Web-based calibration of observers using MRI of the temporomandibular joint

K Hellén-Halme*,1, L Hollender2, M Janda3 and A Petersson1

1Department of Oral and Maxillofacial Radiology, Faculty of Odontology, Malmö University, Malmö, Sweden; 2Department of Oral Medicine, School of Dentistry, University of Washington, Seattle, WA, USA; 3Public Dental Health Service of Skåne, Lund, Sweden

Objectives: Research diagnostic criteria for temporomandibular disorders (RDC/TMDs) were proposed in 1992 with the aim of standardizing and testing methods for diagnosing TMDs. RDC/TMDs have so far been lacking standardized methods for imaging and criteria for imaging diagnosis of disorders of the temporomandibular joint (TMJ). Criteria for disorders of the TMJ have recently been proposed for MRI. The aim of this study was to publish MR images of the TMJ on the web and to test the agreement of five observers using the criteria.

Methods: 20 cases of MRI of the TMJ were published on the web together with the criteria. The MR images were taken in closed and open mouth positions, and sagittal and coronal views. Five observers diagnosed disc position, disc shape, joint effusion and loose calcified bodies in the TMJ.

Results: In all cases except one, three or more observers agreed upon the diagnosis. All agreed on whether a loose calcified body was present or not. The second best agreement was obtained for disc position in the sagittal view, where all observers agreed in 16 of the 20 cases. For disc position in the coronal view and the evaluation of the disc shape, observer agreement was lower.

Conclusion: Criteria were useful in order to standardize and simplify evaluation and thereby probably increase the diagnostic outcome among different observers for MRI of the TMJ. We recommend that the criteria be used internationally to facilitate comparisons between different studies.

Keywords: magnetic resonance imaging; observer variation; temporomandibular disorders; temporomandibular joint

Introduction

Evaluation of research depends on complete and accurate reporting. The Standards for Reporting of Diagnostic Accuracy (STARD) steering committee has taken an initiative to improve reporting of studies of diagnostic accuracy, and a checklist for reporting diagnostic accuracy studies has been published.1 The list includes methods to test the reproducibility of observers.

Research diagnostic criteria for temporomandibular disorders (RDC/TMDs) were proposed by Dworkin and LeResche2 in 1992 with the aim of standardizing and testing methods for diagnosing TMDs. Today the RDC/TMD is an international consortium with the goal of advancing the scientific knowledge of TMDs and related pain conditions through authenticated tools for international use, multisite and cross-cultural research studies based on standardized assessment, annual meetings and international consensus workshops (http://www.rdc-tmdinternational.org/). RDC/TMD has so far been lacking standardized methods for imaging and imaging diagnosis of disorders of the temporomandibular joint (TMJ). Recent reviews have concluded that there is a lack of consistency in diagnostic criteria and the classification system for MRI of the TMJ.3,4 Criteria have been proposed for
CT and MRI within the framework of RDC/TMD, and a high reliability was found for assessing osteoarthritis (OA) using CT and for disc position and effusion using MRI.5 Previously it has been shown that interobserver agreement was increased with the aid of reference films.6 Publishing reference images and creating a calibration program on the web could provide a way of increasing the coherence among observers using MRI of the TMJ.

The aim was therefore to test if the criteria proposed by Ahmad et al5 could be useful when applied by several observers using MR images of the TMJ published on the web. A future aim is to use the images as reference images in calibration of observers using MRI of the TMJ. The hypothesis was that criteria could be a way to facilitate and increase observer agreement for MRI of the TMJ.

Materials and methods

The images in this study were obtained during 2008 and the first 6 months of 2009 at the Department of Radiology, Skåne University Hospital, Malmö, Sweden. All examinations were made in a Siemens Magnetom Vision 1.5 T machine (Siemens, Erlangen, Germany) with a bilateral TMJ surface coil. For the closed mouth position, the patients were instructed to close their mouth with their teeth in maximal contact. In the open mouth position, a stepped plastic bite block was placed between the upper and lower incisors. Use of the bite block was practised before the examination. The patients were asked to open their mouths as much as they could without experiencing unbearable pain. The examination included sagittal and coronal imaging of the TMJ in the closed mouth position, and sagittal images in the open mouth position. Sagittal sections were orientated perpendicular to the long axis of the condyle, and coronal sections were taken parallel to the long axis. The orientation was determined using axial localizers, one for the closed mouth position and one for the open mouth position. Proton density (PD) and T2 weighted images were acquired using a double-echo turbo-spin echo sequence [echo time (TE), 15/105 ms; repetition time (TR), 2400 ms (sagittal) or 2000 ms (coronal); turbo factor, 7; field of view (FOV), 160 × 160 mm2; matrix, 255 × 512; and slice thickness, 3 mm]. Each sequence yielded 8–10 sections.

Selection and evaluation of images

The criteria for the non-osseous components of the TMJ described by Ahmad et al5 were used for selection of MR images in order to create samples of different scenarios relative to disc position, disc shape, effusion and presence of loose calcified bodies in the TMJ. The aim was to select five cases of each scenario, but for loose calcified bodies only one case was found. The different criteria are listed in Table 1. Disc diagnoses for the TMJ using MRI according to Ahmad et al5 are given in Table 2. 25 samples of MR images exemplifying the different criteria were published on the internet. All cases had previously been evaluated and regarded as accept-able examinations at the hospital. Images from different examinations were included with differences in image quality. The images shown were PD sagittal images, closed and open mouth, T2 sagittal images closed mouth and PD coronal image closed mouth (Figure 1). Five observers evaluated the images from the web, four of whom had more than 15 years’ experience in evaluating MRI of the TMJ; the fifth observer was a post-graduate student. In five cases, one or more of the observers thought that the image quality was unacceptable when the cases were evaluated, and these five cases were excluded. The remaining 20 cases were used in the evaluation of the criteria and they are presently published on the internet together with instructions on how to use the program (http://www.mah.se/od/rad/tmj).

The diagnoses in Table 2 were used for determining the disc position in the sagittal plane. In all other evaluations the criteria shown in Table 1 were used. The observer had the opportunity to read the criteria at any time when observing the images (Figure 1).

Results

The results of the assessments of the five observers are shown in Table 3. The agreements of the observers for the 20 cases are shown in Table 4. In all cases except one, three or more observers agreed upon the diagnosis. In one case of joint effusion, two observers agreed that there was no effusion; two observers indicated slight effusion, while one thought it was frank. All observers agreed on whether a loose calcified body was present or not. However, the presence of a loose calcified body was shown only in one case. The second best agreement was obtained for diagnosis of the disc position in the sagittal view, where all 5 observers agreed in 16 of the 20 cases. For disc position in the coronal view and the evaluation of the disc shape, the agreement was lower. Total agreement for the evaluation of disc shape was achieved in only 6 of the 20 cases.

Discussion

Diagnostic radiology is part of a larger system where the goal is to treat patients with efficacy and efficiency.7 The accuracy of a diagnostic method can be defined in many different ways. Sensitivity and specificity are two measures that are often used to describe the diagnostic accuracy. Another important measure is observer agreement, i.e. how different observers agree when reading the same images. A systematic literature review of the efficacy of MRI of the TMJ in diagnosing degenerative and inflammatory joint diseases showed that no conclusion concerning the diagnostic efficacy could be reached.3 One problem was that diagnostic criteria varied between publications. In order to
improve the diagnostic accuracy for MR images of the TMJ, specific imaging criteria have been developed by Ahmad et al.\textsuperscript{5} The criteria are intended to be used in the RDC/TMD classification system. This system is used in many countries in order to simplify and guide how an examination of a patient with TMD should be performed. The criteria can also serve as a base when calibrating different clinicians. This aim of this study was to evaluate whether the guidelines for MRI of the TMJ could be used as a way to calibrate observers for better diagnostic accuracy. The reliability of the criteria in the study by Ahmad et al\textsuperscript{5} was examined with observers who were calibrated in an extensive training programme. The mean kappa values in the study by Ahmad et al\textsuperscript{5} reached a value of 0.84 for any disc displacement and 0.64 for effusion. The kappa value as a statistical measurement is calculated from data which form a square table from two observers with different options to choose from. In this study, we chose therefore to report the observer agreement in percentage figures.

The internet has made it possible to reach out to many people. Its use for the purpose of this study is, however, rather new. The attempt was to make a website with the purpose of presenting different scenarios of TMJ conditions in order to evaluate if criteria such as those published by Ahmad et al\textsuperscript{5} can be helpful to increase the observer agreement for MR images of the TMJ. In order to represent a clinical situation, the images chosen were not optimized with regard to image quality. All MR examinations had been used for diagnosing patients with TMD and approved by the radiologist who made the initial report. Further, the five observers agreed upon using the images in this investigation. It has been reported that higher quality MR images lead to a better interobserver agreement,\textsuperscript{8} and therefore the procedure used in this investigation can be subject to discussion. If, in the future, a website for calibration and learning to diagnose different conditions of the TMJ is published on the internet, the highest quality MR images should be chosen and revised continuously as the MRI technique develops.

All observers in our study except one were experienced in the use of MRI of the TMJ. They had several years of clinical experience and extensive research in the area. The post-graduate student, however, had substantial experience of oral radiography, and the decisions made by the student did not affect the final results. This fact could indicate that MR images of the TMJ can be evaluated with the aid of written criteria if the observer has previous knowledge of radiological interpretation. The MRI technique could further be refined for obtaining the coronal images of the disc,\textsuperscript{9} as this technique was used in the article by Ahmad et al.\textsuperscript{5}

**Table 1** Non-osseous component assessment criteria using MRI of the temporomandibular joint from Ahmad et al\textsuperscript{5}

| Disc position: closed mouth sagittal and axially corrected coronal views |
|-----------------------------|-----------------------------|
| **Normal disc position**    |                             |
| i. In the sagittal plane, relative to the superior aspect of the condyle, the border between the low signal of the disc and the high signal of the retrodiscal tissue is located between the 11:30 and 12:30 clock positions |
| ii. In the sagittal plane, the intermediate zone is located between the anterosuperior aspect of the condyle and the posteroinferior aspect of the articular eminence |
| iii. In the oblique coronal plane, the disc is centred between the condyle and eminence in the medial, central and lateral parts |
| **Indeterminate**           |                             |
| i. In the sagittal plane, relative to the superior aspect of the condyle, the low signal of the disc and the high signal of the retrodiscal tissue are located anterior to the 11:30 position, but the condyle contacts the intermediate zone located between the anterosuperior aspect of the condyle and the posteroinferior aspect of the articular eminence, or |
| ii. In the sagittal plane, relative to the superior aspect of the condyle, the low signal of the disc and the high signal of the retrodiscal tissue are located between the 11:30 and 12:30 clock positions, but the intermediate zone of the disc is located anterior to the condyle, and |
| iii. In the axially corrected coronal plane, the disc is positioned between the condyle and eminence in the medial, central and lateral parts |

**Disc not visible:** Neither signal intensity nor outline make it possible to define a structure as the disc

**Disc position: open mouth sagittal views**

**Normal disc position:** The intermediate zone is located between the condyle and the articular eminence

**Persistent disc displacement:** The intermediate zone is located anterior to the condylar head

**Disc not visible:** Neither signal intensity nor outlines make it possible to define a shape of the disc

**Disc shape: closed mouth sagittal views**

| Normal: The disc in the sagittal plane is biconcave |
| Deformed: All shapes other than biconcave in the sagittal plane |
| Disc not visible: Neither signal intensity nor outlines make it possible to define a shape of the disc |

**Effusion: open or closed mouth sagittal views**

**None:** No bright signal in either joint space in the $T_2$ weighted images

**Slight effusion:** A bright signal in either joint space that conforms to the contours of the disc, fossa/articular eminence and/or condyle

**Frank effusion:** A bright signal in either joint space that extends beyond the osseous contours of the fossa/articular eminence and/or condyle and has a convex configuration in the anterior or posterior recesses

**Loose calcified bodies: closed mouth sagittal views**

Single or multiple discrete low-signal intensity objects are present in the joint spaces, and are not attached to the condyle, fossa or eminence in any plane

Ahmad et al.\textsuperscript{5} The criteria are intended to be used in the TMJ, specific imaging criteria have been developed by Ahmad et al.\textsuperscript{5} The criteria are intended to be used in the RDC/TMD classification system. This system is used in many countries in order to simplify and guide how an examination of a patient with TMD should be performed. The criteria can also serve as a base when calibrating different clinicians. This aim of this study was to evaluate whether the guidelines for MRI of the TMJ could be used as a way to calibrate observers for better diagnostic accuracy. The reliability of the criteria in the study by Ahmad et al\textsuperscript{5} was examined with observers who were calibrated in an extensive training programme. The mean kappa values in the study by Ahmad et al\textsuperscript{5} reached a value of 0.84 for any disc displacement and 0.64 for effusion. The kappa value as a statistical measurement is calculated from data which form a square table from two observers with different options to choose from. In this study, we chose therefore to report the observer agreement in percentage figures.

The internet has made it possible to reach out to many people. Its use for the purpose of this study is, however, rather new. The attempt was to make a website with the purpose of presenting different scenarios of TMJ conditions in order to evaluate if criteria such as those published by Ahmad et al\textsuperscript{5} can be helpful to increase the observer agreement for MR images of the TMJ. In order to represent a clinical situation, the images chosen were not optimized with regard to image quality. All MR examinations had been used for diagnosing patients with TMD and approved by the radiologist who made the initial report. Further, the five observers agreed upon using the images in this investigation. It has been reported that higher quality MR images lead to a better interobserver agreement,\textsuperscript{8} and therefore the procedure used in this investigation can be subject to discussion. If, in the future, a website for calibration and learning to diagnose different conditions of the TMJ is published on the internet, the highest quality MR images should be chosen and revised continuously as the MRI technique develops.

All observers in our study except one were experienced in the use of MRI of the TMJ. They had several years of clinical experience and extensive research in the area. The post-graduate student, however, had substantial experience of oral radiography, and the decisions made by the student did not affect the final results. This fact could indicate that MR images of the TMJ can be evaluated with the aid of written criteria if the observer has previous knowledge of radiological interpretation. The MRI technique could further be refined for obtaining the coronal images of the disc,\textsuperscript{9} as this technique was used in the article by Ahmad et al.\textsuperscript{5}
Several studies of observer agreement at MRI of the TMJ have been published.\textsuperscript{10–17} It is difficult to compare the studies as different criteria were used, for instance for disc position. Widmalm et al\textsuperscript{16} concluded that a diagnosis of TMD based on MRI examination protocols made by a single examiner should not be accepted, as the diagnosis may be misleading. A recent publication concluded that examiners do not demonstrate reproducibility in the interpretation of MRI of the TMJ.\textsuperscript{17} It seems that the diagnosis “anterior disc displacement without reduction” reached the highest kappa values and agreements between observers in different publications.\textsuperscript{8,13,15–17} In 80\% of the cases the five observers agreed on the diagnosis of the disc in the sagittal plane, a finding that is similar to our results.

The observations of disc position in the coronal plane, joint effusion and disc shape reached much lower agreement between the observers. This is consistent with the findings of Takano et al,\textsuperscript{15} Widmalm et al,\textsuperscript{16} Butzke et al\textsuperscript{17} and Schmitter et al.\textsuperscript{8} However, better agreement of the disc shape was found by Tasaki et al.\textsuperscript{10} One way of increasing the agreement between observers is to reduce the number of categories. This was proposed by Takano et al\textsuperscript{15} regarding the finding of joint effusion, and in our study, if the finding of joint fluid was dichotomized to “yes” or “no”, the observer agreement would increase to 80\%. Therefore, we propose that the criteria described by Ahmad et al\textsuperscript{5} are changed for joint effusion to be stated as present or not. Regarding the judgement of disc deformity, a better definition would perhaps be in case of a disc displacement with reduction that the shape of the disc is judged in the open (reduced) position. If the shape of the disc is normal in the open position then the judgement should be that of a normal disc shape, since the deformity seen in the closed was only “temporary”. This was also suggested by Ahmad et al,\textsuperscript{3} and it will be used in the future presentation of the criteria.

Figure 1 Page on the internet showing, in the left and right columns, MR images of the temporomandibular joint (TMJ) in the closed mouth position [proton density (PD) and $T_2$], open mouth position (PD) and coronal view (PD). Images can be shown in larger magnification in the middle fields. In the row in the bottom, alternative diagnoses can be chosen. In the corner to the left, it is possible to click on the different buttons and the written criteria will be shown.
The MR images in the sagittal plane showed closed and open mouth positions of the TMJ. The images formed the bases for giving diagnoses such as disc displacement with or without reduction. The coronal images could only be used for identifying the disc position in the closed mouth position. This limits the number of possible diagnoses of the disc position in the coronal plane and a combination of diagnoses in the sagittal and coronal plane. However, this decision was taken by the RDC/TMD consortium, as more emphasis was put on disc displacement in the anteroposterior direction than in the mediolateral direction. A diagnostic examination should ideally lead to different treatment options. Today there are no such possibilities in RDC/TMD, and this might be a suggestion for future development of the system.

The criteria proposed by Ahmad et al\textsuperscript{5} are important to follow in order to standardize the reporting of TMD in MRI of the TMJ. We suggest that the criteria are used internationally following the intentions of RDC/TMD.

### Table 2
Disc diagnosis in the sagittal plane for the temporomandibular joint using MRI according to Ahmad et al\textsuperscript{5}

<table>
<thead>
<tr>
<th>Case</th>
<th>Disc position sagittal</th>
<th>Disc position coronal</th>
<th>Disc shape</th>
<th>Effusion</th>
<th>Loose calcified body</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 3
Results of the five assessments for Observers a–e

<table>
<thead>
<tr>
<th>Case number</th>
<th>Disc position sagittal</th>
<th>Disc position coronal</th>
<th>Disc shape</th>
<th>Effusion</th>
<th>Loose calcified body</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>5 5 5 5 5 5 5 5 5 5 5 3</td>
<td>3 5 5 5 3 3 3 3 4 3 3 4</td>
<td>3 4 5 3 4 4 4 4 5 3 4 3</td>
<td>3 5 5 5 3 3 3 3 4 3 4 3</td>
<td>3 5 5 5 3 3 3 3 4 3 4 3</td>
</tr>
<tr>
<td>B</td>
<td>3 5 5 5 3 3 3 3 4 3 3 4</td>
<td>3 5 5 5 3 3 3 3 4 3 3 4</td>
<td>3 4 5 3 4 4 4 4 5 3 4 3</td>
<td>3 5 5 5 3 3 3 3 4 3 4 3</td>
<td>3 5 5 5 3 3 3 3 4 3 4 3</td>
</tr>
<tr>
<td>C</td>
<td>5 5 5 5 3 3 3 3 4 3 3 4</td>
<td>3 5 5 5 3 3 3 3 4 3 3 4</td>
<td>3 4 5 3 4 4 4 4 5 3 4 3</td>
<td>3 5 5 5 3 3 3 3 4 3 4 3</td>
<td>3 5 5 5 3 3 3 3 4 3 4 3</td>
</tr>
<tr>
<td>D</td>
<td>3 5 5 5 3 3 3 3 4 3 3 4</td>
<td>3 5 5 5 3 3 3 3 4 3 3 4</td>
<td>3 4 5 3 4 4 4 4 5 3 4 3</td>
<td>3 5 5 5 3 3 3 3 4 3 4 3</td>
<td>3 5 5 5 3 3 3 3 4 3 4 3</td>
</tr>
<tr>
<td>E</td>
<td>5 5 5 5 5 5 5 5 5 5 5 5</td>
<td>5 5 5 5 5 5 5 5 5 5 5 5</td>
<td>5 5 5 5 5 5 5 5 5 5 5 5</td>
<td>5 5 5 5 5 5 5 5 5 5 5 5</td>
<td>5 5 5 5 5 5 5 5 5 5 5 5</td>
</tr>
</tbody>
</table>

### Table 4
Agreements between the five observers

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>All (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>80</td>
</tr>
<tr>
<td>B</td>
<td>3</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>30</td>
</tr>
<tr>
<td>C</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>30</td>
</tr>
<tr>
<td>D</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>55</td>
</tr>
<tr>
<td>E</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>100</td>
</tr>
</tbody>
</table>

Number indicates how many observers agreed.
Diagnosis: A, disc position sagittal; B, disc position coronal; C, disc shape; D, effusion; E, loose calcified body.
The total agreement of all five observers is given in the column “all”.
TMD. Our internet page can be used in order to train and calibrate observers, as in the majority of cases more than three of five trained observers agreed upon the diagnosis. On the website it will be possible to reach our conclusions after the evaluations are made on a separate page.

The conclusion of this study was that the use of criteria was useful in order to standardize and simplify the evaluation and thereby probably increase the diagnostic outcome among different observers for MRI of the TMJ. We recommend that the criteria are used internationally to be able to compare different studies.

Acknowledgments
We want to thank Professor Per-Lennart Westesson and Senior Consultant Lars-Göran Hansson for acting as observers.

References