Interexaminer Reliability and Clinical Validity of the Temporomandibular Index: A New Outcome Measure for Temporomandibular Disorders

Jason Pehling, DDS, MS
TMJ and Orofacial Disorders Center
Seattle, Washington

Eric Schiffman, DDS, MS
Associate Professor

John Look, DDS, PhD
Senior Research Associate

TMJ and Orofacial Pain Clinic
School of Dentistry
University of Minnesota
Minneapolis, Minnesota

J. Shaefer, DDS, MS
Assistant Professor
Department of Oral Surgery
Harvard School of Dental Medicine
Boston, Massachusetts

Pat Lenton, RDH
Research Fellow
Clinical Dental Research Center
School of Dentistry
University of Minnesota
Minneapolis, Minnesota

James Fricton, DDS, MS
Professor
TMJ and Orofacial Pain Clinic
School of Dentistry
University of Minnesota
Minneapolis, Minnesota

Correspondence to:
Dr Jason Pehling
2111 N Northgate Way, Suite 221
Seattle, WA 98133

Based on a thesis submitted to the graduate faculty, University of Minnesota, by Dr Jason Pehling, in partial fulfillment of the requirements for the MS degree.

Aims: *The operational definitions for the Craniomandibular Index (CMI) were redesigned to conform precisely to those of the Research Diagnostic Criteria for Temporomandibular Disorders (RDC/TMD), resulting in a single examination protocol, the Temporomandibular Index (TMI). The objectives were to evaluate interexaminer reliability of the TMI as well as its criteria and construct validity for measurement of TMD severity.* ***Methods:*** *Interexaminer reliability of the TMI was assessed on 12 subjects. Criterion validity of the TMI was evaluated relative to the CMI, the latter having established validity. Construct validity of the TMI was evaluated for its capacity to differentiate TMD patients (n = 79) from normal subjects (n = 20) and to detect changes in severity over time.* ***Results:*** *The examiners' average TMI scores were 0.27 ± 0.19 (SD) and 0.26 ± 0.20. Agreement was excellent, with an intraclass correlation coefficient (ICC) of 0.93. The scores for the TMI and the CMI correlated highly, with an ICC of 0.97. Statistical contrasts between the symptomatic groups and the normal subjects were highly significant (P < .001). In 20 TMD patients who underwent treatment for their disorder, their mean change of 0.12 from their pretreatment TMI scores was highly significant (P < .001).* ***Conclusion:*** *This study has provided statistical evidence for the clinical reliability and validity of the TMI, which indicates that the RDC examination protocol is appropriate for determining TMD severity by the TMI algorithm, and diagnosis of TMD subtypes by the RDC algorithm.*

J OROFAC PAIN 2002;16:296–304.

Key words: temporomandibular disorder, reliability, validity, outcomes, Research Diagnostic Criteria, Craniomandibular Index

High rates of success have been reported for a variety of treatments for temporomandibular disorders (TMD), including medications, intraoral orthotics, physical therapy, cognitive-behavioral interventions, occlusal therapy, and surgery. Consequently, clinicians are faced with the dilemma of basing their treatment choices on sometimes controversial and often conflicting concepts concerning the etiology, diagnosis, and treatment of patients with TMD.[1] The National Institutes of Health (NIH) in the United States have responded to these ambiguities by calling for more observational studies and randomized clinical trials. Furthermore, the NIH recommended that these studies utilize appropriate measures of clinical outcome.[2]

The Research Diagnostic Criteria for TMD (RDC/TMD) were developed to address some of these concerns through their classifi-

cation of TMD subtypes within subjects.[3] When used as an outcome measure, this instrument's limitation is that it was not designed to evaluate the relative severity of the disorders that comprise TMD. Therefore, 2 examinations are required to assess both the diagnosis and the severity of TMD. Measures of severity have most commonly been obtained through the use of either the Helkimo Clinical Dysfunction Index[4,5] or the Craniomandibular Index (CMI).[6,7] To obviate this need for 2 separate exams, the operational definitions for the CMI exam items have been redesigned to conform precisely to those specified for the RDC/TMD. Thus, for the first time, a single clinical examination can be used to determine both the specific diagnosis of a TMD and the relative severity of the disorder. This revised CMI is referred to as the Temporomandibular Index (TMI).

The usefulness of the TMI in both observational and experimental studies is a function of its reliability and validity. Reliability is defined as an estimate of the repeatability or consistency of a measure.[8] Unreliable clinical measures cannot be interpreted with confidence when measuring the effects of a disorder or the efficacy of a particular treatment.[8] For TMI measures to be deemed reliable, there must be demonstrated consistency between the classification outcomes of both asymptomatic and symptomatic persons by both individual and multiple examiners performing repeated examinations.[9]

The validity of an instrument is an estimate of how well it measures what it purports to measure. For a condition such as TMD, there may be considerable uncertainty as to whether an index score adequately characterizes the complex nature of this disorder. Thus, the validity of such instruments is commonly assessed from 4 perspectives: *(1)* face validity, *(2)* content validity, *(3)* construct validity, and *(4)* criterion validity.[8] Face validity merely states that an instrument appears to measure what it is supposed to measure. Content validity refers to the degree to which the items of an instrument are representative of the current knowledge base. Construct validity requires that the instrument be sensitive to and follow a predetermined intuitive concept. Criterion validity of a new measurement method is evaluated by its comparison to an accepted standard.

The specific aims of this study were to evaluate the reliability of the TMI, as well as its criterion and construct validity with respect to estimates of TMD severity. Since the face validity and content validity of the TMI are both well supported in the literature, this study does not address these perspectives. Also, the validity of the RDC/TMD has already been affirmed, as demonstrated by its current use in TMD research.[10–14] Because the TMI instrument includes the same examination items as the RDC/TMD, the diagnostic validity of the TMI would be expected to be similar to that of the RDC/TMD.

Materials and Methods

Description of the TMI

The TMI is composed of 3 subindexes: *(1)* the Function Index (FI), *(2)* the Muscle Index (MI), and *(3)* the Joint Index (JI). The FI includes 12 items related to the range of motion (ROM) of the mandible. These items characterize pain or limitation related to mandibular ROM and deviation of the mandible during opening movement. The MI measures pain associated with bilateral digital palpation of selected intraoral and extraoral masticatory muscles at a total of 20 sites. The JI measures pain evoked by digital palpation of 2 sites for each temporomandibular joint (TMJ) and the incidence of noise in each TMJ (Fig 1).

The specific definition of each examination item and the operational definitions for the measurements included in the TMI are exactly as described for the RDC/TMD.[3] Examination sites with no pain or deviations are scored as 0; those sites positive for pain or deviations are scored as 1. As shown in Fig 1, the FI, MI, and JI are calculated by dividing the sum of positive findings for each subindex by the total number of items examined. Accordingly, the FI is calculated by dividing the sum of the positive findings by 12, the MI is calculated by dividing the sum of the positive findings by 20, and the JI is calculated by dividing the sum of the positive findings by 8. The overall TMI score is the average of the scores for the FI, MI, and JI. Each of the TMI and the 3 subindex scores vary between 0 and 1, with 1 being the highest score possible.

Assessment of Interexaminer Reliability

The examiners included a dentist (ES) experienced in the diagnosis and treatment of TMD and a dental hygienist (PL) who had been trained in the use of the CMI. Both examiners underwent more than 50 hours of education, training, and calibration in the use of the TMI. Just prior to beginning the study examinations, there was an examiner brief-

Mark [0] if the exam item is negative and [1] if the exam item is positive.

I. Function Index: Mandibular range of motion parameters

For range of motion items, the measurement in parentheses indicates normal or baseline values. All opening movements include the incisal to incisal opening measurement plus the vertical overlap of the incisors. If an anterior open bite is present, then the amount of open bite is subtracted from the incisal to incisal opening measurement.

			Pain with movement		
Maximum unassisted opening without pain (≥ 40 mm)	____mm [0]	[1]			
Maximum unassisted opening (≥ 40 mm)	____mm [0]	[1]	Pain	[0]	[1]
Maximum assisted opening (≥ 40 mm)	____mm [0]	[1]	Pain	[0]	[1]
Right lateral (≥ 7 mm)	____mm [0]	[1]	Pain	[0]	[1]
Left lateral (≥ 7 mm)	____mm [0]	[1]	Pain	[0]	[1]
Protrusion (≥ 7 mm)	____mm [0]	[1]	Pain	[0]	[1]

Vertical overlap of incisors ± ____mm

Opening pattern (mark only one box in this section):

Straight	[0]
Corrected deviation	[1]
Uncorrected deviation	[1]
Other (jerky, etc)	[1]

Function Index: Total number of positive responses _____/12 = _____

II. Muscle Index: Masticatory muscle palpitation sites

Right			*Left*		
Anterior temporalis	[0]	[1]	Anterior temporalis	[0]	[1]
Middle temporalis	[0]	[1]	Middle temporalis	[0]	[1]
Posterior temporalis	[0]	[1]	Posterior temporalis	[0]	[1]
Origin of the masseter	[0]	[1]	Origin of the masseter	[0]	[1]
Body of the masseter	[0]	[1]	Body of the masseter	[0]	[1]
Insertion of the masseter	[0]	[1]	Insertion of the masseter	[0]	[1]
Posterior mandibular region	[0]	[1]	Posterior mandibular region	[0]	[1]
Submandibular region	[0]	[1]	Submandibular region	[0]	[1]
Lateral pterygoid area	[0]	[1]	Lateral pterygoid area	[0]	[1]
Tendon of the temporalis	[0]	[1]	Tendon of the temporalis	[0]	[1]

Muscle Index: Total number of positive responses _____/20 = _____

III. Joint Index: TMJ palpation and TMJ noise

TMJ palpation:

Right			*Left*		
Lateral pole	[0]	[1]	Lateral pole	[0]	[1]
Posterior attachment	[0]	[1]	Posterior attachment	[0]	[1]

[For scoring TMJ noise, count only one positive per side for sections A and B.]

Right			*Left*		
A: Reproducible opening click	[0]	[1]	A: Reproducible opening click	[0]	[1]
Reproducible closing click	[0]	[1]	Reproducible closing click	[0]	[1]
Reproducible reciprocal click	[0]	[1]	Reproducible reciprocal click	[0]	[1]
Reproducible laterotrusive click	[0]	[1]	Reproducible laterotrusive click	[0]	[1]
Reproducible protrusive click	[0]	[1]	Reproducible protrusive click	[0]	[1]
Nonreproducible click	[0]	[1]	Nonreproducible click	[0]	[1]

[*Nonreproducible clicks occurring with any of the jaw movements are not used for scoring purposes.]

B: Coarse crepitus	[0]	[1]	B: Coarse crepitus	[0]	[1]
Fine crepitus	[0]	[1]	Fine crepitus	[0]	[1]

Joint Index: Total number of positive responses _____/8 = _____

TMI: Function Index + Muscle Index + Joint Index / 3 = _____

Fig 1 Clinical examination form for data collection on the items of the temporomandibular index.

ing to clarify definitions and review scoring of the TMI. Each examiner evaluated 2 practice subjects; their findings were then compared, and any differences were reconciled. The practice subjects were asked to provide verbal feedback with regard to consistency of palpation pressure between examiners. Following the examiner briefing, the actual study examinations were completed.

Interexaminer agreement was assessed by use of the intraclass correlation coefficient (ICC). The ICC is employed to factor out variation associated with 2 or more examinations on the same subject and to estimate the proportion of the total variability that is explained by true differences between the subjects. It can be assumed that there should be little difference between examination findings performed on the same subject. If this holds true, the ICC would then be close to 1.0. Conversely, if all the variability can be explained by discordant exam scores (ie, disagreement between examiners), the ICC would be 0.0.

The sample size of 12 subjects was based on a power analysis and formulas published by Mian and Shoukri.[15] Specifically, this sample size was adequate to ensure statistical power of the study to detect an ICC of 0.9 with a lower 95% confidence limit of no less than 0.7. This level of correlation, if it existed, would confirm good agreement. Among the 12 subjects for the reliability study, there were 9 women and 3 men with an average age of 29 years (range, 21 to 46 years). Eight of the subjects complained of symptoms suggestive of TMD, while 4 were asymptomatic. In all subjects, however, some positive signs of TMD were observed. Both examiners performed the specified data collection for the TMI on all 12 subjects in a blinded and random order, with 20 to 30 minutes between each examination.

Assessment of Criterion Validity

Ideally, criterion validity would be measured relative to a "gold standard." As no such standard exists for TMD, criterion validity of a new index requires its comparison to an accepted index that measures the same construct. Thus, the original CMI was validated against the Helkimo Dysfunction Index, and a Pearson correlation coefficient of 0.89 was reported.[7] Since the CMI has been used and validated repeatedly in clinical studies, the TMI was compared to it. The 2 examiners (ES and JF) who originally developed the CMI assessed the criterion validity of the TMI for estimation of TMD severity. They had both undergone more than 50 hours of training and calibration on the use of the TMI. For this component of the study, examiner 1 employed the original CMI and examiner 2 employed the TMI. The same 12 subjects were employed as in the reliability study; therefore, this comparison required just 1 additional examination for each subject by examiner 1 using the CMI protocol. Both examiners were blinded to the status of the subjects and the findings of each other, and their examinations were separated by 20 to 30 minutes. Their findings were compared for agreement by use of the ICC.

The issue of whether a 2- or 4-point scale should be used to measure muscle tenderness has also been debated. Therefore, it was decided to evaluate this question as a nested component of this criterion validity assessment. The RDC/TMD guidelines call for a 4-point scale for measuring muscle tenderness (none, mild, moderate, and severe), while the TMI specifies a 2-point scale (presence or absence of pain). In this study, the 4-point response was recorded for the TMI as per RDC/TMD guidelines, while the 2-point response was recorded for the CMI. This permitted a side-by-side comparison of the 2 scales as to the time they required. For the analytic comparison of the instruments, the TMI data were collapsed into a 2-point scale.

Assessment of Construct Validity

The construct validity of the TMI was assessed by its ability to differentiate symptomatic from asymptomatic subjects based on the severity of their TMD signs, and to measure changes in symptomatic patients over time as the severity of their disorder changed. Differentiation of symptomatic from asymptomatic subjects involved 79 symptomatic patients and 20 asymptomatic controls. All symptomatic patients were either referred by their primary care providers or were self-referred to the University of Minnesota TMJ and Orofacial Pain Clinic. All asymptomatic subjects were responders to advertisements displayed at the University of Minnesota Dental School. Three diagnostic groups of symptomatic patients were enrolled, including 30 patients with an RDC/TMD diagnosis of myofascial pain (MFP), 29 patients with an RDC/TMD diagnosis of disc displacement with reduction (DDWR), and 20 patients with an RDC/TMD diagnosis of disc displacement without reduction (DDWOR). Among these 79 patients there were 59 women and 20 men with an average age of 31 years (range, 21 to 48 years). The 20 asymptomatic controls (ASY) were recruited to approximate the age and sex distributions of the 3

symptomatic groups. They included 15 women and 5 men with an average age of 31 years.

Of the 30 MFP subjects, 12 had restricted ROM according to RDC/TMD criteria, while 18 were without restriction. A majority (22 of 30) of these subjects also had TMJ arthralgia, but none had a disc displacement as assessed by RDC/TMD criteria. All of the DDWR subjects presented with both myofascial pain and arthralgia, but there were no non-reducing discs. All of the DDWOR group had disc displacements without reduction for more than 6 months and presented with both myofascial pain and TMJ arthralgia. Six of the DDWOR subjects had limitation in opening. A 2-sample *t* test for independent samples was specified to determine the statistical difference between the mean TMI scores for the symptomatic and asymptomatic groups. The *P* value for statistical significance was set at < .01 to reduce the likelihood of Type I error.

The second method for assessing the construct validity of the TMI involved a comparison of pre- and post-treatment scores in 20 subjects (17 women and 3 men) who underwent treatment at the University of Minnesota TMJ and Orofacial Pain Clinic. These patients had an average age of 25 (range, 18 to 44 years), and they presented with mixed muscle and joint TMD diagnoses as per the RDC/TMD. Treatment consisted of home-based self-care and of 2 or more of the following modalities: medications, intraoral orthotics, physical therapy, cognitive-behavioral intervention, and/or TMJ arthrocentesis. The pretreatment TMI score was measured at the time of the initial evaluation. Post-treatment testing was completed when the patient and treating doctor (who was not one of the examiners, ES or JF) determined that a healing plateau had been achieved. The average time between examinations was 6 weeks (standard deviation [SD] 1.4 weeks). The paired *t* test was used to determine the statistical difference between the pre- and post-treatment scores, and the *P* value for statistical significance was set at < .01 to limit the probability of Type I error.

The issue of the validity of cervical muscle palpation as an indicator of TMD severity was reexamined as a nested addition to the assessment of the construct validity of the TMI. Pre- and post-treatment scores for cervical muscle tenderness were measured on the 20 subjects who underwent treatment for their TMD with the aim of determining whether changes in these symptoms paralleled any other observed changes in TMD severity that were associated with treatment.

Results

Interexaminer Reliability

The 3 subindexes of the TMI were all found to have good interexaminer reliability (an ICC of 0.75 or greater indicates good reliability[16]). Specifically, the subindex ICCs were 0.92 for the FI, 0.84 for the MI, and 0.75 for the JI. The overall score for the TMI, which is the unweighted combination of the 3 subindex scores, showed excellent agreement, with an ICC of 0.93 (95% confidence interval of 0.79 to 0.98). Further delineation of the JI into TMJ noises and TMJ pain on palpation yielded ICCs of 0.69 and 0.79, respectively. The components of the MI extraoral and intraoral muscle pain on palpation were observed to have ICCs of 0.91 and 0.80, respectively. The TMI, FI, MI, and JI mean scores (and their SD) for examiners 1 and 2, respectively, were as follows: for the TMI, 0.27 (0.19) and 0.26 (0.20); for the FI, 0.35 (0.21) and 0.32 (0.21); for the MI, 0.38 (0.30) and 0.34 (0.32); and for the JI, 0.10 (0.12) and 0.14 (0.17).

Criterion Validity

The agreement between the TMI and the CMI for the measurement of TMD severity was highly significant, with an ICC = 0.97 ($P < .001$). The mean CMI score over 12 subjects was 0.26 (SD 0.19). The mean TMI score was 0.26 (SD 0.18).

Construct Validity (Symptomatic Versus Asymptomatic Subjects)

The composite TMI scores for the 3 diagnostic groups of TMD patients did not differ statistically ($P > .05$). The mean (and SD) scores for the MFP, DDWR, and DDWOR groups were 0.46 (0.21), 0.47 (0.23), and 0.50 (0.22), respectively (Table 1). All 3 of these mean scores were significantly different ($P < .001$) from the mean TMI score of 0.08 (0.10) for the ASY group. Similarly, the FI scores for all 3 diagnostic groups were not statistically different ($P > .10$): 0.45 (0.12) for the MFP group, 0.41 (0.11) for the DDWR group, and 0.43 (0.14) for the DDWOR group. However, all were significantly different ($P < .001$) from the FI score of 0.10 (0.05) for the ASY group. As shown in Table 1, the JI mean scores tended to be greater for the DDWR and DDWOR groups (0.45 and 0.44, respectively) than for the MFP group (0.34), although no statistical difference was found between these means ($P > .10$). All 3 had similar

Table 1 Temporomandibular Index and Subindex Mean Values (SD) for 3 Diagnostic Groups of Temporomandibular Disorders and the Asymptomatic Comparison Group

Index/Subindex	ASY (n = 20)	MFP (n = 30)	DDWR (n = 29)	DDWOR (n = 20)
Function Index	0.10 (0.05)	0.45 (0.12)	0.41 (0.11)	0.43 (0.14)
Muscle Index	0.12 (0.14)	0.58 (0.25)	0.55 (0.27)	0.63 (0.26)
Joint Index	0.02 (0.06)	0.34 (0.28)	0.45 (0.30)	0.44 (0.26)
Temporomandibular Index	0.08 (0.10)	0.46 (0.21)	0.47 (0.23)	0.50 (0.22)

ASY = asymptomatic group; MFP = myofascial pain group; DDWR = disc displacement with reduction group; DDWOR = disc displacement without reduction group.
The mean TMI scores for the MFP, DDWR, and DDWOR groups were all statistically different from the mean TMI score for the asymptomatic group ($P < .001$).

Table 2 Temporomandibular Index and the Associated Subindexes: Comparison of Pre- and Post-treatment Mean (SD) Values for 12 Subjects

Index/Subindex	Pretreatment	Post-treatment	Mean difference	*P* value*
Function Index	0.37 (0.13)	0.31 (0.13)	0.06	.02
Muscle Index	0.46 (0.18)	0.34 (0.21)	0.12	.007
Joint Index	0.39 (0.16)	0.21 (0.13)	0.18	.005
Temporomandibular Index	0.41 (0.10)	0.29 (0.11)	0.12	.0001

*Paired *t* test.

standard deviations of around 0.28, and all were significantly different ($P < .001$) from the ASY mean JI of 0.02 (0.06). The MI scores were also statistically similar ($P > .30$) for the MFP group (0.58), the DDWR group (0.55), and the DDWOR group (0.63), with their standard deviations grouped around 0.26. All 3 were significantly different ($P < .001$) from the MI score of 0.12 (0.14) for the ASY group.

Construct Validity (Pretreatment Versus Post-treatment TMI Scores)

The pre- and post-treatment scores of the TMI and its subindexes were evaluated as to their change over time (Table 2). The TMI demonstrated significant change ($P < .001$) from pre- to post-treatment, with the mean score dropping from 0.41 (0.10) to 0.29 (0.11). The subindexes reflected similar changes, but the change in the FI showed only borderline statistical significance ($P < .02$). Changes in cervical muscle symptoms were not observed to mimic any of the other parameters for determining the relative severity of TMD. As the mean scores for the TMI, FI, JI, and MI all decreased, the cervical muscle scores actually increased on average over the treatment period, from a mean of 0.25 to a mean of 0.35 ($P > .10$).

Discussion

The TMI and its 3 subindexes, the FI, MI, and JI, all demonstrated good interexaminer reliability levels in this study. The ICC of 0.93 for the TMI had a 95% confidence interval of 0.79 to 0.98. Based on numerous reliability and calibration studies using these same examiners and similar clinical protocols, the intraexaminer reliability level has consistently exceeded interexaminer reliability. This observation is corroborated in a study[6] in which the intraexaminer reliability was slightly higher than the interexaminer reliability, despite the fact that fewer than half as many subjects were examined. Although intraexaminer reliability was not evaluated in the present study, there is no rea-

son to anticipate that it would have been inferior to the interexaminer reliability that is reported.

The TMI also satisfies the requirements of validity. The face validity of the TMI is unchallenged because its examination protocol has already met with the acceptance of leading researchers and clinicians in the field of TMD. The content validity of the TMI is similarly upheld, since this instrument is consistent with the current understanding of TMD.[17] When developing the CMI, Fricton and Schiffman[6,7] evaluated the literature as well as numerous examination forms used by experts in the field of TMD. Based on a comprehensive list of important TMD-related signs, examination items were selected for the CMI. These same examination items were later employed for the RDC/TMD and the TMI and included *(1)* TMJ noise; *(2)* joint and muscle pain upon palpation; and *(3)* pain, deviation, or limitation associated with mandibular ROM.

Criterion validity requires that an index agree with an accepted standard that measures the same condition. The clinical examination is the best current means of measuring TMD severity, and criterion validity for the TMI was evaluated by comparing it to the original CMI. As reported above, the CMI and the TMI showed excellent agreement, with an ICC of 0.97. This inter-instrument concordance between the parent CMI and the TMI is very comparable to the ICC of 0.95 reported for interexaminer reliability between the same 2 examiners when the CMI was used alone.[6]

Construct validity requires that an index behave according to some predetermined intuitive concept. This aspect of validity was established by assessing 2 constructs, the first of which involved differentiation of asymptomatic subjects from symptomatic subjects representing 3 diagnostic groups of TMD. Intuitively, it makes sense that asymptomatic subjects would have negligible TMD signs and, therefore, a low TMI score, whereas symptomatic subjects would have a variety of TMD signs, with comparatively higher TMI scores. The second construct was based on the sensitivity of pre- and post-treatment scores to reflect changes in severity associated with treatment. For an outcome measurement to be useful in clinical trials, it is essential that its items would change over time in response to a change in the subject's status.

In testing the first construct, the mean TMI score for the asymptomatic subjects was 0.08 (0.10), which is comparable to the CMI mean of 0.07 (0.08) reported for asymptomatic subjects by Fricton and Schiffman and the CMI mean of 0.10 (0.10) reported by Schiffman and colleagues for normal subjects.[6,7,18] Although there was no statistical difference in TMI scores among the 3 diagnostic groups (MFP, DDWR, and DDWOR), all 3 of these groups did differ significantly from the asymptomatic group. The symptomatic groups were all heterogeneous groups. Most of the MFP subjects also had arthralgia, and all of the disc displacement subjects had MFP and arthralgia. This corresponds to what one finds clinically. The fact that the 3 symptomatic groups had similar TMI, JI, FI, and MI scores agrees with the findings of Schiffman and colleagues,[19] who observed that a measure of severity alone cannot discriminate between intra-articular and muscle disorders in TMD.

The construct that the index should change over time with treatment was also confirmed, with a statistically significant mean change of 0.12 for the composite TMI over 6 weeks of treatment. As for the subindexes of the TMI, the changes in the JI and MI were significant, but the change in the FI was only borderline significant ($P < .02$). The items with the least amount of change were found to be jaw opening patterns and the excursive movements that are part of the FI. To observe greater change in these variables, a longer treatment period or treatments that differ from those used in this study, may be required.

The TMI underwent a number of revisions with respect to its parent CMI to conform more precisely to the RDC/TMD. Cervical muscle palpations that had been included in the CMI were dropped from the TMI because they have been held to be questionable markers for a TMD condition and yield poor content validity. This issue was nonetheless revisited in the present study by performing the cervical muscle palpations on each of the 20 symptomatic subjects who were evaluated pre- and post-treatment. Their cervical muscle tenderness was observed to show a mean increase over the treatment period. Although this increase was not statistically significant, its direction was at odds with the other parameters, all of which indicated a decrease in severity. Thus, cervical muscle palpations may not be informative for the assessment of TMD severity.

Non-reproducible clicking was not included in the scoring of the TMI for this study because it also is not listed as one of the RDC/TMD diagnostic criteria. However, it is defined in the operational definitions in the RDC/TMD, and it has been shown[20] to be useful in the diagnosis of

TMD. Thus, it was retained as an optional non-scored item on the TMI examination form.

As noted above, the RDC/TMD guidelines call for the use of a 4-point scale for measuring muscle tenderness to palpation (none, mild, moderate, and severe), although a simple presence or absence of pain is employed for determining an RDC diagnosis of TMD. Some have proposed that a 4-point scale would have better sensitivity and specificity and thus be better able to detect changes in severity over time.[21] During the assessment of criterion validity, the examiner performing the original CMI consistently finished the examination in 5 to 10 minutes, while the examiner performing the TMI needed 10 to 15 minutes to complete the examination. This was due, in large part, to the considerable amount of time required for a subject to rate as mild, moderate, or severe any pain that was felt during palpation. Practically speaking, the usefulness of an index is based not only on adequate reliability and validity. Consideration must also be given to its ease of use, including the time needed to perform the examination protocol. In outcome studies and epidemiologic studies, multiple examinations may need to be performed. If an index is too time-consuming, it becomes impractical. A 2-point scale was clearly adequate for detecting differences within this study. Unless the additional benefit of a 4-point scale can be demonstrated, a 2-point scale should be used due to the decreased time needed for the examination.

All examination items in the TMI are given equal weight in contributing to the overall index score. The question arises as to whether all measures should be considered equally, or if certain items should be differentially weighted, as is the case for the Helkimo Clinical Dysfunction Index.[4,5] There is no conclusive evidence to support the assertion that any single TMD examination item is more predictive or diagnostic than another within each subindex (ie, temporalis tenderness versus masseter tenderness) or between subindexes. Since the purpose of the TMI is to give the clinician/researcher an idea of the severity of TMD signs, an equal weighting of individual items within the subindex scores seems reasonable, as does the decision for each of the 3 subindexes of the TMI to contribute equally to the total TMI score.

The use of summary scores for determining the reliability of muscle palpations has been criticized, since this approach evaluates muscles as a group rather than individually. Summary scores also contribute little to a diagnosis involving a single painful muscle or a subset of muscles.[3] This concern is valid only if the diagnosis and treatment