Int. J. Oral Maxillofac. Surg. 2019; xxx: xxx–xxx
https://doi.org/10.1016/j.ijom.2020.10.012, available online at https://www.sciencedirect.com

International Journal of Oral & Maxillofacial Surgery

Clinical Paper, TMJ

The agreement between magnetic resonance imaging and arthroscopic findings in temporomandibular joint disorders

P.-J. Verhelst[1,2], K. Vervaeke[1,2], K. Orhan[2,3], B. Lund[4,5,6], D. Benchimol[6], W. Coucke[2], F. Van der Cruyssen[1,2], A. De Laat[1,7], C. Politis[1,2], R. Jacobs[1,2,6]

[1]Department of Oral and Maxillofacial Surgery, University Hospitals Leuven, Leuven, Belgium; [2]OMFS IMPATH Research Group, Department of Imaging and Pathology, Faculty of Medicine, KU Leuven, Leuven, Belgium; [3]Department of Dentomaxillofacial Radiology, Faculty of Dentistry, Ankara University, Ankara, Turkey; [4]Department of Clinical Dentistry, Faculty of Medicine, University of Bergen, Bergen, Norway; [5]Department of Oral and Maxillofacial Surgery, Haukeland University Hospital, Bergen, Norway; [6]Department of Dental Medicine, Karolinska Institutet, Stockholm, Sweden; [7]Department of Oral Health Sciences, KU Leuven and Department of Dentistry, University Hospitals, Leuven, Belgium

P.-J. Verhelst, K. Vervaeke, K. Orhan, B. Lund, D. Benchimol, W. Coucke, F. Van der Cruyssen, A. De Laat, C. Politis, R. Jacobs: The agreement between magnetic resonance imaging and arthroscopic findings in temporomandibular joint disorders. Int. J. Oral Maxillofac. Surg. 2019; xxx: xxx–xxx. © 2020 International Association of Oral and Maxillofacial Surgeons. Published by Elsevier Ltd. All rights reserved.

Abstract. The aim of this retrospective observational study was to assess the potential agreement between independent magnetic resonance imaging (MRI) and arthroscopic findings and their respective contributions to a final diagnosis in patients with refractory temporomandibular joint disorders. Two dentomaxillofacial radiologists and two oral and maxillofacial surgeons scored 50 joints. All observers, who were blinded to additional clinical information, used a specific scoring form and selected one or more diagnostic labels. Agreement between MRI and arthroscopy and their contributions to the final diagnosis were assessed as primary outcomes using Fleiss' kappa. Intra-modality agreement and the correlation between signal intensity ratio (SIR) measurements on MRI and synovitis grading on arthroscopy were assessed as secondary outcomes. Agreement between MRI and arthroscopy was poor. A fair level of agreement was only reached for reduction capacity of the disc and disc perforation. Arthroscopic diagnostic labels matched better with the final diagnosis, suggesting a bigger contribution to that diagnosis. Higher SIR measurements correlated with higher synovitis grading scores for the retrodiscal tissue and the posterior band of the disc. Intra-modality agreement was better in arthroscopy. When blinded to clinical information, arthroscopy and MRI observations can lead to different conclusions. The diagnostic outcomes of both examinations should be considered and integrated into a final diagnosis.

Key words: temporomandibular joint; temporomandibular joint disorders; arthroscopy; magnetic resonance imaging; diagnostic imaging.

Accepted for publication 30 October 2020

Introduction

The temporomandibular joint (TMJ) can be affected by a wide range of conditions, of which disc disorders and degenerative joint disease are some of the most prevalent. They can cause pain and dysfunction, and can mostly be treated with non-invasive measures such as behavioural modification, splint therapy, home treatment and physical therapy, and non-steroidal anti-inflammatory drugs[1,2]. A subgroup

 © 2020 International Association of Oral and Maxillofacial Surgeons. Published by Elsevier Ltd. All rights reserved.

Please cite this article in press as: Verhelst PJ, et al. The agreement between magnetic resonance imaging and arthroscopic findings in temporomandibular joint disorders, *Int J Oral Maxillofac Surg* (2020), https://doi.org/10.1016/j.ijom.2020.10.012

of these patients have a condition that is refractory to conservative treatment and needs more invasive measures. In these patients, a lavage of the upper joint compartment is often the first invasive treatment of choice[1,3].

The goal of lavage of the upper joint compartment is to rinse out inflammatory mediators[3,4]. While this effect is achieved by arthrocentesis alone, the introduction of TMJ arthroscopy in 1975 yielded some additional advantages[5]. First, visual confirmation of entry into the upper joint compartment was made possible. Second, therapeutic interventions such as lasering, lysis of adhesions, and direct disc manipulations were made possible. Finally, a visual diagnostic assessment of the upper joint compartment could now be performed, thereby providing additional diagnostic information alongside the clinical examination, history-taking, and radiological studies.

Regarding radiological studies of the TMJ, magnetic resonance imaging (MRI) provides vital information for diagnosis due to its ability to visualize the soft tissues of the joint[6,7]. The location, morphology, and integrity of the disc are evaluated in open- and closed-mouth positions. The status of the retrodiscal tissue, attached muscles, and overall inflammation are also assessed. New advances have made it possible to provide a thorough assessment of the bony and cartilaginous structures in the joint, which could make (cone beam) computed tomography ((CB)CT) evaluations theoretically obsolete in the future[8,9].

Since most of the patients with refractory disease undergo MRI as well as arthroscopy, the upper joint compartment is evaluated twice. However, the two modalities evaluate the joint through different techniques and their findings are interpreted by different specialists, being surgeons for arthroscopy and radiologists for MRI. Furthermore, the available clinical information may introduce bias in their observations. This raises two questions. First, when blinded from clinical information, do the surgeon and radiologist agree on their findings? Second, what is the contribution of their findings in reaching a final diagnosis?

The available literature provides limited data on the agreement between arthroscopic and MRI assessment in TMJ disorders (TMD)[10–15]. To answer the two questions raised, a retrospective observational cohort study was set up. The aim of this study was to assess the agreement between independent MRI and arthroscopic findings and their respective contributions to reaching a final diagnosis in patients with TMD refractory to conservative therapy.

Materials and methods

Patient selection

This single-centre retrospective observational cohort study received ethical approval from the local ethics committee (mp012487). The cases of all patients with a TMD who underwent arthroscopy between October 2013 and October 2018 in the Department of Oral and Maxillofacial Surgery of the University Hospitals of Leuven, Belgium, were reviewed for eligibility. Inclusion criteria were the availability of an intraoperative video recording and a preoperative MRI study. Exclusion criteria were prior TMJ surgery, MRI without the open mouth sequence, and a period of longer than 6 months between the preoperative MRI and arthroscopy. All MRI sequence types were allowed, as long as the TMJ was available for examination in both open- and closed-mouth positions. MRI examinations were not always performed at the same imaging centre, which caused variation in the scanning sequences. However, T1 and T2 sequences were always available. The same surgeon performed the arthroscopy procedure for all patients. A standardized diagnostic sweep of the upper joint compartment was performed every time using a 1.9-mm and 30°-approach arthroscope with two 2.7-mm outer protective cannulas. High definition video recordings of the procedure were available for all included patients.

Although the selection was performed at the patient level, the affected joint itself was the unit of observation. The age and sex of the patient, side of the affected joint, main patient complaint, and final diagnosis were extracted from the patient medical files. The final diagnosis was defined as the result of the integrated findings from the clinical examination according to the Diagnostic Criteria for Temporomandibular Disorders (DC/TMD) guidelines[16], radiological examinations ((CB)CT, panoramic radiographs, and MRI), and arthroscopic information, as stated in the patient's medical file. This final diagnosis was converted to one or more of the study diagnostic labels: anterior disc displacement without reduction (ADD − R), anterior disc displacement with reduction (ADD + R), degenerative joint disease (DJD), inflammatory arthritis (IA), fibrous ankylosis (FA), and 'others'.

Arthroscopy and MRI scoring

For both MRI and arthroscopy data, two independent observers performed an assessment; the observers were blinded to any sort of clinical information. A junior and senior observer were recruited to score each modality. Junior observers had approximately 5 years of experience, whereas senior observers had a minimum of 15 years of experience in interpreting MRI or arthroscopy of the TMJ.

To be able to compare findings between arthroscopy and MRI, a standardized scoring form was constructed for this study (Fig. 1). The presence or absence of certain pathological entities (scoring items) was scored using a Likert scale. Arthroscopy, synovitis, and chondromalacia were scored using a grading scale proposed by McCain et al.[10]. The absence of roofing of the disc as described by McCain et al. was used for arthroscopic determination of disc displacement. After scoring each joint, the observers also gave it one or more of the study diagnostic labels that summarized their findings best. Both oral and maxillofacial surgeons scored each joint one time by reviewing the video footage of the arthroscopy. Both dento-maxillofacial radiologists scored each joint by importing and reviewing the DICOM files in the Sante DICOM Viewer (Santesoft, Nicosia, Cyprus).

Training and calibration sessions were organized for the MRI and arthroscopy observers, where they were shown a number of cases that were discussed in a group with further calibration to expert observations. Following successful calibration, they assessed all joints independently.

Additionally, one MRI observer also performed signal intensity ratio (SIR) measurements of the retrodiscal zone and the anterior and posterior bands of the articular disc. Measurements were taken from T1-weighted images. Brain grey matter (GM) and white matter (WM) were measured in order to normalize differences in signal intensity. SIR values were calculated by custom Java software (JBuilder 10.0; Embarcadero Technologies, Austin, Texas, USA), using the following formula: $SIR = (SI_{measured} - SI_{GM})/(SI_{WM} - SI_{GM})$[17,18]. These values could then be compared to the respective synovitis grading scores of one arthroscopy observer to assess if the two methods showed a correlation when scoring inflammation in these structures.

Statistical analysis

Descriptive statistics were calculated to characterize the study sample. As a first primary outcome, the agreement between MRI and arthroscopy scoring items and diagnostic labels was assessed. Fleiss kappa values for the four combinations of observer pairs were calculated for each comparable scoring item and diagnostic

Please cite this article in press as: Verhelst PJ, et al. The agreement between magnetic resonance imaging and arthroscopic findings in temporomandibular joint disorders, *Int J Oral Maxillofac Surg* (2020), https://doi.org/10.1016/j.ijom.2020.10.012

		Arthroscopy	MRI
Displacement of disc		Likert	Likert
	Anterior	Likert	Likert
	Posterior	Likert	Likert
	Partial	Likert	Likert
	Reduction	Likert	Likert
Deformation of disc		Likert	Likert
	Crumpled	Likert	Likert
	Rounded	Likert	Likert
	Flat	Likert	Likert
	Perforation	Likert	Likert
Adhesions		Likert	Likert
Joint effusion		-	Likert
Condylar pterygoid attachment anomaly		-	Likert
Pterygoid shadow anomaly		Likert	-
Retrodiscal deformation		Likert	Likert
Condylar deformation		Likert	Likert
Glenoid fossa deformation		Likert	Likert
	Chrondromalacia grade	Grading	-
Articular eminence deformation		Likert	Likert
	Chondromalacia grade	Grading	-
Overall synovitis		Likert	-
	Anterior band of disc	Grading	SIR
	Posterior band of disc	Grading	SIR
	Retrodiscal synovium	Grading	SIR
	Anterior recess	Grading	-
	Fossa	Grading	-
	Articular eminence	Grading	-
	Medial synovium	Grading	-
Diagnostic label	ADD-R	Yes/No	Yes/No
	ADD+R	Yes/No	Yes/No
	DJD	Yes/No	Yes/No
	Inflammatory arthritis	Yes/No	Yes/No
	Fibrous ankylosis	Yes/No	Yes/No
	Other	Yes/No	Yes/No

Fig. 1. The study specific scoring form.
Legend:
ADD-R: anterior disc displacement without reduction
ADD + R: anterior disc displacement with reduction
DJD: degenerative joint disease
Likert scale 1 to 5:
1. Absolutely not
2. Probably not
3. Uncertain
4. Probably yes
5. Absolutely yes

Synovitis Grading
0. Normal, with whitish synovial tissue and wire capillaries
1. Light hypervascularity, with larger capillaries
2. Moderate hypervascularity, with the presence of petechiae
3. Severe hypervascularity, with the presence of broad areas of petechiae/ecchymosis.

Chondromalacia Grading
0. Normal, completely smooth, smooth white surface with striations on the surface
1. Surface degeneration
2. Intermediate degeneration, with fibrocartilage with bubbles, but without bone exposure
3. Advanced degeneration, with evident bone exposure throughout the joint

Only scoring items that were included in both the arthroscopy and MRI form were assessed for agreement. If it was not assessed, it is indicated with '-'.

Please cite this article in press as: Verhelst PJ, et al. The agreement between magnetic resonance imaging and arthroscopic findings in temporomandibular joint disorders, *Int J Oral Maxillofac Surg* (2020), https://doi.org/10.1016/j.ijom.2020.10.012

Table 1. Characteristics of the study population.

	Number	%
Patients	48	100%
Sex		
Male	5	10%
Female	43	90%
Age (years)	Mean ± SD	41 ± 16
Time between MRI and arthroscopy (days)	Mean ± SD	71 ± 44
Time from onset of complaint to arthroscopy		
0–6 months	16	33%
7–12 months	10	21%
13–24 months	9	19%
>24 months	13	27%
Joints	50	100%
Side		
Left	28	56%
Right	22	44%
Main complaint		
Pain	45	90%
Limited MMO	4	8%
Sounds	1	2%
Clinical diagnostic label		
ADD − R	20	40%
ADD + R	3	6%
Degenerative joint disease	23	46%
Inflammatory arthritis	28	56%
Fibrous ankylosis	5	10%
Other	2	4%

ADD − R, anterior disc displacement without reduction; ADD + R, anterior disc displacement with reduction; MMO, maximum mouth opening; MRI, magnetic resonance imaging; SD, standard deviation.

label. The highest and lowest kappa values of the four pairs were reported to show the spread of agreement between the observers.

Next, the contribution of unbiased MRI and arthroscopy in reaching a final diagnosis was evaluated as a second primary outcome. For each joint, the agreement between diagnostic labels of one MRI and one arthroscopy observer respectively and those of the final diagnosis was calculated. A high kappa value suggests a bigger contribution of the modality to reaching a final diagnosis.

As a secondary outcome, this study also gauged the relationship between SIR measurements and synovitis grading scores. A Kruskal–Wallis test was used to determine whether patients, grouped by their synovitis grading scores, had significantly different SIR values ($P < 0.05$). Next, Spearman rank correlation coefficients were calculated to test if increasing synovitis grading scores correlated with increasing SIR values. For the final secondary outcome, the agreement within each modality was assessed using Fleiss kappa values.

To interpret the kappa values, the guidelines of Altman were used: poor agreement, $\kappa < 0.20$; fair agreement, $0.21 < \kappa < 0.40$; moderate agreement, $0.41 < \kappa < 0.60$; good agreement, $0.61 < \kappa < 0.80$; and very good agreement, $0.81 < \kappa < 1.00$[19]. The standard error was used to calculate 95% confidence intervals (CI), and *P*-values were calculated as well ($P < 0.05$ as significant).

Results

After screening the medical records of arthroscopy patients, 50 joints in 48 patients were included in the study (Table 1). The most frequently used diagnostic labels for the final diagnosis were inflammatory arthritis (56%), degenerative joint disease (46%), and ADD − R (40%). The average time between preoperative MRI and arthroscopy was 71 days. The average time between the onset of the problem and arthroscopy was more than 1 year for 46% of patients.

Agreement between MRI and arthroscopy

Table 2 provides an overview of the agreement between the MRI and arthroscopy observers. Their findings showed poor agreement, with only two items achieving a fair level of agreement: the capacity of disc reduction on mouth closure, and the presence of a disc perforation. Agreement levels for the diagnostic labels were higher, with ADD − R, ADD + R, and 'other' reaching fair to moderate levels of agreement. Overall, agreement between the arthroscopy and MRI observers was low, meaning that when blinded from clinical information, arthroscopy and MRI observers often come to different conclusions for the same joint.

Contribution of MRI/arthroscopy to the final diagnosis

For five of the six possible diagnostic labels, arthroscopy scored higher levels of agreement with the final diagnosis than MRI (Table 3). These five diagnostic labels also reached fair to moderate levels of agreement. MRI only had two labels that scored fair levels of agreement. Overall, arthroscopy labels had a better match with the final diagnosis, suggesting a bigger contribution to reaching the final diagnosis than MRI.

Correlation between SIR and synovitis grading

The results showed a significant correlation between SIR measurements and synovitis grading for the posterior band of the disc (Kruskal–Wallis-test $P = 0.004$ and Spearman rank correlation coefficient 0.2904 with $P = 0.0018$) and the retrodiscal tissue (Kruskal–Wallis $P = 0.0012$ and Spearman rank correlation coefficient 0.3202 with $P = 0.0008$). This was not the case for the anterior band (Kruskal–Wallis $P = 0.6004$ and Spearman rank correlation coefficient 0.1646 with $P = 0.2101$). These findings showed that higher SIR values calculated on MRI corresponded with higher synovitis grading scores based on arthroscopy for the posterior band and retrodiscal tissue.

Agreement within MRI or arthroscopy

Table 4 shows that poor agreement on scoring items was found more frequently within MRI observers (50% of scoring items) when compared to arthroscopy observers (33% of scoring items). The majority (66%) of the arthroscopy scoring items had fair to moderate agreement levels. In the MRI group, 30% of the scoring items were assessed with fair levels of agreement. Displacement of the disc and the diagnostic label ADD − R had a good level of agreement within MRI observers.

Discussion

TMDs affect between 5% and 12% of the population[20]. A small subset of this sample are patients with TMDs that are re-

Please cite this article in press as: Verhelst PJ, et al. The agreement between magnetic resonance imaging and arthroscopic findings in temporomandibular joint disorders, *Int J Oral Maxillofac Surg* (2020), https://doi.org/10.1016/j.ijom.2020.10.012

Table 2. Agreement between MRI and arthroscopy observer pairs.

	Worst observer pair				Best observer pair				Category of agreement[a]				
	Kappa	95% CI[b]		*P*-value	Kappa	95% CI[b]		*P*-value	Poor	Fair	Moderate	Good	Very good
Displacement of disc	−0.02	−0.08	0.04	0.5228	0.13	−0.03	0.28	0.1053	4				
Anterior	−0.05	−0.16	0.06	0.3914	0.14	−0.03	0.31	0.1020	4				
Posterior	0.08	−0.04	0.20	0.1961	0.14	−0.07	0.34	0.1975	4				
Partial	−0.01	−0.21	0.18	0.8968	0.13	−0.07	0.34	0.1898	4				
Reduction	0.14	−0.06	0.35	0.1758	0.25	0.03	0.46	0.0238*	1	3			
Deformation of disc	0.03	−0.17	0.23	0.7632	0.19	0.02	0.37	0.0324*	4				
Crumpled	−0.05	−0.19	0.10	0.5181	0.20	−0.03	0.43	0.0964	4				
Rounded	−0.03	−0.13	0.06	0.4867	0.02	−0.14	0.18	0.8332	4				
Flat	−0.10	−0.26	0.07	0.2460	0.03	−0.08	0.13	0.5921	4				
Perforation	0.07	−0.05	0.19	0.2754	0.40	0.17	0.63	0.0006*	3	1			
Adhesions	0.08	−0.02	0.17	0.1106	0.19	0.01	0.37	0.0357*	4				
Lateral pterygoid anomaly	−0.02	−0.17	0.13	0.7674	0.02	−0.05	0.09	0.6003	4				
Retrodiscal deformation	−0.02	−0.21	0.17	0.8144	0.16	−0.03	0.35	0.0965	4				
Glenoid fossa deformation	−0.07	−0.17	0.02	0.1250	0.04	−0.01	0.08	0.1272	4				
Articular eminence deformation	0.02	−0.02	0.06	0.3098	0.09	0.00	0.19	0.0470*	4				
Diagnostic label													
ADD − R	0.06	−0.17	0.29	0.6101	0.38	0.12	0.64	0.0044*	2	2			
ADD + R	0.08	−0.20	0.36	0.5674	0.46	0.18	0.74	0.0012*	2	1	1		
DJD	−0.18	−0.44	0.08	0.1742	0.10	−0.15	0.35	0.4234	4				
Inflammatory arthritis	−0.10	−0.29	0.10	0.3359	0.10	−0.10	0.29	0.3359	4				
Fibrous ankylosis	–	–	–	–	–	–	–	–					
Other	0.13	−0.12	0.38	0.3196	0.23	−0.04	0.51	0.0953	3	1			

ADD − R, anterior disc displacement without reduction; ADD + R, anterior disc displacement with reduction; CI, confidence interval; DJD, degenerative joint disease; MRI, magnetic resonance imaging. *Significant *P*-value ($P < 0.05$).

[a] Observer pairs were placed in the poor/fair/moderate/good/very good category of agreement according to their kappa value.

[b] The first column is the lower limit of the 95% CI and the second column is the upper limit.

Table 3. Agreement between diagnostic labels of MRI/arthroscopy and the result of the complete clinical workup (clinical examination + radiology + arthroscopy).

	Clinical workup vs Arthroscopy					Clinical workup vs MRI				
	Kappa	95% CI[a]		*P*-value	Agreement[b]	Kappa	95% CI[a]		*P*-value	Agreement[b]
ADD − R	0.44	0.18	0.71	0.0012*	Moderate	0.21	−0.04	0.46	0.0900	Fair
ADD + R	0.27	0.04	0.50	0.0236*	Fair	0.12	−0.06	0.30	0.1843	Poor
DJD	0.26	0.01	0.51	0.0410*	Fair	0.07	−0.14	0.27	0.5236	Poor
Inflammatory arthritis	0.32	0.06	0.59	0.0148*	Fair	0.01	−0.14	0.17	0.8494	Poor
Fibrous ankylosis	0.56	0.28	0.83	0.0001*	Moderate	–	–	–	–	
Other	−0.06	−0.32	0.20	0.6704	Poor	0.37	0.10	0.64	0.0075*	Fair

ADD − R, anterior disc displacement without reduction; ADD + R, anterior disc displacement with reduction; CI, confidence interval; DJD, degenerative joint disease; MRI, magnetic resonance imaging. *Significant *P*-value ($P < 0.05$).

[a] The first column is the lower limit of the 95% CI and the second column is the upper limit.

[b] Agreement according to the kappa value.

fractory to conservative therapy. These patients are often subjected to MRI and subsequently arthroscopy[21]. This study investigated whether the diagnostic findings of the two examinations match and how they contribute to a final diagnosis. This information could help clinicians in their interpretation of these results

Assessment of the primary outcomes of this study showed that agreement levels between MRI and arthroscopy observers were low. There was far from a perfect match, and MRI and arthroscopy observers appeared to reach different conclusions on the same joint. This is in contrast to other studies that have reported high levels of agreement or correlation between MRI and arthroscopy findings[10–15]. It is, however, important to note that in these previous studies a limited set of joint structures or pathologies was studied and no blinding procedures for observers were mentioned. If the clinical information is available when interpreting MRI or arthroscopy, this will introduce a bias in the result of the examination, making it difficult to draw conclusions about the independent diagnostic merit of MRI or arthroscopy. By blinding the observers in the present study, an attempt was made to overcome this issue and assess the unbiased agreement between the two diagnostic modalities. One should also keep in mind that the correlation between MRI findings and clinical symptoms is not always strong[7,22]. The presence or absence of pathological MRI findings does not always line up with clinical symptoms. This could be a contributory factor to the low agreement between MRI and arthroscopy: a symptomatic patient with a normal MRI could have problems that become evident during arthroscopy.

As for their respective contributions to a final diagnosis, arthroscopy seems to outweigh MRI. A possible explanation for this difference is that in clinical practice, the final diagnosis relies more on information from arthroscopy findings than on information from MRI findings. However, one must keep in mind that arthroscopy reached only fair to moderate agreement levels with the final diagnosis, which means that MRI still brings valuable in-

Please cite this article in press as: Verhelst PJ, et al. The agreement between magnetic resonance imaging and arthroscopic findings in temporomandibular joint disorders, *Int J Oral Maxillofac Surg* (2020), https://doi.org/10.1016/j.ijom.2020.10.012

Table 4. Intra-modality agreement within arthroscopy and MRI.

	Arthroscopy					MRI				
	Kappa	95% CI[a]		*P*-value	Agreement[b]	Kappa	95% CI[a]		*P*-value	Agreement[b]
Displacement of disc	0.19	0.08	0.29	0.0006*	Poor	0.69	0.47	0.91	0.0000*	Good
Anterior	0.29	0.15	0.42	<0.0001*	Fair	0.34	0.12	0.56	0.0021*	Fair
Posterior	0.22	0.05	0.39	0.0108*	Fair	–	0.00	0.00	–	
Partial	−0.10	−0.29	0.10	0.3439	Poor	0.01	−0.09	0.10	0.9126	Poor
Reduction	0.21	0.00	0.42	0.046*	Fair	0.53	0.31	0.76	0.0000*	Moderate
Deformation of disc	0.27	0.12	0.42	0.0005*	Fair	0.34	0.15	0.53	0.0004*	Fair
Crumpled	0.29	0.08	0.51	0.0082*	Fair	0.13	0.00	0.26	0.0440*	Poor
Rounded	0.11	−0.01	0.22	0.0658	Poor	0.28	0.13	0.43	0.0003*	Fair
Flat	−0.03	−0.22	0.16	0.7373	Poor	0.13	−0.03	0.29	0.1116	Poor
Perforation	0.55	0.31	0.79	<0.0001*	Moderate	0.01	−0.08	0.10	0.7778	Poor
Adhesions	0.52	0.32	0.72	<0.0001*	Moderate	0.06	−0.03	0.15	0.2207	Poor
Joint effusion						0.28	0.11	0.45	0.0010*	Fair
Pterygoid shadow anomaly	0.03	−0.11	0.17	0.6947	Poor	0.00	−0.05	0.05	0.9593	Poor
Retrodiscal deformation	0.41	0.23	0.59	<0.0001*	Moderate	0.26	0.10	0.42	0.0011*	Fair
Condylar deformation						0.37	0.18	0.56	0.0001*	Fair
Glenoid fossa deformation	0.29	0.09	0.49	0.004*	Fair	0.00	−0.15	0.15	0.9932	Poor
Chondromalacia grade	0.41	0.22	0.61	<0.0001*	Moderate					
Articular eminence deformation	0.16	0.03	0.29	0.0187*	Poor	−0.03	−0.13	0.06	0.4816	Poor
Chondromalacia grade	0.20	0.08	0.33	0.002*	Poor					
Overall synovitis	0.19	0.05	0.33	0.0064*	Poor					
Anterior band of disc	0.24	−0.08	0.56	0.1419	Fair					
Posterior band of disc	0.24	0.06	0.42	0.0076*	Fair					
Retrodiscal synovium	0.50	0.28	0.71	<0.0001*	Moderate					
Anterior recess	0.41	0.10	0.71	0.0095*	Fair					
Fossa	0.56	0.37	0.76	<0.0001*	Moderate					
Articular eminence	0.46	0.28	0.64	<0.0001*	Moderate					
Medial synovium	0.34	0.16	0.52	0.0002*	Fair					
Diagnostic label										
ADD − R	0.37	0.10	0.63	0.0062*	Fair	0.66	0.40	0.92	0.0000*	Good
ADD + R	0.22	−0.04	0.48	0.0943	Fair	0.43	0.17	0.69	0.0011*	Moderate
DJD	0.20	−0.01	0.40	0.064	Poor	0.05	−0.15	0.25	0.6016	Poor
Inflammatory arthritis	0.34	0.11	0.56	0.0031*	Fair	0.11	−0.17	0.39	0.4321	Poor
Fibrous ankylosis	–	–	–	–		–	–	–	–	
Other	−0.11	−0.38	0.15	0.4002	Poor	–	–	–	–	

ADD − R, anterior disc displacement without reduction; ADD + R, anterior disc displacement with reduction; CI, confidence interval; DJD, degenerative joint disease; MRI, magnetic resonance imaging. *Significant *P*-value ($P < 0.05$). '-' indicates that no kappa value could be calculated, while blank cells indicate items not scored for MRI/arthroscopy.

[a] The first column is the lower limit of the 95% CI and the second column is the upper limit.

[b] Agreement according to the kappa value.

formation to the table. This is emphasized by the observation that the findings of MRI differ from those of arthroscopy. Combining the information from both modalities in diagnosing these refractory patients seems to remain the best option, as illustrated in Fig. 2.

SIRs on MRI provide a standardized way of measuring and comparing signal intensities of structures of interest from different patients and MRI scans[17]. A high signal intensity measurement of the disc and retrodiscal tissue is regarded as an objective sign of inflammation[18]. When there is inflammation of these structures, it is characterized by haemorrhagic infiltration and the appearance of newly formed capillaries[10]. This can be assessed easily and graded during the arthroscopic evaluation. A significant positive relationship between grading scores and SIR measurements for the posterior band of the disc and the retrodiscal tissue was found in the present study. SIR measurements as an objective measurement of inflammation seem to correlate with the detection of inflammation during arthroscopy. However, this was not the case for the anterior band.

Finally, the agreement between the two observers of each modality was also assessed as a secondary outcome. Arthroscopy observers tended to agree more and better on scoring items than MRI observers. Few to no studies evaluating intra-modality reliability or agreement for MRI or arthroscopy are available as a reference to check these results. This was unexpected, as this type of reliability assessment provides valuable information when interpreting the results of diagnostic studies. Furthermore, this indicates the importance of calibrating exercises, ideally on a regular basis, between clinicians.

This study had some significant limitations. First, 50 joints is a limited sample size for making a conclusive statement at the population level. This is reflected in some of the non-significant *P*-values for the kappa values and these results should therefore be interpreted with care. Next, the mean time between MRI and arthroscopy was 71 days. During these 2.5 months, changes could have occurred in the joint resulting in a de facto non-agreement between MRI and arthroscopy. Also, patients with internal derangement on MRI do not always present with symptoms and vice versa. The patient sample in this study consisted of patients who were refractory to conservative measures. Heavily degraded joints pose a challenge for detailed MRI or arthroscopy scoring and could have played a role in lower agreement scores. The assessment of arthroscopy videos is not easy when the observer is

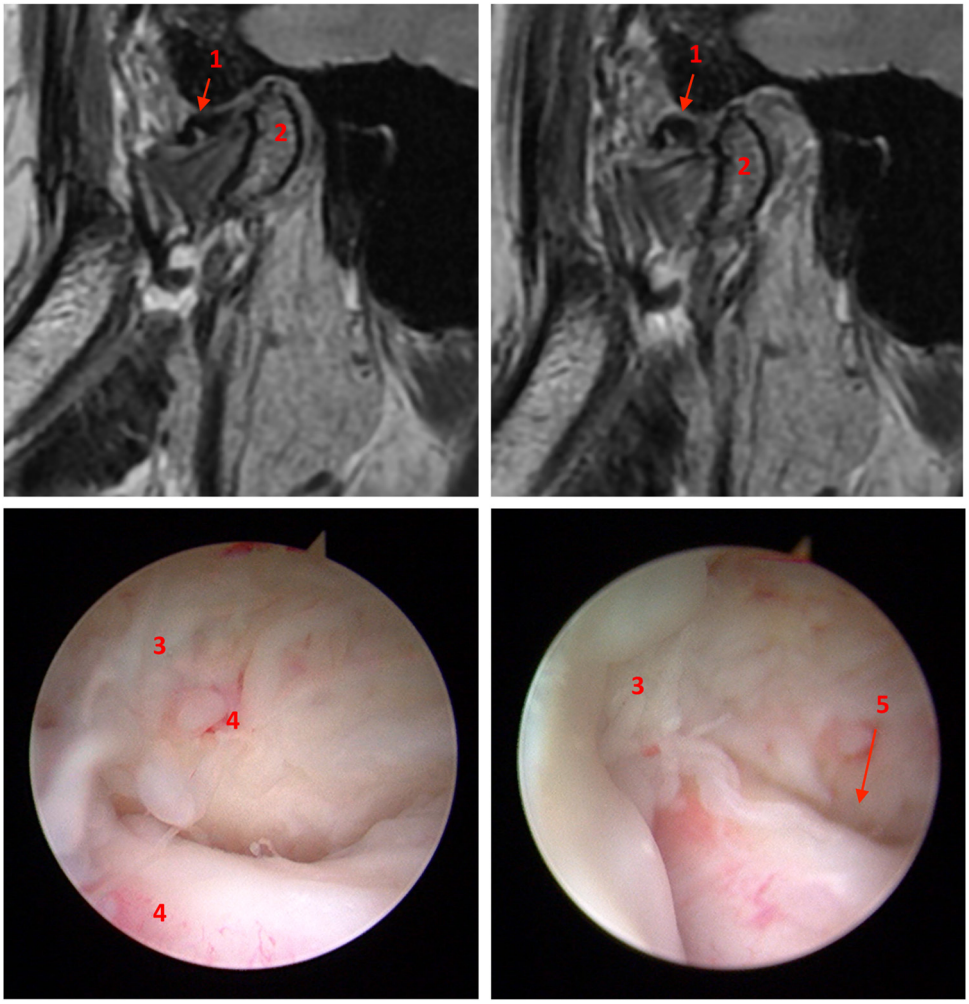


Fig. 2. Illustration of a patient for whom MRI and arthroscopy findings differed and a combination of both led to a better final diagnosis. MRI showed an anteriorly dislocated disc without reduction (1). Some degenerative signs of the condyle (2) were also suspected. Arthroscopy showed advanced degenerative changes of the fossa and articular eminence with flocculations and adhesions (3). Additionally, a significant synovitis component (4) was identified, as well as a perforation of the retrodiscal tissue (5).

not performing the arthroscopy themselves. This makes orientation in the joint and differentiating iatrogenic changes from pathological changes even more difficult and could have played an additional role in the low agreement scores. Finally, although the MRI and arthroscopic data were assessed prospectively, the conclusions of the complete clinical workup were collected retrospectively, which could have introduced bias in the resulting diagnostic labels.

In conclusion, the aim of this retrospective observational cohort study was to assess the potential agreement between independent MRI and arthroscopic findings and their respective contributions to reaching a final diagnosis in patients with TMDs refractory to conservative therapy. Agreement levels between MRI and arthroscopy were low and arthroscopy findings appeared to make a greater contribution to the final diagnosis. Higher SIR measurements of the posterior band of the disc and retrodiscal tissue correlated with higher synovitis grading by arthroscopy. These findings suggest that both MRI and arthroscopy provide complementary information on joint pathology, but should be interpreted with caution and carefully weighed against clinical findings.

Funding

None.

Competing interests

None.

Ethical approval

This retrospective study was approved by the Ethics Committee of UZ Leuven, Belgium (mp012487).

Patient consent

Not required.

References

1. Renapurkar SK. Surgical versus nonsurgical management of degenerative joint disease. *Oral Maxillofac Surg Clin North Am* 2018;**30**:291–7. http://dx.doi.org/10.1016/j.coms.2018.04.005.
2. Murphy MK, MacBarb RF, Wong ME, Athanasiou KA. Temporomandibular disorders: a review of etiology, clinical management, and tissue engineering strategies. *Int J Oral Max-*

illofac Implants 2013;**28**:e393–e414. http://dx.doi.org/10.11607/jomi.te20.
3. Guo C, Shi Z, Revington P. Arthrocentesis and lavage for treating temporomandibular joint disorders. *Cochrane Database Syst Rev* 2015;**12**. http://dx.doi.org/10.1002/14651858.CD004973.pub3.
4. Laskin DM. Arthroscopy versus arthrocentesis for treating internal derangements of the temporomandibular joint. *Oral Maxillofac Surg Clin North Am* 2018;**30**:325–8. http://dx.doi.org/10.1016/j.coms.2018.04.008.
5. Onishi M. Arthroscopy of the temporomandibular joint. *J Stomatol Soc Japan* 1975;**42**:207–13. http://dx.doi.org/10.5357/koubyou.42.207.
6. Aiken A, Bouloux G, Hudgins P. MR Imaging of the temporomandibular joint. *Magn Reson Imaging Clin N Am* 2012;**20**:397–412. http://dx.doi.org/10.1016/j.mric.2012.05.002.
7. De Laat A, Horvath M, Bossuyt M, Fossion E, Baert AL. Myogenous or arthrogenous limitation of mouth opening: correlations between clinical findings, MRI, and clinical outcome. *J Orofac Pain* 1993;**7**:150–5.
8. Yura S, Nobata K, Shima T. Diagnostic accuracy of fat-saturated T2-weighted magnetic resonance imaging in the diagnosis of perforation of the articular disc of the temporomandibular joint. *Br J Oral Maxillofac Surg* 2012;**50**:365–8. http://dx.doi.org/10.1016/j.bjoms.2011.05.017.
9. Sawada K, Amemiya T, Hirai S, Hayashi Y, Suzuki T, Honda M, Sisounthone J, Matsumoto K, Honda K. Diagnostic reliability of 3.0-T MRI for detecting osseous abnormalities of the temporomandibular joint. *J Oral Sci* 2018;**60**:137–41. http://dx.doi.org/10.2334/josnusd.16-0838.
10. McCain JP, de la Rua H, Le Blanc WG. Correlation of clinical, radiographic, and arthroscopic findings in internal derangements of the TMJ. *J Oral Maxillofac Surg* 1989;**47**:913–21. http://dx.doi.org/10.1016/0278-2391(89)90373-X.
11. Rao VM, Farole A, Karasick D. Temporomandibular joint dysfunction: correlation of MR imaging, arthrography, and arthroscopy. *Radiology* 1990;**174**:663–7. http://dx.doi.org/10.1148/radiology.174.3.2305046.
12. Casares G, Benito C, de la Hoz JL, Benito C. Treatment of TMJ static disk with arthroscopic lysis and lavage: a comparison between MRI arthroscopic findings and clinical results. *Cranio* 1999;**17**:49–57. http://dx.doi.org/10.1080/08869634.1999.11746077.
13. Millon-Cruz A, Martín-Granizo R, Encinas A, Berguer A. Relationship between intra-articular adhesions and disc position in temporomandibular joints: magnetic resonance and arthroscopic findings and clinical results. *J Craniomaxillofac Surg* 2015;**43**:497–502. http://dx.doi.org/10.1016/j.jcms.2015.02.010.
14. Sato FRL, Lima CAA, Tralli G, da Silva RA. Is there a correlation between arthroscopic findings and the clinical signs and symptoms of patients with internal derangement of the temporomandibular joint? A prospective study. *Int J Oral Maxillofac Surg* 2019;**48**:233–8. http://dx.doi.org/10.1016/j.ijom.2018.07.004.
15. Moses JJ, Salinas E, Goergen T, Sartoris D. Magnetic resonance imaging or arthrographic diagnosis of internal derangement of the temporomandibular joint. *Oral Surg Oral Med Oral Pathol* 1993;**75**:268–72. http://dx.doi.org/10.1016/0030-4220(93)90132-N.
16. Ohrbach R., Gonzalez Y., List T., Michelotti A., Schiffman E. Diagnostic Criteria for Temporomandibular Disorders (DC/TMD) clinical examination protocol: version 02 June 2013. http://www.rdc-tmdinternational.org [Accessibility verified April 8, 2020].
17. Orhan K, Nishiyama H, Tadashi S, Murakami S, Furukawa S. Comparison of altered signal intensity, position, and morphology of the TMJ disc in MR images corrected for variations in surface coil sensitivity. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 2006;**101**:515–22. http://dx.doi.org/10.1016/j.tripleo.2005.04.004.
18. Cesur E, Özdiler O, Köklü A, Orhan K, Seki U. Effects of wear time differences of removable functional appliances in class II patients: prospective MRI study of TMJ and masticatory muscle changes. *Oral Radiol* 2020;**36**:47–59. http://dx.doi.org/10.1007/s11282-019-00379-0.
19. Laerd Statistics. Fleiss' kappa using SPSS Statistics. Statistical tutorials and software guides, 2019. https://statistics.laerd.com/spss-tutorials/fleiss-kappa-in-spss-statistics.php [Accessibility verified].
20. Schiffman E, Ohrbach R, Truelove E, Look J, Anderson G, Goulet JP, List T, Svensson P, Gonzalez Y, Lobbezoo F, Michelotti A, Brooks SL, Ceusters W, Drangsholt M, Ettlin D, Gaul C, Goldberg LJ, Haythornthwaite JA, Hollender L, Jensen R, John MT, De Laat A, de Leeuw R, Maixner W, van der Meulen M, Murray GM, Nixdorf DR, Palla S, Petersson A, Pionchon P, Smith B, Visscher CM, Zakrzewska J, Dworkin SF, International RDC/TMD Consortium Network. International association for Dental Research. Orofacial Pain Special Interest Group. International Association for the Study of Pain. Diagnostic Criteria for Temporomandibular Disorders (DC/TMD) for Clinical and Research Applications: Recommendations of the International RDC/TMD Consortium Network* and Orofacial Pain Special Interest Group†. *J Oral Facial Pain Headache* 2014;**28**:6–27. http://dx.doi.org/10.11607/jop.1151.
21. Rigon M, Pereira LM, Bortoluzzi MC, Loguercio AD, Ramos AL, Cardoso JR. Arthroscopy for temporomandibular disorders. *Cochrane Database Syst Rev* 2015;**2015**(12). http://dx.doi.org/10.1002/14651858.CD006385.pub3.
22. González-Ballester D, González-García R, García Nogales A, Moreno García C, Monje Gil F. ¿Es el desplazamiento discal sinónimo de patología articular temporomandibular? Correlación clínico-radiológica y prevalencia de trastornos internos en sujetos voluntarios asintomáticos [Is disc displacement synonymous with temporomandibular joint disease? Clinical-radiological correlation and prevalence of internal disorders in asymptomatic volunteers] *Revista Espapola de CirugÚa Oral y Maxilofacial* 2020;**42**:97–106. http://dx.doi.org/10.20986/recom.2020.1133/2020.

Address:
Pieter-Jan Verhelst
Department of Oral and Maxillofacial Surgery
University Hospitals Leuven
Campus Sint-Rafaël
Kapucijnenvoer 33
BE-3000 Leuven
Belgium
Tel: +32 (0)476 33 40 14
E-mail: pieter-jan.verhelst@uzleuven.be