Intra- and inter-rater reliability of the detection of tears of the supraspinatus central tendon on MRI by shoulder surgeons

John A. Grant, PhD, MD, FRCSC\textsuperscript{a}, Bruce S. Miller, MD, MS\textsuperscript{a,\ast}, Jon A. Jacobson, MD\textsuperscript{b}, Yoav Morag, MD\textsuperscript{b}, Asheesh Bedi, MD\textsuperscript{a}, James E. Carpenter, MD\textsuperscript{a}, and the MOON Shoulder Group\textsuperscript{+,\textsuperscript{c}}

\textsuperscript{a}MedSport, Department of Orthopaedic Surgery, University of Michigan, Ann Arbor, MI, USA
\textsuperscript{b}Division of Musculoskeletal Radiology, Department of Radiology, University of Michigan, Ann Arbor, MI, USA
\textsuperscript{c}The Multi-center Orthopaedic Outcomes Network (MOON) Shoulder Group

\textbf{Background:} The purpose of this study was to determine the intra- and inter-rater reliability of detecting full- and partial-thickness tears of the supraspinatus intramuscular central tendon on magnetic resonance imaging (MRI) by orthopaedic shoulder surgeons. Full-thickness tears of this tendon have previously been associated with the failure of nonsurgical management of rotator cuff tears.

\textbf{Methods:} Shoulder MRIs from 40 patients entered into a prospective rotator cuff disease database were independently reviewed by two musculoskeletal (MSK) radiologists in order to determine if there was a partial- or full-thickness tear of the supraspinatus central tendon. The MRIs were randomly sorted and distributed to 16 fellowship-trained shoulder surgeons. The surgeons then similarly diagnosed each patient. After a 1-month interval, surgeons repeated the evaluation with the same set of randomly reordered MRIs. Surgeon intra- and inter-rater reliability was determined with the kappa statistic. Agreement and inter-rater reliability were also determined between the shoulder surgeons and MSK radiologists.

\textbf{Results:} For full-thickness tears, the intra-rater reliability was excellent (0.86 ± 0.1, 95\% confidence interval (CI): 0.81, 0.91) and the agreement was 93.4\% ± 4.6, 95\% CI: 91.1, 95.8. Inter-rater reliability for both rounds was also excellent (0.77 and 0.74). The agreement between the shoulder surgeons and MSK radiologists was 92.9\% ± 3.9, 95\% CI: 90.9, 94.9, and the kappa was 0.85 ± 0.08, 95\% CI: 0.81, 0.89. Including partial-thickness tears resulted in agreement of 65-92\% and kappa values of 0.59-0.72.

\textbf{Conclusion:} The reliability for the MRI detection of full thickness tears of the supraspinatus central tendon among shoulder surgeons and between shoulder surgeons and MSK radiologists was excellent.

\textbf{Level of evidence:} Level III, Nonconsecutive Patients, Diagnostic Study.

\textcopyright\ 2013 Journal of Shoulder and Elbow Surgery Board of Trustees.

\textbf{Keywords:} MRI; rotator cuff tear; reliability; central tendon; intramuscular tendon; agreement
Partial and full thickness rotator cuff tears are common clinical problems and remain challenging to manage because of their complex natural history and broad variations in their clinical presentations. Nonsurgical treatment can provide satisfactory results in some patients with significant improvement in their symptoms. Some patients, however, will fail to improve after a course of nonsurgical management and subsequently require surgical management. The main challenge to clinicians is successfully identifying those individuals who would benefit from immediate surgical intervention and distinguishing them from those who could be successfully treated without surgery.

A recent study reported successful nonsurgical treatment of full thickness rotator cuff tears in patients who had at least 3 of the 4 following attributes: 1) absence of an impingement sign; 2) preserved external rotation range of motion; 3) little or no fatty atrophy of the supraspinatus muscle; 4) an intact intramuscular central tendon of the supraspinatus. The intramuscular tendon was initially identified on magnetic resonance imaging (MRI) as a low signal void centrally in the supraspinatus muscle. This structure has similar signal characteristics to tendon and is identified on magnetic resonance imaging (MRI) as a low signal void centrally in the supraspinatus muscle. Given that the central tendon comprises only a portion of the whole supraspinatus tendon, it is therefore possible to have an intact central tendon in the presence of a tear to another part of the supraspinatus tendon (ie, more anterior or posterior to the central tendon).

The overall reliability of detecting tears of the rotator cuff tendons on MRI has been reported to be highly variable, ranging from 0.37 to 0.80 (Cohen’s kappa). The reliability for the detection of full thickness tears (0.67–0.88) is better than the reliability for the detection of partial thickness tears (0.13–0.44), demonstrating the challenge of differentiating the increased signal characteristics of partial tears compared to inflammatory or granulation tissue. To date, however, there have been no studies evaluating the intra- and inter-rater reliability for the specific detection of supraspinatus intramuscular central tendon tears on MRI.

The purpose of this study was to determine the intra- and inter-rater reliability of detecting a partial- or full-thickness tear of the supraspinatus intramuscular central tendon on MRI by orthopaedic shoulder surgeons.

Materials and methods

A diagnostic test-retest design was used in this study. Fully de-identified shoulder MRIs were obtained from a prospective database of rotator cuff disease patients maintained by the senior author (B.S.M.). These images included sagittal- and coronal-oblique fast-spin echo (FSE) sequences with fat-saturation (TR/TE = 3000-3500/40-61). The echo train length was 12 with a 3-mm slice thickness and 1-mm gap. The field of view was 16 cm² with a 256 × 192 matrix. Images were acquired on either a Phillips (Philips Healthcare, Andover, MA, USA) or GE (GE Healthcare, Waukesha, WI, USA) 1.5 Tesla machine with a dedicated shoulder coil. Sagittal- and coronal-oblique sequences were reviewed by a fellowship-trained musculoskeletal (MSK) radiologist (J.J.) to ensure that the sequences were of diagnostic quality and that the greater tuberosity was parallel to the glenoid. Two MSK radiologists (J.J. and Y.M.) then independently reviewed the MRIs to determine whether the supraspinatus central tendon was intact, partially torn, or fully torn. The criteria for diagnosing a tendon tear was the presence of fluid signal identifying discontinuity of the central tendon on 2 imaging planes. Importance was placed on differentiating fluid signal identifying a discontinuity somewhere in the rotator cuff insertion versus discontinuity specifically in the area of the insertion of the supraspinatus central tendon. The independent diagnoses were then compared to identify cases in which the MSK radiologists had concordant diagnoses. This process identified 10 patients (38.4 ± 13.6 years old, mean ± SD) with an intact central tendon, 15 patients (53.5 ± 11.1 years old) with a partial tear of the central tendon, and 15 patients (60.5 ± 13.4 years old) with a full tear of the central tendon for a total of 40 cases. The complete study sample was 77.5% male with a mean age of 52.3 ± 15.1 years old. As the intramuscular tendon is not visible on arthroscopy, the radiologists’ diagnoses were compared to those of the shoulder surgeons.

The 40 MRIs were then copied onto compact disc (CDs) in random order. The CDs contained software that allowed the MRI images to be displayed independent of a computerized diagnostic imaging system. Typical diagnostic imaging tools such as zoom, contrast, brightness, etc were available as part of this software. A set of CDs was then distributed to 16 fellowship-trained orthopaedic shoulder surgeons who are members of the Multi-center Orthopaedic Outcomes Network (MOON) Shoulder Group. The surgeons were blinded to any patient demographic or clinical data related to the MRI images that they reviewed. An educational PowerPoint (Microsoft, Redmond, WA, USA) presentation was also distributed to all the reviewers. This presentation included information about the study and specific anatomical and radiological images demonstrating the supraspinatus intramuscular central tendon along with typical sagittal- and coronal-oblique MRI images as examples of an intact, partially torn, and fully torn central tendon (Figs. 2-4 are examples from different patients). The surgeons were instructed to focus on the status of the central tendon itself, regardless of whether or not there was a partial- or full-thickness tear of some part of the supraspinatus tendon. The reviewers were asked to use the sagittal- and coronal-oblique MRI images to determine whether or not there was a tear of the central tendon, and if so, if the tear was partial or complete. These diagnoses were marked on a standard data collection form and then returned to the study coordinator. A second set of CDs was produced with the same 40 MRIs randomly reordered. One month following the return of the round one data form, the second set of CDs was sent to each reviewer. The 1-month period was used to help reduce recall bias. The reviewers then completed the data form in the same manner as previously performed.
Double data entry was performed independently by the study coordinator and one of the authors to ensure data accuracy. To assess intra- and inter-observer reliability, responses from each reviewer were cross-tabulated to enable the calculation of agreement and the kappa statistic. The kappa statistic was categorized as poor (<0.50), good (0.51–0.75), or excellent (>0.75) agreement. Inter-observer multi-rater generalized kappa statistics were calculated with Stata V6.0 (Stata Corporation, College

Figure 1 Sagittal-oblique MRI images of the supraspinatus central tendon. Images A through D move successively from medial to lateral.

Figure 2 Coronal-oblique image of an intact supraspinatus central tendon.

Figure 3 Coronal-oblique MRI image of a partially torn supraspinatus central tendon (partially torn laterally).
Intra-rater reliability was calculated within each shoulder surgeon comparing rounds 1 and 2. The individual agreement and intra-rater kappa statistics for each surgeon were then averaged across all 16 surgeons and 95% confidence intervals (CI) for the mean were calculated. Inter-rater reliability was calculated between all 16 shoulder surgeons. These were calculated separately for both rounds 1 and 2. Inter-rater agreement and reliability was also calculated between each shoulder surgeon and the radiologists’ diagnoses. The individual agreement and inter-rater kappa statistics for each surgeon were then averaged (with 95% CI) across all 16 surgeons. This was done for both rounds 1 and 2. All kappa statistic calculations were unweighted as there was no clinical evidence to support a different weighting for partial tears. Measures of agreement and the kappa statistic were first determined with two possible mutually exclusive diagnosis categories: 1) no full tear, or 2) full tear of the supraspinatus central tendon. In this manner, “partial tears” and “no tear” of the central tendon were coded as “no full thickness tear.” In order to determine the reliability of using a more sensitive diagnosis system, an exploratory analysis was performed by comparing the 3 mutually exclusive diagnosis categories of: 1) no tear; 2) partial tear; or 3) full thickness tear of the supraspinatus central tendon.

Results

For the primary analysis of “no full thickness tear” versus “full thickness tear,” the inter-rater reliabilities (kappa) for the shoulder surgeons were 0.77 (round 1) and 0.74 (round 2). The mean (±SD) intra-rater agreement was 93.4% ± 4.6 (95% CI: 91.1, 95.8) and the mean intra-rater kappa statistic was 0.86 ± 0.1 (95% CI: 0.81, 0.91). The mean agreements between the MSK radiologists and the shoulder surgeons for rounds 1 and 2 were 92.9% ± 3.9 (95% CI: 90.9, 94.9) and 91.4% ± 4.1 (95% CI: 89.3, 93.5), respectively. The mean kappa statistics were 0.85 ± 0.08 (round 1; 95% CI: 0.81, 0.89) and 0.82 ± 0.09 (round 2; 95% CI: 0.77, 0.86). The diagnosis-specific agreements between the shoulder surgeons and the MSK radiologists are reported in Table I.

When comparing the diagnoses of no tear, partial tear, or full thickness tear of the supraspinatus central tendon, the mean (±SD) intra-rater agreement between rounds 1 and 2 was 81.7% ± 8.5 (95% CI: 77.4, 86.0) and the mean intra-rater kappa statistic was 0.72 ± 0.13 (95% CI: 0.66, 0.79). Across the 16 shoulder surgeons, the inter-rater kappa statistics for rounds 1 and 2 were 0.60 and 0.59, respectively. The mean agreements between the MSK radiologists and the shoulder surgeons for rounds 1 and 2 were 78.2% ± 6.8 (95% CI: 74.8, 81.7) and 76.7% ± 5.5 (95% CI: 73.9, 79.5), respectively. The associated mean kappa statistics were 0.67 ± 0.1 (round 1; 95% CI: 0.62, 0.72) and 0.65 ± 0.08 (round 2; 95% CI: 0.61, 0.69). The agreement between the MSK radiologists and the shoulder surgeons for each of the 3 diagnosis categories are listed in Table II.

Discussion

The current study is the first to evaluate the reliability and agreement of orthopedic shoulder surgeons in identifying tears of the supraspinatus central tendon on MRI. The agreement values for the determination of full thickness tear or no full thickness tear were all above 90% with excellent reliability kappa statistics of 0.74–0.86. As may be expected, the intra-rater reliability and agreement were better than those of the between surgeon and surgeon-radiologist inter-rater reliability. These agreement and kappa values are in line with a previous evaluation of the reliability of fellowship-trained shoulder surgeons for the differentiation of partial- versus full-thickness tears of the whole rotator cuff (agreement = 89%, kappa = 0.77).22

When the possible diagnostic categories were expanded to differentiate no tear from a partial tear of the central tendon, the lowest agreement levels were seen for the diagnosis of a partial tear (Table II). This may be partially explained by the fact that increased signal in an otherwise intact appearing tendon could represent a true partial thickness tear, tendinosis, or granulation tissue.1,5,19 Kappa values will be lower if there is difficulty in differentiating between diagnosis categories.4 The inter-rater reliability between surgeons

![Figure 4](Coronal-oblique image of a full thickness tear (with retraction) of the supraspinatus central tendon.)
(kappa = 0.60, 0.59) was within the range of those previously demonstrated for rotator cuff tears (kappa = 0.37–0.80). Intra-rater reliability (kappa = 0.72) was found to be better than the inter-rater reliability between surgeons and between surgeons and MSK radiologists (kappa = 0.67, 0.65). The associated agreement was in the range of 77% to 82%.

There have been no previous studies evaluating the reliability of detecting tears of the central tendon of the supraspinatus. There are multiple reports in the radiology literature discussing the reliability of detecting tears of the whole rotator cuff with MRI. These studies have evaluated the reliability between different levels of training (junior residents through experienced faculty), different imaging techniques (indirect MR arthrography with intravenous gadolinium), and different shoulder positions (standard arm at the side position compared to abduction and external rotation of the shoulder). The kappa statistics and level of agreement were generally lower for those reviewers with less experience and for partial thickness tears relative to full thickness tears. Additionally, a study involving fellowship-trained orthopaedic shoulder surgeons demonstrated agreement of 89% and a kappa of 0.77 for the differentiation of partial- versus full-thickness rotator cuff tears on MRI. Although the current study evaluated the status of a small part of the rotator cuff, the findings are consistent, if not better, than the previous rotator cuff studies. Balich et al reported kappa values of 0.63 to 0.80 when using three diagnosis categories (full thickness tear, partial thickness tear, no tear). In comparison, Robertson et al included a fourth diagnostic category (tendinitis) with lower resultant kappa values of 0.37–0.50. These studies demonstrate how the kappa value (and therefore reliability) can be negatively affected by a lack of clear definition or independence of the diagnostic categories included. In their study evaluating the reliability of shoulder surgeons to diagnose shoulder pathology on MRI, Spencer et al have suggested that scoring/rating systems with fewer choices or more objective measures would likely improve agreement. Given the issues with differentiating the MRI signal associated with partial thickness tendon tears from tendon inflammation, tendinosis, or granulation tissue, the categories of partial tear, tendinosis, and no tear are not well defined. In the current study, the agreement for the diagnosis of a partial tear of the central tendon was lower than either the normal or full thickness tear diagnoses (Table II). Combining the partial and no tear diagnoses substantially improved the agreement (91–93%) and the kappa statistics (0.74–0.86), while still allowing the study to attain its goal of determining the ability to reliably identify full-thickness tears.

The current study included a relatively large number of reviewers (16) with a similar level of training (fellowship-trained in shoulder surgery). A similar educational background allows for a more accurate kappa statistic which is therefore likely reproducible for other surgeons with similar training. The patient MRIs were presented in random order and a one month interval between review sessions was used to decrease recall bias. Educational materials, including detailed MRI images demonstrating examples of each of the diagnoses categories, were provided to each reviewer. This is the first study to evaluate the reliability of detecting tears in the intramuscular central tendon of the supraspinatus. Additionally, this study has included the evaluation of reliability between fellowship-trained shoulder surgeons and between those surgeons and MSK radiologists.

The shoulder surgeons were blinded to all patient demographic and clinical information. This may be seen as both a strength and a limitation of this study. The more information that reviewers are given to aid in the diagnosis, the higher the kappa values should be and the study becomes more generalizable to a true clinical scenario. The complete blinding of patient information in the current study provides a baseline, or worst case scenario estimate of the reliability. Given the excellent reliability found in this study, the addition of patient history and physical exam findings should result in highly reliable outcomes in the clinical setting.

The use of the kappa statistic for the determination of reliability is an improvement over the standard agreement calculation in that it takes into account the amount of agreement that would be expected by chance alone. This correction is based on the marginal totals from the data table and is therefore sensitive to the prevalence of the “disease” in the study. In this manner, 2 studies with the same agreement can have widely different kappa values dependent on the prevalence of “disease” in each study. The convenience patient sample used in the current study was chosen both on the quality of the imaging studies and to allow for a similar distribution of patients with a diagnosis (as made by the MSK radiologists) of no tear, partial tear, or full thickness tear of the supraspinatus central

<table>
<thead>
<tr>
<th>Table I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agreement (%) between the MSK radiologists and the shoulder surgeons for the diagnosis of “full tear” or “no full tear” of the supraspinatus central tendon</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>No tear</td>
</tr>
<tr>
<td>Full tear</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agreement (%) between the MSK radiologists and the shoulder surgeons for the status of the supraspinatus central tendon</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>No tear</td>
</tr>
<tr>
<td>Partial tear</td>
</tr>
<tr>
<td>Full tear</td>
</tr>
</tbody>
</table>
tendon. In this regard, a similar study performed with a skewed distribution of these 3 diagnoses may result in different kappa values.

The current study does not directly provide a new clinical decision making tool for use in the management of rotator cuff tear patients but it does provide validity and generalizability to the findings of Tanaka et al. This study has shown that shoulder surgeons can reliably identify whether or not the intramuscular tendon of the supraspinatus is intact. Within the methodological merits of the aforementioned study, shoulder surgeons can now confidently use and further evaluate the recommendations of Tanaka et al in their own practice for the management of patients with full-thickness rotator cuff tears.

The reliability of shoulder surgeons and MSK radiologists to determine whether or not there is a full thickness tear of the supraspinatus intramuscular central tendon has been confirmed. Anatomical and clinical research is now required to further advance the use of this potential outcome for clinical decision making. Given the current lack of a gold standard for a live patient, further work could involve imaging cadaveric shoulders to determine the status of the central tendon, followed by an anatomical dissection of the same shoulders to confirm or refute the MRI diagnosis.

Clinically, there are three possible questions to answer. First, a replication of the study by Tanaka et al should be performed to support or refute their findings that an intact supraspinatus intramuscular tendon is 1 factor that may help predict the successful nonsurgical management of patients with full thickness rotator cuff tears. Second, imaging research is required to help identify ways for differentiating inflammatory, tendinotic, or granulation tissue from partial tears of the intramuscular tendon (or the rotator cuff tendons in general) on MRI. This may allow an evaluation of the role that partial thickness rotator cuff (or central tendon) tears play in the success of non-surgical management of these tears. Finally, the correlation between tears of the supraspinatus tendon as a whole and the status of the central tendon (torn or not torn) needs to be evaluated in order to understand how having an intact central tendon affects the clinical presentation of a patient with a full thickness rotator cuff tear.

There are three main limitations to the present study. Given the inability to arthroscopically view the intramuscular component of the supraspinatus central tendon, there is currently no gold standard with which to compare the MRI diagnosis. As such, the reliability and agreement can be determined, but the sensitivity and specificity of the MRI diagnoses cannot be evaluated. The current study was performed with a 1.5T MRI scanner using a consistent imaging protocol. The reliability may therefore vary with different magnet strengths or imaging protocols. Additionally, this study was not designed to determine the effects of patient age, tear chronicity, and muscle atrophy on the reliability of detecting supraspinatus central tears using MRI.

**Conclusion**

The intra-and inter-rater reliability between shoulder surgeons, and the inter-rater reliability between shoulder surgeons and MSK radiologists, was excellent for the detection of full thickness tears of the central tendon of the supraspinatus. When including the diagnosis of partial-thickness tears, the agreement and reliability was lower. Because there is agreement among orthopaedic surgeons and radiologists about the diagnoses of a full-thickness tear in the supraspinatus central tendon on MRI, this parameter should be further studied as a possible predictor of outcome for the nonsurgical management of full thickness rotator cuff tears.

**Acknowledgment**

The authors would like to thank Theresa Kijeck for her work as the study coordinator and Dr. Adnan Cutuk for his work in the initial planning stages of this study.

**Disclaimer**

None of the other authors, their immediate families and any research foundation with which they are affiliated received any financial payments or other benefits from any commercial entity related to the subject of this article.

**References**


