

Interdisciplinary inter-observer agreement and accuracy of MR imaging of the shoulder with arthroscopic correlation

J. J. Halma · R. Eshuis · Y. M. J. Krebbers ·
T. Weits · A. de Gast

Received: 5 November 2010 / Published online: 13 August 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract

Introduction Differences between radiologists and orthopaedic surgeons in the interpretation of MR images of the shoulder joint are experienced in daily clinical practice. This study set out to evaluate the inter-observer agreement between radiologists and orthopaedic surgeons in assessing pathology on MR imaging of the shoulder joint. Also, we determined the accuracy of the observers with arthroscopy as the standard of reference.

Materials and methods Two radiologists and one orthopaedic surgeon reviewed 50 MR studies—25 conventional MR examinations and 25 MR arthrographies—of patients with shoulder complaints who had undergone MR imaging and subsequently arthroscopic surgery. The assessment was independent and blinded. All observers evaluated the MR examinations twice. Standard evaluation forms were used to score for pathology of rotator cuff, glenoid labrum, tendon of the long head of the biceps brachii and glenohumeral ligaments. The presence or absence of osteoarthritis, SLAP

lesions, Bankart lesions, Hill-Sachs lesions or impingement was also noted. Intra- and inter-observer agreement, the sensitivity and specificity were calculated. Differences in percentages of correctly diagnosed lesions were tested for significance using McNemar's test.

Results There was a poor inter-observer agreement between the orthopaedic surgeon and the radiologists in assessing Bankart lesions and ligamentous lesions. We found significant differences between the radiologists and the orthopaedic surgeon in the assessment of osteoarthritis, Hill-Sachs lesions and impingement.

Conclusion The orthopaedic surgeon and radiologists differed in their interpretation of what defines a Bankart lesion and what defines a ligamentous lesion. The orthopaedic surgeon was significantly more accurate in assessing impingement.

Keywords Inter-observer agreement · Accuracy · Orthopaedic surgeons · Radiologists · Shoulder joint · MR imaging

J. J. Halma · R. Eshuis · A. de Gast
Clinical Orthopaedic Research Center, Department of
Orthopaedics, Diaconessenhuis Hospital, Utrecht/Zeist,
Bosboomstraat 1, 3582, KE, Utrecht, The Netherlands

Present Address:

J. J. Halma
Burg. Reigerstraat 5, 3581, KJ, Utrecht, The Netherlands
e-mail: halma.jelle@gmail.com

Y. M. J. Krebbers · T. Weits
Department of Radiology, Diaconessenhuis Hospital, Utrecht/
Zeist, Bosboomstraat 1, 3582, KE, Utrecht, The Netherlands

A. de Gast (✉)
Postbus 80250, 3508, TG, Utrecht, The Netherlands
e-mail: adgast@diakhuis.nl

Introduction

The incidence of shoulder complaints is high. Rotator cuff pathology is the most common cause of shoulder pain with a reported incidence of 5–40% [8]. Instability is another commonly presented shoulder problem, mostly resulting from lesions of the capsulo-labral complex.

MR imaging is a commonly used non-invasive test for assessing lesions of the glenoid labrum and musculo-tendinous units of the rotator cuff with high accuracy [1, 2, 5, 6, 9–11, 16–18, 21, 23, 25, 26].

However, in daily clinical practice radiologists and orthopaedic surgeons frequently differ in the interpretation

of MR examinations of the shoulder. When comparing arthroscopic findings with the clinical MR report this disagreement occurs even more so. What causes this divergence and whether this is a true difference in interpretation or just a different use of terminology has not been investigated before. The aim of this study is to evaluate the inter-observer agreement between orthopaedic surgeons and radiologists in the assessment of MR examinations of the shoulder in daily clinical practice. Furthermore, we evaluate the accuracy in predicting shoulder pathology based on MR imaging of orthopaedic surgeons and radiologists with arthroscopic findings as the standard of reference.

Materials and methods

Patients

The study subjects in this investigation consisted of a group of patients with clinically suspected shoulder pathology who underwent either unenhanced MR imaging or MR arthrography and subsequently arthroscopy of the shoulder at our institution, from January 2007 to January 2010. A total of 73 patients were considered for enrolment in this study. Patients were excluded if imaging quality was impaired (due to movement-artefacts or otherwise) or if the arthroscopy report was not available. Furthermore, patients were excluded if arthroscopy was performed more than 180 days after MR imaging. After applying these exclusion criteria 65 patients remained. Of these patients 50 were selected, so that 25 had unenhanced MR imaging and 25 had MR arthrography of the shoulder joint. Mean time between imaging and arthroscopy was 76 days (median 65 days, range 22–174 days). Of the 50 patients—30 men and 20 women—that were involved in this study, ages ranged from 17 to 79 years old (mean 44, median 46) at the time of MR imaging. The spectrum of pathology in the patients ranged from no abnormalities to the presence of multiple lesions. Four patients had prior surgery of the investigated shoulder.

As shown in Table 1, the majority of patients (54%) that were included in this study presented with shoulder pain. After physical examination 34% of the patients were suspected of having subacromial impingement syndrome, in 18% a rotator cuff tear was suspected. One patient (2%) presented with frozen shoulder syndrome. 46% of the patients in this study presented with unidirectional shoulder instability. Institutional review board approval was not needed for this retrospective study.

MR imaging protocol

All MR images were obtained at our institution using a 1.5 T MRI (Siemens, either type Avanto or Espree) with a standard shoulder coil. High resolution and small field of view imaging was performed.

For unenhanced MR imaging of the shoulder the following sequences were obtained: T1-weighted turbo-spin-echo images in the axial plane and T1-weighted spin-echo sequences in the coronal–oblique plane. T2-weighted turbo-spin-echo images were acquired in the sagittal plane and T2-weighted turbo spin-echo sequence with fat suppression in the coronal–oblique plane.

MR arthrography was routinely performed with fluoroscopic guidance and in most cases a posterior approach. Intra-articular needle placement was verified with the injection of 1–5 mL Iomeprol (Iomeron® 300 mg/ml). Thereafter 0.5 mL Gadoteridol (ProHance® 2,793 g/10 ml) was diluted in 100 mL 0.9% saline solution. Of this mixture, 15–20 mL was injected in the glenohumeral joint. Following arthrography, T1-weighted turbo-spin-echo sequences with fat suppression were obtained in the axial plane and T1-weighted VIBE sequences in the coronal–oblique plane of which multiplanar reconstructions were obtained in the coronal–oblique, axial and sagittal plane. Proton density and T2-weighted turbo-spin-echo images were acquired in the sagittal plane and the same sequences with fat suppression in the coronal–oblique plane.

Table 1 Patient characteristics (*n* = 50)

Age	17–79 years old (mean 44, median 46)	
Sex	Male	Female
	30 (60%)	20 (40%)
Imaging technique	Unenhanced MR	MR arthrography
	25 (50%)	25 (50%)
Time between imaging and arthroscopy	22–174 days (mean 76, median 65)	
History of symptoms	Shoulder instability	23 (46%)
	Shoulder pain	27 (54%)
		Impingement syndrome 17 (34%)
		Rotator cuff tear 9 (18%)
		Frozen shoulder 1 (2%)

Image analysis

Images were interpreted by two radiologists (radiologist 1 and 2) and one orthopaedic surgeon, all three with vast experience in shoulder pathology. MR examinations were retrospectively reviewed and scored independently by all three observers. The observers were blinded for patients' name, date of birth and patient number. Hereby, it was made sure that none of the observers had access to the arthroscopy report or the clinical MR report. CD-ROM's with the MR images could only be read on "stand alone" computers (that were not linked to the hospital network) with standard monitor quality. Standard evaluation forms, developed by the authors, were used to score for pathology of rotator cuff, glenoid labrum, tendon of the long head of the biceps brachii (biceps tendon), labral–bicipital complex (SLAP-lesion) and glenohumeral ligaments. Furthermore, the presence or absence of glenohumeral osteoarthritis, a Bankart lesion, a Hill-Sachs lesion or impingement of the rotator cuff was noted on the evaluation forms. In case a rotator cuff lesion was found to be present, the location and size was scored on the evaluation forms. The location of a rotator cuff tear was described in terms of which muscle/tendon was affected (supraspinatus, infraspinatus, subscapularis or teres minor). A partial rotator cuff tear was categorised according to the proportion of rotator cuff thickness affected (less or more than 50%). A full thickness rotator cuff tear was categorised according to the size of the biggest gap in the coronal plane (less or more than 3 cm). Furthermore if a lesion of the glenoid labrum was scored on the evaluation forms, the location of the lesion had to be noted in terms of being superior, inferior, anterior or posterior. If a ligamentous lesion was noted to be present, the exact ligament affected (either the superior, middle or inferior glenohumeral ligament), was also noted.

The observers were not provided with instructions about specific criteria to use for interpreting the MR images. Moreover, they were specifically asked to assess to MR examinations as they would in daily practice. Still, the observers are aware of the standard criteria as established in the literature used for the assessment of labral and rotator cuff pathology [9].

All three observers assessed the 50 MR examinations twice, with a 2-week interval between the appraisal of the first and second series. The first series was assessed in a different order than the second series.

Arthroscopy

All arthroscopies were performed at the same institution, Diaconessenhuis Hospital, Utrecht, the Netherlands. 47 of the 50 arthroscopies (94%) included in this study were performed by the same experienced orthopaedic surgeon.

All two shoulder surgeons at our institution routinely perform arthroscopy using a posterior and an anterior portal consecutively, followed by subacromial bursoscopy. The evaluation of the glenohumeral joint is performed at our institution following a protocol that is concordant with the 15-point Anatomy Review as described by Snyder [19]. Subacromial bursoscopy is also performed in a standardised manner concordant with Snyder's Eight-point Bursal Anatomy Review by all our shoulder surgeons [20].

Patients clinically suspected of having subacromial impingement syndrome (positive Neer test) and who were found to have spurs around the inferior portion of the acromion or the acromioclavicular joint on plain radiographs and or arthroscopy, as well as patients with typical soft tissue changes at arthroscopy were considered having subacromial impingement. These typical changes encompass fraying of the bursal floor and rotator cuff tendon, a partial bursal-sided rotator cuff tear or a degenerative full thickness rotator cuff tear.

The arthroscopic findings were noted by the performing surgeon in the arthroscopy report, using the same systematic approach as for performing the arthroscopy itself.

Data collection

Data on the MR findings were obtained from the standard evaluation forms. Data on the arthroscopic findings were collected from the surgical records. If the rotator cuff or glenoid labrum was described to be "degenerative" in the arthroscopy report, this was regarded as negative for the scoring of rotator cuff tears or labral lesions. Furthermore, structures that were not mentioned in the arthroscopy report were assumed to be normal. The arthroscopy reports that were reviewed in this study did not contain any intra-operative photographs.

Statistical analysis

Intra- and inter-observer agreement for the presence or absence of a given lesion on MR imaging was calculated using the kappa statistic in SPSS. Kappa values were calculated manually when asymmetry of cross tables prevented the calculation by SPSS. The kappa values can be interpreted as poor ($K = 0$), slight ($K = 0.00$ – 0.20), fair ($K = 0.21$ – 0.40), moderate ($K = 0.41$ – 0.60), substantial ($K = 0.61$ – 0.80) and almost perfect agreement ($K = 0.81$ – 1.00) [15].

Accuracy of detecting pathology of the glenohumeral joint was determined for all three observers. The sensitivity and specificity for each observer were calculated per lesion type. For each reader, the percentages of correctly diagnosed lesions per lesion type, as confirmed by the arthroscopy report, were calculated. Differences in the

percentage of correct diagnoses among the observers were tested for significance using the McNemar statistic. Differences were considered significant at the 5% level ($p < 0.05$, significant). Statistical evaluations were performed using SPSS 17.0 software.

Results

As shown in Table 2, according to the arthroscopy reports of our 50 study subjects, 16 patients (32%) had rotator cuff pathology, 18 patients (36%) had labral pathology and 7 patients (14%) had a lesion of the biceps tendon. In two patients (4%), several glenohumeral ligaments were ruptured and in three patients (6%) a Hill-Sachs defect was found at arthroscopy. In nine study subjects, degenerative changes of the glenohumeral joint were found at arthroscopy of which there was found to be osteoarthritis in one patient (2%). There was impingement of the rotator cuff in 26 (52%) of our study subjects. In two subjects (4%), no abnormalities were found at arthroscopy. Of the patients with rotator cuff pathology, seven patients had a partial

thickness tear, eight patients had a full thickness tear of the rotator cuff and one patient had both. The surgical reports did not render information on the measures of the partial thickness rotator cuff tears. In six of the eight full thickness rotator cuff tears, the arthroscopy did not give conclusive information in the measure of retraction in the coronal plane.

Table 3 summarises the inter-observer agreement among radiologists 1 and 2 and the orthopaedic surgeon per lesion type assessed on all MR examinations. Most notable finding listed here is the poor inter-observer agreement among the radiologists and the orthopaedic surgeon in assessing whether a Bankart lesion is present or absent or “not interpretable”. The inter-observer agreement on the presence or absence of impingement among the radiologists is moderate. However, the agreement between the radiologists and the orthopaedic surgeon is fair to slight. Furthermore, among the radiologists there is a fair inter-observer agreement in assessing a Hill-Sachs lesion. There is also a fair agreement between radiologist 1 and the orthopaedic surgeon. However, radiologist 2 and the orthopaedic surgeon have a poor agreement on the presence or absence of a Hill-Sachs lesion.

As shown in Table 4, when assessing a Bankart lesion on enhanced MR images alone there is a perfect agreement among radiologists (kappa 1.00). The agreement between the radiologists and the orthopaedic surgeon is poor (kappa 0.0). Also, there is a perfect agreement among radiologists in assessing a labral lesions but the agreement between radiologists and the orthopaedic surgeon is fair. The radiologists have substantial agreement on the presence or absence of a lesion of the glenohumeral ligaments, whereas the agreement between radiologists and the orthopaedic surgeon is slight.

The percentages of correctly diagnosed lesions confirmed by the arthroscopy report are summarised in

Table 2 Incidence of pathology as found at arthroscopy ($n = 50$)

Pathology	<i>N</i> (% of total)
Rotator cuff lesions (7 partial thickness, 8 full thickness)	16 (32)
Labral lesions	18 (36)
Long head of biceps tendon lesions	7 (14)
GHL lesions	2 (4)
Hill-Sachs lesions	3 (6)
Glenohumeral osteoarthritis	1 (2)
Impingement	26 (52)
No abnormalities	2 (4)

Table 3 Inter-observer agreement among radiologist 1 and 2 and the orthopaedic surgeon per lesion type assessed on all MR-examinations ($n = 50$)

Pathology	Rad 1 vs. Rad 2 (kappa)	Rad 1 vs. OrthSurg (kappa)	Rad 2 vs. OrthSurg (kappa)
Rotator cuff lesion (present/absent)	0.74	0.63	0.62
Rotator cuff tear type (full/partial/both)	0.58	0.48	0.33
Partial RC tear location	0.49	0.28	0.17
Partial RC tear grade	0.49	0.19	0.15
Full thickness RC tear location	0.35	0.61	0.18
Full thickness RC tear degree of retraction	0.27	0.41	0.19
Biceps tendon	0.23	0.37	0.26
Bankart lesion	-0.03	-0.07	-0.02
Hill-Sachs lesion	0.21	0.31	-0.01
Osteoarthritis	0.12	0.15	0.23
Impingement	0.43	0.29	0.15
Cause of impingement	0.38	0.22	0.17

Rad 1 radiologist 1,
Rad 2 radiologist 2,
OrthSurg orthopaedic
surgeon, RC rotator cuff

Table 4 Inter-observer agreement among radiologist 1 and 2 and the orthopaedic surgeon per lesion type assessed on enhanced MR images ($n = 25$)

Pathology	Rad 1 vs. Rad 2 (kappa)	Rad 1 vs. OrthSurg (kappa)	Rad 2 vs. OrthSurg (kappa)
Labral lesion	1.00	0.32	0.23
Labral lesion location	0.50	0.16	0.02
SLAP lesion	0.36	0.28	0.44
Ligamentous lesion	0.61	0.10	0.01
Ligamentous lesion location	0.61	0.13	0.04
Bankart lesion	1.0	0.0	0.0

Rad 1 radiologist 1, *Rad 2* radiologist 2, *OrthSurg* orthopaedic surgeon, *SLAP* superior labral anterior-to-posterior lesion

Tables 5, 6. As marked in Table 5, radiologist 2 has the highest percentage of correctly diagnosed Hill-Sachs lesions, which is significantly higher than the other two observers ($p < 0.05$; McNemar's test). Radiologist 1 is significantly less accurate in assessing osteoarthritis compared to the other observers. The orthopaedic surgeon has the highest percentage in correctly diagnosing impingement and is significantly more accurate than radiologist 2. Furthermore, the orthopaedic surgeon is significantly the most accurate observer in determining the cause of impingement. The remaining findings listed in Tables 5 and 6 are not significantly different among the observers.

In Tables 7 and 8, the intra-observer agreement of the radiologists and the orthopaedic surgeon are presented. Intra-observer agreement of the radiologists is almost perfect in assessing tears of the rotator cuff and glenoid labrum, whereas this is substantial in the orthopaedic surgeon. Intra-observer consistency of the orthopaedic surgeon is only slight in assessing ligamentous lesions, although the radiologists have moderate and substantial consistency. The orthopaedic surgeon is more consistent in predicting impingement than the radiologists.

Tables 9 and 10 contains the sensitivity and specificity of each observer in predicting pathology of the glenohumeral joint.

Discussion

We assessed the inter-observer agreement among two radiologists and an orthopaedic surgeon in predicting pathology of the glenohumeral joint on MR examinations. We found a wide range of inter-observer agreements between the radiologists and the orthopaedic surgeon, varying per lesion type.

The inter-observer agreement between the radiologists and the orthopaedic surgeon was remarkably less than the

agreement among the radiologists in assessing impingement, ligamentous lesions, Bankart lesions and labral lesions. These findings indicate that the radiologists and the orthopaedic surgeon have a different interpretation of what defines these lesion types. The orthopaedic surgeon was significantly more accurate than the radiologists in assessing impingement. An explanation for this difference is that orthopaedic surgeons commonly need to determine the cause of impingement when preparing for operative treatment. Radiologists do not routinely assess these features when evaluating an MRI of the shoulder joint. On enquiry, the orthopaedic surgeon had a more dynamic approach to assessing impingement; combining the findings of tendinopathy in one plane and spurs around acromion or acromioclavicular joint in another plane to make the diagnosis.

The radiologists and the orthopaedic surgeon also differed in their interpretation of what defines a ligamentous lesion. The orthopaedic surgeon scored elongation of the glenohumeral ligaments as positive for a (chronic) ligamentous lesion; however, this was regarded as negative in the calculation of the sensitivity, specificity and the percentage of correct diagnoses. The radiologists did not score the elongation of glenohumeral ligaments as positive for ligamentous lesions. None of the observers detected the presence of a ligamentous lesion correctly resulting in a sensitivity of 0.0%. This is the result of the low prevalence of ligamentous lesions in the study subjects.

The results of the observers determining which one of the glenohumeral ligaments was affected are useless since none of the observers detected the presence of a ligament lesion correctly in the first place.

In this study, the radiologists agreed perfectly that no Bankart lesion was present at all, whereas the orthopaedic surgeon found several Bankart lesions to be present on enhanced MR studies. This remarkable difference resulted from a different interpretation of what defines a Bankart lesion. The radiologists only scored for bony Bankart lesions whereas the orthopaedic surgeon scored classic Bankart lesions of the anterior glenoid labrum as well. In the arthroscopy reports a Bankart lesion was defined a classic Bankart lesion. However, these differences did not result in a significant difference considering the percentage of correctly diagnosed lesions. This can be explained by the high false positive rate of the orthopaedic surgeon.

The interpretation of what defines a labral lesion was not really different among the observers. However, the orthopaedic surgeon had a higher sensitivity in detecting labral lesions than the radiologists. Also the percentage of correct diagnoses was the highest for the orthopaedic surgeon, but not significantly different from the radiologists.

Table 5 Percentage of correctly diagnosed lesions confirmed by arthroscopy for radiologist 1 and 2 and the orthopaedic surgeon (in predicting pathology of the glenohumeral joint on MR examinations) ($n = 50$)

Pathology	Correct diagnosis (%)		
	Radiologist 1	Radiologist 2	Orthopaedic surgeon
Rotator cuff lesion (present/absent)	88	84	86
Rotator cuff tear type (full/partial/both)	76	70	70
Partial thickness RC tear location	80	72	78
Partial thickness RC tear grade	83.3	76.2	88.1
Full thickness RC tear location	81.3	83.7	75.5
Full thickness RC tear degree of retraction	86	93	79.1
Biceps tendon	85.7	88.0	78.0
Bankart lesion	73.3	80.0	82.0
Hill-Sachs lesion	69.4	90.0*	70.0
		$p < 0.05$ vs. rad 1 and OrthSurg	
Osteoarthritis	71.4	86.0*	94.0*
		$p < 0.05$ vs. rad 1	$p < 0.05$ vs. rad 1
Impingement	60.0	48.0	76*
			$p < 0.05$ vs. rad 2
Cause of impingement	44.0	46.0	64.0*
			$p < 0.05$ vs. rad 1 and rad 2

Rad 1 radiologist 1, Rad 2 radiologist 2, OrthoSurg orthopaedic surgeon, RC rotator cuff

* McNemar's test for paired proportions; $p < 0.05$, significant

Table 6 Percentage of correctly diagnosed lesions confirmed by arthroscopy report per lesion type assessed on enhanced MR images ($n = 25$)

Pathology	Correct diagnosis (%)		
	Radiologist 1	Radiologist 2	Orthopaedic surgeon
Labral lesion	60.0	60.0	72.0
Labral lesion location	29.2	44.0	24.0
SLAP lesion	72.0	68.0	68.0
Ligamentous lesion	70.8	79.2	50.0
Ligament lesion location	70.8	79.2	50.0
Bankart lesion	68.0	68.0	68.0

SLAP superior labral anterior-to-posterior lesion

Table 7 Intra-observer agreement for radiologist 1 and 2 and the orthopaedic surgeon per lesion type assessed on all MR examinations ($n = 50$)

Pathology	Radiologist 1 (kappa)	Radiologist 2 (kappa)	Orthopaedic surgeon (kappa)
Rotator cuff lesion (present/absent)	0.91	1.00	0.76
Rotator cuff tear type (full/partial/both)	0.72	0.91	0.61
Partial thickness RC tear location	0.47	0.86	0.31
Partial thickness RC tear grade	0.37	0.90	0.27
Full thickness RC tear location	0.73	0.79	0.73
Full thickness RC tear degree of retraction	0.57	0.59	0.71
Biceps tendon	-0.02	0.31	0.38
Bankart lesion	0.27	0.00	0.60
Hill-Sachs lesion	0.31	0.79	0.73
Osteoarthritis	0.39	0.92	0.79
Impingement	0.46	0.45	0.76
Cause of impingement	0.43	0.26	0.74

RC rotator cuff

Table 8 Intra-observer agreement for radiologist 1 and 2 and the orthopaedic surgeon per lesion type assessed on enhanced MR images ($n = 25$)

Pathology	Radiologist 1 (kappa)	Radiologist 2 (kappa)	Orthopaedic surgeon (kappa)
Labral lesion	0.84	1.00	0.74
Labral lesion location	0.58	0.94	0.38
SLAP lesion	0.72	0.50	0.52
Ligamentous lesion	0.57	0.72	0.04
Ligamentous lesion location	NA	0.72	0.11
Bankart lesion	1.00	1.00	0.43

SLAP superior labral anterior-to-posterior lesion, NA not applicable, unable to calculate kappa since radiologist 1 did not score any ligamentous lesion to be present in the second series

No significant differences were found in the percentages of correctly diagnosed Bankart lesions and ligamentous lesions.

The inter-observer agreement among radiologists was in the same range as the agreement between the orthopaedic surgeon and the radiologists in detecting osteoarthritis, lesions of the rotator cuff, lesions of the biceps tendon, a Hill-Sachs lesion and a SLAP lesion. These findings indicate that the observers have the same interpretation of the definition of these types of pathology. There were no significant differences in the percentages of correctly diagnosed rotator cuff tears, lesions of the biceps tendon and SLAP lesions. However, radiologist 1 was significantly less accurate than the other observers in detecting cases of osteoarthritis. This was because of a high false positive rate resulting in low specificity. Radiologist 2 was significantly

Table 9 Sensitivity, specificity of radiologist 1 and 2 and the orthopaedic surgeon in predicting pathology of the glenohumeral joint on MR examinations ($n = 50$)

Pathology	Radiologist 1		Radiologist 2		Orthopaedic surgeon	
	Sens (%)	Spec (%)	Sens (%)	Spec (%)	Sens (%)	Spec (%)
Rotator cuff lesion (present/absent)	87.5	88.2	75.0	88.2	93.8	82.4
Rotator cuff tear type (full/partial/both)	50.0	88.2	31.3	88.2	43.8	82.4
Partial thickness RC tear location	62.5	83.3	50.0	76.2	25.0	88.1
Partial thickness RC tear grade	–	83.3	–	76.2	–	88.1
Full thickness RC tear location	28.6	90.2	25.0	97.6	50.0	80.5
Full thickness RC tear degree of retraction	0.0	90.2	0.0	97.6	50.0	80.5
Biceps tendon	14.3	97.6	42.9	95.3	28.6	86.0
Bankart lesion	0.0	91.7	0.0	97.6	77.8	82.9
Hill-Sachs lesion	0.0	73.9	0.0	95.7	33.3	72.3
Osteoarthritis	100	70.8	100	85.7	0	95.9
Impingement	30.8	91.7	7.7	91.7	61.5	91.7
Cause of impingement	0.0	91.7	3.8	91.7	38.5	91.7

Sens sensitivity, Spec specificity, RC rotator cuff

Table 10 Sensitivity and specificity of radiologist 1 and 2 and the orthopaedic surgeon in predicting pathology of the glenohumeral joint on enhanced MR images ($n = 25$)

Pathology	Radiologist 1		Radiologist 2		Orthopaedic surgeon	
	Sens (%)	Spec (%)	Sens (%)	Spec (%)	Sens (%)	Spec (%)
Labral lesion	70.6	37.5	70.6	37.5	88.2	37.5
Labral lesion location	18.8	37.5	47.1	37.5	17.6	37.5
SLAP lesion	33.3	84.2	16.7	84.2	33.3	78.9
Ligamentous lesion	0.0	77.3	0.0	82.6	0.0	54.5
Ligament lesion location	0.0	77.3	0.0	82.6	0.0	54.5
Bankart lesion	0.0	100	0.0	100	58.8	87.5

Sens sensitivity, Spec specificity, SLAP superior labral anterior-to-posterior lesion

more accurate than the other observers in determining the presence or absence of a Hill-Sachs lesion. The latter was completely due to the high specificity, because radiologist 2 did not once predict the presence of a Hill-Sachs lesion correctly.

In this study, the sensitivity and specificity of the observers in assessing rotator cuff tears, Bankart lesions, lesions of the biceps tendon and SLAP lesions are in the same range as reported in most studies [1, 4, 6, 9–11, 17, 21, 23, 25]. Furthermore, the sensitivity and specificity of the observers in detecting osteoarthritis is in the same range of the results of Guntern et al. [5] on predicting humeral and glenoid cartilage lesions.

However, the sensitivity of the observers in this study in detecting impingement is notably lower than the findings of Iannotti et al. [9]. In the study by Iannotti et al. [9], however, the MR studies were retrospectively reviewed after the findings at surgery proofed the first assessment to be false.

All three observers had remarkably lower sensitivity in detecting Hill-Sachs lesions than reported in other studies [4, 21]. This can be explained by the low prevalence of Hill-Sachs lesions in the arthroscopy reports. It is likely that small Hill-Sachs lesions were not mentioned in the arthroscopy report because these do not need surgical treatment. The prevalence of Hill-Sachs lesions in the arthroscopy reports is therefore an underestimation of the true prevalence in the study subjects. This led to a lower true positive rate and consequently the relatively low sensitivity of the observers in assessing Hill-Sachs lesions.

Also in assessing labral lesions specificity of all three observers was surprisingly lower than in most studies [1, 2, 4, 9, 21]. However, these findings are difficult to compare because of the small number of MR arthrographies in this study. All three observers scored very low percentages of correct diagnoses in determining the location of labral lesions. This is probably due to nonspecific terminology used in the arthroscopy report, noting for example only that a labral tear is anterior. The observers, however, scored several of these labral lesions as being antero-inferior. These cases were considered “false” in the calculation of sensitivity and specificity and the percentage of correctly diagnosed lesions.

Concordant with the recent literature [21], the sensitivity of the observers in detecting lesions of the biceps tendon in this study is low. This can be explained by the technical shortcomings of MR imaging in general. The arched course of the tendon of the long head of the biceps migrating through every plane makes it typically susceptible to the “blind spots” of MR imaging, such as the partial volume effect and magic angle artefacts. Furthermore, anomalous origins of the long head of the biceps brachii can also trouble the evaluation on MR images, however, rarely encountered in daily practice [7, 12, 13, 24].

Intra-observer agreement

Internal consistency of our observers varied from almost perfect-to-slight. The radiologists were more consistent than the orthopaedic surgeon in detecting labral and rotator cuff pathology. The orthopaedic surgeon was most consistent in assessing impingement. Overall, all three observers were most consistent in predicting pathology in which they scored the highest percentages of correct diagnoses. The latter indicates that the differences we found in this study are reproducible.

Limitations

The most important limitations of our study lay in its retrospective character. First of all, the use of arthroscopy reports as the standard of reference is precarious. The quality of these surgical reports is generally moderate, which makes them susceptible to interpretation. Moreover, structures not mentioned in the surgical report were assumed to be normal. Second, only patients with an indication for arthroscopic surgery were included in this study. This is only a certain proportion of all patients who undergo MR examination of the shoulder. The assessment of pathology in these patients is of vital importance to determine whether or not arthroscopic surgery is indicated.

The fact that patients who had prior surgery of the investigated shoulder were included in our study could have blurred the assessment of these MR images. Also, the differentiation between physiologic degenerative change and pathologic degeneration was complicated in this study. The observers could not determine the clinical relevance of the finding of degenerative change, because they were blinded for patient’s age and clinical information. In daily clinical practice, the assessing radiologist and orthopaedic surgeon do have access to the patient’s medical history.

Intra-articular injection for arthrography was performed in most but not in all cases with a posterior approach. According to preference of the performing radiologist, an anterior approach could also have been used. This made it difficult for the observers to differentiate between contrast spillage through the puncture hole and (partial thickness) rotator cuff tears.

The use of MR arthrography in detecting partial lesions of the rotator cuff has been advocated [3, 16, 23]. In our institution, we routinely use MR imaging to assess rotator cuff pathology. We believe that due to technical improvement of image quality in modern MR machines that these unenhanced MR studies are also capable of partial rotator cuff lesions. Also, the use of the ABER view can be considered to improve the visualisation of anterior glenoid labrum, the labral–bicipital complex and the rotator cuff [2, 11, 14, 22]. Unfortunately an appropriate

shoulder coil that is needed for this arm position is not available yet at our institution.

The above-mentioned factors play a roll in the diagnostic performance of the observers. However, this affected all three observers equally and did not affect their differences.

Daily clinical practice

We used a study design that resembled daily clinical practice. The observers were not provided with particular criteria to use for the interpretation of the MR examinations. Therefore, the differences we found in this study are the differences in the observers' own interpretation.

At our institution, MR arthrography is routinely performed when labral or ligamentous pathology is clinically suspected. Unenhanced MR imaging is routinely performed when the patient is clinically suspected of having shoulder pathology other than labral or ligamentous lesions. This helped the observers to look for certain pathology in particular and is therefore similar to daily clinical practice in which the radiologist and orthopaedic surgeon have access to the clinical information.

Conclusion

Radiologists and the orthopaedic surgeon at our institution differed in predicting some but not all types of pathology of the glenohumeral joint on MR imaging. The biggest differences were found in the assessment of Hill-Sachs lesions, osteoarthritis and impingement. The orthopaedic surgeon performed better than the radiologists in the assessment of impingement. Furthermore, the differences in the interpretation of what defines Bankart lesions and ligamentous lesions were found. It is important for orthopaedic surgeons and radiologists to become aware of these differences to obtain mutual understanding and to learn from each others expertise.

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Applegate GR, Hewitt M, Snyder SJ, Watson E, Kwak S, Resnick D (2004) Chronic labral tears: value of magnetic resonance

- arthrography in evaluating the glenoid labrum and labral–bicipital complex. *Arthroscopy* 20(9):959–963
2. Cvitanic O, Tirman PF, Feller JF, Bost FW, Minter J, Carroll KW (1997) Using abduction and external rotation of the shoulder to increase the sensitivity of MR arthrography in revealing tears of the anterior glenoid labrum. *AJR Am J Roentgenol* 169(3):837–844
3. Flannigan B, Kursunoglu-Brahme S, Snyder S, Karzel R, Del Pizzo W, Resnick D (1990) MR arthrography of the shoulder: comparison with conventional MR imaging. *AJR Am J Roentgenol* 155(4):829–832
4. van Grinsven S, Kesselring FO, van Wassenaeer-van Hall HN, Lindeboom R, Lucas C, van Loon CJ (2007) MR arthrography of traumatic anterior shoulder lesions showed modest reproducibility and accuracy when evaluated under clinical circumstances. *Arch Orthop Trauma Surg* 127(1):11–17
5. Guntern DV, Pfirrmann CW, Schmid MR, Zanetti M, Binkert CA, Schneeberger AG, Hodler J (2003) Articular cartilage lesions of the glenohumeral joint: diagnostic effectiveness of MR arthrography and prevalence in patients with subacromial impingement syndrome. *Radiology* 226(1):165–170
6. Holzapfel K, Waldt S, Bruegel M, Paul J, Heinrich P, Imhoff AB, Rummeny EJ, Woertler K (2010) Inter- and intraobserver variability of MR arthrography in the detection and classification of superior labral anterior posterior (SLAP) lesions: evaluation in 78 cases with arthroscopic correlation. *Eur Radiol* 20(3):666–673
7. Hyman JL, Warren RF (2001) Extra-articular origin of biceps brachii. *Arthroscopy* 17(7):E29
8. Iannotti JP, Williams GR (2007) Disorders of the shoulder: diagnosis and management volume 1. In: Iannotti JP, Codsi MJ, Lafosse L, Flatow EL (eds) *Management of rotator cuff disease: intact and repairable cuff*, 2nd edn. Lippincott Williams & Wilkins, Philadelphia, pp 53–100
9. Iannotti JP, Zlatkin MB, Esterhai JL, Kressel HY, Dalinka MK, Spindler KP (1991) Magnetic resonance imaging of the shoulder. Sensitivity, specificity, and predictive value. *J Bone Joint Surg Am* 73:17–29
10. Jee WH, McCauley TR, Katz LD, Matheny JM, Ruwe PA, Daigneault JP (2001) Superior labral anterior posterior (SLAP) lesions of the glenoid labrum: reliability and accuracy of MR arthrography for diagnosis. *Radiology* 218(1):127–132
11. Jung JY, Jee WH, Chun HJ, Ahn MI, Kim YS (2010) Magnetic resonance arthrography including ABER view in diagnosing partial-thickness tears of the rotator cuff: accuracy, and inter- and intra-observer agreements. *Acta Radiol* 51(2):194–201
12. Kim KC, Rhee KJ, Shin HD (2009) A long head of the biceps tendon confluent with the intra-articular rotator cuff: arthroscopic and MR arthrographic findings. *Arch Orthop Trauma Surg* 129(3):311–314
13. Kim KC, Rhee KJ, Shin HD, Kim YM (2008) Biceps long head tendon revisited: a case report of split tendon arising from single origin. *Arch Orthop Trauma Surg* 128(5):495–498
14. Kwak SM, Brown RR, Trudell D, Resnick D (1998) Glenohumeral joint: comparison of shoulder positions at MR arthrography. *Radiology* 208(2):375–380
15. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33(1):159–174
16. Magee T, Williams D, Mani N (2004) Shoulder MR arthrography: which patient group benefits most? *AJR Am J Roentgenol* 183(4):969–974
17. Schreinemachers SA, van der Hulst VP, Willems WJ, Bipat S, van der Woude HJ (2009) Detection of partial-thickness supraspinatus tendon tears: is a single direct MR arthrography series in ABER position as accurate as conventional MR arthrography? *Skeletal Radiol* 38(10):967–975

18. Schreinemachers SA, van der Hulst VP, Willems WJ, Bipat S, van der Woude HJ (2009) Is a single direct MR arthrography series in ABER position as accurate in detecting anteroinferior labroligamentous lesions as conventional MR arthrography? *Skeletal Radiol* 38(7):675–683
19. Snyder SJ (2003) Shoulder arthroscopy. In: *Diagnostic arthroscopy of the shoulder: normal anatomy and variation*, 2nd edn. Lippincott Williams & Wilkins, Philadelphia, pp 22–38
20. Snyder SJ (2003) Shoulder arthroscopy. In: *Diagnostic bursoscopy*, 2nd edn. Lippincott Williams & Wilkins, Philadelphia, pp 39–45
21. Theodoropoulos JS, Andreisek G, Harvey EJ, Wolin P (2010) Magnetic resonance imaging and magnetic resonance arthrography of the shoulder: dependence on the level of training of the performing radiologist for diagnostic accuracy. *Skeletal Radiol* 39(7):661–667
22. Tirman PF, Bost FW, Steinbach LS, Mall JC, Peterfy CG, Sampson TG, Sheehan WE, Forbes JR, Genant HK (1994) MR arthrographic depiction of tears of the rotator cuff: benefit of abduction and external rotation of the arm. *Radiology* 192(3):851–856
23. Toyoda H, Ito Y, Tomo H, Nakao Y, Koike T, Takaoka K (2005) Evaluation of rotator cuff tears with magnetic resonance arthrography. *Clin Orthop Relat Res* 439:109–115
24. Vangsness CT Jr, Jorgenson SS, Watson T, Johnson DL (1994) The origin of the long head of the biceps from the scapula and glenoid labrum. An anatomical study of 100 shoulders. *J Bone Joint Surg Br* 76(6):951–954
25. Waldt S, Bruegel M, Mueller D, Holzapfel K, Imhoff AB, Rummeny EJ, Woertler K (2007) Rotator cuff tears: assessment with MR arthrography in 275 patients with arthroscopic correlation. *Eur Radiol* 17(2):491–498
26. Waldt S, Burkart A, Imhoff AB, Bruegel M, Rummeny EJ, Woertler K (2005) Anterior shoulder instability: accuracy of MR arthrography in the classification of anteroinferior labroligamentous injuries. *Radiology* 237(2):578–583