To review the diagnostic performance of ultrasonography (US) and magnetic resonance imaging (MRI) for the detection of rotator cuff tears, we performed a retrospective audit of patients who underwent shoulder arthroscopy at Hinchingbrooke hospital. The diagnostic accuracies of US for full and partial thickness tears were 82% and 28% respectively. Those of MRI were 82% and 81% respectively. These were lower than expected from the literature. This discrepancy is likely to be the consequence of over-diagnosis in imaging and under-diagnosis at arthroscopy.

Keywords: ultrasonography; magnetic resonance imaging; rotator cuff tear; meta-analysis; accuracy.

INTRODUCTION

Rotator cuff tears are a common cause of shoulder pain and disability and surgery is important in their management (11,12,23). Imaging has a crucial role in surgical planning, and ultrasound (US) and magnetic resonance imaging (MRI) are now both established as accurate investigative techniques for pre-operative diagnosis (7).

Early reports evaluating the use of ultrasound for identifying tears were controversial (2,24). However recent technological and methodological advances have refined the technique, and contemporary reports of its performance have been more consistently favourable (17,20,30,39). In contrast, there has been more agreement about the high accuracy of MRI ever since its introduction in the 80s (9,13-15,19, 26,29,32). Some commentators support the use of MRI as the first line investigation for all musculoskeletal conditions and have emphasized the relative disadvantages of US in the assessment of shoulder disease, such as its operator dependence, the necessity to place the shoulder in potentially painful positions, and the limitations in the structures that can be visualised (34). Other reports highlight the many advantages of US (28,32). For example: it enables dynamic imaging that allows features such as subacromial impingement to be directly observed; it permits direct correlation between visualized pathology and the site of pain; and it typically involves a much shorter waiting list for patients and is far cheaper to perform.

In the context of the UK National Health System, the low cost and short waiting lists associated with US give it considerable appeal (38). Furthermore, the literature supports the conclusion that choice of US over MRI theoretically represents no fall in standards for the diagnosis of rotator cuff pathology (7,8). Thus, it seems appropriate to suggest that US should be used as the initial investigation in cases where the patient’s history and clinical findings are suggestive of appropriate soft tissue pathology. A minority of patients might subsequently require an MRI scan (34), but the majority should receive sufficient information to guide management of their condition without the need for further investigation.

With this study, we aimed to determine whether utility of the two imaging modalities in our own specific clinical context (a district general hospital, where US is currently first-line for the investigation of the rotator cuff) achieved the standards of accuracy that we expected from the literature.

**MATERIALS AND METHODS**

The retrospective study protocol was approved by the hospital’s board. The study period was 19 months from 01/06/2009 to 31/12/2010 and the population of 283 patients who underwent arthroscopy in that period was identified via the clinical coding records. Inclusion criteria were: a) clinical examination findings in keeping with rotator cuff pathology, b) shoulder arthroscopy performed in the DGH, and c) US or MRI, but not both, performed preoperatively in the DGH. The electronic Picture Archiving and Communications Systems (PACS) was used to identify patients who had undergone only one of the two imaging procedures, and the condition of the supraspinatus tendon specified in the formal report was recorded.

All shoulder MRI scans were performed on a 1.5 T Unit (Siemens Magnetom Symphony) using the same protocol: sagittal T2-weighted TSE with fat suppression (TR = 3500/TE = 78); axial T2-weighted med2d (TR = 941/TE = 27); coronal T1-weighted SE (TR = 450/TE = 13) and coronal T1-weighted TIRM (TR = 4000/TE = 29/IR = 150). The scans were reported by two experienced Consultant Radiologists with more than 5 years of experience each. One of them, who reported 68% of the scans (38/56), was a specialist in musculoskeletal radiology. All US scans were performed on the same equipment (Toshiba Aplio XG) by the same experienced Consultant Musculoskeletal Radiologist who reported the majority of the MRIs. Starting from the end of the study period the notes record of each patient was reviewed in sequence, according to the operation date, until a total of 56 patients had been included in each arm of the study; patients with a history of previous shoulder operation were excluded. Information about the condition of the supraspinatus tendon identified at arthroscopy was retrieved from the operation notes. All patients had arthroscopy under the care of the same Consultant Orthopedic Surgeon. The average time between imaging (US or MRI) and surgery was 5 months.

The data regarding the presence and type of tears of the supraspinatus tendon identified via imaging and arthroscopy were cross-tabulated (Tables I and II) and classified according to the correspondence between the findings in the imaging reports and those during arthroscopy. Four partial-thickness tears described as intra-substance tears on imaging were excluded from the accuracy analyses on the understanding that such tears cannot be identified via arthroscopy by visual inspection alone (22). Three cases in which partial-thickness tears were identified in imaging but full-thickness tears were identified at arthroscopy were classified as false negatives in the full-tear and all-tears analyses. All cases (n = 24) in which there was no comment about a tear in the scan report or operation note were categorized as intact tendons.

Microsoft Excel was used for data entry and for calculation of the measures of diagnostic accuracy (sensitivity, specificity, negative and positive predictive values, and overall accuracy). Data were also retrieved from the figures of the latest published meta-analysis (7) and statistical comparisons between results were performed by calculating the 2 proportion z-score in excel using the formula specified below and then retrieving the corresponding two-tailed p-value from an online z-score calculator.

\[
z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\hat{p}(1 - \hat{p}) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}
\]

In this equation \(\hat{p}_1\) and \(\hat{p}_2\) signify each individual proportion to be compared, \(n_1\) and \(n_2\) represents the number in each group respectively and \(\hat{p}\) represents the overall combined proportion.

**RESULTS**

According to the arthroscopy results, 34% (19/56) of patients who underwent a pre-operative MRI had
a supraspinatus tendon tear. Sixteen percent of these (3/19) were specified as partial-thickness tears (Table I). Similar results (p > 0.2) were seen for the group of patients who had US before surgery, in whom 25% (14/56) had a tear and 7% of these (1/14) were partial-thickness (Table II).

In the MRI group, 54% (30/56) of patients were diagnosed with tears through imaging and 40% of them (12/30) were said to be partial-thickness (Table II). MRI agreed with arthroscopy in approximately 80% of cases for both partial and full-thickness tears (Table III). This overall accuracy is equivalent to the meta-analysis result for partial-thickness tears (p > 0.4), despite the fact that the current study demonstrates lower specificity (p < 0.032). In contrast, the reduced specificity observed for full-thickness tears (p < 0.002) was large enough to generate a lower accuracy overall (p < 0.002).

Ultrasound diagnosed tears in almost all patients in the group (86% ; 48/56) and 71% of these (34/48) were identified as partial-thickness (Table II ; fig. 3 and 4). Agreement with arthroscopy was 75% for full-thickness tears and only 24% for partial-thickness tears (Table III). Both of these accuracy values were statistically lower than the equivalent results reported in the meta-analysis. With a difference of over 60%, the partial-thickness tear result was very significantly different (z-score = 10.5 ; p-value = 0).

In accordance with the above results, directly comparing the performance of US and MRI in this study revealed that partial-thickness tears were more accurately identified on MRI (p < 0.0001) and that this difference was due to a difference in

---

Table I. — Identification of supraspinatus tears via pre-operative MRI

<table>
<thead>
<tr>
<th></th>
<th>ARTHROSCOPY</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Articular Tear</td>
</tr>
<tr>
<td>MRI</td>
<td></td>
</tr>
<tr>
<td>Articular Tear</td>
<td>1</td>
</tr>
<tr>
<td>Bursal Tear</td>
<td>0</td>
</tr>
<tr>
<td>Intrasubstance Tear</td>
<td>0</td>
</tr>
<tr>
<td>Partial Tear unspecified</td>
<td>0</td>
</tr>
<tr>
<td>Full Thickness Tear</td>
<td>0</td>
</tr>
<tr>
<td>Intact</td>
<td>0</td>
</tr>
<tr>
<td>No comment</td>
<td>0</td>
</tr>
<tr>
<td>TOTAL</td>
<td>1</td>
</tr>
</tbody>
</table>

Table II. — Identification of supraspinatus tears via pre-operative US

<table>
<thead>
<tr>
<th></th>
<th>ARTHROSCOPY</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Articular Tear</td>
</tr>
<tr>
<td>ULTRASOUND</td>
<td></td>
</tr>
<tr>
<td>Articular Tear</td>
<td>0</td>
</tr>
<tr>
<td>Bursal Tear</td>
<td>0</td>
</tr>
<tr>
<td>Intrasubstance Tear</td>
<td>0</td>
</tr>
<tr>
<td>Partial Tear unspecified</td>
<td>0</td>
</tr>
<tr>
<td>Full Thickness Tear</td>
<td>0</td>
</tr>
<tr>
<td>Intact</td>
<td>0</td>
</tr>
<tr>
<td>No comment</td>
<td>0</td>
</tr>
<tr>
<td>TOTAL</td>
<td>0</td>
</tr>
</tbody>
</table>
Acta Orthopædica Belgica, Vol. 80 - 3 - 2014

EVALUATING THE SUCCESS OF PREOPERATIVE IMAGING FOR DIAGNOSING ROTATOR CUFF TEARS

Fig. 1. — Coronal MRI image in an 85 year-old male patient. STIR sequence (TR = 4000/TE = 29/IR = 150) that shows high signal intensity within the bursal surface of the supraspinatus tendon (arrows) representing a partial-thickness tear.

Fig. 2. — Sixty nine year-old male patient with a history of fall onto the shoulder. Coronal STIR MR image that demonstrates a full-thickness tear of the supraspinatus tendon (short arrow) shown as fluid signal intensity at the level of its insertion to the greater tuberosity. There is a fracture line through the greater tuberosity noticed as well (long arrow).

evaluating the success of preoperative imaging for diagnosing rotator cuff tears

we observed that:

US and MRI exhibited equal accuracy for the diagnosis of full-thickness tears; partial-thickness tears were more accurately diagnosed via MRI; both modalities exhibited lower overall accuracies in the diagnosis of full-thickness tears than expected from the latest meta-analysis in the literature; and US was markedly less accurate than literature standards in the diagnosis of partial-thickness tears. Both US and MRI were excellent at ruling out, and moderately good at ruling in, full-thickness tears. Both modalities also performed well at ruling out partial-thickness tears, but US performed extremely poorly at ruling them in, generating a high proportion of type I (false positive) errors and positive predictive value of only 3%. These results are reminiscent of some of the studies from the 1980s that criticized the performance of US and appear to lead to a conclusion that US should not be used for the diagnosis of rotator cuff tears. However, such a conclusion reflects an

specificity. There was no difference between the imaging modalities for the identification of full-thickness tears (p > 0.6; table III).

Calculations of positive and negative predictive values indicated that, in the population represented by our samples, MRI diagnosis of a tear denoted a 40-61% probability that the tear was subsequently seen in arthroscopy, with full-thickness tears contributing the upper limit (Table IV). In contrast, a positive US scan for partial-thickness tears signified only a 3 in 100 chance of showing a tear at surgery, whilst the positive predictive value for a full-thickness tear was 71% (Table IV). A negative result in either imaging modality for any type of tear demonstrated a good probability of no tear being found at arthroscopy, with negative predictive values of at least 80% (Table IV).

DISCUSSION

In this retrospective review of operation notes and image reports for a selection of patients who underwent shoulder arthroscopy within a specific 18-month period at a single district general hospital we observed that: US and MRI exhibited equal accuracy for the diagnosis of full-thickness tears; partial-thickness tears were more accurately diagnosed via MRI; both modalities exhibited lower overall accuracies in the diagnosis of full-thickness tears than expected from the latest meta-analysis in the literature; and US was markedly less accurate than literature standards in the diagnosis of partial-thickness tears. Both US and MRI were excellent at ruling out, and moderately good at ruling in, full-thickness tears. Both modalities also performed well at ruling out partial-thickness tears, but US performed extremely poorly at ruling them in, generating a high proportion of type I (false positive) errors and positive predictive value of only 3%. These results are reminiscent of some of the studies from the 1980s that criticized the performance of US and appear to lead to a conclusion that US should not be used for the diagnosis of rotator cuff tears. However, such a conclusion reflects an
depending on the specific pathology (25,31). Furthermore, in the assessment of partial-thickness tears, a surgeon reviewing the arthroscopic images over video agreed with the findings of the scope operator on only 55% of cases (25). Other work has highlighted the weakness of arthroscopy for quantifying specific characteristics of rotator cuff tears and, moreover, reported a type I error (false positive) rate of 30% (compared to open surgery) for merely determining the presence of a tear in the supraspinatus tendon (35). Moreover, as indicated above, comprehensive evaluation of the rotator cuff via arthroscopy is undermined by the challenge of identifying intratendinous tears (22). Accordingly, the blind acceptance of arthroscopy as the gold standard has been questioned (31).

The ‘gold standard’ has an important role as the best available representative of reality in the quantitative analysis of diagnostic tests. Nevertheless, there are strong grounds to advocate a more nuanced approach to interpreting the results of such quantitative analyses; i.e. an approach that acknowledges the separation that actually exists between reality and the gold standard as practiced within a particular context (16).

uncritical interpretation of the data provided here which fails to recognize the complexities inherent in the concept and function of the ‘gold standard’, and the challenges of relating results from retrospective audits of medical outcomes to the results of prospective studies in the literature.

**Shoulder arthroscopy as a ‘gold standard’**

Arthroscopy is not a perfect technique for diagnosing rotator cuff tears. The agreement between different orthopedic surgeons in the diagnosis of shoulder pathology through arthroscopy ranges from 100% (perfect agreement) to < 60% of cases, depending on the specific pathology (25,31). Furthermore, in the assessment of partial-thickness tears, a surgeon reviewing the arthroscopic images over video agreed with the findings of the scope operator on only 55% of cases (25). Other work has highlighted the weakness of arthroscopy for quantifying specific characteristics of rotator cuff tears and, moreover, reported a type I error (false positive) rate of 30% (compared to open surgery) for merely determining the presence of a tear in the supraspinatus tendon (35). Moreover, as indicated above, comprehensive evaluation of the rotator cuff via arthroscopy is undermined by the challenge of identifying intratendinous tears (22). Accordingly, the blind acceptance of arthroscopy as the gold standard has been questioned (31).

The ‘gold standard’ has an important role as the best available representative of reality in the quantitative analysis of diagnostic tests. Nevertheless, there are strong grounds to advocate a more nuanced approach to interpreting the results of such quantitative analyses; i.e. an approach that acknowledges the separation that actually exists between reality and the gold standard as practiced within a particular context (16).

![Table III. — Summary of sensitivity, specificity and accuracy for data in current study and data from meta-analysis (‘literature’) (7). Lines and asterisks indicate comparisons producing significant differences](image-url)

![Table IV. — Positive (PPV) and Negative predictive values (NPV) for each imaging modality and each type of tear](image-url)

**Acta Orthopaedica Belgica, Vol. 80 - 3 - 2014**
Evaluating the Success of Preoperative Imaging for Diagnosing Rotator Cuff Tears

Clinical studies vs clinical practice

The strength of evidence obtainable from retrospective studies is weakened by a dependence on reported information, and the impossibility of applying controls against bias and confounding factors (5). Prospective blinded studies provide more robust evidence, but have been criticized for lacking clinical relevance because information is not normally withheld in clinical decision-making (1,21). Prospective studies of clinical practice may also suffer from the possibility that observation, for the purpose of evaluating clinical behaviour, influences the behaviour that is to be evaluated. We can speculate that having knowledge that the results of one’s activities are being included in a study might lead one to perform those activities in a different way.

Accordingly, the outcomes from such studies may not be readily transferable to the context of normal, un-scrutinized practice. So, for example, a surgeon performing arthroscopies in the context of a study might expend more time and effort in conducting a comprehensive evaluation of the anatomy than an individual engaged in his or her routine weekly, unobserved arthroscopy lists.

It must be acknowledged that some retrospective studies in the literature report excellent results for...
the accuracy of both MRI and US in identifying both full and partial-thickness rotator cuff tears (27). However, the possibility remains that accuracies derived from comprehensive reviews of the literature may be distorted by the artificial influence of observation on any of the prospective studies that were included.

This argument may appear to be undermined by the fact that the MRI results obtained in this study were largely in agreement with the results from the literature. Admittedly, the deliberations described above were primarily stimulated by the objective of interpreting the disparity between the US results and the meta-analysis data. However, the issues underlying the proposals put forward remain valid no matter how accurate and similar to published data retrospective observational results might be. Consequently, this discussion highlights the following generalizable principles that should be considered whenever audit data of diagnostic efficacy are assessed: a) to what extent does the performance of the gold standard depart from perfect accuracy; and b) are expected standards retrieved from the literature directly applicable to the particular audit context?

It is not unreasonable to suggest that the results observed in our centre might be representative of any number of similarly sized and similarly situated hospitals across the United Kingdom and perhaps even further afield. Thus, the considerations presented here have a general relevance for the issue of relating performance in such small centers to the wider medical community, particularly if studies contributing to literature standards have been produced in larger institutions.

The patient sample sizes in the study were chosen following a review of similar work in the literature and, with 56 patients in each group, the number analyzed compares favorably with numerous studies in the field (13,10,19,33,36,37). However, there was an extremely low prevalence of partial-thickness tears identified by arthroscopy in this context, and this low number limits the generalizability of some of the results observed.

As with all retrospective studies, the comparison between MRI and US is also limited by the risk of confounding variables. Patients had not been randomly allocated to each imaging modality but had been directed in accordance with normal clinical decision-making. Thus, a large proportion of the patients undergoing MRI either also needed an assessment of the joint space and peri-articular structures or had a very restricted range of movement that precluded use of US. This design has the benefit of reflecting genuine clinical practice, but is vulnerable to referral or allocation bias that might make the group of patients that underwent US fundamentally different to those who had MRI. Some equivalence between the groups can be inferred from the fact that the prevalence of each type of tear as identified by arthroscopy was statistically indistinguishable (p > 0.38; table I and II), however it is impossible to know whether some other unknown confounding factor contributed to the differences in the diagnostic accuracies seen.

**CONCLUSION**

This study evaluates the accuracy of MRI and US in the diagnosis of rotator cuff tears in a single DGH and discusses the issues involved in relating such results to the literature at large. The data show a marked divergence from the results reported in a recent meta-analysis, particularly with regards to the accuracy of the detection of partial-thickness tears by US. Consideration of the challenges that an imperfect gold-standard raises for interpretation of the term ‘accuracy’ lead us to conclude that partial-thickness tears were likely to have been over-diagnosed by US, but also may have been under-diagnosed by arthroscopy, and we therefore recommend that the practitioners of both techniques carry out reflection on the diagnostic challenges that contribute to such errors.

**REFERENCES**


3. Brenneke SL, Morgan CJ. Evaluation of ultrasonography as a diagnostic technique in the assessment of rotator cuff...
EVALUATING THE SUCCESS OF PREOPERATIVE IMAGING FOR DIAGNOSING ROTATOR CUFF TEARS

329


33. Stetson WB, Phillips T, Deutsch A. The use of magnetic resonance arthrography to detect partial-thickness rotator...


