Accuracy of MRI, MR Arthrography, and Ultrasound in the Diagnosis of Rotator Cuff Tears: A Meta-Analysis

OBJECTIVE. The purpose of this study was to compare the diagnostic accuracy of MRI, MR arthrography, and ultrasound for the diagnosis of rotator cuff tears through a meta-analysis of the studies in the literature.

MATERIALS AND METHODS. Articles reporting the sensitivities and specificities of MRI, MR arthrography, or ultrasound for the diagnosis of rotator cuff tears were identified. Surgical (open and arthroscopic) reference standard was an inclusion criterion. Summary statistics were generated using pooled data. Scatterplots of the data sets were plotted on a graph of sensitivity versus (1 – specificity). Receiver operating characteristic (ROC) curves were generated.

RESULTS. Sixty-five articles met the inclusion criteria for this meta-analysis. In diagnosing a full-thickness tear or a partial-thickness rotator cuff tear, MR arthrography is more sensitive and specific than either MRI or ultrasound (p < 0.05). There are no significant differences in either sensitivity or specificity between MRI and ultrasound in the diagnosis of partial- or full-thickness rotator cuff tears (p > 0.05). Summary ROC curves for MR arthrography, MRI, and ultrasound for all tears show the area under the ROC curve is greatest for MR arthrography (0.935), followed by ultrasound (0.889) and then MRI (0.878); however, pairwise comparisons of these curves show no significant differences between MRI and ultrasound (p > 0.05).

CONCLUSION. MR arthrography is the most sensitive and specific technique for diagnosing both full- and partial-thickness rotator cuff tears. Ultrasound and MRI are comparable in both sensitivity and specificity.
all languages and involving human and animal subjects.

**Study Selection**

Our query of the MEDLINE database returned 1,195 hits. The articles were analyzed for concordance with the inclusion criteria. These criteria are English language; absolute (raw) data on rotator cuff tears (full or partial thickness or both) in the form of true-positives (TPs), true-negatives (TNs), false-positives (FPs), and false-negatives (FNs) either provided or extractable; surgical reference standard (arthroscopy or open surgery); and diagnostic imaging studies interpreted by radiologists. In addition, data must not have been published in a prior study. To prevent this possibility, we included only the article with the earlier publication date if two articles with common authors or from the same institution had the earlier publication date if two articles with this possibility, we included only the article with open surgery); and diagnostic imaging studies negatives (FNs) either provided or extractable; the inclusion criteria. These ed 1,195 hits. The articles were analyzed for concordance with the inclusion criteria. These judgments are detailed here.

Non-English-language (n = 160) and animal (n = 1) studies were excluded. The abstracts of the remaining studies were evaluated for relevance to our study. Of these, 270 relevant articles were retrieved. One hundred sixty-two of the 270 were excluded because either raw data were not provided or the data could not be extracted into discrete TPs, TNs, FPs, and FNs. Of the remaining 108 articles with data, 43 were excluded for the following reasons: 18 studies had a nonsurgical standard of reference, 15 had ultrasound read by nonradiologists, one had MRI read by non-radiologists, and nine had overlapping dates of subject inclusion with other studies by common authors from the same institution. Therefore, 65 of 270 (24.1%) of the English-language articles met the inclusion criteria [7–71]. Twenty-five studies analyzed MRI only; five, MRI and MR arthrography; nine, MR arthrography only; five, MRI and ultrasound; one, MR arthrography and ultrasound; and 20, ultrasound only. The most recently published article was in July 2007. The oldest was published in May 1985.

From the 65 articles that fulfilled the inclusion criteria, we retrieved a total of 140 data sets: 48 ultrasound, 67 MRI, and 25 MR arthrography. The breakdown of the data sets by diagnostic end point was as follows: 56 evaluated for the presence of a rotator cuff tear (including full or partial thickness) versus no tear; 49 evaluated full-thickness tear versus non-full-thickness tear (including partial thickness tear or no tear); and 35 evaluated partial-thickness tear versus non-partial-thickness tear (including full-thickness tear or no tear). The heterogeneous nature of these studies required judgments regarding which data to include and how to avoid including duplicate data. These judgments are detailed here.

In one article [16] that assessed partial-thickness tears, three full-thickness tears were identified as partial-thickness tears; however, because a tear was identified, we considered those studies to be TPs.

Another article [21] evaluated the sensitivity and specificity of six MRI findings in detecting a full-thickness rotator cuff tear; TPs, TNs, FPs, and FNs were recorded for each finding. Because we were influenced by the co-authored with techniques and not with specific findings, we picked one finding to represent MRI. For each finding, we calculated the [(sensitivity + specificity) / 2] and picked the finding with the highest value. This finding turned out to be thinning of the tendinous cuff.

In one article [22], some MRI diagnoses by report were equivocal (e.g., partial- versus full-thickness tear). The authors of this study used the more severe diagnosis in the final tally of no disease versus disease with the rationale that a more significant diagnosis is more likely to affect management.

One article [27] compared the accuracy of two MR pulse sequences. Because we were not interested in individual sequences but more in the actual technique, we omitted one of the sequences. The sequence we omitted was trivial because the authors of that article found no diagnostically significant difference between the two sequences, and the reported sensitivities and specificities were identical.

Similarly, another article [28] looked at the accuracy of T2-weighted sequences with and without fat saturation in diagnosing full- and partial-thickness rotator cuff tears. For each of the four sets of data, we took [(sensitivity + specificity) / 2] and omitted the sequence with the lowest value (i.e., the fat-saturated fast spin-echo sequence).

In one article [30], the authors divided patients into two groups on the basis of who performed ultrasound: Group 1 patients underwent ultrasound performed by a sonographer with 5 years of experience and group 2, by a radiologist with 10 years of experience. We omitted the data from the group examined by the sonographer.

Another study [31] examined the interobserver agreement of five readers, each of whom interpreted MR images twice, first as a blinded review and second with knowledge of the surgical outcome. Sensitivities and specificities were calculated for full- and partial-thickness tears for each reader, generating five sets of data. For our purposes, we averaged the TPs, TNs, FPs, and FNs for full- and partial-thickness tears for both the blinded and retrospective readings.

In one study [32], investigators also reported TPs, TNs, FPs, and FNs for two readers and in another study [33], for four readers. In a similar fashion, we took the average TP, average TN, average FP, and average FN and calculated sensitivities and specificities.

One article [36] compared MR arthrography performed on a low-field magnet (0.2-T) and on a high-field magnet (1.5-T); two sets of data were reported. The reported sensitivities and specificities were identical, so we dropped one of the data sets.

Similar to other studies, one study [38] had two sets of data for two independent readers. We averaged the TP, TN, FP, and FN values for those readers.

In one article [45], the diagnosis of a rotator cuff tear was established on the basis of findings from transverse, parasagittal, or both transverse and parasagittal MR images of the shoulder. Two radiologists independently evaluated the planar images for each shoulder. We used the reported data for diagnosis based on both the transverse and parasagittal images. As in prior studies, we averaged the data for the two readers.

Another article [49] compared the results of MR arthrography performed using three different solutions for intraarticular injection: Ringer solution and two different concentrations of gadoteridol. All three solutions were reported as equivalent in diagnostic accuracy. Furthermore, two independent readers evaluated all the studies, and interobserver agreement was calculated as a kappa value for each of the three solutions. We took the data set for the contrast solution with the highest kappa value and averaged the data for the two readers.

In studies with multiple observers and multiple data sets [31–33, 38, 49], we took averages of the contingency data (TPs, TNs, FPs, FNs) to calculate sensitivities and specificities, as described earlier.

In one article [52] that looked at ultrasound of the rotator cuff, the authors established the diagnosis of a tear using published diagnostic criteria and again using a subset of the published criteria. We used the data set that yielded the higher accuracy—namely, the one generated from the more restricted subset of the published criteria.

In references 45–51, multiple reference standards were used in each study. For each article, we used only the data sets with surgically proven findings— that is, either open surgery or arthroscopy.

Meister et al. [61] reported nine full-thickness tears, 28 partial-thickness tears, and 39 intact tendons by arthroscopy. Using MR arthrography, the authors recorded nine full-thickness tears that were “...classified as true-negatives in this analysis of partial-thickness tears” [61]. In the calculation
of specificity of MR arthrography for the diagnosis of partial-thickness tears, a TN is not a partial tear—that is, composed of full-thickness tears and intact tendons. However, the authors reported a specificity of 96% (43/45), which appears instead to correspond to the negative predictive value. Specificity should actually be 90% (43/48) if the nine full-thickness tears are counted as TNS.

In their study, Milosavljevic et al. [64] noted that in the calculation of the sensitivity of ultrasound for partial-thickness tears, “…in seven additional shoulders, ultrasound identified a full-thickness tear instead of a partial-thickness tear. Because a tear was identified, these studies were considered to be true-positive.” For the purpose of calculating sensitivity for partial-thickness tears in this meta-analysis, we counted only the 17 partial-thickness tears as TPs.

Herold et al. [65] recorded data for two readers. These data were averaged in the aforementioned fashion. Additionally the authors noted that each patient underwent MR arthrography twice: once with the patient’s shoulder in a neutral position and a second time with it abducted and externally rotated (ABER). Sensitivity, specificity, and accuracy were compared. As done previously, we calculated [(sensitivity + specificity) / 2] and took the data set with the higher values, which turned out to be the ABER data set.

Fritz et al. [67] evaluated the association of cystic changes at tendon insertion sites with rotator cuff disorders. Overall performance for tears was reported. Data for full- or partial-thickness tears could not be gleaned.

Ferrari et al. [71] classified full-thickness supraspinatus tears as focal, subtotal, or total and classified partial-thickness tears as intratendinous, articular, or bursal-sided. For the meta-analysis, these subcategories of full- and partial-thickness tears were not considered separately.

**Statistical Analysis**

Our central interest was to compare the diagnostic accuracies of MRI, MR arthrography, and ultrasound in the evaluation of rotator cuff tears using a surgical reference standard. Because investigators from different studies reported using different criteria for the diagnosis of rotator cuff tears (full-thickness, partial-thickness, or both), we compared techniques within and across criteria.

Two common approaches appear in the meta-analysis literature for the analysis of diagnostic tests. One approach is to pool data from a number of studies to obtain overall sensitivities and specificities and compare them using the chi-square test. A second approach involves using regression to construct summary ROC curves for each technique and then computing a z test to compare the Q* points of the curve—that is, the points on an ROC curve where sensitivity equals specificity [72]. We analyzed the data for this meta-analysis in both ways.

It should be noted that an article with a small sample size has a small impact on the summary analysis because it is aggregated into a pool of data. However, such a study may appear as an outlier on the scatterplots.

Additionally, in some instances, a value for sensitivity or specificity could not be calculated from the contingency data. For instance, Toyoda et al. [63] report a data set with 41 shoulders and a sensitivity of 100% for 41 TPs and 0 FNs; specificity could not be calculated because there were 0 TNS and 0 FPs. These data sets were not included in either the distribution plots or the ROC curves because an actual value, 0 or a positive value, is needed to generate a point on the plot or curve.

**Results**

Figure 1 shows distribution plots for full-thickness tears, full- and partial-thickness tears, and partial-thickness tears. These scatterplots show sensitivity and (1 – specificity) on the y- and x-axes, respectively. Each point on the diagrams represents a published study. The meta-analysis of an ideal imaging technique—that is, one that is highly sensitive with a low FP rate—would be depicted with a cluster of data points in the upper left area of the plot.

The data points for MR arthrography in all three plots are tightly clustered in the upper left, although there are only 25 points: eight for full-thickness tears, 11 for full-thickness and partial tears, and six for partial-thickness tears.

There is considerable overlap in the distribution of data points for MRI and ultrasound. For full-thickness tears (Fig. 1A), MRI and ultrasound have a grouping of data points in the upper left. However, for full- and partial-thickness tears (Fig. 1B), MRI and ultrasound have data points somewhat loosely dispersed across the upper half of the graph, corresponding to high sensitivity and varying FP rates. Both have prominent outliers. For partial-thickness tears (Fig. 1C), MRI and ultrasound have a low FP rate but a wide variation in sensitivities, with data points dispersed loosely in the left one third of the plot. Again, there are outliers for both.

Table 1 illustrates the pooled sensitivities and specificities of each imaging technique with respect to full-thickness tears; partial-thickness tears; full- or partial-thickness tears; and total tears, which is the sum of the first three groups.

For full-thickness tears, chi-square analysis shows no significant difference in sensitivity among the three techniques. Furthermore, ultrasound and MRI are not significantly different in sensitivity or specificity. MR arthrography is more specific than either MRI or ultrasound (specificity vs MRI, χ² = 40.142, p < 0.0001; specificity vs ultrasound, χ² = 25.836, p < 0.0001).

For partial-thickness tears, MR arthrography is more sensitive and more specific than either MRI or ultrasound (sensitivity and specificity of MR arthrography vs MRI: χ² = 27.358 and 14.134, p < 0.0001 and 0.0002, respectively; vs ultrasound: χ² = 21.635 and 5.111, p < 0.0001 and 0.02). Although there is no statistically significant difference between MRI and ultrasound for the diagnosis of partial-thickness tears, ultrasound tends to be more sensitive and more specific (sensitivity: χ² = 2.057, p = 0.15; specificity: χ² = 3.347, p = 0.067).

For the diagnosis of full- or partial-thickness tears, MR arthrography is more sensitive and more specific than either MRI or ultrasound (sensitivity and specificity of MR arthrography vs MRI: χ² = 8.130 and 54.990, p < 0.0004 and 0.0001, respectively; vs ultrasound: χ² = 14.843 and 28.073, p < 0.0001 and 0.0001). However, there is no statistically significant difference between MRI and ultrasound (p > 0.05).

Figure 2 shows the summary ROC curves for MR arthrography, MRI, and ultrasound for all tears. We followed the method described by Langlotz and Sonnad [73], which consists of fitting a regression line to the sum and difference of the logit transforms of the sensitivities and specificities. MR arthrography has the greatest area under the ROC curve (0.935), followed by ultrasound (0.889) and then MRI (0.878). The curves for ultrasound and MRI cross at the location where sensitivity is approximately 88% and 1 – specificity is 15%, after which MRI is shown to be superior to ultrasound.

Q* is the point on an ROC curve where sensitivity and specificity are equal and has been proposed as a more relevant measure of test efficacy than the entire ROC curve. After Moses et al. [72], we calculated Q* and its SD for each technique: For MR arthrography, these values were 0.92 and 0.057, respectively; MRI, 0.86 and 0.032; and ultrasound, 0.86 and 0.049. The results of z tests show that pairwise comparisons are not significantly different (p > 0.05).
For the construction of the distribution plots, studies in which TP, TN, FP, FN, or a combination of these values equal 0 are not included because 0 values cannot be plotted. Likewise, for the composite ROC curve, these studies are excluded. For this reason, two of the MR arthrography studies [24, 63] are not plotted in the distribution plots and ROC curve. Therefore, MR arthrography is shown to be significantly better than MRI or ultrasound in the chi-square values but not the ROC curves.

Discussion

For the patient with shoulder pain, a host of therapeutic options, ranging from medical management to physiotherapy to open surgery, are available [6, 74]. The role of diagnostic imaging is to help guide surgical or nonsurgical management. The ideal imaging technique should have a high rate of TP’s and an acceptable rate of FP’s to limit unnecessary surgical intervention.

In the 65 studies included in our analysis, most of the reported sensitivities and specificities fall in the range of 60–100% for ultrasound. A similarly wide variation is observed for MRI and MR arthrography. There are a few possible reasons for such a variation in numbers including small sample sizes per study, differing study designs, varying quality of imaging equipment (e.g., a wide range of ultrasound probe frequencies and MRI field strengths), and differing imaging criteria for diagnosis. We sought to draw on the strengths of a meta-analysis to evaluate the large body of literature to overcome the small sample sizes and heterogeneous designs of individual trials by pooling the data and obtaining summary sensitivities and specificities.

Our findings show that MR arthrography is more sensitive and more specific than either ultrasound or MRI in diagnosing both full- and partial-thickness rotator cuff tears. Additionally, there is no statistically significant difference between the sensitivities and specificities of MRI versus ultrasound in diagnosing either full- or partial-thickness tears.

To our knowledge, only one other meta-analysis of rotator cuff tears has been published [6]. In that systematic review published in 2003, Dinnes et al. [6] evaluated the diagnostic effectiveness of MRI, MR arthrography, ultrasound, and clinical examination in the evaluation of a painful shoulder, with rotator cuff tears as the disease end point. They concluded that either MRI or ultrasound could be used for equal detection of full-thickness rotator cuff tears but that ultrasound is the more cost-effective test. They also stated that MR arthrography appeared to perform better than either MRI or ultrasound but that “any such benefit must be set against the invasiveness and potential discomfort to patients of the procedure” [6].

Our inclusion criteria differed from those used for the systematic review by Dinnes et al. [6] because we used data sets with only surgically proven findings. In the study by Dinnes and colleagues, a large number of studies used nonsurgical techniques, such as arthrography, as the reference standard for disease. In one article, investigators had even used MRI as the reference standard for disease [75]. Also, Dinnes et al. included studies.
MRI, MR Arthrography, and Ultrasound of Rotator Cuff Tears

TABLE 1: Summary Sensitivities and Specificities for MR Arthrography, MRI, and Ultrasound

<table>
<thead>
<tr>
<th>Technique</th>
<th>No. of Positive Cases</th>
<th>Sensitivity (%)</th>
<th>CI (%)</th>
<th>No. of Negative Cases</th>
<th>Specificity (%)</th>
<th>CI (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Full-thickness tear</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MR arthrography</td>
<td>227</td>
<td>95.4</td>
<td>2.7</td>
<td>879</td>
<td>98.9</td>
<td>0.7</td>
</tr>
<tr>
<td>MRI</td>
<td>625</td>
<td>92.1</td>
<td>2.1</td>
<td>1,085</td>
<td>92.9</td>
<td>1.5</td>
</tr>
<tr>
<td>Ultrasound</td>
<td>639</td>
<td>92.3</td>
<td>2.1</td>
<td>674</td>
<td>94.4</td>
<td>1.7</td>
</tr>
<tr>
<td><strong>Partial-thickness tear</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MR arthrography</td>
<td>195</td>
<td>85.9</td>
<td>4.9</td>
<td>875</td>
<td>96.0</td>
<td>1.3</td>
</tr>
<tr>
<td>MRI</td>
<td>236</td>
<td>63.6</td>
<td>6.2</td>
<td>916</td>
<td>91.7</td>
<td>1.8</td>
</tr>
<tr>
<td>Ultrasound</td>
<td>249</td>
<td>66.7</td>
<td>5.9</td>
<td>790</td>
<td>93.5</td>
<td>1.7</td>
</tr>
<tr>
<td><strong>Full- or partial-thickness tear</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MR arthrography</td>
<td>485</td>
<td>92.3</td>
<td>2.4</td>
<td>869</td>
<td>94.5</td>
<td>1.5</td>
</tr>
<tr>
<td>MRI</td>
<td>667</td>
<td>87.0</td>
<td>2.6</td>
<td>463</td>
<td>81.7</td>
<td>3.5</td>
</tr>
<tr>
<td>Ultrasound</td>
<td>915</td>
<td>85.1</td>
<td>2.3</td>
<td>481</td>
<td>86.1</td>
<td>3.1</td>
</tr>
<tr>
<td><strong>All tears</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MR arthrography</td>
<td>907</td>
<td>91.7</td>
<td>1.8</td>
<td>2,623</td>
<td>96.5</td>
<td>0.7</td>
</tr>
<tr>
<td>MRI</td>
<td>1,528</td>
<td>85.5</td>
<td>1.8</td>
<td>2,464</td>
<td>90.4</td>
<td>1.2</td>
</tr>
<tr>
<td>Ultrasound</td>
<td>1,803</td>
<td>85.1</td>
<td>1.6</td>
<td>1,945</td>
<td>92.0</td>
<td>1.2</td>
</tr>
</tbody>
</table>

Note—The results are arranged by the type of tear reported on in the individual study. CI = confidence interval.

*Sum of data listed under full-thickness tear, partial-thickness tear, and full- or partial-thickness tear.

Fig. 2—Overall summary receiver operating characteristic curves for detection of rotator cuff tears using MR arthrography (solid line), MRI (dotted line), and ultrasound (dashed line).

in which ultrasound had been performed by nonradiologists, usually orthopedic surgeons or rheumatologists. We included only those studies in which radiologists performed and interpreted the imaging studies. Furthermore, we directly compared MR arthrography, MRI, and ultrasound using ROC curves and single-statistic summaries (Q*). Despite all of these differences, our conclusions are similar to those of Dinnes and colleagues.

Our study has limitations. First, although the criteria for diagnosing a full- or partial-thickness tear of the rotator cuff have been published, the criteria used to make the diagnosis in each study varied. This variability reflects the wide range in time over which the included studies were published and is a manifestation of the maturation process of each technique for which criteria for diagnosis must be refined. The selection bias and workup bias inherent to each individual study also may play an important role in our pooled study. The design of most studies was retrospective to identify patients who underwent the index test. Of this set of patients, a subset who subsequently underwent surgery was identified as the sample population. Because of workup bias, most patients who underwent the index test and subsequent surgery had a high suspicion and probability of, indeed, having a tear; effectively, those studies assessed the accuracy of the index test in detecting a tear in a patient with a high pre-test clinical suspicion of a rotator cuff tear.

Additional forms of bias encountered by Dinnes et al. [6] were present in our study. In partial verification bias, only patients who underwent the reference test were included in a sample, with the results of the remaining patients who underwent the index test (MRI, ultrasound, MR arthrography) unreported.

Publication bias is a particular drawback of meta-analyses. Studies with favorable results have a higher likelihood of being published, creating an inherent selection bias during a literature review [6, 76, 77].

In summary, MR arthrography is more accurate than MRI and ultrasound in diagnosing rotator cuff tears. Ultrasound is as accurate as MRI for both full-thickness tears and partial-thickness tears. These results, combined with the lower cost for ultrasound, suggest that ultrasound may be the most cost-effective imaging method for screening for rotator cuff tears provided that the examiner has been properly trained in this operator-dependent technique. For practitioners without ultrasound expertise, MRI can be used. MR arthrography can be performed in cases in which ultrasound and MRI are not definitive.

References

de Jesus et al.

173:78–91
29. Loew R, Kretiner RF, Runkel M, Zoelner J, Thelen M. MR arthrography of the shoulder: comparison of low-field (0.2 T) vs. high-field (1.5 T) imaging. *Eur Radiol* 2000; 10:989–996
43. Torstensson ET, Hollinshead RM. Comparison of magnetic resonance imaging and arthroscopy in the evaluation of shoulder pathology. *J Shoulder Elbow Surg* 1999; 8:42–45
46. Teeffy SA, Rubin DA, Middleton WD, Hildebolt CF, Leibold RA, Yamaguchi K. Detection and
46. Singson RD, Hoang T, Dan S, Friedman M. MR evaluation of rotator cuff pathology using T2-weighted fast spin-echo technique with and without fat suppression. AJR 1996; 166:1061–1065


