



Published in final edited form as:

*J Orofac Pain*. 2010 ; 24(1): 63–78.

## The Revised Research Diagnostic Criteria for Temporomandibular Disorders: Methods used to Establish and Validate Revised Axis I Diagnostic Algorithms

**Eric L. Schiffman, DDS, MS[Associate Professor],**

University of Minnesota School of Dentistry, Department of Diagnostic and Biological Sciences, 6-320 Moos Tower, 515 Delaware Street SE, Minneapolis, MN 55455, Telephone: 612-625-5146, Fax: 612-626-0138

**Richard Ohrbach, DDS, PhD[Associate Professor],**

University at Buffalo School of Dental Medicine, Department of Oral Diagnostic Sciences, 355 Squire Hall, Buffalo, NY 14214

**Edmond L. Truelove, DDS, MSD[Professor],**

University of Washington School of Dentistry, Department of Oral Medicine, Box 356370, Seattle, WA 98195

**Tai Feng, MS, PhD[Research Assistant],**

Division of Biostatistics, University of Minnesota, A460 Mayo Building, MMC 303, Minneapolis, MN 55455--0347

**Gary C. Anderson, DDS, MS[Associate Professor],**

University of Minnesota School of Dentistry, Department of Diagnostic and Biological Sciences, 6-320 Moos Tower 515 Delaware Street SE, Minneapolis, MN 55455

**Wei Pan, PhD[Professor],**

University of Minnesota, Department of Biostatistics, 420 Delaware Street SE, A428 Mayo, Minneapolis, MN 55455

**Yoly M. Gonzalez, DDS, MS[Assistant Professor],**

University at Buffalo School of Dental Medicine, Department of Oral Diagnostic Sciences, 3435 Main Street, 355 Squire Hall, Buffalo, NY 14214

**Mike T. John, DDS, MPH, PhD[Associate Professor],**

University of Minnesota School of Dentistry/School of Public Health, Department of Diagnostic and Biological Sciences, 6-320 Moos Tower, 515 Delaware Street SE, Minneapolis, MN 55455

**Earl Sommers, DDS, MSD[Clinical Associate Professor],**

University of Washington School of Dentistry, Department of Oral Medicine, Box 356370, Seattle, WA 98195

**Thomas List, DDS, Odont. Dr.[Professor],**

Department of Stomatognathic Physiology, Faculty of Odontology, Malmö University, SE 205 06 Malmö, Sweden

**Ana M. Velly, DDS, MS, PhD[Research Associate],**

University of Minnesota School of Dentistry, Department of Diagnostic and Biological Sciences, Room 6-219, 515 Delaware Street SE, Minneapolis, MN 55455

**Wenjun Kang, MS[Research Fellow], and**

University of Minnesota School of Dentistry, Department of Diagnostic and Biological Sciences, Room 7-546, 515 Delaware Street SE, Minneapolis, MN 55455

**John O. Look, DDS, PhD, MPH[Senior Research Associate]**

University of Minnesota School of Dentistry, Department of Diagnostic and Biological Sciences,  
6-320 Moos Tower, 515 Delaware Street SE, Minneapolis, MN 55455

Eric L. Schiffman: schif001@umn.edu

## Abstract

**AIMS**—To derive reliable and valid revised Research Diagnostic Criteria for Temporomandibular Disorders (RDC/TMD) Axis I diagnostic algorithms for clinical TMD diagnoses.

**METHODS**—The multi-site RDC/TMD Validation Project’s dataset (614 TMD community and clinic cases, and 91 controls) was used to derive revised algorithms for Axis I TMD diagnoses. Validity of diagnostic algorithms was assessed relative to reference standards, the latter based on consensus diagnoses rendered by 2 TMD experts using criterion examination data, including temporomandibular joint imaging. Cut-offs for target validity were sensitivity  $\geq 0.70$  and specificity  $\geq 0.95$ . Reliability of revised algorithms was assessed in 27 study participants.

**RESULTS**—Revised algorithm sensitivity and specificity exceeded the target levels for myofascial pain (0.82, 0.99, respectively) and myofascial pain with limited opening (0.93, 0.97). Combining diagnoses for any myofascial pain showed sensitivity of 0.91 and specificity of 1.00. For joint pain, target sensitivity and specificity were observed (0.92, 0.96) when arthralgia and osteoarthritis were combined as “any joint pain.” Disc displacement without reduction with limited opening demonstrated target sensitivity and specificity (0.80, 0.97). For the other Group II disc displacements and Group III osteoarthritis and osteoarthrosis, sensitivity was below target (0.35 to 0.53), and specificity ranged from 0.80 to meeting target. Kappa for revised algorithm diagnostic reliability was  $\geq 0.63$ .

**CONCLUSION**—Revised RDC/TMD Axis I TMD diagnostic algorithms are recommended for myofascial pain and joint pain as reliable and valid. However, revised clinical criteria alone, without recourse to imaging, are inadequate for valid diagnosis of two of the three disc displacements and osteoarthritis/osteoarthrosis.

## Keywords

reference standard; validity; reliability; research diagnostic criteria; temporomandibular disorders; temporomandibular muscle; joint disorders

## Introduction

The Research Diagnostic Criteria for Temporomandibular Disorders (RDC/TMD) are a widely employed diagnostic protocol for TMD research.<sup>(1)</sup> This taxonomic system includes an Axis I physical assessment and diagnostic protocol and an Axis II assessment of psychological status and pain-related disability.

In the third paper in this series, we reported on the validity of the original RDC/TMD Axis I protocol.<sup>(2)</sup> None of the 8 Axis I diagnostic algorithms met the target sensitivity of  $\geq 0.70$  and specificity of  $\geq 0.95$ . Only myofascial pain with limited opening (Ib) demonstrated target sensitivity (0.79) with specificity close to target (0.92). The criterion validity for the other 7 Axis I diagnostic algorithms showed poor to marginal sensitivity with specificity in the range of 0.86 to 0.99. The only diagnostic category that exhibited adequate levels for both sensitivity and specificity occurred when myofascial pain (Ia) and myofascial pain with limited opening (Ib) were combined into “any myofascial pain.” The Validation Project’s results strongly suggest a need for revising the 8 Axis I RDC/TMD diagnostic algorithms or combining some of the diagnoses.

This paper reports on the model-building methods employed to revise the RDC/TMD diagnostic algorithms, and the procedures for testing their criterion validity against reference-standard diagnoses. The reliability and validity of the revised algorithms are reported, and are compared with the reliability and validity of the original RDC/TMD algorithms. Finally, we describe the changes in the revised diagnostic algorithms and the questions and clinical tests that they incorporate.

## Materials and Methods

### Overview

Using the RDC/TMD Validation Project dataset<sup>(3)</sup>, revised Axis I diagnostic algorithms were derived, tested and validated. Items included in the revised algorithms were selected from among both the original RDC/TMD clinical tests and from the additional clinical tests that were performed as part of the criterion data collection specified for the RDC/TMD Validation Project. All clinical tests employed for the algorithm model building are listed in Table 2 of the first manuscript of this series.<sup>(3)</sup> The validity of the revised algorithms, relative to reference-standard diagnoses, was then compared with that of the original RDC/TMD algorithms in a randomly selected half of the data that had been reserved for the validation purposes. Using only the revised algorithm models that are recommended in this paper for use by TMD clinicians and investigators, inter-examiner reliability assessment was performed using 27 study participants.

### RDC/TMD Validation Study Participants

Seven hundred twenty-four participants completed the data collections required for the RDC/TMD Validation Project study. Five of these participants were unclassifiable using the inclusion and exclusion criteria described in the first paper in this series,<sup>(3)</sup> and were excluded from further analyses. Fourteen additional cases were excluded from the Axis I analyses due to the concurrent presence of chondromatosis (n = 2), reported fibromyalgia (n = 9), or reported rheumatoid arthritis (n = 3). Persons with fibromyalgia and rheumatoid arthritis were included in the study if there was a documented medical diagnosis. Chondromatosis was excluded based on suspicion of the presence of the disorder as detected on MRI by the radiologist. Therefore, the final sample size utilized for Axis I analyses was 705 participants: 614 TMD cases and 91 controls. Please refer to the first paper of this series for additional details on study settings and location, recruitment methods, informed consent process, participant reimbursement and IRB oversight.<sup>(3)</sup>

### Revised RDC/TMD Axis I Diagnostic Algorithm Model-Building Methods

From the dataset collected for the formal validation study, 2 subsets of data were used for revision of the Axis I algorithms: (1) algorithm model-building dataset, and (2) reference-standard diagnoses dataset. Methods for both of these data collections are briefly described here; complete descriptions are available in the first and third paper in this series.<sup>(2,3)</sup>

**Criterion examinations**—Two criterion examiners (CE-1 and CE-2), blinded to each other's findings, evaluated each study participant using the standardized criterion examination that was composed of the original RDC/TMD clinical tests as well as additional clinical tests as previously described.<sup>(1,3)</sup> These additional clinical tests included both questionnaire responses and clinical measurements. The CE-2 data were collected for each participant on the same day that the reference-standard diagnosis was established. The algorithm model-building dataset included all of the criterion examination diagnostic tests performed by the CE-2, the participants' responses to the published RDC/TMD Questionnaire,<sup>(1)</sup> questions from the Supplemental History Questionnaires that were

developed for this project, and the occlusal test data that were collected uniquely by the CE-1.

**Reference-standard diagnoses**—For each study site, the consensus diagnoses for the reference-standard dataset were established by the 2 CEs with the participant present.<sup>(3)</sup> All data collections were considered and verified as necessary. The 2 CEs re-examined the participant together if they encountered any diagnostic disagreement during their consensus deliberations. Their final diagnostic decisions were supported by radiologist-interpreted panoramic radiographs and bilateral temporomandibular joint (TMJ) magnetic resonance imaging (MRI) and computed tomography (CT).<sup>(2,3)</sup> The radiologist was consulted as needed in the determination of the reference-standard diagnoses. Details of the image analysis criteria used by the radiologists to identify MRI-disclosed disc displacements and CT-disclosed osteoarthritis, as well as assessment of their reliability using these criteria, are described elsewhere.<sup>(4)</sup>

**Statistical procedures for revised algorithm model building**—Diagnostic algorithm model building required division of the criterion examination dataset described above into 2 parts with 352 participants in the first part and 353 in the second. Part I was the model-building dataset used to derive the revised algorithms, and Part II was the validation dataset that was reserved for validity testing of the derived models. This procedure avoided the problem of circularity associated with deriving and validating an algorithm in the same dataset. Two methods were used for the revised diagnostic algorithm derivation that can be described as *data-driven* and *expert-driven*. The Part I dataset was used for model building by both methods.

**Data-driven method**—The *rpart* package was used to identify the best combination of variables in order for the data-driven algorithms to have optimum sensitivity and specificity relative to the reference standards.<sup>(5)</sup> The advantage of this package is that it outputs its resultant diagnostic algorithms so that the operator can assess whether the item selection makes clinical sense. This methodology uses procedures described by Breiman, Friedman, Olshen, and Stone<sup>(6)</sup> and Hastie T, Tibshirani R, Friedman J.<sup>(7)</sup> A 10-fold cross-validation strategy was specified for the model-building step. For each of 10 repetitions of this exercise, the Part I dataset was subdivided into 2 parts: 90% was randomly assigned to a *training dataset* and 10% was randomly assigned to a *testing dataset*. The *rpart* method was specified as class. This “class” designation in the formula is appropriate for differentiating the mutually exclusive diagnoses within a single group. For example, the Group II dependent variable for this formula was the reference-standard diagnoses that included IIa, IIb, IIc, and no Group II. The independent variables for this equation were selected as those examination items that best predicted these target diagnoses when arranged in the most mathematically relevant order. This process was repeated 9 times in other randomly selected sets of training and testing data within the Part I dataset. Based on these results, the “best-fit” model was realized for prediction of the reference-standard diagnoses. With the *rpart* method, the operator may choose various complex formulas. However, we employed the simplest *rpart* formula with no weights or other parameters added to the model.

The main usefulness of these data-based studies was to assist in diagnostic item selection, and to provide estimates of sensitivity and specificity associated with this statistical method. However, it was anticipated that the algorithmic outputs might be complex, and that they might include diagnostic items that were mathematically associated with the target diagnoses, but were not consistent with current TMD constructs.

**Expert-driven method**—The Part I dataset was used for empirical trials that substituted new clinical tests for the original tests of the RDC/TMD. In all, 175 clinical tests were

considered, all being tests that were part of the criterion examination. Some were measures related to diagnostic criteria from the American Academy of Orofacial Pain and the American College of Rheumatology.<sup>(8,9)</sup> Other tests included pressure pain threshold algometry, static and dynamic tests, soft and hard end-feel, bite test with unilateral and bilateral placement of cotton rolls, 1-minute clench test, and orthopedic joint play tests including compression, traction, and translation.<sup>(10–17)</sup> The observed improvements in sensitivity and specificity with respect to the reference-standard diagnoses were the basis for subsequent additions and deletions to the clinical tests defining the diagnostic algorithms. The primary goal of the expert-driven method was to preserve the original RDC/TMD classification tree structures as much as possible while redefining the “split condition” nodes with new tests for measuring the same constructs. Additional goals included the development of parsimonious diagnostic trees, and using, as much as possible, parallel tree structures to define the muscle pain (Ia and Ib) and joint pain (IIIa and IIIb) constructs. The revised split-condition nodes are described below.

### Validity Assessment of Revised Diagnostic Algorithms

The primary purpose of this analysis was to test criterion validity of the best performing revisions for the 8 Axis I RDC/TMD diagnostic algorithms. In addition, 4 grouped diagnoses were evaluated: any Group I myofascial pain diagnosis (Ia or Ib); any Group II disc displacement diagnosis (IIa, IIb, or IIc); any joint pain diagnosis (IIIa or IIIb); and any arthrosis (IIIb or IIIc).

**Cut-offs for determination of validity**—Diagnostic validity was estimated by computing sensitivity and specificity of the revised diagnostic models relative to the reference-standard diagnoses. In the development process for the RDC/TMD, the desired goal for sensitivity was set at  $\geq 0.70$  and specificity at  $\geq 0.95$ .<sup>(1)</sup> Assuming a prevalence of 10% for TMD, the authors showed that a positive predictive value of at least 0.75 for this diagnostic protocol would require specificity  $> 0.95$  while sensitivity could be as low as 0.70. Based on these expectations, we intended to declare as valid a given diagnosis if its estimated sensitivity was at least 0.70 and its estimated specificity was at least 0.95.

**Statistical procedures for validity testing**—Validity testing was performed in Part II ( $n = 353$ ) of the randomly divided dataset. Because Group II and III diagnoses are side-specific, the problem of multiple diagnoses within participants involving correlated data (right versus left sides) required the use of generalized estimate equation (GEE) methodology to generate adjusted confidence intervals around point estimates for sensitivity and specificity. The study participants included for specificity estimates were all the controls plus all the TMD cases that did not have the specific TMD diagnosis being assessed.

### Reliability Assessment of Revised Diagnostic Algorithms

Reliability testing was performed only for the expert-driven revised algorithms for reasons discussed below. The reliability assessment was performed in calibration exercises conducted at 2 study sites with a total of 27 participants: University of Minnesota ( $n = 18$ ) and University of Washington ( $n = 9$ ). The study design specified that each participant would be examined by one criterion examiner as well as the study site test examiner (a trained and calibrated dental hygienist). The 2 examiners alternated the order of the first and second examinations, and each was blinded to the exam findings of the other. These methods were similar to those used for reliability testing described in the second paper in this series,<sup>(18)</sup> with the exception that the revised examination protocol was used here.

**Statistical procedures for reliability assessment**—Interexaminer reliability for diagnoses and for the individual clinical tests included in the revised exam protocol was



estimated using generalized estimate equation (GEE) methodology to generate kappa estimates.<sup>(19)</sup> Data from both the right and left temporomandibular joints were used to yield the GEE kappa estimates for Group II and III reliability. In addition, the GEE estimates of variance used in the 95% confidence intervals were adjusted for correlated side-to-side data within participants. Diagnostic percent agreement was computed from contingency tables generated by SAS Version 9.1. The reliability estimates were based on 27 participants for Group I diagnoses and 54 joints for the Groups II and III diagnoses.

## Results

### Validity of the Revised Algorithms: Data-driven and Expert-driven

Table 1 shows the sensitivity and specificity, relative to the reference-standard diagnoses, for the 3 types of diagnostic algorithms (the original RDC/TMD algorithms, the data-driven revised algorithms, and the expert-driven revised algorithms). Since RDC/TMD examination items were included as a part of the CE-2 examination protocol, we were able to use the CE-2 data for the estimates of sensitivity and specificity shown for the RDC/TMD algorithms in Table 1. This allowed for an appropriate comparison of the RDC/TMD algorithms to the expert-driven and data-driven algorithms since the latter algorithms were derived using the CE-2 exam findings. We wish to point out that RDC/TMD algorithmic sensitivity and specificity estimates that are based on CE-2 data in Table 1, column 6, of the third paper of this series<sup>(2)</sup> vary slightly from the parallel results shown in Table 1 of this paper. The reason is that the former results are based on 705 participants whereas the results in this paper are computed using just the 353 participants reserved for the validation of these algorithms. To make strictly comparable estimates, we tested the original RDC/TMD algorithms using the same partial dataset as was used for validation of the data-driven and expert-driven diagnostic algorithms.

### Superior Performance of the Expert-driven Algorithm for Pain Diagnoses

The revised diagnostic algorithms for Group I Muscle Disorders, derived by both the data-driven and expert-driven methods, showed target validity for both sensitivity ( $\geq 0.70$ ) and specificity ( $\geq 0.95$ ) for the individual diagnoses of myofascial pain (Ia) and myofascial pain with limited opening (Ib). When combined into one diagnosis, “any myofascial pain” (Ia or Ib), target validity was obtained only with the expert-driven algorithm. In addition, only the expert-driven algorithm showed target validity for the measurement of joint pain, and then only when arthralgia (IIIa) and osteoarthritis (IIIb) were combined as “any joint pain” (IIIa or IIIb). In summary, the expert-driven revised algorithms alone met target validity for all categories of myofascial pain and for “any joint pain.”

For Groups II and III, only a diagnosis of disc displacement without reduction with limited opening (IIb) demonstrated target sensitivity and specificity for both the expert-driven and the data-driven algorithms. About one-half of the CT-detected cases of osteoarthritis or osteoarthrosis (see the “any arthrosis” estimate in Table 1) failed to be detected by either the expert-driven model or the data-driven model.

Confidence intervals (CI) for sensitivity and specificity of the expert-driven algorithm are presented in Table 1. Despite use of only a partial dataset for this validation testing, the total width of all but two CIs was  $\leq 0.19$ , thus falling within the limit of 0.20 (on a 0.0 to 1.0 scale) that was specified as the desired precision for the validity estimates for the RDC/TMD Validation project.<sup>(2)</sup> The two exceptions were the total CI width of 0.27 for IIb sensitivity, and 0.26 for IIIc sensitivity (Table 1). No CIs are presented for the data-driven model. Due to its incorporation of clinical data that are not consistent with TMD constructs, it could not be proposed as a diagnostic scheme for other investigators. The utility and

problems associated with data-driven algorithms are further explained in the “Model Choice” paragraph of the Discussion section.

### Reliability of the Expert-driven Revised Algorithms

Reliability estimates for the expert-driven algorithms are based on intrasite calibration exercise data from 2 study sites. At each site, independent diagnostic findings of one criterion examiner were compared with those of the test examiner. Combining the data between the 2 sites ( $n = 27$ ), kappas for interexaminer agreement and their percent agreement for each diagnosis are shown in Table 2. The kappa estimates from the original RDC/TMD algorithms were abstracted from “Table 2: Intra-site Reliability” of the second paper of this series<sup>(18)</sup> for comparison to the kappas of the revised algorithms. We note that point estimates for the reliability of the revised algorithms for all diagnoses exceeded the paired estimates for the original RDC/TMD algorithms by differences in kappa ranging from 0.01 to 0.59. Of special note, all kappas for the revised algorithm were  $\geq 0.63$ , and mean percent agreement for the 8 RDC/TMD diagnoses was 92%.

The estimates of reliability of questionnaire and examination items incorporated in the revised protocol are shown in Table 3. The reliability for the detection of muscle pain and detection of disc click is good (0.73 and 0.70, respectively). The reliability for diagnosing joint pain and coarse crepitus is excellent (0.78 and 0.85, respectively).

### The Expert-driven Revised Diagnostic Algorithms

Some of the clinical tests that were selected for the expert-driven algorithms were obtained from item selection by the data-driven methodology. Certain items in the expert-driven algorithms had been previously incorporated in the original RDC/TMD algorithms. Others were newly specified tests that were part of the criterion examination protocol. All model building of the expert-driven algorithms was done in concert with expert opinion from the Validation Project team and from experts outside this project as well as from review of the current literature.<sup>(3)</sup> The basic classification tree designs employed for the original RDC/TMD algorithms were retained for the revised models (Figures 1–3). These classification trees are composed of nodes defined by a “split condition” that can consist of a single variable or several variables. Examination data from an individual either satisfy or fail to satisfy each split condition, leading to a terminal node with its associated diagnosis. The main difference between the original RDC/TMD algorithms and the revised algorithms is that some of the split-condition nodes are defined by new clinical tests, although the condition being evaluated is the same. These differences are described by diagnostic grouping below.

**Revised Group I Muscle Disorders algorithm**—The revised Group 1 myofascial pain algorithm that was derived by the expert-driven method requires a report of pain, the location of which is verified by the examiner to be located in a masticatory structure, plus the report of familiar muscle pain associated with at least one of several clinical tests (Figure 1). Just as in the original RDC/TMD algorithm for myofascial pain, the first node includes the question, “Have you had pain in the face, jaw, temple, in front of the ear, or in the ear in the past month?” Unlike the original algorithm, we now propose that the examiner should confirm that the location of the pain endorsement is in a masticatory structure. A second difference is that the original RDC/TMD’s criterion for a positive finding of muscle pain required  $\geq 3$  of 20 muscle palpation sites to be painful, with one of them on the side of the pain complaint. The revised algorithm requires a minimum of 1 site among a total of 12 sites to be painful to palpation, and these test sites are located in only the masseter or temporalis muscles. A third difference is that the original algorithm did not include pain in these muscles associated with movement. The revised algorithm now includes, as an alternate

method for detecting muscle pain, a report of pain in masseter or temporalis muscles associated with maximum unassisted opening or maximum assisted opening. Finally, the revised algorithm requires that any pain provoked with palpation or opening must be familiar pain.

The sensitivity of the revised algorithm was not improved by the inclusion of the other muscle test sites specified by the original RDC/TMD protocol, nor was performance improved by inclusion of familiar pain with lateral or protrusive movements. These tests provided no additional information to that which was collected by using the tests that were retained. The second node distinguishes Ia (myofascial pain) from Ib (myofascial pain with limited opening). The original RDC/TMD algorithm had a passive stretch measurement requirement. In contrast, the only requirement for rendering a diagnosis of Ia by the revised RDC/TMD is that unassisted opening without pain would be  $\geq 40$  mm. The threshold of 40 mm employed in the original RDC/TMD algorithm was not changed, as there were no compelling data to invalidate it.

**Revised Group II Disc Displacements algorithm**—Group II Disc Displacements diagnoses are typically detected by TMJ clicking noise and limited opening assessed from history or as a current clinical finding. In the revised Group II algorithm derived from the expert-driven method, the first node defines the presence or absence of clicking noise (Figure 2). This node varies from the original RDC/TMD in that it uses a less rigorous criterion for assigning a positive finding for clicking noises, notably, that a click is diagnostic if it occurs during just 1 of 3 repetitions of the specified jaw movements instead of the original requirement for detection of a click during a minimum of 2 of 3 repetitions. The same jaw movements specified for the original RDC/TMD protocol (jaw opening, closing, lateral excursion and protrusion) are retained for the revised protocol. Unlike the original RDC/TMD algorithm, there is no requirement for reciprocal clicking to be eliminated with protrusive positioning of the jaw, or that the distance between the opening and closing clicking be  $\geq 5$  mm since the sensitivity of the revised algorithm was not improved by the inclusion of these tests. The first node differentiates disc displacement with reduction (IIa) from disc displacement without reduction with limited opening (IIb) and disc displacement without reduction without limited opening (IIc) based on the presence or absence of clicking noise. A diagnosis of IIa is defined by the presence of clicking noise as described in the original RDC/TMD as a distinct sound, of brief and very limited duration, with a clear beginning and end, which usually sounds like a “click”.<sup>(1)</sup> For this study we also defined a click as “snap” or “pop”. The absence of joint clicking noises or presence of joint clicking sounds not meeting criteria for IIa are indicative of IIb, IIc, or no group II diagnosis.

The second node differentiates “no group II diagnosis” from IIb and IIc. A determination of “no group II” diagnosis results when the participant has a negative history of jaw locking with limited opening that was severe enough to interfere with eating. This information is based on questions 14a and 14b from the original RDC/TMD Questionnaire.<sup>(1)</sup>

Finally, the third node is based on whether the combined measurement of maximum assisted opening plus vertical incisal overlap is  $< 40$  mm (IIb) or  $\geq 40$  mm (IIc). Maximum unassisted opening is no longer required. The revised third node also differs from the original RDC/TMD for IIb and IIc in that no additional criteria regarding passive stretch and contralateral movement are needed since sensitivity was not improved by including them. Finally, for IIb, there is no requirement for uncorrected deviation to the ipsilateral side on opening as in the original RDC/TMD since sensitivity was not improved by including it.



**Revised Group III Arthralgia, Arthritis, and Arthrosis algorithm**—Group III diagnoses include a combination of joint pain diagnoses of arthralgia (IIIa) and osteoarthritis (IIIb) as well as diagnoses related to osseous joint changes of osteoarthritis or osteoarthrosis (IIIc). In the revised Group III algorithm derived from the expert-driven method, the first node consists of structure that is parallel to the Group I pain diagnoses with a history of pain plus a report of familiar joint pain provoked with palpation or jaw movements (Figure 3). Unlike the original RDC/TMD algorithm for arthralgia, it now includes the same history of pain used for myofascial pain with examiner confirmation that the reported pain involves a masticatory structure. Like the original RDC/TMD algorithm, the revised algorithm includes the report of pain with palpation of the lateral pole with 1 pound of pressure. Palpation of the posterior attachment is replaced with the modified joint palpation test of palpating around the lateral pole with a target of at least 2 pounds of pressure (range 2 to 3 pounds).<sup>(3)</sup> Also retained is a report of pain with maximum unassisted opening, assisted opening, or with lateral excursive movements. However, unlike the original algorithm, the provoked pain must be familiar and joint-related for diagnoses of IIIa (arthralgia) or IIIb (osteoarthritis). The sensitivity of the revised algorithm was not improved by the inclusion of the posterior attachment as specified by the original RDC/TMD protocol, nor was performance improved by inclusion of familiar pain with protrusive movement.

In the second node, the revisions for IIIb (osteoarthritis) and IIIc (osteoarthrosis) now include the addition of a report by the participant of crepitus during any examination jaw movement. Participant report of crunching, grinding or grating sounds was recorded as crepitus. This finding serves as an alternate to the original RDC/TMD criterion that was limited only to coarse crepitus detected by the examiner. Although the original RDC/TMD method for the detection of coarse crepitus by palpation is retained for the revised algorithm, the revised method also stipulates that the noise must be audible to the examiner 6 inches from the joint.

**Observations regarding muscle and joint palpation pressure**—It is recommended above that palpation of the masseter and temporalis muscles as well as the alternative palpation around the lateral pole of the joint employ at least 2 pounds of pressure. Our data analyses demonstrate that diagnostic validity was not necessarily improved by using the myofascial palpation test (allowing 2 to 4 pounds of pressure) with sensitivity for a myofascial pain diagnosis increased by 0.02 and specificity decreased by 0.03.<sup>(3)</sup> But we can conclude that palpation pressures greater than 2 pounds are not associated with significantly more false positive diagnoses, given the familiar pain requirement for a positive diagnosis.

The modified joint palpation test using a range of 2 to 3 pounds of pressure with a target of at least 2 pounds target around the lateral pole accounted alone for a 10-point (0.10) increase in diagnostic sensitivity for arthralgia (IIIa).<sup>(3)</sup> In combination with the other revised examination tests for arthralgia, this test helped increase the overall sensitivity for any joint pain (IIIa or IIIb in Table 1) from 0.42 for the original RDC/TMD to 0.92 for the expert-driven revised algorithm. Finally, reliability for joint palpation as in the original RDC/TMD using 1 pound applied to the lateral pole is equivalent to the reliability of using a target of at least 2 pounds applied around the pole. Their respective estimates for reliability were  $k = 0.86$  and  $k = 0.89$  (Table 3).

**Association between TMD pain and jaw movements**—One of the questions that we tested for its statistical influence was: “In the last month, did any of the following activities change your pain?” The response options were related to jaw movement, function, parafunction, and/or rest. This question reflected the principle that TMD pain can be influenced by these jaw activity states. Ninety-nine percent (495/500) of the participants with a TMD pain diagnosis reported that their pain was modified by movement, function,

parafunction and/or rest. However, 8.6% of these participants indicated that at least one aspect of their pain was improved specifically by movement. Because each pain location (masseter, TMJ, temporalis, ear) was assessed separately, a participant could report that one pain was aggravated by movement while another was improved by movement. The inclusion of this clinical information in our analyses had no effect on the validity of the revised diagnostic algorithms for “any joint pain.” For myofascial pain (Ia or Ib), sensitivity increased by 0.06 and specificity decreased by 0.02.

**Problematic results observed with data-driven algorithms**—Item selection for the data-driven models was observed at times to be clinically illogical. For example, range of motion cutoffs for a number of algorithms ranged from 37 to 51 mm. Other observations that were not consistent with TMD constructs were crepitus as a diagnostic predictor in the data-driven model for Group II, and muscle pain in the Group III model.

## Discussion

### Progress in TMD pain diagnoses

Based on the expert-driven model derivation method, the revised algorithms for Group I Muscle Disorders showed target validity of sensitivity  $\geq 0.70$  and specificity  $\geq 0.95$  for myofascial pain (Ia) and myofascial pain with limited opening (Ib), either as 2 discrete diagnostic disorders or when these 2 disorders were combined. For Group III (Arthralgia, Arthritis, and Arthrosis), the diagnosis or “any joint pain” (i.e., arthralgia) well exceeded the targets for sensitivity and specificity if the distinction was dropped between arthralgia (IIIa) alone versus arthralgia associated with osteoarthritis (i.e., osteoarthritis - IIIb). If IIIa is differentiated from IIIb, then the sensitivity for IIIa was barely adequate (0.74), and its specificity was below target (0.88). Our finding regarding the diagnosis of any joint pain (IIIa or IIIb) is significant. In the third paper in this series, we suggested that the original RDC/TMD examination protocol could be recommended as a simple clinical examination for the detection of muscle pain, but it was inadequate for joint pain.<sup>(2)</sup> This conclusion held true whether for diagnosis of IIIa alone or for the combination of any joint pain (IIIa or IIIb). These new validity data suggest that the revised protocol can now be recommended as a simple clinical test for *muscle and joint* pain in TMD patients. This recommendation is also supported by interexaminer reliability of  $k \geq 0.73$  observed for the diagnoses of myofascial pain (i.e., Ia, Ib or in combination), arthralgia and any joint pain using the revised exam protocol (Table 2).

### Diagnosis of Disc Displacements, Osteoarthritis, and Osteoarthritis

The clinical diagnosis of IIa is dependent on the presence of a diagnostic click. Subject report of clicking noises was assessed with questionnaires as well as during the exam. We also assessed clicking noises during range of motion of the jaw with palpation, auscultation with a stethoscope and during superior loading of the joint. The inadequate sensitivity of the algorithm for disc displacement with reduction likely resulted from the findings that clicking noises are often not present, or are present only with a single movement. The diagnosis of IIa had a sensitivity of 0.46 suggesting that the presence of a diagnostic clicking noise during  $\geq 1$  of 3 repetitions of the specified jaw movements was associated with less than 50% of IIa diagnoses. This result is consistent with a prior report of sensitivity of 0.51 and specificity of 0.83 for the diagnostic accuracy of clicking noises to predict IIa when using MRI as the reference standard.<sup>(20)</sup> The diagnostic importance of clicking is further undermined by the fact that clicking may occur with normal disc position, disc displacement with reduction, or disc displacement without reduction.<sup>(20–29)</sup> Disc displacement without reduction without limited opening often presents with inconsistent or the lack of signs or

symptoms thus making difficult a valid clinical diagnosis.<sup>(20,21,25,30)</sup> Therefore, rendering a valid joint-specific diagnosis of IIa and IIc requires TMJ MRI.

The expert-driven revised algorithm for disc displacement without reduction with limited opening (IIb) had target sensitivity (0.80) and specificity (0.97) using a history of jaw locking, absence of clicking noises or presence of noise not meeting criteria for IIa and limited opening with assisted opening (passive stretch). A prior study, using only maximum unassisted opening, reported disc displacement without reduction to have diagnostic sensitivity of 0.69 and specificity of 0.81.<sup>(20)</sup> A comparison of these two reports suggests that the use of the passive stretch measurement may improve diagnostic accuracy for IIb. One inherent difficulty with this diagnosis is that the TMJs are not independent, and therefore limitation in opening is jaw-specific as well as joint-specific. Given that the jaw is biarticulate, limited condylar translation on one side may be associated with an inaccurate clinical joint diagnosis on the contralateral side.<sup>(20,21,30)</sup>

About one-half of the diagnoses for radiographically-detected frank osseous joint changes, either osteoarthritis (IIIb) or osteoarthrosis (IIIc), were missed by the expert-driven algorithm. In contrast, Table 3 demonstrates compellingly that coarse crepitus can be detected much more reliably ( $k = 0.85$ ) using this revised protocol than was reported for the original RDC/TMD protocol ( $k = 0.53$ ) for coarse crepitus in Table 4 of the second paper of the series.<sup>(18)</sup> Regardless, the problem remains that the clinical finding of crepitus has been shown to be a poor indicator of frank osseous changes revealed by TMJ MRI.<sup>(29,31)</sup> Given that TMJ MRI can miss osseous changes that may be identified by CT<sup>(4,32)</sup>, a definitive ruling in or out for a diagnosis of osteoarthritis (IIIb)/osteoarthrosis (IIIc) should be accomplished with CT imaging.

To sum up this section on Groups II and III, it is reasonable to ask whether a clinical exam based on signs and symptoms for detection of intra-articular disorders of the TMJ is an attainable goal.<sup>(33-37)</sup> This is despite the fact that we used an extensive number of clinical tests including the 16 tests that have been reported to be the most important for differentiating intra-articular disorders.<sup>(38)</sup> Clearly, the original RDC/TMD suggested supplementation of the exam findings with imaging, and that recommendation cannot yet be changed. The one possible exception to this rule is for the diagnosis of IIb.

### Improved Reliability of the Revised Diagnostic Algorithms

In comparison to the original RDC/TMD algorithms, Table 2 shows improved reliability to be associated with the revised diagnostic algorithms. This improvement may be explained by the following considerations: 1) Examination item reliability tends to be higher for the revised diagnostic procedures (Table 3); 2) The revised algorithms include fewer items on which multiple examiners need to agree for their findings to be consistent; 3) Inter-examiner agreement is facilitated by the fact that the first algorithmic node for each of the Groups I-III can be satisfied by any of multiple items. Thus, multiple examiners for a given subject may not detect the same sign, but their finding of an alternate sign allows their agreement on the condition being assessed; and 4) The influence of chance in the improved estimates of reliability must also be considered. The preliminary results in this report should be replicated in other study populations.

### Diagnostic Model Choices and Model Building

Classification tree models are preferred for diagnostic decisions. Both clinicians and investigators prefer a diagnostic structure that is interpretable and intuitively consistent with theoretical constructs that describe the health or disease conditions being studied. However, as noted above in Results, data-driven algorithms can be inconsistent with these constructs.

We emphasize for this study that our purpose for the data-driven methods was to help in selection of candidate items, and to establish likely maximum estimates of sensitivity and specificity that might be attainable when using the diagnostic tests included in our dataset. The focus of this study was to present revised RDC algorithms that were consistent with current constructs in TMD and that would be potentially usable for clinicians and investigators.

All methods for joint noise detection were assessed, including clinician-based tests (auscultation with stethoscope and loading of the joint), participant report of noise (from the history or during the exam), and all combinations of sounds with any jaw movement. No method for detection of joint clicking was associated as highly with Group II diagnoses as the less rigorous clicking criterion used in the expert-driven model. We also tested additional clinical provocation maneuvers for possible inclusion in the revised algorithms for joint or muscle pain.<sup>(10–17)</sup> However, these clinical tests were not selected as strong diagnostic predictors by the data-driven methodology. Since they are also more technique-sensitive than palpation and jaw movement assessment, their inclusion in the expert-driven models was not justified.

### Case Definitions for Musculoskeletal Disorders

Case definitions are a set of diagnostic criteria that establish the boundaries by which a disorder can be considered to be present or absent. The discussion that follows defines and examines the rationales for the “expert-driven” decisions to retain certain clinical tests from the original RDC/TMD, to incorporate some new clinical tests, and to not incorporate others.

**Diagnostic questions for pain diagnoses**—The expert-driven revised diagnostic algorithms for “any joint pain” (IIIa or IIIb) and Group I muscle disorders (Ia or Ib) now include the same global question concerning the presence of pain: “Have you had pain in the face, jaw, temple, in front of the ear, or in the ear in the past month?”<sup>(1)</sup> The original RDC/TMD algorithms employed this question only for Group I Muscle Disorders. As reported in Results, the association between pain in the masticatory structures and jaw movement, function, parafunction and rest was also evaluated. This issue reflects the principle that TMD pain is orthopedic in nature and that it can be influenced by these jaw activity states. Our results suggest that jaw activities do, in fact, modify jaw pain, but that such activities do not necessarily always aggravate the pain complaint. Finally, whether this characteristic is a necessary hallmark of TMD, as a musculoskeletal pain disorder, or is a broader characteristic of trigeminal-mediated pain disorders where convergence from multiple structures occurs in the trigeminal nucleus, could not be determined in the context of this study due to the composition of the study sample.

Our data analysis showed no incremental improvement in the validity of the diagnostic algorithms for “any joint pain” or for myofascial pain that was related to the participants’ reports as to whether their pain condition was changed by jaw movement, function, parafunction or rest. However, when these revised diagnostic criteria are further tested in populations with co-morbid conditions, especially those that are trigeminally-mediated, it may be possible to differentiate TMD from these conditions based on pain associated with jaw movement, function, parafunction or rest. Given our study exclusion criteria, we could not test this hypothesis in our study sample.

**Clinical tests for pain diagnoses**—We eliminated muscle palpation sites with low reliability and those difficult to examine. These included the posterior mandibular and submandibular areas, as well as the intraoral palpation of the lateral pterygoid muscle area

and the temporalis tendon. Apart from the inaccessibility of the lateral pterygoid and posterior digastric muscles, it has been recommended that these sites be dropped from the protocol due to low reliability.<sup>(39–43)</sup> The revised algorithm for myofascial pain (Ia or Ib) now specifies a total of just 12 sites for the bilateral masseter and temporalis muscles while maintaining excellent sensitivity, specificity, and reliability. For myofascial pain with limited opening (Ib), there is no longer a requirement that pain-free unassisted opening would be increased at least 5 mm with assisted opening. This clinical information did not improve sensitivity. Although an unassisted opening that is limited by pain may indicate a different treatment than one without this limitation, future research will need to determine the actual clinical utility of a Ib diagnosis.

An increase of muscle palpation pressure by using 2 to 4 pounds did not improve diagnostic validity for myofascial pain over that which can be detected with the RDC-specified 2-pounds muscle palpation pressure. However, we emphasize that the revised myofascial palpation test is not the same as the method specified by RDC/TMD. The revised palpation method uses motion as described in the first paper of the series<sup>(3)</sup> whereas the RDC/TMD method is static. Although increased palpation pressure was not associated with improved diagnosis of myofascial pain, the addition of the modified joint palpation test with a target of at least 2 pounds pressure around the lateral pole substantially improved the diagnostic validity of the arthralgia (IIIa) algorithm. A previous investigation using algometry suggested that 2.5 pounds of palpation pressure was optimal for diagnosis of TMJ arthralgia (sensitivity, 0.81; specificity, 0.97).<sup>(44)</sup> Consequently, we are recommending a range of 2 to 3 pounds of pressure with a target of at least 2 pounds of pressure for both the myofascial and the modified joint palpation tests. We suggest this 2 to 3 pound range because our collective experience confirmed that it is not always possible to apply an exact pressure, even with static palpation. The only way to apply a specific amount of pressure is to use an algometer. Therefore, giving the examiner a “minimum target” of palpation pressure to use for these structures simply addresses this clinical reality.

**Familiar pain for diagnosis of pain disorders**—Positive provocation tests were followed with a question regarding whether the provoked pain was “familiar,” that is, pain similar to or like that he/she had been experiencing from the target condition outside the examination setting. The intent was to reproduce the participant’s pain complaint, if one was present.<sup>(45,46)</sup> Inclusion of this question in the revised examination protocols significantly improved the sensitivity of the diagnostic algorithms for myofascial pain by 9 points (0.09), and “any joint pain” by > 0.40, with little effect on specificity (Table 1). This question embodies the principle that provoked duplication of the pain complaint suggests the anatomical source of the pain. It also eliminates diagnostic confounding due to false-positive pain endorsements in pain-free participants, because such pain would not be “familiar” to them. The concept of “familiar pain” likewise eliminated the requirement of the original RDC/TMD for pain to palpation to be present on the side of the participant’s pain complaint in the case of myofascial pain. Finally, the “familiar pain” concept addresses the inconsistency of the original RDC/TMD myofascial pain algorithm that resulted in a positive diagnosis based on pain to palpation in any muscle, as long as it was on the side of the pain complaint. The concept of “familiar pain” is well established in the pain literature, and has been used for identifying other musculoskeletal, cardiac, and visceral pain disorders.<sup>(47–54)</sup>

### Application of the revised RDC/TMD diagnostic algorithms

The rules for assigning diagnoses are unchanged from the original RDC/TMD<sup>(1)</sup> and are reiterated here. A *subject* can be assigned a maximum of one muscle (Group I) diagnosis (either myofascial pain or myofascial pain with limited opening). In addition each joint may



be assigned, at most, one diagnosis from each of Group II and Group III, resulting in a maximum possible of 5 diagnoses. Diagnoses within any given group are mutually exclusive. The reader is advised that before applying the revised algorithms, it is necessary to assess for and rule out other pathology, including the conditions that are listed in the exclusion criteria for the RDC/TMD Validation Project.<sup>(3)</sup> Guidelines for ruling out other orofacial pain disorders using a comprehensive history and examination protocol are available.<sup>(55,56)</sup> Further investigation is needed to determine whether other orofacial pain conditions can be differentiated from TMD pain by the fact that TMD pain can be aggravated by jaw function.<sup>(45,46)</sup>

### Limitations of this Report

The recommendation of the revised RDC/TMD algorithms for myofascial pain and “any joint pain” has essentially 3 limitations. First, the derivation of the revised algorithms was accomplished using half of the RDC/TMD Validation Project dataset, because the remaining half of the original dataset was reserved for the validity testing. Although the resulting 95% confidence intervals for the point estimates for sensitivity and specificity in Table 1 are, with two exceptions, associated with the desired precision that was specified for the validity testing of the original RDC/TMD protocol, the diagnostic validity of the revised algorithms should be confirmed by future assessment in other large populations.

Second, the currently reported estimates of reliability of the revised algorithms require additional assessment. Given the limited sample size ( $n = 27$ ), the results should be considered preliminary. For one notable example, the high reliability associated with the IIC diagnosis could be chance-related in a small sample size situation.

Third, generalizability of the results is constrained by the characteristics of the study sample. The sample includes controls and cases without significant co-morbid conditions to meet the recommendations of Standards for Reporting of Diagnostic Accuracy (STARD).<sup>(57)</sup> STARD advocates that preliminary studies of diagnostic accuracy answer the question, “Do the test results in patients with the target condition differ from the results in healthy people?” The RDC/TMD Validation Project addressed this question. Affirmation of the first question then leads to a second question, “Are patients with specific test results more likely to have the target disorder than similar patients with other test results?”<sup>(57)</sup> Investigation of the second question requires a study sample of patients with the target condition, as well as a sufficient number of those participants with possible co-morbid conditions such as odontalgia, fibromyalgia, rheumatoid arthritis, or active migraine headache. So few cases with co-morbid conditions were recruited for the RDC/TMD Validation Project ( $n = 14$ ) that a meaningful estimate of the statistical effect associated with co-morbid conditions was not possible, and they were excluded from the analyses.

### Revised Examination and History Data Collection

The revised history questionnaire, specifications for TMD examination and examination forms for the revised Axis I algorithms are located on the International RDC-TMD Consortium Network web site.<sup>(58)</sup>

### Conclusion

We recommend that the expert-driven algorithms described here be used for Axis I diagnoses of myofascial pain, either as 2 discrete diagnoses or as a combined diagnosis of “any myofascial pain” (Ia or Ib), and for temporomandibular joint pain when there is no need to differentiate between arthralgia (IIIa) and osteoarthritis (IIIb). For two of the three TMJ disc displacement diagnoses, and for osteoarthritis/osteoarthrosis, an accurate diagnosis

of these conditions requires that the revised Axis I diagnostic algorithms be supplemented with the use of MRI and CT imaging, respectively. The one exception to this rule would be for the diagnosis of disc displacement without reduction with limited opening (IIb) that was observed to have adequate sensitivity and specificity based only on the recommended clinical tests and questions, without recourse to imaging.

Given that the diagnosis and management of these TMD pain disorders (muscle and joint pain) is the primary focus of research and clinical care, the results of this study will be useful to both the researcher and clinician. The methods and findings of the entire RDC/TMD Validation Project will facilitate comparisons for outcomes from both observational and experimental projects. These future studies will allow for an improved taxonomic system based on signs and symptoms, and ultimately lead to a diagnostic system based on mechanism and etiology.

## Acknowledgments

### Acknowledgement of Validation Project Study Group

The RDC/TMD Validation Study Group is comprised of: University of Minnesota: Eric Schiffman (Study PI), Mansur Ahmad, Gary Anderson, Quentin Anderson, Mary Haugan, Amanda Jackson, Pat Lenton, John Look, Wei Pan, Feng Tai; University at Buffalo: Richard Ohrbach (Site PI), Leslie Garfinkel, Yoly Gonzalez, Patricia Jahn, Krishnan Kartha, Sharon Michalovic, Theresa Speers; and University of Washington: Edmond Truelove (Site PI), Lars Hollender, Kimberly Huggins, Lloyd Mancl, Julie Sage, Kathy Scott, Jeff Sherman, Earl Sommers. Research supported by NIH/NIDCR U01-DE013331.

## References

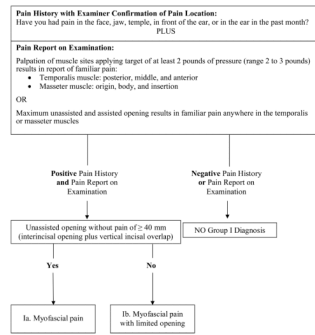
1. Dworkin SF, LeResche L. Research diagnostic criteria for temporomandibular disorders: Review criteria, examinations and specifications, critique. *J of Craniomandib Dis.* 1992; 6:301–355.
2. Truelove E, Pan W, Look J, Mancl L, Ohrbach R, Velly A, John MT, Schiffman EL. Research Diagnostic Criteria for Temporomandibular Disorders: Validity of Axis I Diagnoses. *J Orofac Pain.* 2009 in revision.
3. Schiffman EL, Truelove E, Ohrbach R, Anderson G, John MT, List T, et al. Assessment of the Validity of the Research Diagnostic Criteria for Temporomandibular Disorders: Overview and Methodology. *J Orofac Pain.* 2009 in revision.
4. Ahmad, M.; Hollender, L.; Anderson, Q.; Kartha, K.; Ohrbach, R.; Truelove, E.; Mike, JT.; Schiffman, E. Oral Surg Oral Med Oral Pathol Oral Radiol Endod. Research Diagnostic Criteria for Temporomandibular Disorders (RDC/TMD): Development of Image Analysis Criteria and Examiner Reliability for Image Analysis. Accepted for publication
5. <http://roadrunner.cancer.med.umich.edu/comp/docs/R.rpart.pdf>.
6. Breiman, L.; Freidman, JH.; Olshen, RA.; Stone, CJ. Classification and regression trees. Anonymous Monterey, CA: Wadsworth; 1984.
7. Hastie, T.; Tibshirani, R.; Friedman, JH. The Elements of Statistical Learning: Data Mining, Inference and Prediction. Anonymous New York, NY: Springer-Verlag; 2001.
8. Okeson, JP. Anonymous Carol Stream: Quintessence. 1996. Orofacial pain: Guidelines for assessment, diagnosis, and management.
9. Altman R, Asch E, Bloch D, Bole G, Borenstein D, Brandt K, Christy W, Cooke TD, Greenwald R, Hochberg M. Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Diagnostic and Therapeutic Criteria Committee of the American Rheumatism Association. *Arthritis Rheum.* 1986; 29:1039–1049. [PubMed: 3741515]
10. Steenks, MH.; deWijer, A.; Lobbezoo-Scholte, AM.; Bosman, F. Orthopedic Diagnostic Tests for Temporomandibular and Cervical Spine Disorders. In: Fricton, J.; Dubner, R., editors. *Advances in Pain Research and Therapy Orofacial Pain and Temporomandibular Disorders.* New York, New York: Raven Press; 1995.

11. Lobbezoo-Scholte AM, Steenks MH, Faber JA, Bosman F. Diagnostic value of orthopedic tests in patients with temporomandibular disorders. *J Dent Res.* 1993; 72:1443–1453. [PubMed: 8408888]
12. Lobbezoo-Scholte AM, de Wijer A, Steenks MH, Bosman F. Interexaminer reliability of six orthopaedic tests in diagnostic subgroups of craniomandibular disorders. *J Oral Rehabil.* 1994; 21:273–285. [PubMed: 8057195]
13. Visscher CM, Lobbezoo F, Naeije M. A reliability study of dynamic and static pain tests in temporomandibular disorder patients. *J Orofac Pain.* 2007; 21:39–45. [PubMed: 17312640]
14. Okeson, JP. History and examination for temporomandibular disorders. In: , editor. *Management of Temporomandibular Disorders and Occlusion.* St. Louis: Mosby Year Book; 2008.
15. Ohrbach R, Gale EN. Pressure pain thresholds, clinical assessment, and differential diagnosis: reliability and validity in patients with myogenic pain. *Pain.* 1989; 39:157–169. [PubMed: 2594394]
16. Howard, J. Clinical Diagnosis of Temporomandibular Joint Derangements. In: Moffett, BC., editor. *Diagnosis of Internal Derangements of the Temporomandibular Joint.* Seattle, Washington: Continuing Dental Education, University of Washington; 1984.
17. Wright, EF. *Manual of Temporomandibular Disorders.* Anonymous Ames, Iowa: Blackwell Munksgaard; 2005.
18. Look JO, John MT, Tai F, Huggins K, Lenton PA, Truelove E, Schiffman EL. Research Diagnostic Criteria for Temporomandibular Disorders: Reliability of Axis I Diagnoses and Selected Clinical Measures. *J Orofac Pain.* 2009 Accepted.
19. Williamson JM, Lipsitz SR, Manatunga AK. Modeling kappa for measuring dependent categorical agreement data. *Biostatistics.* 2000; 1:191–202. [PubMed: 12933519]
20. Orsini MG, Kuboki T, Terada S, Matsuka Y, Yatani H, Yamashita A. Clinical predictability of temporomandibular joint disc displacement. *J Dent Res.* 1999; 78:650–660. [PubMed: 10029463]
21. Schiffman E, Anderson G, Friction J, Burton K, Schellhas K. Diagnostic criteria for intraarticular T.M. disorders. *Community Dent Oral Epidemiol.* 1989; 17:252–257. [PubMed: 2791516]
22. Limchaichana N, Nilsson H, Ekberg EC, Nilner M, Petersson A. Clinical diagnoses and MRI findings in patients with TMD pain. *J Oral Rehabil.* 2007; 34:237–245. [PubMed: 17371560]
23. Eriksson L, Westesson PL, Rohlin M. Temporomandibular joint sounds in patients with disc displacements. *Int J Oral Surg.* 1985; 14:428–436. [PubMed: 3932236]
24. Rohlin M, Westesson PI, Eriksson L. The correlation of temporomandibular joint sounds with joint morphology in fifty-five autopsy specimens. *J Oral Maxillofac Surg.* 1985; 43:194–200. [PubMed: 3855989]
25. Manfredini D, Guarda-Nardini L. Agreement between Research Diagnostic Criteria for Temporomandibular Disorders and Magnetic Resonance Diagnoses of Temporomandibular disc displacement in a patient population. *Int J Oral Maxillofac Surg.* 2008; 37:612–616. [PubMed: 18486451]
26. Huddleston-Slater JJ, Van Selms MK, Lobbezoo F, Naeije M. The clinical assessment of TMJ sounds by means of auscultation, Palpation or both. *J Oral Rehabil.* 2002; 29:873–878.
27. Manfredini D, Basso D, Salmaso L, Guarda-Nardini L. Temporomandibular joint click sound and magnetic resonance-depicted disk position: which relationship? *J Dent.* 2008; 36:256–260. [PubMed: 18276055]
28. Marguelles-Bonnet RE, Carpentier P, Yung JP, Defrennes D, Pharaboz C. Clinical diagnosis compared with findings of magnetic resonance imaging in 242 patients with internal derangement of the TMJ. *J Orofac Pain.* 1995; 9:244–253. [PubMed: 8995924]
29. Schmitter M, Kress B, Rammelsberg P. Temporomandibular joint pathosis in patients with myofascial pain: a comparative analysis of magnetic resonance imaging and a clinical examination based on a specific set of criteria. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod.* 2004; 97:318–324. [PubMed: 15024353]
30. Anderson GC, Schiffman EL, Schellhas KP, Friction JR. Clinical vs. arthrographic diagnosis of TMJ internal derangement. *J Dent Res.* 1989; 68:826–829. [PubMed: 2715477]
31. Ohlmann B, Rammelsberg P, Henschel V, Kress B, Gabbert O, Schmitter M. Prediction of TMJ arthralgia according to clinical diagnosis and MRI findings. *Int J Prosthodont.* 2006; 19:333–338. [PubMed: 16900815]

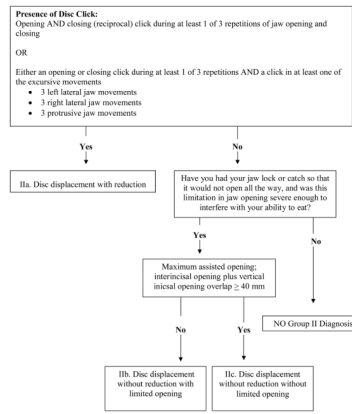
32. Westesson PL, Katzberg RW, Tallents RH, Sanchez-Woodworth RE, Svensson SA. CT and MR of the temporomandibular joint: comparison with autopsy specimens. *AJR Am J Roentgenol.* 1987; 148:1165–1171. [PubMed: 3495142]
33. Barclay P, Hollender LG, Maravilla KR, Truelove EL. Comparison of clinical and magnetic resonance imaging diagnosis in patients with disk displacement in the temporomandibular joint. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod.* 1999; 88:37–43. [PubMed: 10442943]
34. Emshoff R, Rudisch A. Validity of clinical diagnostic criteria for temporomandibular disorders: clinical versus magnetic resonance imaging diagnosis of temporomandibular joint internal derangement and osteoarthritis. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod.* 2001; 91:50–55. [PubMed: 11174571]
35. Yatani H, Sonoyama W, Kuboki T, Matsuka Y, Orsini MG, Yamashita A. The validity of clinical examination for diagnosing anterior disk displacement with reduction. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod.* 1998; 85:647–653. [PubMed: 9638696]
36. Yatani H, Suzuki K, Kuboki T, Matsuka Y, Maekawa K, Yamashita A. The validity of clinical examination for diagnosing anterior disk displacement without reduction. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod.* 1998; 85:654–660. [PubMed: 9638697]
37. Tognini F, Manfredini D, Montagnani G, Bosco M. Is clinical assessment valid for the diagnosis of temporomandibular joint disk displacement? *Minerva Stomatol.* 2004; 53:439–448. [PubMed: 15278022]
38. Schmitter M, Kress B, Leckel M, Henschel V, Ohlmann B, Rammelsberg P. Validity of temporomandibular disorder examination procedures for assessment of temporomandibular joint status. *Amer J of Ortho and Dentofac Orthopedics.* 2008; 133:796–803.
39. Johnson DB, Templeton MCC. The feasibility of palpating the lateral pterygoid muscle. *J Prosthet Dent.* 1980; 40:318–323.
40. Turp JC, Minagi S. Palpation of the lateral pterygoid region in TMD--where is the evidence? *J Dent.* 2001; 29:475–483. [PubMed: 11809325]
41. Sato H, Matusgama T, Ishikawa M, Ukon S, Zeze R, Kodama J. A study of the accuracy of posterior digastric muscle palpation. *J Orofac Pain.* 2000; 14:239.
42. Stratmann U, Mokrys K, Meyer U, et al. Clinical anatomy and palpability of the inferior lateral pterygoid muscle. *J Prosthet Dent.* 2000; 83:548–554. [PubMed: 10793387]
43. Turp JC, Arima T, Minagi S. Is the posterior belly of the digastric muscle palpable? A qualitative systematic review of the literature. *Clin Anat.* 2005; 18:318–322. [PubMed: 15971213]
44. Shaefer JR, Jackson DL, Schiffman EL, Anderson QN. Pressure-pain thresholds and MRI effusions in TMJ arthralgia. *J Dent Res.* 2001; 80:1935–1939. [PubMed: 11706955]
45. Steenks MH, deWijer A. Validity of the research diagnostic criteria for temporomandibular disorders axis I in clinical and research settings. *J Orofac Pain.* 2009; 23:9–16. [PubMed: 19264032]
46. Greene CS. Validity of the Research Diagnostic Criteria for Temporomandibular Disorders Axis I in Clinical and Research Settings. *J Orofac Pain.* 2009; 23:20–23.
47. Schwarzer AC, Derby R, Aprill CN, Fortin J, Kine G, Bogduk N. The value of the provocation response in lumbar zygapophyseal joint injections. *Clin J Pain.* 1994; 10:309–313. [PubMed: 7858361]
48. McFadden JW. The stress lumbar discogram. *Spine.* 1988; 13:931–933. [PubMed: 2973136]
49. Thevenet P, Gosselin A, Bourdonnec C, Gosselin M, Bretagne JF, Gastard J, et al. pHmetry and manometry of the esophagus in patients with pain of the angina type and a normal angiography. *Gastroenterol Clin Biol.* 1988; 12:111–117. [PubMed: 3366313]
50. Janssens J, Vantrappen G, Ghillebert G. 24-hour recording of esophageal pressure and pH in patients with noncardiac chest pain. *Gastroenterology.* 1986; 90:1978–1984. [PubMed: 3699414]
51. Vaksman G, Ducloux G, Caron C, Manouvrier J, Millaire A. The ergometrine test: effects on esophageal motility in patients with chest pain and normal coronary arteries. *Can J Cardiol.* 1987; 3:168–172. [PubMed: 3594299]
52. Davies HA, Kaye MD, Rhodes J, Dart AM, Henderson AH. Diagnosis of oesophageal spasm by ergometrine provocation. *Gut.* 1982; 23:89–97. [PubMed: 7068041]

53. Wise CM, Semble EL, Dalton CB. Musculoskeletal chest wall syndromes in patients with noncardiac chest pain: a study of 100 patients. 1992; 73:147–149.
54. Kokkonen SM, Kurunlahti M, Tervonen O, Ilkko E, Vanharanta H. Endplate degeneration observed on magnetic resonance imaging of the lumbar spine: correlation with pain provocation and disc changes observed on computed tomography diskography. *Spine*. 2002; 27:2274–2278. [PubMed: 12394906]
55. Goulet, J.; Palla, S. The path to diagnosis. In: Sessle, BJ.; Lavigne, GJ.; Lund, JP.; Dubner, R., editors. *Orofacial Pain: From Basic Science to Clinical Management*. Chicago: Quintessence; 2008.
56. de Leeuw, JR. *Reny De Leeuw and American Academy of Orofacial Pain. Orofacial pain: guidelines for assessment, diagnosis, and management*. Chicago: Quintessence; 2008.
57. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonsi CA, Glasziou PP, Irwig LM, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Ann Intern Med*. 2003; 138:W1–W12. [PubMed: 12513067]
58. International Consortium for RDC-TMD-based Research.

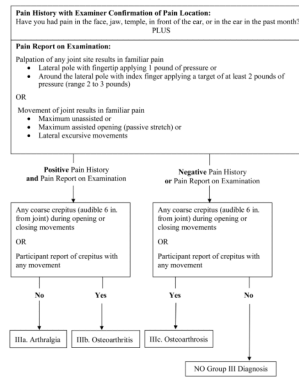




**Figure 1.** Revised Group I Muscle Disorders diagnostic algorithm (expert-driven).



**Figure 2.** Revised Group II Disc Displacements diagnostic algorithm (expert-driven).



**Figure 3.** Revised Group III Arthralgia, Arthritis, and Arthrosis diagnostic algorithm (expert-driven).

Table 1

Sensitivity and Specificity for Original and Revised RDC/TMD Axis I Diagnostic Algorithms, and 95% Confidence Intervals for the Expert-driven Revised Algorithms

Diagnostic Groupings	Algorithm Models					
	Original RDC Point Estimates Sens./Spec.	Data-driven Point Estimates Sens./Spec.	Expert-driven Point Estimates Sens./Spec.	95% CI Expert-driven Sens.	95% CI Expert-driven Spec.	
Any Group I	0.82/0.98	0.95/0.92	0.91/1.00	0.88-0.95	1.00	
Ia	0.75/0.97	0.82/0.97	0.82/0.99	0.74-0.89	0.98-1.00	
Ib	0.83/0.99	0.98/0.95	0.93/0.97	0.89-0.97	0.95-0.99	
Any Group II	0.35/0.96	0.81/0.54	0.71/0.67	0.65-0.76	0.60-0.74	
IIa	0.42/0.92	0.73/0.53	0.46/0.90	0.40-0.52	0.87-0.93	
IIb	0.26/1.00	0.77/0.98	0.80/0.97	0.63-0.90	0.95-0.98	
IIc	0.05/0.99	0.18/0.97	0.53/0.80	0.43-0.62	0.75-0.84	
Any Joint Pain (IIIa or IIIb)	0.42/0.99	0.90/0.83	0.92/0.96	0.88-0.95	0.93-0.98	
Any Arthrosis (IIIb or IIIc)	0.14/0.99	0.46/0.90	0.52/0.86	0.44-0.60	0.82-0.89	
IIIa	0.38/0.90	0.78/0.81	0.74/0.88	0.67-0.79	0.85-0.91	
IIIb	0.13/1.00	0.48/0.92	0.51/0.91	0.42-0.60	0.88-0.94	
IIIc	0.12/0.99	0.20/0.97	0.35/0.95	0.23-0.49	0.93-0.96	

CI = confidence interval. Sens = Sensitivity. Spec = Specificity.

**Table 2**

Diagnostic Reliability of Revised Expert-driven Algorithms: Kappa and Percent Agreement Compared with the Diagnostic Reliability of the Original RDC/TMD Protocol

Diagnostic Grouping	Revised Algorithm Kappa *	Revised Algorithm Percent Agreement **	Original RDC/TMD Intrasite Kappa ***
Any Group I	0.83	93	0.82
Ia	0.73	89	0.60
Ib	0.92	96	0.70
Any Group II	0.84	93	0.58
IIa	0.70	89	0.60
IIb	0.63	93	0.51
IIc	0.72	89	0.13
Any Joint Pain (IIIa or IIIb)	0.85	93	0.55
Any Arthrosis (IIIb or IIIc)	0.87	94	0.33
IIIa	0.81	93	0.52
IIIb	0.71	93	0.36
IIIc	0.79	94	0.28

\* Kappa estimates were computed using generalized estimate equations (GEE).

\*\* Percent agreement was computed from single contingency tables for Group I diagnoses (27 participants) and from two contingency tables (right and left sides) for the Groups II and III diagnoses (total of 54 joints).

\*\*\* Estimates abstracted from Table 2 of the second paper in this series, "Research Diagnostic Criteria for Temporomandibular Disorders: reliability of Axis I diagnoses and selected clinical measures"<sup>4</sup>.



**Table 3**

Reliability of Binary Clinical Tests Incorporated in the Expert-driven Revised Algorithms

Clinical Test	Kappa for Item Reliability *
History of pain (RDC/TMD Question 3 **)	1.00
History of jaw locking (RDC/TMD Question 14 **)	1.00
Temporalis or masseter familiar pain to palpation	0.75
Temporalis or masseter familiar pain with opening	0.63
Detection of muscle pain (Familiar pain to palpation or jaw movement)	0.73
Pain-free unassisted interincisal opening plus vertical incisal overlap < 40 mm	0.92
Detection of disc click	0.70
Maximum assisted interincisal opening (regardless of pain) plus vertical incisal overlap < 40 mm	0.63
Familiar joint pain with 1 lb. palpation pressure	0.86
Familiar joint pain with 2 lb. palpation pressure	0.89
Familiar joint pain with opening	0.61
Familiar joint pain with lateral excursion	0.63
Detection of joint pain (Familiar pain to palpation or with jaw movements)	0.78
Examiner-detected coarse crepitus	0.80
Participant-reported crepitus	0.91
Detection of crepitus (Examiner-detected or participant- reported crepitus)	0.85

\* Kappa estimates were computed using generalized estimate equations (GEE).

\*\* Research Diagnostic Criteria for Temporomandibular Disorders: Review, Criteria, Examinations and Specifications, Critique <sup>1</sup>.