AJSM PreView, published on July 2, 2012 as doi:10.1177/0363546512452714

Interobserver and Intraobserver Reliability of the Goutallier Classification Using Magnetic Resonance Imaging

Proposal of a Simplified Classification System to Increase Reliability

Mark A. Slabaugh,* MD, Nicole A. Friel,† MD, Vasili Karas,† MD, Anthony A. Romeo,† MD, Nikhil N. Verma,† MD, and Brian J. Cole,†‡ MD, MBA
Investigation performed at Rush University Medical Center, Chicago, Illinois

Background: The Goutallier classification of fatty infiltration of the rotator cuff was developed for use in axial computed tomography arthrography. Now the Goutallier classification is being used with magnetic resonance imaging (MRI). Not only is there debate on the validity of the use of this system in MRI, but current literature is unclear as to the clinical use of the Goutallier classification.

Hypothesis: There will be significant inter- and intraobserver variability of the Goutallier classification grading system for patients with chronic rotator cuff tears.

Study Design: Cohort study (diagnosis); Level of evidence, 2.

Methods: An online database consisting of 35 single MRI images from 35 patients with chronic rotator cuff tears was designed and sent to members of the American Shoulder and Elbow Society. Surgeons were asked to identify the stage of rotator cuff fatty infiltration using the Goutallier classification system. Thirty surgeons responded and completed the survey in its entirety. At a minimum of 2 months later, 28 of the 30 initial respondents completed evaluations of the same online database. The responding surgeons were divided dichotomously according to their demographics and the interobserver reliability of the groups compared. A kappa analysis was performed to determine inter- and intraobserver reliability using 95% confidence intervals (95% CIs). A simplified 3-tiered classification was proposed combining Goutallier grades 0 and 1 as well as grades 2 and 3.

Results: Statistical analysis of all respondent data demonstrated moderate intraobserver variability with a κ value of 0.56 (95% CI, 0.53-0.60). In addition, moderate interobserver variability was shown with a κ value of 0.43 (range, 0.16-0.74). With the simplified classification, intraobserver reliability was 0.70 (95% CI, 0.66-0.74) and interobserver reliability was 0.61 (range, 0.21-0.87). Correlation analysis showed no correlation with the presence or absence of fellowship training or board certification with either the Goutallier classification or the proposed modification ($P > .05$). Sports versus shoulder/elbow fellows had statistically better intraobserver variability ($\kappa = 0.63$ vs 0.50) with the Goutallier classification. Years in practice was negatively correlated with the level of agreement for both classifications (–r value, $P < .05$). The number of rotator cuff repairs performed per year negatively correlated with the level of agreement in the proposed modification only ($r = –0.44$, $P = .022$). Percent of practice dedicated to the shoulder did not correlate significantly with either classification ($P > .05$).

Conclusion: There is significant inter- and intraobserver variability observed among experienced shoulder surgeons using the Goutallier classification for assessing fatty infiltration of the rotator cuff muscles after chronic rotator cuff tears. Respondents were more likely to agree with themselves than with other respondents. A simplification of the MRI classification system is proposed that takes into consideration the variability determined by this study.

Keywords: rotator cuff tears; MRI; Goutallier classification; fatty infiltration; reliability

Fatty degeneration or infiltration of the rotator cuff musculature is a consequence of chronic rotator cuff muscle tears. The precise cause of this degeneration is unknown; however, it has been proposed that degeneration results from the lack of neurogenic feedback, imbalance in the muscle tendon stress/strain, or physiologic changes in the muscles themselves resulting from tendon bone discontinuity. Independent of cause, the rotator cuff muscles ultimately display degeneration and atrophy with fatty replacement in addition to scarring and retraction of the muscle-tendon unit.

‡Address correspondence to Brian J. Cole, MD, MBA, Rush University Medical Center, 1611 W Harrison St, Suite 300, Chicago, IL 60612 (e-mail: bcole@rushortho.com).
*Department of Surgery, Aviano AB Hospital, Aviano, Italy.
†Rush University Medical Center, Chicago, Illinois.
The authors declared that they have no conflicts of interest in the authorship and publication of this contribution.

The American Journal of Sports Medicine, Vol. XX, No. X
DOI: 10.1177/0363546512452714

Goutallier et al[6] first described a method for classifying fatty infiltration of the rotator cuff in 1994. This scheme was originally used for assessing fatty infiltration of the infraspinatus and subscapularis muscles using axial computed tomography arthrography (CTA). Fuchs et al[3] later applied this classification scheme to magnetic resonance imaging (MRI) as it became an integral and commonly used method for the evaluation of rotator cuff integrity. Since the publication of these landmark studies, there has been considerable debate regarding the clinical application of this grading scheme in relation to several shoulder surgical procedures, including rotator cuff repairs and both conventional and reverse shoulder arthroplasty. Several authors have shown that the degree of fatty infiltration negatively affects the clinical outcomes of patients when specifically regarding rotator cuff repairs.[2,4-7,13,16] In addition, it has been shown that even after a successful rotator cuff repair (anatomic and clinical), fatty infiltration does not improve in patients.[5,12,17]

Because of the clinical implications of fatty infiltration in rotator cuff tears, there have been several efforts to ensure that the Goutallier scheme is reliable and valid. Despite the widespread use of this system, several authors have questioned its validity, and others have proposed new classification schemes.[8,18] Such questions have emerged because of a number of variables affecting standardization, including which imaging modality and plane to use as well as identifying which rotator cuff muscle most effectively classifies a tear. Furthermore, the lack of knowledge regarding whether atrophy or fatty infiltration of the rotator cuff musculature is responsible for poor outcomes after repair of tears has led to difficulty in defining a classification system that is clinically relevant.

The purpose of this study was to determine the inter- and intraobserver reliability of the Goutallier classification scheme for rotator cuff infiltration using a large cohort of subspecialty trained sports and shoulder surgeons.

MATERIALS AND METHODS

Institutional review board approval was obtained before the inception of this study, and informed consent was obtained from each patient. A power analysis deemed that 35 MRI images were necessary with a minimum of 25 respondent surgeons to obtain significant power ($\beta > 0.8$). Thirty-five patients were selected having a documented preoperative MRI (standardized sequences from our institution using a Siemens 1.5 Tesla magnet) and a large rotator cuff tear (>3 cm) that was confirmed at the time of surgery. These patients were randomly selected from the databases of 3 shoulder- or sports-trained subspecialty orthopaedic surgeons from a single institution. A single T1-weighted oblique sagittal image was selected for each patient. The sagittal image selected was the lateralmost image in which the scapular spine is still in contact with the scapular body.

Because of the controversy regarding imaging modality, imaging plane, and identifying muscle belly that is best to use for the Goutallier classification, MRI and the oblique sagittal plane of the supraspinatus were chosen due to supporting literature and clinical experience.[15] Magnetic resonance imaging was chosen over CTA because of its more frequent use and familiarity among the musculoskeletal radiologists at our institution (CTA was never obtained for evaluating the rotator cuff at our institution). In addition, Oh et al[15] found that the interobserver reliability is higher when using MRI over CTA. Oblique sagittal images were chosen as the plane for this study since most studies to date have used at least the oblique sagittal images in their determination of reliability, whereas not all studies have used the axial or oblique coronal images to determine reliability of the Goutallier classification scheme.[15] The supraspinatus muscle was chosen for grading since all patients in this cohort had at minimal a supraspinatus tear. It has also been demonstrated that a supraspinatus tear is the most common muscle-tendon tear in patients with rotator cuff injury.[9] When designing the online database, 1 image was chosen per patient for a total of 35 images to increase participation in the study. Although fatty infiltration is not evenly distributed throughout the muscle in each image, including multiple images of the 35 patients would have discouraged surgeon participation.

The 35 images were placed in an online database free of any identifying information to ensure patient confidentiality. Approval was obtained from the Shoulder and Elbow Society to solicit participation from their closed membership. An e-mail explaining the purpose and nature of the study was sent to each of the Shoulder and Elbow Society members asking for their participation. For those surgeons willing to participate, an embedded link was incorporated into the e-mail allowing them to open the survey.

The first portion of the survey restated the purposes and objectives of the study. The second portion of the survey consisted of an educational tutorial, which described the Goutallier classification of fatty infiltration of the rotator cuff. The tutorial consisted of sample images that clearly delineated the grading of all 5 stages of the Goutallier classification scheme. Grade 0 was defined as no fatty infiltration, grade 1 as some fatty streaking of the supraspinatus, grade 2 as less fat than muscle, grade 3 as equal amounts of fat and muscle, and grade 4 as more fat than muscle (see Figures 1-5).

Participants then graded the 35 selected MRI images according to this classification scheme. Each image was accompanied by both the grade (ie, 1, 2, 3, 4) and a corresponding description of the grade for an alternative option of selection. Participants were asked to grade only the intramuscular portion of the supraspinatus. Participants were also asked to include demographic data, orthopaedic surgery board certification, specialty training, years in practice, portion of shoulder cases in their practice, and the average number of rotator cuff repairs performed per year. No time limit was set for finishing the survey. Only those participants who completed the entire survey were included in the data analysis. If participants completed only a portion of the survey, a reminder e-mail was sent out 2 days after beginning the survey.

At a minimum of 2 months later (time chosen to prevent possible surgeon image recall), a repeat e-mail was sent to

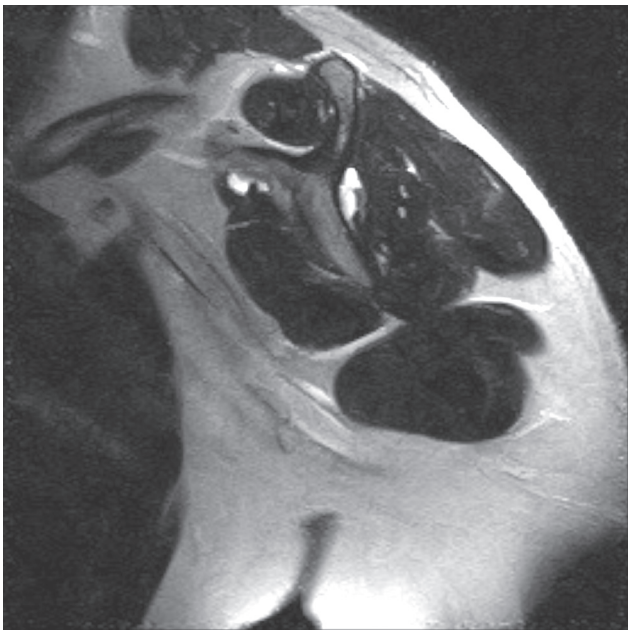

Figure 1. T1-weighted MRI showing grade 0 fatty infiltration in a patient with a large full-thickness supraspinatus tear.

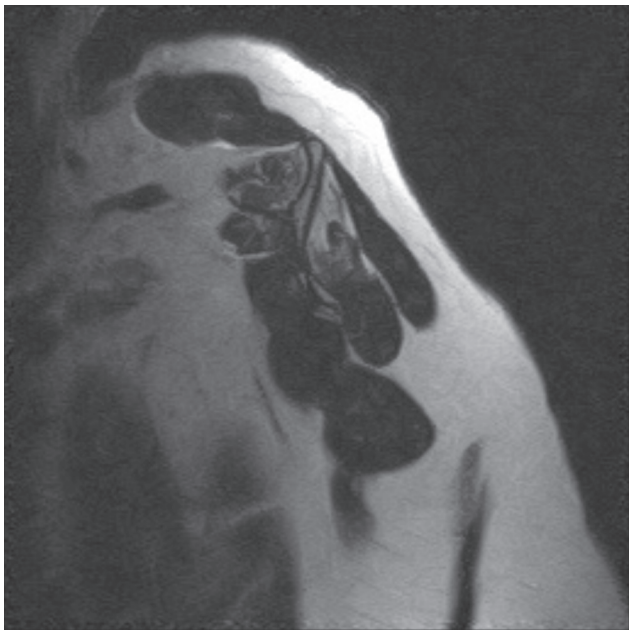

Figure 3. T1-weighted MRI showing grade 2 fatty infiltration in a patient with a large full-thickness supraspinatus tear.

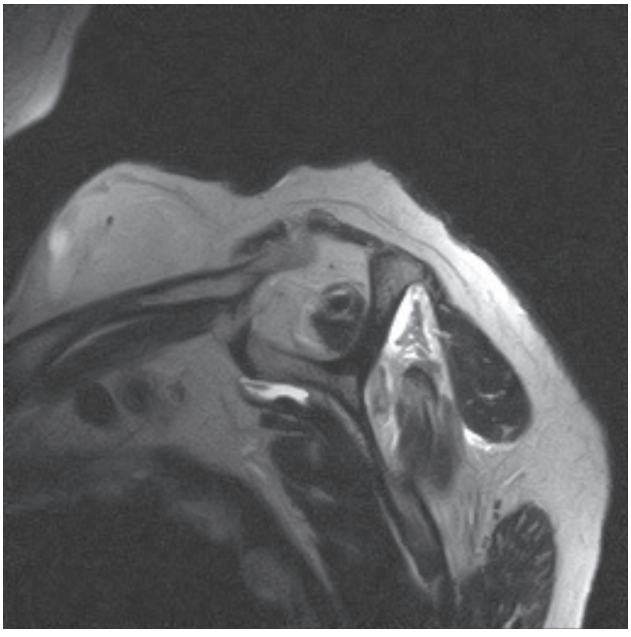

Figure 2. T1-weighted MRI showing grade 1 fatty infiltration in a patient with a large full-thickness supraspinatus tear.

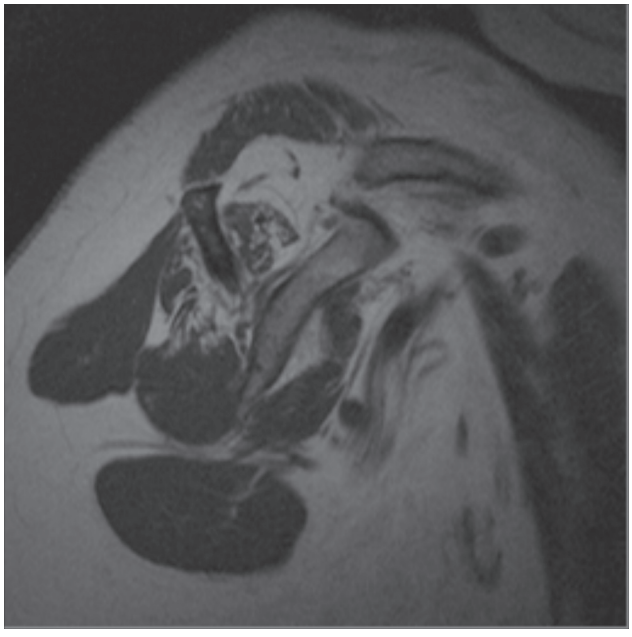

Figure 4. T1-weighted MRI showing grade 3 fatty infiltration in a patient with a large full-thickness supraspinatus tear.

participants who had entirely completed the survey. The purpose of the survey and the objectives of the study were restated, and participation was elicited. A link to the survey was contained in the e-mail. Participants were reeducated about the Goutallier grading system with written descriptions and MRI images of each respective grade (see Figures 1-5). The same 35 images were contained in the survey and tested in the same fashion but in a randomized order to prevent recall bias.

Statistical analysis was completed using the initial and follow-up survey scores from the 28 respondents who graded the 35 MRI images. Kappa analysis with 95% confidence intervals was performed for both the inter- and intraobserver reliability using Excel (Microsoft, Redmond, Washington)

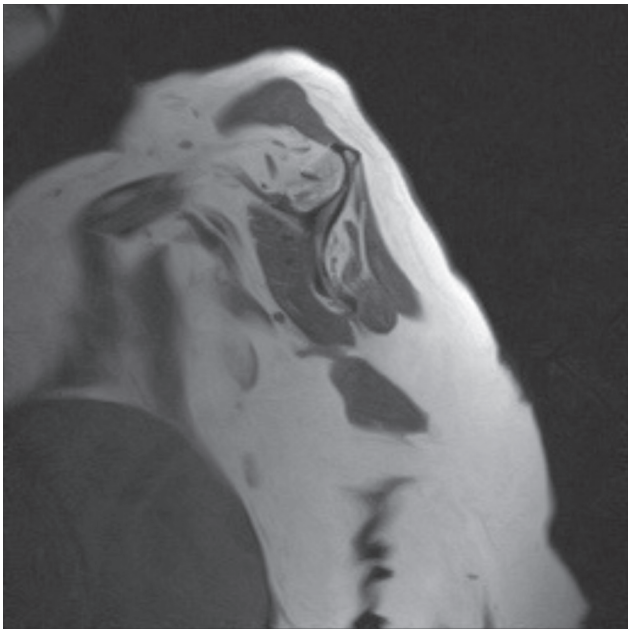

Figure 5. T1-weighted MRI showing grade 4 fatty infiltration in a patient with a large full-thickness supra- and infraspinatus tear.

and SAS (SAS Institute, Cary, North Carolina). Degrees of reliability were used as described by Landis and Koch[10] (see Table 1). To determine the correlation between intraobserver reliability and demographic data of the respondents, a simple *t* test was performed using the same software (Excel, SAS) to compare the κ values with the dichotomous value (ie, board certification vs not, fellowship training vs not) and κ values for types of fellowship (shoulder/elbow vs sports). Statistical significance was set at $P < .05$. Correlation analysis (SAS) was used to measure the effect of years in practice, number of rotator cuffs repaired, and percent of practice dedicated to shoulder cases on κ values. The levels of agreement were tested for normality using the Anderson-Darling Test for Normality.[1] All data from the respondents can be found in the Appendix (available in the online version of this article at http://ajs.sagepub.com/supplemental/).

RESULTS

The survey was sent to the 300+ members of the Shoulder and Elbow Society. Of these, 37 members responded to the survey. Thirty surgeons completed the entire survey initially. Of these 30 respondents, 28 completed the survey 2 months after completing the initial survey (93.3%). Demographic data of the respondents included 16 fellowship-trained shoulder surgeons and 12 sports medicine–trained surgeons with an average of 18.5 years (range, 4-32 years) of experience performing an average of 157.3 (range, 30-250) rotator cuff repairs each year. For the respondents, 20% to 100% of their practice was dedicated to the shoulder. Twenty-six of the respondents were board certified.

As hypothesized, respondents were more likely to agree with themselves than with other surgeons, resulting in higher intraobserver reliability. Kappa statistical analysis revealed moderate intraobserver reliability with a κ value of 0.56 (95% confidence interval [CI], 0.53-0.60) (Table 2). Analysis of the interobserver reliability showed moderate interobserver reliability with a κ value of 0.43 (range, 0.16-0.74). Finally, when the senior author's reading of the images was used as a gold standard, only a small increase in interobserver reliability resulted (0.43-0.46).

Correlation of the demographic data with the reliability of the Goutallier classification demonstrated that the presence/absence of board certification or fellowship training did not have a significant effect on the level of agreement (*P* value of .64 and .33, respectively). The type of fellowship did, however, show a significant effect on the level of agreement. Sports fellowship–trained physicians produced an intraobserver agreement of 0.63, whereas shoulder/elbow fellowship–trained physicians produced an intraobserver agreement of 0.50 ($P = .025$) (Table 2).

In addition, years in practice negatively correlated with the level of agreement ($r = –0.45$, $P = .017$) (Table 3). The number of rotator cuff repairs performed per year also showed a trend toward negative correlation ($r = –0.36$, $P = .065$). Percent of practice dedicated to the shoulder did not correlate significantly either ($r = –0.1260$, $P = .53$). The Anderson-Darling test for normality was used to test the null hypothesis to ensure that the analysis was valid ($P = .63$).

TABLE 1
Degrees of Reliability Determined by κ Values
(From Landis and Koch[10])

κ Value	Reliability
0.00-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost perfect

TABLE 2
Summary of Inter- and Intraobserver Reliability
of the Goutallier Classification and
Proposed Modified Classification
With Correlation to the Demographic Data

	Interobserver Reliability	Intraobserver Reliability
Goutallier classification	0.43	0.56
Proposed modification of classification	0.61	0.70
Comparison of demographic data		
Sports vs shoulder/elbow		0.63/0.50[a]
Board vs non–board certification		0.57/0.51[b]
Fellowship vs nonfellowship		0.55/0.66[b]

[a] $P < .05$.
[b] $P > .05$.

TABLE 3
Summary of Correlation Analysis

	Goutallier Classification		Proposed Classification	
	Correlation (r)	P Value	Correlation (r)	P Value
Years in practice	–0.45	.017	–0.40	.041
Number of RCRs[a] performed	–0.36	.065	–0.44	.0216
% of practice dedicated to shoulder	–0.126	.531	–0.21	.301

[a]RCRs, rotator cuff repairs.

TABLE 4
Comparison of Fatty Infiltration Using the Goutallier and Proposed Modified Classifications

	Goutallier Classification	Proposed Classification
Grade 0	Normal muscle	Normal to mild fatty infiltration
Grade 1	Fatty streaks	
Grade 2	Fat < muscle	Moderate fatty infiltration
Grade 3	Fat = muscle	
Grade 4	Fat > muscle	Fat > muscle

Based on the moderate inter- and intraobserver reliability findings of this study, an alternative scale for the assessment of fatty infiltration of the rotator cuff musculature was proposed. Thus, we proposed and analyzed a modified Goutallier classification scheme that combined grades 0 and 1 (normal to mild fatty infiltration) in addition to grades 2 and 3 (moderate fatty infiltration) while maintaining grade 4 (more fat than muscle) in its own category (Table 4).

Using this alternative classification scheme, the intraobserver reliability increased from 0.56 with the traditional scoring to 0.70, indicating that there was substantial reliability of the modified system. The 95% CI for the new classification system was 0.66 to 0.74. The interobserver reliability also increased from a κ value of 0.43 to 0.61, indicating substantial reliability of the new classification system.

Correlation analysis of the proposed modification of the Goutallier classification had similar findings to the correlation analysis of the Goutallier classification. Correlation of the demographic data demonstrated that the presence/absence of board certification or fellowship training did not affect the level of agreement (P value of .77 and .94, respectively). The type of fellowship, however, did not show a significant effect on the level of agreement ($P = .10$).

Years in practice negatively correlated with the level of agreement ($r = –0.40$, $P = .041$). Although the number of rotator cuff repairs performed per year trended toward negative correlation with the Goutallier classification, our proposed modification showed a negative correlation with the level of agreement ($r = –0.44$, $P = .022$). As with the Goutallier classification, percentage of practice dedicated to the shoulder did not correlate significantly either ($r = –0.21$, $P = .30$) in our proposed modification.

DISCUSSION

The principal findings of this study demonstrate that the inter- and intraobserver reliability of the Goutallier classification is only moderate among shoulder-trained subspecialists. When this classification system is reduced to 3 grades, both inter- and intraobserver reliabilities can be characterized as being substantial.[10] Furthermore, we showed that only type of fellowship training correlated with intraobserver reliability, whereas years in practice negatively correlated with intraobserver reliability.

For a classification system to be of clinical value, the grading scale has to be effective with regard to decisions for treatment and clinical outcomes. Fatty infiltration of the rotator cuff musculature has received significant attention due to several studies that have shown its prognostic value in clinical and radiologic outcome studies.[2,4-7,13,16] Such attention has produced an ongoing interest in evaluating the reliability of the grading of fatty infiltration of the rotator cuff.[3,6,11,15,19]

A study done by Fuchs et al[3] was the first to look at the reliability of their classification scheme for assessing fatty infiltration. Both computed tomography and MRI images of patients with rotator cuff tears were studied, and it was found that MRI was a superior method for assessing fatty infiltration due to higher interobserver reliability. Their study, however, was limited by having only 2 radiologists involved in the grading of the fatty infiltration. In comparison with our results, similar interobserver reliability was found in their study (0.61), yet due to the method used, no intraobserver reliability was possible.

Fuchs et al[3] proposed a 3-stage grading classification as well; however, their system combined grades 3 and 4, whereas our system combined grades 2 and 3. On the basis of the statistical analysis, we chose to combine grades 2 (fat < muscle) and 3 (fat = muscle) since the distinction between these 2 grades was less than if we were to combine grades 3 and 4. In addition, we thought that a significant difference between the amount of fatty infiltration in grades 3 and 4 would affect the outcome and prognosis of a rotator cuff repair (strength, anatomic/functional, etc). We concede that this distinction has not been shown in the literature, and further studies are needed to correlate this modification with clinical outcome.

In patients with rotator cuff ruptures, Williams et al[19] found similar reliability values using the Goutallier

classification system in comparison with those found in our study. This study can extend upon the findings of our study as it tested all 3 imaging planes for each of the 87 total patients. For the axial, coronal, and sagittal images, interobserver reliability values were 0.50, 0.46, and 0.40, respectively. Using the oblique sagittal images on CTA, interobserver reliability had a κ value of 0.40, which compared well with our study (κ = 0.43). The intraobserver reliability seen in the Williams et al[19] study was slightly higher at 0.61 in comparison with our value of 0.56. Despite finding similar reliabilities, it is important to consider that Williams et al[19] used CTA and enrolled patients with all rotator cuff tears irrespective of tear size.

Williams et al[19] also proposed a 3-tier system for grading fatty infiltration that they ultimately found to be more reliable regarding both inter- and intraobserver reliability for all 3 planes studied. As done in our study, Williams et al[19] combined grades 0 and 1 of the Goutallier classification system. Their study, however, combined grades 3 and 4, whereas we chose to combine grades 2 and 3 and leave grade 4 intact.[19] On the basis of clinical experience, we have found it more difficult to make a distinction between grades 2 and 3 than to make a distinction between grades 3 and 4. Also, the interobserver reliability (κ = 0.61) seen in our study, when using our 3-tier classification, was higher in comparison to interobserver reliability (κ = 0.48) seen when using the 3-tier system proposed by Williams et al[19] for axial images. According to Landis and Koch,[10] it can be said that our 3-tier classification system produces substantial reliability, whereas the Williams et al[19] classification system produces only moderate reliability.

Another study performed by Oh et al[15] looked at the reliability of the classification of fatty infiltration by having 3 orthopaedic surgeons and 2 musculoskeletal radiologists grade 75 patients preoperatively with magnetic resonance arthrography (MRA) and postoperatively with CTA. This study concluded that MRA was superior to CTA because of higher resulting inter- and intraobserver reliabilities. These values were higher than those found in our study. It is important to consider that Oh et al[15] used the interclass coefficient (ICC) to determine the inter- and intraobserver reliabilities as opposed to using κ values.[15] The ICC is not as effective as κ in terms of measuring reliability because it does not consider chance and typically overestimates inter- and intrareliability values.[14] Using the ICC to determine our intraobserver reliability would have computed out to a value of 0.88 (a reliability of near-perfect correlation), which, compared with a κ value of 0.56, exaggerates the reliability of our study tremendously. Oh et al[15] also concluded that more experienced surgeons and radiologists had higher intraobserver reliability, which was not seen in our study. This finding could be attributed to the fact that the most experienced observer in the Oh et al[15] study had 7 years in practice. The most experienced respondent in our study had 32 years in practice. Despite conventional thinking, years in practice had a negative correlation with intraobserver reliability. It would seem intuitive that experienced surgeons would be able to grade fatty infiltration more reliably because of an increased number of patients seen with this condition; however, this is not supported by our current study. Perhaps the experience of the senior surgeons made them more cognizant of the variability of fatty infiltration, leading to decreased reliability. Other factors such as number of rotator cuffs repaired each year and the percentage of practice dedicated to the shoulder did not show a significant correlation with intraobserver reliability.

There are several limitations to the current study. First, the response rate to the survey acts as a weakness when determining interobserver reliability. The initial response rate was only <12%; however, the secondary response rate was excellent (93%), with 28 of the 30 respondents participating in the second survey. The short time period between surveys (2 months) could have contributed to this high secondary response rate but also provided enough time as to not become a confounding factor for intraobserver reliability. A second weakness of the study can be seen in the quality and quantity of the MRI testing images. Image variability due to slight deviations in body/coil positioning may have led to response differences in comparison with a series of images of uniform quality. To minimize this variability and maintain conformity, we performed all MRIs at our institution following a standard protocol with a 1.5 Tesla magnet. In addition, because of the nature of the study, subspecialists evaluated only 1 image from each patient to increase participation in the survey. Perhaps interobserver reliability would have increased had the entire study of axial, oblique coronal, and sagittal images been available for each respondent to review. Another limitation of our study is the lack of clinical correlation of the proposed new classification system with patient outcomes. Despite the better reliability of the new classification system, further studies will be needed to validate this classification with clinical outcomes due to the heterogeneity of rotator cuff tears and to confirm its reliability in a different cohort.

The strengths of this study include the incorporation of multiple subspecialty-trained orthopaedic surgeons. To our knowledge, this is the first study to include a large number of raters. In doing so, chance is minimized when determining κ values/reliability, ultimately increasing the power of the study. In addition, this study is the first reliability study that exclusively uses patients with surgically proven large rotator cuff tears (>3 cm); patients with smaller tears were excluded from the study. Larger rotator cuff tears are thought to be more susceptible to fatty infiltration than smaller tears. Also, larger tears with fatty infiltration are clinically the most problematic; thus, a classification for fatty infiltration should be reliable for these larger tears.[16]

Furthermore, this is the first study to our knowledge that recorded demographic data of surgeons to determine factors that correlated with intraobserver reliability. The type of fellowship was a factor that increased intraobserver reliability. Sports-trained surgeons had higher intraobserver reliability than shoulder/elbow-trained surgeons.

In conclusion, this study demonstrates that there is moderate inter- and intraobserver reliability when using the Goutallier classification. If this classification system is condensed to 3 as opposed to 5 stages, as proposed in this study, the reliability is increased substantially.

REFERENCES

1. Anderson TW, Darling DA. Asymptotic theory of certain ''goodness-of-fit'' criteria based on stochastic processes. *Ann Math Stat*. 1952;23:193-212.
2. Burkhart SS, Barth JR, Richards DP, Zlatkin MB, Larsen M. Arthroscopic repair of massive rotator cuff tears with stage 3 and 4 fatty degeneration. *Arthroscopy*. 2007;23(4):347-354.
3. 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;8(6):599-605.
4. Gerber C, Schneeberger AG, Hoppeler H, Meyer DC. Correlation of atrophy and fatty infiltration on strength and integrity of rotator cuff repairs: a study in thirteen patients. *J Shoulder Elbow Surg*. 2007;16(6):691-696.
5. 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(5):719-728.
6. 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;304:78-83.
7. Goutallier D, Postel JM, Gleyze P, Leguilloux P, Van Drissche S. Influence of cuff muscle fatty degeneration on anatomic and functional outcomes after simple suture of full-thickness tears. *J Shoulder Elbow Surg*. 2003;12(6):550-554.
8. Kenn W, Bohm D, Gohlke F, Hummer C, Kostler H, Hahn D. 2D SPLASH: a new method to determine the fatty infiltration of the rotator cuff muscles. *Eur Radiol*. 2004;14(12):2331-2336.
9. Kim HM, Dahiya N, Teefey SA, et al. Location and initiation of degenerative rotator cuff tears: an analysis of three hundred and sixty shoulders. *J Bone Joint Surg Am*. 2010;92(5):1088-1096.
10. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-174.
11. Lesage P, Maynou C, Elhage R, Boutry N, Herent S, Mestdagh H. Reproducibility of CT scan evaluation of muscular fatty degeneration: intra- and interobserver analysis of 56 shoulders presenting with a ruptured rotator cuff muscles. *Rev Chir Orthop Reparatrice Appar Mot*. 2002;88(4):359-364.
12. Melis B, DeFranco MJ, Chuinard C, Walch G. Natural history of fatty infiltration and atrophy of the supraspinatus muscle in rotator cuff tears. *Clin Orthop Relat Res*. 2010;468:1498-1505.
13. Mellado JM, Calmet J, Olona M, et al. Surgically repaired massive rotator cuff tears: MRI of tendon integrity, muscle fatty degeneration, and muscle atrophy correlated with intraoperative and clinical findings. *AJR Am J Roentgenol*. 2005;184(5):1456-1463.
14. Müller R, Büttner P. A critical discussion on intraclass correlation coefficients. *Stat Med*. 1994;13(23-24):2465-2476.
15. 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.
16. Oh JH, Kim SH, Ji HM, et al. Prognostic factors affecting anatomic outcome of rotator cuff repair and correlation with functional outcome. *Arthroscopy*. 2009;25(1):30-39.
17. Uhthoff HK, Matsumoto F, Trudel G, Himori K. Early reattachment does not reverse atrophy and fat accumulation of the supraspinatus—an experimental study in rabbits. *J Orthop Res*. 2003;21(3):386-392.
18. van de, Sande MA, Stoel BC, Obermann WR, et al. Quantitative assessment of fatty degeneration in rotator cuff muscles determined with computed tomography. *Invest Radiol*. 2005;40(5):313-319.
19. Williams MD, Ladermann A, Melis B, Barthelemy R, Walch G. Fatty infiltration of the supraspinatus: a reliability study. *J Shoulder Elbow Surg*. 2009;18(4):581-587.