Skeletal Radiol (2009) 38:21–29
DOI 10.1007/s00256-008-0561-1

SCIENTIFIC ARTICLE

Comparison of spin echo T1-weighted sequences versus fast spin-echo proton density-weighted sequences for evaluation of meniscal tears at 1.5 T

Andrew B. Wolff · Lorenzo L. Pesce · Jim S. Wu · L. Ryan Smart · Michael J. Medvecky · Andrew H. Haims

Received: 31 March 2008 / Revised: 28 June 2008 / Accepted: 1 July 2008 / Published online: 12 August 2008

Abstract

Purpose At our institution, fast spin-echo (FSE) proton density (PD) imaging is used to evaluate articular cartilage, while conventional spin-echo (CSE) T1-weighted sequences have been traditionally used to characterize meniscal pathology. We sought to determine if FSE PD-weighted sequences are equivalent to CSE T1-weighted sequences in the detection of meniscal tears, obviating the need to perform both sequences.

Method and materials We retrospectively reviewed the records of knee arthroscopies performed by two arthroscopy-focused surgeons from an academic medical center over a 2-year period. The preoperative MRI images were interpreted independently by two fellowship-trained musculoskeletal radiologists who graded the sagittal CSE T1 and FSE PD sequences at different sittings with grades 1–5, where 1 = normal meniscus, 2 = probable normal meniscus, 3 = indeterminate, 4 = probable torn meniscus, and 5 = torn meniscus. Each meniscus was divided into an anterior and posterior half, and these halves were graded separately. Operative findings provided the gold standard. Receiver operating characteristic (ROC) analysis was performed to compare the two sequences.

Results There were 131 tears in 504 meniscal halves. Using ROC analysis, the reader 1 area under curve for FSE PD was significantly better than CSE T1 (0.939 vs. 0.902, >95% confidence). For reader 2, the difference met good criteria for statistical non-inferiority but not superiority (0.913 for FSE PD and 0.908 for CSE T1; >95% non-inferiority for difference at most of −0.027).

Conclusion FSE PD-weighted sequences, using our institutional protocol, are not inferior to CSE T1-weighted sequences for the detection of meniscal tears and may be superior.

Keywords MR imaging · Meniscus · Knee · Arthroscopy · Sports medicine

Introduction

Magnetic resonance imaging (MRI) has revolutionized the imaging and treatment of knee injuries. Meniscal tears can be a source of significant knee pain and disability in many individuals. Identification of meniscal tears is the most common indication for MRI of the knee [1]. In a comparison of nine studies with a study volume of at least 200 menisci done by Rubin and Paletta [2], the sensitivity for diagnosing a medial meniscal tear was 86% to 96% with a specificity of 84% to 94%. For lateral meniscal tears, the sensitivity was 68% to 86% and the specificity was 92% to 98%. Taken together, the negative predictive value for MRI for these studies was 91% [2]. Due to its high negative predictive value, MRI can help avoid unnecessary diagnostic arthroscopy [3–7]. MRI can also identify alternative knee pathology, as well as size, location, and characteristics of meniscal tears that will guide treatment. Coupled with a patient's history and physical exam findings, MRI has a high utility in directing the treatment of individuals with knee pain.

This manuscript has not been presented or accepted for presentation at any meeting or in any journal.

A. B. Wolff · L. L. Pesce · J. S. Wu · L. R. Smart · M. J. Medvecky · A. H. Haims (✉)
Department of Diagnostic Radiology, Yale University School of Medicine, P.O. Box 208042, Tompkins East 2, New Haven, CT 06520-8042, USA
e-mail: andrew.haims@yale.edu

Several authors have reported on the accuracy of T1- and PD-weighted sequences in the detection of meniscal tears, but little is reported comparing the two [8]. Reported accuracy in the literature suggests that the two sequences are of similar value in evaluating the meniscus [8–16]. Additionally, fast spin-echo and proton density-weighted sequences are reported to have good accuracy in the evaluation of osteochondral lesions [8, 17–19]. Given the versatility and speed of FSE PD-weighted sequences for the purpose of evaluating various structures about the knee, FSE PD-weighted sequences have become a standard part of routine knee MRI in many centers. While PD-weighted sequences have not been definitively shown to be equivalent or better than T1-weighted sequences for the purpose of meniscal evaluation, the more traditional (and oftentimes slower) CSE T1-weighted sequences have become less popular.

Furthermore, whether fast spin-echo sequences are equivalent to conventional spin-echo sequences for the purpose of meniscal evaluation is still a subject of controversy. Early studies reported results for FSE PD-weighted sequences comparable to those of CSE PD-weighted sequences [11, 14, 20, 21]. In a 2005 study, Blackmon et al. disputed these results, citing an unacceptable drop in sensitivity with FSE PD-weighted sequences and encouraged the abandonment of FSE PD-weighted sequences in favor of CSE PD-weighted sequences for the purpose of detecting meniscal tears [10].

At our institution, we use CSE T1-weighted sagittal sequences in conjunction with FSE PD sagittal imaging sequences as part of our routine knee MRI evaluation. In undertaking this study, we sought to determine whether performing sagittal CSE T1-weighted sequences in conjunction with sagittal FSE PD-weighted sequences was necessary. Our hypothesis was that FSE PD-weighted sagittal sequences are equivalent to CSE T1-weighted sequences in the detection of meniscal tears and that the CSE T1-weighted sagittal sequences could be abandoned for routine knee MRI.

Materials and methods

This study was performed in compliance with the Health Insurance Portability and Accountability Act and with approval from our Human Investigational Committee.

Patient selection and medical records review

We reviewed the records of knee arthroscopies performed by two sports medicine and arthroscopy-focused orthopedic surgeons from an academic medical center over 2 years beginning July 1, 2003. We identified 345 patients who underwent knee arthroscopy at that time. We excluded patients who had undergone prior surgery on the operated knee, patients who had preoperative MRIs at facilities other than our own, and patients who underwent arthrograms. There were 143 patients who met these criteria, of whom 17 were excluded due to problems with their MRI [no high resolution PD sequence (10), no T1 sequence (1), motion artifact (5), and metal artifact (1)]. There were 126 records reviewed (50 women, 76 men) with an average age of 39.6 years (10–77 years). Operative reports and diagrams drawn by the surgeon at the time of surgery were reviewed for presence or absence and location of meniscal tear. Each meniscus was divided into anterior and posterior halves. For meniscal tears that crossed the midline of the meniscus in the coronal plane, both halves were considered torn. A meniscus was labeled as torn if it exhibited any signs of instability or cleavage in any direction regardless of apparent mechanical stability or degenerative changes. Marginal fraying was not considered abnormal for the purposes of this study.

MRI technique

All patients underwent MR imaging of the knee on a 1.5-T system (GE Medical Systems, Milwaukee, Wis.). Imaging was performed in a quadrature knee coil (Invivo, Latham, NY, USA) in the sagittal plane. The parameters for the CSE T1 sequences are as follows: TR/TE 475/14; BW 15.63, matrix 256×192; NEX 1, FOV 14; spacing, 3.3 mm interleaved. The parameters for the FSE PD sequences are as follows: TR/TE 2800/25; ETL 6, BW 20.83, matrix 320×256; NEX 3, FOV 14; spacing, 3:0.5 mm).

MRI analysis

The MRI images were interpreted independently by two fellowship-trained and experienced musculoskeletal radiologists (A.H.H. and J.S.W.) who graded the sagittal CSE T1- and FSE PD-weighted sequences at different sittings on different days. When each sequence was viewed, there were no other sequences available for review. Additionally, the radiologists were blinded to the results of arthroscopy, although they did have knowledge that patients underwent arthroscopy. As with the operative report, each meniscus was arbitrarily divided into anterior and posterior halves, and tears were ascribed to either, both, or neither of each half of each meniscus. Grades 1–5 were assigned, where 1 = normal meniscus, 2 = probable normal meniscus, 3 = indeterminate, 4 = probable torn meniscus, and 5 = torn meniscus. Criteria for a definitive tear were threefold: (1) linear signal contacting the articular surface on two or more images, (2) contour abnormality of the meniscus, and (3)

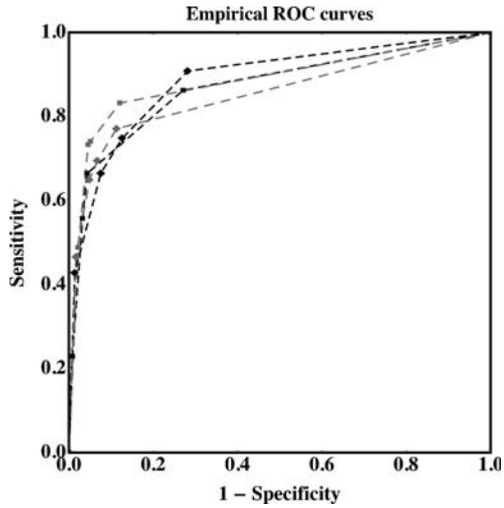

Fig. 1 Empirical ROC curves generated from the grading of the MRI readings of readers 1 and 2 for conventional spin-echo (*CSE*) T1-weighted sequences and fast spin-echo proton density (*PD*)-weighted sequences. The *gray-shaded plots* represent the readings of the FSE PD-weighted sequences (*squares* reader 1, *diamonds* reader 2). The *black-shaded plots* represent the readings of the CSE T1-weighted sequences (*squares* reader 1, *diamonds* reader 2)

abnormal size or truncation of the meniscus. A probably torn meniscus had signal abnormality contacting the articular surface on one image only, and a probably normal meniscus had signal abnormality that did not definitely contact the articular surface.

Statistical analysis

The results of the grades of the MRI sequences by each reader were correlated with the findings at the time of surgery. Receiver operating characteristic (ROC) analysis was performed, and area under curve (AUC) values were calculated for both readers for both MRI sequences [22]. Multiple reader, multiple case (MRMC) analysis was also performed [23, 24] (using DBM-MRMC 2.0, http://www-radiology.uchicago.edu/krl/). The fitting model used was the proper binormal model because of its reliability [25]. One-tailed analysis for the lower bound for 95% confidence of non-inferiority of FSE PD versus CSE T1-weighted sequences was performed for the ROC curves. For the sensitivity and specificity comparisons, as well as for correlated frequencies comparisons (including tests that involved both positive and negative cases at the same time), the statistical analysis was performed using the McNemar test [26]. Multiple comparisons issues were not considered because the purpose of the paper was not to find one or more significant measurement among many non-significant ones. Positive predictive values and negative predictive values were not tested for statistical significance and are reported as confidence intervals.

Results

There were a total of 131 meniscal tears identified in 504 meniscal halves at the time of surgery. Considering all meniscal halves using ROC analysis (Figs. 1 and 2), reader 1 area under curve (AUC) for FSE PD-weighted sequences was significantly better than CSE T1-weighted sequences (0.939 vs. 0.902; >95% confidence). For reader 2, the difference was less and met criteria for statistical non-inferiority but not superiority (0.913 for FSE PD, 0.908 for CSE T1; lower bound of 95% confidence for non-inferiority of FSE PD=−0.027)). Similarly, results for MRMC analysis reflected statistical non-inferiority (0.925 for FSE PD, 0.905 for CSE T1; lower bound of 95% confidence for non-inferiority of FSE PD=−0.057). When the two readers are considered as fixed but together, the lower bound for non-inferiority is −0.002 (i.e., 95% certainty that, at worst, FSE PD-weighted sequences have an AUC of the ROC curve of 0.002 less than the AUC of the ROC curve of CSE T1-weighted sequences).

When the medial and lateral menisci were considered separately, a similar pattern emerged in the ROC curves generated. For the medial meniscus, reader 1 AUC for the ROC curve of FSE PD-weighted sequences was significantly better than that of CSE T1-weighted sequences (0.957 vs. 0.929; ~95% confidence), while for reader 2 AUC for the ROC curve of FSE PD was not significantly worse than that of CSE T1 (0.931 vs. 0.944; lower bound of 95% confidence for non-inferiority of FSE PD=−0.052). In evaluating the lateral meniscus, both readers had a higher

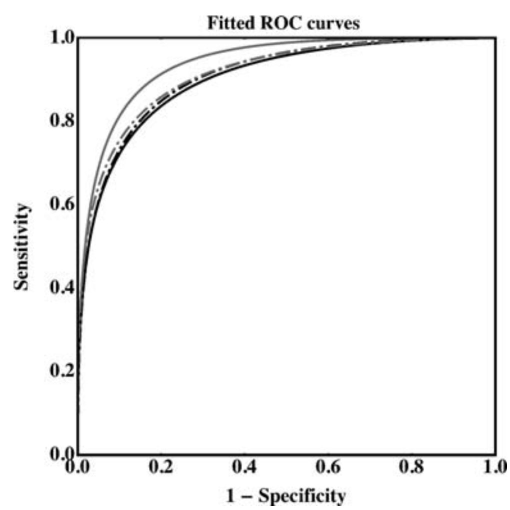

Fig. 2 Fitted ROC curves using the proper binormal model generated from the grading of the MRI readings of readers 1 and 2 for conventional spin-echo (*CSE*) T1-weighted sequences and fast spin-echo proton density (*PD*)-weighted sequences. The *gray-shaded plots* represent the readings of the FSE PD-weighted sequences (*continuous* reader 1, *dotted* reader 2). The *black shaded plots* represent the readings of the CSE T1-weighted sequences (*continuous* reader 1, *dotted* reader 2)

area under the ROC curve for FSE PD-weighted sequences relative to CSE T1-weighted sequences (reader 1, 0.915 vs. 0.876 with lower bound of 95% confidence for non-inferiority of FSE PD=−0.01; reader 2, 0.913 vs. 0.868 with lower bound of 95% confidence for non-inferiority of FSE PD=−0.008).

Because of the low number of tears of the anterior half of the medial meniscus (six), no meaningful ROC curves could be generated from the subdivision of the medial meniscus. For the lateral meniscus, the ROC curves for the anterior and posterior halves were not statistically different for FSE PD- and CSE T1-weighted sequences for either reader.

The data also reflect a stronger diagnostic response for FSE PD-weighted sequences as compared to CSE T1-weighted sequences. Reader 1 assigned the grade of 1 (normal meniscus) to 350 meniscal halves from the FSE PD-weighted sequences compared to 290 for the CSE T1-weighted sequences (>99% confidence). Reader 2 assigned the grade of 1 to 361 meniscal halves from the FSE PD-weighted sequences compared to 280 for the CSE T1-weighted sequences (>99% confidence). Similarly, reader 1 assigned the grade of 5 (torn meniscus) to 72 meniscal halves from the FSE PD-weighted sequences compared to 33 for the CSE T1-weighted sequences (>99% confidence), and reader 2 assigned the grade of 5 to 67 meniscal halves from the FSE PD-weighted sequences compared to 61 for the CSE T1-weighted sequences (not statistically significant). This caused an inward shift on the empirical ROC curve for FSE PD-weighted sequences compared to the CSE T1-weighted sequences (Fig. 1).

Upon examination of the more definitive grades, the positive and negative predictive values of the CSE T1-weighted sequences and the FSE PD-weighted sequences were comparable. When a grade of 5 (torn meniscus) was assigned to a meniscal half, the positive predictive value (PPV) for reader 1 was 89% (CI 81–100%) and 91% (CI 81–96%) for FSE PD- and CSE T1-weighted sequences, respectively. For reader 2, the PPV was 91% (CI 85–98%) and 92% (CI 85–99%) for FSE PD- and CSE T1-weighted sequences, respectively. For readings of grade 5, the specificity for both readers was 98% for FSE PD-weighted sequences and 99% for CSE T1-weighted sequences. These differences were not statistically significant. Similarly, when a grade of 1 (normal meniscus) was assigned to a meniscal half, the negative predictive value (NPV) for reader 1 was 94% (CI 91–97%) for both FSE PD- and CSE T1-weighted sequences. For reader 2, the NPV was 92% (CI 89–95%) and 96% (CI 93–98%) for FSE PD- and CSE T1-weighted sequences, respectively.

When less definitive grades are used, the results are comparable to those of the more definitive grades. When a grade of 4 or 5 (probable torn meniscus or torn meniscus) was assigned to a meniscal half, the positive predictive value for reader 1 was 85% (CI 78–92%) and 86% (CI 78–93%) for FSE PD- and CSE T1-weighted sequences, respectively. For reader 2, the PPV was 83% (CI 75–90%) and 76% (CI 68–84%) for FSE PD- and CSE T1-weighted sequences, respectively. For readings of grade 4 or 5, the specificity for reader 1 was 95% for FSE PD-weighted sequences and 97% for CSE T1-weighted sequences (not statistically significant). For reader 2, the specificity for readings of grade 4 or 5 was slightly but significantly better for FSE PD-weighted sequences than those of CSE T1-weighted sequences (95% vs. 92%; $p<0.05$). Similarly, when a grade of 1 or 2 (normal meniscus or probable normal meniscus) was assigned to a meniscal half, the negative predictive value for reader 1 was 91% (CI 88–94%) for FSE PD-weighted sequences and 89% (CI 86–92%) for CSE T1-weighted sequences. For reader 2, the NPV was 90% (CI 86–92%) and 91% (C.I. 88%-94%) for FSE PD- and CSE T1-weighted sequences, respectively. The sensitivity for readings of grade 1 or 2 for reader 1 was 74% and 64% for FSE PD- and CSE T1-weighted sequences, respectively ($p<0.06$). For reader 2, the sensitivity for readings of grade 1 or 2 was 69% and 75% for FSE PD- and CSE T1-weighted sequences, respectively (not a statistically significant difference).

Discussion

In the current study, the comparison of the two sequences was not an attempt to directly compare FSE and CSE sequences or PD- and T1-weighted sequences with identical parameters as the specific parameters used were markedly different. This was a comparison of our traditional meniscal sequence (CSE T1) to our standard articular cartilage (FSE PD) sequence. With regards to the imaging parameters, the FSE PD was a higher resolution sequence with frequency and phase matrices of 320×256 compared to the CSE T1 256×192. The number of excitations (NEX) of the FSE PD-weighted sequences was also significantly higher (three versus one for CSE T1). These two factors would potentially provide the FSE PD-weighted sequence with an advantage for evaluation of meniscal tears. However, the lack of an interslice gap and the shorter echo time with the CSE T1-weighted sequences provides a potential advantage for meniscal evaluation. Thus, we undertook this study to evaluate the continued need for performing the CSE T1-weighted sequences.

To our knowledge, the only study comparing CSE T1- to FSE PD-weighted sequences was a 1996 study in which Kojima et al. compared parasagittal CSE T1-weighted sequences with coronal FSE PD-weighted sequences (fat suppression FSE with a TR=2,000 and TE=19 and an echo train length of 2) with arthroscopic correlation in 202 patients [8]. They reported that the FSE PD-weighted sequences were superior to the CSE T1-weighted images in depicting the menisci, articular cartilage, and capsular structures. However, their quantitative analysis did not bear this out. For medial meniscal tears, they found a sensitivity and specificity of 95.1% and 93.3%, respectively, for CSE T1-weighted sequences compared to 90.8% and 93.3% for FSE PD-weighted sequences. For lateral meniscal tears, they found a sensitivity and specificity of 72.6% and 97.9%, respectively, for CSE T1-weighted sequences compared to 66.1% and 97.1% for FSE PD-weighted sequences. None of these differences met the threshold for statistical significance.

Other authors have sought to compare FSE PD- to CSE PD-weighted sequences with respect to meniscal evaluation. Many reports concluded that FSE PD-weighted sequences were equivalent to CSE PD-weighted sequences in this regard [11, 14, 20, 21, 27]. Consequent to this, many centers have replaced the conventional spin-echo with fast spin-echo PD imaging for routine knee MRI. In 2005, Blackmon et al. disputed these findings and recommended abandoning the FSE PD-weighted sequences reporting a sensitivity of 93% for CSE PD-weighted sequences versus 80% for FSE PD-weighted sequences [10]. However, in their study, the values for CSE sequences were obtained from a separate retrospective control group of patients who underwent arthroscopy after MRI. In contrast, the sensitivity value for the FSE sequences was generated by comparison with these control group numbers for CSE sequences in a different group of patients who did not undergo arthroscopic correlation. Blackmon et al. also reported a statistically significant rate of discordance of 17% ($p<0.01$) between CSE and FSE sequences (72 of 432 menisci), noting that 40 of the 72 were tears picked up on CSE but not FSE sequences, while the other 32 were "abnormal meniscal signal on fast spin-echo that was not present on conventional spin-echo sequences." This raises the question as to whether these "abnormal meniscal signals" represented tears seen on FSE but not CSE sequences; however, the authors do not comment on this. Thus, the issue of whether conventional spin-echo can be replaced by fast spin-echo sequences is still a subject of controversy.

While the current study was not designed to resolve this issue, it does lend credence to the argument that FSE PD-weighted sequences have a high level of accuracy in the detection of meniscal tears even when compared to a sequence designed for the specific purpose of meniscal evaluation. Indeed, as we had hypothesized, the FSE PD-weighted sequences were statistically at least as good, if not better, than the CSE T1-weighted sequences at detecting meniscal tears.

One of the strengths of this study was the use of ROC analysis to describe our results. This is an effective method of evaluating the quality or performance of a diagnostic test and is widely used in the radiologic literature [28–31]. To our knowledge, this is the first study to utilize ROC analysis to compare accuracy of two MRI sequences in the detection of meniscal tears. Previous studies have described the MRI findings of the meniscus in a binary manner—either torn or not torn, or with the grading scale described by Crues et al. [12] and then analyzed as separate groups categorically with respect to arthroscopic findings depending on their assigned grade. We feel that assigning a grade of 1–5 more accurately reflects the nuances of clinical MRI interpretation and provides a more complete assessment of the relative merits of a diagnostic test. The ROC curve is defined as the plot of test sensitivity as the y coordinate versus 1-specificity or false positive rate (FPR) as the x coordinate. The generated curves (Figs. 1 and 2) demonstrate how, for each sequence, sensitivity varies together with specificity and FPR. To compare the overall accuracy of the two sequences, the area under the curves were calculated.

The area under reader 1's ROC curve for all meniscal halves taken together was statistically significantly greater for FSE PD-weighted sequences as compared with CSE T1-weighted sequences. For reader 2, the differences were less pronounced and allow us only to conclude that, statistically speaking, there is a 95% chance that, at worst, FSE PD-weighted sequences have an area under the ROC curve of 0.027 less than the area under the ROC curve for CSE T1-weighted sequences. In other words, for reader 2, there is a strong case that FSE PD-weighted sequences are not inferior to CSE T1-weighted sequences in the detection of meniscal tears, while for reader 1, the FSE PD-weighted sequences were statistically superior.

Multiple reader, multiple case (MRMC) analysis was also performed using the ROC curves of both readers together. MRMC analysis provides a quantitative measure of the performance of a diagnostic test across a population of readers with varying degrees of skill. It allows the results of the current readers evaluating the current cases to be generalized to all readers of all cases [32]. In this way, it is superior to the collection of single reader ROC analyses. The utility of the MRMC analysis in the current study is limited by the low number of readers (two). Nevertheless, the findings of statistical non-inferiority in the MRMC analysis of the both readers' ROC curves strengthen the conclusion that FSE PD-weighted sequences are not inferior to CSE T1-weighted sequences. When the ROC curves of both readers were considered together without the use of MRMC analysis but rather considering each reader as fixed (and thereby unable to be generalized across a population of readers), a stronger conclusion of statistical

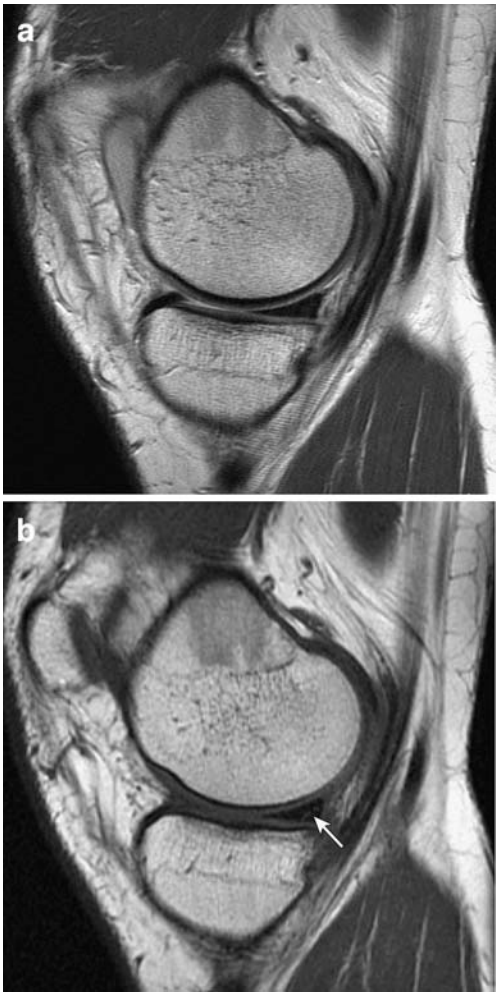

Fig. 3 **a** Sagittal FSE PD-weighted image of a 19-year-old male patient shows no signal or contour abnormality within the posterior horn of the medial meniscus in this arthroscopically confirmed normal meniscus. This study was rated as a 1 (normal) by both readers. **b** The ratings on the sagittal CSE T1-weighted sequences were less definitive (2, probably normal) by the respective readers, as there was trace signal abnormality (*arrows*)

non-inferiority was reached with a lower bound of essentially identical performance.

To assess relative accuracy of the two sequences in differing anatomic locations, data were collected and analyzed for the anterior and posterior halves of each meniscus. For the medial meniscus, the number of tears in the anterior half of the meniscus was too low to perform any meaningful statistical analysis of the anterior half versus the posterior half, while for the lateral, no significant differences were found between the two halves with either sequence.

The results of the analysis of detection of tears of the medial meniscus and lateral meniscus considered separately echo those of the medial and lateral menisci considered together. When considering tears of the medial meniscus, the accuracy of reader 1 was statistically significantly better for FSE PD-weighted sequences versus CSE T1-weighted sequences. For reader 2, the area under the ROC curve for FSE PD-weighted sequences was slightly less than that that of CSE T1-weighted sequences, but still met reasonable criteria for statistical non-inferiority—i.e., there is a 95% chance that, at worst, the area under the ROC curve for FSE PD-weighted sequences is 0.052 less than that of CSE T1-weighted sequences. Taken together, these data suggest that FSE PD-weighted sequences are very likely not inferior and

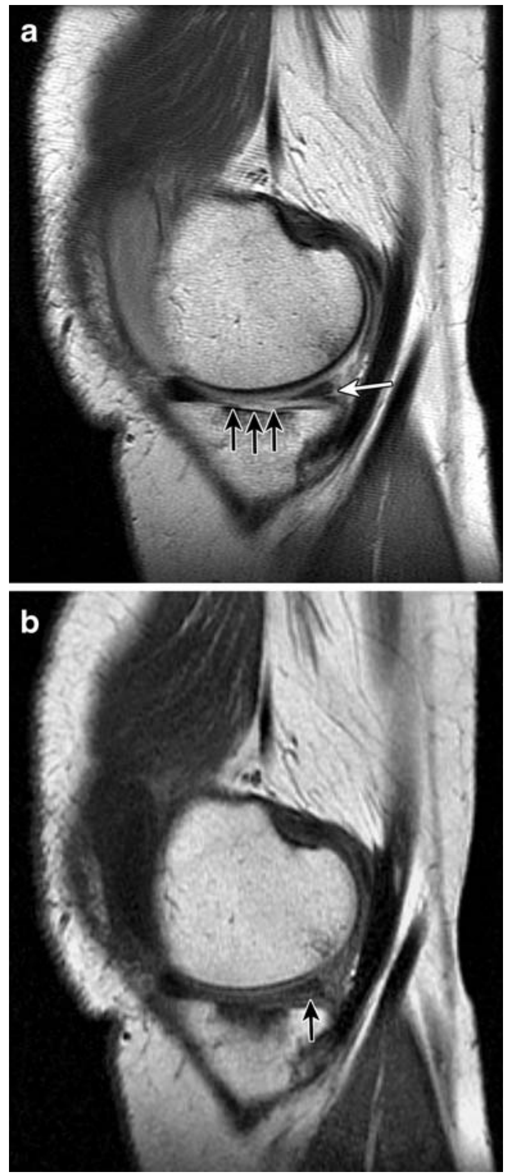

Fig. 4 **a** Sagittal FSE PD-weighted image of a 67-year-old female patient shows signal abnormality extending through the posterior horn of the medial meniscus (*white arrows*). This was an arthroscopically confirmed tear and was rated as a 5 (torn) by both readers. Also, note the articular cartilage defect in the medial tibia (*black arrows*). **b** The signal abnormality on the sagittal CSE T1-weighted sequences is more subtle (*arrow*), and the ratings were less definitive (3 indeterminate and 4 probably torn) by the respective readers

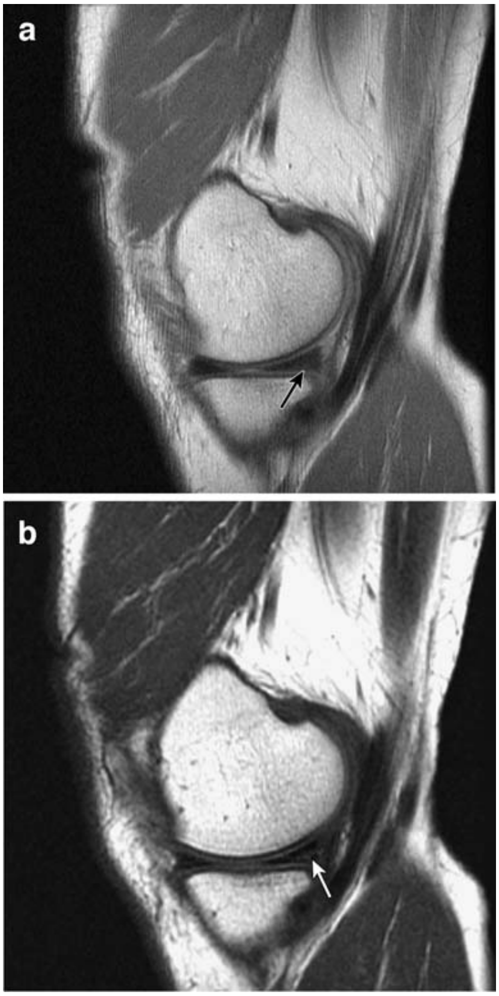

Fig. 5 **a** Sagittal FSE PD-weighted image of a 53-year old female patient shows trace signal abnormality within the posterior horn of the medial meniscus (*arrow*). This study was rated as a 1 (normal) by both readers but was shown to be torn at surgery. **b** The similar trace signal abnormality (*arrow*) on the sagittal T1-weighted sequences were similar and had similar ratings (1 normal and 2 probably normal) by the respective readers

possibly superior to CSE T1-weighted sequences in the detection of medial meniscal tears.

Similarly, with regards to lateral meniscal tears, the results reflect an excellent case for non-inferiority of FSE PD versus CSE T1-weighted sequences. The area under the ROC curve for both readers was greater for the FSE PD as compared with the CSE T1-weighted sequences. The differences did not meet the criteria to say that the FSE PD-weighted sequences were superior, only that they were not inferior—i.e., there is a 95% chance that, at worst, the AUC for FSE PD-weighted sequences is 0.01 (reader 1) or 0.008 (reader 2) less than the AUC for CSE T1-weighted sequences, which essentially implies a lower bound of identical performance.

Using the empirical ROC curves, sensitivity, specificity, PPV, and NPV were determined at various grades. Upon examination of the definitive cut points [grades 1 (normal) and 5 (torn)], there was no difference between FSE PD- and CSE T1-weighted sequences except for reader 2 who, for readings of grade 1, CSE T1-weighted sequences were significantly more sensitive than FSE PD-weighted sequences [91% vs. 77% ($p<0.001$)]. Nevertheless, both sequences demonstrated very good performance from the standpoint of PPV and NPV for grades 5 and 1, respectively. The PPV for a meniscus read as "torn" ranged from 89–92% for both readers of both sequences, while the NPV for a meniscus read as "normal" ranged from 92–96%. Thus, a clinician can predict that a patient with a definitive reading of "normal" on either sequence will have at least a 92% chance of having a normal meniscus on arthroscopic evaluation (Fig. 3).

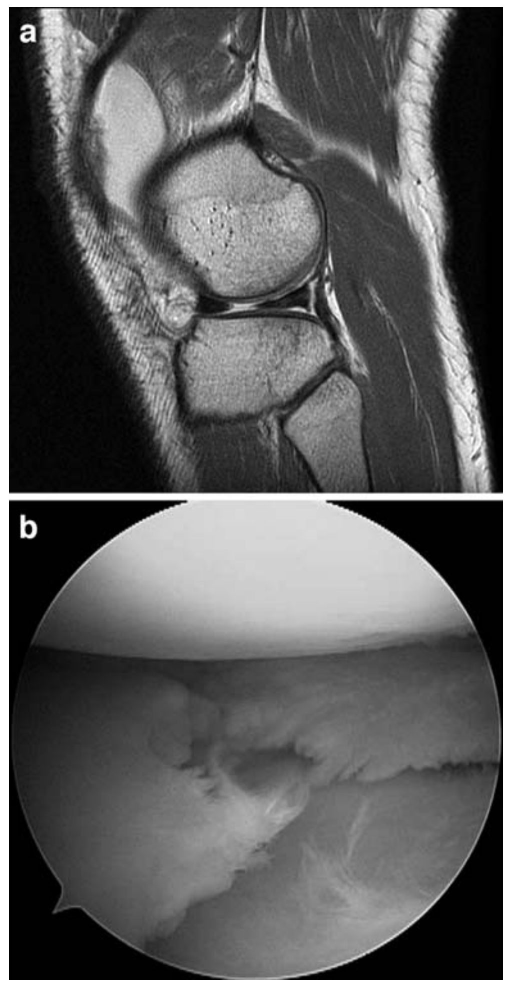

Fig. 6 **a** Sagittal FSE PD-weighted image of a 17-year-old male patient shows a contour abnormality in the peripheral aspect of the anterior horn/body junction of the lateral meniscus. This was rated as a tear by both readers in the anterior half of the lateral meniscus. **b** Arthroscopic image confirms the presence of a radial type tear. In the operative report, this tear was described as posterior to midline in the meniscus and was, thus, a false positive in the anterior half and a false negative in the posterior half for the statistical evaluation

Interestingly, there were significantly more definitive grades (1 and 5) generated by the FSE PD-weighted sequences compared with the CSE T1-weighted sequences (with the exception of grades 5 by reader 2 where the difference was not statistically significant). While this stronger diagnostic response does not imply a greater level of accuracy, it does imply that the readers attributed more confidence to their readings of the FSE PD-weighted sequences when compared with the CSE T1-weighted sequences (Fig. 4). This could be potentially related to the short TE values used with the CSE T1-weighted sequence. Peh et al. have shown, in a porcine model, that TE values below 16 create increased intrasubstance signal in normal menisci [33].

For comparison with other studies as well as for application of the findings of the current study to clinical practice, the less definitive cut points on the ROC curve are useful. When examining the more specific end of the ROC curves at a cut point of grade 4 or greater (probably torn or torn meniscus), FSE PD-weighted sequences were either significantly more specific (reader 2) or not significantly different (reader 1) than CSE T1-weighted sequences. The specificities found for the two sequences in this area of the curves range from 92–97%, which are consistent with results reported in the literature [2]. At the more sensitive end of the ROC curves at a cut point of grade 2 or less (probably normal or normal meniscus), there was no statistically significant difference between the sensitivities of the two sequences. The range of sensitivities (65–76%) is lower than the range reported in the literature [2] (Fig. 5). This is likely caused by several factors. One is that these data reflect a combination of medial and lateral menisci evaluation. In their review of the literature, Rubin and Paletta report the range of sensitivity for the detection of medial meniscal tears of 86% to 96%, while the range of sensitivity for the detection of lateral meniscal tears as 68% to 86% [2]. Another is that the readers did not have access to coronal imaging to confirm negative or equivocal findings. Additionally, lack of access to coronal imaging may have lead to an increased false negative rate (and thus lower sensitivity) due to missing radial tears and tears of the meniscal root, which are often better visualized in the coronal plane. Similarly, radial tears at the midpoint of the meniscus may have been ascribed to the incorrect half of the meniscus and may have lowered the accuracy metrics relative to if the menisci had been evaluated as a whole (Fig. 6). Furthermore, the interpretations of the gold standard of arthroscopic findings will have an effect on what abnormalities are deemed “tears” and what are not. In the current study, all abnormal findings with the exception of marginal “fraying” were classified as tears. In many of the studies including most of those cited by Rubin and Paletta [2], the arthroscopic criteria for meniscal tear or abnormality are not described [8–12, 14, 16, 34, 35]. Others classified a meniscus as torn only if it could be shown to have mechanical instability [13], while still others classified “frayed or degenerated” menisci as intact [15]. The relatively liberal definition of a meniscal tear in the current study likely led to an increased number of false negative MRI results and thereby decreased the overall sensitivity.

There are a number of limitations to this study. While the use of arthroscopic findings as a gold standard has been debated [36], it remains the most reliable means of detecting clinically significant meniscal tears. Prior knowledge of an ongoing study with respect to meniscal tear detection by MRI may have induced the surgeons to be more diligent in the assignment of anatomic location of the tear. It also would have allowed the surgeon to grade the tears in a similar manner (grades 1–5) as the MRI images were graded. However, it may have introduced bias into their reporting by changing their routine assessment of the meniscus. The use of arthroscopic findings also introduced the possibility of bias into the interpretation of the MRI readers as they knew that all of the patients underwent arthroscopy subsequent to their MRI. This was somewhat mitigated by their lack of knowledge of the surgery performed as well as by the relatively high number of normal versus torn meniscal halves that were evaluated (373 versus 131).

Finally, given the wide variation in the parameters of the two sequences, this study neither definitively end the controversy over whether all fast spin-echo sequences can safely replace all conventional spin-echo sequences for meniscal evaluation nor does it definitively find that all PD-weighted sequences are at least as accurate as all T1-weighted sequences for meniscal evaluation. The results of this study do, however, inform both of these issues. Furthermore, they allow us to conclude that for evaluation of the meniscus, our traditional CSE T1-weighted sequence, used primarily for detection of meniscal pathology, can be safely replaced by our FSE PD-weighted sequence that was designed for evaluation of articular cartilage. Finally, it demonstrates the use of ROC analysis to provide a thorough evaluation of the relative accuracy of two MRI sequences for the detection of meniscal tears.

References

1. Fu FH, Harner CD, Vince KG. Knee surgery. Baltimore, MD: Williams & Wilkins; 1994.
2. Rubin DA, Paletta GA Jr. Current concepts and controversies in meniscal imaging. Magn Reson Imaging Clin N Am 2000; 8: 243–270.
3. Mackenzie R, Dixon AK, Keene GS, Hollingworth W, Lomas DJ, Villar RN. Magnetic resonance imaging of the knee: assessment of effectiveness. Clin Radiol 1996; 51: 245–250.

4. Bui-Mansfield LT, Youngberg RA, Warme W, Pitcher JD, Nguyen PL. Potential cost savings of MR imaging obtained before arthroscopy of the knee: evaluation of 50 consecutive patients. AJR Am J Roentgenol 1997; 168: 913–918.
5. Carmichael IW, MacLeod AM, Travlos J. MRI can prevent unnecessary arthroscopy. J Bone Joint Surg Br 1997; 79: 624–625.
6. Ruwe PA, Wright J, Randall RL, Lynch JK, Jokl P, McCarthy S. Can MR imaging effectively replace diagnostic arthroscopy? Radiology 1992; 183: 335–339.
7. Reicher MA, Hartzman S, Duckwiler GR, Bassett LW, Anderson LJ, Gold RH. Meniscal injuries: detection using MR imaging. Radiology 1986; 159: 753–757.
8. Kojima KY, Demlow TA, Szumowski J, Quinn SF. Coronal fat suppression fast spin echo images of the knee: evaluation of 202 patients with arthroscopic correlation. Magn Reson Imaging 1996; 14: 1017–1022.
9. De Smet AA, Graf BK. Meniscal tears missed on MR imaging: relationship to meniscal tear patterns and anterior cruciate ligament tears. AJR Am J Roentgenol 1994; 162: 905–911.
10. Blackmon GB, Major NM, Helms CA. Comparison of fast spin-echo versus conventional spin-echo MRI for evaluating meniscal tears. AJR Am J Roentgenol 2005; 184: 1740–1743.
11. Cheung LP, Li KC, Hollett MD, Bergman AG, Herfkens RJ. Meniscal tears of the knee: accuracy of detection with fast spin-echo MR imaging and arthroscopic correlation in 293 patients. Radiology 1997; 203: 508–512.
12. Crues JV 3rd, Mink J, Levy TL, Lotysch M, Stoller DW. Meniscal tears of the knee: accuracy of MR imaging. Radiology 1987; 164: 445–448.
13. Fischer SP, Fox JM, Del Pizzo W, Friedman MJ, Snyder SJ, Ferkel RD. Accuracy of diagnoses from magnetic resonance imaging of the knee. A multi-center analysis of one thousand and fourteen patients. J Bone Joint Surg Am 1991; 73: 2–10.
14. Kowalchuk RM, Kneeland JB, Dalinka MK, Siegelman ES, Dockery WD. MRI of the knee: value of short echo time fast spin-echo using high performance gradients versus conventional spin-echo imaging for the detection of meniscal tears. Skeletal Radiol 2000; 29: 520–524.
15. Rubin DA, Kettering JM, Towers JD, Britton CA. MR imaging of knees having isolated and combined ligament injuries. AJR Am J Roentgenol 1998; 170: 1207–1213.
16. Justice WW, Quinn SF. Error patterns in the MR imaging evaluation of menisci of the knee. Radiology 1995; 196: 617–621.
17. Bredella MA, Tirman PF, Peterfy CG, et al. Accuracy of T2-weighted fast spin-echo MR imaging with fat saturation in detecting cartilage defects in the knee: comparison with arthroscopy in 130 patients. AJR Am J Roentgenol 1999; 172: 1073–1080.
18. Broderick LS, Turner DA, Renfrew DL, Schnitzer TJ, Huff JP, Harris C. Severity of articular cartilage abnormality in patients with osteoarthritis: evaluation with fast spin-echo MR vs arthroscopy. AJR Am J Roentgenol 1994; 162: 99–103.
19. Potter HG, Linklater JM, Allen AA, Hannafin JA, Haas SB. Magnetic resonance imaging of articular cartilage in the knee. An evaluation with use of fast-spin-echo imaging. J Bone Joint Surg Am 1998; 80: 1276–1284.
20. Anderson MW, Raghavan N, Seidenwurm DJ, Greenspan A, Drake C. Evaluation of meniscal tears: fast spin-echo versus conventional spin-echo magnetic resonance imaging. Acad Radiol 1995; 2: 209–214.
21. Escobedo EM, Hunter JC, Zink-Brody GC, Wilson AJ, Harrison SD, Fisher DJ. Usefulness of turbo spin-echo MR imaging in the evaluation of meniscal tears: comparison with a conventional spin-echo sequence. AJR Am J Roentgenol 1996; 167: 1223–1227.
22. Metz CE, Herman BA, Roe CA. Statistical comparison of two ROC-curve estimates obtained from partially-paired datasets. Med Decis Making 1998; 18: 110–121.
23. Dorfman DD, Berbaum KS, Metz CE. Receiver operating characteristic rating analysis. Generalization to the population of readers and patients with the jackknife method. Invest Radiol 1992; 27: 723–731.
24. Hillis SL, Berbaum KS. Monte Carlo validation of the Dorfman–Berbaum–Metz method using normalized pseudovalues and less data-based model simplification. Acad Radiol 2005; 12: 1534–1541.
25. Pesce LL, Metz CE. Reliable and computationally efficient maximum-likelihood estimation of “proper” binormal ROC curves. Acad Radiol 2007; 14: 814–829.
26. Pepe MS. The statistical evaluation of medical tests for classification and prediction. New York: Oxford University Press; 2004.
27. Rubin DA, Kneeland JB, Listerud J, Underberg-Davis SJ, Dalinka MK. MR diagnosis of meniscal tears of the knee: value of fast spin-echo vs conventional spin-echo pulse sequences. AJR Am J Roentgenol 1994; 162: 1131–1135.
28. Iinuma G, Ushio K, Ishikawa T, Nawano S, Sekiguchi R, Satake M. Diagnosis of gastric cancers: comparison of conventional radiography and digital radiography with a 4 million-pixel charge-coupled device. Radiology 2000; 214: 497–502.
29. Park SH, Goo JM, Jo CH. Receiver operating characteristic (ROC) curve: practical review for radiologists. Korean J Radiol 2004; 5: 11–18.
30. Obuchowski NA. Receiver operating characteristic curves and their use in radiology. Radiology 2003; 229: 3–8.
31. Mushlin AI, Detsky AS, Phelps CE, et al. The accuracy of magnetic resonance imaging in patients with suspected multiple sclerosis. The Rochester–Toronto Magnetic Resonance Imaging Study Group. JAMA 1993; 269: 3146–3151.
32. Wagner RF, Beiden SV, Campbell G, Metz CE, Sacks WM. Assessment of medical imaging and computer-assist systems: lessons from recent experience. Acad Radiol 2002; 9: 1264–1277.
33. Peh WCG, Chan JHM, Shek TWH, Wond JWK. The effect of using shorter echo times in MR imaging of knee menisci: a study using a porcine model. AJR 1999; 172: 485–488.
34. Quinn SF, Brown TR, Szumowski J. Menisci of the knee: radial MR imaging correlated with arthroscopy in 259 patients. Radiology 1992; 185: 577–580.
35. De Smet AA, Tuite MJ. Use of the “two-slice-touch” rule for the MRI diagnosis of meniscal tears. AJR Am J Roentgenol 2006; 187: 911–914.
36. Quinn SF, Brown TF. Meniscal tears diagnosed with MR imaging versus arthroscopy: how reliable a standard is arthroscopy? Radiology 1991; 181: 843–847.