ORIGINAL ARTICLE

Andréa S. Doria
Cláudio C. de Castro
Maria Helena B. Kiss
Renato A. Sernik
Luís F. Vitule
Carlos H. M. Silva
Cristiano A. F. Zerbini
Paula R. Arantes
Leandro Lucato
Marco A. N. Germano
Giovanni G. Cerri

Inter- and intrareader variability in the interpretation of two radiographic classification systems for juvenile rheumatoid arthritis

A.S. Doria (✉) · C.C. de Castro
R.A. Sernik · L.F. Vitule · P.R. Arantes
L. Lucato · M.A.N. Germano · G.G. Cerri
Department of Diagnostic Imaging,
Heart Institute (InCor) and Radiology
Institute (InRad), Hospital das Clínicas da
Universidade de São Paulo,
São Paulo, Brazil
E-mail: andrea.doria@sickkids.ca

M.H.B. Kiss
Division of Rheumatology,
Children's Institute, Hospital das Clínicas
da Universidade de São Paulo,
São Paulo, Brazil

C.H.M. Silva
Division of Rheumatology,
Department of Pediatrics,
Faculdade de Medicina da Universidade
Federal de Uberlândia MG,
Uberlândia, Brazil

C.A.F. Zerbini
Division of Rheumatology,
Hospital Heliópolis, São Paulo, Brazil

Abstract *Objective*: To evaluate the inter- and intrareader variability for interpretation of a modified Larsen's radiographic classification system for juvenile rheumatoid arthritis (JRA) focused on osteochondral lesions and a conventional Larsen's classification system, compared to a reference MR scoring system of corresponding images. *Materials and methods*: Seventy-five radiographs of 60 children with JRA, performed within a short interval of time from the MR examinations, were independently evaluated by three experienced radiologists, three diagnostic imaging residents and three rheumatologists, in two separate sessions, according to the two different classification methods, blinded to the corresponding MR images. *Results*: The inter- and intrareader concordance rates between the two radiographic classification systems and the MR-related radiographs were respectively poor and poor/moderate. The interobserver range of weighted kappa values for the conventional and the modified Larsen's system respectively was 0.25–0.37 vs 0.19–0.39 for radiologists, 0.25–0.37 vs 0.18–0.30 for residents and 0.19–0.51 vs 0.17–0.29 for rheumatologists. The intrareader rate ranged from 0.17–0.55 for radiologists, 0.2–0.56 for residents, and 0.14–0.59 for rheumatologists. *Conclusion*: Although the proposal of a new radiographic classification system for JRA focused on osteochondral abnormalities sounds promising, the low inter- and intrareader concordance rates with an MR-related radiographic system makes the clinical applicability of such a radiographic system less suitable.

Keywords Classification systems · Scoring systems · Juvenile rheumatoid arthritis · Radiography · MR imaging · Inter- and intrareader variability

Introduction

Plain radiographs are widely used to monitor the anatomic progression of juvenile rheumatoid arthritis (JRA), which is characterized by the extension of hypertrophic synovium (pannus) from the peripheral synovial recesses centripetally into the joint producing marginal and subchondral lesions [1, 2]. The Steinbrocker's [3], Sharp's [4], and Larsen's [5] rheumatoid arthritis radiographic classification systems, as well as

the Petterson's JRA classification system [6], are able to score the degree of severity of radiographic findings, although subtle signs of osteochondral progression may be missed. The sensitivity of Larsen's stages 0–V for measurement of JRA progression is low because of the wide variability on grades, especially II and III, with several pathological changes developing between two grades, without the joint qualifying for an osteochondral progression from one grade to the next one. The fact that the conventional Larsen's scoring system does not take into consideration minimal osteochondral changes may cause underestimation of the need for therapy.

A well-conducted treatment for JRA is associated with a reduced rate of radiological progression [7]. Therefore, a rigorous evaluation of osteocartilaginous abnormalities (bone erosions and subchondral changes), which are responsible for most disability in JRA, is of outstanding value for follow-up and evaluation of clinical outcomes post-therapy. The rationale for creating a new scoring system focused on the detection of minimal osteochondral abnormalities was based on the fact that such a scoring system would be able to improve the measurement precision of the conventional Larsen's score by means of increasing the number of distinct levels enumerated by the scale (fineness of specification). Such a scoring system would enable the measurement of smaller longitudinal changes on the dimension of interest in an individual providing a better evaluative index for assessment of early changes in the disease.

The aim of this study was to investigate the possibility of improvement in the rate of identification of bone erosions and subchondral lesions among blinded readers with different levels of radiologic experience through the inclusion of specific scores for radiographic assessment of osteochondral abnormalities into the conventional Larsen's classification system. Evaluation of the inter- and intrareader variability for interpretation of the conventional and the modified Larsen's system could validate the use of this new radiographic classification system for follow-up of children with JRA.

Materials and methods

Patients

Patients with the diagnosis of JRA according to the revised criteria of the American College of Rheumatology [8] and admitted to the outpatient clinic of Pediatric Rheumatology at the Children's Institute of Universidade de São Paulo were enrolled in a prospective study over 4.5 years (February 1995 to August 1999). Requests for radiographic assessment were based on clinical indication. Requests for MRI assessment in patients who had history of previous or current knee involvement were based on clinical or research purposes. Clinical indications for an MRI examination included: (1) persistent clinically or laboratory active disease despite the appropriate use of conventional therapy; (2) pre-biopsy procedure assessment; (3) follow-up of asymptomatic patients for

1 year upon termination of therapy. Sedation was warranted only for patients with clinical indications (1) and (2) for an MRI examination. Exclusion criteria were: (1) history of a previous surgery in the arthritic knee; (2) need of sedation for patients undergoing the examination for exclusive research purposes; (3) impossibility to undergo an MR examination within a maximum interval of time of 3 months from a previous plain film of the knees. Forty-six girls and 14 boys, aged 4.8–20.1 years (mean, 11.9 years), agreed to participate in the study with written consent from the parents. There were 3 patients (4%) ≤ 5 years old, 13 children (17.3%) > 5 and ≤ 9 years old, and 29 (38.7%) > 9 and ≤ 12 years old, and 30 (40%) > 12 years old. According to the type of onset, 20 (26.7%) patients had polyarticular onset of disease, 21 (28%), pauciarticular and 34 (45.3%), systemic onset of JRA. The duration of the arthritic process prior to the radiological assessment ranged from 0.3 to 14.6 years (mean 6 years). The knees imaged showed clinical evidence of arthritis at variables stages of the disease. The research protocol was approved by the Ethics Board Review of our university.

Methods

Seventy-five anteroposterior non-weight-bearing position radiographs and 75 MR examinations of the knee of 60 children and adolescents were reviewed. Of those patients, 15 underwent 2 radiographies of the same knee in different occasions (interval range between both radiographic examinations: 0.8–2.4 years; mean: 1.3 years). The radiographs of the knee were obtained within 3 months of interval (59.1% within 1 month; 28.2% within 2 months and 12.7% within 3 months) from the corresponding MR study. In patients with bilateral knee involvement, the knee with more severe clinical involvement was the one chosen for examination.

MR images were performed on a 1.5-T magneton unit (Gyroscan, Philips Medical Systems, Best, The Netherlands) using a dedicated knee coil. The MR pulse sequences used for this study were part of a clinical research protocol used for evaluation of disease progression [9]. Contiguous 4-mm-thick coronal T1-weighted spin-echo (SE) and T2-weighted gradient-echo (FFE) images (TR/TE: 525/20 and 40/14 ms) were obtained with 18 cm of field of view, flip angles of 90° and 10°, respectively in 3:18 and 1:32 min (scan time).

Our protocol investigated the degree of inter- and intrareader variability between two radiographic classification systems for JRA: a well-established one (Larsen's) (Table 1) [5] and another one based on modifications of the Larsen's system developed at the Division of Diagnostic Imaging of Heart Institute (Incor) (A.S.D.), Hospital das Clínicas da Universidade de São Paulo (Table 1).

Guidelines for categorization of readers

Figure 1 contains three flowcharts.The same plain films underwent blinded multireader (nine readers) interpretation (flowchart 1) and were retrospectively evaluated by two unblinded radiologists in comparison to the corresponding MR images (flowchart 2). The unblinded radiologists determined which features from the two grading systems were able to be identified in the plain radiographs upon retrospective comparison with the corresponding MR images, which played the role of "gold-standard" models. For example, if a small lesion were depicted on an MR image in the distal medial femoral condyle, each of the reviewers would independently try to identify a lesion on this topography of the knee on the corresponding plain film. If none of the readers were able to identify that lesion on the plain film, they would consider it as undepictable from the radiographic point of view and would classify that plain film accordingly based on the scoring system of the two radiographic

Table 1 Radiographic classification systems of the degree of joint destruction.[a] Larsen's method [5] and modified Larsen's method

(A) Larsen's method
Grade 0: Normal conditions. Abnormalities not related to arthritis may be present.
Grade I: Slight abnormality. Phase of indefinite arthritis. One or more of the following lesions are present:
  periarticular soft tissue swelling, periarticular osteoporosis and slight joint space narrowing.
Grade II: Definite early abnormality. Erosions (not obligatory in the knees) and joint space narrowing.
Grade III: Medium destructive abnormality. Erosions (obligatory in all joints) and joint space narrowing.
Grade IV: Severe destructive abnormality. Erosion, joint space narrowing and bone deformation (in the knees).
Grade V: Mutilating abnormality. Disappearing of the original articular surfaces and bone deformation.
  Bony ankylosis not considered in the grading.

(B) Modified Larsen's method for the knee joints
Grade 0: Larsen's scoring method.
Grade I: Larsen's scoring method.
Grade IIa: Definite early abnormality. Presence of definite marginal erosions in a single location (only one of condyles)
  or indefinite (doubtful) erosions in more than one location (two or more condyles).
Grade IIb: Definite early abnormality. Presence of definite marginal erosions in two or more locations.
Grade IIIa: Medium destructive abnormality. Presence of at least one indefinite (doubtful) subchondral lesion.
Grade IIIb: Medium destructive abnormality. Presence of at least one definite subchondral lesion
  or at least 2 indefinite subchondral lesions.
Grade IV: Larsen's scoring method.
Grade V: Larsen's scoring method.

[a]Number of marginal erosions and subchondral lesions: according to the number of femoral (2) and tibial (2) condyles involved: lateral and medial distal femoral condyles, lateral and medial proximal tibial condyles

grading systems. Discrepancies in opinion were solved by consensus. The "gold-standard-related" list of 75 cases would be utilized in the future as a reference for comparison with the scoring of radiographic reports from the nine blinded readers.

Although not specific for depiction of osteochondral abnormalities, the T2-weighted FFE images helped identify intraosseous abnormalities and peripheral erosions. However, the direct radiographic-MR imaging correlation was performed using T1-weighted SE images, which provided a better anatomic correlation.

Guidelines for categorization of radiographic evaluation
for both classification systems

The readers were expected to grade each set of 75 plain films within an interval of 75 minutes (1 min/radiograph Fig. 1 containing three flowcharts). The plain films were randomly ordered for both evaluations, and all radiographs were viewed with similar light boxes under identical conditions. The radiologists and residents who participated in the study had no previous knowledge about the Larsen's system. One pediatric rheumatologist had some experience in evaluation of radiographs according to the conventional Larsen's method, but the other two rheumatologists had no prior radiologic experience. On the two reading occasions, the readers received a copy of the classification system to be used at that time (Table 1), illustrations of the descriptions of the stages according to Larsen's publication [5], a form containing the rules for radiographic reading (time framework) and a detailed explanation using ten tutorial plain films, which included an example of each grade on the conventional and modified Larsen's system.

The modified Larsen's system received the inclusion of parameters for characterization of marginal bone erosions on grade II (IIa and IIb), as shown in Figs. 2 and 3, and subchondral lesions (IIIa and IIIb), as shown in Figs. 4, 5 and 6 and Table 1 on the medial and lateral compartments of both femora and tibiae (Fig. 1, flowchart 1).

Osseous and cartilaginous lesions resultant from progression of the disease included subchondral erosions and cysts, as well as marginal erosions. On MR imagi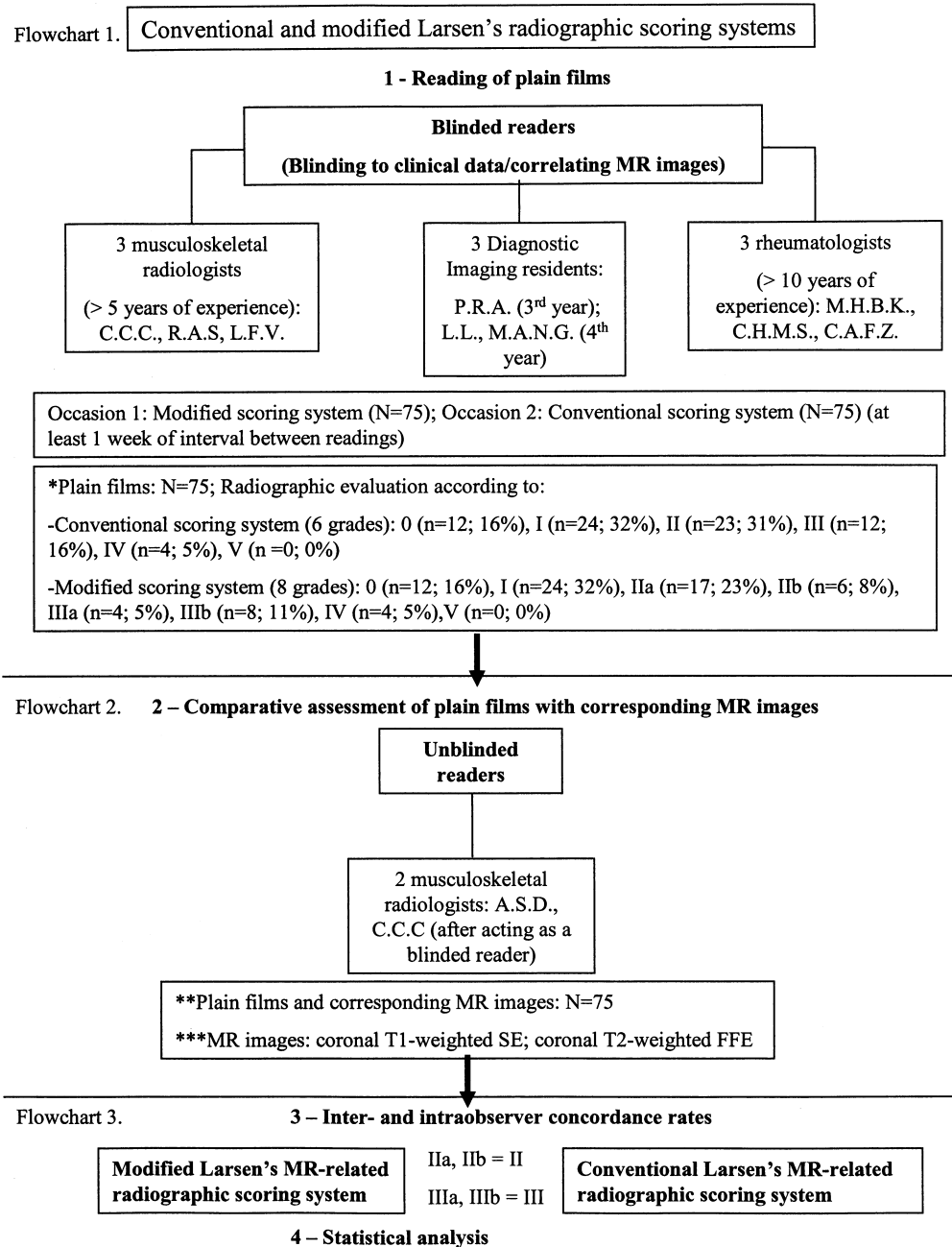ng, subchondral erosions appear as subcortical areas of abnormal signal intensity in the marrow adjacent to the bone margin or beneath the chondral surface on T1-weighted images [10]. The content of an erosion can present with either low-signal intensity on T1-weighted images and bright signal intensity on T2-weighted images or intermediate signal on both T1- and T2- weighted images [10]. Subchondral cysts are intraosseous round structures typically situated beneath the cartilaginous-bone interface presenting with low-intensity signal on T1-weighted images and high-intensity signal on T2- or T2*- weighted images [11]. Marginal erosions represent focal cortical defects situated at the edges of the femoral and tibial condyles produced by the effect of invasive pannus [10]. Only marginal erosions situated along the external margins of both femoral and tibial condyles were considered for evaluation. For the effect of comparison with the plain films, no differentiation was made between subchondral erosions and cysts, both being considered as osteochondral lesions.

According to the modified system, the maximum number of erosions and cysts per joint was four, including one lesion in each of the distal femoral condyles (lateral and medial) and proximal tibial condyles (lateral and medial). Abnormalities in the patellofemoral compartment were not assessed since a reliable evaluation would require specific radiographic and MR planes of acquisition [12].

Guidelines for categorization of criteria for analysis
of inter- and intrareader concordance rates

In Fig. 1, flowchart 3, one can see that the only grades that did not match berween the modified Larsen's scoring system and the MR-related radiographic scoring system and between the conventional Larsen's scoring system and the MR-related radiographic scoring system were grades II and III. For intrareader scoring purposes, both grades IIa and IIb on the modified Larsen's system corresponded to grade II on the conventional system and grades IIIa and IIIb to grade III. For example, if one of the blinded readers graded a particular plain film as IIa on the first reading using the modified grading system and as II on the second reading using the conventional system, it was considered as intrareader agreement.

Fig. 1 Three flowcharts. Flowchart 1 relates to the reading of plain films, flowchart 2 to the comparative assessment of plain films with corresponding MR images, and flowchart 3 contains inter- and intrareader concordance rates

Flowchart 1.
| Conventional and modified Larsen's radiographic scoring systems |

1 - Reading of plain films

Blinded readers
(Blinding to clinical data/correlating MR images)

| 3 musculoskeletal radiologists (> 5 years of experience): C.C.C., R.A.S, L.F.V. | 3 Diagnostic Imaging residents: P.R.A. (3rd year); L.L., M.A.N.G. (4th year) | 3 rheumatologists (> 10 years of experience): M.H.B.K., C.H.M.S., C.A.F.Z. |

Occasion 1: Modified scoring system (N=75); Occasion 2: Conventional scoring system (N=75) (at least 1 week of interval between readings)

*Plain films: N=75; Radiographic evaluation according to:

-Conventional scoring system (6 grades): 0 (n=12; 16%), I (n=24; 32%), II (n=23; 31%), III (n=12; 16%), IV (n=4; 5%), V (n =0; 0%)

-Modified scoring system (8 grades): 0 (n=12; 16%), I (n=24; 32%), IIa (n=17; 23%), IIb (n=6; 8%), IIIa (n=4; 5%), IIIb (n=8; 11%), IV (n=4; 5%),V (n=0; 0%)

Flowchart 2. **2 – Comparative assessment of plain films with corresponding MR images**

Unblinded readers

2 musculoskeletal radiologists: A.S.D., C.C.C (after acting as a blinded reader)

**Plain films and corresponding MR images: N=75

***MR images: coronal T1-weighted SE; coronal T2-weighted FFE

Flowchart 3. **3 – Inter- and intraobserver concordance rates**

| **Modified Larsen's MR-related radiographic scoring system** | IIa, IIb = II
IIIa, IIIb = III | **Conventional Larsen's MR-related radiographic scoring system** |

4 – Statistical analysis

Statistical analysis

The level of agreement between readers was characterized by weighted kappa values (kw), which provided a measure of inter- and intrareader agreement adjusted for chance of agreement. Values ≤ 0.40 indicated poor, > 0.40 and ≤ 0.60 moderate, > 0.60 and ≤ 0.80 good and > 0.80 very good agreement [13]. To test whether kappa scores differed statistically, P values and standard 95% confidence intervals (CI) were used for testing [14]. The two highest values of concordance between radiologists, residents, and rheumatologists on both radiographic scoring methods according to the standard reference MR scoring system were assessed using k-statistics.

Results

The mean time spent by the nine blinded readers to report the plain films according to the modified Larsen's classification system was about 30% longer than the amount of time required for evaluation of radiographs according to the conventional Larsen's system. The mean time required by the blinded readers to score the 75 plain films according to the conventional Larsen's grading system was 48.4 min (radiologists, mean 45.3,

Fig. 2a, b Subtle marginal bone erosion is demonstrated on the medial proximal tibial condyle (*arrows*) on this frontal view of the right knee (**a**) in an 18-year-old girl with long-standing JRA (14-year duration) and systemic subtype. The radiograph was scored as grade IIa by the modified Larsen's classification and as grade II by the conventional method. This finding can be better appreciated on the respective coronal T1-weighted SE image (**b**) obtained within an interval of 10 days from the plain radiograph
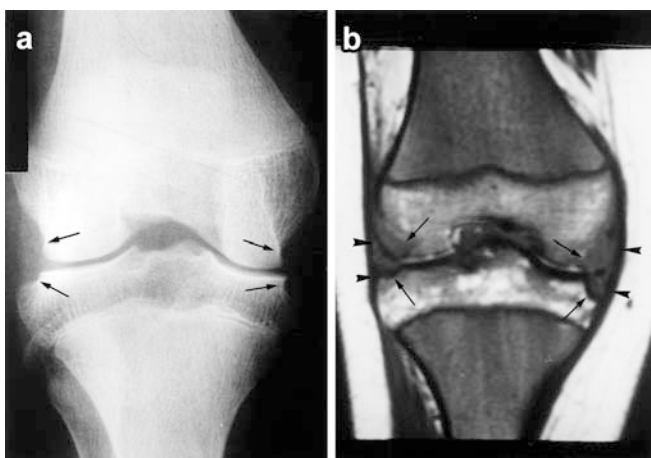
Fig. 3a, b The presence of marginal bone erosions on the external edges of both distal femoral and proximal tibial condyles (*arrows*) on this AP radiograph of the right knee (**a**) of an 11-year-old girl with pauciarticular JRA and 3.6 years of duration of disease was assigned as grade IIb by the modified system and to grade II by the conventional method. The respective coronal T1-weighted SE image (**b**) obtained 3 months following acquisition of the plain film confirms the presence of these bony defects (*arrows*). Note the presence of low signal intensity pannus proliferation overlying the region of bone defects (*arrowheads*). Note that there is a significant delay in the conversion of fatty to hematopoietic bone marrow in the distal femoral metaphysis, and epiphysis and focal areas of residual hematopoietic bone marrow are seen most notably in the proximal tibial epiphysis. These abnormalities in the bone marrow are likely related to the associated chronic anemic status of the patient

range 28–63 min; residents, mean 49, range 46–55 min; rheumatologists, mean 51, range 50–53 min). On the other hand, the mean time required by the readers to
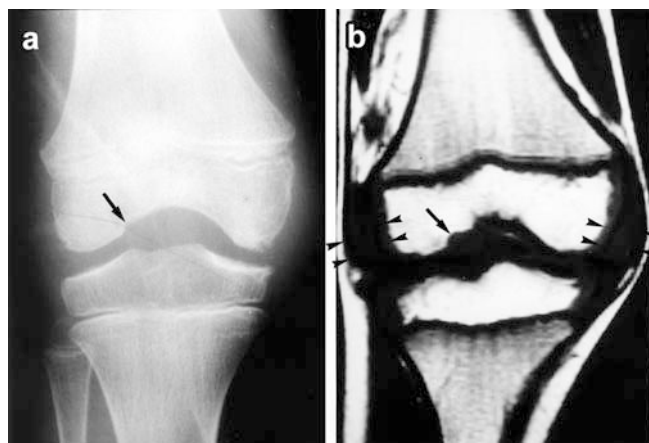

Fig. 4a, b On this frontal radiograph of the right knee (**a**) an ill-defined subchondral lesion is seen on the lateral femoral condyle of the right knee (*arrow*) of a 6-year-old girl with systemic JRA of 3 years' duration. This subtle osteocartilaginous abnormality was graded as IIIa by the modified system and as III by the conventional method. Note that this radiographic finding is represented by a subcortical area of low-signal intensity in the bone marrow beneath the chondral surface on the coronal T1-weighted image (**b**) obtained within an interval of 1 month from the corresponding plain film. Low-signal intensity soft tissue material in keeping with synovial proliferation is noted involving the joint along the lateral and medial recesses of the knee (*arrowheads*) on the MR image

Fig. 5 This is an AP plain film of the right knee of the same JRA patient as in Fig. 3 obtained within 15 months after the initial radiograph. The intraosseous lesion on the lateral aspect of the distal femoral condyle has progressed (*arrow*), but the conventional Larsen's system was not able to qualify this change for an increase in stage and both images (Fig. 3a and 4) were scored as grade III. By the modified Larsen's system, the knee of the patient of Fig. 3a was scored as grade IIIa and of Fig. 4, as grade IIIb

score the radiographs according to the modified Larsen's classification was 71 min (radiologists, mean 58.3, range 43–67 min; residents, mean 66.3, range 65–68 min; rheumatologists, mean 88.3, range 85–90 min).
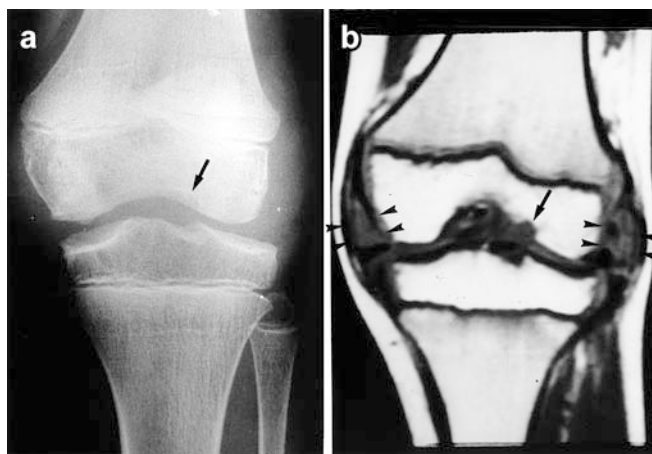
Fig. 6a, b AP radiograph (**a**) and coronal T1-weighted SE image (**b**) of the left knee of the patient of Fig. 3 demonstrate a well-defined subchondral abnormality (*arrow*) on the lateral aspect of the distal femoral condyle. The radiograph was scored as grade IIIb by the modified classification and as grade III by the conventional system. This radiograph was acquired on the same day from the plain film shown in Fig. 3, and the corresponding MR image was obtained 1 month after acquisition of the plain film. Similarly to what is noted on the contralateral knee, pannus is seen along the lateral and medial recesses of the knee (*arrowheads*) on the MR image

For the group of experienced radiologists the interobserver kappa concordance index was poor for both classification systems, ranging from 0.25 to 0.37 when comparing the conventional Larsen's scoring system with the MR-related radiographic scoring system and from 0.19 to 0.39 when comparing the modified Larsen's scoring system with the MR-related radiographic scoring system (Table 2).

For the group of residents the interreader kappa corcondance index was also poor for both the modified and conventional Larsen's system, ranging from 0.25 to 0.37 for comparison between the conventional Larsen's scoring system and the MR-related radiographic scoring system and from 0.18 to 0.30 for comparison between the modified Larsen's scoring system and the MR-related radiographic scoring system (Table 2).

For the group of rheumatologists the interreader kappa corcondance index was poor to moderate for the conventional method and poor for the modified system, ranging from 0.19 to 0.51 for comparative analysis between the conventional Larsen's scoring system and the MR-related radiographic scoring system and from 0.17 to 0.29 for comparison between the modified Larsen's scoring system and the MR-related radiographic scoring system (Table 2).

The highest weighted kappa values of interreader concordance between the conventional Larsen's scoring system and the MR-related radiographic scoring system among the blinded readers respectively were 0.43 (moderate agreement; $P < 0.001$, CI 0.31–0.55) for radiologists, 0.38 for residents (poor agreement: $P < 0.001$, CI, 0.26–0.51) and 0.37 (poor agreement: $-P < 0.001$, CI 0.25–0.50) for rheumatologists. For comparative analysis between the modified Larsen's scoring system and the standard reference MR-related radiographic scoring system these values respectively were 0.49 (moderate agreement: $P < 0.001$, CI 0.36–0.62) for radiologists, 0.28 (poor agreement – $P < 0.001$; CI, 0.18–0.39) for residents and 0.38 (poor agreement: $P < 0.001$, CI 0.25–0.51) for the rheumatologists.

The intrareader kappa concordance index between the conventional and the modified Larsen's scoring system was poor to moderate for all groups of readers, ranging from 0.28 to 0.43 for the group of experienced radiologists, from 0.30 to 0.43 for the group of residents

Table 2 Analysis of interreader concordance rates between the conventional/modified Larsen's scoring system and the MR-related radiographic scoring system (standard reference parameter)

Experienced radiologists		Kw	Qualitative equivalent	P	Confidence limits
Conventional Larsen's system	1	0.25	Poor	< 0.001	0.14–0.36
	2	0.36	Poor	< 0.001	0.24–0.48
	3	0.37	Poor	< 0.001	0.25–0.49
Modified Larsen's system	1	0.19	Poor	0.0003	0.08–0.29
	2	0.31	Poor	< 0.001	0.28–0.41
	3	0.39	Poor	< 0.001	0.29–0.49
Residents					
Conventional Larsen's system	1	0.25	Poor	< 0.001	0.13–0.37
	2	0.29	Poor	< 0.001	0.17–0.40
	3	0.37	Poor	< 0.001	0.24–0.50
Modified Larsen's system	1	0.18	Poor	< 0.001	0.09–0.26
	2	0.19	Poor	0.0001	0.09–0.28
	3	0.30	Poor	< 0.001	0.20–0.40
Rheumatologists					
Conventional Larsen's system	1	0.19	Poor	0.0007	0.08–0.31
	2	0.21	Poor	0.0002	0.10–0.32
	3	0.51	Moderate	< 0.001	0.39–0.64
Modified Larsen's system	1	0.17	Poor	0.0005	0.07–0.27
	2	0.24	Poor	< 0.001	0.14–0.33
	3	0.29	Poor	< 0.001	0.20–0.38

[a]Kw = kappa weighted values
*Values ≤ 0.40 = poor, > 0.40 and ≤ 0.60 = moderate, > 0.60 and ≤ 0.80 = good and > 0.80 = very good agreement

and from 0.25 to 0.48 for the group of rheumatologists (Table 3).

Discussion

Several studies recommend the use of MR imaging to monitor JRA progression and response to therapy [15, 16, 17]. The advantage of using MR imaging, which has been proved to be superior to radiographs for detection of intraarticular changes in rheumatoid arthritis [18, 19, 20, 21], for retrospective comparison with plain films is that subtle osteochondral lesions that could be missed on the radiographic evaluation could be reanalyzed more carefully by unblinded readers upon their identification on the corresponding MR images. Fast spin-echo T2 or intermediate-weighted imaging [16] and 3D-spoiled gradient echo (SPGR) T1-weighted sequence with fat supression [16, 22] maximize the contrast between bone and cartilage, being so far the best MR pulse sequences to accomplish the criteria for maximal contrast and spatial resolution acquisition. Although none of these MR sequences has been used in our study, previous reports have considered spin-echo T1- and T2- weighted images acceptable or at least able to identify osteochondral lesions in adult rheumatoid arthritis and JRA [10, 23]. Additional MR findings including synovial proliferation and enhancement, joint effusion, meniscal damage, and bone marrow abnormalities, were not considered in this study, since its scope was the evaluation of the variability of interpretation of two radiographic classification systems. Standard radiography is not able to show inflamed and enhanced synovium, meniscal or bone-marrow change, but may show bone erosions and evidence of cartilage destruction.

The Larsen's radiographic system for rheumatoid arthritis classification is one of the methods most often used for radiographic evaluation of patients with arthritis and has proved to be of high reproducibility in adults [5]. Nevertheless, this scoring system has limitations for use in children because of the cartilaginous structure of the epiphyses, which may mask the real timing of development of bone erosions. In addition, the irregularity of the epiphyseal cartilage corroborates missing bone erosions on radiographs even after extensive subchondral bone damage has occurred. With the maturation of the joint, the epiphyseal cartilage thickness decreases, enabling more accurate evaluation of osteochondral destruction. Although Petterson [6] has described a radiographic method for evaluation of JRA, some pediatric centers in the world, including ours, still use the Larsen's system for radiographic follow-up of JRA in the routine practice.

Findings in our study show a poor colinearity between the blinded (nine readers) and the unblinded (two readers) grading of radiologic findings for both the conventional and modified Larsen's classification systems (Table 2). A wide range of kappa values (0.19–0.51 for the conventional system and 0.17–0.39 for the modified system) was obtained from the blinded readers, regardless of their previous experience in Diagnostic Imaging (Table 2). A variable that might have accounted for the low interreader concordance rates was the high number of blinded readers who participated in the interpretation of radiographic findings. Previous studies have demonstrated that a better level of interreader agreement in image interpretation is obtained from a lower number of readers [24, 25, 26]. The increase in the number of readers or in the number of categories to be analyzed would have a direct implication in the increase of frequency of opportunities for disagreement.

The highest kappa values of concordance for the conventional Larsen's system were obtained by the group of radiologists, with the values obtained from the groups of residents and rheumatologists following respectively. On the other hand, despite the highest values of agreement between the modified Larsen's system and the MR-related standard radiographic system have also been obtained by the group of experienced radiologists, there was an inversion in the concordance rates for the two other groups. These results likely reflect the role of the level of radiographic experience on the consistency of interpretation of random plain films, showing that in-training readers (residents) and occasional readers of radiographs (rheumatologists) seem to be more prone to variations in the interpretation of radiographic findings on the basis of different classification systems.

Table 3 Analysis of intrareader concordance rates between the conventional and the modified Larsen's scoring systems for the three groups of blinded readers (experienced radiologists, residents and rheumatologists)

Experienced radiologists	Kw[a]	QE	Confidence limits*,**	Res	Kw	QE	Confidence limits	Rheum	Kw	QE	Confidence limits
1	0.28	P	0.17–0.39	1	0.30	P	0.20–0.40	1	0.25	P	0.14–0.36
2	0.36	P	0.25–0.47	2	0.36	P	0.24–0.48	2	0.27	P	0.15–0.39
3	0.43	M	0.32–0.55	3	0.43	M	0.30–0.56	3	0.48	M	0.37–0.59

[a]Kw = kappa weighted values; QE = qualitative equivalent; Res = residents; Rheum = rheumatologists
*P < 0.001: for all readers

**Values ≤ 0.40 = poor (P), > 0.40 and ≤ 0.60 = moderate (M), > 0.60 and ≤ 0.80 = good (G) and > 0.80 = very good (VG) agreement

Table 3 shows that the intrareader concordance rates for the conventional and modified Larsen's radiographic methods did not change significantly for the three groups of blinded readers, converse to the interreader results. Although the overall reproducibility indices for all blinded readers have been low to moderate, these indices tended to be higher than the interreader concordance indices (Tables 2). The intrareader results demonstrate a tendency for the observer, regardless of his/her level of experience in diagnostic imaging, to repeat his / her way of interpreting radiographic findings or, in other words, to repeat errors of interpretation and deficits of attention skills.

A limiting factor for this study was the interval discrepancy between acquisition of plain films and the corresponding MR images, which was related to difficulties for performance of both examinations on the same day in some cases and to delay in re-scheduling the following MR examination in the remaining cases. Wallace and Levinson [1] have reported that the median time to development of radiographic changes of joint-pace narrowing and marginal erosions was 5 years in patients with pauciarticular JRA onset, 2 years in patients with polyarticular onset, and 2 years in patients with systemic onset disease. These results demonstrated that there is a relatively long interval of time until the development of radiographic changes. Moreover, in cases in which MR images showed osteochondral lesions not identifiable on the corresponding plain films, the MR-related radiographic scoring system did not categorize those lesions since they were not visible from the radiographic point of view. Based on these two assumptions, we considered acceptable a maximum interval of 3 months between acquisition of plain films and MR imaging as a criterion for inclusion of a patient in the study.

Another limitation of the study was a verification bias, since both volunteers and clinically indicated cases were included in our protocol, which most likely resulted in the evaluation of radiographic and MR examinations from a spectrum of JRA patients with greater severity of disease. However, both patients with favorable (post-treatment follow-up) and unfavorable clinical progression of disease were included in the study, reducing the extension of the bias effect and directing the results of the study towards the evaluation of a limited number of cases with presence of severe osteochondral changes. This effect favoured the evaluation of small abnormalities with the modified scoring system. If fewer cases with presence of osteochondral abnormalities were available, a larger sample size might have been required for the study.

Neither the conventional and nor the modified Larsen's classification systems take into account the acceleration of bone maturity owing to the inflammatory process involving the joint. Knees with previous inflammatory activity and consequent accelerated maturity are expected to present with smaller thickness of epiphyseal cartilage for the patient's age group and gender. This factor has an important effect on the radiographic assessment since cartilaginous abnormalities cannot be directly assessed on plain films. As a result, osteochondral lesions are more easily identifiable in patients with advanced bone maturation (smaller thickness of epiphyseal cartilage). Conversely, in younger patients, especially in those aged ≤ 5 years ($n = 3$; 4% in our study) who have a spherical or elliptical ossification center surrounded by a bulk of epiphyseal cartilage [27] and at a lesser extent in those ≤ 9 years old ($n = 16$; 21.3% in our study) osteochondral abnormalities may be missed, making radiographs less valuable in the younger population.

Problems for designing classification systems include the difficulty to obtain a reproducible system that is able to describe findings in a progressive and additive way. Pathological changes in JRA do not seem to follow a standardized pattern; thus, patients may present some abnormalities from a higher grade in the classification system, without presenting the ones from the lower grades.

In summary, findings in this study show that although the level of experience in Diagnostic Imaging might have had an implication in the interreader concordance rate for interpretation of plain radiographs, the intrareader concordance rates did not change significantly according to the system utilized and to the level of experience of the readers. This demonstrates a trend towards repetition of the way the reader interprets images in subsequent reports.

The low concordance rates obtained from all groups of blinded readers for both the conventional and the modified Larsen's systems discourage the clinical application of these radiographic classification systems for a careful evaluation of osteochondral progression in the knees of children with JRA. Rather than increasing the sensitivity of Larsen's system, the inclusion of additional parameters emphasizing the search for osteochondral changes into a modified system increased the degree of complexity of the method. For the reasons previously described, the plain film classification schemes should be replaced by MR imaging classification systems using optimized cartilage sequences in the near future, which will be particularly beneficial for younger patients. Since more specific therapies for JRA have been developed in recent times and patients can be "fine-tuned" by imaging evaluation and placed into specific treatment algorithms, specific imaging classification schemes are needed for reproducibility.

References

1. Wallace CA, Levinson JE (1991) Juvenile rheumatoid arthritis: outcome and treatment for the 1990s. Rheum Dis Clin North Am 17:891–905
2. Stanley P, Senac M (1990) Miscellaneous disorders of the musculoskeletal system. In: Cohen MD, Edwards MK (eds) Magnetic resonance imaging of children. Decker, Philadelphia, pp 944–995
3. Steinbrocker O, Traeger CH, Batterman RC (1949) Therapeutic criteria in rheumatoid arthritis. JAMA 140:659–662
4. Sharp JT, Lidsky MD, Collins LC, et al (1971) Methods of scoring the progression of radiologic changes in rheumatoid arthritis: correlation of radiologic, clinical, and laboratory abnormalities. Arthritis Rheum 14:706–720
5. Larsen A, Dale K, Eek M (1977) Radiographic evaluation of rheumatoid arthritis and related conditions by standard reference films. Acta Radiol Diagnosis 18: 481–491
6. Petterson H, Rydholm U (1984) Radiologic classification of knee joint destruction in juvenile chronic arthritis. Pediatr Radiol 14:719–722
7. Scott DL, Grindulis KA, Struthers GR, et al (1984) Progression of radiological changes in rheumatoid arthritis. Ann Rheum Dis 43:8–17
8. Brewer EJ, Bass J, Baum J, et al (1977) Current proposed revision of JRA criteria. Arthritis Rheum 20[Suppl37]:195–199
9. Doria AS, Kiss MHB, Sallum AM, et al (2001) Correlation between osteochondral changes depicted by magnetic resonance imaging and disease progression. Rev Hosp Clin Fac Med S Paulo 56:107–114
10. Winalski CS, Palmer WE, Rosenthal DI, et al (1996) Magnetic resonance imaging of rheumatoid arthritis. Radiol Clin North Am 34:243–258
11. Gubler FM, Algra PR, Maas M, et al (1993). Gadolinium-DTPA enhanced magnetic resonance imaging of bone cysts in patients with rheumatoid arthritis. Ann Rheum Dis 52:716–719
12. Burk DL Jr, Dalinka MK, Schliebler ML, et al (1988) Strategies for musculoskeletal magnetic resonance imaging. Radiologic Clin North Am 26:653–672
13. Altman DG (1991) Practical statistics for medical research. Chapman and Hall, London, pp 404–408
14. Cicchetti H (1981) A computer program for determining the significance of the difference between pairs od independently derived values of kappa or weighted kappa scores. Educational Psychol Meas 41:189–193
15. Graham TB, Blebea JS, Gylys-Morin V, et al (1997) Magnetic resonance imaging in juvenile rheumatoid arthritis. Semin Arthritis Rheum 27:161–168
16. Gylys-Morin VM, Graham TB, Blebea JS, et al (2001) Knee in early juvenile rheumatoid arthritis: MR imaging findings. Radiology 220:696–706
17. Cakmakci H, Kovanlikaya A, Unsal E (2001) Short-term follow-up of the juvenile rheumatoid knee with fat-saturated 3D MRI. Pediatr Radiol 31:189–195
18. Beltran J, Caudill JL, Herman LA, et al (1987) Rheumatoid arthritis: MR imaging manifestations. Radiology 165:153–157
19. Foley-Nolan D, Stack JP, Ryan M, et al (1991) Magnetic resonance imaging in the assessment of rheumatoid arthritis—a comparison with plain film radiographs. Br J Rheumatol 30:101–106
20. Poleksic L, Zdravkovic D, Kablanovic D, et al (1993) Magnetic resonance imaging of bone destruction in rheumatoid arthritis: comparison with radiography. Skeletal Radiol 22:577–580
21. Gilkeson G, Polisson R, Sinclair H, et al (1988) Early detection of carpal erosions in patients with rheumatoid arthritis: a pilot study of magnetic resonance imaging. J Rheumatol 15:1361–1366
22. Disler DG (1997) Fat-suppressed three-dimensional spoiled gradient-recalled MR imaging: assessment of articular and physeal hyaline cartilage. AJR 169:1117–1123
23. Senac MO, Deutsch D, Bernstein BH, et al (1988) MR imaging in juvenile rheumatoid arthritis. AJR 150:873–878
24. Vuillez JP, Peltier P, Mayer JC, et al (1991) Reproducibility of image interpretation in immunoscintigraphy performed with indium-111- and iodine-131-labeled OC125 F(ab')2 antibody injected into the same patients. J Nucl Med 32:221–226
25. Brunett A, Tedeschi G, Di Constanzo, et al (1997) White matter lesion detection in multiple sclerosis: improved interobserver concordance with multispectral MRI display. J Neurol 244:586–590
26. Robinson PJA, Wilson D, Coral A, et al (1999) Variation between experienced readers in the interpretation of accident and emergency radiographs. Br J Radiol 72:323–330
27. Varich LJ, Laor T, Jaramillo D (2000) Normal maturation of the distal femoral epiphyseal cartilage: age-related changes at MR imaging. Radiology 214:705–709