The Research Diagnostic Criteria for Temporomandibular Disorders. III: Validity of Axis I Diagnoses

Edmond Truelove, DDS, MSD
Professor and Chairman
Department of Oral Medicine
University of Washington
Seattle, Washington

Wei Pan, PhD
Professor
Department of Biostatistics

John O. Look, DDS, PhD, MPH
Senior Research Associate
Department of Diagnostic and Biological Sciences

University of Minnesota

Lloyd A. Mancl, PhD, MS, BA
Research Associate Professor
Department of Dental Public Health Sciences
University of Washington
Seattle, Washington

Richard K. Ohrbach, DDS, PhD
Associate Professor
Department of Oral Diagnostic Sciences
University at Buffalo
Buffalo, New York

Ana M. Velly, DDS, MS, PhD
Research Associate
University of Minnesota and
Assistant Professor
McGill University
Montreal, Quebec, Canada

Kimberly H. Huggins, RDH, BS
Affiliate Instructor
Department of Oral Medicine
University of Washington
Seattle, Washington

Patricia Lenton, RDH, MA
Research Fellow

Eric L. Schiffman, DDS, MS
Associate Professor

Department of Diagnostic and Biological Sciences
University of Minnesota

Correspondence to:
Dr Edmond L. Truelove
Department of Oral Medicine
University of Washington
Box 356370
Seattle, WA 98195
Fax: 206-685-8412
Email: edmondt@u.washington.edu

Aims: *To estimate the criterion validity of the Research Diagnostic Criteria for Temporomandibular Disorders (RDC/TMD) Axis I TMD diagnoses.* ***Methods:*** *A combined total of 614 TMD community and clinic cases and 91 controls were examined at three study sites. RDC/TMD Axis I diagnoses were algorithmically derived from an examination performed by calibrated dental hygienists. Reference standards ("gold standards") were established by means of consensus diagnoses rendered by two TMD experts using all available clinical data, including imaging findings. Validity of the RDC/TMD Axis I TMD diagnoses was estimated relative to the reference-standard diagnoses (gold standard diagnoses). Target sensitivity and specificity were set a priori at ≥ 0.70 and ≥ 0.95, respectively.* ***Results:*** *Target sensitivity and specificity were not observed for any of the eight RDC/TMD diagnoses. The highest validity was achieved for Group Ia myofascial pain (sensitivity 0.65, specificity 0.92) and Group Ib myofascial pain with limited opening (sensitivity 0.79, specificity 0.92). Target sensitivity and specificity were observed only when both Group I diagnoses were combined (0.87 and 0.98, respectively). For Group II (disc displacements) and Group III (arthralgia, arthritis, arthrosis) diagnoses, all estimates for sensitivity were below target (0.03 to 0.53), and specificity ranged from below to on target (0.86 to 0.99).* ***Conclusion:*** *The RDC/TMD Axis I TMD diagnoses did not reach the targets set at sensitivity of ≥ 0.70 and specificity of ≥ 0.95. Target validity was obtained only for myofascial pain without differentiation between normal and limited opening. Revision of the current Axis I TMD diagnostic algorithms is warranted to improve their validity.* J OROFAC PAIN 2010;24:35–47

Key words: diagnostic criteria, gold standard, reference standard, temporomandibular disorders, temporomandibular muscle and joint disorders, validity

The Research Diagnostic Criteria for Temporomandibular Disorders (RDC/TMD) constitute a diagnostic protocol that is widely employed by clinical and research personnel.[1] It is a standardized and a well-operationalized diagnostic scheme. However, its full acceptance as a taxonomic system requires a rigorous assessment of its diagnostic validity.

Reliability of the RDC/TMD Examination Protocol

The level of measurement reliability (agreement or measurement reproducibility) associated with any diagnostic instrument can be a limiting factor for its diagnostic validity (accuracy). Reliability

results for the RDC/TMD algorithmic diagnoses are reported in the second article in this series of articles.[2] It was concluded that the reliability of the RDC/TMD protocol is excellent when myofascial pain diagnoses are combined into one diagnosis. Reliability was good for all the common diagnoses, including myofascial pain, myofascial pain with limited opening, arthralgia, and disc displacement with reduction. Of the less common diagnoses, reliability was good for disc displacement without reduction with limited opening. However, reliability was fair to poor for disc displacement without reduction without limited opening, osteoarthritis, and osteoarthrosis. These estimates of reliability are consistent with the results of the most comprehensive study to date that assessed the reliability of the RDC/TMD Axis I TMD diagnoses.[3]

Validity of the RDC/TMD Examination Protocol

Ongoing efforts to investigate the validity of the RDC/TMD taxonomic system were anticipated at its inception in 1992, as it was not intended to be an end product.[1] However, the diagnostic accuracy of the RDC/TMD has not been tested comprehensively. Only limited comparisons of components of the RDC/TMD algorithms have been made to objective evidence or "gold standards." The RDC/TMD diagnosis of arthralgia has been compared to magnetic resonance imaging (MRI) findings for disc displacement, joint effusions or osteoarthritis.[4–6] The RDC/TMD diagnosis for disc displacement has been compared to MRI-detected disc displacement.[7–11] A clinical diagnosis of disc displacement has also been compared to joint sound recordings.[12] The aim of the latter study was to investigate the validity of the RDC/TMD criterion for disc displacement with reduction that stipulates a reciprocal joint click should occur at an interincisal mouth opening measurement that is 5 mm greater than for the closing click. However, no comprehensive assessment of the validity of the RDC/TMD has been performed.

Validity Testing

The definition of validity is the degree to which an index test correctly classifies the presence/absence of a disorder in individuals when compared with a reference standard, also referred to as a gold standard or criterion measure. For such a comparison, both the index test and the reference standard are measured in participants who are suspected of having the condition of interest.[13] With use of a reference standard, the criterion validity of an index test is established. Validity can be evaluated in terms of sensitivity and specificity. Sensitivity is the proportion of participants with the target disorder who have a positive test outcome. Specificity is the proportion of participants without the target disorder who have a negative test outcome. The term "test" refers to any method used to obtain diagnostic information relevant to a patient's health status. It may include information from history and physical examination, laboratory tests, imaging tests, function tests, and histopathology.[13]

The aim of this study was to determine the criterion validity of the RDC/TMD Axis I algorithm-derived TMD diagnoses when compared with reference-standard diagnoses. The reference standards were established by two expert TMD and orofacial pain dentists who independently performed the standardized criterion examination protocol, rendered their diagnoses, and then came together for a consensus diagnosis. This procedure is described in the first article of this series.[14]

Materials and Methods

Study Setting, Locations, and Examiners

Data collections were carried out at three sites: the University at Buffalo (UB), the University of Minnesota (UM), and the University of Washington (UW). Based on terminology recommended by the Standards for Reporting of Diagnostic Accuracy (STARD),[13] these data collections were prospective in that all history, examination, and imaging data collections were planned before the index test (RDC/TMD procedures) and the criterion procedures for the reference standard were performed. No validation data measures were collected retrospectively. A total of nine clinicians served as the examiners for the RDC/TMD Validation Project, including two criterion examiners (CEs) and one test examiner/dental hygienist (TE) for each study site. All six CEs were experienced TMD and orofacial pain dentists. Their training, experience, and reliability have been reported in the first article in this series.[14] The three dental hygienists who served as the TEs were trained and calibrated to perform the RDC/TMD examination protocol.

Study Participants

During August 2003 to September 2006, the three participating study sites recruited a combined total of 628 TMD cases and 91 controls from clinic and community sources. The site principal investigator

or co-investigator at each study site determined that the participants fully understood the study procedures and obtained consent. All procedures were reviewed and approved by the Institutional Review Boards overseeing each study site. Fourteen participants, distributed across the three sites and originally categorized as cases were excluded due to the presence of chondromatosis (n = 2), reported fibromyalgia (n = 9), or reported rheumatoid arthritis (n = 3). Although persons with a documented medical diagnosis of fibromyalgia or rheumatoid arthritis were initially included in the study, they were later excluded from the analyses. Chondromatosis was excluded based on suspicion of the presence of the disorder as detected on MRI by the participating radiologist. It is possible that a small number of subjects may have had fibromyalgia or emerging symptoms of fibromyalgia but potential subjects reporting widespread pain as required in fibromyalgia were excluded to reduce the number of undiagnosed cases. A total of 614 cases remained for the primary analysis. The first article of this series has presented a complete description of the recruitment of the participants and their compensation, the inclusion and exclusion criteria used, the study sample demographics, and baseline clinical characteristics.[14] This first article established that the participant population was appropriate for assessing the validity of the Axis I TMD diagnoses.

Index Test

The index test specified for this validation project was the published RDC/TMD Axis I diagnostic examination procedure that employs a set of standardized clinical and questionnaire items.[1] This procedure allows for assignment of TMD participants to any of three diagnostic groups that include eight subdiagnoses:

- **Group I** Muscle Disorders: (Ia) myofascial pain; (Ib) myofascial pain with limited opening.
- **Group II** Disc Displacements: (IIa) disc displacement with reduction; (IIb) disc displacement without reduction with limited opening; (IIc) disc displacement without reduction without limited opening.
- **Group III** Arthralgia, Arthritis, Arthrosis: (IIIa) arthralgia; (IIIb) osteoarthritis; (IIIc) osteo - arthrosis.

RDC/TMD nomenclature, especially for Group III, is not universally employed. For purposes of this article, the term "arthralgia" is used to describe temporomandibular joint (TMJ) pain without osseous joint changes as determined with computed tomography (CT). The term "arthrosis" is used to signify the presence of any osseous joint change, whether accompanied by joint pain (RDC/TMD osteoarthritis, IIIb) or without joint pain (RDC/TMD osteoarthrosis, IIIc).

Reference Standard (Gold Standard)

The reference-standard diagnoses were derived from the consensus diagnoses of two TMD experts who independently assessed all participants using the criterion examination protocol specified for this study. This protocol incorporated a comprehensive history and examination, as well as imaging that included a panoramic radiograph, bilateral TMJ MRI, and bilateral TMJ CT. All imaging was interpreted by a calibrated board-certified radiologist. These methods have been described in detail in the first article in this series.[14]

Clinical Examination Sequence

The formal validation of the RDC/TMD first required an assessment of each participant by a CE who performed the standardized criterion evaluation and ordered imaging that was completed at the second visit. At the third visit, within 1 to 2 weeks after visit 1, the TE completed the RDC/TMD examination protocol while blinded to the results of the first examination. This was followed on the same day by a second criterion examination performed by the second CE who was blinded to the results of the previous examinations. The reference-standard criterion diagnosis was then established by the two CEs coming together to determine their consensus diagnosis with the participant still present. The radiologist was also involved if the CEs' interpretation of the TMJ MRI or CT findings differed from the radiologist's interpretation. The radiologist rendered the final interpretation of the images following these consultations.

Primary Outcome Variables and Covariates

One of the primary outcomes for this project, and the primary focus of this report, is the validity assessment for the RDC/TMD Axis I TMD diagnoses. Validity was assessed in terms of the sensitivity and specificity of diagnostic outcomes based on the TE examination data collection. The diagnoses were not rendered directly by the TE, but instead were derived by applying the RDC/TMD algorithms to the examinations findings of the TE.

Sensitivity and specificity of these diagnoses were determined relative to the reference-standard diagnoses rendered by the consensus of the two CEs. As is conventional, participants with a positive reference standard diagnosis for a given disorder were included in the sensitivity analyses and those with a negative reference standard diagnosis for a disorder were included in the specificity analyses.

The authors were interested not only in the sensitivity and specificity of the eight RDC/TMD-specified diagnoses, but also in four combinations of these diagnoses, including any Group I muscle pain (Ia or Ib), any Group II disc displacement (IIa, IIb, or IIc), any Group III joint pain (IIIa or IIIb), and any Group III arthrosis (IIIb or IIIc). The combined diagnoses for disc displacement, joint pain, and arthrosis were joint-specific within each participant.

Measurement error associated with the TE examination data collection was evaluated by comparing its agreement with the data collection performed the same day on the same study participant by the second CE. This comparison was facilitated because the criterion examination included all the examination items specified by the RDC/TMD, in addition to a much-expanded examination protocol described in the first paper in this series.[14] Thus, the sensitivity and specificity of the RDC/TMD data collection that was embedded in the criterion data collection could be compared to the TE's results.

The RDC/TMD diagnostic algorithms used for these analyses were published in 1992 as decision and classification tree models.[1] A classification tree is composed of nodes defined by a "split condition" that may consist of a single variable or a combination of variables. Each node in the tree is either satisfied or not satisfied using the data collection for a participant, thus allowing for this individual's diagnostic status to be established by following out the tree to its terminal node. Such diagnostic structures are interpretable and intuitively consistent with theoretical constructs that describe the conditions, or disorders, being diagnosed.

Factors Influencing Diagnostic Sensitivity and Specificity

The planned analysis included reporting the observed variation across study sites in terms of their ranges of sensitivity and specificity estimates and any statistical differences among them, and the assessment of the influence on sensitivity and specificity estimates that might be associated with the number of concurrent TMD diagnoses, right versus left side differences within participants for Group II and III diagnoses, and other baseline variables of interest. The latter included age, gender, education,[1] income,[1] characteristic pain intensity,[1,15] duration of TMD symptoms,[1] depression,[1,16,17] nonspecific physical symptoms,[16,17] pain-related disability,[15] and current or recent treatment for TMD. For purposes of this analysis, education with a range of 7 to 19 years was dichotomized by its median value, ≤ 15 years versus > 15. Income was dichotomized at its median value of US $40,000, as was characteristic pain intensity (0 to 40 versus > 40), duration of TMD pain (≤ 72 months versus > 72 months), depression (0 to 0.3 versus > 0.3), and nonspecific physical symptoms (0 to 0.4 versus > 0.4). Pain-related disability was differentiated by Graded Chronic Pain Scale (GCPS) grades 0 to II versus III to IV.[15] Current treatment for TMD was differentiated into three categories: none, during last 6 months, and prior to 6 months ago.

Validation Study Sample Size Calculations

Sample size requirements were determined as a function of the precision that was stipulated a priori for the sensitivity and specificity estimates in this project. The required precision specified that neither the upper nor lower confidence bound should differ from the point estimate by more than 0.10. Assuming symmetrical confidence bounds, the half-width for each confidence interval (CI) was expressed as $2\sqrt{p(1 - p)/N}$, where p is the estimated sensitivity or specificity, and N is the number of participants truly positive for a diagnosis as determined by the reference-standard diagnosis. This gave a conservative estimate for the precision for Group II and Group III diagnoses, because it assumed only one diagnosis per participant, whereas both joints in each participant were assessed for the validation study. Based on an observed sensitivity or specificity of 0.5 (when the binomial variance is the largest), and with the desired precision defined by upper and lower confidence bounds no greater than 0.10 for all sensitivity and specificity point estimates, 100 participants were required for each diagnosis.

This design resulted in the following recruitment methods:

- Participant recruitment was to include at least 100 cases for each of the eight TMD diagnoses. Given that some participants would have multiple diagnoses for the various TMD subtypes, total case recruitment was planned to be 600.

Initially, study participants were recruited consecutively with no attempt being made to selectively enrich the participant sample for the less common diagnoses (IIb, IIc, IIIb, IIIc). However, after three-fourths (approximately 550) of the participants had been enrolled, selective recruitment was implemented at all sites to ensure an adequate sample size for the less prevalent diagnoses.

- To test rigorously the RDC/TMD, the study also planned to recruit 100 subthreshold cases, ie, participants whose signs and symptoms did not qualify for an RDC/TMD diagnosis, but who nonetheless had at least one of the three cardinal symptoms or signs of TMD: jaw pain, limited mouth opening, or TMJ noise. In addition, these participants were eligible if they had any of the additional TMD diagnoses listed in the Expanded TMD Taxonomy that was developed specifically for this study and reported in Table 3 in the first article in this series.[14]
- Finally, the study planned to recruit 100 controls, that is, participants without signs and symptoms of TMD, to populate four age strata: 18 to 30, 31 to 40, 41 to 50, and 51 to 70 years of age. Equal distributions of controls in each age category was deemed unlikely, given the low prevalence of completely normal joints among older participants, but the goal was to recruit as many controls as possible in the two older age categories so as to be able to select a subset of control participants that reasonably matched the age distributions of the eight TMD diagnostic groups. For example, it was anticipated that the average age of participants with TMJ osteoarthrosis would be greater than that of participants without this diagnosis.

Statistical Procedures

The prevalence of the reference-standard diagnoses in the study sample and the overall percent agreement between the test and reference-standard diagnoses were computed. A logistic regression approach was used to estimate the sensitivity and specificity and utilized generalized estimate equation (GEE) methodology to account for the multiple diagnoses per participant for Group II and Group III diagnoses when CIs and statistical significance were computed.[18] To assess the influence of study site, number of concurrent TMD diagnoses, right versus left side differences for Group II and III diagnoses within participants, and other baseline variables, additional analyses were performed in which each covariate was added separately to the logistic regression model.

Cutoff points for determinations of validity. In the development process for the RDC/TMD, the desired goal for sensitivity was set at ≥ 0.70 and specificity at ≥ 0.95. In the original RDC/TMD monograph, it was shown that, when assuming a prevalence of 10% for TMD, 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.[1] Based on these expectations, the authors intended to declare as valid a given diagnosis—either one of the four diagnostic groupings or one of the eight subdiagnoses—if its estimated sensitivity was at least 0.70 and its estimated specificity was at least 0.95. They planned to apply these validity thresholds to the estimated sensitivities and specificities, even though the lower 95% confidence bounds might include values that were less than the validity threshold.

Secondary analyses. For the secondary analyses involving the 13 specified covariates, 16 tests were planned for each: 8 sensitivity estimates for the Axis I RDC/TMD diagnoses and 8 specificity estimates. No a priori hypotheses could be advanced for these tests based on previous studies. Given that these multiple analyses were purely exploratory, and with each covariate of independent interest, the *P* value for significance was set at ″ .005 to minimize Type I error, but also to allow for detection of associations of potential importance. All analyses were performed with SAS Version 9.1 statistical software (SAS Institute).

Results

Recruitment was monitored over the course of the study relative to distributions and sample sizes for the cases, subthreshold cases, and control participants. When it was determined that a sufficient number of participants had been recruited to satisfy the aims for precision in this study, recruitment was closed with 81 participants fewer than the 800 participants who were originally proposed. An additional 14 participants were excluded from the primary Axis I analysis (see Materials and Methods), leaving a total of 705 participants. This study sample included 614 cases, 579 of which were frank TMD cases, and 35 were clinically normal participants who had a disc displacement with reduction confirmed by MRI. In addition, there were 91 control participants who had no TMD diagnosis. Recruitment totals were fairly evenly distributed across study sites, with

Table 1 Validity of the RDC/TMD Algorithmic Diagnoses and Prevalence of the Reference Standards

Diagnosis	Sensitivity of RDC/TMD*	95% CI for sensitivity*	Specificity of RDC/TMD*	95% CI for specificity*	Percent agreement†	CE sensitivity/ specificity‡	Prevalence of reference-standard diagnoses§
Any Group I	0.87	0.84 – 0.90	0.98	0.94 – 0.99	90	0.84/0.98	0.70
Ia	0.65	0.58 – 0.71	0.92	0.89 – 0.94	84	0.80/0.98	0.30
Ib	0.79	0.74 – 0.84	0.92	0.89 – 0.94	87	0.84/0.99	0.40
Any Group II	0.36	0.32 – 0.39	0.94	0.91 – 0.96	57	0.35/0.95	0.64
IIa	0.38	0.34 – 0.43	0.88	0.85 – 0.90	69	0.40/0.91	0.38
IIb	0.22	0.14 – 0.32	0.99	0.99 – 1.00	94	0.22/1.00	0.06
IIc	0.03	0.01 – 0.06	0.99	0.99 – 1.00	81	0.06/1.00	0.20
Any joint pain (IIIa or IIIb)	0.57	0.52 – 0.61	0.95	0.93 – 0.97	77	0.46/0.98	0.49
Any arthrosis (IIIb or IIIc)	0.15	0.11 – 0.20	0.98	0.97 – 0.99	78	0.15/0.98	0.24
IIIa	0.53	0.48 – 0.58	0.86	0.84 – 0.88	76	0.40/0.88	0.33
IIIb	0.15	0.10 – 0.21	0.99	0.98 – 0.99	86	0.12/0.99	0.16
IIIc	0.10	0.05 – 0.18	0.99	0.98 – 0.99	91	0.16/0.99	0.09

*Data derived from the RDC/TMD test examination performed by the TE.
†The rate of agreement between the TE diagnosis (RDC/TMD algorithm) and the consensus reference-standard diagnosis, taking into account both positive and negative diagnoses.
‡The sensitivity and specificity of the RDC/TMD algorithmic diagnoses from the criterion examination. One CE performed a criterion examination on the same day as the test examination was done.
§Prevalence rates shown are for the total participant sample (n = 705 with 614 cases and 91 controls). For Group I, there is only one diagnosis per subject and the rate denominator is 705, eg, 495/705 = 0.70. For Groups II and III, there are two joints per subject and the rate denominator is 1,410, eg, 532/1,410 = 0.38 for IIa.

35.6% recruited at UB, 30.1% at UM, and 34.3% at UW. The proportion of females among the clinically and radiographically normal participants was 63%. Among the cases, the proportion of females was 85%. The 614 TMD cases had 2,202 RDC/TMD diagnoses assigned by the consensus process, or an average of 3.6 diagnoses per TMD case, with the maximum possible being 5 (1 Group I, 2 Group II, and 2 Group III diagnoses). Thirty-two percent of the 705 participants had 0 to 2 diagnoses, and 68% had 3 to 5 diagnoses. Table 4 of the first article of this series has reported the number of diagnoses within each diagnostic grouping among all study participants.[14] The study sample prevalence rate for each diagnosis is indicated in column 7 of Table 1 in the present article.

Based on the combined data from the three study sites, the overall sensitivity and specificity results for each of the diagnoses and groupings of diagnoses were determined (Table 1). The precision for all sensitivity and specificity estimates in Table 1 was very high. Only one confidence bound differed by as much as 0.10 from the point estimate, that one being the upper bound for sensitivity of IIb. The upper confidence interval differed by 0.06 and the lower confidence interval differed by 0.07.

When the analyses used the validity thresholds defined above as sensitivity ≥ 0.70 and specificity ≥ 0.95, the TE examination attained both target sensitivity and target specificity only for the collapsed group labeled "any Group I" diagnosis (Ia or Ib) (Table 1). Ia sensitivity was slightly deficient at 0.65, as was its specificity (0.92). Ib sensitivity was on target at 0.79 but the specificity of 0.92 was less than the target level. Sensitivity of IIa, IIb, IIc, or any Group II was low (0.03 to 0.38). Specificity was close to the target validity threshold for any Group II (0.94), somewhat deficient for IIa (0.88), and excellent for IIb and IIc (0.99). Sensitivity for any joint pain (IIIa or IIIb) failed to reach target (0.57), whereas specificity was on target (0.95). Sensitivity for IIIa alone was 0.53, and specificity was also not on target at 0.86. Sensitivity for any arthrosis (IIIb or IIIc) was low at 0.15, as was the sensitivity of the individual entities within this grouping. However, specificity was excellent at 0.98 for the combined IIIb or IIIc, as well as for IIIb alone and IIIc alone (0.99).

Percent agreement between the TE results and the reference standards ranged from 81% to 94% for six of the eight RDC/TMD diagnoses (Table 1). This decreased to 69% for disc displacement with reduction (IIa) and to 76% for arthralgia (IIIa).

Measurement error associated with the TE examination data collection was evaluated by comparing its sensitivity and specificity (columns 1 and 3, Table 1) to parallel estimates derived from the second criterion (CE-2) examination data (column 6, Table 1). For IIb and the combined category for any arthrosis (IIIb or IIIc) diagnoses, the TE estimates were virtually identical to the CE-2 estimates. The greatest differences between TE and CE-2 were as follows: the TE sensitivity (0.65) for myofascial pain without limited opening (Ia) was

Table 2 Statistical Influence of Study Site, Age, and Gender on RDC/TMD Sensitivity and Specificity Estimates

	Study site				Age (y)					Gender		
Dx/type	1	2	3	*P*	18–30	31–40	41–50	51–70	*P*	Female	Male	*P*
Ia												
Sensitivity	0.61	0.71	0.63	.45	0.66	0.64	0.63	0.65	.98	0.62	0.78	.06
Specificity	0.89	0.94	0.94	.15	0.93	0.91	0.90	0.94	.67	0.92	0.93	.63
Ib												
Sensitivity	0.76	0.78	0.83	.38	0.82	0.78	0.73	0.81	.58	0.81	0.68	.13
Specificity	0.92	0.96	0.87	.03	0.90	0.96	0.89	0.93	.19	0.89	0.99	< .001*
IIa												
Sensitivity	0.29	0.39	0.50	.002*	0.42	0.35	0.40	0.30	.31	0.40	0.27	.02
Specificity	0.89	0.91	0.84	.04	0.89	0.89	0.87	0.86	.78	0.87	0.92	.04
IIb												
Sensitivity	0.00	0.39	0.20	NC	0.27	0.20	0.18	0.21	.87	0.22	0.22	1.00
Specificity	0.99	1.00	0.99	NC	1.00	0.99	1.00	1.00	.87	1.00	0.99	.70
IIc												
Sensitivity	0.03	0.05	0.01	.23	0.04	0.04	0.02	0.00	NC	0.03	0.00	NC
Specificity	0.99	0.99	1.00	.35	0.99	1.00	1.00	1.00	NC	0.99	1.00	.49
IIIa												
Sensitivity	0.36	0.76	0.52	< .001*	0.60	0.45	0.58	0.37	.02	0.56	0.36	.02
Specificity	0.90	0.81	0.87	.03	0.85	0.89	0.82	0.89	.26	0.84	0.96	< .001*
IIIb												
Sensitivity	0.05	0.12	0.22	.01	0.05	0.16	0.20	0.19	.08	0.14	0.17	.82
Specificity	1.00	1.00	0.98	.05	1.00	0.98	0.99	0.99	.48	0.99	0.99	.87
IIIc												
Sensitivity	0.09	0.00	0.15	NC	0.00	0.08	0.04	0.21	NC	0.13	0.00	NC
Specificity	0.99	1.00	0.97	.008	0.99	1.00	0.98	0.96	.053	0.99	0.99	.40

NC = Due to zero cell counts, GEE procedures could not estimate statistical differences.
*Statistically significant *P* values (≤ .005).
Dx = diagnosis.

lower than the CE-2 sensitivity (0.80). In contrast, TE sensitivity (0.53) for arthralgia (IIIa) was higher than the CE-2 sensitivity (0.40). Relative to the eight RDC/TMD diagnoses, overall TE percent agreement with the reference standard was 84% (derived from Table 1 data), and CE-2 percent agreement averaged 85% (data not shown). This comparison demonstrates that the RDC/TMD examination skills of the TEs were highly comparable to those of the CEs, and that low primary outcome estimates of sensitivity and specificity based on the TE examination data collection were not due to lack of agreement between the TEs and CEs.

Based on the cutoff points for statistical significance at $P \leq .005$, and with the trend toward statistical significance defined as < .01, eight covariates were statistically influential with respect to certain sensitivity and specificity estimates (Tables 2 to 5). Table 2 shows that study site effects were essentially limited to sensitivity estimates, but with no definite pattern that would suggest a systematic difference among them. Site-specific estimates for sensitivity differed for IIa (range: 0.29 to 0.50, $P = .002$) and for IIIa (range: 0.36 to 0.76, $P < .001$). In addition, site-specific differences for IIb sensitivity ranged from 0.00 to 0.39 and, for IIIc sensitivity, 0.00 to 0.15. For the latter two diagnoses, *P* values for an overall difference between sites could not be estimated due to the zero estimates. Specificity estimates for IIIc tended toward a statistical difference between sites ($P < .008$), but the actual magnitude of the difference in estimates was negligible (0.97 to 1.00). Gender was associated with statistically different specificity estimates for Ib and IIIa. The female-male differences in specificity were 0.89 versus 0.99 for Ib and 0.84 versus 0.96 for IIIa ($P = .001$). Sensitivity and specificity did not differ statistically ($P \geq .02$) by age (Table 2).

Comparisons of the 0 to 2 diagnostic category of concurrent diagnoses to the 3 to 5 diagnostic category revealed statistically more ($P < .001$) false-negative diagnoses for a IIa disc displacement when fewer diagnoses were present (sensitivity of 0.21 versus 0.43) (Table 3). In contrast, with fewer diagnoses present, specificity was higher for Ia, Ib, IIa, and IIIa; the 0 to 2 diagnostic category showed specificities of 0.96 to 0.99, while the 3 to 5 diagnostic category had specificities of 0.77 to 0.89 ($P < .001$). When TMD symptoms had been present for 0 to 72 months, there was a trend ($P = .007$)

Table 3 Statistical Influence of Number of Concurrent TMD Diagnoses, Duration of TMD, Side Affected, and Recent Treatment for TMD on RDC/TMD Sensitivity and Specificity Estimates

	Concurrent diagnoses			Duration of TMD (mo)			Side affected			Recent treatment for TMD			
Dx/type	0–2 Diag	3–5 Diag	*P*	0–72	> 72	*P*	Left	Right	*P*	None	< 6 mo	> 6 mo	*P*
Ia													
Sensitivity	0.67	0.65	.83	0.66	0.64	.75	NA	NA	NA	0.64	0.68	0.65	.90
Specificity	0.97	0.89	.001*	0.94	0.87	.007	NA	NA	NA	0.87	0.86	0.89	.74
Ib													
Sensitivity	0.75	0.80	.63	0.76	0.83	.16	NA	NA	NA	0.80	0.75	0.84	.24
Specificity	0.99	0.85	< .001*	0.96	0.80	< .001*	NA	NA	NA	0.85	0.86	0.83	.85
IIa													
Sensitivity	0.21	0.43	< .001*	0.31	0.49	< .001*	0.39	0.37	.54	0.43	0.43	0.46	.81
Specificity	0.96	0.83	< .001*	0.89	0.85	.15	0.89	0.87	.33	0.86	0.83	0.86	.59
IIb													
Sensitivity	0.50	0.21	.48	0.21	0.24	.74	0.30	0.14	.07	0.11	0.27	0.14	.34
Specificity	1.00	0.99	.22	0.99	1.00	.74	0.99	1.00	.12	1.00	0.98	1.00	.34
IIc													
Sensitivity	0.00	0.03	NC	0.00	0.06	NC	0.03	0.02	.47	0.04	0.03	0.03	.98
Specificity	1.00	0.99	.10	1.00	0.99	.16	0.99	1.00	.02	0.99	0.99	0.99	.69
IIIa													
Sensitivity	0.47	0.53	.61	0.49	0.57	.12	0.54	0.52	.54	0.46	0.55	0.56	.36
Specificity	0.98	0.77	< .001*	0.89	0.79	.001*	0.87	0.85	.37	0.83	0.76	0.78	.31
IIIb													
Sensitivity	0.00	0.15	NC	0.14	0.16	.71	0.14	0.15	.61	0.10	0.16	0.14	.75
Specificity	1.00	0.99	NC	0.99	0.98	.11	0.99	0.99	.07	0.99	0.97	0.99	.15
IIIc													
Sensitivity	0.13	0.09	.64	0.10	0.08	.67	0.10	0.10	.77	0.00	0.06	0.04	NC
Specificity	0.99	0.98	.11	0.99	0.99	.98	0.98	0.99	.22	0.99	0.99	0.98	.77

NC = Due to zero cell counts, GEE procedures could not estimate statistical differences.
NA = Not applicable because Group I diagnoses are not side-specific.
*Statistically significant *P* values (≤ .005).

for Ia myofascial pain specificity to be greater (0.94) than when the condition had lasted more than 72 months (0.87). Shorter duration was also associated with increased specificity ($P < .001$) for myofascial pain with limited opening (Ib) and arthralgia (IIIa). In contrast, Table 3 shows that, with increased duration of TMD (> 72 months), sensitivity for IIa disc displacement increased from poor (0.31) to fair (0.49) ($P < .001$). Table 3 also indicates that sensitivity and specificity did not differ statistically by right versus left sides within participants ($P > .02$) or by the report of recent treatment for TMD ($P > .09$).

For the characteristic pain intensity covariate (Table 4), lower level scores (0 to 40) were associated with specificities in the range of 0.92 to 0.97 for Ia, Ib, IIa, and IIIa, whereas the higher category for characteristic pain intensity (index scores > 40) was associated with more false-positive pain diagnoses, and specificity decreased to a range of 0.74 to 0.88 ($P \leq .002$). Nonspecific physical symptoms (0 to 0.4 versus > 0.4) had similar effects on specificity estimates for Ia, Ib, and IIIa. Specificities associated with the lower level ranged from 0.92 to 0.97 while the higher level had specificities of 0.78 to 0.88 ($P < .004$). Depression showed the same effects on specificity for Ib and IIIa (lower level with specificities of 0.90 to 0.96; higher levels with 0.82 to 0.86, $P < .003$). Pain-related disability (Table 5) tended to mimic the pattern of influence of nonspecific physical symptoms and depression on IIIa specificity. However, the difference in IIIa specificity estimates between a state of no disability (GCPS scores of 0 to II) versus a state of disability (scores of III or IV) only tended toward statistical significance ($P = .01$). Table 5 also shows that sensitivity and specificity did not differ statistically according to income ($P = .05$) or education ($P = .04$).

In summary, seven covariates had significant effects on sensitivity and specificity. One covariate, pain related disability, tended toward statistical significance. However, these variables did not change the current report for on-target sensitivity and specificity of the RDC/TMD algorithms except for the effect of gender on Ib sensitivity. This sensitivity estimate decreased from on target overall (0.79) to just below target for males (0.68).

Table 4 Statistical Influence of Characteristic Pain Intensity Score, Nonspecific Physical Symptoms Score, and Depression Score on RDC/TMD Sensitivity and Specificity Estimates

	Characteristic pain intensity			Nonspecific physical symptoms			Depression		
Dx/type	0–40	> 40	*P*	0–0.4	> 0.4	*P*	0–0.3	> 0.3	*P*
Ia									
Sensitivity	0.62	0.67	.44	0.61	0.67	.42	0.64	0.65	.93
Specificity	0.96	0.88	.002*	0.95	0.88	.004*	0.94	0.90	.19
Ib									
Sensitivity	0.75	0.81	.29	0.77	0.81	.48	0.79	0.79	.93
Specificity	0.97	0.81	< .001*	0.97	0.84	< .001*	0.96	0.86	< .001*
IIa									
Sensitivity	0.34	0.42	.06	0.35	0.41	.22	0.38	0.38	.93
Specificity	0.92	0.83	< .001*	0.91	0.85	.01	0.90	0.85	.04
IIb									
Sensitivity	0.12	0.24	.22	0.23	0.22	.90	0.19	0.25	.54
Specificity	1.00	0.99	.25	0.99	1.00	.73	1.00	0.99	.37
IIc									
Sensitivity	0.01	0.04	.15	0.01	0.04	.15	0.02	0.03	.68
Specificity	1.00	0.99	.05	1.00	0.99	.05	0.99	0.99	.51
IIIa									
Sensitivity	0.47	0.56	.14	0.50	0.55	.33	0.50	0.55	.35
Specificity	0.94	0.74	< .001*	0.92	0.78	< .001*	0.90	0.82	.003*
IIIb									
Sensitivity	0.14	0.15	.90	0.11	0.16	.29	0.12	0.17	.38
Specificity	1.00	0.98	.13	1.00	0.98	.03	0.99	0.99	.38
IIIc									
Sensitivity	0.13	0.03	.07	0.12	0.07	.41	0.10	0.10	.96
Specificity	0.99	0.99	.74	0.99	0.99	.78	0.99	0.99	.82

*Statistically significant *P* values (≤ .005).

Table 5 Statistical Influence of Pain-related Disability, Annual Income, and Education on RDC/TMD Sensitivity and Specificity Estimates

	Pain-related disability			Annual income			Education		
Dx/type	GCPS 0 to 11	GCPS III or IV	*P*	≤ $40,000	> $40,000	*P*	≤ 15 y	> 15 y	*P*
Ia									
Sensitivity	0.65	0.65	.98	0.61	0.69	.21	0.61	0.68	.26
Specificity	0.93	0.86	.14	0.92	0.93	.64	0.90	0.94	.05
Ib									
Sensitivity	0.78	0.86	.25	0.79	0.79	.97	0.76	0.83	.14
Specificity	0.92	0.75	.02	0.90	0.93	.39	0.92	0.91	.79
IIa									
Sensitivity	0.38	0.39	.94	0.36	0.40	.37	0.39	0.37	.62
Specificity	0.89	0.81	.09	0.86	0.90	.05	0.89	0.87	.22
IIb									
Sensitivity	0.22	0.24	.88	0.24	0.19	.63	0.16	0.26	.24
Specificity	0.99	1.00	.88	0.98	0.99	.19	0.99	1.00	.55
IIc									
Sensitivity	0.03	0.00	NC	0.02	0.04	.42	0.03	0.02	.81
Specificity	0.99	0.99	.79	0.99	0.99	.91	1.00	0.99	.30
IIIa									
Sensitivity	0.52	0.65	.09	0.52	0.55	.58	0.50	0.56	.25
Specificity	0.87	0.71	.01	0.86	0.87	.57	0.89	0.84	.04
IIIb									
Sensitivity	0.13	0.23	.32	0.15	0.13	.66	0.16	0.13	.65
Specificity	0.99	0.98	.40	0.99	0.99	.20	0.99	0.99	.14
IIIc									
Sensitivity	0.12	0.00	NC	0.07	0.15	.31	0.13	0.06	.25
Specificity	0.99	0.99	.57	0.99	0.99	.68	0.99	0.99	.92

NC = Due to zero cell counts, GEE procedures could not estimate statistical differences.

Discussion

The validity of the RDC/TMD Axis I TMD algorithmic diagnoses was estimated by their sensitivity and specificity relative to reference-standard consensus diagnoses of two TMD experts using a comprehensive assessment. The RDC/TMD algorithmic diagnoses were derived from an examination performed by a trained dental hygienist. For a diagnostic procedure to be declared valid, both target sensitivity (≥ 0.70) and target specificity (≥ 0.95) had to be met as proposed by Dworkin and LeResche.[1] Based on these criteria, only a diagnosis of "myofascial pain" achieved the desired diagnostic accuracy for both sensitivity and specificity when both Ia and Ib were not differentiated. Sensitivity for Group II disc displacement diagnoses was below target, and specificity ranged between below target to on target. Sensitivity for any joint pain was below target, as was sensitivity for any arthrosis. These results suggest a need for revising the Axis I TMD diagnostic algorithms. The study attempted to exclude subjects with other forms of regional pain including odontogenic pain, any specific craniofacial neuralgia, nonspecific neuropathic pain, and pain arising from recent trauma (see Table 1, reference 14) in addition to pain associated with fibromyalgia and rheumatoid arthritis. Therefore, the findings should be generally applicable to assessment of patients in routine clinical practice where such disorders are excluded during the normal diagnostic process.

Statistical Effects of Covariates on Sensitivity and Specificity Estimates

The reason for an association between the female gender and lower specificity for Ib and IIIa pain diagnoses may relate to self-reported TMD pain in women being approximately twice that of men.[19] The study did not analyze this finding further to see if other factors may have influenced these gender differences including longer pain duration, more sites of pain to palpation, higher characteristic pain intensity, depression, and nonspecific physical symptoms. These factors are related to higher sensitivity to noxious stimulation and may influence reporting of responses to physical pain measures. As for the impact of the number of concurrent diagnoses, the observed pattern for sensitivity and specificity estimates suggests that the rater was more likely to detect a true disc displacement (higher sensitivity) when more diagnoses were present. One might speculate that when a higher number of diagnoses are present, the participant would be more likely to have a Group II diagnosis, and that the higher diagnostic sensitivity for Group II could be a function of the increased opportunity for this diagnosis. It is noted, however, that sensitivity and specificity estimates are theoretically independent of the prevalence of a given condition,[20] and that the influence of the concurrent conditions on such higher sensitivity is not clear. The specificity for Ia, Ib, IIa, and IIIa was significantly greater in the presence of two or fewer diagnoses than when compared with three or more. The RDC/TMD was less likely to give a false-positive diagnosis when fewer diagnoses were present.

Duration of symptoms, characteristic pain intensity, and nonspecific physical symptoms all had a similar effect on the specificity of pain diagnoses (Ia, Ib, and IIIa), with depression showing the same effect on specificity for Ib and IIIa. Lower levels were consistently associated with higher specificity. The same pattern, supported by a trend toward statistical significance, was observed for pain-related disability relative to its effect on specificity for IIIa. However, additional analyses to be completed may reveal more complex interactions associated with these Axis II covariates.

The statistical differences observed between study sites are more difficult to interpret. To render the estimates of validity more generalizable, the authors intentionally recruited participants from both coasts as well as the north-central area of the US to obtain a more heterogeneous participant sample. Accordingly, they anticipated that there would be some differences in study populations recruited from widely separated locations across the US. To study the site-specific differences in IIa and IIIa sensitivity estimates, preliminary adjusted analyses have been performed using a GEE multivariate model. When the effects of age, gender, number of concurrent diagnoses, left- versus right-side differences, and recent treatment for TMD were controlled for, the significance of site differences for IIa sensitivity did not remain. In contrast, the significance of the site differences for IIIa sensitivity has been maintained throughout the analyses to date. It is the authors' intention to report the *observed* differences between sites at this time. A better understanding of these differences may be found after further data analyses using other regression models with different combinations of variables and assessment of their interactions. With respect to the observed site differences, the following are noted:

- Use of the RDC/TMD protocol revealed that, the diagnostic detection rate was low for the diagnoses IIa, IIb, IIIa, and IIIc, likely due to lack of imaging data in the RDC/TMD protocol. Low examiner detection rates due to the design of the diagnostic system result in greater variability for the estimates of agreement.
- Considerable variation has also been reported across the 10 sites in the International Consortium for RDC/TMD-based Research. The site-specific diagnostic agreement showed intraclass correlation coefficient (ICC) ranges of 0.29 to 0.71 for IIa, 0.00 to 0.92 for IIb, 0.22 to 0.66 for IIIa, and 0.00 to 0.74 for IIIc.[3]
- None of the three sites for this study was consistently high or low as to diagnostic accuracy.

Generalizability of Study Estimates

The diagnostic validity estimates presented in this article are both credible and generalizable. As reported in the first article in this series, the demographics and characteristics of the study participants are comparable to other studies that have employed the RDC/TMD and appropriate for testing this examination protocol.[14] The broad geographical distribution of the participants provides credibility for these estimates based on the heterogeneity of the study sample. In addition, the large sample size in this study made it the first to obtain adequately precise confidence intervals for its point estimates. Although this project was not designed to be a definitive assessment of the influence of covariates on sensitivity and specificity estimates, these findings should be viewed as preliminary data to be taken into consideration for future study designs. To the extent that these effects are not due to chance, small study samples may be more susceptible to chance-related perturbations that can affect estimates of validity for TMD.

Validity Study Design

Case definitions for musculoskeletal disorders are a set of diagnostic criteria that establish the boundaries by which a disorder can be considered to be present or absent. For TMD, these criteria are determined by a clinical examination based on muscle and TMJ palpation, auscultation of the TMJ, and measurement of mandibular range of movement. This protocol may also be supplemented, as the 1992 RDC/TMD publication states, with imaging for detection of certain intra-articular disorders.[1] Some investigators may suggest that only an objective biological measurement can constitute a reference standard for the presence/absence of a disorder. In the absence of such an objective reference standard, however, the closest reference standard to date for TMD, especially TMD pain, is the comprehensive clinical assessment that was specified for this study. The usefulness of a diagnostic protocol for TMD that is necessarily based on clinical signs and symptoms is limited in the presence of competing diagnoses that share some or all of the same signs and symptoms (ie, comorbid conditions such as fibromyalgia). Because this is the first study to comprehensively assess the TMD diagnoses with an adequate sample size, STARD recommendations were followed to compare participants presenting only with the target condition (TMD) with healthy controls.[13] Cases with potentially confounding comorbid conditions were eliminated through specific exclusion criteria as reported in the first article in this series.[14]

Parallels Between Validity Testing in This Study and Other Areas of Medicine

Developing a credible reference standard to assess the construct of muscle and joint pain is challenging, given that pain is a subjective experience. Psychology and psychiatry have successfully dealt with this issue through structured interviews.[21] In rheumatology, the reference standard for the diagnosis of fibromyalgia is a consensus diagnosis established by two rheumatologists using all available clinical tests. The resultant diagnostic criteria, however, were simplified to 11 of 18 points being reported by the participant as "tender" to palpation, with a concurrent self-report of widespread pain.[22] Other areas of medicine have used anesthetic blocks to make diagnostic decisions,[23,24] although the placebo effect of blocks complicates their interpretation. Still other approaches have been to make diagnostic inference based on the results of interventions. Headache is one such disorder that is "confirmed" when a treatment mitigates the pain.[25] Such interventions can also have many nonspecific effects that limit the conclusions from these tests. Some diagnostic criteria for musculoskeletal pain include the construct of asking the participant if a pain-provocation test produces "familiar pain."[23,24,26–32] This method takes into consideration that many tests can produce pain, but that "familiar pain" allows the participant to identify the pain as reproducing the chief complaint. Other methods are based on the participant pointing to the location of his/her pain when location is informative in making a diagnosis. Because

the criterion validity of the RDC/TMD Axis I diagnostic algorithms for myofascial pain and arthralgia had not been previously assessed, the reference standard specified in this project utilized many of these approaches.[14] Relative to TMJ disc displacements and arthrosis, the standard in medicine for assessing intra-articular soft and hard tissue is based upon MRI and CT findings, respectively. In the current project, the reference standard specified the use of these imaging techniques for assessing disc displacement and osteoarthritis, and required that experienced board-certified radiologists interpret these images.

Conclusions

Clinical examination methods used in the RDC/TMD and the diagnoses derived from them are able to detect myofascial TMD pain diagnoses with acceptable validity. Therefore, this simple and well-standardized clinical examination method can be recommended for the diagnosis of myofascial TMD pain. However, when the RDC/TMD protocol is employed, arthralgia can only be diagnosed at a level that is clearly below desirable goals for sensitivity and specificity. Validity for diagnoses of disc displacements and arthrosis (that is, osseous joint changes) based upon the RDC/TMD protocol was found to be poor; valid diagnoses for these disorders must be based on the synthesis of patient-reported, clinical, and radiological data. Overall, the results of this study support the need for revising the current Axis I TMD diagnostic algorithms to improve their validity.

Secondary analyses of the data revealed statistically significant influences on diagnostic accuracy associated with study site, gender, number of concurrent TMD diagnoses, duration of TMD symptoms, characteristic pain intensity, depression, nonspecific physical symptoms, and possibly, pain-related disability. Future TMD studies should continue to evaluate the influence of these covariates.

Acknowledgments

The authors thank the following personnel of the RDC/TMD Validation Project: at the University of Minnesota, Mansur Ahmad, Quentin Anderson, Mary Haugan, Amanda Jackson, Wenjun Kang, and Feng Tai; at the University at Buffalo, Carla Beneduce, Leslie Garfinkel, Yoly Gonzalez, Patricia Jahn, Krishnan Kartha, Sharon Michalovic, and Theresa Speers; and, at the University of Washington, Lars Hollender, Julie Sage, Kathy Scott, Jeff Sherman, and Earl Sommers. Research supported by NIH/NIDCR U01-DE013331 and N01-DE-22635.

References

1. Dworkin SF, LeResche L. Research diagnostic criteria for temporomandibular disorders: Review, criteria, examinations and specifications, critique. J Craniomandib Disord Facial Oral Pain 1992;6:301–355.
2. Look J, John M, Tai F, et al. The research diagnostic criteria for temporomandibular disorders. II. Reliability of axis I diagnoses and selected clinical measures. J Orofac Pain 2010;24:25–34.
3. John MT, Dworkin SF, Mancl LA. Reliability of clinical temporomandibular disorder diagnoses. Pain 2005;118: 61–69.
4. Shaefer JR, Jackson DL, Schiffman EL, Anderson QN. Pressure-pain thresholds and MRI effusions in TMJ arthralgia. J Dent Res 2001;80:1935–1939.
5. Emshoff R, Brandlmaier I, Bosch R, Gerhard S, Rudisch A, Bertram S. Validation of the clinical diagnostic criteria for temporomandibular disorders for the diagnostic subgroup—Disc derangement with reduction. J Oral Rehabil 2002;29:1139–1145.
6. 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.
7. 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.
8. 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.
9. 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.
10. 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.
11. 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.
12. Huddleston Slater JJ, Lobbezoo F, Naeije M. Mandibular movement characteristics of an anterior disc displacement with reduction. J Orofac Pain 2002;16:135–142.
13. Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. Clin Chem 2003;49:1–6.
14. Schiffman EL, Truelove E, Ohrbach R, et al. The research diagnostic criteria for temporomandibular disorders. I. Overview and methodology for assessment of validity. J Orofac Pain 2010;24:7–24.
15. Von Korff M, Ormel J, Keefe FJ, Dworkin SF. Grading the severity of chronic pain. Pain 1992;50:133–149.
16. Derogatis L. SCL-90-R: Symptom checklist-90-R. Administration, scoring and procedures manual. Psychopharmacol Bull 1994;9:12–28.

17. Derogatis LR, Lipman RS, Covi L. SCL-90: An outpatient psychiatric rating scale—Preliminary report. Psychopharmacology 1973;9:13–28.
18. Smith PJ, Hadgu A. Sensitivity and specificity for correlated observations. Stat Med 1992;11:1503–1509.
19. LeResche L. Epidemiology of orofacial pain. In: Lund JP, Levigne GJ, Dubner R, Sessle BJ (eds). Orofacial Pain. From Basic Science to Clinical Management. Chicago: Quintessence, 2001:15–25.
20. Fletcher RH, Fletcher SW,Wagner EH. Clinical Epidemiology: The Essentials. Baltimore: Wilkins & Wilkins, 1996.
21. Rogers R. Handbook of Diagnostic and Structured Interviewing. New York: Guilford, 2001.
22. Wolfe F, Smythe HA, Yunus MB, et al. The American College of Rheumatology 1990 criteria for the classification of fibromyalgia. Report of the multicenter criteria committee. Arthritis Rheum 1990;33:160–172.
23. Laslett M, Aprill CN, McDonald B, Young SB. Diagnosis of sacroiliac joint pain: Validity of individual provocation tests and composites of tests. Man Ther 2005;10:207–218.
24. 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.
25. Headache Classification Subcommittee of the International Headache Society. The international classification of headache disorders ICHD-II tension-type headache (TTH). Cephalalgia 2004;24 (suppl 1):37–43.
26. McFadden JW. The stress lumbar discogram. Spine 1988;13:931–933.
27. Thevenet P, Gosselin A, Bourdonnec C, 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.
28. 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.
29. Vaksmann 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.
30. Davies HA, Kaye MD, Rhodes J, Dart AM, Henderson AH. Diagnosis of oesophageal spasm by ergometrine provocation. Gut 1982;23:89–97.
31. Wise CM, Semble EL, Dalton CB. Musculoskeletal chest wall syndromes in patients with noncardiac chest pain: A study of 100 patients. Arch Phys Med Rehabil 1992; 73:147–149.
32. 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.