MRI in the Assessment of TMJ-Arthritis in Children with JIA;
Repeatability of a Newly Devised Scoring System

Oskar W. Angenete, MD, Thomas A. Augdal, MD, Marite Rygg, MD, PhD, Karen Rosendahl, MD, PhD

Rationale and Objectives: The temporomandibular joint (TMJ) is commonly involved in children with juvenile idiopathic arthritis. The diagnosis and evaluation of the disease progression is dependent on medical imaging. The precision of this imaging is under debate. Several scoring systems have been proposed but transparent testing of the precision of the constituents of the scoring systems is lacking. The present study aims to test the precision of 25 imaging features based on magnetic resonance imaging (MRI).

Materials and Methods: Clinical data and imaging were obtained from the Norwegian juvenile idiopathic arthritis study, The NorJIA study. Twenty-five imaging features of the TMJ in MRI datasets from 86 study participants were evaluated by two experienced radiologists for inter- and intraobserver agreement. Agreement of ordinal variables was measured with Cohen’s linear or weighted Kappa as appropriate. Agreement of continuous measurements was assessed with 95% limit of agreement according to Bland-Altman.

Results: In the osteochondral domain, the ordinal imaging variables “loss of condylar volume,” “condylar shape,” “condylar irregularities,” “shape of the eminence/fossa,” “disk abnormalities,” and “condylar inclination” showed inter- and intraobserver agreement above Kappa 0.5. In the inflammatory domain, the ordinal imaging variables “joint fluid,” “overall impression of inflammation,” “synovial enhancement” and “bone marrow oedema” showed inter- and intraobserver agreement above Kappa 0.5. Continuous measurements performed poorly with wide limits of agreement.

Conclusion: A precise MRI-based scoring system for assessment of TMJ in JIA is proposed consisting of seven variables in the osteochondral domain and four variables in the inflammatory domain. Further testing of the clinical validity of the variables is needed.

Key Words: arthritis; juvenile; observer variation; scoring system; precision.

© 2021 The Association of University Radiologists. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

INTRODUCTION

Juvenile idiopathic arthritis (JIA) is a chronic rheumatic disease of unknown origin, with an onset before the age of 16 and a reported incidence of 15 (7-23) per 100,000 children/year in the Nordic countries (1). In patients with JIA, the temporomandibular joint (TMJ) is affected in 39-78% (2-5) depending on definitions used for involvement, disease duration, and on the methods used for ascertainment. On imaging, TMJ arthritis is characterized by synovial inflammation, bone marrow- and soft tissue oedema and joint effusion, subsequently followed by destructive changes of cartilage and bone. Left untreated, or in treatment-resistant cases, arthritis of the TMJ can lead to facial asymmetry, orofacial pain and reduced quality of life (6-10).

The diagnosis of TMJ involvement in JIA is based on clinical findings, magnetic resonance imaging (MRI), cone beam computed tomography (CBCT) or a combination of these (11-15). The accuracy of clinical findings and clinical
monitoring of the disease course, both active inflammation and permanent damage, is under debate (16,17) and much effort has been made during the past years to establish a valid, MRI-based imaging protocol and classification system. However, methodological difficulties, including lack of references for normal findings, low image resolution and imprecise scoring systems have led to both over- and underreporting of findings in the TMJ for both MRI and CBCT. The authors reported that the system was not hampered with significant intra- or inter-reader differences but did not present any data to confirm their statement. Vaid and colleagues (22) proposed an MRI-based scoring system based on 20 patients, classifying changes into acute or chronic (structural damage). The grading system included measurements of small, intra-articular components under 3 mm in size, however, the precision of these measurements was not presented. The overall interobserver agreement for acute and chronic changes, based on composite variables, was moderate to good, with weighted kappa values of 0.51 and 0.68, respectively.

In 2015, a third MRI-based scoring system was published by Kellenberger and co-authors (13). The system is progressive on a 0-4 scale and divided into an inflammatory domain and a deformity domain. The system is in part built on the experiences drawn from the systems published by Koos and Vaid, but in the publication from 2015, a full scale, adequate test of intra- and interobserver agreement is lacking. In 2018, Tolland et al proposed an MRI-based scoring system (23) founded on the experiences drawn from the systems published by Koos, Vaid and Kellenberger. The system was developed by a multi-institutional consensus process finally proposing eight imaging items including both the inflammatory and osteochondral damage domains. Each item was assigned either a binary, ordinal grade (0–1) or a 3-graded, ordinal grade (0–2). The grades of each, individual item were then added, resulting in a total score. The authors performed a reliability exercise of the system in 21 selected cases and chose to measure reliability along an intra-class correlation scale (ICC). However, measuring agreement of ordinal variables with ICC is debatable (24). Furthermore, the selection of patients and low number of patients (n = 21) leaves unanswered questions on the transferability of the results to the JIA population.

In 2018, Kellenberger published a pictorial essay on JIA-related, temporomandibular changes on MRI (14). This publication presents a thorough explanation of the scoring systems already proposed by Tolland and Kellenberger, both through written explanations and through a wide range of MRI examples. Used as a common ground-reference this publication might help reduce interobserver variability.

To date, however, no MRI-based scoring system of the TMJ is proven precise and valid. We therefore aimed to examine the precision of MRI-based measurements and scores used to describe anatomy, structural damage and inflammation of the TMJ in a large cohort of children and adolescents with JIA. Next, to indicate markers holding sufficient precision to be included in a future scoring system for active arthritis and structural deformity.

**MATERIAL AND METHODS**

**Patients**

The participants in this study constitute a subset of 86 children and adolescents selected from a prospective, longitudinal observational study addressing TMJ involvement in children with JIA (n = 228), the Norwegian JIA Study (NorJIA), NCT number NCT03904459 in www.clinicaltrials.gov. Participants in the NorJIA study were recruited from three tertiary pediatric university hospitals (Haukeland University hospital, Bergen, St Olav University hospital, Trondheim and University hospital of North Norway, Tromsø). Inclusion criteria were a diagnosis of JIA according to the ILAR criteria (25) performed by experienced pediatric rheumatologists, and age between 4 and 16 years at inclusion. According to the study protocol, all of the included participants in the NorJIA study were referred to MRI of the TMJ, regardless of clinical symptoms from the TMJ. In cases of clinical TMJ symptoms, and when an MRI was judged to be of specific clinical importance, sedation was used for the younger children. For this particular sub-study, we included MRIs performed between March 2015 and May 2018. Exclusion criteria for this study were suboptimal examinations due to artefacts and the use of braces.

To test the scoring system regarding skeletal development and varied pathology, an a priori, balanced selection of patients from the NorJIA cohort was made, based on the radiology report and patient age. The selection consisted of approximately 33% participants with moderate/severe findings, 33% participants with mild findings and 33% participants with subtle or no findings.

**Imaging**

All MRI examinations were performed on a 3 Tesla system (Skyra, Siemens healthineers, Erlangen, Germany), using a 64-channel head coil (32-channel at St Olav). An extensive protocol, including nine sequences was performed to allow for comparisons of different sequences, either alone or in combination, in the assessment of pathology. The MRI protocol takes into account the recommendations given by Miller (26) and Kellenberger (27), including sagittal T1-weighted MPRAGE, sagittal/oblique proton density-weighted, sagittal/oblique fat-saturated T2-weighted, sagittal/oblique fat-saturated T1-weighted, coronal T1-weighted and coronal T1-weighted two-point Dixon sequences.
Following intravenous gadolinium contrast injection, a dynamic coronal sequence, a sagittal/oblique fat-saturated T1-weighted sequence and a sagittal/oblique proton density-weighted sequence (open mouth) were performed. Intravenous gadolinium contrast was injected in a standardized way in an antecubital vein (Dotarem 279.3 mg/ml, 0.2 ml/kg body weight, 2 ml/s with 20 ml saline chaser). A detailed protocol description is provided in Appendix A.

Image Review

For the present study, the following seven sequences were used; coronal T1-weighted, sagittal T1-weighted MPRAGE, sagittal/oblique fat-saturated T2-weighted, sagittal/oblique fat-saturated T1-weighted, sagittal/oblique proton density-weighted with closed and open mouth and sagittal/oblique fat-saturated T1-weighted after intravenous contrast. The images were assessed independently by two consultant radiologists, twice (at an interval of at least 4 weeks) by O.A. (12 years of experience) and once by T.A.A. (13 years of experience), without any additional information available. Before scoring was performed, previous publications on scoring protocols and imaging atlas were thoroughly studied (14,21-23). The readers calibrated their interpretation of the chosen scoring protocol during two 1-day meetings and 2 video conferences, followed by consensus scoring of five TMJ MR examinations from a cohort of children with JIA, not included in the present study.

Five imaging markers describing anatomical features, seven describing structural changes (damage) and 13 markers describing inflammation were analyzed for the right and for the left TMJ, separately (Appendix B, C, D and E). To explore the usefulness of an extended MRI protocol, assessment of condylar irregularities was made, first on a minimal (core) set of sequences and second, on an extended (ideal) set of sequences, as suggested by Miller et al (26).

Statistical Analysis

Continuous data were presented as means (±SD), ordinal data as medians (ranges) and dichotomous data as proportions. Intra- and interobserver agreement were analyzed using a simple or a weighted (linear) Cohen’s Kappa coefficient with 95% confidence interval. A kappa score of <0.2 was considered poor, 0.21-0.40 fair, 0.41-0.60 moderate, 0.61-0.80 good and 0.81-1.00 very good. Absolute agreement was reported as proportions. Differences in measurements were analyzed using 95% limits of agreement (termed repeatability coefficient, when used for repeat measurements) as per Bland-Altman. Bland-Altman plots are generally interpreted informally, and a clinically acceptable agreement was set at 15%. A significance level of 0.05 was decided a priori and all the reported p values are two-tailed. Statistical analyses were performed using IBM SPSS Statistics, version 26.

Ethics

The NorJIA study was approved by the Regional Ethics Committee; REK nr 2012/542. Informed consents were given by the children if ≥16 years, and by the parents if the child were <16 years. Data was collected and stored according to the General Data Protection Regulation.

RESULTS

One set of MRIs from a total of 86 children (51 females) with JIA, median age 13 years (IQR 5), were included. Median age at diagnosis was 6 years (IQR 8) and the median duration of disease at the time of MRI imaging was 4,5 years (IQR 6) (Table 1). The distribution of findings for each of the 25 MRI-features assessed are shown in Figure 1 and 2 (right side first reading).

Osteochondral Domain

Assessment of loss of condylar volume on a 0-1 scale, condylar shape/flattening in the sagittal (0-3 and 0-2 scale) and in the coronal plane (0-2 scale), condylar irregularities on a 0-2 scale (both based on a core and an ideal protocol), disk abnormalities on a 0-1 scale and the shape of the articular eminence and glenoid fossa on a 0-2 scale showed good to very good agreement for the same reader, with kappa values of 0.67-0.80 (Table 2) (Fig 3). The inter-reader agreement was also good to very good except for condylar irregularities (both the ideal and the core protocols) and shape of the articular eminence and glenoid fossa, showing moderate agreement with kappa values of 0.57, 0.47 and 0.55, respectively (Table 2) (Figs 4a–b and Figure 5a–b).

Assessment of condylar inclination on a 0-2 scale showed good intra- and interobserver agreement, with kappa values of 0.74 and 0.61.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Girls, n (%)</td>
<td>51 (59%)</td>
</tr>
<tr>
<td>Age at MRI examination, median years (IQR)</td>
<td>13.0 (6)</td>
</tr>
<tr>
<td>Age at JIA diagnosis, median years (IQR)</td>
<td>6.0 (8)</td>
</tr>
<tr>
<td>Disease duration, median years (IQR)</td>
<td>4.5 (6)</td>
</tr>
<tr>
<td>JIA categories</td>
<td></td>
</tr>
<tr>
<td>Systemic</td>
<td>3 (3%)</td>
</tr>
<tr>
<td>Oligoarticular persistent</td>
<td>25 (29%)</td>
</tr>
<tr>
<td>Oligoarticular extended</td>
<td>8 (9%)</td>
</tr>
<tr>
<td>Polyarticular RF negative</td>
<td>27 (31%)</td>
</tr>
<tr>
<td>Psoriatic arthritis</td>
<td>3 (3%)</td>
</tr>
<tr>
<td>Enthesitis-related arthritis</td>
<td>11 (13%)</td>
</tr>
<tr>
<td>Undifferentiated arthritis</td>
<td>9 (10%)</td>
</tr>
</tbody>
</table>

JIA, juvenile idiopathic arthritis; MRI, Magnetic resonance imaging; RF, Rheumatoid factor.
Figure 1. Distribution of findings in the osteochondral domain, right side, first reading. The x-axis denotes number of patients. (Color version of figure is available online.)

Figure 2. Distribution of findings in the inflammatory domain, right side, first reading. The x-axis denotes number of patients. (Color version of figure is available online.)
Assessment of the position of the condyle on a 0–6-point location scale showed fair agreement for the same reader and poor inter-reader agreement (Table 2).

As for disk position on a 0-5 scale, with the mouth closed, there was a fair intra-reader and a poor inter-reader agreement (Table 2).

### Inflammatory Domain

**Joint fluid:** The intra-observer agreement for assessment of joint fluid on a 0-2 scale was good, both for the whole joint, and for the lower compartment, with kappa values of 0.74 and 0.69, respectively, while the agreement for upper compartment was moderate (kappa 0.51) (Table 3) (Fig 6). Agreement between observers was good for the whole joint, moderate for the upper and poor for the lower compartment (Table 3). Assessment of pathological fluid on a 0-1 scale performed well for the same observer, and moderately between observers.

**Synovial inflammation/enhancement/thickening:** There was moderate agreement for grading overall impression of inflammation on a 0-2 scale, with kappa values of 0.59 for the same reader and 0.57 between readers (Table 3). Assessing synovial enhancement on a 0-2 scale showed good to moderate agreement, with kappa values of 0.68 for the same reader and 0.54 between readers (Fig 7). Similar, the
agreement for grading inflammation on a 0-4 scale according to the progressive system as suggested by Kellenberger (13) was good to moderate, with kappa values of 0.61 for the same reader and 0.45 between readers. The agreement for assessment of synovial thickening on a 0-2 scale and joint enhancement on a 0-2 scale, as suggested by Tolend (23), was moderate with kappa values of 0.43-0.44 both between readers and for the same reader. Subjective impression of thickened synovium was assessed with moderate agreement for the same reader and fair agreement between readers (Kappa 0.23).

**Bone marrow oedema/enhancement:** Assessment of bone marrow oedema on a 0-1 scale showed fair to moderate agreement, with kappa values of 0.35 for the same reader and 0.54 between readers.

The analysis of agreement of the variable bone marrow enhancement on a 0-2 scale was hampered by severely skewed distribution in one of the readings. Therefore, kappa analysis could not be performed. The variable showed a high proportion of absolute agreement (89%).

**Direct measurements of joint fluid:** The mean measurement of joint fluid in the upper compartment was 0.2 mm (median 0.1), with 95% limits of agreement of -0.6 to 0.4 mm between readers. The mean measurement of joint fluid in the lower compartment was 0.3 mm (median 0.1) with 95% limits of agreement of -1.0 to 0.7 mm between readers.

Based on the presented results a scoring system consisting of the following, precise imaging features could be considered (Table 4).

**DISCUSSION**

Of 25 commonly used MRI-based markers for TMJ changes in children with JIA, 13 showed sufficient precision, of which
11 were judged the more relevant to be included in a robust scoring system; seven within the osteochondral domain and four within the inflammatory domain (Table 4). An additional six markers performed well for the same reader, indicating that these be used with caution. Interestingly, several of the commonly used markers performed poorly, in particular assessment of synovial thickness and joint enhancement, as well as measurements of joint fluid.

**Osteochondral Domain**

In the present study, the most precise MRI marker suggestive of osteochondral damage was condylar volume on a 0–1 scale;
0 being within normal and 1 representing a clearly deformed condyle in the sagittal and/or coronal views, a feature not seen in children without JIA (14,19,28).

Likewise, assessment of osseous deformity as suggested by Tolend and Kellenberger using a progressive scoring system performed well, however, this grading system is based on a sequence of pathological changes, starting with a mildly flattened mandibular condyle and/or temporal bone (grade 1), followed by moderate flattening of the same structures (grade 2). Grade 3 is characterized by severe flattening of the mandibular condyle with loss of height, and/or completely flat temporal bone, and/or presence of small erosions/irregularities while grade 4 is defined as destruction of the temporomandibular joint by large erosions, fragmentation of the mandibular condyle, intra-articular ossification or bone apposition on mandibular condyle or temporal bone (13).

We have previously shown that a mildly flattened condyle is seen in around 20% of children without JIA, and as such represents a normal variation rather than early destructive change (19). Moreover, we experienced that both condylar irregularities and erosions may be present before severe condylar flattening, thus biasing a progressive system.

To overcome the abovementioned challenges, we suggest that the different markers are scored separately, and summarized. More specifically, that the most precise markers, such as loss of condylar volume, condylar shape and irregularities, and shape of articular eminence and glenoid fossa are used to construct a total damage score. Ideally, each of these components should be weighted, for example by using CBCT scores that are more fine-meshed in the osteochondral domain.

Several authors have explored the importance and incidence of disk abnormalities in TMJ (29–32), however, without addressing the precision of findings. We have now shown that assessing the disk as either normal or pathological represents a precise variable.

Subjective assessment of the condylar inclination showed good intra- and interobserver agreement. Previous studies have shown that the condylar inclination is symmetrical, and that it normally increases with age (14,19,28). Thus, the finding of asymmetric condylar inclination in a child with JIA could indicate growth disturbances secondary to the disease.

Inflammation

Four markers within the inflammation domain were considered of sufficient precision, both within and between readers, to be included in a future scoring system, namely joint fluid on a 0–2 scale, overall impression of inflammation on a 0–2 scale, synovial enhancement on a 0–2 scale and bone marrow oedema on a 0–1 scale (Table 4).

As for evaluation of joint fluid, the hybrid assessment with both continuous measurements and semi-qualitative evaluation suggested by Tolend (23) performed well in contrast to the subjective grading of the upper and lower compartments.
severe; high signal intensity diffusely involving 1 or both joint compartments. Adapted from reference 23

signal intensity focally exceeding signal perimeter of physiologic amount of joint synovial thickening, 2=Moderate/severe; entire joint compartment and/or presence of bone marrow oedema. 2= moderate in

≥ near by vessel) (33,34), re

can turn out to be rather inaccurate, with significant ve riations. These results are in line with others

1 mm width. No enhancement or enhancement confined to physiological joint fluid. 1= mild inflammation: Extension of joint enhancement exceeds that of physiological joint fluid but does not involve entire joint compartment and/or presence of bone marrow oedema. 2= moderate inflammation: Joint enhancement involves entire joint compartment or there is an enhancing joint effusion, 3= severe inflammation: Detectable synovial thickening in addition to increased joint enhancement or effusion, 4= joint space filled with and enlarged by pannus. Adapted from reference 13

1 mm width), 1=mild; >1 mm and ≤2 mm thickness at the point of maximum synovial thickening, 2= Moderate/severe; >2 mm thickness at the point of maximum synovial thickening. Adapted from reference 23

8= normal; high signal intensity confined to signal perimeter of normal amount of fluid on corresponding fluid-sensitive image, 1=mild; high signal intensity focally exceeding signal perimeter of physiologic amount of joint fluid on corresponding fluid-sensitive image, 2= moderate/severe; high signal intensity diffusely involving 1 or both joint compartments. Adapted from reference 23

Bone marrow enhancement (0-2)12 0.85 (0.64-1.00) n/a 62/70 (89%)

Abbreviations: JIA juvenile idiopathic arthritis; MRI, magnetic resonance imaging; TMJ, temporomandibular joint.

1 mm and

Bone marrow oedema (0-1)11 0.35 (0.18-0.51) 0.54 (0.28-0.80) 76/85 (89%)

Joint fluid

Joint fluid (0-2)1 0.74 (0.55-0.93) 0.71 (0.48-0.95) 76/81 (94%)

Overall impression of pathological joint fluid (0-1)4 0.71 (0.50-0.91) 0.40 (0.14-0.66) 68/82 (83%)

Subjective impression of joint fluid, lower compartment (0-2)3 0.69 (0.53-0.85) 0.29 (0.14-0.44) 41/81 (51%)

Synovial inflammation/enhancement/thickening

Overall impression of inflammation (0-2)5 0.59 (0.41-0.76) 0.57 (0.43-0.72) 59/79 (75%)

Synovial enhancement (0-2)6 0.68 (0.52-0.83) 0.54 (0.40-0.69) 53/78 (68%)

Inflammation, progressive system (0-4)7 0.61 (0.43-0.79) 0.45 (0.31-0.60) 49/74 (66%)

Synovial thickening (0-2)8 0.44 (0.22-0.65) 0.44 (0.22-0.66) 63/78 (81%)

Joint enhancement (0-2)9 0.43 (0.23-0.62) 0.44 (0.25-0.62) 58/78 (74%)

Subjective impression of synovial thickening (0-2)10 0.51 (0.32-0.70) 0.23 (0.08-0.38) 46/78 (59%)

Bone marrow oedema/enhancement

Bone marrow oedema (0-1)11 0.35 (0.18-0.51) 0.54 (0.28-0.80) 76/85 (89%)

Bone marrow enhancement (0-2)12 0.85 (0.64-1.00) n/a 62/70 (89%)

Overall impression of inflammation (0-2)5 0.59 (0.41-0.76) 0.57 (0.43-0.72) 59/79 (75%)

Inflammation, progressive system (0-4)7 0.61 (0.43-0.79) 0.45 (0.31-0.60) 49/74 (66%)

Joint enhancement (0-2)9 0.43 (0.23-0.62) 0.44 (0.25-0.62) 58/78 (74%)

Subjective impression of synovial thickening (0-2)10 0.51 (0.32-0.70) 0.23 (0.08-0.38) 46/78 (59%)

Bone marrow oedema/enhancement

Bone marrow oedema (0-1)11 0.35 (0.18-0.51) 0.54 (0.28-0.80) 76/85 (89%)

Bone marrow enhancement (0-2)12 0.85 (0.64-1.00) n/a 62/70 (89%)

Abbreviations: JIA juvenile idiopathic arthritis; MRI, magnetic resonance imaging; TMJ, temporomandibular joint.

1 mm fluid in joint recess, 1= small; >1 mm and ≤2 mm fluid in recess or involving entire joint compartment, 2=large; >2 mm fluid in recess or involving entire joint compartment. Adapted from reference 23

2= no joint fluid, 1= a thin line of fluid, 2= more than a thin line of fluid

3= no joint fluid, 1= a thin line of fluid, 2= more than a thin line of fluid

4= no, 1=yes

5= normal, includes normal synovial enhancement and a thin line of joint fluid, 1= mild inflammation, considered pathological, 2= moderate/severe inflammation

6= subtle synovial enhancement, 1= mildly increased synovial enhancement, 2= moderately to severe synovial enhancement (signal intensity ≥ near by vessel)

7= no inflammation: No or small amounts of joint fluid in any recess, with ≤ 1 mm width. No enhancement or enhancement confined to physiological joint fluid. 1= mild inflammation: Extension of joint enhancement exceeds that of physiological joint fluid but does not involve entire joint compartment and/or presence of bone marrow oedema. 2= moderate inflammation: Joint enhancement involves entire joint compartment or there is an enhancing joint effusion, 3= severe inflammation: Detectable synovial thickening in addition to increased joint enhancement or effusion, 4= joint space filled with and enlarged by pannus. Adapted from reference 13

8= no; synovium visible (apparent joint compartment ≤1 mm width), 1= mild; >1 and ≤2 mm thickness at the point of maximum synovial thickening, 2= Moderate/severe; >2 mm thickness at the point of maximum synovial thickening. Adapted from reference 23

9= normal; high signal intensity confined to signal perimeter of normal amount of fluid on corresponding fluid-sensitive image, 1= mild; high signal intensity focally exceeding signal perimeter of physiologic amount of joint fluid on corresponding fluid-sensitive image, 2= moderate/severe; high signal intensity diffusely involving 1 or both joint compartments. Adapted from reference 23

10= no; no enhancement, 1= mild thickening, 2= moderate/severe thickening

11= absent, 1= present

12= No enhancement, 1= subtle enhancement, what is considered normal, 2= increased, pathological enhancement

separately. However, direct continuous measurement of joint fluid turned out to be rather inaccurate, with significant variation between observers. These results are in line with others (33,34), reflecting difficulties in measuring small distances. To overcome the challenges associated with continuous measurements, we tested the subjective variable “overall impression of joint fluid,” although with disappointing results between readers. In conclusion, the mechanisms providing high precision to the variable “joint fluid” are not fully understood, but probably depends on a thorough understanding of the normal appearances of fluid in the recesses and joint compartments.

The variable “overall impression of inflammation 0–2” depends explicitly on the subjective understanding of normal, age-related and physiologic findings in the TMJ. At the same time the variable demands the reader to define, from his/her own understanding, the difference between normal findings and inflammation. Like the binary variables “overall impression of pathological joint fluid” and “loss of condylar volume” this type of variables has not been tested in other publications. This study shows that the variable as such is precise enough to be studied further.

Opposite to the marker synovial enhancement, which was based on pre and post T1-fat suppressed images only,
assessment of joint enhancement, as suggested by Tolend et al (23), is based on both fluid-sensitive images as well as post-contrast T1-weighted fat-suppressed images. According to their 0-2 score, mild inflammation is defined as high signal intensity focally exceeding signal perimeter of physiologic amount of joint fluid on corresponding fluid-sensitive image.

Figure 6. Sagittal/oblique T2 weighted image with fat saturation of a 15-year-old boy with persistent oligoarthritic JIA and disease duration 4 years, showing large amount of joint fluid (grade 2) in the right temporomandibular joint.

Figure 7. Sagittal/oblique T1 weighted image with fat saturation after intravenous contrast of a 6-year-old girl with polyarthritic JIA and disease duration 4 years, showing moderate to severe synovial enhancement (grade 2) (white arrow).
while moderate to severe inflammation is characterized by high signal intensity diffusely involving one or both joint compartments. We observed numerous cases showing subtle, focal, synovial contrast enhancement on T1-weighted fat-suppressed images, with no fluid seen on T2-weighted images, i.e., a grade 0 according to the synovial enhancement score and a grade 1 according to the joint enhancement score. Thus, it seems that combining pre-gadolinium fluid-sensitive images with post-gadolinium fat-suppressed T1-weighted images tends to overestimate pathology, possibly accentuated by slightly different imaging parameters on T1- and T2-weighted images. These difficulties are reflected in the slightly lower agreement between readers for the joint enhancement score as compared to the synovial enhancement score.

We found acceptable agreement between readers for the assessment of condylar bone marrow oedema. In adults with rheumatoid arthritis of the wrist, the precision of this variable is addressed in numerous publications (35-37) with results supporting the findings in our study. However, the precision in these studies is measured as a sum of scores along an ICC-scale so the transferability of the results to the mandibles of a pediatric population is questionable. In their study on MRI and CBCT Koos and colleagues report “no relevant interobserver differences” which per se supports our findings, even though their statement could be more elaborated (21). In 2014, Vaid studied the composite variable including contrast enhancement, joint fluid, synovial thickening and bone marrow oedema with a weighted kappa of 0.51. The complexity of their composite variable makes it hard to say if their results support or contradict our findings (22). Lastly, Tolend tested both a binary and a 4-graded version of the variable bone marrow oedema with ICC-results that do not support our findings (dICC 0.01 and 0.06, avICC 0.61 and 0.57). Still, bone marrow oedema is considered an important marker, as oedema/osteitis is believed to represent relevant pathology in rheumatology. Taken together, we suggest the variable should be part of a future scoring system.

As for the progressive inflammation score, this is based on a fixed sequence of changes, like that described for the osteochondral domain. We experienced, in a small number of TMJs, that this sequence was violated, in that subtle synovial thickening was present without synovial enhancement or joint effusion. Thus, according to the progressive system, these joints should be scored as a grade 3 inflammation. Seen together with the difficulties in defining synovial thickening this represents a bias in the progressive system. As in the osteochondral domain, we suggest that each variable be scored separately, and subsequently summarized.

Similar to bone marrow oedema, the variable “bone marrow enhancement” aims to describe an important and closely related part of the rheumatologic pathology, namely osteitis and increased perfusion of the intraosseous part of the condyle. However, we noted that virtually all condyles demonstrated some degree of enhancement, also when compared to the mandibular ramus, which corresponds to a grade 1 in the binary system as proposed by Tolend. The 3-graded system proposed in this study shows a slight differentiation between assumed normal and pathological enhancement with a high proportion of absolute agreement, although kappa analysis could not be performed due to skewed distribution of the findings. We note that Tolend and co-workers do not present data on the repeatability of the binary variant of this variable. The assumed importance of the pathological process, in combination with the paucity of data on the precision of the variable makes it an interesting topic for further research, but as per today it should not be included in a robust scoring system.

Except for the inflammation score in the progressive system, all these scores are relatively crude, however, previous studies have demonstrated difficulties in establishing reliable, fine-meshed imaging markers for the inflammatory domain (23).

In general, we found that the intra-observer agreement was better than agreement between observers, despite thorough calibration and the use of a reference atlas. This is not unexpected and similar results has been shown in numerous earlier publications. Still, we assume that this finding underscores the importance of performing clinical, JIA-related radiology reporting in a small environment of subspecialists with a special interest in JIA.

Limitations and Strengths

We acknowledge that our study has shortcomings. First, the use of Cohens Kappa has limitations especially in analysis of datasets with skewed distribution (38,39). To compensate for this, we chose to both present the proportion of absolute agreement and
### APPENDIX A. MRI Protocol for TMJ Imaging in the NorJIA-Study

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Plane</th>
<th>Fat Saturation</th>
<th>TR (ms)</th>
<th>TE (ms)</th>
<th>Slice Thickness (mm)</th>
<th>Gap (mm)</th>
<th>FOV (mm)</th>
<th>Matrix</th>
<th>Number of Averages</th>
<th>Flip Angle</th>
<th>ETL</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1-MPRAGE$^e$</td>
<td>Sagittal</td>
<td>No</td>
<td>2000</td>
<td>2.26</td>
<td>1</td>
<td>250 × 250</td>
<td>256 × 256</td>
<td>1</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1-TSE$^f$</td>
<td>Coronal</td>
<td>No</td>
<td>826</td>
<td>7.9</td>
<td>2</td>
<td>2.2</td>
<td>179 × 179</td>
<td>448 × 359</td>
<td>3</td>
<td>131</td>
<td>4</td>
</tr>
<tr>
<td>T2-TSE</td>
<td>Sagittal/oblique</td>
<td>Yes (CHESS)$^g$</td>
<td>3530</td>
<td>71</td>
<td>2</td>
<td>2.2</td>
<td>150 × 150</td>
<td>448 × 314</td>
<td>2</td>
<td>150</td>
<td>9</td>
</tr>
<tr>
<td>PD$^h$-TSE</td>
<td>Sagittal/oblique</td>
<td>No</td>
<td>3470</td>
<td>22</td>
<td>2</td>
<td>2.2</td>
<td>150 × 150</td>
<td>448 × 314</td>
<td>2</td>
<td>139</td>
<td>10</td>
</tr>
<tr>
<td>T1-TSE</td>
<td>Sagittal/oblique</td>
<td>Yes</td>
<td>774</td>
<td>8.1</td>
<td>2</td>
<td>2.2</td>
<td>150 × 150</td>
<td>384 × 269</td>
<td>3</td>
<td>122</td>
<td>4</td>
</tr>
<tr>
<td>T1-TSE-Dixon</td>
<td>Coronal</td>
<td>2-point Dixon</td>
<td>650</td>
<td>12</td>
<td>2</td>
<td>2.2</td>
<td>180 × 180</td>
<td>448 × 358</td>
<td>1</td>
<td>144</td>
<td>4</td>
</tr>
<tr>
<td>Vibe-Twist</td>
<td>Coronal</td>
<td>No</td>
<td>3.78</td>
<td>1.03</td>
<td>2</td>
<td></td>
<td>210 × 210</td>
<td>160 × 144</td>
<td>1</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>T1-TSE</td>
<td>Sagittal/oblique</td>
<td>Yes (CHESS)</td>
<td>774</td>
<td>8.1</td>
<td>2</td>
<td>2.2</td>
<td>150 × 150</td>
<td>384 × 269</td>
<td>2</td>
<td>122</td>
<td>4</td>
</tr>
<tr>
<td>PD-TSE, open mouth</td>
<td>Sagittal/oblique</td>
<td>No</td>
<td>3470</td>
<td>22</td>
<td>2</td>
<td>2.2</td>
<td>150 × 150</td>
<td>448 × 314</td>
<td>2</td>
<td>139</td>
<td>10</td>
</tr>
</tbody>
</table>

MRI, magnetic resonance imaging; NorJIA, the Norwegian juvenile idiopathic arthritis study; TMJ, temporomandibular joint.

$^a$ Repetition time (TR)
$^b$ Echo time (TE)
$^c$ Field of view (FOV)
$^d$ Echo train length (ETL)
$^e$ Magnetization prepared rapid gradient echo (MPRAGE)
$^f$ Turbo spin-echo (TSE)
$^g$ Chemical shift selective fat saturation (CHESS)
$^h$ Proton density (PD)
the distribution of findings for each variable. We assume this to be a more correct and transparent way of presenting the data than other statistical models which would introduce other sources of error. Next, the study was performed with two readers only, aiming to examine the potential of a scoring system given optimal conditions, rather than assessing its performance in a clinical setting. And lastly, the distribution of findings for some of the variables under investigation was skewed, thus hindering statistical analysis to be performed. The strengths of our study include the high numbers, the meticulous standardization of scoring systems and measurements, and the construction of an atlas for optimizing precision.

**CONCLUSION**

We propose a robust scoring system for the assessment of TMJ involvement in children with JIA including four variables in the inflammatory domain and seven variables in the osteochondral domain. Further studies on clinical validity of these markers are needed.

**ACKNOWLEDGMENTS**

This work was partially funded by the Liaison Committee between the Central Norway Regional Health Authority.
### APPENDIX C. Imaging Features for Scoring of Temporomandibular Joints in the Inflammatory Domain by Magnetic Resonance Imaging

<table>
<thead>
<tr>
<th>Imaging Feature</th>
<th>Definition/Image Plane</th>
<th>Grading</th>
</tr>
</thead>
</table>
| Overall impression of inflammation                   | All available images                                                                   | 0=Normal; includes normal synovial enhancement and a thin line of joint fluid  
1=Mild inflammation; considered pathological  
2=Moderate to severe inflammation                     |
| Overall impression of pathological joint fluid       | All available images                                                                   | 0=No  
1=Yes                                                                 |
| Synovial enhancement                                 | Signal intensity of the synovium, based on sagittal/oblique T1-fs pre contrast and sagittal/oblique T1-fs post-contrast images | 0= Subtle synovial enhancement, what is believed as normal  
1=Mildly increased synovial enhancement  
2= Moderately to severely increased synovial enhancement (signal intensity ≥ nearby vessel) |
| Subjective impression of joint fluid, upper compartment | Sagittal/oblique T2-fat saturated images                                                | 0=No signal  
1=A thin line of signal  
2=More than a thin line                                      |
| Subjective impression of joint fluid, lower compartment | Sagittal/oblique T2-fat saturated images                                                | 0=No signal  
1=A thin line of signal  
2=More than a thin line                                      |
| Joint enhancement<sup>a</sup>                          | Signal intensity of the synovium, capsule and joint fluid higher than that of muscle on post contrast T1-fat saturated images | 0=Normal; high signal intensity confined to signal perimeter of normal amount of fluid on corresponding fluid-sensitive image  
1=Mild; high signal intensity focally exceeding signal perimeter of physiologic amount of joint fluid on corresponding fluid-sensitive image  
2= Moderate/ severe; high signal intensity diffusely involving 1 or both joint compartments |
| Joint fluid<sup>b</sup>                                | Increased joint fluid with isointense signaling of joint space compared to that of cerebrospinal fluid on fluid-sensitive images | 0=Absent; ≤1mm fluid in recess  
1=Small; >1 and ≤2 mm in recess or involving entire joint compartment  
2=Large; >2 mm fluid in recess or involving entire joint compartment |
| Synovial enhancement                                 | Sagittal/oblique T1-fat saturated images post iv contrast                              | 0=Subtle synovial enhancement  
1=Mildly increased synovial enhancement  
2=Moderate to severe synovial enhancement (signal intensity ≥ nearby vessel) |
| Synovial thickening<sup>a</sup>                        | Sagittal/oblique T2 fat-saturated images                                                | 0=Absent; no synovium visible (apparent joint compartment ≤1 mm width)  
1=Mild; >1 and <2 mm thickness at the point of maximum synovial thickening,  
2=Moderate/severe; >2 mm thickness at the point of maximum synovial thickening |
| Joint enhancement<sup>a</sup>                          | Sagittal/oblique T1-fat saturated images post iv contrast and sagittal/oblique T2 fat-saturated images | 0=Normal; high signal intensity confined to signal perimeter of normal amount of fluid on corresponding fluid-sensitive image  
1=Mild; high signal intensity focally exceeding signal perimeter of physiologic amount of joint fluid on corresponding fluid-sensitive image  
2= Moderate/severe; high signal intensity diffusely involving 1 or both joint compartments |

<sup>a</sup>Joint enhancement involves the use of post-contrast images to assess enhancement.

<sup>b</sup>Joint fluid quantification involves measuring the amount and distribution of joint fluid using fluid-sensitive images.

(continued)
APPENDIX C. (Continued)

<table>
<thead>
<tr>
<th>Imaging Feature</th>
<th>Definition/Image Plane</th>
<th>Grading</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjective impression of synovial thickening</td>
<td>Sagittal/oblique T2 fat-saturated images</td>
<td>0=No thickening</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1=Mild thickening</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2=Moderate/severe thickening</td>
</tr>
<tr>
<td>Bone marrow oedema</td>
<td>Coronal T1 images and sagittal/oblique T2 fat-saturated images</td>
<td>0=Absent</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1=Present</td>
</tr>
<tr>
<td>Bone marrow enhancement</td>
<td>Sagittal/oblique T1-fat saturated images before and post iv contrast</td>
<td>0=No enhancement</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1=Subtle enhancement, considered normal</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2=Increased, pathological enhancement</td>
</tr>
</tbody>
</table>

* Adapted from reference 23.

APPENDIX D. Progressive Scoring System for Assessing Inflammation and Osseous Deformity of Temporomandibular Joint by Magnetic Resonance Imaginga

<table>
<thead>
<tr>
<th>Inflammation</th>
<th>Osseous Deformity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade 0</td>
<td>No inflammation:</td>
</tr>
<tr>
<td></td>
<td>No or small amounts of fluid in any recess with ≤1 mm width.</td>
</tr>
<tr>
<td></td>
<td>No enhancement or enhancement confined to physiological joint fluid.</td>
</tr>
<tr>
<td>Grade 1</td>
<td>Mild inflammation:</td>
</tr>
<tr>
<td></td>
<td>Extension of joint enhancement exceeds that of physiological joint fluid but does not involve entire joint compartment and/or presence of bone marrow oedema.</td>
</tr>
<tr>
<td>Grade 2</td>
<td>Moderate inflammation:</td>
</tr>
<tr>
<td></td>
<td>Joint enhancement involves entire joint compartment or there is an enhancing joint effusion</td>
</tr>
<tr>
<td>Grade 3</td>
<td>Severe inflammation:</td>
</tr>
<tr>
<td></td>
<td>Detectable synovial thickening in addition to increased joint enhancement or effusion.</td>
</tr>
<tr>
<td>Grade 4</td>
<td>Joint space filled with and enlarged by pannus</td>
</tr>
<tr>
<td></td>
<td>Normal shape of temporal bone and mandibular condyle according to age:</td>
</tr>
<tr>
<td></td>
<td>S-shaped articular eminence/glenoid fossa.</td>
</tr>
<tr>
<td></td>
<td>Round condyle (young patient)</td>
</tr>
<tr>
<td></td>
<td>Less rounded, more angular appearing condyle (older patient)</td>
</tr>
<tr>
<td></td>
<td>Smooth subchondral bone contour</td>
</tr>
<tr>
<td>Grade 1</td>
<td>Mild flattening of the mandibular condyle and/or temporal bone.</td>
</tr>
<tr>
<td>Grade 2</td>
<td>Moderate flattening of the mandibular condyle and/or temporal bone</td>
</tr>
<tr>
<td>Grade 3</td>
<td>Severe flattening of the mandibular condyle with loss of height, and/or completely flat temporal bone, and/or presence of small erosions/irregularities</td>
</tr>
<tr>
<td>Grade 4</td>
<td>“Destruction” of temporomandibular joint by large erosions, fragmentation of the mandibular condyle, intra-articular ossification or bone apposition on mandibular condyle or temporal bone.</td>
</tr>
</tbody>
</table>

* Adapted from reference 14.

APPENDIX E. Grading of Position of the Mandibular Condyle and Disc Displacement

<table>
<thead>
<tr>
<th>Position of the Condyle in the Glenoid Fossa</th>
<th>Disc Displacement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade 0</td>
<td>Neutral</td>
</tr>
<tr>
<td>Grade 1</td>
<td>Anterior</td>
</tr>
<tr>
<td>Grade 2</td>
<td>Posterior</td>
</tr>
<tr>
<td>Grade 3</td>
<td>Medial</td>
</tr>
<tr>
<td>Grade 4</td>
<td>Lateral</td>
</tr>
<tr>
<td>Grade 5</td>
<td>Superior</td>
</tr>
<tr>
<td>Grade 6</td>
<td>Inferior</td>
</tr>
</tbody>
</table>

1376
REFERENCES


36. Olech E, Crues JV, 3rd Yocum DE, et al. Bone marrow edema is the most frequent finding for rheumatoid arthritis (RA) on noncontrast magnetic resonance imaging of the hands and wrists: a comparison of patients with RA and healthy controls. J Rheumatol 2010; 37(2):265–274.

