Comparison of Arthroscopic and MRI Findings of Osteochondral Damage in Knees


Abstract

Magnetic resonance imaging (MRI) scans are widely used in the assessment of knees, often prior to arthroscopic procedures. The reporting of osteochondral damage on MRI scans can be variable. The correlation between MRI reports of osteochondral damage and that found at arthroscopy is often inconsistent.

A retrospective case-note review of a single-surgeon series of 175 arthroscopic procedures was performed. Eighty-three patients were included in the study. The remainder were excluded if an MRI scan had not been performed or had been performed more than 3-months before surgery. The condition of the articular cartilage demonstrated by MRI was compared to that found at arthroscopy. Data was analysed for presence and extent of osteochondral damage.

Comparison between MRI and arthroscopy findings showed high specificity (90%) and negative predictive values (89%) for osteochondral damage but low sensitivity (46%). Cohen’s kappa values < 0.2 revealed very poor correlation for the extent of damage.

This study demonstrates MRI as a good identifier of osteochondral damage but an unreliable descriptor for such change.

Magnetic resonance imaging (MRI) scans are widely used in the orthopaedic assessment of knee pathology. In UK hospitals, radiologists commonly report osteochondral and articular damage on knee MRI scans. These reports aid pre-operative planning, and counselling of patients, but the reliability of this information is unknown. This retrospective study assesses the accuracy of these reports by comparing them with arthroscopic findings for a series of patients.

The role of MRI scans in the diagnosis of meniscal and ligamentous pathology is well-documented. However, the role of MRI in describing osteochondral damage is less clear-cut. Some investigators suggest MRI lacks the diagnostic strengths of arthroscopy, whereas others suggest MRI is a reliable indicator of osteochondral damage.

A problematic issue in this area of research are the varied radiological scoring systems used in osteochondral description, which include the following:

- BLOKS (Boston Leeds Osteoarthritis Knee Score)
- KOSS (Knee Osteoarthritis Scoring System)
- SFA-MR (French Society of Arthroscopy MRI Score)

None of these directly correlated with the widely used Outerbridge arthroscopic grading.

As a result, the literature is not directly applicable to clinical practice due to the different scoring methodology.

This study has used a simplified osteochondral scoring system, in-line with the Outerbridge classification, to assess our series of patients at Musgrove Park Hospital (MPH).

Method

The operation notes from every arthroscopic knee procedure performed by the senior investigator (JLW) from December 2006 to July 2009 were collated (N = 175). Patients with reported MRI scans within 3-months of arthroscopy were included in the study (N = 83). Patients without an MRI or those MRIs performed more than 3-months preoperatively were excluded from the study (N = 92). All scans were available on the hospital digital picture archiving and communications system (PACS).

MRI at our hospital uses a 0.2 Tesla magnetic resonance scanner giving sagittal PD/IW fat-saturated, coronal T1W/IW fat-saturated, and axial T2 TSE sequences.
Knee procedures performed included meniscectomy, meniscal repair, ACL reconstruction, chondroplasty, and EUA. Patient details were anonymized. Information on age, gender, operation date, arthroscopy findings, and MRI reports of osteochondral damage were recorded.

A novel classification system was used in this study (Table 1). This scoring system is in-line with the Outerbridge Arthroscopic scoring system. Osteochondral damage was classified as none, mild, moderate, or severe and scored from 0 to 3, according to damage. The medial, lateral, and patellofemoral compartments were scored separately. The scoring of written MRI reports (available on PACS) and the operation notes was performed by one researcher. Where no comment was made on the MRI report regarding the articular surface, this was classified as normal cartilage (0). Where there was difficulty in ascribing a classification to a report, consensus was gained by discussion with two senior orthopaedic surgeons.

Data was gathered to enable sensitivity, specificity, and agreement analysis. Arthroscopy was defined as the gold standard for sensitivity and specificity testing. For sensitivity and specificity, a study score of 0 was described as a negative test result and scores of 1 to 3 a positive test result. Cohen’s Kappa was calculated to indicate the level of agreement between the MRI and arthroscopy scores (0-3).

Results
There were 83 patients (57 male, 26 female) included in the study with a mean age of 37 (Table 2).

All operation notes commented on osteochondral damage (N = 175). All MRI scans were reported (N = 83). Tables 3 and 4 show the results.

The results show high specificity and high negative predictive value (NPV) and low sensitivity and low positive predictive values (PPV). The high specificity indicates MRI is a strong tool for confirming the presence of osteochondral damage. The confidence intervals (mean lower border = 80%) strengthen the assertion that MRI reporting is adequately specific.

The high NPV suggests MRI does pick up a high proportion of healthy patients, but the sample population was predominantly healthy. This does not reflect the sensitivity of the test, however, which is shown to be weak. Sensitivity of 46% suggests MRI cannot be used to rule out the presence of osteochondral damage.

The positive predictive values (PPV) are significantly

<table>
<thead>
<tr>
<th>Outerbridge Grade</th>
<th>Description of Articular Cartilage</th>
<th>Study Grade of Osteochondral Damage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Normal cartilage</td>
<td>None</td>
</tr>
<tr>
<td>I</td>
<td>Cartilage with softening and swelling</td>
<td>Mild</td>
</tr>
<tr>
<td>II</td>
<td>A partial-thickness defect with fissures on the surface that do not reach subchondral bone and are &lt; 1.5 cm in diameter</td>
<td>Moderate</td>
</tr>
<tr>
<td>III</td>
<td>Fissuring to the level of subchondral bone in an area with a diameter &gt; 1.5 cm</td>
<td>Moderate</td>
</tr>
<tr>
<td>IV</td>
<td>Exposed subchondral bone</td>
<td>Severe</td>
</tr>
</tbody>
</table>

Table 1 Study Scoring System

<table>
<thead>
<tr>
<th>Joint Compartment</th>
<th>Sensitivity</th>
<th>Confidence Interval</th>
<th>Specificity</th>
<th>Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patellofemoral</td>
<td>57%</td>
<td>37% – 76%</td>
<td>90%</td>
<td>80% – 95%</td>
</tr>
<tr>
<td>Lateral</td>
<td>40%</td>
<td>12% – 77%</td>
<td>85%</td>
<td>75% – 91%</td>
</tr>
<tr>
<td>Medial</td>
<td>41%</td>
<td>22% – 64%</td>
<td>94%</td>
<td>85% – 98%</td>
</tr>
</tbody>
</table>

Table 3 Sensitivity, Specificity, and 95% Confidence Intervals

<table>
<thead>
<tr>
<th>Joint Compartment</th>
<th>Prevalence</th>
<th>Positive Predictive Value</th>
<th>Negative Predictive Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patellofemoral</td>
<td>25%</td>
<td>67%</td>
<td>86%</td>
</tr>
<tr>
<td>Lateral</td>
<td>6%</td>
<td>14%</td>
<td>96%</td>
</tr>
<tr>
<td>Medial</td>
<td>13%</td>
<td>64%</td>
<td>86%</td>
</tr>
</tbody>
</table>
worse than the NPV, which is a reflection of the predominantly healthy population. The poor PPV (mean = 48%) suggests that although MRI is an adequately specific test, it fails to recognize half of the unhealthy patients in a healthy population (mean disease prevalence = 15%).

Cohen’s Kappa was calculated as a measure of agreement between MRI and arthroscopy findings (Table 5).

A Kappa of less than or equal to 0 indicates no agreement.12

The Kappa of 0.23 in comparison of findings in the patellofemoral compartment indicates slight agreement, though in both medial and lateral compartments the value was less than 1. These findings show that MRI compares poorly with arthroscopic findings.

The variability of results between the different joint compartments suggests MRI is a more specific test in the patellofemoral and medial compartments. Although the prevalence of lateral joint degeneration was low in this study, the widths of the confidence intervals are uniform across the three compartments, suggesting the estimates for specificity are equally reliable. The reduced prevalence of lateral degeneration does have an impact on the PPV and NPV for this compartment, suggesting these values are not comparable across compartments.

Discussion

This study does have various weaknesses relating to its design. Firstly, the relatively small sample size (N = 83) means that the study is underpowered. The exclusion criterion (lack of PACS MRI report) used in this study also potentially introduced selection bias. An alternative method would have been the analysis of the scans themselves and scoring directly from this analysis.

Another weakness is the quantification of subjective data (i.e., converting subjective comments made on MRI reports into the categorical data for normal, mild, moderate, and severe). In the absence of a validated comparison between arthroscopic and radiological scoring, this approach has the advantage of utility but the disadvantage of introducing potential bias.

A single researcher performed the scoring, with consensus sought only in cases of ambiguity; this potentially introduces measurement bias but eliminates the possibility of inter-observer error. Another approach to the data collection process would be for multiple assessors to score the data sets independently.

Due to the retrospective nature of this study, the surgeons were not blinded to the MRI reports. The results may, therefore, be subject to measurement bias; however, it would seem unethical and inappropriate to blind surgeons to important clinical information.

Despite these limitations, our findings are comparable with other studies. In a study by von Englehardt and associates,13 high specificity (90% to 95%) and low sensitivity (29% to 74%) with regards to the MRI grading of osteochondral defects in all compartments of the knee was found. This group carried out further research that also demonstrated high specificities and NPVs.14

These findings are similar to the study performed by Friemert and coworkers,5 who also found MRI had high specificity (97% to 99%) and NPVs when looking at articular cartilage in the knee.

A study by Vaz and associates15 showed high specificity 94.9% and poorer sensitivity 76.1%, when assessing articular damage. They also showed good positive predictive value (94.7%) and positive negative value (76.9%).

Bredella and colleagues16 showed that they were able to achieve high sensitivity and specificity, 94% and 99% respectively, in a 7 spin sequence MRI protocol. However, looking at the sequences in isolation, it is clear that individually the sensitivity is lower (range: 40% to 94%; mean: 69%) than the specificity (range: 99% to 100%; mean: 99%).

Conclusion

The results of this study demonstrate that MRI is an adequately specific diagnostic tool for identifying the presence of osteochondral or articular damage, but once recognized, this damage is not reliably described.

MRI also shows poor sensitivity, meaning damage cannot be ruled out.

Despite limitations in terms of study size and standardization of subjective outcome measurements, this study does reflect the situation in a UK DGH Orthopaedic Department.

This study shows MRI is a highly specific test for osteochondral knee joint damage. However, MRI is a poor at excluding (low sensitivity) or describing (poor agreement) any pathology.

Future research in this area should aim to provide suitably powered studies using data from multiple surgeons. A validated MRI grading systems correlated with the Outerbridge classification system would also be of benefit.

Disclosure Statement

None of the authors have a financial or proprietary interest in the subject matter or materials discussed, including, but not limited to, employment, consultancies, stock ownership, honoraria, and paid expert testimony.

References