E ORIGINAL ARTICLE

Shoulder & Elbow

Shoulder & Elbow
2020, Vol. 12(4) 284–293

DOI: 10.1177/1758573218821151
journals.sagepub.com/home/sel
SAGE

Interobserver reliability of the classification of capitellar osteochondritis dissecans using magnetic resonance imaging

Rens Bexkens[1,2], F. Joseph Simeone[3], Denise Eygendaal[1,4], Michel PJ van den Bekerom[5] and Luke S Oh[1]; on behalf of Shoulder and Elbow Platform*

Abstract

Aim: (1) To determine the interobserver reliability of magnetic resonance classifications and lesion instability criteria for capitellar osteochondritis dissecans lesions and (2) to assess differences in reliability between subgroups.

Methods: Magnetic resonance images of 20 patients with capitellar osteochondritis dissecans were reviewed by 33 observers, 18 orthopaedic surgeons and 15 musculoskeletal radiologists. Observers were asked to classify the osteochondritis dissecans according to classifications developed by Hepple, Dipaola/Nelson, Itsubo, as well as to apply the lesion instability criteria of DeSmet/Kijowski and Satake. Interobserver agreement was calculated using the multirater kappa (k) coefficient.

Results: Interobserver agreement ranged from slight to fair: Hepple ($k = 0.23$); Dipaola/Nelson ($k = 0.19$); Itsubo ($k = 0.18$); DeSmet/Kijowksi ($k = 0.16$); Satake ($k = 0.12$). When classifications/instability criteria were dichotomized into either a stable or unstable osteochondritis dissecans, there was more agreement for Hepple ($k = 0.52$; $p = .002$), Dipaola/Nelson ($k = 0.38$; $p = .015$), DeSmet/Kijowski ($k = 0.42$; $p = .001$) and Satake ($k = 0.41$; $p < .001$). Overall, agreement was not associated with the number of years in practice or the number of osteochondritis dissecans cases encountered per year ($p > .05$).

Conclusion: One should be cautious when assigning grades using magnetic resonance classifications for capitellar osteochondritis dissecans. When making treatment decisions, one should rather use relatively simple distinctions (e.g. stable versus unstable osteochondritis dissecans; lateral wall intact versus not intact), as these are more reliable.

Keywords

Osteochondritis dissecans, capitellum, magnetic resonance imaging, classification system, interobserver, reliability

Date received: 11th September 2018; accepted: 17th November 2018

Introduction

Treatment strategies and operative planning for osteochondritis dissecans (OCD) lesions of the capitellum are based on stability, size and location of the lesion, in addition to the severity of symptoms and capitellar

[1]Department of Orthopaedic Surgery, Sports Medicine Service, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA
[2]Department of Orthopaedic Surgery, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands
[3]Department of Radiology Division of Musculoskeletal Imaging and Intervention, Massachusetts General Hospital, Boston, MA, USA
[4]Department of Orthopaedic Surgery, Amphia Hospital, Breda, The Netherlands
[5]Department of Orthopaedic Surgery, Shoulder and Elbow Unit, Amsterdam, The Netherlands
*Names of the collaborators are given at the end of the article.

Corresponding author:
Rens Bexkens, Department of Orthopaedic Surgery, Sports Medicine Service, Massachusetts General Hospital, Harvard Medical School, 175 Cambridge Street, Boston, MA 02114, USA.
Email: rensbexkens@gmail.com

physis status, among others.[1–3] Non-operative treatment is advocated for a stable OCD (i.e. intact cartilage) in the setting of an open capitellar growth plate.[4] A surgical approach is indicated in an unstable OCD.[3,5,6] An OCD is considered unstable if magnetic resonance (MR) images demonstrate discontinuity of the cartilage, a high signal intensity interface between the fragments and their bed or articular defects.[7–9] Arthroscopic debridement with bone marrow stimulation[3,5,10,11] or fragment fixation[6,12] may lead to satisfactory outcomes in an unstable OCD without involvement of the lateral margin of the capitellar wall. More invasive treatment by means of osteochondral autologous transplantation is suggested in large (>10 mm), unstable lesions that involve the lateral wall of the capitellum.[13–17]

Various imaging modalities have been used to characterize OCD of the capitellum including radiography, ultrasonography and computed tomography (CT), but most commonly MR imaging is performed.[7,9,18–21] Both CT and MR have shown to correlate well with intraoperative findings[7,9]; however, classifying the stage of an OCD on CT images has been shown to be interpreted inconsistently among surgeons specialized in upper extremity injuries.[22] Because surgical decision making and preoperative planning highly depend on lesion stability, size and location of the lesion, it is important to know if physicians interpret MR images regarding these characteristics in a consistent manner.

The goals of this study were: (1) to determine the interobserver reliability of existing MR classification systems and lesion instability criteria for capitellar OCD lesions; (2) to assess differences in interobserver reliability between subgroups (e.g. years in practice, number of OCD cases per year, specialty).

Materials and methods

This study was approved by our institutional review board (protocol no. 2009P001019/MGH).

Physicians from different continents were invited to participate in this interobserver study via an invitation e-mail that included a study description. Invited were orthopaedic surgeons who were fellowship trained in shoulder and elbow injuries and/or sports-related injuries. Invited as well were radiologists who were fellowship trained in musculoskeletal imaging. Invitations were sent only to physicians who were known by at least one of the authors.

Participating physicians (i.e. observers) were asked to review MRs of 20 patients who were selected from our retrospective database. MR selection was performed by one of the authors who is a musculoskeletal fellowship trained radiologist (FJS), using the Kijowski/DeSmet lesion instability criteria.[8,23] We sought to select a representative variety of OCDs. Accordingly, we selected 12 MRs with an unstable OCD (Figure 1), 6 MRs with a stable OCD (Figure 2) and 2 patients with an unremarkable MR. In these two patients, an MR was performed because of ongoing pain localized at the radio-capitellar joint; however, MR images demonstrated no abnormalities. Not more than 20 MRs were selected because reviewing 20 MRs according to multiple classifications/criteria is time consuming for the observers.[24] Selecting more MRs would have resulted in fewer observers completing the study.

The mean age of patients at the time of MR was 15.4 years (range, 11 to 17), including 11 males and nine females.

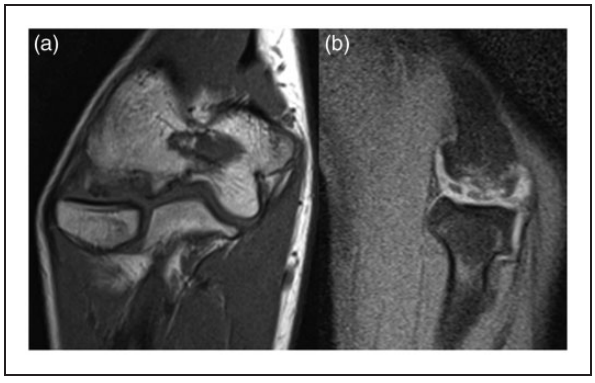

Figure 1. Images of an *unstable* OCD in the left elbow of a 15-year-old male patient. (a) Coronal T1 and (b) sagittal PDFS (proton density fat suppressed) images from MR of the left elbow showing articular surface collapse with fluid undercutting a cortical ossific fragment on the sagittal image. Mild surrounding bone marrow edema in addition to cartilage irregularity and loss are also seen.

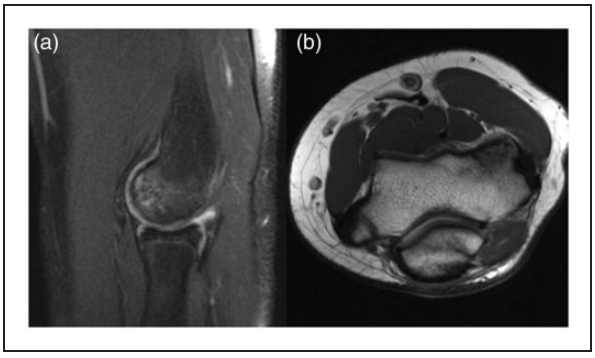

Figure 2. Images of a *stable* OCD in the left elbow of a 13-year-old male patient. (a) Sagittal proton density fat saturated and (b) axial proton density images from MR of the left elbow showing subchondral bone marrow edema of the capitellum with intact overlying cortical margin, lack of fluid signal undercutting the cortex or cartilage, and no cystic change in the capitellum.

MRs which were obtained locally were performed on a 1.5T or 3T scanner, using standard departmental protocol including proton density (PD) axial, T2FS (fat suppression) axial, T2FS coronal, T1 coronal, T2 GRE (gradient recalled echo) coronal and PDFS sagittal pulse sequences. As a tertiary referral center, we often receive and interpret MRs from outside institutions with variable pulse sequences, vendors and magnet strengths. All MRs, either obtained locally or from an outside institution, included fat-saturated, fluid-sensitive sequences (T2FS or PDFS) in the coronal and sagittal planes for adequate assessment of the capitellum. MRs with intra-articular contrast were not selected.

One of the authors (RB) not involved in patient care removed all identifying information from the MR images and uploaded the Digital and Communications in Medicine (DICOM) files to a web-based study platform (www.shoulderelbowplatform.com). Observers evaluated MR images using a built-in, web-based DICOM viewer and could adjust brightness, contrast, window leveling, zoom and measure distance. All questions related to one case had to be completed to proceed to the next case. Observers completed the study at their own pace on various computers if needed.

Upon website login, observers were asked about demographics and professional information: sex (male or female), location of practice (Europe, North America, South America, Asia, other), specialty (musculoskeletal fellowship trained radiologist or surgeon's specialty), years in practice (0–5, 6–10, 11–15, 16–20, >20), number of capitellar OCD cases per year (0–5, 6–10, 11–15, 16–20, >20). Subsequently, observers were asked to review all 20 MRs. For each MR, observers were asked to classify the stage of the OCD according to three existing classifications (Hepple,[25] Dipaola/Nelson[18,26] and Itsubo[7]), as well as to apply the criteria for lesion instability as defined by DeSmet/Kijowski[8,23] and Satake[9] (Table 1). Itsubo et al. have developed an accurate classification system to stage capitellar OCD.[7] Satake et al. have defined instability criteria for capitellar OCD and reported substantial surgical correlation.[9] DeSmet/Kijowski et al. have developed instability criteria for knee and ankle OCD with high surgical correlation, which were later confirmed in patients with capitellar OCD.[8,23] Dipaola/Nelson et al. reported a similar correlation of their proposed staging system in knee and ankle OCD.[18,26] Lastly, the Hepple classification is the most commonly used tool to stage OCD of the talus.[25] The aforementioned classification systems/criteria are frequently used tools to characterize an OCD involving the elbow, knee or ankle. This list is intended to be as inclusive as possible; however, it is not exhaustive. For observers to complete this

Table 1. MR classification systems and lesion instability criteria for capitellar OCD.

Hepple classification	
1	Articular cartilage damage only (stable lesion)
2a	Cartilage injury with underlying fracture and surrounding bony edema (unstable lesion)
2b	Cartilage injury with underlying fracture, but without surrounding bony edema (unstable lesion)
3	Detached but non-displaced fragment (unstable lesion)
4	Detached and displaced fragment (unstable lesion)
5	Subchondral cyst formation (unstable lesion)
Dipaola/Nelson classification	
1	Thickening of articular cartilage and low signal changes (stable lesion)
2	Articular cartilage breached, low signal rim behind fragment (unstable lesion)
3	Articular cartilage breached, high signal changes behind fragment indicating synovial fluid between fragment and underlying subchondral bone (unstable lesion)
4	Loose body (unstable lesion)
Itsubo classification	
1	Normally shaped capitellum with several spotted areas of high signal intensity that is lower than that of the cartilage (stable lesion)
2	As with stage 1 but with several spotted areas of higher intensity than that of the cartilage (stable lesion)
3	As with stage 2 but with both discontinuity and non-circularity of the chondral surface signal of the capitellum and no high signal interface apparent between the lesion and the floor (unstable lesion)
4	Lesion separated by a high-intensity line in comparison with cartilage (unstable lesion)
5	Capitellar lesion displaced from the floor or defect of the capitellar lesion noted (unstable lesion)

(continued)

Table 1. Continued

DeSmet/Kijowksi instability criteria	
-	A thin ill-defined or well-defined line of high signal intensity at the interface between the lesion and the underlying bone (unstable lesion)
-	A discrete round area of high signal intensity beneath the lesion indicating a cyst (unstable lesion)
-	A focal defect in the articular surface of the lesion (unstable lesion)
-	A high signal intensity line through the articular cartilage and subchondral bone plate into the lesion (unstable lesion)
Satake instability criteria	
-	Irregular contours of the articular surface as a low signal abnormality (unstable lesion)
-	Articular defect of the capitellum as a high signal abnormality (unstable lesion)
-	High signal intensity interface (compared with bone) between fragments and their bed (unstable lesion)
-	A high signal intensity line through the articular cartilage (unstable lesion)

MR: magnetic resonance; OCD: osteochondritis dissecans.

interobserver study using five classification systems/criteria is time consuming. Adding more classification systems would have led to fewer observers completing the study. Observers were also asked to assess lesion size (largest diameter among three planes; 1–5, 6–10, 11–15, >15 mm) and involvement of the lateral capitellar wall (i.e. is the lateral cartilage margin intact or not intact?). After each question, observers were asked if they were certain about their choice on a scale from 1 to 4.

Statistical analysis

Agreement among observers was calculated using the multirater kappa (k) coefficient as described by Siegel and Castellan. Point estimates and two-sided 95% confidence intervals (CIs) were calculated as well.[22,27–29] The multirater kappa is a commonly used statistic to describe chance-corrected agreement in interobserver studies. A value of 0 indicates no agreement beyond chance alone.[30,31] A value of 0.01 to 0.20 is defined as slight agreement; 0.21 to 0.40, fair agreement; 0.41 to 0.60, moderate agreement; 0.61 to 0.80, substantial agreement; and more than 0.80, near-perfect agreement.[30,31] Using a Z-test, differences in agreement between specific subgroups (e.g. years in practice, number of OCD cases per year, specialty) were analyzed.[22,27,29] Statistical analysis was performed with the use of Stata 12.0 (StataCorp LP, College Station, TX, USA).

Results

Participants

Thirty-three observers, 18 orthopaedic surgeons and 15 radiologists, completed the interobserver study on the online platform (Table 2).

Interobserver agreement

Interobserver agreement for the classification systems ranged from slight to fair: Hepple (fair, $k = 0.23$), Dipaola/Nelson (slight, $k = 0.19$) and Itsubo (slight, $k = 0.18$) (Table 3). There was slight agreement for the lesion instability criteria as developed by DeSmet/Kijowski ($k = 0.16$) and Satake ($k = 0.12$).

When observers' responses for classification systems were dichotomized into either a stable or unstable OCD, there was more agreement for the Hepple (moderate, $k = 0.52$; $p = .002$) and Dipaola/Nelson classification (fair, $k = 0.38$; $p = .015$). Similarly, when observers' responses for instability criteria were dichotomized into either a stable OCD (i.e. none of criteria present) or unstable OCD (i.e. one or more criteria present), agreement significantly improved for DeSmet/Kijowski (moderate, $k = 0.42$; $p = .001$) and Satake (moderate, $k = 0.41$; $p < .001$).

Interobserver agreement was fair for lesion size assessment ($k = 0.24$); agreement improved to moderate when lesion size was dichotomized into lesions $\leq$10 mm versus >10 mm ($k = 0.41$; $p < .001$) (Table 3). Agreement was slight with regard to involvement of the lateral capitellar wall ($k = 0.16$).

Observers' confidence about their responses ranged from 2.5 to 2.7 (range, 1 to 4) (Table 3). None of the classification systems or instability criteria demonstrated significantly higher or lower confidence levels compared to the others ($p > .05$).

Factors associated with interobserver agreement

Besides for the Hepple classification, there was no difference in agreement between observers in practice 10 years or fewer in comparison to observers more than 10 years in practice (Table 4). Interobserver agreement was not associated with the number of OCD cases per year (0–10 versus >10) or location of practice

Table 2. Observer demographics (n = 33).

Demographic	No. (%)
Sex	
Male	28 (85)
Female	5 (15)
Area	
North America	19 (58)
Europe	9 (27)
South America	4 (12)
Asia	1 (3)
Specialty	
Musculoskeletal radiology	15 (46)
Orthopaedic surgery	18 (54)
Shoulder and elbow	7 (21)
Shoulder and elbow, hand and wrist	3 (9)
Shoulder and elbow, traumatology	2 (6)
Sports medicine	2 (6)
Shoulder and elbow, sports medicine	1 (3)
Shoulder and elbow, sports medicine, hand and wrist	1 (3)
Shoulder and elbow, sports medicine, hand and wrist, traumatology	1 (3)
Sports medicine, traumatology	1 (3)
Years in practice	
0–5	14 (43)
6–10	5 (15)
11–15	5 (15)
16–20	7 (21)
>20	2 (6)
Capitellar osteochondritis dissecans cases per year	
0–5	16 (49)
6–10	7 (21)
11–15	4 (12)
16–20	2 (6)
>20	4 (12)

(North America versus Europe) (Tables 5 and 6). For two classification systems (Hepple and Itsubo) and lesion size assessment, there was more agreement among radiologists (fair) than surgeons (slight; $p < .05$) (Table 7).

Discussion

The present investigation is the first that evaluated the interobserver reliability of existing MR classifications among a large group of physicians. Overall, there was limited interobserver reliability for classifications and lesion instability criteria, as well as for lesion size assessment and lateral capitellar wall involvement. Reliability significantly improved (fair to moderate) when classification systems and instability criteria were simplified (i.e. stable lesion versus unstable lesion). Similarly, reliability significantly improved (moderate) when OCD size was subdivided into small (≤10 mm) and large lesions (>10 mm). Overall, reliability was independent of the number of years in practice, as well as independent of the number of capitellar OCD cases a physician encountered per year.

The interobserver reliability for existing classification systems was lower in our investigation compared to two previous studies.[7,32] Itsubo et al. reported an intraclass coefficient, of their own developed classification, that ranged from 0.82 to 0.88 (i.e. good to excellent reliability).[7] This was based on three orthopaedic surgeons who evaluated 52 MRs of capitellar OCD. The fact that there was substantially more agreement in their group may be the result of comprehensive experience in using its own classification since 2006.[7] This discordance may also be due to the fact that much more observers were included in the present study. Ellerman et al. determined the agreement for the Dipaola/Nelson classification based on MRs of knee OCD.[32] The authors reported moderate agreement among two musculoskeletal-trained radiologists and one musculoskeletal radiology fellow, compared to slight agreement in our investigation concerning capitellar OCD. Higher reliability reported by the authors may be due to the fact that the Dipaola/Nelson classification was originally developed for knee and talar OCD.[18,26]

Interestingly, our results showed substantially more agreement when classification systems[7,25] and instability criteria[8,9,23] were simplified. In other words, there is more consistency among observers in determining whether the OCD was either stable or unstable (fair to moderate) rather than classifying its specific stage (slight to fair). This finding suggests that one should be cautious when assigning grades using MR classifications. When making treatment decisions, one should rather use simplified distinctions (e.g. stable versus

Table 3. Interobserver reliability of characterization of capitellar osteochondritis dissecans using MR imaging.

	Category	Kappa	95% CI	Confidence level (1–4)	Category if dichotomized	Kappa if dichotomized	95% CI if dichotomized
Hepple	Fair	0.23	0.19–0.29	2.7	Moderate[a]	0.52	0.36–0.71
Dipaola/Nelson	Slight	0.19	0.14–0.25	2.7	Fair[a]	0.38	0.26–0.55
Itsubo	Slight	0.18	0.12–0.25	2.6	Fair	0.30	0.19–0.43
DeSmet/Kijowski	Slight	0.16	0.11–0.21	2.6	Moderate[a]	0.42	0.28–0.56
Satake	Slight	0.12	0.072–0.17	2.5	Moderate[a]	0.41	0.28–0.54
Lesion size	Fair	0.24	0.18–0.30	2.7	Moderate[b]	0.41	0.33–0.51
Involvement of the lateral capitellar wall	Slight	0.16	0.098–0.24	2.8	–	–	–

MR: magnetic resonance; CI: confidence interval.
[a]Significantly more agreement when observers' responses were dichotomized: stable OCD versus unstable OCD.
[b]Significantly more agreement when observers' responses were dichotomized: lesions ≤10 mm versus >10 mm.

Table 4. Interobserver agreement by years in practice: 0–10 years versus >10 years.

	0–10 years (n = 19)			>10 years (n = 14)		
	Category	Kappa	95% CI	Category	Kappa	95% CI
Hepple	Fair[a]	0.31	0.24–0.40	Slight	0.16	0.13–0.20
Dipaola/Nelson	Fair	0.23	0.16–0.31	Slight	0.16	0.12–0.21
Itsubo	Fair	0.24	0.17–0.30	Slight	0.14	0.093–0.20
DeSmet/Kijowski	Fair	0.20	0.14–0.25	Slight	0.14	0.094–0.20
Satake	Slight	0.14	0.077–0.20	Slight	0.091	0.055–0.14
Lesion size	Fair	0.31	0.23–0.40	Slight	0.19	0.13–0.26
Involvement of the lateral capitellar wall	Fair	0.23	0.14–0.35	Slight	0.12	0.064–0.20

CI: confidence interval.
[a]Significantly more agreement among observers 0–10 years in practice than observers >10 years in practice.

unstable OCD; lateral wall intact versus not intact), as these seem more reliable. The finding that relatively simple distinctions result in more reliability compared to classifications consisting of multiple subgroups is in line with previous studies who determined the reliability of the classification of proximal femur fractures.[29,33]

Overall, interobserver reliability did not differ between experienced and less experienced physicians: neither more years in medical practice nor a higher number of OCD cases per year were associated with more agreement ($p > .05$). The fact that agreement did not differ between experienced and less experienced physicians is consistent with two studies investigating the reliability of fracture characteristics on radiographs.[29,34] It could be that the learning curve of assigning grades or measuring the width of an OCD is more steep in the first few years of practice and subsequently reaches a plateau after a certain point in time. Or it may be that assigning grades or measuring the width of an OCD is not experience dependent. Interestingly, only for the Hepple classification, we found more agreement among observers up to 10 years in practice compared to observers >10 years in practice. We hypothesize that this is the result of a relative large number of radiologists in the first group (11 radiologists) compared to the last (3 radiologists).

Table 5. Interobserver agreement by capitellar OCD cases per year: 0–10 cases versus >10 cases.

	0–10 cases (n = 23)			>10 cases (n = 10)		
	Category	Kappa	95% CI	Category	Kappa	95% CI
Hepple	Fair	0.27	0.23–0.34	Slight	0.19	0.16–0.24
Dipaola/Nelson	Fair	0.26	0.18–0.36	Slight	0.14	0.11–0.20
Itsubo	Fair	0.22	0.15–0.30	Slight	0.15	0.10–0.21
DeSmet/Kijowski	Fair	0.20	0.15–0.27	Slight	0.12	0.079–0.17
Satake	Slight	0.14	0.089–0.20	Slight	0.094	0.057–0.14
Lesion size	Fair	0.27	0.20–0.34	Fair	0.21	0.14–0.29
Involvement of the lateral capitellar wall	Fair	0.21	0.14–0.30	Slight	0.14	0.083–0.21

CI: confidence interval.

Table 6. Interobserver agreement by location of practice: North America versus Europe.

	North America (n = 19)			Europe (n = 9)		
	Category	Kappa	95% CI	Category	Kappa	95% CI
Hepple	Fair	0.25	0.19–0.33	Slight	0.15	0.10–0.21
Dipaola/Nelson	Fair	0.21	0.15–0.30	Slight	0.17	0.10–0.27
Itsubo	Fair	0.20	0.14–0.29	Slight	0.16	0.11–0.23
DeSmet/Kijowski	Slight	0.17	0.12–0.23	Slight	0.16	0.11–0.23
Satake	Slight	0.12	0.071–0.18	Slight	0.091	0.045–0.16
Lesion size	Fair	0.34	0.25–0.46	Fair	0.24	0.17–0.33
Involvement of the lateral capitellar wall	Slight	0.18	0.11–0.27	Slight	0.15	0.068–0.29

CI: confidence interval.

Subgroup analysis demonstrated that radiologists interpreted two classification systems (including Hepple) and lesion size more consistent than surgeons. This indicates that, although agreement is still limited (fair), consulting a musculoskeletal radiologist should be part of routine clinical care in patients suspected for capitellar OCD.

One of the strengths of this study is that it investigated the interobserver reliability MR classifications among a large group of observers (n = 33), which allowed subgroup analysis. Furthermore, this is the first study that compared agreement among orthopaedic surgeons and musculoskeletal radiologists in such a large group. Lastly, all available classification systems were evaluated simultaneously by the same group of observers. However, the findings of this study should be interpreted by considering some limitations. First, observers did not have any additional training regarding the classification systems and instability criteria. Observers who were not familiar with these may have been more consistent if they had some form of training.[28,35] For instance, the use of an image atlas at the time of assigning grades may have been helpful. Second, neither patient history nor physical examination was given, whereas treatment decisions in the orthopaedic practice are based on the whole patient instead of solely MR imaging. Although we aimed to determine the helpfulness of MR without any potential bias, adding this information would have led to more agreement. Third, most reviewers were not familiar with the web-based

Table 7. Interobserver agreement by specialty: orthopaedic surgeons versus musculoskeletal radiologists.

	Surgeons (n = 18)			Radiologists (n = 15)		
	Category	Kappa	95% CI	Category	Kappa	95% CI
Hepple	Slight	0.16	0.13–0.22	Fair[a]	0.34	0.27–0.42
Dipaola/Nelson	Slight	0.15	0.11–0.22	Fair	0.27	0.19–0.36
Itsubo	Slight	0.13	0.089–0.19	Fair[a]	0.30	0.22–0.38
DeSmet/Kijowski	Slight	0.15	0.10–0.22	Fair	0.21	0.15–0.26
Satake	Slight	0.086	0.045–0.13	Slight	0.16	0.092–0.22
Lesion size	Slight	0.17	0.11–0.24	Fair[a]	0.35	0.26–0.44
Involvement of the lateral capitellar wall	Slight	0.13	0.072–0.22	Fair	0.22	0.13–0.33

CI: confidence interval.
[a]Significantly more agreement among radiologists than surgeons.

DICOM viewer which, though adequate for the purposes of this study, was not a fully functioning PACS workstation and lacked certain functionality such as the ability to cross-reference images, leading to difficulty navigating the cases and decreasing confidence.

The present study highlights the need to develop and test the most relevant MR classifications. Improved diagnostic MR protocols and training may lead to more consistency among physicians and ultimately to classifications that are reliable enough to be used for prognostic and therapeutic studies. Also, when comparing various studies, it is important that we use reliable distinctions to classify an OCD to ensure that patient characteristics are similar between studies.

Conclusions

This investigation adds to a growing body of evidence indicating that relatively simple distinctions on MR images are more reliable. One should be cautious when assigning grades using MR classifications in the assessment of capitellar OCD. When making treatment decisions, one should rather use simplified distinctions (e.g. stable versus unstable OCD; lateral wall intact versus not intact), among other factors such as the severity of symptoms and capitellar physis status.

Collaborators

Patzer T., Ovesen J., Oliviveira A.M., Meehan T.M., Bryan Jr. R.G., Lenobel S., Simeone F.J., Godoy I.R., Taneja A.K., Lambers Heerspink F.O., Porcellini G., Duncan S.F., Patino J.M., Marinelli A., Barco R., Bhatia D.N., Bilsel K., Chang C.Y., van Eck C.F., Kaar S.G., Palmer W.E., Shafritz A.B., Torriani M., Wong T.T., Oppenheimer J.D., Caputo A.E., Waryasz G.R., Arrigoni, P., Freehill M.T., van Bergen C.J., Lin D.J., Shahid K.R., and Vicentini J.

IRB approval

Massachusetts General Hospital IRB; protocol #: 2009P001019/MGH.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship and/or publication of this article: One author (RB) receives support from small grants to cover expenses for living and housing during research activities: amount less than USD 10,000 from International Society of Arthroscopy, Knee, Surgery and Orthopaedic Sports Mendicine (ISAKOS) Foundation (San Ramon, USA), amount less than USD 10,000 from Marti-Keuning Eckhart Foundation (Lunteren, Netherlands), amount less than USD 10,000 from Hendrik Muller Foundation, and an amount less than USD 10,000 from Anna Foundation (Oegstgeest, Netherlands).

ORCID iD

Rens Bexkens http://orcid.org/0000-0001-6993-0941
Michel PJ van den Bekerom

References

1. Takahara M, Mura N, Sasaki J, et al. Classification, treatment, and outcome of osteochondritis dissecans of the

humeral capitellum. *J Bone Joint Surg Am* 2007; 89: 1205–1214.
2. Kolmodin J and Saluan P. Osteochondritis dissecans of the humeral capitellum: the significance of lesion location. *Orthop J Sports Med* 2014; 2: 2325967114530840.
3. Bexkens R, van den Ende KIM, Ogink PT, et al. Clinical outcome after arthroscopic debridement and microfracture for osteochondritis dissecans of the capitellum. *Am J Sports Med* 2017; 45: 2312–2318.
4. Mihara K, Tsutsui H, Nishinaka N, et al. Nonoperative treatment for osteochondritis dissecans of the capitellum. *Am J Sports Med* 2009; 37: 298–304.
5. Miyake J and Masatomi T. Arthroscopic debridement of the humeral capitellum for osteochondritis dissecans: radiographic and clinical outcomes. *J Hand Surg* 2011; 36: 1333–1338.
6. Uchida S, Utsunomiya H, Taketa T, et al. Arthroscopic fragment fixation using hydroxyapatite/poly-L-lactate acid thread pins for treating elbow osteochondritis dissecans. *Am J Sports Med* 2015; 43: 1057–1065.
7. Itsubo T, Murakami N, Uemura K, et al. Magnetic resonance imaging staging to evaluate the stability of capitellar osteochondritis dissecans lesions. *Am J Sports Med* 2014; 42: 1972–1977.
8. Kijowski R and De Smet AA. MRI findings of osteochondritis dissecans of the capitellum with surgical correlation. *Am J Roentgenol* 2005; 185: 1453–1459.
9. Satake H, Takahara M, Harada M, et al. Preoperative imaging criteria for unstable osteochondritis dissecans of the capitellum. *Clin Orthop Relat Res* 2013; 471: 1137–1143.
10. Wulf CA, Stone RM, Giveans MR, et al. Magnetic resonance imaging after arthroscopic microfracture of capitellar osteochondritis dissecans. *Am J Sports Med* 2012; 40: 2549–2556.
11. Bexkens R, van Bergen CJA, van den Bekerom MPJ, et al. Decreased defect size and partial restoration of subchondral bone on computed tomography after arthroscopic debridement and microfracture for osteochondritis dissecans of the capitellum. *Am J Sports Med* 2018; 46: 2954–2959.
12. Hennrikus WP, Miller PE, Micheli LJ, et al. Internal fixation of unstable in situ osteochondritis dissecans lesions of the capitellum. *J Pediatr Orthop* 2015; 35: 467–473.
13. Maruyama M, Takahara M, Harada M, et al. Outcomes of an open autologous osteochondral plug graft for capitellar osteochondritis dissecans: time to return to sports. *Am J Sports Med* 2014; 42: 2122–2127.
14. Iwasaki N, Kato H, Ishikawa J, et al. Autologous osteochondral mosaicplasty for capitellar osteochondritis dissecans in teenaged patients. *Am J Sports Med* 2006; 34: 1233–1239.
15. Lyons ML, Werner BC, Gluck JS, et al. Osteochondral autograft plug transfer for treatment of osteochondritis dissecans of the capitellum in adolescent athletes. *J Shoulder Elb Surg* 2015; 24: 1098–1105.
16. Bexkens R, Ogink PT, Doornberg JN, et al. Donor-site morbidity after osteochondral autologous transplantation for osteochondritis dissecans of the capitellum: a systematic review and meta-analysis. *Knee Surg Sports Traumatol Arthrosc* 2017; 25: 2237–2246.
17. Bexkens R, van den Bekerom MPJ, Eygendaal D, et al. Topographic analysis of 2 alternative donor sites of the ipsilateral elbow in the treatment of capitellar osteochondritis dissecans. *Arthroscopy* 2018; 34: 2087–2093.
18. Nelson DW, DiPaola J, Colville M, et al. Osteochondritis dissecans of the talus and knee: prospective comparison of MR and arthroscopic classifications. *J Comput Assist Tomogr* 1990; 14: 804–808.
19. Kida Y, Morihara T, Kotoura Y, et al. Prevalence and clinical characteristics of osteochondritis dissecans of the humeral capitellum among adolescent baseball players. *Am J Sports Med* 2014; 42: 1963–1971.
20. Minami M, Nakashita K, Ishii S, et al. Twenty-five cases of osteochondritis dissecans of the elbow. *Rinsho Seikei Geka* 1979; 14: 805–810.
21. Bexkens R, Oosterhoff JH, Tsai TY, et al. Osteochondritis dissecans of the capitellum: lesion size and pattern analysis using quantitative 3-dimensional computed tomography and mapping technique. *J Shoulder Elb Surg* 2017; 26: 1629–1635.
22. Claessen FM, van den Ende KI, Doornberg JN, et al. Osteochondritis dissecans of the humeral capitellum: reliability of four classification systems using radiographs and computed tomography. *J Shoulder Elb Surg* 2015; 24: 1613–1618.
23. De Smet AA, Ilahi OA and Graf BK. Reassessment of the MR criteria for stability of osteochondritis dissecans in the knee and ankle. *Skeletal Radiol* 1996; 25: 159–163.
24. van Leeuwen WF, Janssen SJ, Guitton TG, et al. Interobserver agreement in diagnosing early-stage Kienböck disease on radiographs and magnetic resonance imaging. *Hand* 2017; 12: 573–578.
25. Hepple S, Winson IG and Glew D. Osteochondral lesions of the talus: a revised classification. *Foot Ankle Int* 1999; 20: 789–793.
26. Dipaola JD, Nelson DW and Colville MR. Characterizing osteochondral lesions by magnetic resonance imaging. *Arthroscopy* 1991; 7: 101–104.
27. Guitton TG and Ring D. Interobserver reliability of radial head fracture classification: two-dimensional compared with three-dimensional CT. *J Bone Joint Surg Am* 2011; 93: 2015–2021.
28. Mellema JJ, Kuntz MT, Guitton TG, et al. The effect of two factors on interobserver reliability for proximal humeral fractures. *J Amer Acad Orthop Surg* 2017; 25: 69–76.
29. Crijns TJ, Janssen SJ, Davis JT, et al. Reliability of the classification of proximal femur fractures: does clinical experience matter? *Injury* 2018; 49: 819–823.
30. Posner KL, Sampson PD, Caplan RA, et al. Measuring interrater reliability among multiple raters: an example of methods for nominal data. *Stat Med* 1990; 9: 1103–1115.
31. Landis JR and Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159–174.
32. Ellermann JM, Donald B, Rohr S, et al. Magnetic resonance imaging of osteochondritis dissecans: validation study for the ICRS classification system. *Acad Radiol* 2016; 23: 724–729.

33. Jin WJ, Dai LY, Cui YM, et al. Reliability of classification systems for intertrochanteric fractures of the proximal femur in experienced orthopaedic surgeons. *Injury* 2005; 36: 858–861.
34. Janssen SJ, Molleman J, Guitton TG, et al. What middle phalanx base fracture characteristics are most reliable and useful for surgical decision-making? *Clin Orthop Relat Res* 2015; 473: 3943–3950.
35. Buijze GA, Guitton TG, van Dijk CN, et al. Training improves interobserver reliability for the diagnosis of scaphoid fracture displacement. *Clin Orthop Relat Res* 2012; 470: 2029–2034.