Magnetic Resonance Imaging Identification of Rotator Cuff Retears After Repair

Interobserver and Intraobserver Agreement

Michael Khazzam,*† MD, John E. Kuhn,‡ MD, Ed Mulligan,† DPT, Joseph A. Abboud,§ MD, Keith M. Baumgarten,** MD, Robert H. Brophy,†† MD, Grant L. Jones,‡‡ MD, Bruce Miller,§§ MD, Matthew Smith,¶¶ MD, and Rick W. Wright,** MD

Investigation performed at Vanderbilt University Medical Center
Sports Medicine & Shoulder Service, Nashville, Tennessee

Background: Magnetic resonance imaging (MRI) is the most commonly used imaging modality to assess the rotator cuff. Currently, there are a limited number of studies assessing the interobserver and intraobserver reliability of MRI after rotator cuff repair.

Hypothesis: Fellowship-trained orthopaedic shoulder surgeons will have good inter- and intraobserver agreement with regard to features of the repaired rotator cuff (repair integrity, fat content, muscle volume, number of tendons involved, tear size, and retract) on MRI.

Study Design: Cohort study (diagnosis); Level of evidence, 3.

Methods: Seven fellowship-trained orthopaedic shoulder surgeons reviewed 31 MRI scans from 31 shoulders from patients who had previous rotator cuff repair. The scans were evaluated for the following characteristics: rotator cuff repair status (full-thickness retear vs intact repair), tear location, tendon thickness, fatty infiltration, atrophy, number of tendons involved in retear, tendon retraction, status of the long head of the biceps tendon, and bone marrow edema in the humeral head. Surgeons were asked to review images at 2 separate time points approximately 9 months apart and complete an evaluation form for each scan at each time point. Multirater kappa (κ) statistics were used to assess inter- and intraobserver reliability.

Results: The interobserver agreement was highest (80%, κ = 0.60) for identifying full-thickness retears, tendon retear retraction (64%, κ = 0.45), and cysts in the greater tuberosity (72%, κ = 0.43). All other variables were found to have fair to poor agreement. The worst interobserver agreement was associated with identifying rotator cuff footprint coverage (47%, κ = –0.21) and tendon signal intensity (29%, κ = –0.01). The mean intraobserver reproducibility was also highest (77%-90%, κ = 0.71) for full-thickness retears, quality of the supraspinatus (47%-83%, κ = 0.52), tears of the long head of the biceps tendon (58%-94%, κ = 0.49), presence of bone marrow edema in the humeral head (63%-87%, κ = 0.48), cysts in the greater tuberosity (70%-83%, κ = 0.47), signal in the long head of the biceps tendon (60%-80%, κ = 0.43), and quality of the infraspinatus (37%-90%, κ = 0.43). The worst intraobserver reproducibility was found in identification of the location of bone marrow edema (22%-83%, κ = –0.03).

Conclusion: The results of this study indicate that there is substantial variability when evaluating MRI scans after rotator cuff repair. Intact rotator cuff repairs or full-thickness retears can be identified with moderate reliability. These findings indicate that additional imaging modalities may be needed for accurate assessment of the repaired rotator cuff.

Keywords: rotator cuff repair; magnetic resonance imaging; shoulder; interobserver agreement; intraobserver agreement

Although rotator cuff repair is considered a successful procedure with regard to pain relief and restoration of function, the retear rate has been reported as high as 40%.†† The advancement of magnetic resonance imaging (MRI) has allowed a more detailed examination of the rotator cuff. Currently, several methods, based on imaging modalities reported in the literature,9,13,14,34-36 have been used to analyze the characteristics of the rotator cuff to better classify its properties. Similarly, factors such as tear size, muscle atrophy, fatty degeneration, and tear configuration have been used to determine prognosis after rotator cuff repair surgery.

Several studies have reported the interobserver agreement for assessment of the preoperative rotator cuff as visualized on MRI.1,26,29,30,33 This information is an important factor for determining the reliability of the diagnostic test being used (in this case MRI). These studies demonstrated interobserver agreement to be highest in identifying full-thickness tears from partial-thickness tears and the number of tendons involved in full-thickness tears. Only 1 study has reported the accuracy in evaluating the
repaired rotator cuff via MRI.\textsuperscript{25} To our knowledge, no studies have assessed the interobserver and intraobserver reliability of MRI after rotator cuff repair reported in the literature.

The purpose of this study is to evaluate the MRI features of the rotator cuff after rotator cuff repair. We hypothesize that fellowship-trained orthopaedic shoulder surgeons will have good to excellent interobserver and intraobserver agreement with regard to assessing the characteristics of the repaired rotator cuff on MRI.

MATERIALS AND METHODS

Institutional review board approval was obtained before performing this study. Seven fellowship-trained orthopaedic shoulder surgeons were asked to review 31 MRI scans from patients evaluated for shoulder pain and who had previously undergone rotator cuff repair. All surgeons had completed a 1-year fellowship in either sports medicine or shoulder and elbow surgery. The 31 patients represent a consecutive series of patients who were evaluated by 1 surgeon over a 7-year period (December 2003 to March 2010) and had a previous rotator cuff repair, an available MRI scan of the shoulder after repair, a rotator cuff repair performed without the use of metallic implants (ie, metal anchors that would produce image artifact and impair the visibility of the rotator cuff footprint), and an MRI scan using a magnet strength at a minimum 1.5-Tesla without contrast (19 patients, 1.5 Tesla magnet; 12 patients, 3.0 Tesla magnet). Slice thickness ranged between 2.5 and 4.0 mm (2.5 mm, 1 patient; 3.0 mm, 21 patients; 3.5 mm, 7 patients; and 4.0 mm, 2 patients). Images were compiled by an independent reviewer who did not participate in the MRI scan assessment.

All surgeons were sent compact disks containing all 31 MRI scans and were required to complete standardized evaluation forms at their leisure. All reviewers were blinded to all clinical data related to each patient and knew only that each patient had undergone a rotator cuff repair in the past. The assessment form was developed by the authors (please refer to Appendix 1, available in the online version of this article at http://ajs.sagepub.com/supplemental/) and consisted of variables that have been reported in previous studies evaluating rotator cuff repair integrity.\textsuperscript{11}\textsuperscript{11} These variables included the following: full-thickness retear or intact rotator cuff repair (Figures 1 and 2); location of partial-thickness tears (if present); “tear classification” (articular or bursal surface); footprint size; tendon thickness; tendon signal intensity; presence of long head of the biceps abnormality (increased signal, longitudinal split); presence of cysts in the greater tuberosity; glenohumeral joint effusion; bone marrow edema in the humeral head and location; Goutallier classification\textsuperscript{13}; quality of the supraspinatus, infraspinatus, subscapularis, and teres minor (atrophy grading); acromiohumeral distance; location of retear (tuberosity or musculotendinous junction); tear size (coronal and sagittal planes); number of tendons involved in the retear; and amount of retraction. Amount of retraction was classified as minimal retraction (within 5 mm of the greater or lesser tuberosity), retracted to the humeral head, retracted to the glenohumeral joint, or retracted medial to the glenoid. Muscle quality was classified according to the Goutallier\textsuperscript{13} system. Muscle atrophy of the supraspinatus was assessed as described by Zanetti et al.\textsuperscript{36} Atrophy of the infraspinatus, teres minor, and subscapularis was evaluated using methods described previously.\textsuperscript{9,23,30} Each reviewer was instructed to return all images after the first assessment to decrease bias on repeat evaluation for intraobserver reliability.

Reviewers were sent a second MRI evaluation series with the same MRI scans and standardized form approximately 6 months after the return of the first assessment.
Statistical Methods

Multirater kappa (κ) statistics were used to assess intraobserver and interrater reliability among 7 orthopaedic surgeons. Interobserver reliability was calculated using each reviewer’s initial assessment of the MRI scans. The κ value is a chance-adjusted measure of agreement. Fleiss κ coefficients were used for the interrater assessment, whereas the Cohen κ coefficient was used to report the intraobserver agreements. A κ of 0.00 represents agreement equivalent with random chance alone, whereas a κ of 1.00 represents perfect agreement. A negative κ represents worse than what would be expected due to chance alone, whereas a κ of –1.00 represents complete discordance between observers. Analysis was performed using AgreeStat 2011 (Advanced Analytics, LLC, Gaithersburg, Maryland). Kappa values were interpreted according to the guidelines proposed by Landis and Koch.19 Excellent agreement occurs for values between 0.81 and 1.00, good for values between 0.61 and 0.80, moderate for values between 0.41 and 0.60, fair for values between 0.21 and 0.40, and poor for values under 0.20.

RESULTS

Complete results are shown in Tables 1 and 2 and Appendix 2 (available online). All 7 reviewers completed evaluation of the 31 MRI at 2 separate time points with an average of 9 months (range, 7-10 months) between each assessment.

The interobserver agreement was highest (80%, κ = 0.60; moderate agreement) for the identification of a full-thickness rotator cuff retear as compared with intact repairs. The reviewers were also able to identify the amount of tendon retraction with moderate agreement (64%, κ = 0.45) and the presence of cysts in the greater tuberosity (72%, κ = 0.43; moderate agreement). Fair interobserver agreement was found for number of tendons involved in full-thickness tears (69%, κ = 0.40), size of tear in the sagittal plane (49%, κ = 0.31), presence of tears involving the long head of the biceps tendon (71%, κ = 0.35), Goutallier classification (46%, κ = 0.24), presence and size of a glenohumeral joint effusion (53%, κ = 0.24), bone marrow edema in the humeral head (63%, κ = 0.22), acromiohumeral distance (59%, κ = 0.31), and muscle quality of the supraspinatus (54%, κ = 0.31) and infraspinatus (68%, κ = 0.28). Poor interobserver agreement was found regarding tear classification, repair integrity, footprint coverage, tendon thickness, tendon signal intensity, retear location, tear size in the coronal plane, presence of signal involving the long head of the biceps tendon, and muscle quality of the subscapularis and teres minor.

The intraobserver reliability was strongest for detecting full-thickness retears of the repaired rotator cuff versus intact rotator cuff repairs (range, 77%-90%; κ = 0.71; good to moderate agreement). Muscle quality of the supraspinatus (47%-83%; κ = 0.52) and infraspinatus (37%-90%; κ = 0.43), presence of a split in the long head of the biceps tendon (58%-94%; κ = 0.49), cysts in the greater tuberosity (70%-83%; κ = 0.47), and bone marrow edema involving the humeral head (63%-87%; κ = 0.48) all demonstrated fair to moderate agreement. All of the remaining variables had fair to poor intraobserver reliability.
TABLE 2
Intraobserver Reliability

<table>
<thead>
<tr>
<th>Variable</th>
<th>Agreement Range</th>
<th>( \kappa )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intact repair vs full-thickness retear present</td>
<td>0.77-0.90</td>
<td>0.71</td>
</tr>
<tr>
<td>Supraspinatus</td>
<td>0.47-0.83</td>
<td>0.52</td>
</tr>
<tr>
<td>Long head of the biceps (LHB) split</td>
<td>0.58-0.94</td>
<td>0.49</td>
</tr>
<tr>
<td>Bone marrow edema humeral head</td>
<td>0.63-0.87</td>
<td>0.48</td>
</tr>
<tr>
<td>cysts of the greater tuberosity</td>
<td>0.70-0.83</td>
<td>0.47</td>
</tr>
<tr>
<td>LHB signal</td>
<td>0.60-0.80</td>
<td>0.43</td>
</tr>
<tr>
<td>infraspinatus</td>
<td>0.37-0.90</td>
<td>0.43</td>
</tr>
<tr>
<td>Retraction</td>
<td>0.43-0.73</td>
<td>0.39</td>
</tr>
<tr>
<td>acromiohumeral distance</td>
<td>0.50-0.83</td>
<td>0.38</td>
</tr>
<tr>
<td>Tendon signal intensity</td>
<td>0.24-0.44</td>
<td>0.09</td>
</tr>
<tr>
<td>Teres minor</td>
<td>0.83-0.97</td>
<td>0.36</td>
</tr>
<tr>
<td>Goutallier grade</td>
<td>0.22-0.80</td>
<td>0.34</td>
</tr>
<tr>
<td>Subscapularis</td>
<td>0.53-0.98</td>
<td>0.32</td>
</tr>
<tr>
<td>Glenohumeral joint effusion</td>
<td>0.37-0.77</td>
<td>0.30</td>
</tr>
<tr>
<td>No. of tendons involved</td>
<td>0.40-0.73</td>
<td>0.29</td>
</tr>
<tr>
<td>Sagittal tear size</td>
<td>0.25-0.53</td>
<td>0.23</td>
</tr>
<tr>
<td>Retear location</td>
<td>0.55-0.87</td>
<td>0.21</td>
</tr>
<tr>
<td>Coronal tear size</td>
<td>0.15-0.57</td>
<td>0.20</td>
</tr>
<tr>
<td>Tendon thickness has previous rotator repair</td>
<td>0.22-0.67</td>
<td>0.13</td>
</tr>
<tr>
<td>Rotator cuff repair integrity</td>
<td>0.17-0.73</td>
<td>0.11</td>
</tr>
<tr>
<td>Footprint coverage</td>
<td>0.31-0.72</td>
<td>0.10</td>
</tr>
<tr>
<td>Tendon signal intensity</td>
<td>0.24-0.44</td>
<td>0.09</td>
</tr>
<tr>
<td>Tear classification</td>
<td>0.10-0.53</td>
<td>0.06</td>
</tr>
<tr>
<td>Bone marrow edema location</td>
<td>0.22-0.83</td>
<td>-0.03</td>
</tr>
</tbody>
</table>

**DISCUSSION**

The structural healing of the repaired rotator cuff to the anatomic footprint has traditionally been evaluated with the use of MRI. Multiple studies have demonstrated that clinical success does not often correlate with structural tendon-to-bone healing. With the continued development of repair techniques and postoperative physical therapy protocols aimed at improving tendon-to-bone healing success, it is important to know how accurate the diagnostic test (ie, MRI) is at identifying abnormalities and the reliability of these findings. The purpose of this study was to assess the interobserver and intraobserver reliability of MRI findings in shoulders that have had previous rotator cuff repair among fellowship-trained shoulder surgeons.

We found moderate interobserver agreement for predicting full-thickness retears, tendon retraction, and presence of greater tuberosity cysts on MRI of previously repaired rotator cuffs. Upon repeat MRI evaluation, the reviewers agreed with their previous evaluations most closely in detecting full-thickness tears, long head of the biceps tendon abnormalities (presence of signal or tear), quality of the supraspinatus and infraspinatus tendons, presence of cysts in the greater tuberosity, and bone marrow edema in the humeral head. Owen et al examined the accuracy of MRI in evaluating the shoulder after surgery in 31 patients before reoperation and correlated operative findings with MRI findings. These authors found 90% accuracy for findings consistent with full-thickness retearing of the repaired rotator cuff. This study used the presence of fluid signal intensity on T2-weighted images extending through the rotator cuff or the nonvisualized portion of the rotator cuff as criteria for identifying a retear.

In this study, the Goutallier grading system to assess muscle quality and fatty degeneration demonstrated fair interobserver agreement (46%, \( \kappa = 0.24 \)) and intraobserver reliability (54%, \( \kappa = 0.34 \)). Although we are unaware of any studies examining the interobserver or intraobserver agreement in this setting, Spencer et al, Fuchs et al, and Oh et al reported interobserver reliability for rotator cuff tears before surgical repair. Spencer et al reported poor interobserver agreement (\( \kappa = 0.10 \)), Fuchs et al reported excellent interobserver agreement (\( \kappa = 0.86\)–0.93), and Oh et al reported good interobserver agreement (\( \kappa = 0.60\)–0.72) and fair to excellent intraobserver reliability (\( \kappa = 0.26\)–0.81). The results of our study demonstrated worse agreement than what has been reported for MRI fatty atrophy evaluation of the rotator cuff before repair. Given the variability in the literature regarding reported Goutallier grade, we believe these findings further support our results that it is difficult to assess rotator cuff muscle quality and fatty degeneration, especially in the postsurgical setting. Not only do individuals not agree with each other, but also studies have not demonstrated agreement in the reported results. We do not believe that the differences found on Goutallier grading can be attributed to imaging technique or quality. The MRI scan quality in the previous studies is of comparable magnet strength and slice thickness. The previous studies that examined the Goutallier grade of the rotator cuff examined shoulders before surgical repair but also demonstrated inconsistent findings ranging from excellent to poor agreement. We cannot hypothesize as to why there is such variability in the reliability of the Goutallier grading as demonstrated upon comparison of our results with what has been previously reported in the literature.

Use of MRI is a reliable modality for evaluating full-thickness rotator cuff tears before surgical repair with good interobserver agreement (\( \kappa > 0.61 \)) but is not as reliable for determining partial-thickness tearing (fair to poor interobserver agreement), especially in relation to side (bursal or articular) and grade of the partial-thickness tear. This is similar to our findings for repaired rotator cuff tears with moderate (\( \kappa = 0.60 \)) interobserver agreement and excellent (\( \kappa = 0.85 \)) intraobserver reliability for detection of full-thickness tears. Identification of full-thickness retears provided the highest interobserver reliability results in our study, indicating that after rotator cuff repair, shoulder surgeons can predictably identify full-thickness retears on MRI.

Several factors may contribute to the difficulty encountered on agreement of MRI findings after rotator cuff repair as opposed to the shoulder that has not undergone surgery. These variables include artifact from the suture anchor; increased signal within the tendon that could be degeneration, partial retear, or scar; and tendon-to-bone healing that does not restore the normal footprint anatomy, which may make it difficult to determine a normal-healing tendon from abnormal failure to heal or retear.

\[\text{References 2-4, 8, 10, 17, 18, 21, 28, 35.}\]
One weakness of this study is that shoulders were assessed only using MRI. Therefore, we did not have a comparison group to determine if another imaging modality (magnetic resonance arthrogram, computed tomography [CT] arthrogram, ultrasound) improves interobserver and intraobserver reliability in evaluating the status of the repaired rotator cuff. In addition, this study did not include arthroscopic assessment of these shoulders to determine if the findings on MRI correctly predict the status of the rotator cuff in this setting. Therefore, we were unable to comment on the accuracy of MRI following rotator cuff repair at detecting rotator cuff abnormalities.

In conclusion, this study demonstrates that fellowship-trained shoulder surgeons had variable interobserver agreement and intraobserver variability in interpreting MRI scans after rotator cuff repair. Although best agreement was found regarding full-thickness retearing, this was only moderate, but the reviewers did agree with themselves when they reevaluated the scans with excellent intraobserver reliability. It is important to point out that many variables frequently reported in outcomes studies on the success of rotator cuff repair demonstrated only fair to poor interobserver and intraobserver agreement—namely, tear size, Goutallier classification, number of tendons involved, tear size, footprint coverage, and rotator cuff repair integrity. The findings of this study indicate that future investigations on the outcome of rotator cuff repair should consider using other imaging modalities to better define success of tendon-to-bone healing and repair status such as ultrasound, CT, or MR arthrogram to provide more reliable data.

**REFERENCES**


For reprints and permission queries, please visit SAGE’s Web site at http://www.sagepub.com/journalsPermissions.nav