416 patients had a weight-bearing Rosenberg radiograph (315 had bilateral radiographs). Three hundred and seventy-eight patients had both an anteroposterior and a Rosenberg radiograph. Because of various clinical protocols and insurance reimbursements, not all patients had these additional radiographs. Furthermore, not all patients had bilateral anteroposterior and/or Rosenberg radiographs.

Arthroscopic Findings

Four hundred and fifteen patients had arthroscopic evidence of articular cartilage degeneration in one or both tibiofemoral compartments. Moderate to severe (grade-2, 3, or 4) articular cartilage changes were more common in the medial femoral condyle (283 [44.8%] of the 632 patients) than in the lateral femoral condyle (197 [31.2%] of the 632 patients). One hundred and sixty-two (25.6%) of the 632 patients had chondral changes in both the medial and lateral compartments. Chondral lesions involving the medial and lateral tibial plateaus were noted in 21.5% (136) and 14.4% (ninety-one) of the 632 involved knees, respectively. When chondral disease of the tibial

TABLE IV Relationship Between the Radiographic Osteoarthritis Classification Systems and the Degree of Articular Cartilage Degeneration within the Tibiofemoral Joint Identified at the Time of Arthroscopy: Interobserver Reliability and Correlation with Arthroscopic Findings*

		Correlation					
Osteoarthritis Classification	Radiograph	Sum (MFC + MTP + LFC + LTP)†	95% Confidence Interval		Maximum (MFC, MTP, LFC, or LTP)†	95% Confidence Interval	
Kellgren-Lawrence	AP	0.30	0.23	0.38	0.27	0.19	0.34
	R	0.42	0.33	0.49	0.38	0.30	0.46
IKDC	AP	0.32	0.25	0.39	0.28	0.20	0.35
	R	0.37	0.29	0.45	0.34	0.25	0.42
Fairbank	AP	0.32	0.24	0.39	0.28	0.20	0.35
	R	0.36	0.27	0.44	0.33	0.24	0.42
Brandt et al.	AP	0.31	0.23	0.38	0.28	0.20	0.35
	R	0.35	0.26	0.43	0.31	0.22	0.40
Ahlbäck	AP	0.23	0.16	0.31	0.20	0.12	0.27
	R	0.29	0.20	0.37	0.26	0.17	0.35
Jäger-Wirth	AP	0.28	0.21	0.36	0.25	0.17	0.33
	R	0.34	0.25	0.42	0.28	0.19	0.37
Unilateral radiographs	AP (n =164)	0.33	0.27	0.39	0.29	0.23	0.34
	R (n =100)	0.36	0.29	0.43	0.31	0.24	0.38
Bilateral radiographs	AP (n = 428)	0.28	0.25	0.32	0.25	0.21	0.29
	R (n = 315)	0.35	0.31	0.39	0.32	0.28	0.36
Total radiographs	AP (n = 592)	0.29	0.26	0.32	0.26	0.23	0.29
	R (n = 415)	0.36	0.32	0.39	0.32	0.28	0.35

*MFC = medial femoral condyle, LFC = lateral femoral condyle, MTP = medial tibial plateau, LTP = lateral tibial plateau, AP = anteroposterior, R = Rosenberg (posteroanterior weight-bearing radiograph made with the knee in 45° of flexion), and IKDC = International Knee Documentation Committee. †Both the sum and the maximum of the arthroscopic grades of articular cartilage were used to represent the severity of osteoarthritis.

plateau was noted, it was an isolated lesion in 14.7% of the lateral and 5.5% of the medial plateaus.

Interobserver Reliability of Radiographic Classification Systems for Osteoarthritis

Six different radiographic classification systems were used to grade tibiofemoral osteoarthritis on both anteroposterior and Rosenberg radiographs. The results in Table III show the interobserver reliability (intraclass correlation coefficients) and 95% confidence intervals (CIs) for these classification systems. The IKDC had the most favorable combinations, with moderate reliability for the anteroposterior radiographs (intraclass correlation coefficient = 0.59; 95% CI, 0.55 to 0.63) and good reliability for the Rosenberg radiographs (intraclass correlation coefficient = 0.66; 95% CI, 0.62 to 0.71), according to Bland and Altman[20]. Overall, the interobserver reliability for classifying osteoarthritis on the anteroposterior images was moderate (intraclass correlation coefficient = 0.55; 95% CI, 0.53 to 0.56), while Rosenberg images demonstrated good reliability (intraclass correlation coefficient = 0.63; 95% CI, 0.61 to 0.65). Bilateral Rosenberg images were interpreted with significantly better reliability (intraclass correlation coefficient = 0.65; 95% CI, 0.63 to 0.67) than were unilateral images (intraclass correlation coefficient = 0.54; 95% CI, 0.49 to 0.58). This difference was not seen in the groups with unilateral (intraclass correlation coefficient = 0.55; 95% CI, 0.51 to 0.58) and bilateral (intraclass correlation coefficient = 0.55; 95% CI, 0.52 to 0.57) anteroposterior radiographs.

Correlation of Radiographic Classification and Arthroscopic Findings

Table IV shows the Spearman rho values for the relationship between the radiographic grades of osteoarthritis according to the six different classification systems and the degree of articular cartilage degeneration within the tibiofemoral joint identified with use of arthroscopy. Overall, a Rosenberg radiograph had a better correlation in predicting total chondral disease (Spearman rho = 0.36; 95% CI, 0.32 to 0.39) compared with an anteroposterior radiograph (Spearman rho = 0.29; 95% CI, 0.26 to 0.32). According to Cohen[21], a small correlation is on the order of 0.10, a medium correlation is on the order of 0.30, and a large correlation is on the order of 0.50. With this standard for the magnitude of the correlation coefficient, the overall estimate of correlation between radiographic classification with use of the Rosenberg radiographs and arthroscopic grading was medium. Both for the anteroposterior and the Rosenberg radiographs, the use of bilateral images did not lead to a better correlation compared with when only unilateral images were used.

Discussion

This study confirmed our primary hypothesis that no radiographic classification system has very good (intraclass correlation coefficient of 0.8 to 1.0) interobserver reliability, but rather has moderate (0.4 to 0.6) or good (0.6 to 0.8) reliability, for classifying tibiofemoral osteoarthritis of the knee. The IKDC system, which places more emphasis on joint space narrowing than does the more traditional Kellgren-Lawrence system, had good reliability and performed superior to all but the Ahlbäck system with use of Rosenberg radiographs. Some of these osteoarthritis classification scales have been previously analyzed for reliability but, to our knowledge, never at the magnitude or comprehensiveness of the current study. In only one previous study that we are aware of was the interobserver reliability investigated with use of the IKDC. Those authors found slightly poorer intraclass correlation coefficients, of 0.46 and 0.45, for the medial and lateral tibiofemoral joints, respectively, with use of unilateral Rosenberg radiographs[22]. The Kellgren-Lawrence classification has been the most-studied classification system, with previous investigations demonstrating a wide range of interobserver reliability (0.51 to 0.89)[11,23-27]. This large range may likely be due to the various techniques used, the broad range of patient age groups, and the wide variation in the degree of osteoarthritis.

The classification scale by Brandt et al., also strongly based on joint space narrowing, demonstrated moderate interobserver reliability in this study. This classification scale breaks joint space narrowing into 25% increments, which perhaps makes agreement more difficult among interpreters. However, to our knowledge no previous studies have measured interobserver reliabilities. The Ahlbäck classification is also based on joint space narrowing. Nonetheless, there is more emphasis on bone loss, which is less applicable to a patient population with early osteoarthritis, such as in this study. Previous studies of the Ahlbäck system have also demonstrated worse interobserver reliabilities (0.11 to 0.23) compared with those in this study[17,28]. No reliability studies that we are aware of have been performed for the Fairbank and Jäger-Wirth classifications.

We believe that the present study is unique in that it is the first time a wide range of articular cartilage changes have been investigated both radiographically and arthroscopically. We found that all six radiographic classification systems had small to medium correlations with arthroscopic findings. Previous studies with weight-bearing radiographs and joint space narrowing without assigning classification scales have had similar findings[4,5,29-31]. The authors of one previous study noted similar correlation coefficients for arthroscopic disease (0.49 to 0.56) when classified with use of the systems of Kellgren-Lawrence and Brandt et al., which were not significantly different from each other[5].

Our hypothesis that Rosenberg radiographs would demonstrate higher reliability than anteroposterior radiographs was true for all classification scales except the Jäger-Wirth, for which reliability with Rosenberg radiographs was essentially equal to that with anteroposterior radiographs. Only one previous study that we are aware of has compared the interobserver reliability between anteroposterior (0.96) and Rosenberg (0.83) radiographs in regard to osteoarthritis classification scales, and this comparison was done with use of only the Ahlbäck classification[32]. However, the previous study had two observers instead of three (as in the current study) and it looked at one classification system. The majority of cases of arthroscopic disease were classified as stage 1 or 2. When looking at each individual radiographic technique in our study, bilateral Rosenberg radiographs were interpreted with significantly better reliability than were unilateral images. However, this was not the case with the anteroposterior radiograph groups.

Our study demonstrated that overall and within most classification systems, the Rosenberg radiograph was better for predicting chondral disease when compared with an anteroposterior radiograph. Previous studies have demonstrated the Rosenberg radiograph to more consistently predict osteoarthritis, given its ability to visualize the mid-flexion surface of the femoral condyles, which is a common site of degenerative wear[4,33]. However, these prior studies involved patient populations that were smaller and had more advanced osteoarthritis than the patient population in the current study.

Bilateral Rosenberg results confirmed our final hypothesis that unilateral images would demonstrate less reliability than bilateral images did, but this was not noted with anteroposterior weight-bearing radiographs. To our knowledge, the interobserver reliability of unilateral compared with bilateral knee radiographs has never been investigated. The bilateral images did not correlate any better than unilateral images did when predicting intra-articular disease, although their use did improve the reliability of osteoarthritis classification scales. To our knowledge, this is the first time that this has been demonstrated, and it is contrary to our belief that comparison with the normal knee would allow a better prediction of osteoarthritis found arthroscopically.

We believe that our study has numerous strengths and only minor weaknesses. It uses prospectively collected intra-articular data from forms and a system previously demonstrated to show reasonable agreement[9]. Our arthroscopic data set was complete, so we lost no patients because of incomplete data. The study contains a large sample size with a spectrum of articular cartilage disease in addition to normal knees. It might be assumed that a patient population of relatively young and active ACL revision patients would not demonstrate severe disease; however, as noted earlier, 70% had articular cartilage damage of grade 2 or worse. There were many patients with a disease classification that was at the most severe designation. The radiographs were standard of care at both the academic and the private-practice sites. We believe that the findings of this study have generalizability to the physician care-provider and knee patient population. Weaknesses of the study include a lack of standardized radiographic quality among sites, which could not be well controlled. Thus, some subjects were excluded because of poor image quality. Also, there was a selection bias in the cohort, as all of the patients were undergoing revision ACL reconstruction. Patient insurance restrictions limited our ability to make both anteroposterior and Rosenberg radiographs for all subjects. We overcame this limitation by having a large sample size in our study.

Conclusions

Our study is novel in that we not only investigated the interobserver reliabilities of the various osteoarthritis classification systems commonly used for the knee, but we also analyzed their correlation with cartilage status assessed arthroscopically across a wide range of articular disease. We demonstrated that the commonly used radiographic classification systems have moderate to good interobserver reliability and medium correlation with arthroscopic findings. The IKDC classification assessed with use of Rosenberg radiographs had the most favorable combination of good reliability and medium correlation. Further, we clearly demonstrated that, overall, Rosenberg radiographs had higher interobserver reliability and higher correlation with arthroscopic findings of chondral disease than did anteroposterior radiographs. ■

NOTE: This project received funding from the AOSSM, Smith & Nephew, the NFL Foundation, and the Musculoskeletal Transplant Foundation. Also, it was partially funded by grant 5R01-AR060846 from the National Institute of Arthritis and Musculoskeletal and Skin Diseases of the National Institutes of Health.

Rick W. Wright, MD
MARS Group Principal Investigator,
Department of Orthopaedic Surgery,
Washington University School of Medicine,
660 South Euclid Avenue,
Campus Box 8233,
St. Louis, MO 63110.
E-mail address: Wright@wudosis.wustl.edu

References

1. Felson DT. Clinical practice. Osteoarthritis of the knee. N Engl J Med. 2006 Feb 23;354(8):841-8.

2. American College of Rheumatology Subcommittee on Osteoarthritis Guidelines. Recommendations for the medical management of osteoarthritis of the hip and knee: 2000 update. Arthritis Rheum. 2000 Sep;43(9):1905-15.

3. Hochberg MC, Altman RD, Brandt KD, Clark BM, Dieppe PA, Griffin MR, Moskowitz RW, Schnitzer TJ; American College of Rheumatology. Guidelines for the medical management of osteoarthritis. Part II. Osteoarthritis of the knee. Arthritis Rheum. 1995 Nov;38(11):1541-6.

4. Brandt KD, Fife RS, Braunstein EM, Katz B. Radiographic grading of the severity of knee osteoarthritis: relation of the Kellgren and Lawrence grade to a grade based on joint space narrowing, and correlation with arthroscopic evidence of articular cartilage degeneration. Arthritis Rheum. 1991 Nov;34(11):1381-6.

5. Kijowski R, Blankenbaker D, Stanton P, Fine J, De Smet A. Arthroscopic validation of radiographic grading scales of osteoarthritis of the tibiofemoral joint. AJR Am J Roentgenol. 2006 Sep;187(3):794-9.

6. Wright RW, Huston LJ, Spindler KP, Dunn WR, Haas AK, Allen CR, Cooper DE, DeBerardino TM, Lantz BB, Mann BJ, Stuart MJ; MARS Group. Descriptive epidemiology of the Multicenter ACL Revision Study (MARS) cohort. Am J Sports Med. 2010 Oct;38(10):1979-86.

7. Rosenberg TD, Paulos LE, Parker RD, Coward DB, Scott SM. The forty-five-degree posteroanterior flexion weight-bearing radiograph of the knee. J Bone Joint Surg Am. 1988 Dec;70(10):1479-83.

8. Ahlbäck S. Osteoarthrosis of the knee. A radiographic investigation. Acta Radiol Diagn (Stockh). 1968;Suppl 2777-72.

9. Marx RG, Connor J, Lyman S, Amendola A, Andrish JT, Kaeding C, McCarty EC, Parker RD, Wright RW, Spindler KP; Multicenter Orthopaedic Outcomes Network. Multirater agreement of arthroscopic grading of knee articular cartilage. Am J Sports Med. 2005 Nov;33(11):1654-7. Epub 2005 Aug 10.

10. Curl WW, Krome J, Gordon ES, Rushing J, Smith BP, Poehling GG. Cartilage injuries: a review of 31,516 knee arthroscopies. Arthroscopy. 1997 Aug;13(4):456-60.

11. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. Ann Rheum Dis. 1957 Dec;16(4):494-502.

12. Schiphof D, Boers M, Bierma-Zeinstra SM. Differences in descriptions of Kellgren and Lawrence grades of knee osteoarthritis. Ann Rheum Dis. 2008 Jul;67(7):1034-6. Epub 2008 Jan 15.

13. Irrgang JJ, Anderson AF, Boland AL, Harner CD, Neyret P, Richmond JC, Shelbourne KD; International Knee Documentation Committee. Responsiveness of the International Knee Documentation Committee Subjective Knee Form. Am J Sports Med. 2006 Oct;34(10):1567-73. Epub 2006 Jul 26.

14. Hefti F, Müller W, Jakob RP, Stäubli HU. Evaluation of knee ligament injuries with the IKDC form. Knee Surg Sports Traumatol Arthrosc. 1993;1(3-4):226-34.

15. Fairbank TJ. Knee joint changes after meniscectomy. J Bone Joint Surg Br. 1948 Nov;30(4):664-70.

16. Tapper EM, Hoover NW. Late results after meniscectomy. J Bone Joint Surg Am. 1969 Apr;51(3):517-26 passim.

17. Galli M, De Santis V, Tafuro L. Reliability of the Ahlbäck classification of knee osteoarthritis. Osteoarthritis Cartilage. 2003 Aug;11(8):580-4.

18. Scheller G, Sobau C, Bülow JU. Arthroscopic partial lateral meniscectomy in an otherwise normal knee: Clinical, functional, and radiographic results of a long-term follow-up study. Arthroscopy. 2001 Nov-Dec;17(9):946-52.

19. Schroeder-Boersch H, Töws P, Jani L. [Reproducibility of radiologic markers of osteoarthritis. Evaluating a score for radiologic classification of degenerative changes in osteoarthritis of the knee joint] [German]. Z Orthop Ihre Grenzgeb. 1998 Jul-Aug;136(4):293-7.

20. Bland JM, Altman DG. A note on the use of the intraclass correlation coefficient in the evaluation of agreement between two methods of measurement. Comput Biol Med. 1990;20(5):337-40.

21. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillside: Lawrence Erlbaum; 1998.

22. Mehta VM, Paxton LW, Fornalski SX, Csintalan RP, Fithian DC. Reliability of the international knee documentation committee radiographic grading system. Am J Sports Med. 2007 Jun;35(6):933-5. Epub 2007 Mar 22.

23. Felson DT, Naimark A, Anderson J, Kazis L, Castelli W, Meenan RF. The prevalence of knee osteoarthritis in the elderly. The Framingham Osteoarthritis Study. Arthritis Rheum. 1987 Aug;30(8):914-8.

24. Bagge E, Bjelle A, Valkenburg HA, Svanborg A. Prevalence of radiographic osteoarthritis in two elderly European populations. Rheumatol Int. 1992;12(1):33-8.

25. Scott WW Jr, Lethbridge-Cejku M, Reichle R, Wigley FM, Tobin JD, Hochberg MC. Reliability of grading scales for individual radiographic features of osteoarthritis of the knee. The Baltimore longitudinal study of aging atlas of knee osteoarthritis. Invest Radiol. 1993 Jun;28(6):497-501.

26. Kessler S, Guenther KP, Puhl W. Scoring prevalence and severity in gonarthritis: the suitability of the Kellgren & Lawrence scale. Clin Rheumatol. 1998;17(3):205-9.

27. Gossec L, Jordan JM, Mazzuca SA, Lam MA, Suarez-Almazor ME, Renner JB, Lopez-Olivo MA, Hawker G, Dougados M, Maillefert JF; OARSI-OMERACT task force "total articular replacement as outcome measure in OA". Comparative evaluation of three semi-quantitative radiographic grading techniques for knee osteoarthritis in terms of validity and reproducibility in 1759 X-rays: report of the OARSI-OMERACT task force. Osteoarthritis Cartilage. 2008 Jul;16(7):742-8. Epub 2008 Apr 15.

28. Weidow J, Cederlund CG, Ranstam J, Kärrholm J. Ahlbäck grading of osteoarthritis of the knee: poor reproducibility and validity based on visual inspection of the joint. Acta Orthop. 2006 Apr;77(2):262-6.

29. Blackburn WD Jr, Bernreuter WK, Rominger M, Loose LL. Arthroscopic evaluation of knee articular cartilage: a comparison with plain radiographs and magnetic resonance imaging. J Rheumatol. 1994 Apr;21(4):675-9.

30. Fife RS, Brandt KD, Braunstein EM, Katz BP, Shelbourne KD, Kalasinski LA, Ryan S. Relationship between arthroscopic evidence of cartilage damage and radiographic evidence of joint space narrowing in early osteoarthritis of the knee. Arthritis Rheum. 1991 Apr;34(4):377-82.

31. Wright RW, Boyce RH, Michener T, Shyr Y, McCarty EC, Spindler KP. Radiographs are not useful in detecting arthroscopically confirmed mild chondral damage. Clin Orthop Relat Res. 2006 Jan;442(442):245-51.

32. Pires e Albuquerque R, Carvalho AC, Giordano V, Djahjah MC, do Amaral NP. [Comparative study between different radiographic plans in knee osteoarthritis] [Portuguese]. Acta Reumatol Port. 2009 Apr-Jun;34(2):380-7.

33. Dervin GF, Feibel RJ, Rody K, Grabowski J. 3-Foot standing AP versus 45 degrees PA radiograph for osteoarthritis of the knee. Clin J Sport Med. 2001 Jan;11(1):10-6.