

Ahlbäck grading of osteoarthritis of the knee

Poor reproducibility and validity based on visual inspection of the joint

Jonas Weidow¹, Claes-Göran Cederlund², Jonas Ranstam³ and Johan Kärrholm⁴

^{1,4}Department of Orthopedics, Sahlgrenska University Hospital, SE-413 45 Göteborg, ²Department of Radiology, Central Hospital Halmstad, SE-301 85 Halmstad, Sweden, ³Hialøs International ApS, Grønningen 17, DK-1270 Copenhagen, Denmark
Correspondence JW: jonas@weidow.se
Submitted 05-01-26. Accepted 05-05-31

Background Staging of osteoarthritis (OA) of the knee is commonly based on the Ahlbäck classification. Its value has been questioned, however. We therefore evaluated the reproducibility and validity of this classification of knee osteoarthritis.

Patients and methods 48 patients (48 knees) (medial OA: n = 30; lateral OA: n = 8) operated with total knee prostheses were studied. Weight-bearing radiographs were evaluated twice by 4 observers. Presence of bone attrition on radiographs was compared with observations of the resected parts of the distal femur and proximal tibia.

Results When the same observer classified the radiographs twice according to Ahlbäck, the repeatability was fair in both medial OA (kappa values = 0.15–0.65) and lateral OA (0.59–0.76), and between different observers it was poor (kappa: 0.1).

Comparison between radiographic classification and classification based on visual inspection of the bone pieces removed during arthroplasty revealed an acceptable sensitivity in both medial (67–95%) and lateral OA (43–86%), but the specificity was low (medial: 11–67%; lateral: 25–75%).

Interpretation The main problem with the Ahlbäck classification was that a joint space could often be seen radiographically despite the presence of bone attrition on the preparations. According to our study, conventional radiographs do not give sufficient information for correct grading.

In 1968, Ahlbäck showed the advantage of standing radiographic examinations to demonstrate the

presence of knee OA (Ahlbäck 1968). Since then, a number of studies have focused on the problem of visualizing the smallest width of the knee joint when taking weight-bearing radiographs (Leach et al. 1970, Siu et al. 1991, Buckland-Wright et al. 1994, 1995, Ravaut et al. 1996). Ahlbäck used 3 alternatives to regard narrowing of the articular space as an established fact. The joint space was classified as narrowed when it was (1) “narrower than half the width of the articular space in the other articulation of the same knee or the same articulation of the other knee”, (2) when “it decreased between non-weight-bearing and weight-bearing positions”, and/or 3) when “it was narrower than 3 mm”. It has since been shown that joint space width less than 3 mm reflects cartilage defects on MR (Boegard et al. 1997). Ahlbäck also recorded 4 types of bone changes: sclerosis, cysts, attrition and osteophytes.

In 1980, the so-called Ahlbäck classification was described as follows (Ahlbäck and Rydberg 1980). Grade 1: narrowing of the articular space; grade 2: obliteration or almost obliteration of the articular space; grade 3: bone attrition less than 5 mm; grade 4: bone attrition between 5 and 15 mm, and grade 5: bone attrition greater than 15 mm. The authors pointed out that this scoring system had not been evaluated on cases with lateral OA.

In early OA, radiographic observations have been correlated to arthroscopic findings (DeHaven and Collins 1975, Lysholm et al. 1987, Brandt et al. 1991), emphasizing the insensitivity of the radiographs for detecting early articular cartilage loss, but no such comparison between radiography

and morphological changes has been described for more advanced disease. The Ahlbäck classification has recently been reported to have poor reliability (Galli et al. 2003).

We evaluated the reproducibility of the Ahlbäck classification based on routine weight-bearing radiographs, and its validity in the classification of bone attrition by comparison with osteocartilagenous preparations removed during total knee replacement.

Patients and methods

48 patients (48 knees) were operated with tricompartmental total knee prosthesis at the Department of Orthopedics, Halmstad Hospital, between 1998 and 2002. There were 30 knees with medial OA (18 women, median age 73 (55–85) years; 12 men, median age 71 (59–84) years) and 18 knees (18 women) with lateral OA (median age 73.5 (59–86) years).

Weight-bearing radiographs were taken for all patients, median 4 (1–29) days before the operation. The patients were standing with equal weight on both legs, and with their knees semiflexed according to the routines at the Department of Radiology. The patients stood with their toes roughly in a vertical line from the X-ray film. They flexed their knees until they reached the film, corresponding to about 15–20 degrees of flexion.

All patients received the Freeman Samuelson total knee prosthesis (Sulzer, Switzerland). Standard instrumentation was used with an intramedullary guide on the femoral side and an extramedullary guide on the tibial side. In all cases, the bone cut could be placed a few millimetres below the level of the most worn part of the medial or lateral compartment. The tibial part of the joint could always be obtained in one piece. The cartilage/bone pieces were marked for orientation and were stored at –70°C.

Ahlbäck classification

In all 48 cases, the preoperative radiographs were graded twice according to the Ahlbäck classification by 4 observers. 2 observers (observers 1 and 2) were orthopedic surgeons who frequently performed knee arthroplasty and who were familiar

with interpretation of radiographs of knee OA. The 2 other observers (observers 3 and 4) were experienced radiologists with orthopedic radiology as their main speciality. These gradings were done with a minimum interval of 2 weeks. When the degree of OA was determined, the presence of bone attrition took priority over joint space width. Thus, when there was obvious bone attrition on the radiographs but a visible joint space in the same compartment, this case was classified as stage 3 or more (rather than stage 1). The observers recorded separately whether they believed that there was any bone attrition on the femoral side and the tibial side.

Morphological analysis of OA

Analysis of joint area morphology was done after the operation. When the pieces had thawed, the medial and lateral parts of the femoral and tibial joint area were analyzed separately. Visual inspection was not expected to enable differentiation of Ahlbäck grades 3 and 4. We therefore condensed the Ahlbäck classification to 3 grades: cartilage destruction without visible bone (“Ahlbäck grade 1”), visible bone with no or minimum bone attrition (“grade 2”), and obvious bone attrition (“grade 3–4”).

Statistics

The concordance between the different observations and observers regarding the radiographic classification was evaluated using kappa statistics. In addition, the sensitivity and specificity were determined using visual inspection of the resected joint as reference. Cases with medial and lateral OA were studied separately.

Results (Table)

Ahlbäck classification

When the same observer classified the radiographs twice, the reproducibility varied from a high to an acceptable variation (kappa for medial OA: 0.15–0.65, mean = 0.46; kappa for lateral OA: 0.59–0.76, mean = 0.67). The interobserver variability was, however, consistently low (combined kappa values for all 4 observers: medial, 0.11; lateral, 0.12). The sensitivity (comparison between

The two values for sensitivity and specificity refer to the first and second observations, respectively

Observer	1	2	3	4
<i>Medial OA</i>				
Ahlbäck grading (kappa value)				
Intraobserver variability	0.45	0.65	0.59	0.15
Interobserver variability:				
Individual	0.05	0.20	0.05	0.17
All combined		0.11		
Sensitivity (%)	86 / 95	71 / 67	– / –	– / –
Specificity (%)	22 / 11	56 / 67	– / –	– / –
Attrition tibia				
Sensitivity (%)	82 / 94	59 / 53	94 / 94	88 / 94
Specificity (%)	15 / 8	77 / 77	15 / 23	31 / 23
Attrition femur				
Sensitivity (%)	74 / 74	63 / 68	95 / –	– / –
Specificity (%)	64 / 46	64 / 73	36 / –	– / –
<i>Lateral OA</i>				
Ahlbäck grading (kappa value)				
Intraobserver variability	0.68	0.76	0.65	0.59
Interobserver variability:				
Individual	0.10	–0.01	0.21	–0.01
All combined		0.12		
Sensitivity (%)	64 / 79	43 / 50	– / –	86 / 86
Specificity (%)	50 / 50	75 / 50	– / –	25 / 25
Attrition tibia				
Sensitivity (%)	69 / 77	31 / 38	– / –	85 / 92
Specificity (%)	60 / 40	80 / 60	– / –	40 / 40
Attrition femur				
Sensitivity (%)	25 / 25	– / 25	50 / 38	62 / 88
Specificity (%)	40 / 60	– / 40	30 / 30	20 / 30
– could not be evaluated				

radiographs and visual inspection of resected bone ends) was acceptable both in medial OA (67–95%) and in lateral OA (50–86%). For most comparisons, the specificity was lower (medial: 11–67%; lateral: 25–75%).

Bone attrition (radiographic diagnosis vs. visual inspection of resected bone)

In medial OA, the sensitivity of detection of any tibial and femoral attrition varied between 53–94% and 63–95%, respectively. The corresponding variation of specificity for the different combinations of observers varied between 8–77% (tibia) and 36–73% (femur).

In lateral OA, the sensitivity of detection of tibial and femoral attrition varied between 38–92% and 25–87% (only one observer had values greater than 50%). The specificity varied between 40–80% (tibia) and 20–60% (femur).

Discussion

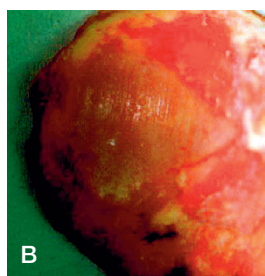
The Ahlbäck classification is probably the most frequently used system for classification of knee osteoarthritis in stages that are relevant for treatment decisions. Some problems with the Ahlbäck classification have been recognized in clinical practice, and support a study (Galli et al. 2003) in which the intraobserver reliability and the interobserver reproducibility showed low agreement, especially when Ahlbäck grade 5 was excluded. In our evaluation, the interobserver error values and the specificity were astonishingly poor.

The staging used today in the Ahlbäck classification was not described until 1980 (Ahlbäck and Rydberg 1980). In that study of 359 patients with medial OA, the distribution between the 5 stages varied between 3% and 38%, a scatter which Ahlbäck regarded as acceptable to test the value of his classification system.

There were no cases with lateral OA, and Ahlbäck grading of this subgroup has not been studied in the subsequent literature.

According to radiographs presented in his thesis in 1968, Ahlbäck was aware of cases with obvious bone attrition despite the presence of a visible joint opening, but did not comment on this problem. In the later publication, this inconsistency was still not addressed. In our experience, the correct use of joint space width in the Ahlbäck classification is difficult. When the joint space has not been obliterated, the tibial or femoral edge facing the joint line is comparatively smooth and it is often uncertain whether there has been bone destruction or not, sometimes resulting in an incorrect choice between grades 1 and 3 (Figure).

Heightened precision and accuracy of radiographic classification of knee OA would be beneficial in facilitation of preoperative planning. It would reduce the degree of uncertainty concern-



Right knee: lateral arthrosis. Remaining joint space on standing radiograph (A). Photograph (B) of the lateral tibial condyle in the same knee after TKR replacement shows bone attrition.

ing choice of technique and implant, and also the surgical skills necessary to perform the operation. During preparations for this study, we repeatedly tried to define radiographic changes that indicated the presence of bone destruction, but failed; nor was there any trend toward improvement in the repeated classification done in the intraobserver analysis. This supports our view that the information available on the radiographs is insufficient. Improvements in radiographic classification of knee OA must be sought through changes to the examination technique, or in methods other than conventional radiography.

Determination of femoral attrition was less reliable in lateral OA than in medial OA. This was the only obvious difference between these two variations of the disease in terms of reproducibility of the radiographic classification. This finding could be explained by a more posterior location of the wear on the lateral condyle (Boegard et al. 1997). This part of the femur will not articulate against the tibia in the position usually used to expose weight-bearing radiographs, but requires exposure at further flexion of the knee joint, i.e. the tunnel view (Resnick and Vint 1980).

In summary, the Ahlbäck classification has variable reproducibility and validity. As expected, the same observer could reproduce the Ahlbäck classification more precisely on different occasions than two different observers. When evaluating the radiographs, there are two major problems. One is to determine whether a visible joint space repre-

sents remaining cartilage. The other is to determine whether there is bone attrition.

Author contributions

JW main author and study design. C-GC radiographic evaluation. JR statistics. JK author and study design.

No competing interests declared.

Ahlbäck S. Osteoarthritis of the knee. A radiographic investigation. *Acta Radiol [Diagn] (Stockh) (Suppl 277)* 1968; 7-72.

Ahlbäck S, Rydberg J. Röntgenologisk klassifikation och undersökningsteknik vid gonartros. *Läkartidningen* 1980; 77 (22): 2091-6.

Boegard T, Rudling O, Petersson I F, Sanfridsson J, Saxne T, Svensson B, Jonsson K. Postero-anterior radiogram of the knee in weight-bearing and semiflexion. Comparison with MR imaging. *Acta Radiol* 1997; 38 (6): 1063-70.

Brandt K D, Fife R S, Braunstein E M, 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; 34 (11): 1381-6.

Buckland-Wright J C, Macfarlane D G, Jasani M K, Lynch J A. Quantitative microfocal radiographic assessment of osteoarthritis of the knee from weight bearing tunnel and semiflexed standing views. *J Rheumatol* 1994; 21 (9): 1734-41.

Buckland-Wright J C, Macfarlane D G, Lynch J A, Jasani M K, Bradshaw C R. Joint space width measures cartilage thickness in osteoarthritis of the knee: high resolution plain film and double contrast macroradiographic investigation. *Ann Rheum Dis* 1995; 54 (4): 263-8.

DeHaven K E, Collins H R. Diagnosis of internal derangements of the knee. The role of arthroscopy. *J Bone Joint Surg (Am)* 1975; 57 (6): 802-10.

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

Leach R E, Gregg T, Siber F J. Weight-bearing radiography in osteoarthritis of the knee. *Radiology* 1970; 97 (2): 265-8.

Lysholm J, Hamberg P, Gillquist J. The correlation between osteoarthritis as seen on radiographs and on arthroscopy. *Arthroscopy* 1987; 3 (3): 161-5.

Ravaud P, Auleley G R, Chastang C, Rousselin B, Paolozzi L, Amor B, Dougados M. Knee joint space width measurement: an experimental study of the influence of radiographic procedure and joint positioning. *Br J Rheumatol* 1996; 35 (8): 761-6.

Resnick D, Vint V. The "Tunnel" view in assessment of cartilage loss in osteoarthritis of the knee. *Radiology* 1980; 137 (2): 547-8.

Siu D, Cooke T D, Broekhoven L D, Lam M, Fisher B, Saunders G, Challis T W. A standardized technique for lower limb radiography. Practice, applications, and error analysis. *Invest Radiol* 1991; 26 (1): 71-7.