Arthritis Care & Research
Vol. 65, No. 5, May 2013, pp 686–694
DOI 10.1002/acr.21890

ORIGINAL ARTICLE

Assessment of Cartilage Changes Over Time in Knee Osteoarthritis Disease-Modifying Osteoarthritis Drug Trials Using Semiquantitative and Quantitative Methods: Pros and Cons

LUKAS MARTIN WILDI,[1] JOHANNE MARTEL-PELLETIER,[2] FRANÇOIS ABRAM,[3] THOMAS MOSER,[1] JEAN-PIERRE RAYNAULD,[1] AND JEAN-PIERRE PELLETIER[2]

Objective. **To evaluate the impact of 2 magnetic resonance imaging (MRI) sequences on cartilage defect assessment in knee osteoarthritis (OA) patients and the sensitivity to change over time comparing cartilage defect (semiquantitative) with cartilage volume loss (quantitative) methods.**

Methods. **Gradient-echo (GRE) and intermediate-weighted fast spin-echo (IW-FSE) sequences were compared. Knee OA MRIs were from two 2-year studies (cohort 1, n = 55; cohort 2, n = 143). For both cohorts, a GRE sequence was used and patients in cohort 1 underwent an additional IW-FSE sequence. Cohort 2 included patients from a previous trial. Cartilage defects and cartilage volume were evaluated.**

Results. **The cartilage defect assessment provided consistently significantly higher scores in IW-FSE than in GRE sequences at baseline and 2 years. However, there was no difference in the change at 2 years between the sequences. The standardized response mean (SRM) for change did not show a difference between the 2 sequences, but was consistently higher (2–2.5-fold) for the quantitative method. The cartilage defect score change between the 2 treatment groups revealed a trend toward significance only in the medial tibial plateau, whereas the change in cartilage volume loss demonstrated a significant difference in the global knee, global femur, lateral femur, and lateral compartment. The SRMs for the treatment groups combined were markedly higher for cartilage volume loss than for the defect scoring by 4.3- to 6.0-fold.**

Conclusion. **The direct comparison between GRE and IW-FSE sequences did not suggest superior sensitivity to cartilage defect change over time of one sequence over the other. Interestingly, the quantitative cartilage volume assessment was more sensitive than the semiquantitative scoring in the detection of treatment effect on OA cartilage changes.**

INTRODUCTION

Osteoarthritis (OA) is the most common form of arthritis. For the development of disease-modifying OA drugs (DMOADs), investigators are in need of technologies that allow reliable and sensitive assessment of joint structural changes over time to study treatment effects. Such development is challenging in a disease known to be of insidious onset and slow progression. For decades, the technical means were limited to conventional radiographs, which allow the articular cartilage to be evaluated only in an indirect manner. This method was shown to have low sensitivity, even though it has been improved with computer-assisted methods and new imaging protocols. Moreover, radiographic measurement comprises the joint space

The original clinical study was supported in part by grants from Merckle (Ulm, Germany) and ArthroLab Inc. (Montreal, Quebec, Canada).

[1]Lukas Martin Wildi, MD, Thomas Moser, MD, Jean-Pierre Raynauld, MD: University of Montreal Hospital Research Centre, Notre-Dame Hospital, Montreal, Quebec, Canada; [2]Johanne Martel-Pelletier, PhD, Jean-Pierre Pelletier, MD: University of Montreal Hospital Research Centre, Notre-Dame Hospital, and ArthroLab Inc., Montreal, Quebec, Canada; [3]François Abram, PhD: ArthroLab Inc., Montreal, Quebec, Canada.

Dr. Wildi has received a bursary from the Swiss National Science Foundation. Drs. Martel-Pelletier and Pelletier own ArthroLab Inc. Dr. Moser has received consultant fees, speaking fees, and/or honoraria (less than $10,000) from ArthroLab Inc. Dr. Raynauld has received consultant fees, speaking fees, and/or honoraria (less than $10,000) from ArthroLab Inc.

Address correspondence to Jean-Pierre Pelletier, MD, Osteoarthritis Research Unit, University of Montreal Hospital Research Centre, Notre-Dame Hospital, 1560 Sherbrooke Street East, Montreal, Quebec, Canada H2L 4M1. E-mail: dr@jppelletier.ca.

Submitted for publication May 1, 2012; accepted in revised form October 11, 2012.

Significance & Innovations

- Direct comparison between gradient-echo and intermediate-weighted fast spin-echo magnetic resonance imaging sequences did not suggest superior sensitivity of either sequence to detect cartilage defect change over time.
- For the detection of treatment effect on osteoarthritis cartilage changes, the quantitative cartilage volume assessment was more sensitive than the semiquantitative scoring.

narrowing (JSN) for which reliability is compromised by a number of factors, including the positioning of the joint and meniscal extrusion. In addition, joint space assessment is subjected to a very small annual change (1,2).

The introduction of magnetic resonance imaging (MRI) to the field of OA allowed direct, precise, and reliable assessment of cartilage and other joint structures and their changes over time. Several different semiquantitative scoring systems using different acquisition protocols have been developed and proposed to assess cartilage defects in patients with knee OA, including the Whole-Organ Magnetic Resonance Imaging Score (WORMS) (3), the Knee Osteoarthritis Scoring System (4), the Boston Leeds Osteoarthritis Knee Score (5–7), the system published by Ding et al (8), and the recently published MRI Osteoarthritis Knee Score (9). These scoring systems were used for the understanding of the pathophysiology of OA and the relationship between joint tissue structural changes and disease symptoms. They can assess cartilage defects in a number of subregions of the joint using a grading scale. Quantitative cartilage assessments have also been developed (10–12) and successfully applied in OA clinical trials (13–15). However, there is discussion about which method is optimal to assess cartilage changes, semiquantitative (scores) or quantitative (continuous values), particularly in the context of a DMOAD trial.

In contrast to quantitative cartilage assessment, for which the gradient-echo (GRE) sequences are widely accepted (12,16), the question about the best sequence to be used for detecting cartilage defects and their change over time is still under discussion. There are obvious differences in the depiction of joint structures, depending on the choice of sequence and its weighting. Clefts in the cartilage surface, for example, seem to be better depicted using water-sensitive intermediate-weighted fast spin-echo (IW-FSE) sequences with fat suppression (FS), which favor contrast resolution, i.e., differentiation of tissue with different water content (17–19). On the other hand, the spatial resolution is superior with GRE sequences, which allow for less partial volume effects and a more precise delineation of the cartilage (12,20–22), as reflected by a better specificity than FSE when using arthroscopic grading as a reference (23,24). Peterfy et al (3) suggested the use of 2 sequences simultaneously for the scoring (WORMS): an FS–3-dimensional (3-D) spoiled gradient-recalled (SPGR) sequence and an FS–T2-weighted FSE sequence, unfortunately without specifically indicating how to merge the findings. For a cross-sectional assessment, both of the sequences may be considered to be appropriate due to high specificity for cartilage pathology (25) as long as no histopathology-controlled study has proven superiority. For the assessment of cartilage defect change over time, however, there are no data to date from a head-to-head comparison to answer the question of whether one particular type of sequence is superior to another.

The lack of information on this important issue prompted us to proceed to a head-to-head study in a knee OA patient population comparing the sensitivity to cartilage defect change of a water-sensitive FSE-type sequence with a high-resolution 3-D GRE sequence as well as, for the GRE sequence, the cartilage defect with the cartilage volume assessment. In a second step, we compared, in the context of an OA clinical trial setting, the sensitivity in assessing the effect of drug treatment on the OA cartilage degradation of the semiquantitative (cartilage defect WORMS) and the quantitative (cartilage volumetry) methods.

PATIENTS AND METHODS

Patient selection. Post hoc analyses were performed on 2 cohorts from recent knee OA studies (13,26). For both cohorts, patients with primary symptomatic knee OA of the tibiofemoral compartment diagnosed according to the American College of Rheumatology criteria (27) were recruited from outpatient rheumatology clinics. The studies were approved by the ethics committees and all patients gave their oral and written informed consent to participate. The study period for cohort 1 was from 2006–2011 and for cohort 2 was from 2001–2007. Major inclusion criteria were a Kellgren/Lawrence grade of 2 or 3, a minimal medial joint space width of 2.5 mm and 2.0 mm for cohorts 1 and 2, respectively, and medial tibiofemoral OA.

Cohort 1 (n = 55) was used to evaluate the impact of 2 MRI sequences on the cartilage defect assessment. Patients had knee OA and had a mean ± SD age of 62.0 ± 7.4 years, had a mean ± SD body mass index of 30.2 ± 4.2 kg/m^2, and 84% were women. The 2 sequences that were both performed on each patient were a 3-D GRE and an IW-FSE.

Cohort 2 (n = 143) was used to compare the detectability of treatment effect on structural changes between semiquantitative scoring (3) and quantitative volumetry (10,11,13). This study was a subanalysis of a published trial in knee OA patients comparing oral intake of licofelone, a lipoxygenase–cyclooxygenase inhibitor to naproxen (13). Briefly, the patients were randomly assigned to receive either therapeutic doses of licofelone (200 mg twice a day) or naproxen (500 mg twice a day) for 2 years. Only the patients who had taken all of the study medication provided throughout the 2-year period and had MRI examinations at baseline and 2 years (according to the protocol population) were used (13,28). Patient characteristics were equally distributed between the 2 treatment groups (licofelone, n = 70; naproxen, n = 73) (13). Further information concerning these clinical studies is found in previous publications (13,26,29).

Knee MRI acquisition. MRI examinations were performed at baseline and 2 years using commercially available whole-body 1.5T scanners with a dedicated knee coil. All of the details about these examinations have been previously published (13). All subjects were imaged with one of two 3-D GRE sequences with similar characteristics, owing to the different scanners of the participating study centers. In addition, patients from cohort 1 were imaged with an IW-FSE sequence.

For the GRE sequences, 2 apparatuses were used. A sagittal fat-suppressed spoiled GRE sequence (SPGR, General Electric) was performed with the following parameters: repetition time (TR) 42 msec, echo time (TE) 7 msec, flip angle 20°, slice thickness 1.5 mm, slice gap 0 mm, number of excitations 1, matrix size 512 × 512, field of view 160 mm, receiver bandwidth 122 Hz/pixel, and phase-encoding direction anteroposterior (AP). A sagittal steady-state GRE sequence with water excitation (fast imaging with steady-state precession, Siemens) was performed with the following parameters: TR 23 msec, TE 11 msec, flip angle 14°, slice thickness 1.5 mm, slice gap 0 mm, number of excitations 1, matrix size 512 × 384, field of view 160 mm, receiver bandwidth 120 Hz/pixel, and phase-encoding direction AP. These 2 GRE sequences are similar to the fast low-angle shot and double-echo steady-state sequences, respectively, which proved equivalent for quantitative cartilage analysis in a recent pilot study for the Osteoarthritis Initiative (30).

For the IW-FSE sequence, a sagittal fat-suppressed IW-FSE sequence was performed with the following parameters: TR 4,760 msec, TE 43 msec, flip angle 90°, slice thickness 3 mm, slice gap 0 mm, number of excitations 1, matrix size 256 × 192, field of view 140 mm, resolution 0.547 × 0.791 mm, receiver bandwidth 122 Hz/pixel, and phase-encoding direction head–feet (HF) for General Electric, and TR 4,870 msec, TE 31 msec, flip angle slice 170°, slice thickness 3.5 mm, slice gap 0 mm, number of excitations 1, matrix size 256 × 256, field of view 140 mm, resolution 0.547 × 0.547 mm, receiver bandwidth 100 Hz/pixel, and phase-encoding direction HF for Siemens.

Cartilage assessment. Cartilage assessment was performed blinded to the treatment. Cartilage defects were assessed by a single reader (trained and data validated by a musculoskeletal radiologist) with more than 10 years of experience in scoring according to the WORMS scoring system. The scoring values were expressed as the median (minimum, maximum) as well as the mean ± SD. The cartilage volume was measured using proprietary software (Cartiscope, ArthroLab) as previously described (10,11,31), using cartilage maps for each slice lying within a mask covering the same areas in baseline and followup MRI. In brief, the technology involves 5 main steps (10,11,31): 1) cartilage inner and outer boundary 2-dimensional delineation in all of the images presenting bone epiphysis (semiautomated), 2) generation of a standardized view of the 3-D cartilage geometry for the inner and outer surfaces (automated), 3) anatomic labeling of cartilage geometry (automated), 4) computation of cartilage volume for all regions (automated), and 5) registration of baseline data sets and corresponding visit images (for followup evaluation). The change in knee cartilage volume was obtained by subtracting the initial (baseline) volume from the followup volume and expressed as the percentage of baseline volume. The reproducibility of the method was previously demonstrated to be excellent (11).

The cartilage defect scores and volume determinations were performed on 8 knee subregions (32,33): the medial and lateral trochlea, the central and posterior femoral condyle, and the medial and lateral tibial plateau. The summation of the medial or lateral femur subregions formed the medial or lateral femur (trochlea and condyle), and summation of the medial and lateral femur subregions or tibial plateau comprised the global femur or global tibial plateau. The medial and lateral compartments each consisted of the summation of the femur subregions with the tibial plateau, and the 2 compartments (medial and lateral) combined comprised the global knee. For the cartilage defect, the maximal scores were 6 for the medial or lateral tibial plateau, 18 for the medial or lateral femur, 24 for the medial or lateral compartment, 12 for the global tibial plateau, 36 for the global femur, and 48 for the global knee.

Statistical analyses. Data were entered into a computerized database using a blinded double-entry procedure. Intrareader reliability was assessed in 20 patients from cohort 1 with an interval of 4 weeks using Cohen's weighted kappa, where <0 = no agreement, 0–0.20 = slight agreement, 0.21–0.40 = fair agreement, 0.41–0.60 = moderate agreement, 0.61–0.80 = substantial agreement, and 0.81–1 = almost perfect agreement (34). The significance of differences in the cartilage defect scores between the scoring in GRE and IW-FSE sequences was assessed by Wilcoxon's signed rank test at baseline and 2 years, and for the change between the 2 time points. For the data from cohort 2, the cartilage defect differences were assessed between treatment groups with the Wilcoxon-Mann-Whitney test, whereas normally distributed cartilage volume was assessed using independent-sample Student's *t*-tests at all time points. All tests were 2-sided, and a *P* value less than or equal to 0.05 was considered statistically significant. For both methods, the standardized response mean (SRM) was calculated as the mean change at 2 years divided by the SD of the change (35–37). The advantage of the SRM lies in its independence of sample sizes and the direct comparability of the values obtained through different tools. All statistical analyses were performed using SPSS statistics software, version 19, and Matlab R2009a.

RESULTS

Intra- and interreader reliability of cartilage defect. Intrareader reliability as reflected by the respective weighted kappa values showed very good to excellent agreement for the cartilage defect assessment in the GRE and IW-FSE sequences in the 8 subregions (GRE = 0.622–0.956, IW-FSE = 0.650–0.931).

Assessment of cartilage defects in GRE and IW-FSE sequences. The comparison of the distribution of OA cartilage defect scores between the GRE and IW-FSE se-

Table 1. Comparison of OA cartilage defect scores between the GRE and IW-FSE sequences (cohort 1) at baseline and 2 years, and changes at 2 years*

		GRE cartilage defect score†		IW-FSE cartilage defect score†			GRE cartilage volume,	SRM‡		
	Score	Mean ± SD	Median (min, max)	Mean ± SD	Median (min, max)	*P*§	mean ± SD mm^3†	GRE CD	IW-FSE CD	GRE CV
Baseline										
Global knee	0–48	26.6 ± 2.4	26 (21, 33)	29.2 ± 2.3	29 (25, 35)	< 0.001	10,132.5 ± 2,008.9			
Global femur	0–36	20.0 ± 2.0	20 (16, 26)	22.1 ± 1.9	22 (19, 28)	< 0.001	6,938.4 ± 1,457.1			
Global tibia	0–12	6.6 ± 0.8	7 (4, 8)	7.1 ± 0.8	7 (6, 10)	< 0.001	3,194.1 ± 638.9			
Medial compartment	0–24	13.8 ± 1.4	14 (11, 16)	15.0 ± 1.5	15 (12, 18)	< 0.001	4,863.0 ± 951.9			
Lateral compartment	0–24	12.8 ± 1.4	13 (8, 17)	14.2 ± 1.4	14 (12, 18)	< 0.001	5,269.5 ± 1,153.5			
Medial femur	0–18	10.2 ± 1.1	10 (8, 12)	11.2 ± 1.1	11 (9, 14)	< 0.001	3,423.7 ± 691.4			
Lateral femur	0–18	9.8 ± 1.2	9 (7, 14)	11.0 ± 1.2	11 (9, 14)	< 0.001	3,514.6 ± 838.9			
Medial plateau	0–6	3.6 ± 0.5	4 (3, 4)	3.9 ± 0.5	4 (3, 6)	< 0.001	1,439.3 ± 308.3			
Lateral plateau	0–6	3.0 ± 0.5	3 (1, 4)	3.2 ± 0.5	3 (3, 5)	0.005	1,754.8 ± 393.6			
2 years										
Global knee	0–48	28.0 ± 2.6	28 (24, 34)	30.7 ± 2.2	31 (26, 37)	< 0.001	9,594.4 ± 1,888.6			
Global femur	0–36	21.2 ± 2.1	21 (18, 27)	23.3 ± 1.7	23 (19, 28)	< 0.001	6,570.2 ± 1,364.3			
Global tibia	0–12	6.8 ± 0.7	7 (6, 8)	7.4 ± 0.8	7 (6, 10)	< 0.001	3,024.2 ± 597.7			
Medial compartment	0–24	14.4 ± 1.4	15 (12, 17)	15.8 ± 1.2	16 (12, 19)	< 0.001	4,571.4 ± 888.1			
Lateral compartment	0–24	13.6 ± 1.6	13 (12, 17)	14.9 ± 1.4	15 (12, 18)	< 0.001	5,023.1 ± 1,094.5			
Medial femur	0–18	10.8 ± 1.1	11 (9, 13)	11.8 ± 0.9	12 (9, 14)	< 0.001	3,201.0 ± 650.3			
Lateral femur	0–18	10.4 ± 1.4	10 (9, 14)	11.5 ± 1.1	12 (9, 14)	< 0.001	3,369.2 ± 796.5			
Medial plateau	0–6	3.6 ± 0.5	4 (3, 4)	4.1 ± 0.5	4 (3, 6)	< 0.001	1,370.3 ± 279.5			
Lateral plateau	0–6	3.2 ± 0.4	3 (3, 4)	3.4 ± 0.5	3 (3, 5)	0.007	1,653.9 ± 372.2			
Change at 2 years										
Global knee		1.5 ± 1.5	1 (−2, 6)	1.5 ± 1.4	1 (−2, 5)	0.929	−538.1 ± 312.7	0.969	1.036	−1.720
Global femur		1.2 ± 1.2	1 (−1, 5)	1.2 ± 1.2	1 (−2, 4)	0.949	−368.2 ± 248.9	0.969	0.993	−1.479
Global tibia		0.3 ± 0.7	0 (−1, 3)	0.3 ± 0.6	0 (−1, 2)	0.658	−169.9 ± 136.3	0.379	0.486	−1.246
Medial compartment		0.6 ± 0.8	1 (−1, 3)	0.8 ± 1.0	1 (−1, 4)	0.329	−291.6 ± 204.9	0.751	0.802	−1.424
Lateral compartment		0.8 ± 1.1	1 (−1, 4)	0.7 ± 1.0	1 (−2, 3)	0.600	−246.4 ± 195.9	0.718	0.676	−1.258
Medial femur		0.6 ± 0.7	0 (−1, 3)	0.6 ± 0.8	0 (−1, 3)	0.699	−222.7 ± 164.8	0.789	0.768	−1.351
Lateral femur		0.6 ± 1.0	0 (−1, 3)	0.6 ± 0.9	1 (−2, 2)	0.906	−145.4 ± 163.4	0.636	0.649	−0.890
Medial plateau		0.1 ± 0.4	0 (−1, 1)	0.2 ± 0.4	0 (0, 1)	0.134	−68.9 ± 86.0	0.136	0.439	−0.801
Lateral plateau		0.2 ± 0.5	0 (0, 2)	0.1 ± 0.4	0 (−1, 1)	0.384	−101.0 ± 78.3	0.411	0.294	−1.289

* OA = osteoarthritis; GRE = gradient-echo; IW-FSE = intermediate-weighted fast spin-echo; SRM = standardized response mean; min = minimum; max = maximum; CD = cartilage defect; CV = cartilage volume.
† Cartilage defects were assessed in the knee OA patients (n = 55) using the Whole-Organ Magnetic Resonance Imaging Score system (3) and cartilage volume was assessed as previously described in the GRE sequence (10,11,31). Values are the number of voxels.
‡ Calculated as the mean change divided by the SD of the change.
§ *P* values were assessed by Wilcoxon's signed rank test.

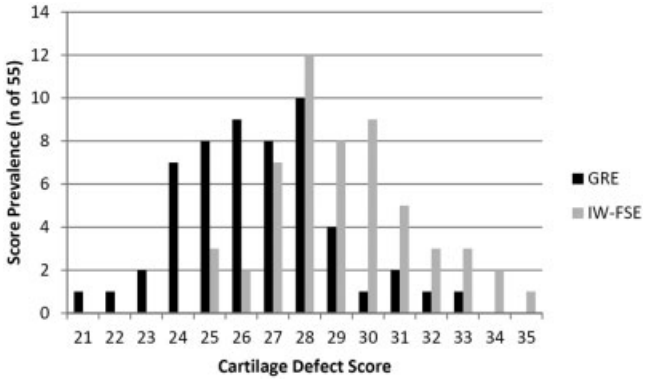

Figure 1. Bar graph of the prevalence of cartilage defect scores in osteoarthritis patients (cohort 1) in the global knee at baseline showing the difference in the scores assessed in the gradient-echo (GRE) and the intermediate-weighted fast spin-echo (IW-FSE) sequences.

quences was done using data from cohort 1. In all regions assessed, the mean score, both at baseline and 2 years, was significantly higher in the IW-FSE sequence than in the GRE sequence (Table 1). This is depicted in Figure 1, which shows a greater prevalence for each defect score for the cohort assessed in the IW-FSE sequence than in the GRE sequence, and in Figure 2, which represents the comparison of the same cartilage defect assessed in GRE and IW-FSE sequences. Interestingly, as shown in Figure 3 and Table 1, although the scores were higher in the IW-FSE

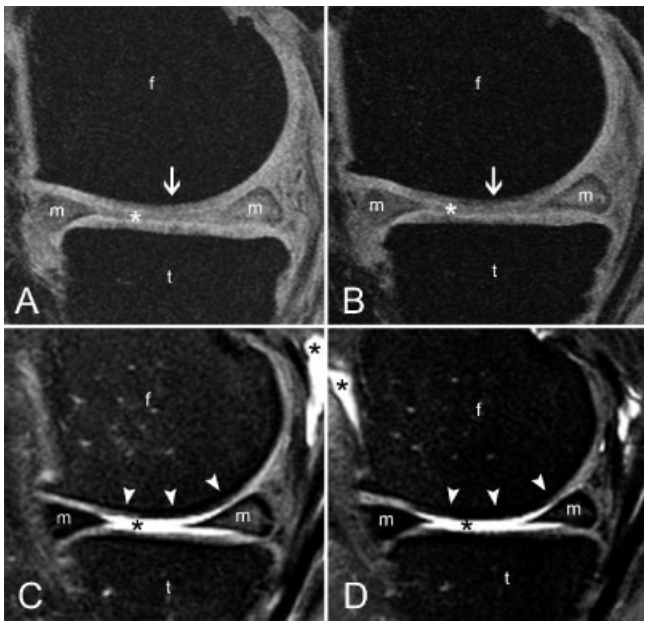

Figure 2. Representative magnetic resonance images of the same knee in the sagittal plane taken at the same level (medial central condyle) in gradient-echo (**A** and **B**) and intermediate-weighted fast spin-echo (**C** and **D**) sequences. The images show a cartilage defect of **A**, grade 3 at baseline (**arrow**), **B**, grade 4 at 2 years (**arrow**), **C**, grade 4 at baseline, and **D**, grade 5 at 2 years. **C** and **D** show the partial volume effect of the bright synovial fluid in the joint space outshining to the top layer of the cartilage surface (**asterisks**), which is not seen in **A** and **B** (**asterisks**). **C** and **D** also show the bone–cartilage interface of the deep calcified layer obscuring the interface by a thick black line (**arrowheads**). f = femur; t = tibia; m = meniscus.

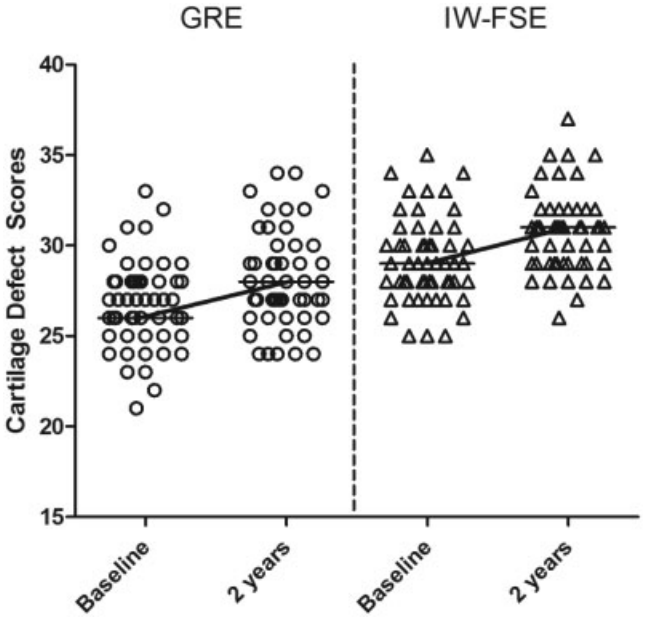

Figure 3. Scatter plot of the global knee cartilage defect scores for each individual assessed in the gradient-echo (GRE) and the intermediate-weighted fast spin-echo (IW-FSE) sequences at baseline and 2 years. The change from baseline at 2 years is indicated by a line between the medians.

sequence at baseline and 2 years than in the GRE sequence, the change in scores was similar for both sequences. This is further depicted in Figure 4, which shows the distribution of score change over time in the 2 sequences. Moreover, there was no meaningful difference in the SRMs between GRE and IW-FSE sequences (Table 1). Compared to the cartilage volume assessment, the SRM ratios of cartilage volume to cartilage defect were on average 2.5-fold lower (ratios in [1.4, 5.9], depending on the region) for the GRE sequence scoring system and were on average 2-fold lower (ratios in [1.5, 4.4], depending on the region) for the IW-FSE sequence, implying an inferior sensitivity to change of the semiquantitative method (Table 1). Similarly, the sensitivity to change over 2 years expressed as the SRM for the minimal and mean JSN, which is still the

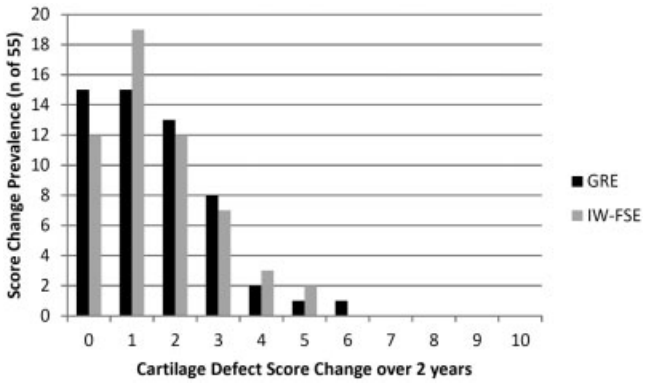

Figure 4. Bar graph of the prevalence of change in cartilage defect scores in osteoarthritis patients (cohort 1) in the global knee over 2 years showing the similar change assessed in the gradient-echo (GRE) and the intermediate-weighted fast spin-echo (IW-FSE) sequences.

Table 2. Cartilage defect change at 2 years according to treatment (cohort 2) in all patients*

	Naproxen (n = 73)		Licofelone (n = 70)		
	Mean ± SD	Median (min, max)	Mean ± SD	Median (min, max)	*P*†
Global knee	1.1 ± 2.0	1 (−4, 7)	0.5 ± 2.2	0.5 (−6, 6)	0.104
Global femur	0.7 ± 1.6	1 (−3, 5)	0.3 ± 1.9	0 (−6, 4)	0.179
Global tibia	0.4 ± 0.9	0 (−2, 3)	0.2 ± 0.7	0 (−2, 2)	0.107
Medial compartment	0.7 ± 1.6	1 (−3, 6)	0.2 ± 1.5	0 (−3, 4)	0.091
Lateral compartment	0.4 ± 1.1	0 (−2, 4)	0.2 ± 1.5	0 (−4, 5)	0.271
Medial femur	0.5 ± 1.2	0 (−2, 4)	0.2 ± 1.2	0 (−3, 3)	0.272
Lateral femur	0.3 ± 0.9	0 (−2, 2)	0.1 ± 1.2	0 (−3, 4)	0.189
Medial plateau	0.3 ± 0.7	0 (−1, 2)	0.0 ± 0.5	0 (−1, 1)	0.054
Lateral plateau	0.1 ± 0.5	0 (−2, 2)	0.1 ± 0.5	0 (−2, 2)	0.789

* Min = minimum; max = maximum.
† *P* values were assessed by the Wilcoxon-Mann-Whitney test.

accepted gold standard, in all of the patients in cohort 2 (n = 143) was −0.530 and −0.541, respectively, which is nearly 3 times lower compared to the corresponding cartilage volume change in the medial compartment, which was −1.473.

Drug treatment evaluation using cartilage defects and cartilage volume loss. Since neither the GRE nor the IW-FSE sequence demonstrated superiority in assessing cartilage defect change over time, we pursued our investigation using cohort 2, for which only a GRE sequence was performed, by comparing the sensitivity to detect treatment effects on cartilage of semiquantitative (cartilage defect) and quantitative (cartilage volume loss) methods.

The cartilage defect scores revealed no statistically significant difference between the treatment groups at baseline and 2 years (data not shown) or for the change at 2 years (Table 2). Of note, a trend toward significance ($P < 0.054$) was found in the medial tibial plateau (Table 2). As shown in Table 3, significant differences in favor of the licofelone group were found for the global knee and femur, the lateral compartment, and the lateral femur.

The SRMs for the treatment groups combined were consistently and markedly higher for the cartilage volume loss assessment than for the defect scoring (minimum, maximum 1.10, 2.0 and 0.18, 0.37, respectively) by 4.3- to 6.0-fold.

Table 3. Percent cartilage volume loss at 2 years according to treatment (cohort 2) in all patients

	Naproxen (n = 73), mean ± SD	Licofelone (n = 70), mean ± SD	*P**
Global knee†	−7.4 ± 3.3	−5.9 ± 3.2	0.009
Global femur	−6.6 ± 3.6	−5.1 ± 3.6	0.015
Global tibia	−9.5 ± 5.1	−8.2 ± 4.5	0.105
Medial compartment†	−8.8 ± 5.8	−7.5 ± 5.2	0.155
Lateral compartment†	−6.3 ± 4.1	−4.8 ± 4.1	0.026
Medial femur	−8.1 ± 6.4	−6.6 ± 6.1	0.159
Lateral femur	−5.4 ± 3.9	−3.6 ± 4.4	0.014
Medial plateau	−10.6 ± 7.0	−9.2 ± 7.7	0.257
Lateral plateau	−8.9 ± 8.3	−7.6 ± 6.6	0.292

* *P* values were assessed by 2-sided independent-sample *t*-tests.
† Data for the global knee and the medial and lateral compartments were previously published (13).

DISCUSSION

This is the first head-to-head study in knee OA patients comparing the sensitivity to change of a cartilage defect scoring system using 2 different types of MRI sequences, a high-resolution GRE sequence and a water-sensitive IW-FSE sequence. Data revealed that neither of the MRI sequences demonstrated superiority in the assessment of cartilage defect change over time. Moreover, comparison between treatment groups in a cohort of patients with knee OA demonstrated that assessing cartilage volume loss provides a much greater sensitivity to treatment response globally than the semiquantitative scoring method. This study has provided important information helping to answer 2 very important questions that are relevant not only to longitudinal studies, but more importantly to DMOAD clinical trials involving the assessment of cartilage changes over time between treatment groups, and therefore response to treatment.

The 2 methods assessed in the present study using MRI technology, cartilage volumetry and a semiquantitative scoring system, have previously shown good reliability (6,7,31). However, questions remained, particularly in the context of clinical trials, as to the sensitivity of each of these methods to assess the cartilage changes over time and, importantly, the effects of treatment in an OA patient cohort, as well as the impact of MRI sequences used to evaluate the cartilage defect changes. First, the results of the present study revealed that, in cross-sectional analysis, the water-sensitive IW-FSE sequence provided consistently higher scores for the cartilage defect assessment, as seen in a similar recent study (17). This may result from different physical phenomena, including partial volume artifact and chemical shift artifact. With an IW-FSE sequence, the voxels located at the interface between the superficial cartilage and articular fluid may be registered

as being fluid only due to the high intensity of the partial fluid volume, which may outshine the partial cartilage volume. This partial volume artifact is more pronounced when using a larger voxel size, i.e., lower spatial resolution, which is usually the case for IW-FSE sequences compared to GRE sequences in most protocols, including those performed in our study. In contrast to their superior spatial resolution, GRE sequences have an inherent inferior contrast resolution that is even lower when using a small voxel size. This may obscure the margins between cartilage and articular fluid and lead the reader to overlook cartilage defects and overestimate the cartilage thickness, therefore resulting in a lower defect score. The depiction of the interface between deep cartilage and bone may be influenced by partial volume effect as well. This artifact was probably minimal in our study, since our IW sequences were FSE and fat suppressed and some of them were performed with an HF phase-encoding direction, but this also may have contributed to variations in appearance of the cartilage interface.

Interestingly, although the score obtained with the IW-FSE sequence is higher than with the GRE sequence and the precision of grading is allegedly better in the GRE sequence (22,24), neither sequence demonstrated a superior sensitivity to detect cartilage defect change over 2 years, since the SRM values were similar between the 2 sequences. These present data do not support the finding that for the scoring of cartilage defects, the water-sensitive FSE sequences are superior to the high-resolution GRE sequences (20,38). Results could have possibly been different if IW-FSE sequences were performed with a higher spatial resolution (i.e., with the same voxel size as for the GRE sequences). However, such an approach, particularly in the context of a clinical trial, would not have been convenient using a 1.5T scanner, since it would have very significantly increased the acquisition time. Such a long acquisition time would be inconvenient for the patients, resulting in higher dropout rates in longitudinal trials. 3T scanners, which are unfortunately not yet readily available, can achieve sufficiently high resolution (as good as with the GRE sequences in 1.5T scanners) in an acceptable acquisition time, but with substantially higher costs. Therefore, the present findings are of interest in that a GRE sequence may suffice to effectively assess cartilage change with either quantitative (cartilage volume loss) or semiquantitative (cartilage defect) means.

With regard to the sensitivity to detect treatment effect using the semiquantitative versus the quantitative method, the present results showed a statistically significant difference in the percentage of cartilage volume loss between the 2 treatment groups in the global knee, global femur, lateral compartment, and lateral femur (13), whereas for the cartilage defects, a close to significant difference was observed only in the medial plateau. In addition, the semiquantitative assessment demonstrated a lower sensitivity to change than the quantitative evaluation of the cartilage volume loss indicated by lower SRM values (4.3–6-fold for the global knee and subregions). This is well in line with the sensitivity to change of the current gold standard JSN, which was shown to be inferior to the cartilage volumetry, as well as with a recently published study using 3T MRI that demonstrated the superiority of quantitative measurements over semiquantitative scores to assess change over time (39).

Among many possible explanations for the difference between these methods is that they may not completely address similar structural changes and/or the stage/severity of the disease. For the structural changes, the cartilage volume evaluation is a combination of both the diffuse thinning of the tissue and the focal changes, representing a more global view. In contrast, the cartilage defect scoring is more focused on lesions, which are often concentrated on weight-bearing areas of cartilage, and therefore do not fully account for the diffuse tissue loss. One may believe that scoring the lesional areas could result in a more sensitive detection of treatment effect by the scoring system than a more inclusive evaluation, i.e., the volumetry method. One may find an answer to this most important question in a recent post hoc analysis of the licofelone study published recently (32). This report showed that the assessment of cartilage loss performed in subregions, including the weight-bearing regions of condyles and plateaus, was much less sensitive to assess changes over time and response to treatment. This finding was mainly related to the greater interpatient variability. This could likely apply to the semiquantitative scoring system.

The limitations of this study include a small sample size for the cartilage defect assessment in 2 different sequences and the comparison of these results, in addition to the threshold for significance not being corrected for multiple comparisons. However, the fact that we could identify a significant difference in the cross-sectional view leads us to believe that the power was sufficient. Our technique does not allow for the assessment of the patellar cartilage because the volumetric assessment as developed (10) does not permit this measurement. Moreover, the cartilage defect assessment using the IW-FSE had slice thicknesses of up to 3.5 mm, which might have led to an underestimation of cartilage defects but also prevented an overreading of minor defects. However, the slice thicknesses of up to 4 mm are in use in the main semiquantitative scoring systems (3,5,9). Furthermore, there was no histopathologic control to confirm the findings. This is simply not possible in human clinical trials addressing longitudinal views. In addition, the choice of the parameter for the assessment of sensitivity to change (SRM) may have influenced the findings in favor of the measurement technique with the higher precision. However, the use of multiple parameters for the same data is not recommended (37,40).

In conclusion, the findings of this study do not suggest superior sensitivity of one sequence over the other in evaluating the progression of cartilage defects over time. In the context of clinical trials assessing differences between treatment groups in patients with knee OA, the quantitative cartilage volume evaluation appeared to be a more sensitive method for identifying the treatment effect. As the field of musculoskeletal imaging using MRI is rapidly evolving, new sequences may eventually be developed for such purposes that are superior to those used in this study.

ACKNOWLEDGEMENTS

The authors thank Virginia Wallis for her assistance with the manuscript preparation and André Pelletier for the MRI reading.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Dr. Pelletier had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Wildi, Martel-Pelletier, Raynauld, Pelletier.

Acquisition of data. Martel-Pelletier, Abram, Pelletier.

Analysis and interpretation of data. Wildi, Martel-Pelletier, Abram, Moser, Raynauld, Pelletier.

ROLE OF THE STUDY SPONSOR

Merckle and ArthroLab had no role in the study design, data collection, data analysis, and writing of the manuscript, as well as approval of the content of the submitted manuscript. Publication of this article was not contingent on the approval of Merckle and ArthroLab.

REFERENCES

1. Buckland-Wright JC, Ward RJ, Peterfy C, Mojcik CF, Leff RL. Reproducibility of the semiflexed (metatarsophalangeal) radiographic knee position and automated measurements of medial tibiofemoral joint space width in a multicenter clinical trial of knee osteoarthritis. J Rheumatol 2004;31:1588–97.
2. Guermazi A, Burstein D, Conaghan P, Eckstein F, Hellio Le Graverand-Gastineau MP, Keen H, et al. Imaging in osteoarthritis. Rheum Dis Clin North Am 2008;34:645–87.
3. Peterfy CG, Guermazi A, Zaim S, Tirman PF, Miaux Y, White D, et al. Whole-Organ Magnetic Resonance Imaging Score (WORMS) of the knee in osteoarthritis. Osteoarthritis Cartilage 2004;12:177–90.
4. Kornaat PR, Ceulemans RY, Kroon HM, Riyazi N, Kloppenburg M, Carter WO, et al. MRI assessment of knee osteoarthritis: Knee Osteoarthritis Scoring System (KOSS)—inter-observer and intra-observer reproducibility of a compartment-based scoring system. Skeletal Radiol 2005;34:95–102.
5. Hunter DJ, Lo GH, Gale D, Grainger AJ, Guermazi A, Conaghan PG. The reliability of a new scoring system for knee osteoarthritis MRI and the validity of bone marrow lesion assessment: BLOKS (Boston Leeds Osteoarthritis Knee Score). Ann Rheum Dis 2008;67:206–11.
6. Lynch JA, Roemer FW, Nevitt MC, Felson DT, Niu J, Eaton CB, et al. Comparison of BLOKS and WORMS scoring systems part I. Cross sectional comparison of methods to assess cartilage morphology, meniscal damage and bone marrow lesions on knee MRI: data from the Osteoarthritis Initiative. Osteoarthritis Cartilage 2010;18:1393–401.
7. Felson DT, Lynch J, Guermazi A, Roemer FW, Niu J, McAlindon T, et al. Comparison of BLOKS and WORMS scoring systems part II. Longitudinal assessment of knee MRIs for osteoarthritis and suggested approach based on their performance: data from the Osteoarthritis Initiative. Osteoarthritis Cartilage 2010;18:1402–7.
8. Ding C, Garnero P, Cicuttini F, Scott F, Cooley H, Jones G. Knee cartilage defects: association with early radiographic osteoarthritis, decreased cartilage volume, increased joint surface area and type II collagen breakdown. Osteoarthritis Cartilage 2005;13:198–205.
9. Hunter DJ, Guermazi A, Lo GH, Grainger AJ, Conaghan PG, Boudreau RM, et al. Evolution of semi-quantitative whole joint assessment of knee OA: MOAKS (MRI Osteoarthritis Knee Score). Osteoarthritis Cartilage 2011;19:990–1002.
10. Kauffmann C, Gravel P, Godbout B, Gravel A, Beaudoin G, Raynauld JP, et al. Computer-aided method for quantification of cartilage thickness and volume changes using MRI: validation study using a synthetic model. IEEE Trans Biomed Eng 2003;50:978–88.
11. Raynauld JP, Martel-Pelletier J, Berthiaume MJ, Labonte F, Beaudoin G, de Guise JA, et al. Quantitative magnetic resonance imaging evaluation of knee osteoarthritis progression over two years and correlation with clinical symptoms and radiologic changes. Arthritis Rheum 2004;50:476–87.
12. Eckstein F, Cicuttini F, Raynauld JP, Waterton JC, Peterfy C. Magnetic resonance imaging (MRI) of articular cartilage in knee osteoarthritis (OA): morphological assessment. Osteoarthritis Cartilage 2006;14:A46–75.
13. Raynauld JP, Martel-Pelletier J, Bias P, Laufer S, Haraoui B, Choquette D, et al. Protective effects of licofelone, a 5-lipoxygenase and cyclo-oxygenase inhibitor, versus naproxen on cartilage loss in knee osteoarthritis: a first multicentre clinical trial using quantitative MRI. Ann Rheum Dis 2009;68:938–47.
14. Wildi LM, Raynauld JP, Martel-Pelletier J, Beaulieu A, Bessette L, Morin F, et al. Chondroitin sulphate reduces both cartilage volume loss and bone marrow lesions in knee osteoarthritis patients starting as early as 6 months after initiation of therapy: a randomised, double-blind, placebo-controlled pilot study using MRI. Ann Rheum Dis 2011;70:982–9.
15. Raynauld JP, Martel-Pelletier J, Beaulieu A, Bessette L, Morin F, Choquette D, et al. An open-label pilot study evaluating by magnetic resonance imaging the potential for a disease-modifying effect of celecoxib compared to a modelized historical control cohort in the treatment of knee osteoarthritis. Semin Arthritis Rheum 2010;40:185–92.
16. Peterfy CG, Gold G, Eckstein F, Cicuttini F, Dardzinski B, Stevens R. MRI protocols for whole-organ assessment of the knee in osteoarthritis. Osteoarthritis Cartilage 2006;14 Suppl: A95–111.
17. Roemer FW, Kwoh CK, Hannon MJ, Crema MD, Moore CE, Jakicic JM, et al. Semiquantitative assessment of focal cartilage damage at 3T MRI: a comparative study of dual echo at steady state (DESS) and intermediate-weighted (IW) fat suppressed fast spin echo sequences. Eur J Radiol 2011;80:126–31.
18. Mohr A. The value of water-excitation 3D FLASH and fat-saturated PDw TSE MR imaging for detecting and grading articular cartilage lesions of the knee. Skeletal Radiol 2003;32:396–402.
19. Stahl R, Luke A, Ma CB, Krug R, Steinbach L, Majumdar S, et al. Prevalence of pathologic findings in asymptomatic knees of marathon runners before and after a competition in comparison with physically active subjects: a 3.0 T magnetic resonance imaging study. Skeletal Radiol 2008;37:627–38.
20. Hayashi D, Roemer FW, Guermazi A. Choice of pulse sequences for magnetic resonance imaging–based semiquantitative assessment of cartilage defects in osteoarthritis research: comment on the article by Doré et al [letter]. Arthritis Rheum 2010;62:3830–1.
21. Dore D, Winzenberg T, Ding C, Cicuttini F, Jones G. Reply [letter]. Arthritis Rheum 2010;62:381–2.
22. Masi JN, Sell CA, Phan C, Han E, Newitt D, Steinbach L, et al. Cartilage MR imaging at 3.0 versus that at 1.5 T: preliminary results in a porcine model. Radiology 2005;236:140–50.
23. Yoshioka H, Stevens K, Hargreaves BA, Steines D, Genovese M, Dillingham MF, et al. Magnetic resonance imaging of articular cartilage of the knee: comparison between fat-suppressed three-dimensional SPGR imaging, fat-suppressed FSE imaging, and fat-suppressed three-dimensional DEFT imaging, and correlation with arthroscopy. J Magn Reson Imaging 2004;20:857–64.
24. Lee SY, Jee WH, Kim SK, Koh IJ, Kim JM. Differentiation between grade 3 and grade 4 articular cartilage defects of the knee: fat-suppressed proton density-weighted versus fat-suppressed three-dimensional gradient-echo MRI. Acta Radiol 2010;51:455–61.
25. Bauer JS, Barr C, Henning TD, Malfair D, Ma CB, Steinbach L,

et al. Magnetic resonance imaging of the ankle at 3.0 Tesla and 1.5 Tesla in human cadaver specimens with artificially created lesions of cartilage and ligaments. Invest Radiol 2008;43: 604–11.
26. Cooper C, Reginster JY, Chapurlat R, Christiansen C, Genant H, Bellamy N, et al. Efficacy and safety of oral strontium ranelate for the treatment of knee osteoarthritis: rationale and design of randomised, double-blind, placebo-controlled trial. Curr Med Res Opin 2012;28:231–9.
27. Altman R, Asch E, Bloch D, Bole G, Borenstein D, Brandt K, et al. Development of criteria for the classification and reporting of osteoarthritis: classification of osteoarthritis of the knee. Arthritis Rheum 1986;29:1039–49.
28. Wildi LM, Raynauld JP, Martel-Pelletier J, Abram F, Dorais M, Pelletier JP. Relationship between bone marrow lesions, cartilage loss and pain in knee osteoarthritis: results from a randomised controlled clinical trial using MRI. Ann Rheum Dis 2010;69:2118–24.
29. Raynauld JP, Martel-Pelletier J, Haraoui B, Choquette D, Dorais M, Wildi LM, et al. Risk factors predictive of joint replacement in a 2-year multicentre clinical trial in knee osteoarthritis using MRI: results from over 6 years of observation. Ann Rheum Dis 2011;70:1382–8.
30. Eckstein F, Hudelmaier M, Wirth W, Kiefer B, Jackson R, Yu J, et al. Double echo steady state magnetic resonance imaging of knee articular cartilage at 3 Tesla: a pilot study for the Osteoarthritis Initiative. Ann Rheum Dis 2006;65:433–41.
31. Raynauld JP, Kauffmann C, Beaudoin G, Berthiaume MJ, de Guise JA, Bloch DA, et al. Reliability of a quantification imaging system using magnetic resonance images to measure cartilage thickness and volume in human normal and osteoarthritic knees. Osteoarthritis Cartilage 2003;11:351–60.
32. Raynauld JP, Martel-Pelletier J, Abram F, Dorais M, Haraoui B, Choquette D, et al. Analysis of the precision and sensitivity to change of different approaches to assess cartilage loss by quantitative MRI in a longitudinal multicentre clinical trial in patients with knee osteoarthritis. Arthritis Res Ther 2008;10: R129.
33. Raynauld JP, Martel-Pelletier J, Berthiaume MJ, Beaudoin G, Choquette D, Haraoui B, et al. Long term evaluation of disease progression through the quantitative magnetic resonance imaging of symptomatic knee osteoarthritis patients: correlation with clinical symptoms and radiographic changes. Arthritis Res Ther 2006;8:R21.
34. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.
35. Hudelmaier M, Wirth W, Wehr B, Kraus V, Wyman BT, Hellio Le Graverand MP, et al. Femorotibial cartilage morphology: reproducibility of different metrics and femoral regions, and sensitivity to change in disease. Cells Tissues Organs 2010; 192:340–50.
36. Conrozier T, Brandt K, Piperno M, Mathieu P, Merle-Vincent F, Vignon E. Reproducibility and sensitivity to change of a new method of computer measurement of joint space width in hip osteoarthritis: performance of three radiographic views obtained at a 3-year interval. Osteoarthritis Cartilage 2009;17: 864–70.
37. Stratford PW, Riddle DL. Assessing sensitivity to change: choosing the appropriate change coefficient. Health Qual Life Outcomes 2005;3:23.
38. Hayashi D, Guermazi A, Hunter DJ. Osteoarthritis year 2010 in review: imaging. Osteoarthritis Cartilage 2011;19:354–60.
39. Stahl R, Jain SK, Lutz J, Wyman BT, Le Graverand-Gastineau MP, Vignon E, et al. Osteoarthritis of the knee at 3.0 T: comparison of a quantitative and a semi-quantitative score for the assessment of the extent of cartilage lesion and bone marrow edema pattern in a 24-month longitudinal study. Skeletal Radiol 2011;40:1315–27.
40. Middel B, van Sonderen E. Statistical significant change versus relevant or important change in (quasi) experimental design: some conceptual and methodological problems in estimating magnitude of intervention-related change in health services research. Int J Integr Care 2002;2:e15.