Surgeon Agreement on the Presence of Pathologic Anterior Instability on Shoulder Imaging Studies

Austin M. Beason, MD, Ryan J. Koehler, MD, Rosemary A. Sanders, BA, Brooke E. Rode, BA, Travis J. Menge, MD, Kirk A. McCullough, MD, Natalie A. Glass, PhD, MHCDS, Carolyn M. Hettrich, MD, MPH, MOON Shoulder Group, and Charles L. Cox,* MD, MPH

Investigation performed at the Department of Orthopaedics and Rehabilitation, Vanderbilt University Medical Center, Nashville, Tennessee, USA

Background: In the setting of anterior shoulder instability, it is important to assess the reliability of orthopaedic surgeons to diagnose pathologic characteristics on the 2 most common imaging modalities used in clinical practice: standard plain radiographs and magnetic resonance imaging (MRI).

Purpose: To assess the intra- and interrater reliability of diagnosing pathologic characteristics associated with anterior shoulder instability using standard plain radiographs and MRI.

Study Design: Cohort study (diagnosis); Level of evidence, 3.

Methods: Patient charts at a single academic institution were reviewed for anterior shoulder instability injuries. The study included 40 sets of images (20 radiograph sets, 20 MRI series). The images, along with standardized evaluation forms, were distributed to 22 shoulder/sports medicine fellowship–trained orthopaedic surgeons over 2 points in time. Kappa values for inter- and intrarater reliability were calculated.

Results: The overall response rate was 91%. For shoulder radiographs, interrater agreement was fair to moderate for the presence of glenoid lesions ($\kappa = 0.49$), estimate of glenoid lesion surface area ($\kappa = 0.59$), presence of a Hill-Sachs lesion ($\kappa = 0.35$), and estimate of Hill-Sachs surface area ($\kappa = 0.50$). Intrarater agreement was moderate for radiographs ($\kappa = 0.48\text{-}0.57$). For shoulder MRI, interrater agreement was fair to moderate for the presence of glenoid lesions ($\kappa = 0.44$), glenoid lesion surface area ($\kappa = 0.35$), Hill-Sachs lesion ($\kappa = 0.33$), Hill-Sachs surface area ($\kappa = 0.28$), humeral head edema ($\kappa = 0.41$), and presence of a capsulolabral injury ($\kappa = 0.36$). Fair agreement was found for specific type of capsulolabral injury ($\kappa = 0.21$). Intrarater agreement for shoulder MRI was moderate for the presence of glenoid lesion ($\kappa = 0.59$), presence of a Hill-Sachs lesion ($\kappa = 0.52$), estimate of Hill-Sachs surface area ($\kappa = 0.50$), humeral head edema ($\kappa = 0.51$), and presence of a capsulolabral injury ($\kappa = 0.53$), and agreement was substantial for glenoid lesion surface area ($\kappa = 0.63$). Intrarater agreement was fair for determining the specific type of capsulolabral injury ($\kappa = 0.38$).

Conclusion: Fair to moderate agreement by surgeons was found when evaluating imaging studies for anterior shoulder instability. Agreement was similar for identifying pathologic characteristics on radiographs and MRI. There was a trend toward better agreement for the presence of glenoid-sided injury. The lowest agreement was observed for specific capsulolabral injuries.

Keywords: shoulder instability; MRI; radiographs; agreement; diagnosis

Anterior shoulder instability is the most common type of shoulder instability and is often secondary to traumatic injury to the glenohumeral joint resulting in dislocation. Although the incidence and prevalence of anterior glenohumeral instability are not well established, the incidence rate in the general population and military personnel population is estimated to be 0.08 and 1.69 per 1000 person-years, respectively.[3,10,11] Despite the high incidence of shoulder instability, little evidence is available to confirm the most reliable means of diagnosis.

The complexity of the glenohumeral joint and existence of normal anatomic variants complicate the diagnosis of pathologic shoulder instability.[13] Moreover, shoulder pathology classification is often described as a continuum of severity, making precise diagnosis difficult. Imaging studies in the form of radiographs, computed tomography (CT), and magnetic resonance imaging (MRI) are often used to supplement the history and physical examination, and

the severity of an osseous or capsulolabral injury pattern as seen on imaging may influence the initial treatment plan. Kirkley et al[4] evaluated the agreement between MRI and arthroscopy for 16 patients and found complete agreement for the presence of Hill-Sachs lesions and Bankart lesions but only fair agreement on the presence of capsular injury. Momenzadeh et al[9] also looked at the sensitivity of shoulder MRI compared with arthroscopic findings and found high sensitivity for Hill-Sachs lesions but low sensitivity for labral injury.

It is generally accepted that these imaging studies aid in treatment planning for shoulder instability, but the agreement between surgeons when interpreting these studies has not been evaluated. The purpose of this study was to determine the level of agreement between orthopaedic surgeons when interpreting traditional imaging modalities associated with anterior shoulder instability. Our hypothesis was that moderate intra- and interrater agreement will be found among shoulder/sports medicine fellowship–trained orthopaedic surgeons regarding shoulder pathologic characteristics encountered in the setting of an anterior shoulder instability event.

METHODS

Institutional review board approval was obtained for a retrospective review of patient charts with a history of an anterior shoulder instability event. Patient charts at a single academic institution were reviewed from January 1, 2005, to December 31, 2008. Patients were identified by searching International Classification of Diseases, 9th Revision, codes for anterior shoulder instability (830.00, 830.01, 830.02). Patient charts were reviewed for the availability of radiographic and MRI data and for instability-related pathologic findings as identified by the fellowship-trained musculoskeletal radiologists who initially interpreted the studies. After de-identification, these imaging studies were then reviewed by 2 surgeons (C.L.C., T.J.M.) at our institution to confirm the presence of pathologic findings. After review, we selected 40 imaging sets (20 radiograph sets and 20 MRI series) to send to raters. These sets were selected to represent a spectrum of osseous and soft tissue shoulder abnormalities associated with anterior shoulder instability, including glenoid bone loss, Hill-Sachs lesions, humeral head bone marrow edema, and specific capsulolabral injuries.

For the radiographs, a complete set included anteroposterior, scapular-Y, and axillary views, post reduction if indicated. To replicate clinical practice, we used MRI data sets with which the patients presented at the clinical visit. This included studies obtained on a variety of 1.5-T MRI machines at our institution. The use of contrast was not standard for all MRI data sets. Of the included MRI studies, 14 used contrast (10 intra-articular, 4 intravenous) and 6 were obtained without contrast. Raters were provided with coronal and axial series with T2-weighting. All images were standardized by size and transferred into PowerPoint format (Microsoft Inc). The PowerPoint file was transferred to compact discs for distribution.

We designed 2 standard evaluation forms (1 radiograph-specific and 1 MRI-specific) to allow participating surgeons to select the presence or absence of various shoulder instability abnormalities (Figures 1 and 2). Evaluators were also asked to assess the categorical extent or severity of that abnormality using their preferred measurement method.

A memory disc with images and an evaluation form for each set of images were sent to 22 orthopaedic surgeons. All recipients were shoulder/sports medicine fellowship–trained orthopaedic surgeons who had previously agreed to participate in an anterior instability imaging study and were members of the MOON Shoulder Group. Raters were assigned a number for the purpose of tracking participation, and only key study personnel were given access to the rater names and corresponding numbers. All forms were generated by use of scanning technology (TELEform Software) and labeled with a unique identification number. Approximately 6 months after raters received the first-round surveys, the images were reorganized in a new, random order on the memory disc and redistributed to the raters with the same imaging modality–specific standard evaluation forms. Data were analyzed from the surgeons who completed both rounds of the study.

Statistical Methods

Multirater kappa (κ) statistics were used to quantify both intrarater and interrater agreement among the participating orthopaedic surgeons. Kappa statistics reflect the proportion of actual agreement achieved (observed accuracy) to the potential agreement achievable by chance alone (expected accuracy). A kappa value of 0.00 represents agreement completely due to chance, and a value of 1.00 represents perfect agreement. Kappa values were interpreted by use of the definitions described by Landis and Koch[6] and are listed in Table 1.

Radiographic Study Questionnaire Form
This questionnaire includes an example page to assist reviewers in making assessments.
(Please mark one answer for each question)

REVIEWER NUMBER_____ DATE_____

RADIOGRAPH NUMBER_____

1A. IS AN OSSEOUS GLENOID LESION PRESENT?
☐ Yes (proceed to 1B)　　☐ No (proceed to 2A)

Instructions: Evaluate axillary lateral radiograph for presence or absence of bony glenoid defect?

1B. OF TOTAL GLENOID SURFACE AREA, DOES THE OSSEOUS GLENOID LESION REPRESENT
☐ < 25%　　☐ >=25%

Instructions: Evaluate axillary lateral radiograph for size of bony glenoid defect?

2A. IS THERE EVIDENCE OF A HILL-SACHS LESION?
☐ Yes (proceed to 2B)　　☐ No (stop)

Instructions: Evaluate all radiographs in series for evidence of a Hill-Sachs defect?

2B. OF TOTAL HUMERAL JOINT SURFACE, DOES THE HILL-SACHS LESIONS REPRESENT
☐ <20%　　☐ 20-40%　　☐ >40%

Instructions: Evaluate AP view size of bony glenoid defect?

Figure 1. Radiographic evaluation form.

RESULTS

Raters

A total of 22 surgeons returned the first round of evaluation forms. Of these raters, 20 surgeons completed the second round of surveys. This resulted in a 91% (20/22) total response rate.

Shoulder Radiographs

Fair to moderate intra- and interrater agreement was found on shoulder radiograph sets (Table 2). Interrater reliability was moderate for the presence of osseous glenoid lesions ($\kappa = 0.49$) and the estimate of osseous glenoid lesion surface area ($\kappa = 0.59$). When images were reevaluated by raters, intrarater agreement was moderate for the presence of glenoid lesions ($\kappa = 0.57$) and osseous glenoid lesion surface area ($\kappa = 0.57$). Interrater agreement was fair for the presence of a Hill-Sachs lesion ($\kappa = 0.35$) and moderate for the estimate of the surface area of the Hill-Sachs lesion ($\kappa = 0.50$). When examined a second time, raters showed a moderate intrarater agreement for the presence of a Hill-Sachs lesion ($\kappa = 0.48$) and estimate of Hill-Sachs lesion surface area ($\kappa = 0.53$).

Shoulder MRI

Intra- and interrater agreement was fair to moderate for most of the pathologic features evaluated with shoulder MRI (Table 3). Interrater agreement was moderate for the presence of osseous glenoid lesions ($\kappa = 0.44$) but only fair for the estimate of surface area ($\kappa = 0.35$). When the MRI series were reviewed again, intrarater agreement was moderate for the presence of osseous glenoid lesions ($\kappa = 0.59$) and substantial for the estimate of surface area ($\kappa = 0.63$). Similarly, interrater agreement was fair for the presence of a Hill-Sachs lesion ($\kappa = 0.33$) and the estimate of Hill-Sachs surface area ($\kappa = 0.28$). Intrarater agreement was again moderate for the presence of a Hill-Sachs lesion ($\kappa = 0.52$) and the estimate of surface area ($\kappa = 0.50$). When raters evaluated more detailed pathologic findings on MRI, interrater agreement was moderate for the presence of bone edema in the humeral head ($\kappa = 0.41$) and fair for the presence of a capsulolabral injury ($\kappa = 0.36$) and the specific type of capsulolabral injury ($\kappa = 0.21$). Intrarater agreement was moderate for the presence of bone edema in the humeral head ($\kappa = 0.51$) and the presence of a capsulolabral injury ($\kappa = 0.53$). Intrarater agreement was fair for determining the specific type of capsulolabral injury ($\kappa = 0.38$).

Figure 2. Magnetic resonance imaging (MRI) evaluation form.

DISCUSSION

In this study, we were able to demonstrate that orthopaedic surgeons had a variable level of agreement when interpreting radiographic and MRI studies in the setting of anterior shoulder instability. The use of shoulder imaging provides orthopaedic surgeons with a supplement to the clinical history and physical examination when making treatment decisions. We found surgeons had a fair to moderate level of agreement on both imaging modalities presented. The diagnosis of specific capsulolabral injury patterns on MRI presented a unique challenge, and we found only fair interrater agreement.

The interrater agreement in this study ranged from fair to moderate for both radiographic sets and MRI series reviewed. Overall, agreement for the presence or absence of osseous abnormality (Hill-Sachs, glenoid lesions, humeral head edema) was better than that for

TABLE 1
Degrees of Reliability Determined by κ Values
From Landis and Koch[6]

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

TABLE 2
Rater Agreement (κ Values) for Shoulder Radiographs

Pathologic Finding	Interrater Agreement	Intrarater Agreement
Osseous glenoid lesion	0.49	0.57
Glenoid lesion surface area	0.59	0.57
Hill-Sachs lesion	0.35	0.48
Hill-Sachs surface area	0.50	0.53

TABLE 3
Rater Agreement (κ Values) for Shoulder MRI[a]

Pathologic Finding	Interrater Agreement	Intrarater Agreement
Osseous glenoid lesion	0.44	0.59
Glenoid lesion surface area	0.35	0.63
Hill-Sachs lesion	0.33	0.52
Hill-Sachs surface area	0.28	0.50
Humeral head bone edema	0.41	0.51
Capsulolabral injury	0.36	0.53
Type of capsulolabral injury	0.21	0.38

[a]MRI, magnetic resonance imaging.

capsulolabral injury. This likely occurred because these were dichotomous assessments on both radiographs and MRI. We also found a trend toward better agreement for the presence of glenoid-sided abnormality. When raters evaluated the size of osseous lesions on MRI there was less agreement, which is possibly attributable to the categorical nature of this item on our evaluation form, since many of the lesions likely approached the categorical cutoff points. However, radiographic evaluation for osseous lesion size proved to be more reliable. It is possible that agreement for osseous lesion size would have varied if different categorical cutoff points had been chosen at the outset of the study. However, we chose the glenoid (25%) and humeral head (<20%, 20%-40%, >40%) values as arbitrary thresholds based on ranges previously reported in the literature that may alter surgical treatment options. It is also likely that agreement for bone loss would have been improved with incorporation of CT and 3-dimensional (3D) CT images for the reviewers, but these imaging modalities were not routinely used at our institution during standard patient evaluation.

The level of agreement we found in this study is similar to that found in several prior studies looking at the agreement between surgeons when interpreting shoulder imaging in the setting of rotator cuff disease.[1,5,7,12] Spencer et al[12] found poor to substantial agreement between 10 orthopaedic surgeons when reviewing MRI series for rotator cuff tears. Those investigators found that increased complexity and subjectivity in their classification of injury led to worse agreement between surgeons.[12] Our findings support this, as we found less agreement when surgeons were asked to specify the type of capsulolabral injuries. Overall, surgeons in our study demonstrated a trend toward more reliably identifying the presence of a capsulolabral injury compared with making a specific diagnosis, which likely represents a spectrum of injury leading to increased variability when a surgeon is asked to differentiate a glenoid articular rim disruption from an anterior labral and periosteal sleeve avulsion. Halma et al[2] identified differences of opinion for the definition of Bankart lesions and ligamentous lesions as one of the main reasons for disagreement between the radiologists and surgeon in their study. One reason for this lack of agreement may be the lack of MRI standardization. We did not standardize this study to include only MRI with contrast, but prior studies have found improved diagnosis of capsulolabral injury when MRI arthrography is performed.[8,14]

The history and physical examination findings serve as the main determinants of the definitive treatment for anterior shoulder instability. Clinicians should use imaging results as an adjunct to determine the treatment strategy, possibly incorporating CT or 3D CT when the volume of bone loss approaches thresholds that may alter the surgical plan. The definitive capsulolabral injury pattern is identified by arthroscopic evaluation and may influence the method of surgical stabilization (eg, treatment of a Bankart lesion vs a humeral avulsion of the glenohumeral ligament [HAGL] lesion), and thus imaging interpretation may not be as important for these specific patterns.

Study Limitations

We recognize several limitations to our study. The imaging used in this study was collected in a retrospective fashion at a single academic institution, possibly limiting the generalizability of our findings. The selection of the included imaging studies was not random, thus leading to incorporation of more rare findings than typically would be seen in general practice. Also, this study assessed the agreement among orthopaedic surgeons using only 1.5-T MRI scans. The inclusion of musculoskeletal radiologists and larger strength magnets would likely influence the results.

Additionally, we provided evaluators in this study with images, but the evaluators were not given the patient's history or physical examination findings. The radiographic sets and MRI series were not from the same patient, so these could not be used to supplement each other in making a diagnosis, possibly leading to lower agreement than would be observed in a clinical situation. Sagittal MRI scans, sequences beyond T2-weighted images, and CT or

3D CT scans were not included in this study, and participants were asked to quantify bone loss using the axial sequences for measurement while considering the coronal imaging findings. The additional MRI sequences and measurement options were excluded in an attempt to minimize responder burden, and thus agreement on the presence and size of an osseous lesion may be improved if these additional sequences and options are used in the clinical setting.

Also, the mix of MRI with and without contrast is a limitation of this study. However, frequently patients are referred to our clinics with MRI scans already completed. From a cost-effectiveness standpoint, it is not feasible to always obtain a second MRI with contrast if the first MRI was performed without contrast. Therefore, our aim was to look at agreement among surgeons based on a sample of images more representative of the real-world scenarios encountered in patient care. Finally, the imaging studies were initially selected based on the interpretation at the coordinating institution and assessed with a simple evaluation form created for this study based on common, previously described injury patterns. Arthroscopy was not used to confirm the exact diagnosis. Without this gold standard for diagnosis, we cannot confirm the specific presence of the underlying abnormality, which allowed for determination of the agreement but not the accuracy for the specific abnormalities. The clinical decision-making process regarding operative versus nonoperative management of anterior shoulder instability is multifaceted, with the history and physical examination serving as the primary determinants of treatment. Osseous lesions may influence the surgical approach, but the specific capsulolabral abnormality is likely more important for intraoperative treatment choices.

CONCLUSION

We found fair to moderate agreement by a group of fellowship-trained sports medicine surgeons when evaluating imaging studies in the setting of anterior shoulder instability without accompanying clinical data. Agreement was similar for identifying abnormalities on radiographs and MRI. We noted a trend toward better agreement for the presence of glenoid-sided lesions. The lowest agreement was observed for making the diagnosis of specific capsulolabral injuries based on MRI imaging alone. This suggests that isolated reliance on imaging may be limited in the diagnosis of anterior shoulder instability, and thus history and physical examination supplemented by imaging findings should serve as the primary determinants of the treatment strategy.

AUTHORS

Austin M. Beason, MD (Southern Illinois University School of Medicine, Springfield, Illinois, USA); Ryan J. Koehler, MD, Rosemary A. Sanders, BA, Brooke E. Rode, BA (Vanderbilt University Medical Center, Nashville, Tennessee, USA); Travis J. Menge, MD (SHMG Orthopedics & Sports Medicine, Grand Rapids, Michigan, USA); Kirk A. McCullough, MD (Orthopaedic & Sports Medicine Clinic of Kansas City, Kansas City, Kansas, USA); Natalie A. Glass, PhD, MHCDS (University of Iowa, Iowa City, Iowa, USA); Carolyn M. Hettrich, MD, MPH (University of Kentucky, Lexington, Kentucky, USA); Charles L. Cox, MD, MPH (Vanderbilt University Medical Center, Nashville, Tennessee, USA); MOON Shoulder Group (contributing authors): Matthew J. Bollier, MD, Brian R. Wolf, MD (University of Iowa, Iowa City, Iowa, USA); Edwin E. Spencer, MD (Knoxville Orthopedic Clinic, Knoxville, Tennessee, USA); John A. Grant, MD (University of Michigan, Ann Arbor, Michigan, USA); Julie Y. Bishop, MD, Grant L. Jones, MD, Jonathan D. Barlow, MD (The Ohio State University, Columbus, Ohio, USA); Keith M. Baumgarten, MD (Orthopedic Institute, Sioux Falls, South Dakota, USA); John D. Kelly, MD, Brian J. Sennett, MD, Milt Zgonis, MD (University of Pennsylvania, Philadelphia, Pennsylvania, USA); Joseph A. Abboud, MD, Surena Namdari, MD (Rothman Institute, Philadelphia, Pennsylvania, USA); Christina Allen, MD (UCSF Medical Center, San Francisco, California, USA); John E. Kuhn, MD, Jaron P. Sullivan, MD (Vanderbilt University Medical Center, Nashville, Tennessee, USA); Rick W. Wright, MD, Robert H. Brophy, MD, Matthew V. Smith, MD (Washington University, St Louis, Missouri, USA); Warren R. Dunn, MD (Fondren Orthopedic Group, Houston, Texas, USA).

REFERENCES

1. Grant JA, Miller BS, Jacobson JA, Morag Y, Bedi A, Carpenter JE. Intra- and inter-rater reliability of the detection of tears of the supraspinatus central tendon on MRI by shoulder surgeons. *J Shoulder Elbow Surg.* 2013;22(6):725-731.
2. Halma JJ, Eshuis R, Krebbers YM, Weits T, de Gast A. Interdisciplinary inter-observer agreement and accuracy of MR imaging of the shoulder with arthroscopic correlation. *Arch Orthop Trauma Surg.* 2012;132(3):311-320.
3. Hovelius L. Incidence of shoulder dislocation in Sweden. *Clin Orthop Relat Res.* 1982;166:127-131.
4. Kirkley A, Litchfield R, Thain L, Spouge A. Agreement between magnetic resonance imaging and arthroscopic evaluation of the shoulder joint in primary anterior dislocation of the shoulder. *Clin J Sport Med.* 2003;13(3):148-151.
5. Kuhn JE, Dunn WR, Ma B, et al. Interobserver agreement in the classification of rotator cuff tears. *Am J Sports Med.* 2007;35(3):437-441.
6. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159-174.
7. Lee CS, Davis SM, Doremus B, Kouk S, Stetson WB. Interobserver agreement in the classification of partial-thickness rotator cuff tears using the Snyder classification system. *Orthop J Sports Med.* 2016;4(9):2325967116667058.
8. Magee T. 3-T MRI of the shoulder: is MR arthrography necessary? *AJR Am J Roentgenol.* 2009;192(1):86-92.
9. Momenzadeh OR, Gerami MH, Sefidbakht S, Dehghani S. Assessment of correlation between MRI and arthroscopic pathologic findings in the shoulder joint. *Arch Bone Joint Surg.* 2015;3(4):286-290.
10. Owens BD, Dawson L, Burks R, Cameron KL. Incidence of shoulder dislocation in the United States military: demographic considerations from a high-risk population. *J Bone Joint Surg Am.* 2009;91(4):791-796.

11. Owens BD, Duffey ML, Nelson BJ, DeBerardino TM, Taylor DC, Mountcastle SB. The incidence and characteristics of shoulder instability at the United States Military Academy. *Am J Sports Med*. 2007; 35(7):1168-1173.

12. 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(1):99-103.

13. Tischer T, Vogt S, Kreuz PC, Imhoff AB. Arthroscopic anatomy, variants, and pathologic findings in shoulder instability. *Arthroscopy*. 2011;27(10):1434-1443.

14. Waldt S, Burkart A, Imhoff AB, Bruegel M, Rummeny EJ, Woertler K. Anterior shoulder instability: accuracy of MR arthrography in the classification of anteroinferior labroligamentous injuries. *Radiology*. 2005; 237(2):578-583.