

# Diagnostic Performance and Reliability of Ultrasonography for Fatty Degeneration of the Rotator Cuff Muscles

Lindley B. Wall, MD, Sharlene A. Teefey, MD, William D. Middleton, MD, Nirvikar Dahiya, MD, Karen Steger-May, MA, H. Mike Kim, MD, Daniel Wessell, MD, PhD, and Ken Yamaguchi, MD

*Investigation performed at the Department of Orthopaedic Surgery, the Mallinckrodt Institute of Radiology, and the Division of Biostatistics, Washington University School of Medicine, St. Louis, Missouri*

**Background:** Diagnostic evaluation of rotator cuff muscle quality is important to determine indications for potential operative repair. Ultrasonography has developed into an accepted and useful tool for evaluating rotator cuff tendon tears; however, its use for evaluating rotator muscle quality has not been well established. The purpose of this study was to investigate the diagnostic performance and observer reliability of ultrasonography in grading fatty degeneration of the posterior and superior rotator cuff muscles.

**Methods:** The supraspinatus, infraspinatus, and teres minor muscles were prospectively evaluated with magnetic resonance imaging (MRI) and ultrasonography in eighty patients with shoulder pain. The degree of fatty degeneration on MRI was graded by four independent raters on the basis of the modified Goutallier grading system. Ultrasonographic evaluation of fatty degeneration was performed by one of three radiologists with use of a three-point scale. The two scoring systems were compared to determine the diagnostic performance of ultrasonography. The interobserver and intraobserver reliability of MRI grading by the four raters were determined. The interobserver reliability of ultrasonography among the three radiologists was determined in a separate group of thirty study subjects. The weighted Cohen kappa, percentage agreement, sensitivity, and specificity were calculated.

**Results:** The accuracy of ultrasonography for the detection of fatty degeneration, as assessed on the basis of the percentage agreement with MRI, was 92.5% for the supraspinatus and infraspinatus muscles and 87.5% for the teres minor. The sensitivity was 84.6% for the supraspinatus, 95.6% for the infraspinatus, and 87.5% for the teres minor. The specificity was 96.3% for the supraspinatus, 91.2% for the infraspinatus, and 87.5% for the teres minor. The agreement between MRI and ultrasonography was substantial for the supraspinatus and infraspinatus (kappa = 0.78 and 0.71, respectively) and moderate for the teres minor (kappa = 0.47). The interobserver reliability for MRI was substantial for the supraspinatus and infraspinatus (kappa = 0.76 and 0.77, respectively) and moderate for the teres minor (kappa = 0.59). For ultrasonography, the interobserver reliability was substantial for all three muscles (kappa = 0.71 for the supraspinatus, 0.65 for the infraspinatus, and 0.72 for the teres minor).

**Conclusions:** The diagnostic performance of ultrasonography in identifying and grading fatty degeneration of the rotator cuff muscles was comparable with that of MRI. Ultrasonography can be used as the primary diagnostic imaging modality for fatty changes in rotator cuff muscles.

**Level of Evidence:** Diagnostic Level II. See Instructions for Authors for a complete description of levels of evidence.

Fatty degeneration is a detrimental change in the muscles of the rotator cuff and is a negative prognostic factor in rotator cuff surgical reconstruction<sup>1-4</sup>. Knowledge of the

presence and extent of fatty changes of the rotator cuff muscles is useful clinical information to guide the treatment options for individuals affected by rotator cuff tears. Magnetic resonance

**Disclosure:** One or more of the authors received payments or services, either directly or indirectly (i.e., via his or her institution), from a third party in support of an aspect of this work. In addition, one or more of the authors, or his or her institution, has had a financial relationship, in the thirty-six months prior to submission of this work, with an entity in the biomedical arena that could be perceived to influence or have the potential to influence what is written in this work. No author has had any other relationships, or has engaged in any other activities, that could be perceived to influence or have the potential to influence what is written in this work. The complete **Disclosures of Potential Conflicts of Interest** submitted by authors are always provided with the online version of the article.

imaging (MRI) has become the standard imaging modality for identifying and quantifying the amount of fatty degeneration of the rotator cuff musculature. Ultrasonography has been used for many years in the evaluation of rotator cuff tendon tears and has an accuracy for this purpose that is comparable with that of MRI<sup>5-7</sup>. Ultrasonography can also be used to identify fatty degeneration; however, the associated diagnostic performance and reliability have not yet been determined for the three posterior rotator cuff muscles. The purpose of this study was to investigate the diagnostic performance and reliability of ultrasonography in detecting and grading fatty degeneration of the supraspinatus, infraspinatus, and teres minor muscles, using MRI as the reference standard. We hypothesized that ultrasonography would demonstrate diagnostic performance and reliability in the detection and grading of fatty degeneration that were comparable with those of MRI.

## Materials and Methods

### Study Subjects

Institutional review board approval was obtained for the study prior to patient recruitment. Eighty-three patients were prospectively enrolled from the shoulder and elbow or sports clinics at our institution. An initial sample size of 100 was chosen on the basis of an a priori power analysis that assumed the performance of an equivalence test. However, because of a refinement of the study goals, this statistical test was not ultimately used in the analysis. Recruitment was stopped when the final number of eighty-three patients was reached because of logistical barriers that included substantial recruitment difficulties. All patients who were approached for the study had initially presented with a painful shoulder and were suspected clinically to have pathology of the rotator cuff tendons. Patients who had previously had an MRI of the shoulder performed at our institution or were scheduled for an MRI for clinical purposes were eligible for the study if they were also scheduled for a sonogram of the same shoulder, either for this study or for clinical purposes. Demographic information, including sex and age, and information on the status of the rotator cuff were recorded. The rotator cuff status was categorized on the basis of the ultrasonography as a full-thickness tear, a partial-thickness tear, or no tear. Exclusion criteria included (1) a neuromuscular disorder, (2) an MRI that was of poor quality or did not extend medial to the spinoglenoid notch, (3) an MRI that was acquired at another institution, (4) a sonogram and an MRI that were obtained more than three months apart, and (5) metallic implants from prior shoulder surgery. Consent was obtained from all patients in the clinic setting or over the telephone with use of a detailed telephone script.

Three of the eighty-three patients who had been prospectively enrolled into the study were excluded because of inadequate MRI sequences, and the remaining eighty patients (thirty-four female and forty-six male) were included in the study. The mean age at the time of the sonogram was fifty-four years (range, eighteen to seventy-seven years). Ultrasonography detected an identifiable rotator cuff tear in fifty-eight patients; thirty-eight patients had a full-thickness tear, seventeen patients had a partial-thickness tear, and it was not possible to determine whether the tear was full or partial-thickness in three patients.

The final diagnoses were made on the basis of MRI, ultrasonography, radiographs, and physical examination. Eight patients had a subscapularis tear in conjunction with a posterior rotator cuff tear. Although all patients were initially thought to have primary rotator cuff tears, some patients without rotator cuff tears were identified as having different or related diagnoses: thirteen had rotator cuff tendinopathy or tendinitis, two had tendinitis and a labral tear, one had a frozen shoulder, two had degenerative changes of the glenohumeral joint, one had a greater tuberosity contusion, one had biceps tendinitis, one had pectoralis major tendinitis, and one did not have any identifiable pathology. Two patients had both shoulders evaluated by MRI; in

order to maintain statistical independence, the data from one side were randomly chosen to be eliminated from the analysis.

The group of subjects examined to determine the interobserver reliability of ultrasonography had a mean age of sixty-nine years (range, forty to eighty-five years); twenty-two were male and eight were female.

### MRI

The MRI examinations were performed with use of a standardized protocol for the evaluation of rotator cuff pathology. All patients were placed in the supine position with the arm at the side of the body in as much external rotation as was comfortably tolerated. A dedicated shoulder coil was positioned over the patient's shoulder. The standard MRI sequences included axial spin-echo T1, axial fast spin-echo T2 with fat saturation, oblique coronal spin-echo T1, oblique coronal fast spin-echo T2 with fat saturation, oblique sagittal spin-echo T1, and oblique sagittal fast spin-echo T2 with fat saturation. The oblique sagittal views were extended at least 1 cm medial to the spinoglenoid notch in order to include sections showing the muscle bulks in the supraspinatus and infraspinatus fossae.

The images were graded, in a blinded fashion, by three orthopaedic surgeons (raters 1, 2, and 3) and one musculoskeletal radiologist (rater 4). Rater 1 was an attending physician with shoulder and elbow fellowship training, rater 2 was a fellow completing shoulder and elbow training, and rater 3 was an orthopaedic resident. The variety of raters was chosen to represent the spectrum of individuals who use this system in clinical practice. A single T1-weighted oblique sagittal section within 1 cm medial to the spinoglenoid notch was chosen for the grading on the basis of previous literature<sup>8-11</sup>. This image was copied from each patient's MRI examination and placed into a PowerPoint (Microsoft, Redmond, Washington) slide. Patient names were removed from all images, and the images were then numbered sequentially from one to eighty-three. The file preparation was performed by rater 3, who was blinded to the report of the radiology examination. The file was then sent by e-mail to each of the raters, who graded the images independently.

The amount of fatty degeneration was graded according to the modified Goutallier five-point grading scale<sup>8,12</sup>; grade 0 = no fatty deposits, grade 1 = some fatty streaks, grade 2 = less fat than muscle, grade 3 = as much fat as muscle, and grade 4 = more fat than muscle (Figs. 1-A through 1-E). Additionally, a direct comparison with the three-point ultrasonography grading scale was performed by collapsing the five-point MRI grading scale to a three-point scale (i.e., Goutallier grades 0 and 1 were converted to grade 0 on the three-point scale, Goutallier grade 2 became grade 1, and Goutallier grades 3 and 4 became grade 2).

The interobserver reliability among all four raters was determined with use of the kappa statistic for multiple observers<sup>13</sup>. Intraobserver reliability was also determined for raters 2 and 3 by repeat grading of the magnetic resonance images. The images were reordered into a different PowerPoint file by an independent party and presented to these raters in a mixed, blinded fashion more than two weeks after the initial grading. The final MRI grade of each muscle was obtained by calculating the mean grade of the four raters and rounding to the nearest integer number.

### Ultrasonography

Shoulder ultrasonography was performed in a standardized manner, as previously described<sup>14</sup>, by one of three radiologists (S.A.T., W.D.M., and N.D.) with extensive experience in musculoskeletal ultrasonography. The radiologist who performed the sonogram was blinded to the patient's MRI results.

All ultrasonographic examinations were performed with an Elegra (Siemens Medical Systems, Issaquah, Washington), Antares (Siemens), iU22 (Philips, Bothell, Washington), or E9 (GE, Milwaukee, Wisconsin) scanner and a variable high-frequency linear-array transducer (7.5 to 15 MHz). The biceps, subscapularis, supraspinatus, infraspinatus, and teres minor tendons were examined as previously described<sup>14,15</sup>. To evaluate for fatty degeneration, the echogenicity and architecture of each muscle were examined with use of a three-point scale, which was modified from the scale previously described by Strobel et al.<sup>10</sup> (Table I). The echogenicity of the supraspinatus was determined in comparison with the



Fig. 1-A



Fig. 1-B



Fig. 1-C

**Fig. 1-A** T1-weighted MRI showing Goutallier grade-0 fatty degeneration (no fatty deposits) of the supraspinatus muscle (arrow), as graded by all three raters. **Fig. 1-B** T1-weighted MRI showing Goutallier grade-1 fatty degeneration (some fatty streaks) of the supraspinatus muscle (arrow). **Fig. 1-C** T1-weighted MRI showing Goutallier grade-2 fatty degeneration (less fat than muscle) of the supraspinatus muscle (arrow).

echogenicity of the overlying trapezius. The echogenicity of the infraspinatus and that of the teres minor were determined in comparison with the overlying deltoid. The architecture was determined on the basis of the visibility of the intramuscular tendons and of the normal muscle pennate pattern. When necessary, the contralateral muscles were scanned to detect subtle asymmetry. The mean of

the grades for echogenicity and architecture was calculated to determine a single grade (0 to 2) for the extent of fatty degeneration of each rotator cuff muscle. Figure 2-A provides an example of a normal infraspinatus muscle, Figure 2-B represents muscle with grade-1 fatty degeneration, and Figure 2-C represents muscle with grade-2 fatty degeneration.

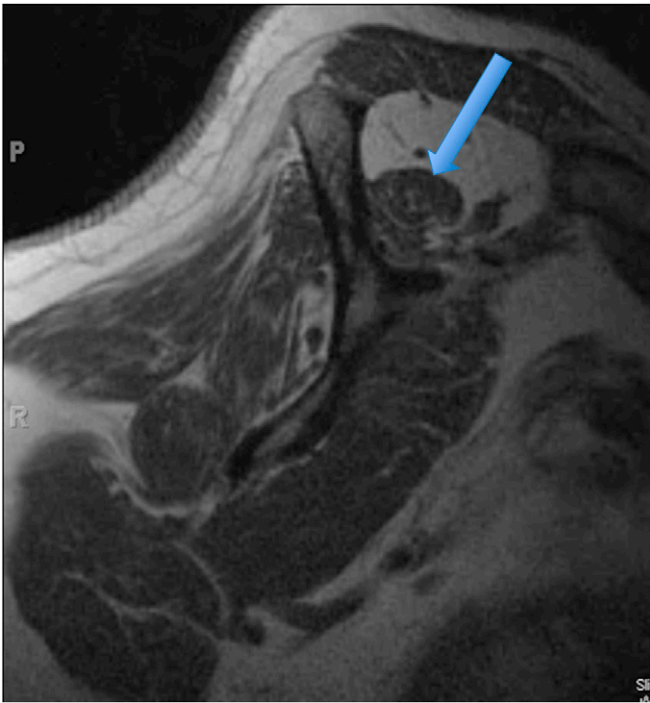


Fig. 1-D

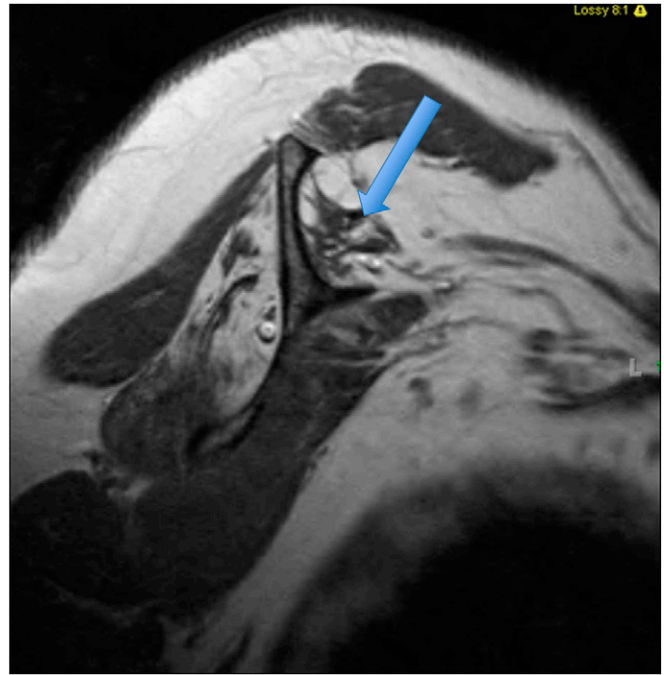


Fig. 1-E

**Fig. 1-D** T1-weighted MRI showing Goutallier grade-3 fatty degeneration (as much fat as muscle) of the supraspinatus muscle (arrow). **Fig. 1-E** T1-weighted MRI showing Goutallier grade-4 fatty degeneration (more fat than muscle) of the supraspinatus muscle (arrow).

The interobserver reliability of ultrasonography in the grading of fatty degeneration was determined by randomized blinded examination of a separate group of thirty individuals (sixty shoulders). Twenty-six of these thirty individuals were participants in a different study in which the natural history of rotator cuff tears in 196 individuals with a symptomatic rotator cuff tear in one shoulder and an asymptomatic cuff tear in the contralateral shoulder were studied. These twenty-six individuals were selected by a person not involved in the reliability portion of the present study to represent a broad spectrum of degeneration of the rotator cuff muscles. The other four participants were volunteers with asymptomatic shoulders. The status of the rotator cuff tendons and musculature was not known to the radiologist prior to the examination. The radiologist scanned the patients in a random order and was blinded to the results from the other two radiologists. The side examined first was also randomized. The supraspinatus, infraspinatus, and teres minor muscles in the sixty shoulders were evaluated by all three radiologists.

### Statistical Methods

The agreement between ultrasonography and MRI for the detection and grading of the degree of fatty degeneration was determined with use of two different grading scales. First, the three-point ultrasonography grading scale was compared with the MRI grading scale that had been collapsed from a five-point to a three-point scale on which grade 0 = Goutallier grades 0 and 1, grade 1 = Goutallier grade 2, and grade 2 = Goutallier grades 3 and 4. This division of MRI grades has been used in other studies<sup>8,16</sup>. It is also based on the clinical experience of the authors, which revealed that clinically relevant fatty degeneration is greater than Goutallier grade 1 and that grade-3 and grade-4 degeneration are treated similarly. Agreement between the two modalities for grading the degree of fatty degeneration was assessed with use of the weighted Cohen kappa ( $\kappa$ ) coefficient. The weighted kappa coefficient represents the fraction of agreement beyond that expected by chance, and it accounts for the magnitude of the disagreement between grades. Second, both MRI and ultrasonography grades were collapsed to a dichotomous scale (i.e., absence or presence of fatty infiltration) to investigate the agreement on the presence of

fatty degeneration. Goutallier MRI grades 0 and 1 and ultrasonography grade 0 became “absence.” Goutallier grades 2, 3, and 4 and ultrasonography grades 1 and 2 became “presence.”

The ultrasonography and MRI classifications were tabulated, and agreement between modalities was assessed on the basis of the percentage agreement (i.e., accuracy), agreement adjusted for the agreement expected by chance (i.e.,  $\kappa$ ), sensitivity, specificity, negative predictive value, and positive predictive value, using MRI as the reference criterion. Interobserver and intraobserver reliability were assessed with use of the weighted Cohen kappa coefficient. The kappa values were interpreted with use of the guidelines suggested by Landis and Koch<sup>17</sup>: 0 = poor, 0 to 0.20 = slight, 0.21 to 0.40 = fair, 0.41 to 0.60 = moderate, 0.61 to 0.80 = substantial and 0.81 to 1.0 = almost perfect agreement. Although the adequacy of agreement should be interpreted on the basis of the gravity of the context-specific consequences of errors, these divisions provide useful benchmarks. The data are presented as the estimate and the accompanying 95% confidence interval (CI). The data analysis was performed with use of published statistical software.

### Source of Funding

Funding for this study was received from the National Institutes of Health (grant R01 AR051026-01A1). This funding was used to perform approximately thirty of the shoulder ultrasonography examinations and to support the statistical assistance needed for the study.

### Results

#### Agreement Between Ultrasonography and MRI in Detection and Grading of Fatty Degeneration

The comparisons between the MRI and ultrasonographic muscle grading are summarized in Tables II, III, and IV. For the supraspinatus muscle, the agreement between the three-point ultrasonography and MRI scales (Table III) was  $\kappa = 0.78$  (95% CI, 0.65 to 0.90). When the dichotomous scales were used

**TABLE I The Three-Point Ultrasound Grading Scale for Rotator Cuff Muscle Fatty Degeneration\***

Grade	Echogenicity†	Architecture
0	Isoechoic to the overlying muscle	Clearly visible intramuscular tendons and identifiable muscle pennate pattern
1	Slightly increased echogenicity compared with the overlying muscle	Partially visible intramuscular tendons and muscle pennate pattern
2	Markedly increased echogenicity compared with the overlying muscle	No discernible intramuscular tendons or muscle pennate pattern

\*Modified from Strobel et al.<sup>10</sup>. †The trapezius was used as a reference for determining the echogenicity of the supraspinatus, and the deltoid was used for determining the echogenicity of the infraspinatus and teres minor.

(Table IV), the agreement was  $\kappa = 0.83$  (95% CI, 0.69 to 0.96). The percentage agreement for the dichotomous scales was 92.5%, the sensitivity was 84.6% (95% CI, 65.1% to 95.6%), the specificity was 96.3% (95% CI, 87.2% to 99.6%), the positive predictive value was 91.7% (95% CI, 73.0% to 99.0%),

and the negative predictive value was 92.9% (95% CI, 82.7% to 98.0%).

For the infraspinatus muscle, the agreement between the three-point ultrasonography and MRI scales was  $\kappa = 0.71$  (95% CI, 0.59 to 0.83) (Table III). With the dichotomous scales, the

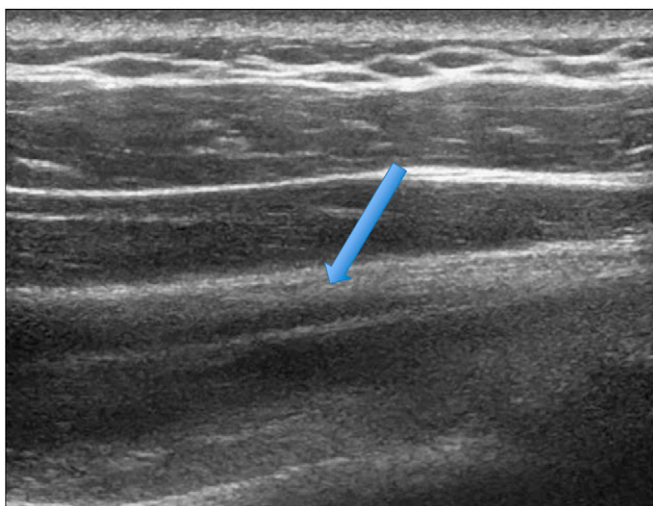


Fig. 2-A

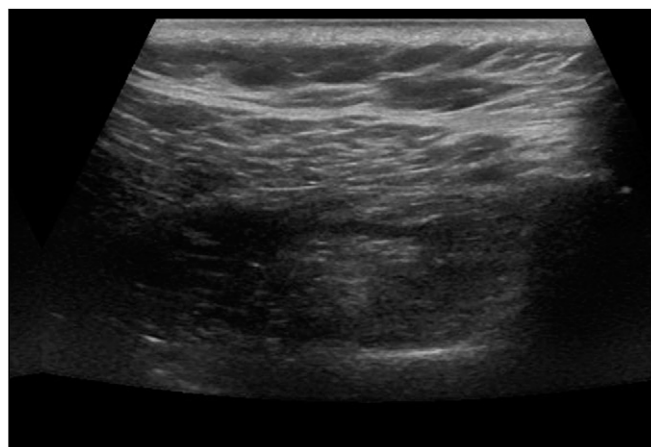


Fig. 2-B



Fig. 2-C

**Fig. 2-A** Long-axis ultrasonographic view showing a grade-0 (normal) infraspinatus muscle (arrow). Note the well-defined central tendon. **Fig. 2-B** Ultrasonographic view showing a grade-1 infraspinatus muscle (moderate fatty degeneration). The central tendon and muscle fibers are less clearly distinguished than in Figure 2-A, and the muscle reveals increased echogenicity. **Fig. 2-C** Ultrasonographic view showing a grade-2 infraspinatus muscle (severe fatty degeneration). The central tendon and muscle fibers seen in Figure 2-A are no longer visible.

**TABLE II Classification with the Three-Point Ultrasound and Five-Point MRI Grading Systems**

Muscle	Ultrasound Grade	MRI Grade (Goutallier)					Total
		0	1	2	3	4	
Supraspinatus	0	38	14	3	1	0	56
	1	0	2	9	2	2	15
	2	0	0	1	3	5	9
	Total	38	16	13	6	7	80
Infraspinatus	0	34	18	0	1	0	53
	1	0	5	6	4	3	18
	2	0	0	2	1	6	9
	Total	34	23	8	6	9	80
Teres minor	0	51	12	1	0	0	64
	1	0	8	3	2	1	14
	2	1	0	0	0	1	2
	Total	52	20	4	2	2	80

agreement was  $\kappa = 0.83$  (95% CI, 0.69 to 0.96) (Table IV). The percentage agreement for the dichotomous scales was 92.5%, the sensitivity was 95.6% (95% CI, 78.0% to 99.9%), the specificity was 91.2% (95% CI, 80.7% to 97.1%), the positive predictive value was 81.5% (95% CI, 61.9% to 93.7%), and the negative predictive value was 98.1% (95% CI, 89.9% to 99.9%).

For the teres minor muscle, the agreement between the three-point ultrasonography and MRI scales was  $\kappa = 0.47$  (95%

CI, 0.25 to 0.70) (Table III). With the dichotomous scales (Table IV), the agreement was  $\kappa = 0.52$  (95% CI, 0.27 to 0.77). The percentage agreement for the dichotomous scales was 87.5%, the sensitivity was 87.5% (95% CI, 47.4% to 99.7%), the specificity was 87.5% (95% CI, 77.6% to 94.1%), the positive predictive value was 43.8% (95% CI, 19.8% to 70.1%), and the negative predictive value was 98.4% (95% CI, 91.6% to 99.9%).

**TABLE III Classification with the Three-Point Ultrasound and Three-Point MRI Grading Systems**

Muscle	Ultrasound Grade	MRI Grade			Total
		0 (Goutallier 0, 1)	1 (Goutallier 2)	2 (Goutallier 3, 4)	
Supraspinatus	0	52	3	1	56
	1	2	9	4	15
	2	0	1	8	9
	Total	54	13	13	80
Infraspinatus	0	52	0	1	53
	1	5	6	7	18
	2	0	2	7	9
	Total	57	8	15	80
Teres minor	0	63	1	0	64
	1	8	3	3	14
	2	1	0	1	2
	Total	72	4	4	80

TABLE IV Classification with the Dichotomous Ultrasound and MRI Grading Systems

Muscle	Ultrasound Grade	MRI Grade		
		Absence (Goutallier 0, 1)	Presence (Goutallier 2, 3, 4)	Total
Supraspinatus	Absence (grade 0)	52	4	56
	Presence (grades 1, 2)	2	22	24
	Total	54	26	80
Infraspinatus	Absence (grade 0)	52	1	53
	Presence (grades 1, 2)	5	22	27
	Total	57	23	80
Teres minor	Absence (grade 0)	63	1	64
	Presence (grades 1, 2)	9	7	16
	Total	72	8	80

**Interobserver and Intraobserver Reliability of MRI**

The interobserver reliability among all four raters had a weighted kappa of 0.76 (95% CI, 0.72 to 0.80) for the supraspinatus, 0.77 (95% CI, 0.74 to 0.81) for the infraspinatus, and 0.59 (95% CI, 0.51 to 0.66) for the teres minor. The intraobserver agreement was determined for raters 2 and 3. Rater 2 exhibited a weighted kappa of 0.90 (95% CI, 0.83 to 0.96) for the supraspinatus, 0.80 (95% CI, 0.72 to 0.89) for the infraspinatus, and 0.75 (95% CI, 0.62 to 0.89) for the teres minor. Rater 3 exhibited a weighted kappa of 0.77 (95% CI, 0.68 to 0.86) for the supraspinatus, 0.84 (95% CI, 0.77 to 0.91) for the infraspinatus, and 0.71 (95% CI, 0.53 to 0.90) for the teres minor.

**Interobserver Reliability of Ultrasonography**

Using the three-point grading scales, the weighted kappa of the three radiologists was 0.71 (95% CI, 0.61 to 0.81) for the supraspinatus, 0.65 (95% CI, 0.56 to 0.74) for the infraspinatus, and 0.72 (95% CI, 0.64 to 0.80) for the teres minor.

**Discussion**

As a diagnostic imaging tool for rotator cuff muscles, ultrasonography has many potential benefits over MRI. In addition to its well-known advantages of being less expensive, well tolerated, and reliable in patients with metallic implants or claustrophobia, it provides a dynamic and global evaluation of the cuff muscles in real time, whereas MRI provides a static and more limited evaluation of the cuff muscles and does not image the more medial aspect of the muscle bellies. The present study was performed to investigate the diagnostic performance and reliability of ultrasonography for detecting and grading fatty degeneration of the rotator cuff muscles, using MRI as the reference standard.

On the basis of the dichotomous scales, ultrasonography had excellent agreement with MRI for the detection of fatty degeneration in the supraspinatus and infraspinatus muscles ( $\kappa = 0.83$  for both) and moderate agreement for the teres

minor muscle ( $\kappa = 0.52$ ). With the three-point scales, there was substantial agreement for the supraspinatus and infraspinatus ( $\kappa = 0.78$  and  $0.71$ , respectively) and moderate agreement for the teres minor ( $\kappa = 0.47$ ). It should be noted that the level of agreement between ultrasonography and MRI was substantially higher than the agreement between MRI and computed tomography (CT) reported by Fuchs et al.<sup>8</sup>. In that study, the mean weighted kappa value was 0.45 for both the supraspinatus and infraspinatus muscles. Although the original Goutallier grading system was based on axial CT images<sup>12</sup> and the agreement between CT and MRI was not satisfactory, MRI has become the accepted modality for the diagnosis and grading of fatty degeneration<sup>9,10</sup>. The agreement between ultrasonography and MRI in the present study was better than the reported agreement between CT and MRI.

The present study also showed that the diagnostic performance of ultrasonography for fatty degeneration was excellent. With MRI as the reference standard, the percentage agreement (i.e., accuracy) was 92.5% for both the supraspinatus and infraspinatus and 87.5% for the teres minor. This was higher than the accuracy reported by Strobel et al.<sup>10</sup>. In that study, the accuracy was 72% to 75% for the supraspinatus muscle and 80% to 85% for the infraspinatus. The use of static ultrasonography images for the muscle evaluations may explain the relatively low accuracy in that study. In contrast, Khoury et al.<sup>9</sup> evaluated the cuff muscles in real time as we did, with a technique similar to ours, and demonstrated an accuracy comparable with that in the present study.

The interobserver reliability of ultrasonography for fatty degeneration was comparable with that of MRI for the supraspinatus muscle, slightly worse for the infraspinatus, and slightly better for the teres minor. However, none of these differences was significant, as indicated by the overlap of the 95% confidence intervals. The interobserver reliability of ultrasonography was  $\kappa = 0.71$  for the supraspinatus muscle and  $\kappa = 0.65$  for the

infraspinatus in our study. These findings were comparable with the interobserver reliability of CT reported in the study of Fuchs et al.<sup>8</sup>, in which the reliability was 0.72 for the supraspinatus muscle and 0.69 for the infraspinatus. Williams et al. also studied the interobserver reliability of CT for fatty degeneration of the supraspinatus muscle and reported kappa coefficients of 0.48 to 0.59<sup>16</sup>. Although it is not possible to directly compare the different grading systems in these studies, our findings suggest that the interobserver reliability of ultrasonography may be at least comparable with those of CT and MRI.

To our knowledge, our study is the first to investigate the diagnostic performance of ultrasonography in grading fatty degeneration of the teres minor muscle. The quality of the teres minor is an important prognostic factor for functional outcome following reverse total shoulder arthroplasty and latissimus dorsi transfer in patients with an irreparable rotator cuff tear<sup>18-22</sup>. In the present study, the percentage agreement between ultrasonography and MRI for the teres minor (87.5%) was slightly lower than that for the supraspinatus and infraspinatus (92.5% for each). The agreement with MRI was moderate ( $\kappa = 0.47$ ) for the teres minor with use of the three-point scales although the interobserver reliability for ultrasonography was substantial ( $\kappa = 0.72$ ). Although these data are supportive of ultrasonography grading of the teres minor muscle, the statistical assessment must be considered thoughtfully since the prevalence of fatty changes in the teres minor muscle is relatively low. The limited number of subjects made it difficult to quantify the discriminatory ability of ultrasonography. The kappa value may not be the best measure for these data and should be interpreted cautiously since chance agreement is high, and kappa is reduced accordingly, when the prevalence is very high or very low<sup>23</sup>. In contrast to the kappa value, however, the percentage agreement does reflect substantial observer accuracy.

One of the differences between ultrasonography and MRI is that ultrasonography uses the overlying muscles as its reference for grading fatty degeneration, whereas MRI takes into account only the absolute ratio of fat to muscle tissue within a given muscle. It is a well known phenomenon that aging muscles accumulate fat<sup>24-26</sup>, and it is not uncommon to see fatty streaks in the deltoid and trapezius muscles in the MRIs of elderly patients. This is a possible source of the discrepancy between ultrasonography and MRI in distinguishing a grade-1 muscle from a grade-0 muscle.

Ultrasonography and MRI have certain intrinsic limitations that should be mentioned. First, ultrasonography relies much more on the experience and skills of the operator than MRI and CT do, and it has a long, steep learning curve. An experienced radiologist may not be available to carry out this examination. Ultrasonography is also difficult to perform in obese patients because of the low penetration rate of ultrasound into the deep tissue. Additionally, ultrasonography cannot evaluate the subscapularis muscle because of its medial location, and this limits the scope of ultrasonography as a comprehensive imaging modality for rotator cuff pathology.

MRI, in contrast to ultrasonography, is a static examination. This is epitomized by the use of the Goutallier grading

system, which utilizes a single parasagittal image from which the grade of fatty degeneration is determined. MRI is also limited by the presence of metallic implants, which can generate scatter that causes the images to be unreadable. Furthermore, most shoulder MRI examinations do not image the entirety of the rotator cuff musculature on sagittal images, as the medial aspect of the muscle bellies is usually not fully visualized. Motion artifact can also be a problem with MRI examinations, and this can become very problematic in restless or claustrophobic patients.

This study was designed to evaluate the diagnostic performance of ultrasonography, compared with MRI, for grading fatty degeneration of the posterior rotator cuff muscles. Our findings showed that ultrasonography was comparable in accuracy with MRI, which is the accepted gold standard. However, it is notable that ultrasonography may in fact be better than MRI because of its capability of evaluating the cuff muscles globally from their insertions to their origins in real time. Overall, these findings suggest that ultrasonography could be used as the primary imaging modality for the evaluation of not only tears but also fatty infiltration of the rotator cuff. Future studies comparing the muscles in their entirety using MRI and ultrasonography are needed to confirm this.

There are certain limitations inherent in the present study. Intraobserver reliability of ultrasonography was not investigated in this study since it was deemed impractical to ask patients to return for a second ultrasonographic examination. Also, it was thought to be technically difficult to blind the radiologists. A second limitation is that the subscapularis muscle, a critical rotator cuff muscle, was not examined in this study. A third limitation involves the collapse of the MRI scale. In order to directly compare the five-point MRI scale with the three-point ultrasonography scale, the five-point scale was collapsed, based on previous literature and the authors' clinical experience. The authors acknowledge that this reduction may have introduced bias, and the effects of such bias on the reliability are unknown. Lastly, there is also a limitation involving the data analysis utilizing the dichotomized scales; the agreement statistic for the dichotomized scales may be artificially inflated simply as a result of the dichotomization process.

In summary, the diagnostic performance and observer reliability of ultrasonography were comparable with those of MRI. In addition to ultrasonography's well-known advantages of being less expensive, being well tolerated by patients, requiring less time for the examination, and being reliable in patients with metal implants or claustrophobia, it provides dynamic and global evaluation in real time. The satisfactory diagnostic performance shown in the present study suggests that ultrasonography can be used as a primary diagnostic imaging modality for the rotator cuff muscles. ■

Lindley B. Wall, MD  
H. Mike Kim, MD  
Ken Yamaguchi, MD  
Department of Orthopaedic Surgery,  
Washington University School of Medicine,



1 Barnes-Jewish Hospital Plaza,  
11300 West Pavilion,  
Campus Box 8233, St. Louis, MO 63110.  
E-mail address for K. Yamaguchi: yamaguchi@wudosis.wustl.edu

Sharlene A. Teefey, MD  
William D. Middleton, MD  
Nirvikar Dahiya, MD

Daniel Wessell, MD, PhD  
Mallinckrodt Institute of Radiology,  
Washington University School of Medicine,  
510 South Kingshighway Boulevard, St. Louis, MO 63110

Karen Steger-May, MA  
Division of Biostatistics Washington University School of Medicine,  
660 South Euclid Avenue, Campus Box 8067, St. Louis, MO 63110-1093

## References

1. Cho NS, Rhee YG. The factors affecting the clinical outcome and integrity of arthroscopically repaired rotator cuff tears of the shoulder. *Clin Orthop Surg*. 2009 Jun;1(2):96-104. Epub 2009 May 30.
2. 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 May;35(5):719-28. Epub 2007 Mar 2.
3. Goutallier D, Postel JM, Gleyze P, Leguilloux P, Van Driessche S. Influence of cuff muscle fatty degeneration on anatomic and functional outcomes after simple suture of full-thickness tears. *J Shoulder Elbow Surg*. 2003 Nov-Dec;12(6):550-4.
4. Shen PH, Lien SB, Shen HC, Lee CH, Wu SS, Lin LC. Long-term functional outcomes after repair of rotator cuff tears correlated with atrophy of the supraspinatus muscles on magnetic resonance images. *J Shoulder Elbow Surg*. 2008 Jan-Feb;17(1 Suppl):1S-7S. Epub 2007 Nov 1.
5. Middleton WD, Teefey SA, Yamaguchi K. Sonography of the rotator cuff: analysis of interobserver variability. *AJR Am J Roentgenol*. 2004 Nov;183(5):1465-8.
6. Prickett WD, Teefey SA, Galatz LM, Calfee RP, Middleton WD, Yamaguchi K. Accuracy of ultrasound imaging of the rotator cuff in shoulders that are painful postoperatively. *J Bone Joint Surg Am*. 2003 Jun;85-A(6):1084-9.
7. Teefey SA, Rubin DA, Middleton WD, Hildebolt CF, Leibold RA, Yamaguchi K. Detection and quantification of rotator cuff tears. Comparison of ultrasonographic, magnetic resonance imaging, and arthroscopic findings in seventy-one consecutive cases. *J Bone Joint Surg Am*. 2004 Apr;86-A(4):708-16.
8. Fuchs B, Weishaupt D, Zanetti M, Hodler J, Gerber C. Fatty degeneration of the muscles of the rotator cuff: assessment by computed tomography versus magnetic resonance imaging. *J Shoulder Elbow Surg*. 1999 Nov-Dec;8(6):599-605.
9. Khoury V, Cardinal E, Brassard P. Atrophy and fatty infiltration of the supraspinatus muscle: sonography versus MRI. *AJR Am J Roentgenol*. 2008 Apr;190(4):1105-11.
10. Strobel K, Hodler J, Meyer DC, Pfirmann CW, Pirkl C, Zanetti M. Fatty atrophy of supraspinatus and infraspinatus muscles: accuracy of US. *Radiology*. 2005 Nov;237(2):584-9. Epub 2005 Sep 28.
11. Zanetti M, Gerber C, Hodler J. Quantitative assessment of the muscles of the rotator cuff with magnetic resonance imaging. *Invest Radiol*. 1998 Mar;33(3):163-70.
12. 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 Jul;(304):78-83.
13. Fleiss JL. *Statistical methods for rates and proportions*. 2nd ed. New York: Wiley; 1981.
14. Teefey SA, Hasan SA, Middleton WD, Patel M, Wright RW, Yamaguchi K. Ultrasonography of the rotator cuff. A comparison of ultrasonographic and arthroscopic findings in one hundred consecutive cases. *J Bone Joint Surg Am*. 2000 Apr;82(4):498-504.
15. Kim HM, Dahiya N, Teefey SA, Keener JD, Yamaguchi K. Sonography of the teres minor: a study of cadavers. *AJR Am J Roentgenol*. 2008 Mar;190(3):589-94.
16. Williams MD, Lädermann A, Melis B, Barthelemy R, Walch G. Fatty infiltration of the supraspinatus: a reliability study. *J Shoulder Elbow Surg*. 2009 Jul-Aug;18(4):581-7.
17. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977 Mar;33(1):159-74.
18. Boileau P, Watkinson D, Hatzidakis AM, Hovorka I. Neer Award 2005: the Grammont reverse shoulder prosthesis: results in cuff tear arthritis, fracture sequelae, and revision arthroplasty. *J Shoulder Elbow Surg*. 2006 Sep-Oct;15(5):527-40.
19. Costouros JG, Espinosa N, Schmid MR, Gerber C. Teres minor integrity predicts outcome of latissimus dorsi tendon transfer for irreparable rotator cuff tears. *J Shoulder Elbow Surg*. 2007 Nov-Dec;16(6):727-34. Epub 2007 Nov 5.
20. Matsen FA 3rd, Boileau P, Walch G, Gerber C, Bicknell RT. The reverse total shoulder arthroplasty. *J Bone Joint Surg Am*. 2007 Mar;89(3):660-7.
21. Simovitch RW, Helmy N, Zumstein MA, Gerber C. Impact of fatty infiltration of the teres minor muscle on the outcome of reverse total shoulder arthroplasty. *J Bone Joint Surg Am*. 2007 May;89(5):934-9.
22. Sirveaux F, Favard L, Oudet D, Huquet D, Walch G, Molé D. Grammont inverted total shoulder arthroplasty in the treatment of glenohumeral osteoarthritis with massive rupture of the cuff. Results of a multicentre study of 80 shoulders. *J Bone Joint Surg Br*. 2004 Apr;86(3):388-95.
23. Brennan P, Silman A. Statistical methods for assessing observer variability in clinical measures. *BMJ*. 1992 Jun 6;304(6840):1491-4.
24. Forsberg AM, Nilsson E, Werneman J, Bergström J, Hultman E. Muscle composition in relation to age and sex. *Clin Sci (Lond)*. 1991 Aug;81(2):249-56.
25. Rice CL, Cunningham DA, Paterson DH, Lefcoe MS. Arm and leg composition determined by computed tomography in young and elderly men. *Clin Physiol*. 1989 Jun;9(3):207-20.
26. Tsubahara A, Chino N, Akaboshi K, Okajima Y, Takahashi H. Age-related changes of water and fat content in muscles estimated by magnetic resonance (MR) imaging. *Disabil Rehabil*. 1995 Aug-Sep;17(6):298-304.