

Routine clinical knee MR reports: comparison of diagnostic performance at 1.5 T and 3.0 T for assessment of the articular cartilage

Jacob C. Mandell¹ · Jeffrey A. Rhodes¹ · Nehal Shah¹ · Glenn C. Gaviola¹ · Andreas H. Gomoll² · Stacy E. Smith¹

Received: 13 April 2017 / Revised: 28 May 2017 / Accepted: 3 July 2017 / Published online: 17 July 2017
© ISS 2017

Abstract

Objective Accurate assessment of knee articular cartilage is clinically important. Although 3.0 Tesla (T) MRI is reported to offer improved diagnostic performance, literature regarding the clinical impact of MRI field strength is lacking. The purpose of this study is to compare the diagnostic performance of clinical MRI reports for assessment of cartilage at 1.5 and 3.0 T in comparison to arthroscopy.

Materials and methods This IRB-approved retrospective study consisted of 300 consecutive knees in 297 patients who had routine clinical MRI and arthroscopy. Descriptions of cartilage from MRI reports of 165 knees at 1.5 T and 135 at 3.0 T were compared with arthroscopy. The sensitivity, specificity, percent of articular surfaces graded concordantly, and percent of articular surfaces graded within one grade of the arthroscopic grading were calculated for each articular surface at 1.5 and 3.0 T. Agreement between MRI and arthroscopy was calculated with the weighted-kappa statistic. Significance testing was performed utilizing the z-test after bootstrapping to obtain the standard error.

Results and conclusions The sensitivity, specificity, percent of articular surfaces graded concordantly, and percent of articular surfaces graded within one grade were 61.4%, 82.7%, 62.2%, and 77.5% at 1.5 T and 61.8%, 80.6%, 59.5%, and 75.6% at

3.0 T, respectively. The weighted kappa statistic was 0.56 at 1.5 T and 0.55 at 3.0 T. There was no statistically significant difference in any of these parameters between 1.5 and 3.0 T. Factors potentially contributing to the lack of diagnostic advantage of 3.0 T MRI are discussed.

Keywords Articular cartilage · MR field strength · Clinical MRI reports · Diagnostic performance · Knee

Introduction

Accurate interpretation of knee articular cartilage MR imaging is clinically important. Osteoarthritis and post-traumatic cartilage defects are common and contribute to increasing medical costs [1]. There is a wide range of non-arthroplasty surgical treatment options for cartilage disease [2, 3] including autologous chondrocyte implantation, osteochondral auto- and allografting, and marrow stimulation. Lifestyle modification and medical treatment options may also slow the onset of osteoarthritis [4, 5]. MRI is one of the most commonly used modalities for the assessment of the articular cartilage of the knee [6, 7] because of its excellent soft-tissue contrast, although the reported sensitivity of MRI to detect cartilage abnormalities is lower than that of other internal derangements of the knee, such as ligamentous or meniscal injuries [6].

There have been multiple previous experimental, retrospective, and prospective studies demonstrating the diagnostic performance of MRI in assessment of articular cartilage utilizing conventional 2D fast-spin-echo sequences [8–13] as well as of specialized techniques including biochemical imaging [14], 3D gradient sequences [15, 16], and 3D double echo steady state (DESS) [17, 18]. While many of these specialized techniques are not in widespread clinical use, 3.0 Tesla (T)

✉ Jacob C. Mandell
jmandell@partners.org

¹ Division of Musculoskeletal Imaging and Intervention, Department of Radiology, Brigham and Women's Hospital, 75 Francis St, Boston, MA 02115, USA

² Cartilage Repair Center, Department of Orthopedic Surgery, Brigham and Women's Hospital, 75 Francis St, Boston, MA 02115, USA

magnets are increasingly common in routine clinical practice, and 3.0 T MRI has been advocated as offering better diagnostic performance in comparison to 1.5 T MRI for assessment of knee articular cartilage [19–23]. The advantages of 3.0 T MRI include improved spatial resolution and contrast resolution, without increasing acquisition time. While prior studies in the established literature on cartilage imaging have utilized dedicated readers to evaluate the articular cartilage, no previous studies have assessed the diagnostic performance of routine clinical reports to evaluate the articular cartilage. The purpose of this study is to retrospectively compare the diagnostic performance of routine clinical knee MRI reports in the evaluation of articular cartilage at 1.5 and 3.0 T, with arthroscopy as the standard of reference.

Materials and methods

This retrospective study was performed with approval from our institutional review board with waiver of informed consent and was in compliance with Health Insurance Portability and Accountability Act regulations.

Study group

A sample size of 1372 articular surfaces (229 knees with 6 articular surfaces per knee) was calculated for a 10% confidence interval width and 95% confidence level, presuming a prevalence of articular lesions of 28% based on a study of 31,516 knee arthroscopies [24]; this sample size was then increased to 1800 articular surfaces (300 knees). The study consists of MRI and arthroscopy reports of 300 knees from 297 consecutive symptomatic patients who underwent arthroscopic knee surgery from 12 March 2012 to 29 March 2013 and who also had a routine (non-contrast, non-arthrogram) MRI examination of the same knee performed at the same institution within 365 days prior to arthroscopic surgery. Three patients had bilateral knee MRI and arthroscopic surgery. The list of consecutive patients who underwent knee arthroscopy was obtained from searching procedure codes in the radiology Picture Archiving and Communication System (PACS), which is also used to store photographic

arthroscopic images. Patients were excluded from the study if non-routine MRI was performed (e.g., direct or indirect arthrogram or intravenous contrast-enhanced examination) or if the patient had undergone a previous cartilage repair surgery. Of the 297 patients included in the study (Table 1), there were 161 males and 136 females (mean age 42.8 years, range 15–83 years). There were 165 knees that underwent MRI at 1.5 T from 163 patients and 135 knees that underwent MRI at 3.0 T from 134 patients. The mean time between MRI and arthroscopy was 68.2 days (range: 1–330 days) for all patients, with the mean for 1.5 T of 73.6 days (range: 1–330 days) and the mean for 3.0 T of 59.0 days (range: 7–252 days).

MR scanning

MR imaging was performed on one of six clinical MRI scanners (four 1.5 T and two 3.0 T) manufactured by Siemens and General Electric Healthcare. The decision to perform imaging at 1.5 or 3 T was determined by a combination of scanner availability, patient location preference, and referrer physician preference. The imaging protocols were not specifically tailored toward imaging cartilage but were instead designed as a standard, comprehensive assessment of the knee joint for evaluation of internal derangement. Imaging protocols vary slightly by scanner; however, the standard institutional protocol includes the following conventional 2D fast-spin-echo sequences: axial T2-weighted or PD-weighted images with fat suppression; sagittal and coronal fast-spin-echo proton-density weighted images with fat suppression; coronal and sagittal fast-spin-echo proton-density weighted images without fat suppression. Additionally, a 3D isotropic sequence with proton-density weighting without fat suppression was performed in the majority of examinations [Siemens Sampling Perfection with Application optimized Contrast using different flip angle Evolution (SPACE)], which was acquired in the sagittal plane and reformatted in axial, coronal, and coronal-oblique trochlear planes. Of the examinations performed at 1.5 T, 105/165 (63.6%) included the 3D isotropic sequence, as did 118/135 (87.4%) of the examinations performed at 3.0 T.

The number of examinations performed on each specific scanner, as well as the proportion of examinations

Table 1 Number, sex, age, and days between MR and arthroscopy for the study group at 1.5 and 3.0 T

	Number and sex	Age (range, standard deviation)	Days between MR and arthroscopy (range, standard deviation)
1.5 T	163 (91 male, 72 female)	41.6 years (16–83; 15.0)	73.6 days (1–330; 74.2)
3.0 T	134 (70 male, 64 female)	44.3 years (15–69; 15.2)	59.0 days (7–252; 52.1)
Total	297 (161 male, 136 female)	42.8 years	68.2 days

including the 3D isotropic sequence per machine, is demonstrated in Table 2. Typical scanning parameters for the most commonly utilized MR machines at 3.0 T (scanner 1) and 1.5 T (scanner 2) are shown in Tables 3 and 4, respectively. Together, these two Siemens scanners comprised 231/300 of the knee MRIs included in this study. There are a few notable differences between the sequences and hardware of these two scanners in addition to differences in field strength. Specifically, the routine sequence for scanner 1 (3.0 T) included an axial T2-weighted sequence with fat suppression, while the axial fluid-sensitive sequence for scanner 2 (1.5 T) was PD-weighted with fat suppression. Additionally, scanner 1 (3.0 T) utilized an eight-channel transmit-receive coil, while the coil for scanner 2 (1.5 T) was a 15-channel transmit-receive unit. The typical scan duration for scanner 1 (3.0 T) was 22 min in comparison to 24 min for scanner 2 (1.5 T).

Correlation of articular surfaces from MRI reports and arthroscopic surgery

All knee MRI examinations were interpreted by one of eight fellowship-trained musculoskeletal radiologists, with 5 to 41 years of experience (mean: 15.4 years, standard deviation: 15.9 years). Most examinations were interpreted in conjunction with a radiology resident or musculoskeletal imaging fellow. Knee arthroscopy was performed by one of nine orthopedic surgeons, with 6 to 25 years of experience (mean 15.0 years, standard deviation 7.2 years). Arthroscopy was performed with standard medial and lateral joint line portals and a medial suprapatellar inflow portal, and complete diagnostic assessment of the joint was performed per routine practice, which was recorded in the clinical operative note.

In both the MRI reports and operative notes from arthroscopic surgery, the articular surfaces of the knee were divided into six regions (patella, trochlea, medial femoral condyle, medial tibial plateau, lateral femoral condyle, and lateral tibial plateau). The MRI reports were structured, and the arthroscopy reports were freeform. All data for this study were only obtained from MRI and arthroscopy reports; neither images from the MR nor stored

images from arthroscopies were evaluated routinely. MRI and arthroscopic descriptions of each articular surface were both classified using a modified Outerbridge classification, which is the system used by the orthopedic surgeons at this institution (Fig. 1). The adaptation of the Outerbridge classification is defined as follows: grade 0 is normal; grade I is heterogeneous signal (correlating to cartilage softening at arthroscopy); grade II is superficial fissuring, thinning, or defect <50% in depth; grade III is deep fissuring, thinning, or defect >50% in depth; grade IV is a full thickness defect. All musculoskeletal radiologists in the department are familiar with this staging system, and all routinely use descriptive terminology to describe the articular cartilage that would allow easy conversion to the Outerbridge classification. If two or more lesions were present on a single articular surface, the highest-graded lesion was recorded. If a specific articular surface was not explicitly described in the arthroscopy report or if several articular surfaces were described with a blanket statement such as “the remainder of the articular surfaces appeared normal,” then these unnamed articular surfaces were presumed normal. All data were entered into an Excel (Microsoft Corp., Seattle, WA) spreadsheet.

Statistical analysis

Statistics were performed in R version 3.3.2 (R Foundation for Statistical Computing, Vienna, Austria). Two-sample t-tests were performed to compare the patient ages and days between MRI and arthroscopy in the two groups who underwent MRI at 1.5 and 3.0 T. Pearson’s chi-squared test was performed to evaluate the sex distribution between 1.5- and 3.0-T groups, the proportion of examinations at 1.5 and 3.0 T that included a 3D isotropic sequence, and the proportion of examinations at 1.5 and 3.0 T read by each radiologist.

To determine sensitivity and specificity, articular cartilage was divided into “disease negative” (grade 0) and “disease positive” (grades I–IV). Sensitivity and specificity with 95% confidence intervals were calculated for each articular surface at 1.5 and 3.0 T. The percentages of articular surfaces graded concordantly and within one grade of the arthroscopic grade

Table 2 Number of examinations performed on each specific scanner and proportion of examinations including a 3D isotropic sequence

Scanner	Field strength (T)	Manufacturer	Number of exams	Number including 3D sequence
Scanner 1	3.0	Siemens	125	121/125
Scanner 2	1.5	Siemens	106	102/106
Scanner 3	1.5	General Electric	48	0/48
Scanner 4	3.0	General Electric	12	0/12
Scanner 5	1.5	Siemens	5	0/5
Scanner 6	1.5	Siemens	4	0/4

Table 3 Specific scanning parameters for the most commonly utilized 3.0-T MRI machine (scanner 1 in Table 2). An eight-channel transmit-receive coil was utilized. PD = proton density; FS = fat saturated;

SPACE = Sampling Perfection with Application optimized Contrast using different flip angle Evolution; NEX = number of excitations

Sequence	Slice thickness	Repetition time	Echo time	NEX	Matrix
Ax T2FS	3.0 mm	3800	78	1	384 × 288
Cor PDFS	3.0 mm	3180	33	2	384 × 288
Cor PD	3.0 mm	2760	33	2	384 × 288
Sag PDFS	3.0 mm	3530	36	1	384 × 288
Sag PD	3.0 mm	3500	36	1	384 × 288
Sag PD SPACE (3D)	0.6 mm	1100	35	1	256 × 254

were calculated for both field strengths and for each articular surface, with 95% confidence intervals. For each radiologist reading a significantly different proportion of examinations at 1.5 and 3.0 T, the per-radiologist sensitivity and specificity were calculated with 95% confidence intervals. Additionally, the knees were divided into two equal groups based on the time between arthroscopy and imaging, and sensitivity and specificity with 95% confidence intervals were calculated for knees with arthroscopy performed both early and late.

The agreement between the grading of articular cartilage in MR reports and arthroscopy was evaluated with the weighted kappa statistic with 95% confidence intervals, which was interpreted according to the recommendations of Landis and Koch [25] as follows: kappa 0.00–0.20, slight agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; 0.81–1.00, almost perfect agreement.

Significance testing was performed by bootstrapping individual articular surfaces to obtain a standard error, after which the Z-test was performed to obtain *p*-values. Significance testing was performed for all articular surfaces between the 1.5 and 3.0 T groups (rather than subdivided into multiple individual analyses of each articular surface or by individual radiologists) to maximize statistical power and to reduce the probability of obtaining a spuriously significant *p*-value. Sensitivity, specificity,

percentage of articular surfaces graded concordantly, percentage of articular surfaces graded within one grade, and kappa values were compared between the 1.5 and 3.0 T groups in this method. Significance testing was also performed between the groups with arthroscopy performed early and late. A *p*-value of <0.05 was considered statistically significant for all tests.

Results

Comparison of the 1.5 and 3.0 T groups

There was no significant difference between the 1.5 and 3.0 T groups for patient age (*p* = 0.13), sex distribution (*p* = 0.74), or time between MR and arthroscopy (*p* = 0.06). A significantly lower proportion of examinations included the 3D isotropic sequence at 1.5 T (63.6%) in comparison to 3.0 T (87.4%) (*p* = 5.2×10^{-6}).

Comparison of diagnostic performance at 1.5 and 3.0 T






There were a total of 674 cartilage lesions in 1800 articular surfaces, for a prevalence of 37.4%. Of 1800 articular surfaces, 1532 were explicitly described in the arthroscopy reports, and the remainder were either described as normal in a

Table 4 Specific scanning parameters for the most commonly utilized 1.5-T MRI machine (scanner 2 in Table 2). A 15-channel transmit-receive coil was utilized. PD = proton density; FS = fat saturated;

SPACE = Sampling Perfection with Application optimized Contrast using different flip angle Evolution; NEX = number of excitations

Sequence	Slice thickness	Repetition time	Echo time	NEX	Matrix
Ax PDFS	3.5 mm	2000	25	1	256 × 256
Cor PDFS	3.0 mm	2080	35	1	384 × 257
Cor PD	3.0 mm	2000	23	1	384 × 257
Sag PDFS	3.5 mm	3190	34	1	384 × 257
Sag PD	3.5 mm	3070	27	1	512 × 256
Sag PD SPACE (3D)	0.6 mm	1150	40	1	256 × 244

Fig. 1 Illustration of MRI and arthroscopic grading of cartilage lesions, per the modified Outerbridge classification

	Grade	MRI findings	Arthroscopy findings
	0	normal	normal
	1	heterogeneous signal	softening
	2	superficial defect or fissuring < 50%	superficial defect or fissuring < 50%
	3	deep defect or fissuring > 50%	deep defect or fissuring > 50%
	4	full-thickness defect	full-thickness defect

blanket statement or presumed normal. There were 365 lesions in the group scanned at 1.5 T (prevalence of 36.9%) and 309 lesions in the group scanned at 3.0 T (prevalence of 38.1%). The distribution of cartilage lesions per grade per articular surface at both 1.5 and 3.0 T is given in Table 5.

The sensitivity, specificity, percent of articular surfaces graded exactly, and percent of articular surfaces graded within one grade are 61.4%, 82.7%, 62.2%, and 77.5% at 1.5 T, respectively, and 61.8%, 80.6%, 59.5%, and 75.6% at 3.0 T. For all articular surfaces at both field strengths, the sensitivity, specificity, percent of articular surfaces graded exactly, and percent of articular surfaces within one grade are 61.6%, 81.8%, 61.0%, and 76.6%, respectively. These parameters per articular surface at 1.5 and 3.0 T are shown in Table 6. Graphs of the sensitivity and specificity per articular surface at 1.5 and 3.0 T are shown in Figs. 2 and 3. Sensitivity was

highest for the patella (78.4% at 1.5 T and 80.5% at 3.0 T) and lowest for the medial tibial plateau (47.9% at 1.5 T and 36.8% at 3.0 T). Specificity was highest for the trochlea (88.3% at 1.5 T and 91.1% at 3.0 T) and lowest for the patella (70.0% at 1.5 T and 79.7% at 3.0 T). The 95% confidence intervals of sensitivity and specificity were widely overlapping for each articular surface at both 1.5 and 3.0 T.

When comparing all articular surfaces, there was no significant difference in sensitivity, specificity, percent of articular surfaces graded exactly, and percent of articular surfaces graded within one grade between the groups scanned at 1.5 and 3.0 T (Table 7).

The agreement between MRI reports and arthroscopy for all articular surfaces was moderate (weighted kappa of 0.56 at 1.5 T and 0.55 at 3.0 T). Weighted kappa values per articular surface are given in Table 8. There

Table 5 Number of cartilage lesions of each articular surface, with the prevalence in parentheses. MFC: medial femoral condyle; MTP: medial tibial plateau; LFC: lateral femoral condyle; LTP: lateral tibial plateau

	Normal	Grade 1	Grade 2	Grade 3	Grade 4	Total lesions
Patella 1.5 T	77 (46.7%)	12 (7.3%)	36 (21.8%)	22 (13.3%)	18 (10.9%)	88 (53.3%)
Patella 3.0 T	58 (43.0%)	14 (10.4%)	32 (23.7%)	25 (18.5%)	6 (4.4%)	77 (57.0%)
Trochlea 1.5 T	103 (62.4%)	5 (3.0%)	21 (12.7%)	16 (9.7%)	20 (12.1%)	62 (37.6%)
Trochlea 3.0 T	79 (58.5%)	8 (5.9%)	16 (11.9%)	20 (14.8%)	12 (8.9%)	56 (41.5%)
MFC 1.5 T	93 (56.4%)	4 (2.4%)	17 (10.3%)	38 (23.0%)	13 (7.9%)	72 (43.6%)
MFC 3.0 T	67 (49.6%)	6 (4.4%)	23 (17.0%)	29 (21.5%)	10 (7.4%)	68 (50.4%)
MTP 1.5 T	117 (70.9%)	9 (5.5%)	22 (13.3%)	14 (8.5%)	3 (1.8%)	48 (29.1%)
MTP 3.0 T	97 (71.9%)	8 (5.9%)	14 (10.4%)	10 (7.4%)	6 (4.4%)	38 (28.1%)
LFC 1.5 T	120 (72.9%)	4 (2.4%)	21 (12.7%)	10 (6.1%)	10 (6.1%)	45 (27.3%)
LFC 3.0 T	107 (79.3%)	6 (4.4%)	9 (6.7%)	3 (2.2%)	10 (7.4%)	28 (20.7%)
LTP 1.5 T	115 (69.7%)	6 (3.6%)	24 (14.5%)	12 (7.3%)	8 (4.8%)	50 (30.3%)
LTP 3.0 T	93 (68.9%)	11 (8.1%)	18 (13.3%)	7 (5.2%)	6 (4.4%)	42 (31.1%)
All 1.5 T	625 (63.1%)	40 (4.0%)	141 (14.2%)	112 (11.3%)	72 (7.3%)	365 (36.9%)
All 3.0 T	501 (61.9%)	53 (6.5%)	112 (13.8%)	94 (11.6%)	50 (6.2%)	309 (38.1%)
All	1126 (62.6%)	93 (5.2%)	253 (14.1%)	206 (11.4%)	122 (6.8%)	674 (37.4%)

Table 6 Sensitivity, specificity, percent of articular surfaces graded exactly, and percent of articular surfaces graded within one grade at 1.5 and 3.0 T (95% confidence intervals in parentheses)

	Sensitivity (%)		Specificity (%)		% Graded exactly		% Graded within 1 grade	
	1.5 T	3.0 T	1.5 T	3.0 T	1.5 T	3.0 T	1.5 T	3.0 T
Patella	78.4 (68–86)	80.5 (70–89)	70.0 (59–80)	70.7 (57–82)	49.1 (41–57)	45.9 (36–54)	77.6 (71–84)	70.4 (62–78)
Trochlea	56.5 (43–69)	57.1 (43–70)	88.3 (81–94)	91.1 (83–96)	67.9 (60–75)	61.5 (53–70)	79.4 (73–85)	74.1 (66–82)
MFC	63.9 (52–75)	73.5 (61–83)	79.6 (70–87)	70.1 (58–81)	55.2 (47–63)	51.1 (43–60)	73.3 (66–80)	76.3 (70–83)
MTP	47.9 (33–63)	36.8 (22–54)	85.5 (78–91)	79.4 (70–87)	69.1 (62–76)	61.5 (53–70)	79.4 (73–86)	73.3 (66–80)
LFC	55.6 (40–70)	50.0 (31–69)	82.5 (75–89)	80.4 (72–87)	65.5 (58–73)	70.4 (62–78)	74.5 (67–81)	77.0 (70–84)
LTP	52.0 (37–66)	45.2 (30–61)	86.1 (78–92)	87.1 (79–93)	66.7 (59–74)	66.7 (58–75)	80.6 (74–87)	82.2 (75–89)
All	61.4 (56–66)	61.8 (56–67)	82.7 (80–86)	80.6 (77–84)	62.2 (59–65)	59.5 (51–68)	77.5 (75–80)	75.6 (73–79)
	61.6 (58–65)		81.8 (79–84)		61.0 (59–63)		76.6 (75–79)	

was no significant difference between agreement at 1.5 or 3.0 T ($p = 0.659$).

Comparison of patients undergoing arthroscopy early and late

In comparing patients who underwent arthroscopy early (mean: 24.0 days, standard deviation: 9.8 days) and late (mean: 110.1 days, standard deviation: 69.2 days), there was no significant difference in sensitivity (early: 62.2%; late: 61.1%; $p = 0.658$) or specificity (early: 82.8%; late: 80.7%; $p = 0.267$).

Comparison of proportion of examinations read at 1.5 and 3.0 T per radiologist

The number of examinations read at 1.5 and 3.0 T for each radiologist is shown in Table 9. For each radiologist, Pearson’s chi-squared test was utilized to determine whether there was a significant difference in the number of examinations read at

each field strength. Three radiologists did read a significantly different proportion of examinations at 1.5 and 3.0 T; for these radiologists, individual sensitivity and specificity were calculated along with 95% confidence intervals (data also presented in Table 9). The sensitivities ranged from 55.0% to 69.8%; however, the 95% confidence intervals were overlapping. The specificities ranged from 78.2% to 83.2%, with widely overlapping 95% confidence intervals.

Discussion

In clinical practice, conventional 2D fast-spin-echo (FSE) sequences are performed most commonly for global assessment of the intra-articular structures of the knee, including evaluation of the articular cartilage. There is a wide reported range of diagnostic performance of 2D FSE MR for assessment of the knee cartilage, with sensitivity ranging from 26%–96%, specificity 50%–100%, and accuracy 49%–94% [8–13].

Fig. 2 Sensitivity of MR reports at 1.5 and 3.0 T per articular surface with error bars representing 95% confidence intervals

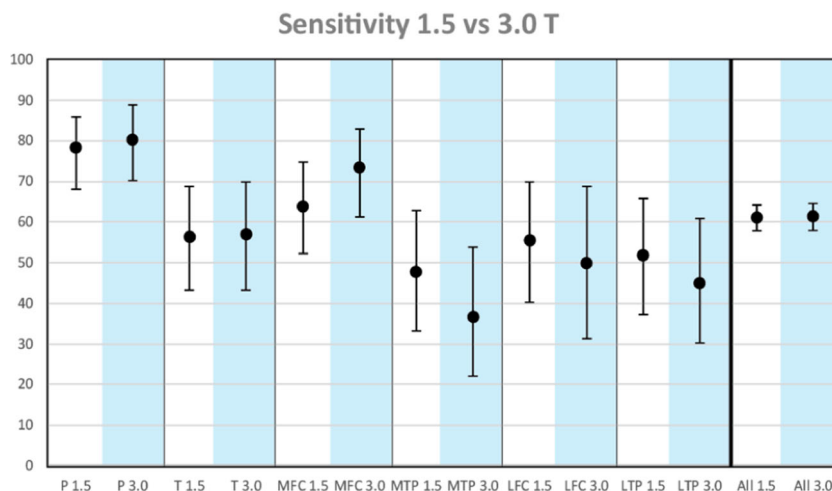
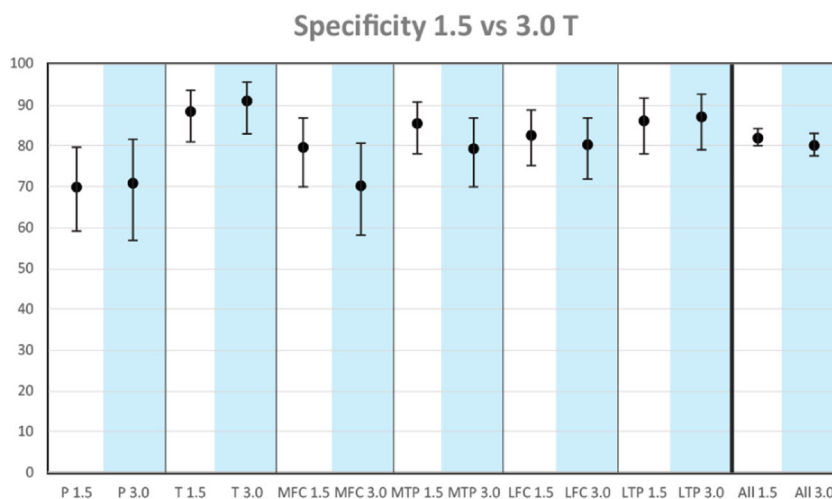


Fig. 3 Specificity of MR reports at 1.5 and 3.0 T per articular surface with error bars representing 95% confidence intervals



Several studies with dedicated readers have shown increased diagnostic performance for assessment of knee articular cartilage at 3.0 T in comparison to 1.5 T with 2D FSE imaging, although the performance metrics have varied among studies. Kijowski et al. [22] reported a significant increase in specificity and accuracy at 3.0 T in comparison to 1.5 T (specificity of 78.0% at 1.5 T and 85.9% at 3.0 T; accuracy of 74.5% at 1.5 T and 80.1% at 3.0 T) in 200 patients with arthroscopic correlation, but no significant difference in sensitivity (69.3% and 70.5%). In contrast, Van Dyck et al. [23] showed 3.0 T to have significantly higher sensitivity in assessment of articular cartilage in 200 patients (69% at 3.0 T compared to 60% at 1.5 T), but not higher specificity or accuracy. In a study of 26 patients, Wong et al. [26] showed 3.0-T MRI to have higher sensitivity (75.6% at 3.0 T and 70.6% at 1.5 T) and accuracy (51.3% at 3.0 T and 42.9% at 1.5 T).

The 3.0-T MRI has also been shown to have better diagnostic performance in experimental animal studies. Link et al. [20] showed a statistically significant increased area under ROC curves at 3.0 T in comparison to 1.5 T in a study of 27 porcine knees. Similarly, Masi et al. [19] showed higher accuracy at 3.0 T in a study of ten porcine knees and two human volunteers.

Although the diagnostic performance of MRI reports for assessment of articular cartilage in this study is within the range of the literature utilizing dedicated readers, the

previously demonstrated diagnostic advantage of 3.0 T MRI for assessment of knee articular cartilage was not achieved. The lack of improved diagnostic performance at 3.0 T was despite scanning parameters at 3.0 that potentially would allow for increased image quality, including a slight increase in matrix size at 3.0 T and increased number of acquisitions in two sequences (Tables 3 and 4). There are several reasons why the potential diagnostic advantage of 3.0 T was not demonstrated in this study, including some that are likely generalizable, and additional factors that may be unique to this institution.

Interobserver variability for imaging interpretation

Imaging evaluation of articular cartilage is challenging, and there may be considerable interobserver variability. A recent study of 5 readers independently evaluating 962 articular surfaces of the knee using CT arthrography demonstrated only fair global interobserver agreement on assessment of the cartilage (kappa = 0.35), and accuracy was shown to be improved with greater experience [27]. The MRI reports in our study were created by a relatively heterogeneous group of eight radiologists with musculoskeletal imaging experience ranging from 5–41 years. Additionally, three of the eight radiologists read a significantly different proportion of examinations at 1.5 and 3.0 T, with the sensitivity of these radiologists ranging from 55.0% to 69.8%, albeit with overlapping 95% confidence intervals. Although the low sample size per

Table 7 Sensitivity, specificity, percent of articular surfaces graded exactly, and percent of articular surfaces graded within one grade of all articular surfaces at both 1.5 and 3.0 T. Standard error in parentheses. *P*-

values were produced by the z-test. There was no significant difference between any of these parameters at 1.5 or 3.0 T

Sensitivity (%)		Specificity (%)		% Graded exactly		% Graded within 1 grade	
1.5 T	3.0 T	1.5 T	3.0 T	1.5 T	3.0 T	1.5 T	3.0 T
61.4 (1.5)	61.8 (1.7)	82.7 (1.2)	80.6 (1.4)	62.2 (1.5)	59.5 (1.7)	77.5 (1.4)	75.6 (1.5)
<i>P</i> = 0.860		<i>P</i> = 0.255		<i>P</i> = 0.234		<i>P</i> = 0.354	

Table 8 Agreement between MRI reports and arthroscopy with the weighted kappa statistic (95% confidence intervals in parentheses)

	Weighted kappa (95% confidence intervals)	
	1.5 T	3.0 T
Patella	0.61 (0.50–0.70)	0.42 (0.27–0.55)
Trochlea	0.66 (0.56–0.73)	0.57 (0.33–0.67)
MFC	0.47 (0.35–0.58)	0.66 (0.56–0.75)
MTP	0.44 (0.31–0.56)	0.33 (0.17–0.47)
LFC	0.45 (0.32–0.56)	0.51 (0.37–0.62)
LTP	0.53 (0.41–0.63)	0.57 (0.45–0.68)
All	0.56 (0.51–0.60)	0.55 (0.50–0.60)

radiologist and potential confounding factors precluded us from testing for significance in the diagnostic performance between individual readers, varying performance may play a role in obscuring the potential diagnostic advantage of 3.0 T.

General clinical factors In contrast to a dedicated research study where the participating readers can focus on the articular cartilage to the exclusion of the remainder of the knee joint, several clinical factors may contribute to varied diagnostic performance. For example, in a typical clinical workday, the radiologist may be required to concurrently integrate information from the medical record, manage interruptions, and interpret images. It is therefore possible that common clinical pitfalls such as fatigue, distraction, and satisfaction of search may impact the diagnostic performance in this environment [28–30].

Interobserver variability for arthroscopic evaluation The interobserver agreement of grading articular cartilage lesions at arthroscopy in the knee has been shown to be generally good, but not perfect, with a study by Marx et al. [31] consisting of 6 surgeons reviewing 31 videos of articular

cartilage demonstrating a kappa statistic ranging from 0.35 to 0.87. In our study, arthroscopic surgery was performed by a total of nine different surgeons with 4 to 25 years of experience, possibly leading to heterogeneity in the reporting of chondral lesions in the arthroscopy reports.

Limitations of arthroscopy The agreement between arthroscopy and histology has been shown to be good, but not perfect, with a reported weighted kappa statistic of 0.619 [32]. A similar kappa statistic of 0.602 was demonstrated for the agreement between arthroscopic grading of knee cartilage in comparison to arthrotomy [33]. There has also been recent work showing a potential blind spot of the far posterior femoral condyle articular cartilage that may not be visible at arthroscopy [34], with a possible such lesion demonstrated in Fig. 4. Additionally, in the freeform arthroscopy reports, 268/1800 of the articular surfaces were either described as normal in a blanket statement or were not explicitly described. For the statistical analysis of this study, these 268 articular surfaces were presumed normal. It is possible that a low-grade chondral lesion in a different compartment from the main site of injury may not have been described by the arthroscopist. However, this potential omission would not be expected to affect the sensitivity of MR or the percentage of cartilage lesions graded within one grade.

Effect of time between imaging and arthroscopy Although cartilage is generally considered to have limited healing potential [35], spontaneous healing of cartilage has been demonstrated clinically, for instance in medial opening-wedge high tibial osteotomy without additional cartilage surgery [35, 36]. It is conceivable that if a patient is treated with conservative measures such as bracing, physical therapy, or activity modification, then some healing may take place, especially if there is a long interval between imaging and arthroscopy. Conversely (and probably much more

Table 9 Number of examinations read at 1.5 and 3.0 T by each radiologist (the order of radiologists is sorted by descending total number of examinations read). Significance testing was performed with Pearson's chi-squared test (* $P < 0.05$). Sensitivity and specificity were

Radiologist	1.5-T exams (N)	3.0-T exams (N)	P = value	Sensitivity	Specificity
Radiologist 1	21	36	0.004*	55.0 (45–65)	83.2 (78–88)
Radiologist 2	34	22	0.421		
Radiologist 3	35	16	0.046*	62.5 (52–72)	78.2 (72–84)
Radiologist 4	25	22	0.911		
Radiologist 5	25	9	0.034*	69.8 (57–81)	80.9 (73–87)
Radiologist 6	12	15	0.248		
Radiologist 7	13	14	0.584		
Radiologist 8	2	2	n/a		

calculated for each radiologist who read a significantly different proportion of examinations at 1.5 and 3.0 T (95% confidence intervals in parentheses)

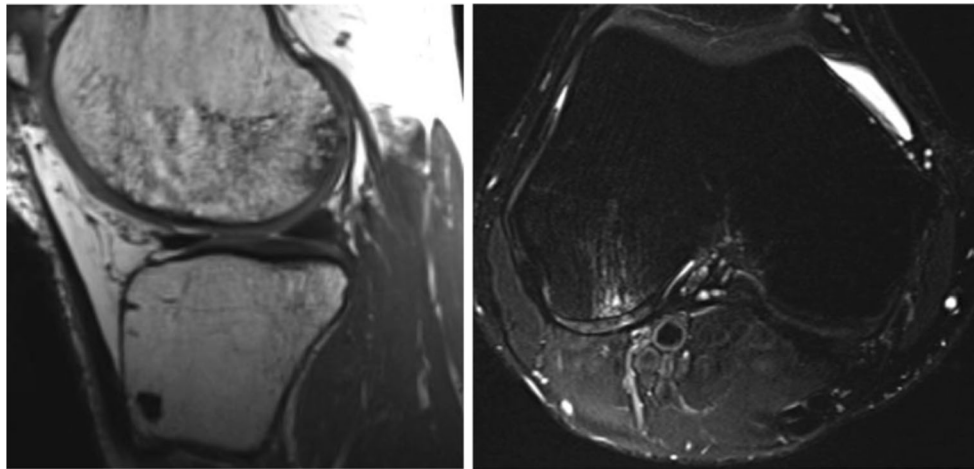


Fig. 4 MRI performed at 3.0 T with an apparent far-posterior cartilage abnormality of the lateral femoral condyle. Arthroscopy performed 29 days later did not reveal any cartilage defects. Sagittal PD SPACE (left image) and axial T2-weighted image with fat suppression (right image) demonstrate signal heterogeneity and superficial fissuring of the far

posterior non-weightbearing lateral femoral condyle with subchondral cystic change. This region of the far posterior femoral condyle may not be visible during routine arthroscopy. In the current study, this was classified as a false-positive MR grade II lesion

likely), further cartilage damage or new cartilage lesions may occur in the interval between imaging and arthroscopy. However, our study showed no significant difference in diagnostic performance if imaging was performed early (mean time interval of 24.0 days) or late (mean time interval of 110.1 days). Figure 5 demonstrates an apparent full thickness cartilage defect (grade IV) of the patella on MRI, while arthroscopy performed 170 days later demonstrated only softening (grade I) of the patellar cartilage. Figure 6 demonstrates a case where the articular surfaces were described as normal on MRI report (with apparent signal

heterogeneity of the medial femoral condyle possibly evident in retrospect), while arthroscopy performed 50 days later demonstrated a delaminating cartilage lesion of the medial femoral condyle without exposed bone (grade III). It is conceivable but purely conjectural that these two discrepancies may reflect interval partial healing and interval worsening, respectively.

Unique to this institution is the fact that a 1.5 T MR scanner (scanner 2 in Table 2) is located in the same outpatient building as the clinic location of two surgeons who specialize in cartilage repair surgery, which accounted for 106/300

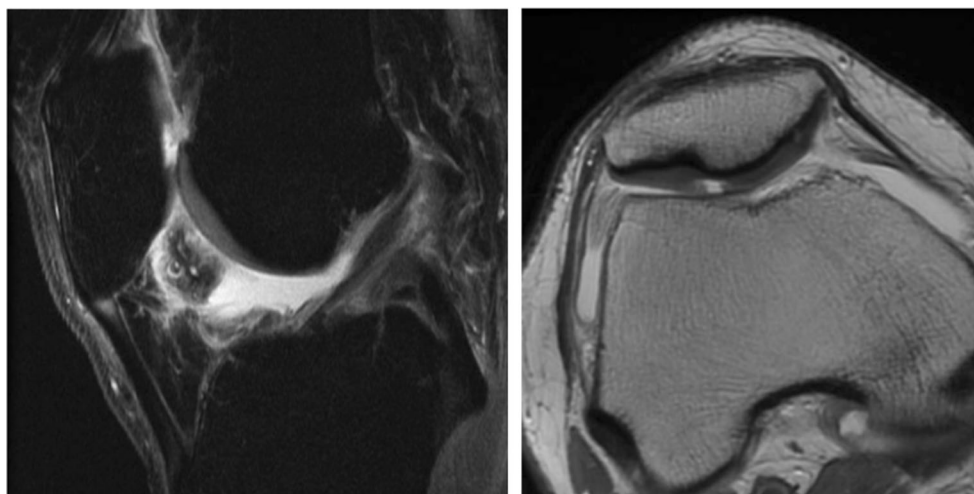


Fig. 5 MRI performed at 1.5 T with an apparent full-thickness defect in the medial facet/median ridge of the patella (grade IV), as demonstrated on the sagittal proton-density-weighted image with fat suppression (left) and reformatted axial PD SPACE image (right). Arthroscopy performed

170 days later revealed some softening (grade I) of the patellar cartilage. This was classified as a true positive (cartilage disease present at both MR and arthroscopy), but grading was not within one grade



Fig. 6 MRI performed at 1.5 T, with articular surfaces classified as normal. Some signal heterogeneity of the medial femoral condyle may be evident in retrospect as evident on the sagittal PD SPACE (left) and coronal PD-weighted image with fat suppression (middle). Arthroscopy

performed 50 days after MRI revealed a delaminating cartilage lesion without exposed bone (grade III) in the medial femoral condyle. This was classified as an MR false-negative grade III lesion

examinations in this study. Since these surgeons are well known to the staff radiologists, and patients of these surgeons often prefer the convenient availability of this scanner, it is possible that radiologists spend extra time and attention assessing the articular surfaces of these patients.

Another factor that deserves discussion is the inclusion of a 3D fast spin-echo isotropic sequence (SPACE on the Siemens scanners) in the majority of examinations. This sequence results in intermediate-weighted images with hyperintense synovial fluid with minimal blurring and allows the creation of multiplanar reformations in any orientation after a single acquisition [7]. SPACE and the similar Cube sequence from General Electric are non-gradient 3D sequences that have shown promise for assessment of intra-articular structures of the ankle [37] and knee [38], including the articular cartilage [39–46]. The musculoskeletal radiologists at this institution typically use the 3D sequence as a problem-solving tool rather than for primary interpretation, and these sequences have shown variable diagnostic performance for assessment of articular cartilage in the literature. Kijowski et al. showed that Cube offered increased sensitivity but lower specificity in detecting cartilage lesions of the knee, in comparison to 2D FSE images [47], which was born out in a similar study by Ristow et al. [48]. Milewski et al. concluded that there was no significant difference in accuracy comparing 3D fat-suppressed FSE imaging with conventional 2D FSE in 34 patients [49]. Schaefer et al. showed that SPACE showed decreased accuracy in comparison to conventional 2D FSE sequences [42]. In the present study, given that a 3D sequence was performed more commonly at 3.0 T (87.4% in comparison to 63.6% at 1.5 T), it is unlikely that the inclusion or exclusion of this sequence could explain the absence of performance benefit at 3.0 T.

This study has important limitations. While many of the factors hypothesized to affect the correlation between clinical MRI

reports and arthroscopic surgery are generalizable to various practice environments, the institution-specific factor of having a 1.5 T scanner that is commonly utilized in imaging of patients referred by the two cartilage repair surgeons likely introduces selection and interpretation bias. It would be difficult to quantify the degree of potential bias due to the potential confounding variables (such as different patient population, different scanner hardware, different scanning parameters, etc.) precluding the ability to perform subgroup analyses reliably. Additionally, the study may be underpowered despite the fact that the number of MRI reports correlated with arthroscopy (300) is greater than prior dedicated reader studies assessing knee articular cartilage. Although a sample size calculation was performed for a 10% ($\pm 5\%$) confidence interval width, it may be necessary to power the study for a narrower confidence interval width given the heterogeneity of radiologists, orthopedic surgeons, and other clinical factors.

Conclusion

This study shows no significant difference in diagnostic performance of clinical MRI reports for assessment of the knee articular cartilage imaged at 1.5 and 3.0 T using arthroscopy as the standard of reference at our institution. The confluence of several patient-, machine-, physician-, and institution-specific factors in a large and busy clinical practice most likely outweigh the inherent diagnostic advantage of 3.0 T MRI that has been previously demonstrated in the literature.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Johnson VL, Hunter DJ. The epidemiology of osteoarthritis. *Best Pract Res Clin Rheumatol*. 2014;28(1):5–15.
- Smith GD, Knutsen G, Richardson JB. A clinical review of cartilage repair techniques. *J Bone Joint Surg Am*. 2005;87(4):715–24.
- Roemer FW, Winalski CS. State of the art: MR imaging after knee cartilage repair surgery I. 2015;277(1).
- Bondeson J. Are we moving in the right direction with osteoarthritis drug discovery? *Expert Opin Ther Targets*. 2011;15(12):1355–68.
- Chevalier X, Eymard F, Richette P. Biologic agents in osteoarthritis: hopes and disappointments. *Nat Rev Rheumatol*. 2013;9(7):400–10.
- Crema MD, Roemer FW, Marra MD, Burstein D, Gold GE, Eckstein F, et al. Articular cartilage in the knee: current MR imaging techniques and applications in clinical practice and research. *Radiographics*. 2011;31(1):37–61.
- Kijowski R. Clinical cartilage imaging of the knee and hip joints. *AJR Am J Roentgenol*. 2010;195(3):618–28.
- Figuerola D, Calvo R, Vaisman A, Carrasco MA, Moraga C, Delgado I. Knee chondral lesions: incidence and correlation between arthroscopic and magnetic resonance findings. *Arthroscopy*. 2007;23(3):312–5.
- Friemert B, Oberländer Y, Schwarz W, Häberle HJ, Bähren W, Gemgroß H, et al. Diagnosis of chondral lesions of the knee joint: can MRI replace arthroscopy?: a prospective study. *Knee Surg Sports Traumatol Arthrosc*. 2004;12(1):58–64.
- Quatman CE, Hettrich CM, Schmitt LC, Spindler KP. The clinical utility and diagnostic performance of magnetic resonance imaging for identification of early and advanced knee osteoarthritis: a systematic review. *Am J Sports Med*. 2011;39(7):1557–68.
- Smith TO, Drew BT, Toms AP, Donell ST, Hing CB. Accuracy of magnetic resonance imaging, magnetic resonance arthrography and computed tomography for the detection of chondral lesions of the knee. *Knee Surg Sports Traumatol Arthrosc*. 2012;20(12):2367–79.
- Sonin AH, Pensy RA, Mulligan ME, Hatem S. Grading articular cartilage of the knee using fast spin-echo proton density-weighted MR imaging without fat suppression. *Am J Roentgenol*. 2002 Nov;179(5):1159–66.
- von Engelhardt LV, Lahner M, Klussmann A, Bouillon B, Dávid A, Haage P, et al. Arthroscopy vs. MRI for a detailed assessment of cartilage disease in osteoarthritis: diagnostic value of MRI in clinical practice. *BMC Musculoskelet Disord*. 2010;11:75.
- Kijowski R, Blankenbaker DG, Munoz Del Rio A, Baer GS, Graf BK. Evaluation of the articular cartilage of the knee joint: value of adding a T2 mapping sequence to a routine MR imaging protocol. *Radiology*. 2013;267(2):503–13.
- Kijowski R, Blankenbaker DG, Woods MA, Shinki K, De Smet AA, Reeder SB. 3.0-T evaluation of knee cartilage by using three-dimensional IDEAL GRASS imaging: comparison with fast spin-echo imaging. *Radiology*. 2010;255(1):117–27.
- Lee SY, Jee WH, Kim SK, Kim JM. Proton density-weighted MR imaging of the knee: fat suppression versus without fat suppression. *Skelet Radiol*. 2011;40(2):189–95.
- Lavdas E, Topalzikis T, Mavroidis P, Kyriakis I, Roka V, Kostopoulos S, et al. Comparison of PD BLADE with fat saturation (FS), PD FS and T2 3D DESS with water excitation (WE) in detecting articular knee cartilage defects. *Magn Reson Imaging*. 2013;31(8):1255–62.
- Kohl S, Meier S, Ahmad SS, Bonel H, Exadaktylos AK, Krismser A, et al. Accuracy of cartilage-specific 3-Tesla 3D-DESS magnetic resonance imaging in the diagnosis of chondral lesions: comparison with knee arthroscopy. *J Orthop Surg Res*. 2015;10:191.
- 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 Jul;236(1):140–50.
- Link TM, Sell CA, Masi JN, Phan C, Newitt D, Lu Y, et al. 3.0 vs 1.5 T MRI in the detection of focal cartilage pathology—ROC analysis in an experimental model. *Osteoarthr Cartil*. 2006 Jan;14(1):63–70.
- Barr C, Bauer JS, Malfair D, Ma B, Henning TD, Steinbach L, et al. MR imaging of the ankle at 3 Tesla and 1.5 Tesla: protocol optimization and application to cartilage, ligament and tendon pathology in cadaver specimens. *Eur Radiol*. 2007;17(6):1518–28.
- Kijowski R, Blankenbaker DG, Davis KW, Shinki K, Kaplan LD, De Smet AA. Comparison of 1.5- and 3.0-T MR imaging for evaluating the articular cartilage of the knee joint I. *Radiology*. 2009;250(3):839–48.
- Van Dyck P, Kenis C, Vanhoenacker FM, Lambrecht V, Wouters K, Gielen JL, et al. Comparison of 1.5- and 3-T MR imaging for evaluating the articular cartilage of the knee. *Knee Surg Sports Traumatol Arthrosc*. 2014;22(6):1376–84.
- Curl WW, Krome J, Gordon ES, Rushing J, Smith BP, Poehling GG. Cartilage injuries: a review of 31,516 knee arthroscopies. *Arthroscopy*. 1997;13(4):456–60.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977 Mar;33(1):159–74.
- Wong S, Steinbach L, Zhao J, Stehling C, Ma CB, Link TM. Comparative study of imaging at 3.0 T versus 1.5 T of the knee. *Skelet Radiol*. 2009;38(8):761–9.
- Omoumi P, Michoux N, Larbi A, Lacoste L, Lecouvet FE, Perlepe V, et al. Multirater agreement for grading the femoral and tibial cartilage surface lesions at CT arthrography and analysis of causes of disagreement. *Eur J Radiol*. 2017;88:95–101.
- Palazzetti V, Guidi F, Ottaviani L, Valeri G, Baldassarre S, Giuseppetti GM. Analysis of mammographic diagnostic errors in breast clinic. *Radiol Med*. 2016;121(11):828–33.
- Berbaum KS, Krupinski EA, Schartz KM, Caldwell RT, Madsen MT, Hur S, et al. Satisfaction of search in chest radiography 2015. *Acad Radiol*. 2015;22(11):1457–65.
- Schartz KM, Madsen MT, Kim J, Ohashi R, Ohashi K, El-Khoury GY, et al. Trauma in CT: the role of severe injury on satisfaction of search revised. *J Am Coll Radiol*. 2016;13(8):973–978.e4.
- Marx RG, Connor J, Lyman S, Amendola A, Andrish JT, Kaeding C, et al. Multirater agreement of arthroscopic grading of knee articular cartilage. *Am J Sports Med*. 2005;33(11):1654–7.
- Acebes C, Roman-Blas JA, Delgado-Baeza E, Palacios I, Herrero-Beaumont G. Correlation between arthroscopic and histopathological grading systems of articular cartilage lesions in knee osteoarthritis. *Osteoarthr Cartil*. 2009;17(2):205–12.
- Cameron ML, Briggs KK, Steadman JR. Reproducibility and reliability of the outerbridge classification for grading chondral lesions of the knee arthroscopically. *Am J Sports Med*. 2003;31(1):83–6.
- Sando MJ, Rajae SS, Liu JP, Banffy M, Limpisvasti O, Crues JV. Identifying hidden zones of the far posterior cartilage of the femoral Condyles not visible during knee arthroscopy. 2017;W1–6.
- Fritz J, Janssen P, Gaismaier C, Schewe B, Weise K. Articular cartilage defects in the knee-basics, therapies and results. *Injury*. 2008;39(1 SUPPL):50–7.
- Kumagai K, Akamatsu Y, Kobayashi H, Kusayama Y, Koshino T, Saito T. Factors affecting cartilage repair after medial opening-wedge high tibial osteotomy. *Knee Surg Sports Traumatol Arthrosc*. 2017;25(3):9–14.
- Stevens KJ, Busse RF, Han E, Brau ACS, Beatty PJ, Beaulieu CF, et al. Ankle: isotropic MR imaging with 3D-FSE-cube—initial experience in healthy volunteers. *Radiology*. 2008;249(3):1026–33.
- Van Dyck P, Gielen JL, Vanhoenacker FM, De Smet E, Wouters K, Dossche L, et al. Diagnostic performance of 3D SPACE for comprehensive knee joint assessment at 3 T. *Insights Imaging*. 2012;3(6):603–10.
- Chen CA, Kijowski R, Shapiro LM, Tuite MJ, Davis KW, Klares JL, et al. Cartilage morphology at 3.0T: assessment of three-

- dimensional magnetic resonance imaging techniques. *J Magn Reson Imaging*. 2010;32(1):173–83.
40. Crema MD, Nogueira-Barbosa MH, Roemer FW, Marra MD, Niu J, Chagas-Neto FA, et al. Three-dimensional turbo spin-echo magnetic resonance imaging (MRI) and semiquantitative assessment of knee osteoarthritis: comparison with two-dimensional routine MRI. *Osteoarthr Cartil*. 2013;21(3):428–33.
 41. Friedrich KM, Reiter G, Kaiser B, Mayerhöfer M, Deimling M, Jellus V, et al. High-resolution cartilage imaging of the knee at 3 T: basic evaluation of modern isotropic 3D MR-sequences. *Eur J Radiol*. 2011;78(3):398–405.
 42. Schaefer FKW, Kurz B, Schaefer PJ, Fuerst M, Hedderich J, Graessner J, et al. Accuracy and precision in the detection of articular cartilage lesions using magnetic resonance imaging at 1.5 Tesla in an in vitro study with orthopedic and histopathologic correlation. *Acta Radiol*. 2007;48(10):1131–7.
 43. Kijowski R, Gold GE. Routine 3D magnetic resonance imaging of joints. *J Magn Reson Imaging*. 2011;33(4):758–71.
 44. Notohamiprodjo M, Homg A, Kuschel B, Paul D, Li G, Raya JG, et al. 3D-imaging of the knee with an optimized 3D-FSE-sequence and a 15-channel knee-coil. *Eur J Radiol*. 2012;81(11):3441–9.
 45. Van Dyck P, Vanhevel F, Vanhoenacker FM, Wouters K, Grodzki DM, Gielen JL, et al. Morphological MR imaging of the articular cartilage of the knee at 3 T—comparison of standard and novel 3D sequences. *Insights Imaging*. 2015;6(3):285–93.
 46. 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(5):857–64.
 47. Kijowski R, Davis KW, Woods MA, Lindstrom MJ, De Smet AA, Gold GE, et al. Knee joint: comprehensive assessment with 3D isotropic resolution fast spin-echo MR imaging—diagnostic performance compared with that of conventional MR imaging at 3.0 T. *Radiology*. 2009;252(2):486–95.
 48. Ristow O, Steinbach L, Sabo G, Krug R, Huber M, Rauscher I, et al. Isotropic 3D fast spin-echo imaging versus standard 2D imaging at 3.0 T of the knee—image quality and diagnostic performance. *Eur Radiol*. 2009;19(5):1263–72.
 49. Milewski MD, Smitaman E, Moukaddam H, Katz LD, Essig DA, Medvecky MJ, et al. Comparison of 3D vs. 2D fast spin echo imaging for evaluation of articular cartilage in the knee on a 3 T system scientific research. *Eur J Radiol*. 2012;81(7):1637–43.