



Published in final edited form as:

PM R. 2015 March ; 7(3): 245–254.e3. doi:10.1016/j.pmrj.2014.08.949.

## Reliability of Magnetic Resonance Imaging Assessment of Rotator Cuff: The ROW Study

Nitin B. Jain, MD, MSPH<sup>1,2,3</sup>, Jamie Collins, PhD<sup>2</sup>, Joel S. Newman, MD<sup>5</sup>, Jeffrey N. Katz, MD, MS<sup>2,4</sup>, Elena Losina, PhD<sup>2,4,6</sup>, and Laurence D. Higgins, MD<sup>2</sup>

<sup>1</sup>Department of Physical Medicine and Rehabilitation, Spaulding Rehabilitation Hospital and Harvard Medical School, Boston, MA

<sup>2</sup>Department of Orthopaedic Surgery, Brigham and Women's Hospital and Harvard Medical School, Boston, MA

<sup>3</sup>Departments of Physical Medicine and Rehabilitation and Orthopaedics, Vanderbilt University Medical Center, Nashville, TN

<sup>4</sup>Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital and Harvard Medical School, Boston, MA

<sup>5</sup>Department of Radiology, New England Baptist Hospital, Boston, MA

<sup>6</sup>Department of Biostatistics, Boston University School of Public Health, Boston, MA

### Abstract

**Background**—Physiatrists encounter patients with rotator cuff disorders and imaging is frequently an important component of their diagnostic assessment. However, there is paucity of literature on reliability of MRI assessment between shoulder specialists and musculoskeletal radiologists.

**Objective**—We assessed inter- and intra-rater reliability of MRI characteristics of the rotator cuff.

**Design**—Cross-sectional secondary analyses in a prospective cohort study

**Setting**—Academic tertiary care centers

**Patients**—Subjects with shoulder pain recruited from orthopedic and physiatry clinics

---

© 2014 by the American Academy of Physical Medicine and Rehabilitation. All rights reserved.

**Corresponding Author:** Nitin B. Jain, MD, MSPH, Department of Physical Medicine and Rehabilitation, Vanderbilt University Medical Center, 2201 Children's Way, Suite 1319, Nashville, TN 37202, Phone: (615) 936-8508; Fax: (615) 322-7454, nitin.jain@vanderbilt.edu.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

**Disclosure:** This study was presented in an abstract form at the Association of Academic Physiatrists (AAP) annual meeting in February 2014 in Nashville, TN.

**Methods**—Two shoulder fellowship trained physicians (a physiatrist and a shoulder surgeon) jointly performed a blinded composite MRI review by consensus on 31 subjects with shoulder pain. Subsequently, MRI was reviewed by one fellowship trained musculoskeletal radiologist.

**Main Outcome Measures**—We calculated Cohen’s kappa coefficients and percent agreement among the two reviews (composite review of two shoulder specialists versus that of the musculoskeletal radiologist). Intra-rater reliability was assessed among the shoulder specialists by performing a repeat blinded composite MRI review. In addition to this repeat composite review, only one of the physiatry shoulder specialists performed an additional review.

**Results**—Inter-rater reliability (shoulder specialists versus musculoskeletal radiologist) was substantial for the presence or absence of tear (kappa=0.90; 95% CI=0.72, 1.00), tear-thickness (kappa=0.84;95% CI=0.70, 0.99), longitudinal size of tear (kappa=0.75;95% CI=0.44, 1.00), fatty infiltration (kappa=0.62; 95% CI=0.45, 0.79), and muscle atrophy (kappa=0.68; 95% CI=0.50, 0.86). There was only fair inter-rater reliability of transverse size of tear (kappa=0.20; 95% CI=0.00, 0.51). The kappa for intra-rater reliability was high for tear thickness (0.88; 95% CI=0.72, 1.00), longitudinal tear size (0.61; 95% CI=0.22, 0.99), fatty infiltration (0.89; 95% CI=0.80, 0.98), and muscle atrophy (0.87; 95% CI=0.76, 0.98). Intra-rater reliability for the individual shoulder specialist was similar to that of the composite reviews.

**Conclusions**—There was high inter-rater and intra-rater reliability for most findings on shoulder MRI. Our data supports the reliability of MRI assessment by shoulder specialists for rotator cuff disorders.

**Level of Evidence**—Level I (testing of a previously developed diagnostic criteria in a series of consecutive patients with an accepted “gold” standard)

## Keywords

rotator cuff; reliability; MRI

## INTRODUCTION

Shoulder symptoms accounted for approximately 11.5 million ambulatory care visits to physician offices in 2010 in the United States.<sup>1</sup> Rotator cuff tears are one of the leading causes of shoulder pain.<sup>2</sup> An estimated 272,148 rotator cuff repairs were performed in 2006 on an ambulatory basis.<sup>3</sup> Factors associated with outcomes of rotator cuff repair include size and thickness of tear<sup>4,5</sup>, tendon retraction<sup>6,7</sup>, muscle atrophy<sup>8,9</sup>, and fatty infiltration<sup>8,10,11</sup>. These characteristics are assessed by imaging modalities such as Magnetic Resonance Imaging (MRI). Surgical decision-making at the point of care is made by shoulder specialists on the basis of a range of factors including imaging findings. While imaging findings are important to clinical decision making in patients with rotator cuff disorders, limited data exist on the reliability of these findings and in particular on agreement between clinicians and musculoskeletal radiologists.

Thus, the objective of our study was to assess inter- and intra-rater reliability of MRI characteristics of the rotator cuff. We examined inter-rater reliability between shoulder specialists and a musculoskeletal radiologist as well as the intra-rater reliability of a shoulder

specialist. The MRI characteristics included rotator cuff tear size and thickness, presence of tendonitis/tendinopathy, and grades of fatty infiltration, muscle atrophy, and supraspinatus tendon retraction.

## MATERIALS AND METHODS

### Patient Population

The investigators are recruiting a longitudinal cohort of patients with shoulder pain (with and without rotator cuff tears) from orthopedic and physiatry clinics at two academic medical centers. Eligibility criteria for this cohort study, termed the ROW (Rotator Cuff Outcomes Workgroup) Study include age 45 years and older and shoulder symptoms for at least 4 weeks. Exclusion criteria include a current shoulder fracture, prior shoulder surgery, evidence of cervical radiculopathy (assessed by neck pain radiating to the shoulder/arm/hand), and presence of claustrophobia, pacemaker, defibrillator, or other surgical hardware that would be a relative MRI contraindication. All of the eligibility and exclusion criteria were applied to the symptomatic shoulder in patients 45 years and older. In cases with bilateral shoulder involvement, the most painful shoulder was assessed. We obtained approval for this study from our Institutional Review Board and obtained informed consent from all participants.

From 02/2011 to 07/2012, we recruited 153 patients with shoulder pain and recruitment is ongoing. For the current study, we selected all subjects who had a shoulder MRI at our institution (n=31). This sub-set was selected to maintain consistency in the equipment used for imaging and MRI sequences available for review.

### Standardized MRI Assessments

MRI was performed on a General Electric 1.5-T magnet (Waukesha, WI, USA) using a dedicated shoulder coil (Invivo, Gainesville, FL, USA). Fields of view ranged from 14 to 16 cm with sequences obtained in the sagittal oblique, coronal oblique, and axial planes. Slice thickness was 4 mm in the coronal and sagittal planes and 3 mm in the axial plane. The following sequences were obtained: coronal oblique fast-spin-echo (FSE) proton-density weighted images, coronal oblique FSE short tau inversion-recovery (STIR) images, sagittal FSE proton-density or T2-weighted images with fat suppression, sagittal FSE T1-weighted images, axial T1-weighted images, and axial T-2 weighted gradient echo images.

A standardized MRI reading form (Appendix A) was developed based on published literature and input from our imaging and shoulder experts prior to recruitment for the study. MRI review was performed jointly by two shoulder experts (LDH, a shoulder surgeon, with about 15 years of experience after shoulder fellowship training and NBJ, a physiatrist and recent shoulder fellowship graduate). Thus, these two clinicians, henceforth referred to as Reviewers 1, provided a composite review. Any differences were resolved by consensus. The MRI reviewers were blinded to patient identifiers and other clinical information. The MRI review sessions were performed on approximately a monthly basis. A blinded MRI review was subsequently performed independently by a musculoskeletal radiologist with over 15 years of experience (JN; henceforth referred to as Reviewer 2). To assess intra-rater

reliability, select characteristics of the supraspinatus including tear-thickness, tear size, and tendon retraction were assessed during a second review by Reviewers 1 occurring at least four months after the original review. The supraspinatus was selected since it was the most commonly torn tendon in our study. Fatty infiltration and muscle atrophy was assessed for all 4 rotator cuff muscles. Finally, the physiatrist among the two shoulder experts (NBJ) who took part in the composite review performed an additional unassisted review, at least four months after the last review. The reliability of this review by the physiatrist versus Reviewers 1 and Reviewer 2 was also assessed.

The following parameters were assessed during the MRI review:

**Tear Thickness**—A full-thickness tear was diagnosed when there was complete disruption of all tendon fibers or when the signal within the cuff tendons was isointense compared with fluid on the T2-weighted images and extended from the articular to the bursal surface on one or more images.<sup>12,13</sup> A partial-thickness tear was diagnosed when fluid-intensity signal within the tendons was in contact with only one of the surfaces or if there was a discontinuity of some but not all tendon fibers.<sup>12,13</sup> For the purposes of this investigation, mild fraying of the tendon by itself was not sufficient to constitute a partial-thickness tear. Tendinosis/tendinopathy was diagnosed if the tendon showed increased signal intensity on proton-density or T1-weighted images without further increase in signal on T-2 weighted images, and without disruption of the tendon.<sup>14,15</sup> On T2 weighted images, the abnormal signal had to be lower than fluid intensity (since fluid intensity would suggest a tear).<sup>13–15</sup> Additional findings of tendinopathy included tendon thickening, whether focal or fusiform.

**Tear Size**—Tear size was graded as small (<1 cm), medium (1–3 cm), large/massive (>3 cm).<sup>16</sup> Tear size was assessed both in the longitudinal and transverse planes.

**Number of Tendons**—All rotator cuff tendons (supraspinatus, infraspinatus, subscapularis, and teres minor) were assessed (except during second reading performed by Reviewers 1 for intra-rater reliability). Single versus multiple tendon tears were noted. For purposes of this analysis, only the largest tear for each patient was analyzed.

**Fatty Infiltration**—Fatty infiltration was evaluated on the basis of fatty streaks within the muscle belly observed on a T1-weighted oblique sagittal image. It was graded as: grade 0, no fat; grade 1, thin streaks of fat; grade 2, less fat than muscle; grade 3, equal amounts of fat and muscle; and grade 4, more fat than muscle as described by Goutallier et al.<sup>10</sup> The original article by Goutallier et al. described fatty infiltration based on Computed Tomography (CT) scan findings. However, MRI offers superior resolution of muscle as compared with CT scan and multiple prior studies have used MRI for fatty infiltration grading.<sup>17–19</sup> Moreover, it is also standard clinical practice to use MRI for rotator cuff assessment as opposed to CT scan. All four tendons were evaluated for our study.

**Muscle Atrophy**—Muscle atrophy was graded according to the scale by Warner et al.<sup>20</sup> based on an oblique sagittal plane in the most lateral image where the coracoid and scapular spine meet the scapular body. This position has been found to be easily reproducible.<sup>21</sup>

Atrophy was graded as none, mild, moderate, and severe. Although the original study used CT scan for muscle atrophy grading, MRI offers muscle resolution that is better than CT scan. Prior studies have also used MRI for muscle atrophy grading.<sup>17</sup>

**Tendon Retraction**—Tendon retraction in the coronal plane was classified as described by Boileau et al.<sup>22</sup>. A tear was classified as stage I tear if the medial edge of the torn tendon was over the greater tuberosity. Stage II tears exposed the humeral head but did not retract to the glenoid. If the tendon retracted to the glenoid, the tear was classified as Stage III. Stage IV tears were retracted medial to the glenoid.

### Statistical Analysis

To quantify inter- and intra-rater reliability, we calculated percent agreement and Cohen's kappa coefficients with 95% confidence intervals.<sup>23</sup> Cohen's kappa is a measure of agreement calculated based on expected versus observed values; it corrects for agreement based on chance alone.<sup>24</sup> For ordinal outcomes we calculated weighted kappas, which account for the degree of disagreement between ordinal outcomes.<sup>25</sup> While there is no standardized guideline for the kappa value that constitutes acceptable agreement, Landis and Koch recommend the following divisions: poor agreement with kappa < 0; slight agreement with kappa between 0–0.2; fair agreement with kappa from 0.2–0.4; moderate agreement with kappa of 0.4–0.6; substantial agreement if kappa is between 0.6 and 0.8, and; almost perfect agreement for kappa 0.8–1.0.<sup>26</sup> Since the precision of point estimates in our study was low (wide confidence intervals), we have used the term substantial agreement to describe kappa values between 0.6 and 1.0.

*Kappa paradox:* Occasionally even in the setting of high agreement, a measurement may have poor kappa simply because of the lack of variability in the population and not because of the intrinsic inaccuracy of the measurement itself.<sup>27</sup> If inter-subject variability is small (the prevalence of a trait is very rare or exceedingly common), then the expected or chance agreement becomes so large that the kappa statistic is difficult to interpret. Therefore, in addition to kappa we have reported percent agreement. These are especially relevant for fatty infiltration and muscle atrophy, as very few patients in our study have severe grades of muscle degradation. Since the grading categories of fatty infiltration and atrophy are identical for all four rotator cuff muscles, we combined the four muscles for inter- and intra-rater reliability analysis of fatty infiltration and atrophy. We have also provided results stratified by the rotator cuff muscle for fatty infiltration and atrophy in the Appendix.

An *a priori* sample size calculation was performed to ensure reasonable precision in the estimate of agreement statistics. For an *a priori* estimate of 90% agreement, a sample size of 27 participants provided a 95% confidence interval of 78.5% to 100% agreement. We performed statistical analyses using SAS for Windows (version 9.2), SAS Institute Inc., (Cary, NC).

## RESULTS

A majority of participants in our study were female (58%) and the mean age was  $60.7 \pm 8.7$  years (Table 1). The mean duration of symptoms was  $19.8 \pm 35.5$  months.

There was substantial inter-rater reliability for the presence or absence of a rotator cuff tear ( $\kappa=0.90$ ; 95% CI=0.72, 1.00; Table 2). Even when reviewers differentiated between tear-thickness (full versus partial versus no tear), there was substantial inter-rater reliability ( $\kappa=0.84$ ; 95% CI=0.70, 0.99). Few examples of disagreement between reviewers' ratings of partial-thickness versus full-thickness tears are presented in Figures 1, 2, and 3. When reviewers' further differentiated between tear-thickness and presence of tendinopathy, the kappa value dropped to 0.68 (95% CI=0.54, 0.82). Most of the differences in reviewers' ratings were explained by disagreement between tendinopathy versus normal tendon.

There were variable results for reliability of tear size. The kappa for inter-rater reliability of longitudinal size of rotator cuff tear was 0.75 (95% CI=0.44, 1.00) whereas that for transverse size of tear was 0.20 (95% CI=0.00, 0.51). There was substantial reliability between the reviewers' rating of tendon retraction stages as described by Boileau et al ( $\kappa=0.77$ ; 95% CI=0.64, 0.90).

Grading of fatty infiltration of the rotator cuff had a kappa of 0.62 (Table 3) although the precision of the estimate was low (95% CI=0.45, 0.79). An example of agreement is presented in Figure 4. When fatty infiltration was stratified by rotator cuff tendons, there was substantial reliability between the raters for supraspinatus and subscapularis, and moderate reliability for infraspinatus and teres minor (Appendix B; Table 3.a). The percent agreement between the raters was high (a range of 0.69–0.87 across the four tendons). Similarly, ratings of muscle atrophy had a kappa of 0.68 (95% CI=0.50, 0.86). When stratified by rotator cuff tendons, there was substantial reliability for teres minor, supraspinatus and subscapularis, and moderate reliability for infraspinatus (Appendix B; Table 3.b).

The intra-rater reliability for shoulder experts in this study was substantial across all variables that were assessed except for transverse tear size (Table 4). Kappa for tear thickness was 0.88 (95% CI=0.72, 1.00), for tendon retraction was 0.88 (95% CI=0.78, 0.97), for fatty infiltration was 0.89 (95% CI=0.80, 0.98), and for muscle atrophy was 0.87 (95% CI=0.76, 0.98). The reliability when only one of the two shoulder experts reviewed the MRI versus the composite read (of Reviewers 1) was also substantial across all variables except for transverse tear size (Appendix B; Table 5). Kappa for tear thickness was 0.86 (95% CI=0.70, 1.00), for tendon retraction was 0.85 (95% CI=0.72, 0.98), and for fatty infiltration was 0.73 (95% CI=0.53, 0.93). When the review by one shoulder expert was assessed against the review by Reviewer 2, the results were similar except for fatty infiltration where the kappa decreased to 0.50 (95% CI=0.30, 0.70) (Appendix B; Table 5).

## DISCUSSION

We studied reliability of routinely assessed MRI characteristics of the rotator cuff. These characteristics are an essential part of the decision-making process in treatment of patients with rotator cuff tears. We found that there was substantial agreement between shoulder experts and a musculoskeletal radiologist in assessment of the presence and thickness of tear. There was substantial agreement in rating longitudinal tear size (though not for transverse tear size), tendon retraction, fatty infiltration, and muscle atrophy. The intra-rater

reliability among shoulder experts was also substantial. These findings were also consistent when only one of the two shoulder experts reviewed the MRI. Our data has to be interpreted with caution since confidence intervals around the point estimates in our study were wide. Our data provides clinicians with valuable information on the reliability of MRI interpretation by expert clinicians as compared with musculoskeletal radiologists.

Only few prior studies have reported on reliability of tear size, tear thickness, and presence of tendinopathy. Balich et al. reported good to excellent inter-rater reliability for the presence and thickness of rotator cuff tear among five radiologists.<sup>28</sup> In a retrospective study of 67 patients, two radiologists had near perfect inter-observer reliability ( $\kappa=0.91$ ) for diagnosis of full-thickness tears and moderate reliability ( $\kappa=0.49$ ) for partial-thickness tears on Magnetic Resonance Arthrography (MRA).<sup>29</sup> In another retrospective study of 97 patients, inter-rater reliability among radiologists for full-thickness tear was moderate whereas that for partial-thickness tear and tendinitis was poor to fair.<sup>30</sup> Sein et al. assessed reliability of supraspinatus tendinopathy among musculoskeletal radiologists.<sup>15</sup> The authors reported an intraclass correlation of 0.55 for inter-observer reliability on MRI. A lower inter-rater reliability for transverse as compared with longitudinal size of tear was observed in our study. Possible reasons include greater technical complexity of measuring transverse tear size and the use of different image sequences during each of the reviews.

Prior studies have also provided variable results on the reliability of assessing the preservation of rotator cuff muscle bulk and degree of fatty infiltration. Wall et al. recently assessed the diagnostic performance of ultrasonography as compared with MRI for assessment of fatty infiltration in 80 participants.<sup>31</sup> The authors also reported on inter-observer reliability of MRI among a shoulder expert, a shoulder fellow, an orthopedic resident, and a musculoskeletal radiologist. They presented a  $\kappa$  of 0.76 for the supraspinatus, 0.77 for the infraspinatus, and 0.59 for the teres minor. Values for subscapularis were not reported and other characteristics of the rotator cuff such as tear size and thickness, atrophy, and tendon retraction were not reported. Our study shows a similar trend with higher inter-observer reliability for supraspinatus and infraspinatus, and lower values for teres minor. The kappa paradox applies to the lower kappa value for teres minor in our study since few patients have higher grades of fatty infiltration. The percent agreement between the reviewers in our study was high for teres minor (83%). Wall et al. also assessed intra-observer agreement for a shoulder fellow and an orthopedic resident, and reported kappa values between 0.71 and 0.90. In another study of 31 patients, three surgeons rated supraspinatus fatty infiltration, supraspinatus and infraspinatus muscle atrophy, and tendon retraction in the frontal plane.<sup>17</sup> In this study, only tendon retraction showed moderate agreement. The kappa for fatty infiltration and atrophy of the supraspinatus were 0.41 and 0.25, respectively. Oh et al. reported kappa values of 0.60–0.75 for inter-rater reliability when grading fatty infiltration in 75 subjects with full-thickness rotator cuff tears among two musculoskeletal radiologists and three shoulder surgeons.<sup>18</sup> Other studies reported low inter-rater reliability among shoulder surgeons for fatty infiltration.<sup>19,32</sup>

A more comprehensive study was performed by Spencer et al. to assess inter-rater reliability of MRI among surgeons in 27 patients.<sup>33</sup> However, the reviewers had prior knowledge that these patients all had rotator cuff surgery for a tear and details on muscle quantity

classification were not provided. Fatty infiltration had only fair inter-rater reliability. Sagittal (transverse) and coronal (longitudinal) sizes of tear also had low kappa values of 0.42. Classification of full versus partial-thickness had substantial to almost perfect reliability.

Thus, prior studies have reported variable results on intra- and inter-rater reliability of MRI assessment of the rotator cuff. Most of these studies assessed inter-rater reliability among shoulder surgeons or among radiologists (as opposed to between shoulder specialists and radiologists), and reported on only a few, select characteristics of the rotator cuff. Our study addresses these limitations, follows a standardized and rigorous research protocol, and is the only study that assesses all of the essential MRI variables in a single analysis. The somewhat higher inter-rater reliability documented in our study as compared to many previous investigations may simply be due to chance, as our confidence intervals were wide, or may also relate to the use of strict protocolized definitions of the salient imaging findings in this study. Although the shoulder experts in our study did not receive specialized musculoskeletal radiology training in MRI review, our study was performed at an academic setting which may bias our results.

Our study has a few limitations. Although two shoulder experts performed a composite review in our study which is not typical in clinical practice, we also assessed the reliability when only one of these two shoulder experts performed the review and the results were similar. It is also possible that the time spent on reviewing each MRI was greater in our study than is typical in a busy clinical practice. In addition, our study did not address anatomic or pathologic findings that were not directly related to rotator cuff injury. These include the capsulolabral complex, glenohumeral, or acromioclavicular joints.

## CONCLUSION

In summary, our data show that MRI assessment of the rotator cuff has good inter-rater and intra-rater reliability for most variables among shoulder experts and a musculoskeletal radiologist. These data offer support for the use of MRI interpretations by expert clinicians in studies of rotator cuff disorders.

## Acknowledgements

We thank the entire ROW team (Abigail Byrne, Caitlyn McCarthy, Doris Strnad, Elana Siegel, Emily Curry, Laurel Donnell-Fink, Li Chen, Yan Dong, Peter Douglass, Alex Girden, Lindsay Miller, and Swastina Shrestha) for their efforts. We also thank our clinical staff at the Orthopedic and Arthritis Center at Brigham and Women's Hospital and the Harvard Shoulder Service at Massachusetts General Hospital for their efforts and cooperation.

**Funding:** Dr. Jain is supported by funding from National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) project number 1K23AR059199, Foundation for PM&R, and Biomedical Research Institute at Brigham and Women's Hospital. Drs. Katz and Losina are in part supported by NIAMS P60 AR 0-47782. Dr. Losina is also supported by NIAMS K24 AR 057827 and Ms. Collins is supported by T32 AR 055885.

The project described was also supported by Grant Number 8 UL1 TR000170-05, Harvard Clinical and Translational Science Center, from the National Center for Advancing Translational Science. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Center for Advancing Translational Science or the National Institutes of Health.



## References

1. CDC/NCHS. National Ambulatory Medical Care Survey: 2010 Summary Tables. 2012
2. Meislin RJ, Sperling JW, Stitik TP. Persistent shoulder pain: epidemiology, pathophysiology, and diagnosis. *Am J Orthop*. 2005; 34:5–9. [PubMed: 16450690]
3. Colvin AC, Egorova N, Harrison AK, Moskowitz A, Flatow EL. National trends in rotator cuff repair. *J Bone Joint Surg Am*. 2012; 94:227–233. [PubMed: 22298054]
4. Bartolozzi A, Andreychik D, Ahmad S. Determinants of outcome in the treatment of rotator cuff disease. *Clin Orthop Relat Res*. 1994:90–97. [PubMed: 7955708]
5. Romeo AA, Hang DW, Bach BR Jr, Shott S. Repair of full thickness rotator cuff tears. Gender, age, and other factors affecting outcome. *Clin Orthop Relat Res*. 1999:243–255. [PubMed: 10546622]
6. Castagna A, Delle Rose G, Conti M, Snyder SJ, Borroni M, Garofalo R. Predictive factors of subtle residual shoulder symptoms after transtendinous arthroscopic cuff repair: a clinical study. *Am J Sports Med*. 2009; 37:103–108. [PubMed: 19059897]
7. Meyer DC, Wieser K, Farshad M, Gerber C. Retraction of supraspinatus muscle and tendon as predictors of success of rotator cuff repair. *Am J Sports Med*. 2012; 40:2242–2247. [PubMed: 22926748]
8. Gladstone JN, Bishop JY, Lo IK, Flatow EL. Fatty infiltration and atrophy of the rotator cuff do not improve after rotator cuff repair and correlate with poor functional outcome. *Am J Sports Med*. 2007; 35:719–728. [PubMed: 17337727]
9. Vad VB, Warren RF, Altchek DW, O'Brien SJ, Rose HA, Wickiewicz TL. Negative prognostic factors in managing massive rotator cuff tears. *Clin J Sport Med*. 2002; 12:151–157. [PubMed: 12011722]
10. Goutallier D, Postel JM, Bernageau J, Lavau L, Voisin MC. Fatty muscle degeneration in cuff ruptures. Pre- and postoperative evaluation by CT scan. *Clin Orthop Relat Res*. 1994:78–83. [PubMed: 8020238]
11. Oh JH, Kim SH, Ji HM, Jo KH, Bin SW, Gong HS. Prognostic factors affecting anatomic outcome of rotator cuff repair and correlation with functional outcome. *Arthroscopy*. 2009; 25:30–39. [PubMed: 19111216]
12. Opsha O, Malik A, Baltazar R, et al. MRI of the rotator cuff and internal derangement. *Eur J Radiol*. 2008; 68:36–56. [PubMed: 18387766]
13. Recht MP, Resnick D. Magnetic resonance-imaging studies of the shoulder. Diagnosis of lesions of the rotator cuff. *J Bone Joint Surg Am*. 1993; 75:1244–1253. [PubMed: 8354686]
14. Rafii M, Firooznia H, Sherman O, et al. Rotator cuff lesions: signal patterns at MR imaging. *Radiology*. 1990; 177:817–823. [PubMed: 2243995]
15. Sein ML, Walton J, Linklater J, et al. Reliability of MRI assessment of supraspinatus tendinopathy. *Br J Sports Med*. 2007; 41:e9. [PubMed: 17289860]
16. Gomoll AH, Katz JN, Warner JJ, Millett PJ. Rotator cuff disorders: recognition and management among patients with shoulder pain. *Arthritis Rheum*. 2004; 50:3751–3761. [PubMed: 15593187]
17. Lippe J, Spang JT, Leger RR, Arciero RA, Mazzocca AD, Shea KP. Inter-rater agreement of the Goutallier, Patte, and Warner classification scores using preoperative magnetic resonance imaging in patients with rotator cuff tears. *Arthroscopy*. 2012; 28:154–159. [PubMed: 22019235]
18. Oh JH, Kim SH, Choi JA, Kim Y, Oh CH. Reliability of the grading system for fatty degeneration of rotator cuff muscles. *Clin Orthop Relat Res*. 2010; 468:1558–1564. [PubMed: 19347412]
19. Slabaugh MA, Friel NA, Karas V, Romeo AA, Verma NN, Cole BJ. Interobserver and intraobserver reliability of the Goutallier classification using magnetic resonance imaging: proposal of a simplified classification system to increase reliability. *Am J Sports Med*. 2012; 40:1728–1734. [PubMed: 22753846]
20. Warner JJ, Higgins L, Parsons IMt, Dowdy P. Diagnosis and treatment of anterosuperior rotator cuff tears. *J Shoulder Elbow Surg*. 2001; 10:37–46. [PubMed: 11182734]
21. Thomazeau H, Rolland Y, Lucas C, Duval JM, Langlais F. Atrophy of the supraspinatus belly. Assessment by MRI in 55 patients with rotator cuff pathology. *Acta Orthop Scand*. 1996; 67:264–268. [PubMed: 8686465]

22. Boileau P, Brassart N, Watkinson DJ, Carles M, Hatzidakis AM, Krishnan SG. Arthroscopic repair of full-thickness tears of the supraspinatus: does the tendon really heal? *J Bone Joint Surg Am*. 2005; 87:1229–1240. [PubMed: 15930531]
23. Fleiss JL, Cohen J, Everitt BS. Large sample standard errors of kappa and weighted kappa. *Psychological Bulletin*. 1969; 72:323–327.
24. Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*. 1960; 20:37–46.
25. Cohen J. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*. 1968; 70:213–220. [PubMed: 19673146]
26. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977; 33:159–174. [PubMed: 843571]
27. Kraemer HC. Ramifications of a population-model for kappa as a coefficient of reliability. *Psychometrika*. 1979; 44:461–472.
28. Balich SM, Sheley RC, Brown TR, Sauser DD, Quinn SF. MR imaging of the rotator cuff tendon: interobserver agreement and analysis of interpretive errors. *Radiology*. 1997; 204:191–194. [PubMed: 9205245]
29. Van Dyck P, Gielen JL, Veryser J, et al. Tears of the supraspinatus tendon: assessment with indirect magnetic resonance arthrography in 67 patients with arthroscopic correlation. *Acta Radiol*. 2009; 50:1057–1063. [PubMed: 19863417]
30. Robertson PL, Schweitzer ME, Mitchell DG, et al. Rotator cuff disorders: interobserver and intraobserver variation in diagnosis with MR imaging. *Radiology*. 1995; 194:831–835. [PubMed: 7862988]
31. Wall LB, Teefey SA, Middleton WD, et al. Diagnostic performance and reliability of ultrasonography for fatty degeneration of the rotator cuff muscles. *J Bone Joint Surg Am*. 2012; 94:e83. [PubMed: 22717835]
32. Williams MD, Ladermann A, Melis B, Barthelemy R, Walch G. Fatty infiltration of the supraspinatus: a reliability study. *J Shoulder Elbow Surg*. 2009; 18:581–587. [PubMed: 19559372]
33. Spencer EE Jr, Dunn WR, Wright RW, et al. Interobserver agreement in the classification of rotator cuff tears using magnetic resonance imaging. *Am J Sports Med*. 2008; 36:99–103. [PubMed: 17932406]

## APPENDIX A

### MRI READING FORM

**Provider Initials:** \_ \_ \_ \_ **Study ID:** \_ \_ \_ \_ \_

**G. Imaging Reviewed:**  1- MRI  2- MRA  3 - CTA  4 - CT  
 5 - X-Ray  6 - No imaging

**A. Rotator Cuff (Tear 1):** (*If there are more than two rotator cuff tears, please rate the biggest two.*)

1. Full thickness tear?  YES  NO

2. Partial thickness tear?  YES  NO

3. If partial thickness tear  Bursal  Articular (or intrasubstance)

3. aIf partial thickness tear  <50%  50-75%  >75%

4. Size (if full-thickness): a. Longitudinal \_ \_ . \_ cm b. Transverse \_ \_ . \_ cm

5. Tendonitis/tendinopathy  YES  NO

(without tear)

6. Tendon involved (tear or tendonitis/tendinopathy):

- a. Subscapularis  YES  NO  
 b. Supraspinatus  YES  NO  
 c. Infraspinatus  YES  NO  
 d. Teres Minor  YES  NO

**B. Rotator Cuff (Tear 2; if present)**

1. Full thickness tear?  YES  NO  
 2. Partial thickness tear?  YES  NO  
 3. If partial thickness tear  Bursal  Articular (or intrasubstance)  
 4. Size (if full-thickness): a. Longitudinal \_\_ \_\_. \_\_ cm b. Transverse \_\_  
 \_\_. \_\_ cm

5. Tendonitis/ tendinopathy  YES  NO

(without tear)

6. Tendon involved (tear or tendonitis/tendinopathy):

- a. Subscapularis  YES  NO  
 b. Supraspinatus  YES  NO  
 c. Infraspinatus  YES  NO  
 d. Teres Minor  YES  NO

**C. Biceps Tendon**

1. Partial Tear  YES  NO  
 2. Fluid  YES  NO  
 3. Absent  YES  NO  
 4. Subluxation  YES  NO  
 5. If medial subluxation,  Normal  Fraying  Tear  
 Subscapularis

**D. Other**

1. Labral tear  YES  NO  
 2. Ganglion Cyst  YES  NO  
 3. Glenohumeral Osteoarthritis  YES  NO  
 4. Bankart Lesion  YES  NO  
 5. Hills-Sachs Lesion  YES  NO  
 6. AC joint arthritis  YES  NO  
 7. Acromion  Type 1  Type 2  Type 3  
 Lateral downslope  Os Acromiale  
 8. Calcific Tendonitis  YES  NO

**E. Standard Evaluations**

(Please complete for all tendons)

1. Supraspinatus Goutallier (stage):

2. Infraspinatus Goutallier (stage):

- 0 - No fatty deposits  0 - No fatty deposits  
 1 - Some fatty streaks  1 - Some fatty streaks  
 2 - More muscle than fat  2 - More muscle than fat  
 3 - Equal muscle and fat  3 - Equal muscle and fat  
 4 - Less muscle than fat  4 - Less muscle than fat

3. Teres Minor Goutallier (stage): 4. Subscapularis Goutallier (stage):
- 0 - No fatty deposits  0 - No fatty deposits
- 1 - Some fatty streaks  1 - Some fatty streaks
- 2 - More muscle than fat  2 - More muscle than fat
- 3 - Equal muscle and fat  3 - Equal muscle and fat
- 4 - Less muscle than fat  4 - Less muscle than fat
5. Boileau Retraction (stage):
- Not applicable  I  II
- III  IV
6. Muscle Atrophy Grading:
- None Mild Moderate Severe
- a. Supraspinatus
- b. Infraspinatus
- c. Teres Minor
- d. Subscapularis

**APPENDIX B**

**Table 3.a**

Inter-Rater Reliability of Rotator Cuff Muscle Fatty Infiltration

		Reviewer 2					Kappa (95% CI)	Percent Agreement (95% CI)
<b>Fatty Infiltration *</b>								
<b>Reviewers 1 (Composite)</b>		<b>Supraspinatus</b>						
		Grade 0	Grade 1	Grade 2	Grade 3	Grade 4		
<b>Supraspinatus</b>	Grade 0	21	3	0	0	0	0.64 (0.31, 0.96)	0.80 (0.61, 0.92)
	Grade 1	0	2	1	0	0		
	Grade 2	1	1	0	0	0		
	Grade 3	0	0	0	0	0		
	Grade 4	0	0	0	0	1		
		<b>Infraspinatus</b>						
		Grade 0	Grade 1	Grade 2	Grade 3	Grade 4		
<b>Infraspinatus</b>	Grade 0	16	7	1	0	0	0.59 (0.28, 0.91)	0.69 (0.49, 0.85)
	Grade 1	1	1	0	0	0		
	Grade 2	0	0	1	0	0		
	Grade 3	0	0	0	1	0		
	Grade 4	0	0	0	0	1		
		<b>Teres Minor</b>						

		Reviewer 2					Kappa (95% CI)	Percent Agreement (95% CI)
		Grade 0	Grade 1	Grade 2	Grade 3	Grade 4		
<b>Teres Minor</b>	Grade 0	25	2	1	0	0	0.47 (0.00, 0.94)	0.83 (0.65, 0.94)
	Grade 1	1	0	0	0	0		
	Grade 2	0	0	0	0	0		
	Grade 3	0	0	0	0	0		
	Grade 4	0	0	0	1	0		
<b>Subscapularis</b>								
		Grade 0	Grade 1	Grade 2	Grade 3	Grade 4		
<b>Subscapularis</b>	Grade 0	25	2	0	0	0	0.72 (0.48, 0.97)	0.87 (0.69, 0.96)
	Grade 1	0	0	1	0	0		
	Grade 2	0	0	0	0	0		
	Grade 3	0	0	1	1	0		
	Grade 4	0	0	0	0	0		

\* Fatty infiltration grades as described by Goutallier et al.<sup>10</sup>

Note: There was too much motion artifact on one study to perform grading and one rater commented that there was too much edema of the infraspinatus on another study to perform grading of this muscle.

**Table 3.b**

Inter-Rater Reliability of Rotator Cuff Muscle Atrophy

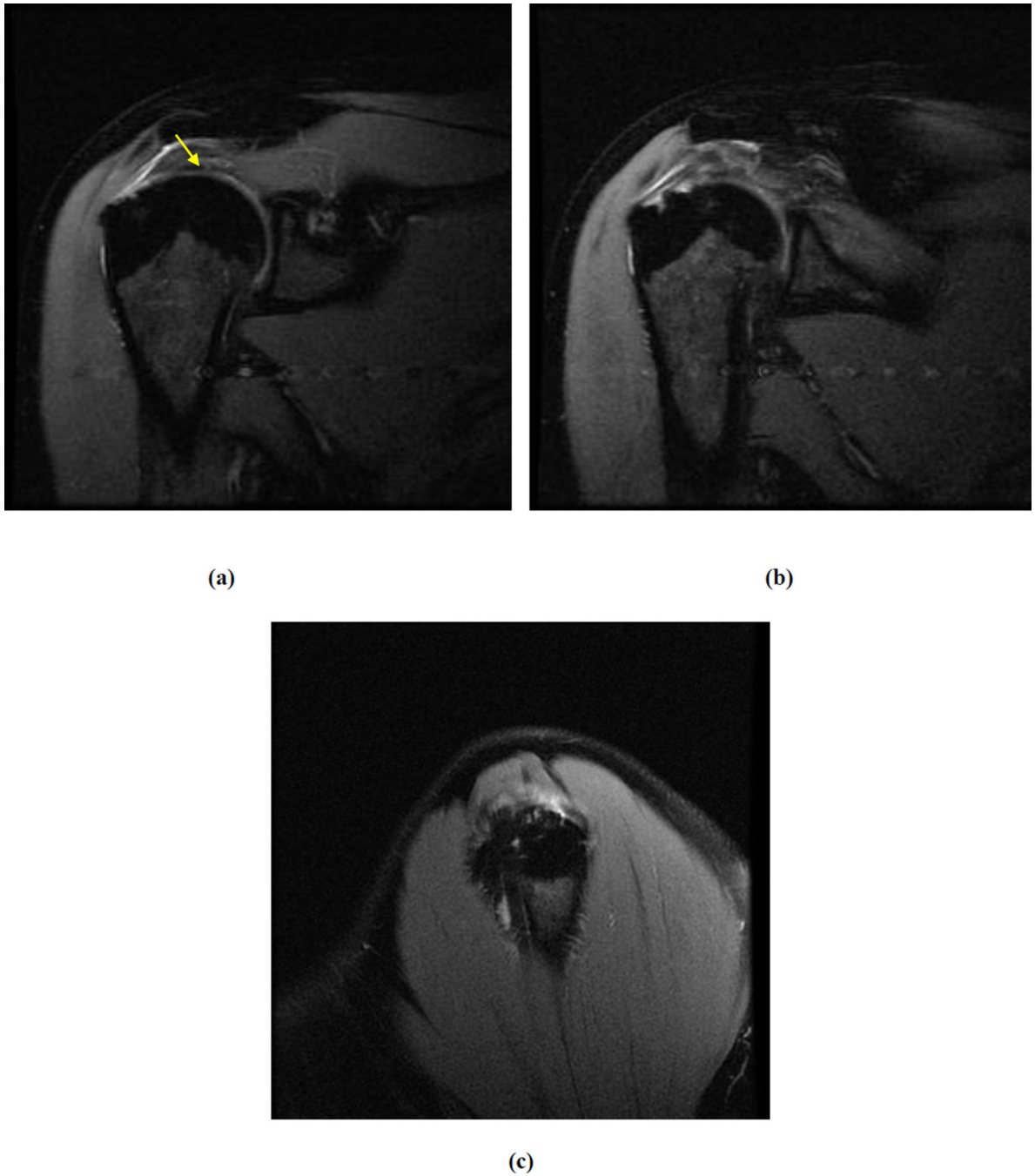
		Reviewer 2				Kappa (95% CI)	Percent Agreement (95% CI)
		None	Mild	Mod	Severe		
<b>Muscle Atrophy*</b>							
<b>Supraspinatus</b>							
<b>Supraspinatus</b>	None	24	2	0	0	0.72 (0.45, 0.98)	0.87 (0.69, 0.96)
	Mild	0	1	1	0		
	Mod	0	1	0	0		
	Severe	0	0	0	1		
<b>Infraspinatus</b>							
<b>Infraspinatus</b>	None	21	4	1	0	0.54 (0.11, 0.96)	0.79 (0.60, 0.92)
	Mild	1	0	0	0		
	Mod	0	0	1	0		
	Severe	0	0	0	1		

Reviewers 1 (Composite)

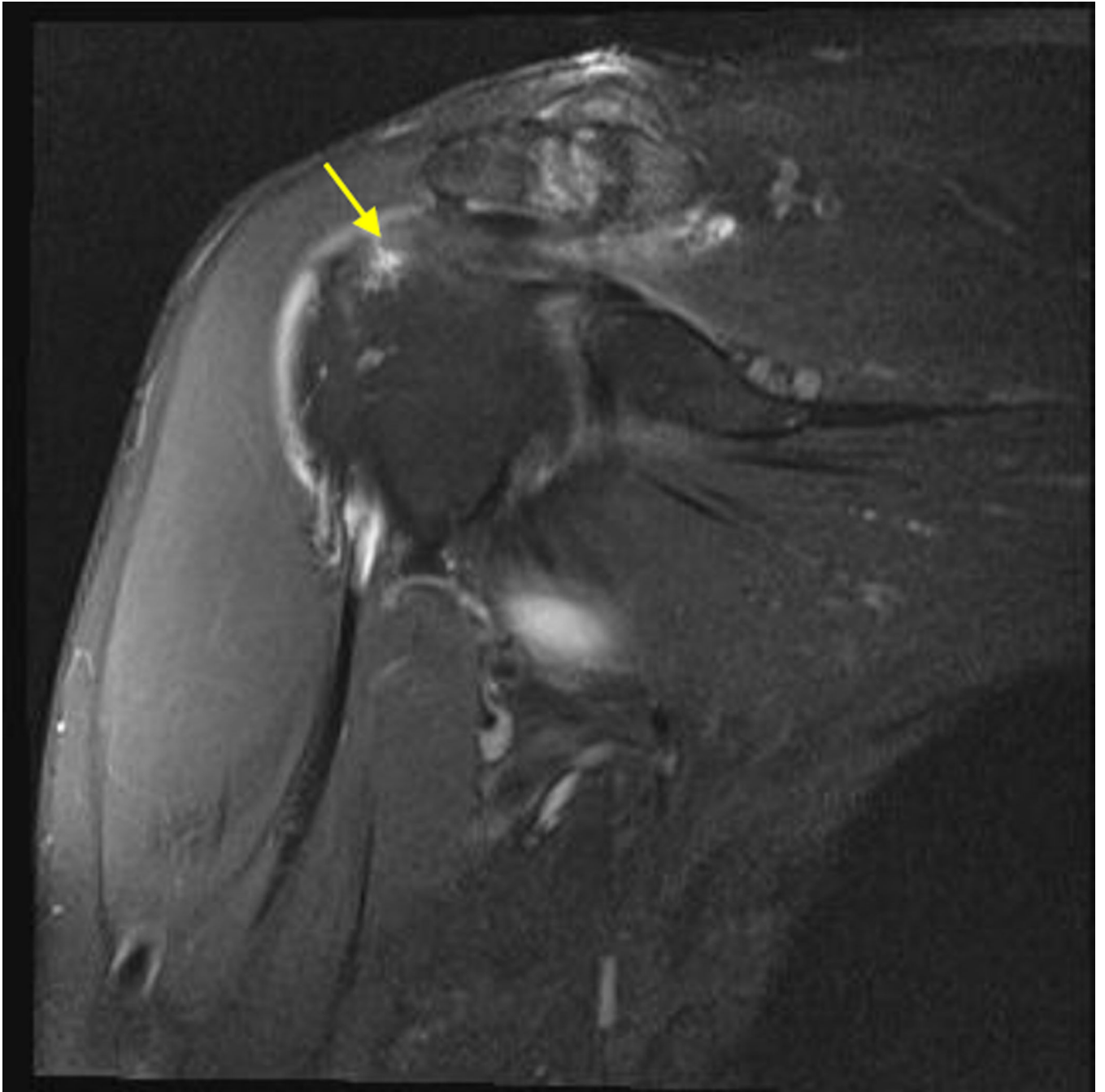
		Reviewer 2				Kappa (95% CI)	Percent Agreement (95% CI)
<b>Teres Minor</b>							
		None	Mild	Mod	Severe		
<b>Teres Minor</b>	None	28	0	0	0	1.00 (N/A)	1.00 (N/A)
	Mild	0	1	0	0		
	Mod	0	0	1	0		
	Severe	0	0	0	0		
<b>Subscapularis</b>							
		None	Mild	Mod	Severe		
<b>Subscapularis</b>	None	26	2	0	0	0.64 (0.27, 1.00)	0.90 (0.73, 0.98)
	Mild	0	0	0	0		
	Mod	0	1	1	0		
	Severe	0	0	0	0		

\* Muscle atrophy stages as described by Warner et al.<sup>20</sup>

Note: There was too much motion artifact on one study to perform grading and one rater commented that there was too much edema of the infraspinatus on another study to perform grading of this muscle.

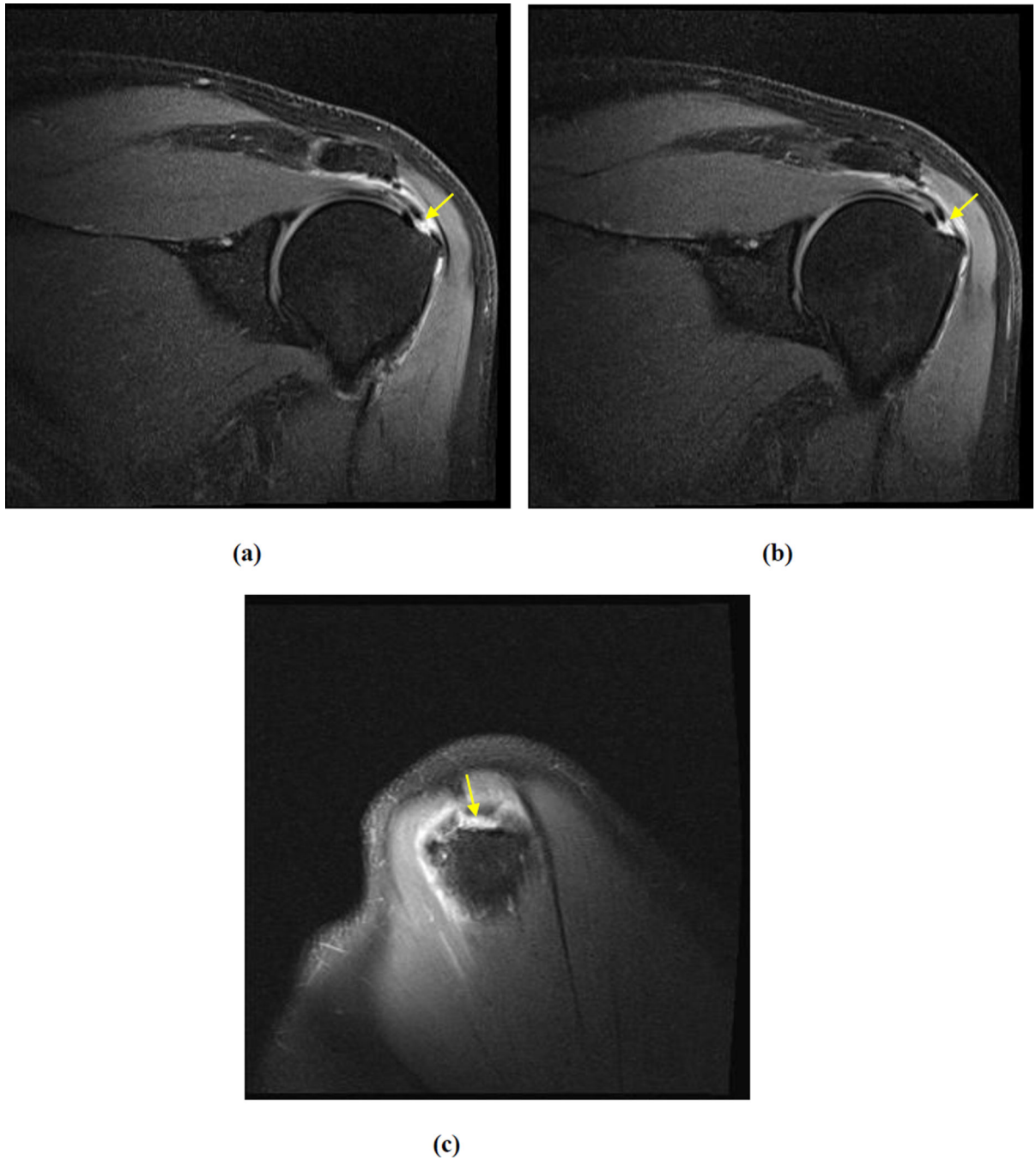


**Figure 1. Inter-Rater Discrepancy (Full-Thickness versus Partial-Thickness Tear)**  
Select Coronal Oblique STIR FSE and Sagittal Oblique Proton Density Fat Saturated images of the supraspinatus in a study read as full-thickness tear by one reviewer and partial-thickness tear (arrow) by other reviewer

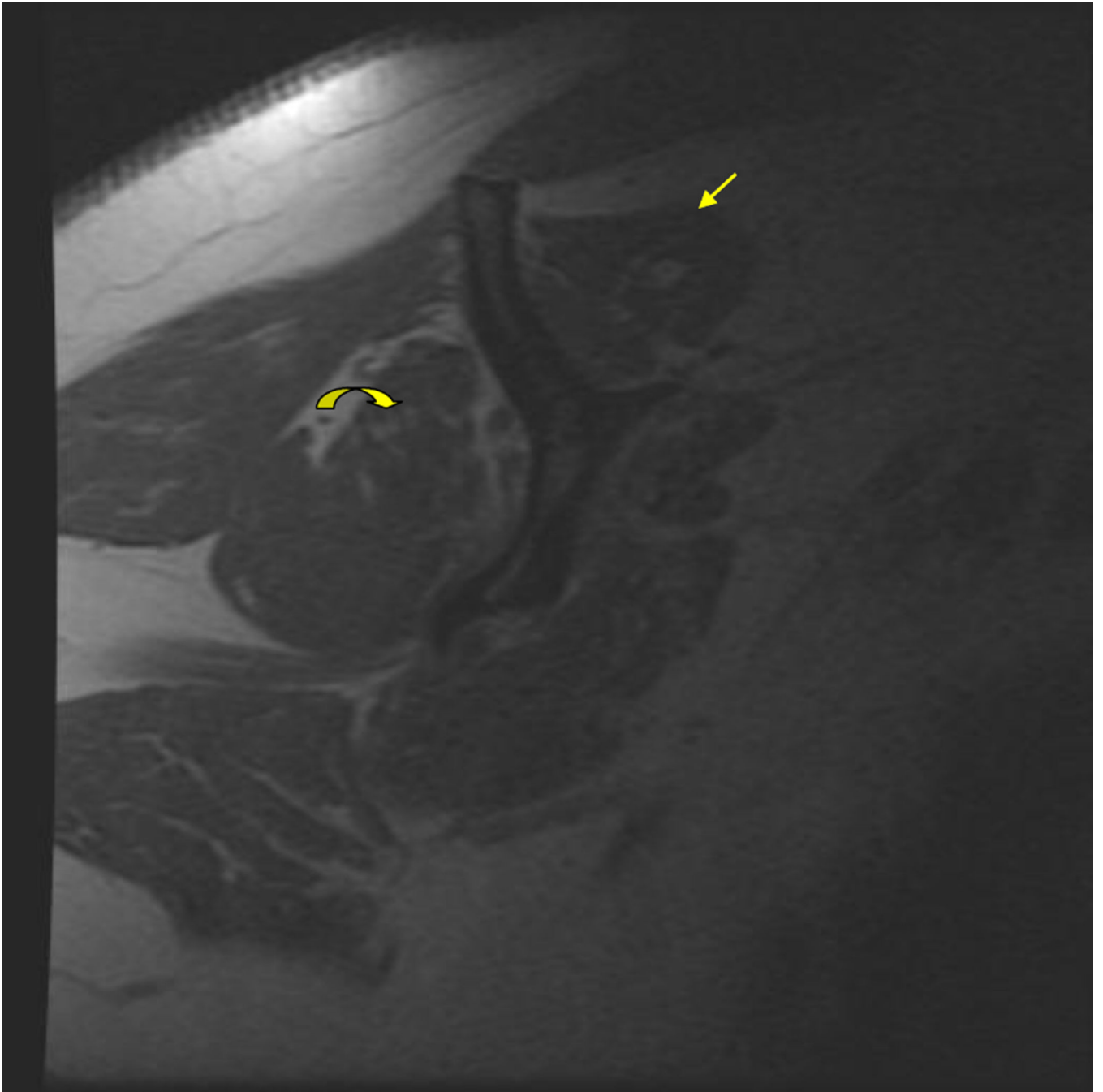


**Figure 2. Inter-Rater Discrepancy (Full-Thickness versus Partial-Thickness Tear)**  
Select Coronal Oblique Proton Density Fat Saturated image of the supraspinatus in a study read as full-thickness tear by one reviewer and partial-thickness tear (arrow) by other reviewer





**Figure 3. Inter-Rater Discrepancy (Full-Thickness versus Partial-Thickness Tear)**  
Select Coronal and Sagittal Oblique Proton Density Fat Saturated images of the supraspinatus in a study read as full-thickness tear by one reviewer and partial-thickness tear (arrow) by other reviewer



**Figure 4. Inter-Rater and Intra-Rater Agreement for Fatty Infiltration**

Select Sagittal Oblique T1-weighted images of the rotator cuff muscle in a study read as Grade I fatty infiltration for supraspinatus (arrow) and Grade 2 fatty infiltration for infraspinatus (curved arrow). This was consistent across Reviewer 1 and Reviewer 2, and during the second read by Reviewer 1

**Table I**

Baseline Characteristics of Study Participants (n=31)

Characteristics	Number (%)
Age (years)*	60.8 ± 8.7
Sex	
Female	18 (58.1%)
Male	13 (41.9%)
Race	
White (non-Hispanic)	26 (83.9%)
Other	5 (16.1%)
Laterality	
Right	20 (64.5%)
Left	11 (35.5%)
Duration of Symptoms (months)*	19.8 ± 35.5
Shoulder Visual Analog Pain Score*	48.4 ± 24.8

\* Expressed as mean ± standard deviation

**Table II**

Inter-Rater Reliability of Rotator Cuff Tear Diagnosis and Characteristics

	Reviewer 2		Reviewer 1 (Composite)		Kappa Agreement (95% CI)	Percent Agreement (95% CI)
Tear	No	Yes	Tear		0.90 (0.72, 1.00)	0.96 (0.83, 0.99)
	6	0	No	Yes		
Yes	1	24	Tear Thickness			
			Full	Partial	Tendinopathy	Normal
Tear Thickness	Full	1	0	0		
	Partial	2	6	0	1	0.68 (0.54, 0.82)
	Tendinopathy	0	0	1	2	0.70 (0.51, 0.85)
	Normal	0	0	3	0	
Tendon Involved	Tendon Involved ≠		Subscapularis	Supraspinatus	Infraspinatus	Teres Minor
	Subscapularis	1	0	0	0	0
	Supraspinatus	0	23	0	0	0
	Infraspinatus	0	0	0	0	0
	Teres Minor	0	0	0	0	0
Tear Size	Tear Size* (Longitudinal)		Tear Size		0.75 (0.44, 1.00)	0.86 (0.59, 0.98)
	<1cm	1-3cm	>3cm			
Tear Size	<1 cm	1	0	0		
	1-3 cm	1	9	0		

	Reviewer 2			Kappa (95% CI)	Percent Agreement (95% CI)
>3cm	0	1	3		
<b>Tear Size * (Transverse)</b>					
	<1cm	1-3cm	>3		
<b>Tear Size</b>					
<1 cm	0	6	0	0.20 (0.0, 0.51)	0.53 (0.26, 0.78)
1-3 cm	0	7	1		
>3cm	0	0	1		
<b>Tendon Retraction**</b>					
	N/A	Stage I	Stage II	Stage III	Stage IV
N/A	14	2	0	0	0
<b>Tendon Retraction</b>					
I	1	3	1	0	0
II	0	2	3	0	0
III	0	0	2	1	0
IV	0	0	0	1	1
				0.77 (0.64, 0.90)	0.70 (0.51, 0.85)

# n=24 since only includes patients with a full- or partial-thickness rotator cuff tear

\* n=15 since only includes patients with a full-thickness rotator cuff tear

\*\* Stages as described by Boileau et al.<sup>22</sup>

**Table III**

**Inter-Rater Reliability of Rotator Cuff Muscle Atrophy and Fatty Infiltration**

	<b>Reviewer 1</b>				<b>Reviewer 2</b>				<b>Kappa</b> (95% CI)	<b>Percent Agreement</b> (95% CI)
<b>Fatty Infiltration *</b>										
	Grade 0	Grade 1	Grade 2	Grade 3	Grade 4	Grade 0	Grade 1	Grade 2	Grade 3	Grade 4
Grade 0	87	14	2	0	0	87	14	2	0	0
Grade 1	2	3	2	0	0	2	3	2	0	0
Grade 2	1	1	1	0	0	1	1	1	0	0
Grade 3	0	0	1	2	0	0	0	1	2	0
Grade 4	0	0	0	1	2	0	0	0	1	2
<b>Reviewers 1 (Composite)</b>										
<b>Muscle Atrophy **</b>										
	None	Mild	Moderate	Severe	None	Mild	Moderate	Severe		
None	99	8	1	0	99	8	1	0	0.68	0.89
Mild	1	2	1	0	1	2	1	0	(0.50, 0.86)	(0.82, 0.94)
Moderate	0	2	3	0	0	2	3	0		
Severe	0	0	0	2	0	0	0	2		

Aggregate statistics for all four rotator cuff tendons are presented

Note: There was too much motion artifact on one study to perform grading and one rater commented that there was too much edema of the infraspinatus on another study to perform grading of this muscle.

\* Fatty infiltration grades as described by Goutallier et al.<sup>10</sup>

\*\* Muscle atrophy stages as described by Warner et al.<sup>20</sup>

**Table IV**

Intra-Rater Reliability\*\* of Select Supraspinatus Tear and Muscle Characteristics

		Reviewers 1 (Second Read)		Kappa (95% CI)	Percent Agreement (95% CI)	
<b>Tear Thickness</b>						
		Full	No Tear			
Full		15	1	0.88 (0.72, 1.00)	0.94 (0.79, 0.99)	
Partial		1	0			
No Tear		0	6			
<b>Tear Size* (Longitudinal)</b>						
		<1 cm	>3cm			
<1 cm		1	0	0.61 (0.22, 0.99)	0.80 (0.51, 0.95)	
1-3 cm		0	0			
>3cm		0	2			
<b>Reviewers 1 (First Read)</b>						
		<1 cm	>3 cm			
<1 cm		1	0	0.21 (0.0, 0.57)	0.66 (0.38, 0.88)	
1-3 cm		0	0			
>3cm		0	0			
<b>Tendon Retraction</b>						
		N/A	Stage I	Stage II	Stage III	Stage IV
N/A		15	1	0	0	0
Stage I		0	4	1	0	0
Stage II		0	0	4	1	0
Stage III		0	0	1	2	0
					0.88 (0.78, 0.97)	0.83 (0.66, 0.94)

Stage IV	Reviewers 1 (Second Read)				Kappa (95% CI)	Percent Agreement (95% CI)
	0	0	1	1		
<b>Fatty Infiltration<sup>±</sup></b>						
	Grade 0	Grade 1	Grade 2	Grade 3	Grade 4	
Grade 0	101	3	0	0	0	
Grade 1	2	3	2	0	0	
Grade 2	0	0	3	0	0	0.89 (0.80, 0.98)
Grade 3	0	0	0	3	0	0.94 (0.88, 0.98)
Grade 4	0	0	0	0	3	
<b>Muscle Atrophy<sup>¶</sup></b>						
	None	Mild	Moderate	Severe		
None	108	1	0	0	0.87 (0.76, 0.98)	0.96 (0.91, 0.99)
Mild	2	1	1	0		
Moderate	0	0	4	1		
Severe	0	0	0	2		

Note: There was too much motion artifact on one study to perform fatty infiltration and atrophy grading. Aggregate statistics for all four rotator cuff tendons are presented.

\* n=15 since only includes patients with a full-thickness rotator cuff tear

\*\* for composite reviewer 1

<sup>±</sup>Fatty infiltration grades as described by Goutallier et al.<sup>10</sup>

<sup>¶</sup>Muscle atrophy stages as described by Warner et al.<sup>20</sup>



**Table V**

Reliability of Rotator Cuff Assessment on MRI by One Shoulder Specialist

	One Shoulder Specialist versus Reviewers 1 <sup>±</sup>		One Shoulder Specialist versus Reviewer 2 <sup>±</sup>	
	Kappa (95% CI)	Percent Agreement (95% CI)	Kappa (95% CI)	Percent Agreement (95% CI)
Supraspinatus Tear Thickness	0.86 (0.70, 1.00)	0.90 (0.74, 0.98)	0.69 (0.47, 0.91)	0.77 (0.59, 0.90)
Supraspinatus Tear Size (Longitudinal)	0.79 (0.52, 1.00)	0.87 (0.60, 0.98)	0.75 (0.44, 1.00)	0.86 (0.57, 0.98)
Supraspinatus Tear Size (Transverse)	0.29 (0.00, 0.72)	0.67 (0.38, 0.88)	0.00 (0.00, 0.00)	0.71 (0.42, 0.92)
Supraspinatus Tendon Retraction	0.85 (0.72, 0.98)	0.84 (0.66, 0.95)	0.76 (0.63, 0.90)	0.71 (0.52, 0.86)
Fatty Infiltration *	0.73 (0.53, 0.93)	0.93 (0.89, 0.98)	0.50 (0.30, 0.70)	0.79 (0.72, 0.86)
Muscle Atrophy **	0.81 (0.67, 0.95)	0.94 (0.90, 0.98)	0.64 (0.43, 0.85)	0.90 (0.85, 0.95)

Aggregate statistics for all four rotator cuff tendons for fatty infiltration and muscle atrophy are presented

<sup>±</sup> Reviewers 1 represent composite assessment by two shoulder specialists whereas Reviewer 2 represents a musculoskeletal radiologist

\* Fatty infiltration grades as described by Goutallier et al.<sup>10</sup>

\*\* Muscle atrophy stages as described by Warner et al.<sup>20</sup>