The Research Diagnostic Criteria for Temporomandibular Disorders. I: Overview and Methodology for Assessment of Validity

Eric L. Schiffman, DDS, MS
Associate Professor
Department of Diagnostic and
    Biological Sciences
University of Minnesota
Minneapolis, Minnesota

Edmond L. Truelove, DDS, MSD
Professor
Department of Oral Medicine
University of Washington
Seattle, Washington

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

Gary C. Anderson, DDS, MS
Associate Professor
Department of Diagnostic and
    Biological Sciences
University of Minnesota
Minneapolis, Minnesota

Mike T. John, DDS, MPH, PhD
Associate Professor
Department of Diagnostic and
    Biological Sciences
University of Minnesota
Minneapolis, Minnesota

Thomas List, DDS, Odont Dr
Professor
Department of Stomatognathic
    Physiology
Malmö University
Malmö, Sweden

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

Correspondence to:
Dr Eric L. Schiffman
University of Minnesota School of
    Dentistry
6-320 Moos Tower
515 Delaware Street SE
Minneapolis, MN 55455
Fax: 612-626-0138
Email: schif001@umn.edu

Aims: The purpose of the Research Diagnostic Criteria for Temporomandibular Disorders (RDC/TMD) Validation Project was to assess the diagnostic validity of this examination protocol. The aim of this article is to provide an overview of the project's methodology, descriptive statistics, and data for the study participant sample. This article also details the development of reliable methods to establish the reference standards for assessing criterion validity of the Axis I RDC/TMD diagnoses. **Methods:** *The Axis I reference standards were based on the consensus of two criterion examiners independently performing a comprehensive history, clinical examination, and evaluation of imaging. Intersite reliability was assessed annually for criterion examiners and radiologists. Criterion examination reliability was also assessed within study sites.* **Results:** *Study participant demographics were comparable to those of participants in previous studies using the RDC/TMD. Diagnostic agreement of the criterion examiners with each other and with the consensus-based reference standards was excellent with all kappas ≥ 0.81, except for osteoarthrosis (moderate agreement, κ = 0.53). Intrasite criterion examiner agreement with reference standards was excellent (κ ≥ 0.95). Intersite reliability of the radiologists for detecting computed tomography-disclosed osteoarthrosis and magnetic resonance imaging-disclosed disc displacement was good to excellent (κ = 0.71 and 0.84, respectively).* **Conclusion:** *The Validation Project study population was appropriate for assessing the reliability and validity of the RDC/TMD Axis I and II. The reference standards used to assess the validity of Axis I TMD were based on reliable and clinically credible methods.* J OROFAC PAIN 2010;24:7–24

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)[1] specifies a dual-axis diagnostic system for temporomandibular disorders (TMD) supported by a well-operationalized history and examination protocol. The Axis I clinical assessment protocol is designed to render TMD diagnoses, and the Axis II screening instruments assess psychological status and pain-related disability. Together, Axis I and Axis II assessments constitute a comprehensive evaluation consistent with the biopsychosocial health model.[2]

Advancement in our understanding of the prevalence, etiologies, natural progression, and treatment of TMD is dependent on having reliable and valid diagnostic criteria for these disorders. The 1996 US National Institutes of Health Technology Assessment

Conference Statement on the Diagnosis and Management of Temporomandibular Disorders noted that an ideal diagnostic classification system should be based on etiology.[3] However, the RDC/TMD Axis I diagnostic protocol, which is based on measurement of signs and symptoms, is the best and most used classification system to date for the epidemiological studies that are needed to understand TMD etiology and mechanisms.[4]

The current RDC/TMD taxonomic system was not intended to be an end product. Ongoing efforts to investigate its validity were anticipated and encouraged when the RDC/TMD was first established in 1992.[1] To date, no comprehensive investigation of the Axis I diagnostic reliability and validity has been reported. Axis II instrument reliability has been demonstrated,[5] but the validity of the Axis II screening instruments for assessing psychological status and pain-related disability in TMD cases has not been adequately demonstrated. Thus, a comprehensive evaluation of RDC/TMD reliability and validity was needed.

Numerous publications have suggested aspects of the RDC/TMD that could be improved to more effectively distinguish TMD cases from controls and differentiate diagnostic subgroups.[6–20] The first aim of the current RDC/TMD Validation Project was to rigorously establish the reliability and validity of the RDC/TMD diagnostic protocol in its published form. The second aim was to propose modifications for the protocol that would improve its reliability and validity as a taxonomic system.

This article is one of six consecutive articles describing the RDC/TMD Validation Project.

- The current article presents an overview of the entire project with an emphasis on methods and reliability of the Axis I reference standards.
- The second article presents the reliability of the original RDC/TMD examination protocol for rendering eight Axis I diagnoses, and discusses selected examination items from which these diagnoses are derived.[21]
- The third article presents the validity of the eight Axis I diagnoses assessed against the reference standard diagnoses.[22]
- The fourth article presents the evaluation of the psychometric properties of the Axis II psychological status and pain-related disability assessment instruments.[23]
- The fifth article presents proposed revised diagnostic algorithms for the eight Axis I diagnoses and the validity and reliability of the revised algorithms. The revised algorithms are based on

new reliable test items that improve the diagnostic validity of the taxonomic system.[24]
- The sixth article presents a proposal for further revision of Axis I, in terms of diagnostic nomenclature and expansion of the scope of the diagnoses, and additional domains of assessment for Axis II. This article also identifies the need for input and discussion from the international TMD community regarding the future of RDC/TMD research.[25]

The purpose of the current article is to present *(1)* an overview of the study methods used in the series of articles for assessment of reliability and validity of the RDC/TMD Axis I and II as a taxonomic system when used to distinguish TMD subtypes from each other and from normal controls; *(2)* descriptive data for the RDC/TMD Validation Project study sample; *(3)* the procedures used for establishing credible Axis I reference standard diagnoses; and *(4)* clinical examiner and radiologist reliability data supporting the Axis I reference standards.

Materials and Methods

Nomenclature

The Standards for Reporting of Diagnostic Accuracy (STARD) nomenclature[26] was used to allow for clarity in the reporting. The terms and their definitions as they pertain to this study are:

- Target condition: Disease or other condition (disorder) that may prompt further diagnostic testing, or initiation, modification, or termination of treatment. In the context of this project, the target condition is TMD or, as termed by the US National Institute of Dental and Craniofacial Research (NIDCR), temporomandibular muscle and joint disorders (TMJD).
- Test: Any method used to obtain diagnostic information relevant to a patient's health status.
- Index test: Test being evaluated (ie, Axis I and II of the RDC/TMD).
- Reference standard: Best available method for establishing the presence or absence of target condition in order to evaluate the criterion validity of an index test. The reference standard is commonly referred to as the "gold standard."

The RDC/TMD Validation Project assessed the *criterion validity* of the index tests that include the original RDC/TMD Axis I diagnostic algorithms

and the Axis II psychological status and pain-related disability instruments. Criterion validity is the measure of the validity of an index test when assessed against a credible reference standard. For assessment of Axis I criterion validity, the reference standard diagnoses were based on the consensus of two criterion examiners (CEs) at each site. The CEs were TMD experts with between 12 and 38 years experience who independently rendered their TMD diagnoses using the criterion examination protocol that was considerably more comprehensive than that specified by the original RDC/TMD protocol. The elements of this comprehensive examination are discussed below.

Study Setting and Locations

The evaluation of the RDC/TMD protocol for Axis I was a multisite collaboration among researchers at the University of Minnesota (UM), the University of Washington (UW), and the University at Buffalo (UB). The study took place at research centers at each of these institutions.

Study Participants

Participant recruitment. Beginning in August 2003, study participants were consecutively recruited until three fourths (approximately 550) of the study sample had been enrolled. At this point, it was necessary to institute selective recruitment in order to fill the recruitment goals for the less common TMD diagnoses, including certain of the disc displacement and arthrosis cases. Other subgroups of participants requiring selective recruitment were older age categories for normal participants and TMD pain cases needed for completing Axis II studies. Selective recruitment was continued until study closure in September 2006. Participants were drawn from two sources: direct referrals from local health care providers to the respective university-based TMD centers (ie, clinic cases) and responses to community advertisements (ie, community controls and cases). Thus, the study sample was a convenience sample that was recruited from both clinic and community sources.

Inclusion and exclusion criteria. Recruitment was designed to include cases with a full spectrum of TMD signs and symptoms. Participants, ages 18 to 70 years old, entered the study as putative TMD cases or controls based on the inclusion and exclusion criteria listed in Table 1. The inclusion criteria for study eligibility differed from the published RDC/TMD diagnostic criteria by assigning putative case status to individuals who reported a minimum

Table 1	Inclusion and Exclusion Criteria

Inclusion criteria
1) Inclusion criteria for TMD cases:
   Participant reports or presents with at least one of the three cardinal signs or symptoms of TMD: Jaw pain, limited mouth opening, or TMJ noise.
2) Inclusion criteria for controls:
  I. History
     a. No lifetime history of TMD symptoms ("supercontrols")
        1. Absence of TMJ noise, locking or catching of the jaw, and
        2. Absence of pain in the jaw or the temporal area, and
        3. Absence of headaches affected by jaw movement, function, or parafunction.
     b. Prior history of TMD symptoms ("controls")
        1. In the last 6 months, no history of TMD symptoms
        2. Prior to 6 months ago:
           a. No more than five isolated episodes of TMJ noise, with each episode lasting less than 1 day and not associated with jaw pain or limited mouth opening, and
           b. No more than one to two isolated episodes of locking or catching of the jaw in the wide-open mouth position, and
           c. No headaches in the temporal area affected by jaw movement, function, or parafunction.
  II. Clinical examination (including assesment for warmth, swelling, and redness of the tissue)
     a. Any pain produced by procedures must be nonfamiliar, and
     b. No TMJ clicking, popping, or snapping noises with more than one movement, and
     c. No coarse crepitus with any movement.
  III. Imaging
     a. TMJ MRI is negative for disc displacements, and
     b. TMJ CT is negative for osteoarthrosis.

Exclusion criteria for cases and controls
  I. History
     a. Systemic rheumatic, neurologic/neuropathic, endocrine, or immune/autoimmune diseases or wide spread pain. (Exception: participants with medical documentation of rheumatoid arthritis or fibromyalgia).
     b. Radiation treatment to head and neck.
     c. TMJ surgery.
     d. Trauma to jaw in the last 2 months (exclusion regardless of time: Jaw trauma from auto accident).
     e. Presence of non-TMD orofacial pain disorders.
     f. Pregnancy.
     g. Unable to participate due to language barrier or mental/intellectual incompetence.
     h. Use of narcotic pain medication, muscle relaxants or steroid therapy unless discontinued for 1 week prior to examination.
     i. Use of antidepressant drugs unless the participant has been on a stable dose for 60 days.
     j. Use of prescription or over-the-counter nonsteroidal anti-inflammatory medications unless the medication(s) were discontinued for 3 days prior to the examination (use of acetaminophen was allowed as a rescue drug).
     k. Drug abuse.
     l. Ongoing dental treatments.
     m. Wearing dentures.
     n. Contraindications for imaging.
     o. Ongoing TMD treatments unless on a stable regimen for at least 2 months.
     p. Unable or unwilling to give informed consent.
  II. Clinical examination
     a. Presence of non-TMD orofacial pain disorders.
  III. Imaging
     a. MRI is positive for pathology (exception for cases: TMJ disc displacements).
     b. CT is positive for osseous pathology (exception for cases: TMJ osteoarthrosis).
     c. Panoramic radiograph is positive for osseous (non-TMJ related) or odontogenic lesions.

of one of the three cardinal symptoms of TMD: jaw pain, limited mouth opening, or temporomandibular joint (TMJ) noise. Participants who denied currently having any of these symptoms were enrolled as controls.

Institutional Review Board (IRB) oversight. IRB approval was obtained at each of the three study sites prior to initiating this project. Participants were compensated $200 for their participation in the Axis I and II clinical assessment, $25 for participation in Axis I and Axis II questionnaire test-retest reliability substudies, $75 for participation in the Axis II criterion substudy, and $50 to $200 for participation in examiner reliability substudies. Patient informed consent was obtained for all aspects of the project at each study site.

Methods for the Axis I and an Overview of the Axis II Procedures

Axis I Methods

Sample size requirements. Sample size requirements stipulated *a priori* for the sensitivity and specificity estimates in this project 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 halfwidth for each confidence interval is 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. 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 cases were required for each diagnosis. Each TMD case could potentially present with up to five TMD diagnoses: a Group I muscle diagnosis, a Group II disc displacement diagnosis for each of two TMJs, and a Group III diagnosis of arthralgia, arthritis, or arthrosis for each of two TMJs. Recruitment of 600 cases was expected to provide a minimum of 100 TMD diagnoses representing each of the eight original RDC/TMD subdiagnoses. In addition, the project planned for an additional 100 participants with minimal symptoms who would be subclinical with respect to the RDC/TMD diagnostic protocol, but who could qualify as TMD cases based on the consensus of the CEs. Finally, it also planned to recruit 100 controls, that is, participants with no current signs

or symptoms of TMD who represented four age strata: 18 to 30, 31 to 40, 41 to 50, and 51 to 70 years of age. This stratification allowed for selection of a "pool" of controls that, at the time of analysis, could match the age distribution of participants in each of the eight TMD subgroups. Given the study sample design above, a total of 800 participants was the initial estimated requirement for the study. Further description of the recruitment objectives resulting from this design is presented in the third article in this series.[22]

Tests and Measures

Participant demographic data and baseline measures. Demographic measures of the study population included gender, age, race, education level, and income.[1] Baseline measures to describe the clinical characteristics of the study participants included characteristic pain intensity,[1,27] duration of pain,[1] depression,[1,28,29] nonspecific physical symptoms,[1,28,29] Graded Chronic Pain Scale (GCPS) scores,[1,27] and the number of RDC/TMD Axis I diagnoses present for each case.

Axis I index test. One of the index tests to be validated by this project was the published RDC/TMD Axis I diagnostic examination procedure that employs a set of standardized clinical and questionnaire items. Each of the clinical measurements has been well defined with operational criteria[1] and allows for assignment of TMD participants to any of three diagnostic groups that include eight subdiagnoses:

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

Axis I Reference Standard

Development. It was required that tests included as part of the reference standards derived from the criterion examination protocol would be simple, reliable, easy to perform, and appropriate for the research setting. Potential Axis I diagnostic tests were drawn from *(1)* recommendations in the 1992 RDC/TMD monograph[1]; *(2)* conclusions from other research published since 1992; *(3)* tests

Table 2	Axis I and II Measures	
	Axis I	Axis II
RDC/TMD History Questionnaire		Graded Chronic Pain Scale
Medical History Inventory		Symptom Checklist 90R
Supplemental History and Supplemental History Follow-up Questionnaires		Jaw Functional Limitation Scale
RDC/TMD clinical examination protocol		Multidimensional Pain Inventory
Joint sounds assessed during superior joint loading		SF-12 Health Survey
Joint sounds assessed with stethoscope		Perceived Stress Scale – 10
Myofascial palpation		McGill Pain Questionnaire
Modified joint palpation		Oral Behaviors Checklist
Algometry		Pittsburgh Sleep Quality Index
Clench test		Explanatory Model Scale
Bite stick test		Mandibular Function Impairment Questionnaire
Joint-play tests; traction, translation, and compression		State-trait Anxiety Inventory
Orthopedic tests: Static and dynamic		Center for Epidemiologic Studies – Depression Scale
Soft/hard end–feel		General Health Questionnaire
Occlusal assessment		
Radiologic examination: Panoramic radiograph, bilateral TMJ CT and MRI		

recommended by the study's External Advisory Panel (AP) composed of clinical and research specialists appointed by the NIDCR; and *(4)* suggestions solicited from members of TMD organizations, including the American Academy of Orofacial Pain. From these recommendations, a list of candidate history questions and examination tests was developed to be considered by the AP. Some proposed tests were ruled out by the AP as being beyond the scope of this study. Such tests included electronic diagnostic systems for assessing joint vibration to potentially detect disc displacements and osteoarthrosis. The AP-vetted diagnostic tests were then operationalized and tested for reliability. The final list of Axis I procedures constituting the criterion examination that was performed by the CEs is shown in Table 2.

Operationalization of Axis I criterion history data collection. The criterion history data collection included the published RDC/TMD History Questionnaire[1] along with the Supplemental History Questionnaires that were developed and used by the CEs as part of their semistructured participant interview. This Supplemental History Questionnaire consisted of 61 questions assessing multiple dimensions of pain in the jaw muscles, TMJ, ear, and temple including whether the pain was changed with jaw movement, function, parafunction and/or rest. It also assessed tension-type headache by using operationalized International Headache Society criteria,[30] and history of joint noise, jaw locking, and perceived occlusal change. To measure changes in these variables occurring between study visits, a Supplemental History Follow-up Questionnaire was also developed for

use at the second CE visit. These supplemental questionnaires will be described and evaluated in a future publication that will include estimates of their test-retest reliability and their capacity to predict the reference standard diagnoses.

Operationalization of Axis I criterion clinical examination. The criterion examination protocol included all the measures as operationalized in the RDC/TMD. These measures were performed according to the published RDC/TMD specifications.[1] In addition, the criterion examination was composed of several previously described examination procedures, including joint-play tests (ie, traction, translation, and compression),[31–33] static and dynamic tests,[31–34] soft and hard end-feel,[35] algometry,[36] bite test with unilateral and bilateral placement of cotton rolls,[35,37] and a 1-minute clench.[38] New tests for the criterion protocol were the myofascial palpation test and the modified joint palpation test. The myofascial palpation test performed at the RDC/TMD-specified muscle sites in the masseter and temporalis used a range of 2 to 4 pounds of pressure rather than the 2 pounds specified by the RDC/TMD examination protocol for muscle palpation. The examiner used the spadelike pad of one finger to apply this pressure to the surface of the muscle while moving the finger back and forth across the long axis of the muscle fibers. This palpation technique was maintained for no more than 5 seconds. To locate areas associated with potential pain referral, the examiner: *(1)* placed the muscle on a slight stretch; *(2)* located so-called "taut bands" in the temporalis and masseter muscles by palpating across or along the long axis of the muscle fiber; *(3)* slid the finger

across the muscle fibers or along the muscle fibers (with muscle slightly stretched); and/or *(4)* asked the subject to clench his/her back teeth together while the area of greatest muscle bulk during the contraction was examined. The modified joint palpation test for evaluating joint pain was as follows: the examiner requested the participant to "Open slightly so your teeth are not touching." The examiner then located the lateral pole of the TMJ and, keeping an edge of the palpating finger on the lateral pole of the participant's TMJ, the examiner orbited his/her finger around the lateral pole and used a range of 2 to 3 pounds of pressure with a target of at least 2 pounds. A range of palpation pressure was used for this latter test because, like the myofascial technique, it required motion while applying the pressure and the authors' collective experience was that it is not always possible to apply an exact pressure. Joint loading with opening[31] and a stethoscope were additional methods for assessing joint noise that were used to supplement the published RDC/TMD auscultation method. The participants' report of examination-induced joint noise was also recorded. If the participant reported distinct sounds such as clicking, popping or snapping sounds, these were recorded as a "click" and longer duration sounds including crunching, grinding or gratings sounds were recorded as "crepitus." If any examination test elicited a report of pain, or if pain occurred with clicking noises, then the participant was asked if this pain was a "familiar pain," that is pain similar to or like what he or she had been experiencing from the target condition outside the examination setting. Participants with a report of pain were also asked to indicate if the pain was referred and, if so, at what other site it was felt. The occlusal assessment included recording the number of teeth, overbite, crossbite, and midline discrepancy.[39,40] Occlusal intercuspal contacts were assessed using Shim stock (Almore International) in maximum intercuspal position (MIP).[41] Centric relation position (CR), and CR to MIP slides were assessed.[42]

Imaging of participants included a panoramic radiograph, bilateral TMJ magnetic resonance imaging (MRI), and bilateral TMJ computed tomography (CT) scans. Details of the image analysis criteria used by the radiologists to identify MRI-disclosed disc displacements and CT-disclosed osteoarthrosis are briefly noted in Table 1 and described in detail elsewhere.[43] Briefly, the criteria for osteoarthritis/osteoarthrosis was the presence of deformation due to subcortical cyst, surface erosion, osteophyte, or generalized sclerosis.

Osseous flattening and/or subcortical sclerosis were considered indeterminate for these diagnoses. The criteria in the sagittal plane for a normal disc position in the closed mouth position was that the border between the low signal of the disc and the high signal of the retrodiscal tissue was located between the 11:30 and 12:30 o'clock positions and the intermediate zone was located between the condyle and the articular eminence. For the closed mouth position, a diagnosis of disc displacement was rendered when these two criteria were not met. In the open mouth position, to be normal, the intermediate zone was located between the condyle and eminence. Persistent disc displacement was diagnosed when the intermediate zone was anterior to the superior aspect of the condyle.

Establishing the reference standard. The CEs used questionnaires and a semistructured interview to review the medical history and pain characteristics in order to rule out possible non-TMD pain conditions and to exclude individuals with comorbid conditions (see exclusion criterion in Table 1). Participants reporting a history consistent with migraine were not excluded. However, if a participant presented for evaluation while having an active migraine headache, the subject was rescheduled at a later date for the clinical examination. In addition, panoramic radiography and a clinical examination, including assessment for warmth, swelling, and redness of the tissue, were used to rule out odontogenic, soft tissue, and hard tissue pathology. Other pathology not targeted for inclusion in the project was ruled out with TMJ MRI and CT. In establishing the reference standard diagnoses, the CEs considered self-report of pain in the last month; effect of jaw function, movement, parafunction, and rest on the reported pain over the past month; replication of the reported pain on provocation using clinical tests (see Table 2); and the TMJ CT and MRI studies. The CEs also considered both common and uncommon TMD conditions that were operationalized by the consensus of the CEs (see Table 3).

The CEs performed their evaluations within the following procedural framework. Each of two CEs interviewed and examined each participant blinded to each other's findings. Using all available clinical information including the imaging studies with the radiologist's interpretations, they independently rendered their criterion diagnoses. They then compared their findings and, if either CE differed with the other's findings or diagnoses, the participant was reexamined by both of them to resolve the area of disagreement. If either CE disagreed with the radiologist's interpretation, the radiologist was

consulted for further review of the images with the CEs. The reference standard diagnoses were then established by consensus between the CEs. The study's requirement of a consensus between two independent examiners was designed to reduce the likelihood of diagnostic error. The estimated absolute error associated with a single examination is reported in the Results section.

Training and expertise of the examiners. A total of nine clinicians served as the examiners for the Axis I validation study, including two CEs and one dental hygienist (test examiner; TE) at each study site. All six of the CEs were specialists in TMD and orofacial pain dentistry; the CEs had between 12 and 38 years of experience in research and clinical management of TMD. The three dental hygienists who served as the TEs were trained and calibrated to perform the RDC/TMD examination protocol. The radiologists at the UM and UW were diplomates of the American Board of Oral and Maxillofacial Radiology and the radiologist at UB was a diplomate of the American Board of Radiology and Neuroradiology; the radiologists had between 12 and 23 years of experience at interpreting TMJ images.

Data collection design. Based on STARD terminology,[26] the data collection for this project was *prospective* in that all history, examination, and imaging data collections were planned before the index test (RDC/TMD procedures) and the criterion examination procedures for the reference standard were performed.

Identical data collection protocols were performed at each study site (Fig 1). Participants who met initial screening criteria, as assessed by the study coordinator using a structured interview, were scheduled for Visit 1. They were asked to complete the baseline self-report instruments 1 day prior to their first appointment. The *baseline data collection instruments* included the RDC/TMD History Questionnaire,[1] Medical History Inventory, and Supplemental History Questionnaire (Table 2).

- **Study participant screening.** The telephone screening process was standardized across the three study sites as a questionnaire/interview composed of 31 questions, 19 of which had multiple response categories. This screening instrument is to be posted on the web site of the International RDC-TMD Consortium Network.[4] The rationale for this extensive screening process was to ensure that participants who were invited to present for Visit 1 were likely to be eligible for accession to the study as either a case or a control.

Table 3   CE-Expanded TMD Taxonomy

I. Muscle Disorders
- I.a.   Myofascial Pain
- I.a.1   Myofascial Pain with Referral
- I.b.   Myofascial Pain with Limited Opening
- I.b.1   Myofascial Pain with Referral with Limited Opening
- I.c.   Muscle Spasm
- I.d.   Myositis
- I.e.   Muscle Contracture

II. Disc Displacements
- II.a.   Disc Displacement with Reduction
- II.a.1   Disc Displacement with Reduction with Intermittent Limited Opening
- II.b.   Disc Displacement without Reduction with Limited Opening
- II.c.   Disc Displacement without Reduction without Limited Opening

III. Arthralgia, Arthritis, Arthrosis
- III.a.   Arthralgia (Synovitis)
- III.a.1   Capsular Ligamentous Sprain
- III.b   Osteoarthritis
- III.c.   Osteoarthrosis

IV. Uncommon TMD
- IV.a.   Dislocation
- IV.b.   Fibrous Ankylosis
- IV.b.1   Bony Ankylosis
- IV.c.   Pigmented Villonodular Synovitis
- IV.d.   Chondromatosis
- IV.e.   Idiopathic Condylar Lysis
- IV.f.   Subcondylar Fracture
- IV.g.   Neoplasms

V. Rheumatic Disorders
- V.a.   Rheumatoid Arthritis with potential concurrent involvement of the TMJ
- V.b.   Juvenile Rheumatoid Arthritis with potential concurrent involvement of TMJ
- V.c.   Miscellaneous rheumatic diseases with potential concurrent TMJ involvement
- V.d.   Fibromyalgia with potential concurrent involvement of masticatory muscles

VI. Tension-Type Headaches
- VI.a.   Infrequent episodic tension-type headache associated with the temporalis muscle
- VI.b.   Frequent episodic tension-type headache associated with the temporalis muscle
- VI.c.   Chronic tension-type headache associated with the temporalis muscle

- **Visit 1.** The two CEs rotated between successive participants at the first appointment for the initial assessment of each participant. They explained the study, obtained informed consent, and reviewed the participant's medical history, particularly with reference to the exclusion criteria. The examiner fulfilling this function is referred to as CE-1 in the text that follows. A panoramic radiograph was obtained, and interpreted by the study radiologist to rule out dental and osseous diseases unrelated to the TMJ. After establishing the participant's eligibility, CE-1 did
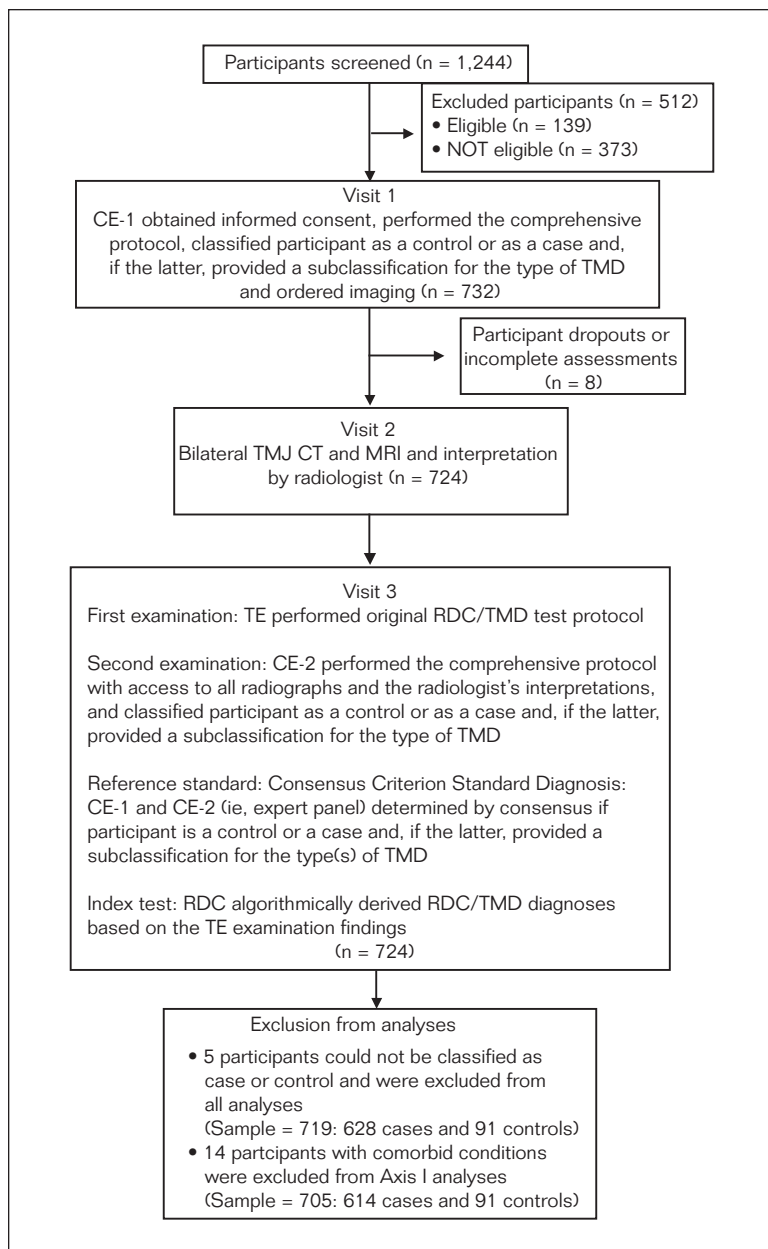
Fig 1 Flow chart for participants in the validation project assessing the reliability and validity of the RDC/TMD Axis I TMD clinical diagnoses.

a complete TMD history by using the RDC/TMD History and the Supplemental History Questionnaires to guide a semistructured interview. CE-1 then completed the criterion clinical assessment protocol as previously described (Table 2). At the end of this assessment, CE-1 classified the participant as a control or as a case with a subclassification for the types of TMD (Table 3).

- **Visit 2.** Bilateral TMJ MRI and CT images were obtained using standardized acquisition protocols for all study participants. The study radiologists at each site interpreted all images.
- **Visit 3.** Visit 3 was scheduled typically within 14 days after Visit 1. Participants were asked to

complete follow-up questionnaires 1 day before Visit 3. The Supplemental History Follow-up Questionnaire was used to assess for any changes in symptoms between Visit 1 and 3. Visit 3 consisted of four components. First, the TE, an RDC/TMD-trained and -calibrated dental hygienist, conducted the published RDC/TMD examination protocol while blinded to the findings of the CE-1 examination and all radiological assessments as well as to the participants' responses to the RDC/TMD History questions numbered 3, 14a, and 14b.[1] Second, the second criterion examiner (CE-2), blinded to findings of both the CE-1 and TE then repeated the criterion assessment. After recording the appropriate diag-

noses based only on examination findings, the CE-2 reviewed the panoramic radiographs and bilateral TMJ MRI and CT scans and updated the diagnoses if necessary. The CE-2 then reviewed the radiologist's interpretation of the images and recorded his/her final diagnoses. Third, with the participant still present, reference standard diagnoses were established based on a consensus between CE-1 and CE-2, using all available clinical information to classify the participant as a control or as a TMD case along with the subtype(s) of TMD. Fourth, one of the CEs then debriefed the participant.

The *index test*, ie, the algorithmically derived RDC/TMD diagnoses based on the TE examination findings, and the *reference standard*, ie, the consensus diagnoses rendered by the two CEs, were both performed on the same day. The index test examination was always completed before the reference standard diagnosis was established.

Assessment of diagnostic agreement for criterion data collections. Criterion examiner reliability: Beginning at baseline and over the course of the project, three sessions were planned for which a single CE from each study site came to UM for assessment of criterion examination diagnostic reliability. Each examiner performed the same criterion protocol on each study participant prior to all three examiners coming together to render a consensus diagnosis. This study design allowed for an overall estimate of diagnostic agreement between the individual criterion examination diagnoses and the consensus-based reference standard. It also provided an estimate of interexaminer reliability by comparing the individual criterion exam findings across the three examiners. Twenty-six participants were assessed over these three sessions that were programmed to occur after one of the annual calibration exercises, as described in the second paper in this series.[21]

In addition, within each study site, assessment of diagnostic agreement between the criterion examination and the reference standard was made possible because, for all study participants, the CE-2 criterion examination and the reference standard consensus were performed the same day.

Radiologist reliability. At baseline and on a yearly basis over the course of the study, four exercises were planned for the assessment of the reliability of the study radiologists.[43] Calibration of the radiologists from the three sites began with their review and discussion of a representative sample of panoramic radiographs, TMJ CT and TMJ MRI scans showing all osseous characteristics from normal to frank osteoarthrosis. In addition, MRI was used for demonstrating normal disc position, disc displacement with reduction, and disc displacement without reduction as well as effusions. For reliability assessment, each radiologist viewed panoramic radiographs; representative axially corrected coronal and sagittal slices from CT; and open- and closed-mouth sagittal views of PD-MRI and T2-MRI. For the initial reliability study, the images were collected from prior studies or teaching files from the three research locations. For the three subsequent annual reliability studies, the images used were from the participants in the current project that were selected by one of the UM radiologists to represent all the intra-articular disorders. The selected images represented the full scope of possible diagnoses presented in random order. Each of the radiologists interpreted panoramic radiographs, CT and MRI scans, blinded to each other's findings and the clinical data. The images were scored according to the criteria developed for RDC/TMD Validation Project. For the initial reliability assessment, 59 joints seen on panoramic radiographs, 70 joints on CT, and 70 joints on MRI scans were used to assess for osteoarthritis; and 68 joints on MRI scans for disc position. For the subsequent reliability studies, 20 panoramic radiographs, 25 joints on CT scans in the closed mouth position, and 25 joints on MRI scans in the closed and open mouth positions were selected to represent all the intra-articular disorders. These CT, MRI, and panoramic radiographs were grouped as sets, but a given set did not represent the same participant. All responses on the data collection forms were categorical.

Test-retest reliability of diagnostic questions. Among all the questionnaires employed in this project, only three questions were used as required determinants for Axis I diagnoses. All three were part of the published RDC/TMD History Questionnaire.[1] These were: Question number 3, "Have you had pain in the face, jaw, temple, in front of the ear, or in the ear in the past month?"; Question number 14a, "Have you ever had your jaw lock or catch so that it won't open all the way?"; and Question number 14b, "Was this limitation in jaw opening severe enough to interfere with your ability to eat?" Test-retest reliability assessment of the RDC/TMD History Questionnaire and the Supplemental History Questionnaire was performed on a subset of participants who participated in Axis I assessment at UB and UW. Reliability results for only Questions 3, 14a, and 14b are reported in this paper.

Statistical procedures. Proc Freq (SAS Institute) was employed to compute percent agreement between examiners. Kappa (k) was specified as the primary measure of reliability of diagnostic renderings. Kappa was also the primary measure for estimating diagnostic agreement between the criterion exam protocol and the reference standards. These estimates were computed by techniques using generalized estimate equations (GEE) and based on a procedure described by Williamson et al.[44] These GEE procedures provided adjustment for side-to-side correlation within participants for diagnostic renderings.

Reliability for the radiograph interpretations was computed using simple kappa, because there was no issue of correlated data in these data sets. The radiographs and scans employed for all radiology calibration exercises were either the right or left sided TMJ for any given participant, but not both sides. Stata statistical software was employed to obtain these estimates across the three examiners.[45]

Axis II Methods Overview

Three separate studies were performed for assessing Axis II of the RDC/TMD. All Axis II measures are listed in Table 2. Briefly, these studies addressed the following:

- At all three study sites, self-report questionnaires were administered for assessment of pain, mood, pain-related disability, health-related disability, stress reactivity, sleep, and behaviors.[27–29,46–57] The Axis II data collection measures were specifically selected to characterize convergent and discriminant validity of the published RDC/ TMD Axis II screening measures. In addition, the Axis II studies were designed to assess incremental increases in validity associated with expanding the domains of assessment specified for the RDC/TMD Axis II screeners. Finally, the selection of measures was structured to allow generalizability of this study's findings to other studies using the same or similar measures.
- At UB and UW, structured psychiatric interviews, self-administration of personality disorders assessment, and mental status testing were also performed using validated instruments. These tests served as reference standards for Axis II depression and nonspecific physical symptoms against which the findings from the RDC/TMD Axis II screening instruments were compared.
- At UB and UW, assessment of temporal stability of the self-report instruments was also performed.

For the entire evaluation of the RDC/TMD Axis II instruments, two study psychologists supervised the biobehavioral data collection and trained the psychometrists. The detailed methods used in these three studies and the Axis II validity results for the published RDC/TMD protocol are presented in the fourth article in this series.[23] Future papers will report on the other self-report measures, particularly as they relate to potentially expanding the domains for the RDC/TMD Axis II assessment.

Results

Study Participants

Over the three study sites, a total of 1,244 potential participants were screened. Of the 512 potential participants who did not enter the study, 373 were not eligible for the following reasons: current use of excluded medications or recreational drugs (79), failure to meet selection criteria at the time when selective recruitment was initiated in order to fulfill diagnostic recruitment goals of 100 of each TMD subgroup diagnosis (64), failure to meet the initial screening criteria (seven questions) for potential cases or controls (63), excluded medical conditions (40), inability to undergo MRI due to body metal (23), non-TMD orofacial pain disorder (21), dentures (18), ongoing litigation for jaw condition (14), ongoing TMD or dental treatments (12), ineligible age (10), medical history exclusion including TMJ surgery (8), trauma to jaw in last 2 months (8), pregnancy (7), and language barrier (6). One hundred and thirty-nine potential participants were eligible but did not enter the study with the primary reasons being no time or time conflict (48), they changed their mind (35), they did not present for a scheduled visit (28), or they did not want to have imaging done (including patients with claustrophobia) (28). A total of 732 participants were enrolled and 724 completed the study, with 8 dropouts or incomplete assessments (Fig 1). Of these 724 participants, there was insufficient evidence to classify 5 of them as either case or control and they were excluded from the analysis. The remaining 719 participants included 628 TMD cases and 91 controls. Fourteen of these 628 cases were subsequently excluded from the Axis I analyses due to the presence of chondromatosis (n = 2), reported fibromyalgia (n = 9), or reported rheumatoid arthritis (n = 3). (Participants with a documented medical diagnosis of fibromylagia or rheumatoid arthritis were eligible for the study.) Chondromatosis was excluded based on suspicion

Table 4   Axis I: Distribution of TMD Cases and Controls

Totals representing all study sites
Study sample: 724 participants (628 cases, 91 controls, and
5 unclassifiable)
Initial case count: 628
Case exclusions
  Systemic/other pathology: 14
  Final case count for Axis I: 614

Diagnoses Appropriate for the Axis I Analyses*			
RDC/TMD diagnoses	Cases	Joints	Diagnoses
Group I			
Ia Myofascial pain	210		210
Ib Myofascial pain with limited opening	285		285
Group II			
IIa Disc displacement with reduction		532	532
IIb Disc displacement without reduction with limited opening		91	91
IIc Disc displacement without reduction without limited opening		275	275
Group III			
IIIa Arthralgia		466	466
IIIb Osteoarthritis		223	223
IIIc Osteoarthrosis		120	120
Diagnostic counts	495	1,707	2,202

*The total number of diagnoses is greater than the number of cases because an individual can have more than one diagnosis. Individual participants averaged 2,202/614 = 3.6 diagnoses/case.

of the presence of the disorder as detected on MRI by the radiologist. Thus, a total of 614 cases remained for the Axis I analysis; these participants presented with a total 2,202 diagnoses, or an average of 3.6 diagnoses per case (Table 4). The Axis II analyses included all 628 cases, excluding only those with insufficient evidence to be classified as case or control. The 91 controls had no signs of TMD and had a negative current history, examination, and imaging (MRI, CT, and panoramic radiograph) findings. Of these 91 controls, 80 had no lifetime history of TMD symptoms (ie, "supercontrols") and 11 of the controls had no current history (within the past 6 months), but had a prior history of symptoms potentially consistent with TMD (see inclusion criteria in Table 1). Of the 614 TMD cases used for the Axis I analyses, 24% were direct referrals from local health-care providers to the university-based TMD clinics at the three sites (clinic cases), and 76% were respondents to study flyers and advertisements (community cases). Figure 2 is a Venn diagram presenting the distribution of 614 TMD cases with Group I Muscle Disorders; Group II Disc Displacements; and Group III Arthralgia, Arthritis, Arthrosis, based on the CE consensus diagnoses.
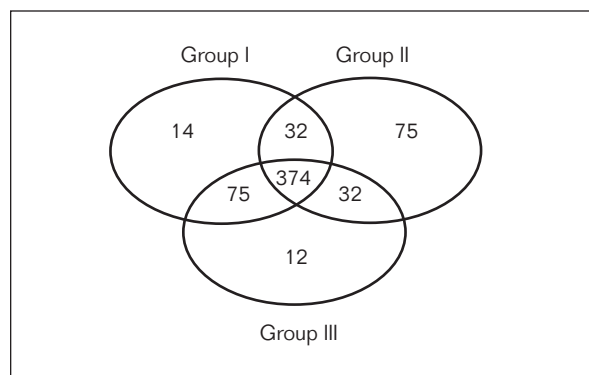

Fig 2 Diagnostic distributions of the cases relative to Group I Muscle Disorders; Group II Disc Displacements; and Group III Arthralgia, Arthritis, Arthrosis.

Study population demographic and clinical characteristics. Table 5 summarizes the participant demographic variables including gender, age, race, education level, and income, and also the Axis II clinical characteristics including characteristic pain intensity,[1] duration of pain,[1] depression,[1,28,29] nonspecific physical symptoms,[1,28,29] graded chronic pain severity,[27] and number of RDC/TMD diagnoses.[1]

Table 5 Participant Demographic and Clinical Characteristics

	Controls	Cases without pain		Cases with pain	
	Community controls (n = 91)	Community cases (n = 107)	Clinic cases (n = 7)	Community cases (n = 359)	Clinic cases (n = 141)
Gender					
Male	37%	21%	43%	14%	10%
Female	63%	79%	57%	86%	90%
Age (years)					
Mean ± SD	36 ± 13	39 ± 13	47 ± 14	36 ± 12	39 ± 15
Range	18-66	18-60	19-63	18-67	18-67
Race					
White	86%	91%	100%	94%	93%
Other	14%	9%	0%	6%	7%
Education					
No college	16%	16%	0%	14%	19%
≥ 1year of college	84%	84%	100%	86%	81%
Income					
Less than $50,000	70%	61%	57%	61%	41%
$50,000 – 79,999	21%	22%	14%	23%	26%
$80,000 or more	9%	17%	29%	16%	33%
Characteristic pain intensity (0 – 100)					
Mean ± SD	N/A	N/A	N/A	51 ± 20	54 ± 20
Range	N/A	N/A	N/A	7-100	10-100
Duration of pain (months)					
Mean ± SD	N/A	N/A	N/A	126 ±103	100 ±107
Range	N/A	N/A	N/A	7-480	9-480
Depression					
Missing	1%	0%	0%	0%	1%
Normal	81%	85%	72%	58%	65%
Moderate	14%	13%	14%	28%	23%
Severe	4%	2%	14%	14%	11%
Nonspecific physical symptoms					
Missing	1%	0%	0%	1%	1%
Normal	91%	90%	57%	45%	54%
Moderate	7%	9%	29%	34%	31%
Severe	1%	1%	14%	20%	14%
Graded chronic pain severity					
I (%)	N/A	N/A	N/A	44%	39%
II (%)	N/A	N/A	N/A	45%	44%
III–IV (%)	N/A	N/A	N/A	11%	17%
Number of RDC/TMD diagnoses (%/n)					
0 Diagnosis	100%/91	0%/0	0%/0	0%/0	0%/0
1 Diagnosis	0%/0	32%/34	14%/1	4%/13	1%/2
2 Diagnoses	0%/0	48%/51	43%/3	6%/23	6%/8
3 Diagnoses	0%/0	7%/8	29%/2	26%/94	19%/27
4 Diagnoses	0%/0	13%/14	14%/1	25%/90	33%/46
5 Diagnoses	0%/0	0%/0	0%/0	39%/139	41%/58

SD = standard deviation.

Adverse events. Only one adverse event occurred, when a participant's jaw locked closed during the examination. This condition was addressed at the time of the event. The participant was advised to return if this symptom reoccurred and she did not return.

Axis I CE Agreement

Intersite interexaminer reliability (n = 26) for the criterion examination was excellent (κ = 0.81 to 0.91) for seven of the eight RDC diagnoses; for osteoarthrosis (IIIc), reliability was good (κ = 0.59).

Table 6 Diagnostic Agreement Associated with the Criterion Examinations: Kappa and Percent Agreement

RDC/TMD diagnosis	Percent of study sample with diagnosis	Intersite criterion calibrations (n = 26)				Intrasite monitoring (n = 724)	
		Inter-CE agreement (kappa)	Inter-CE agreement (PA)	Overall CE agreement with consensus (kappa)	Overall CE agreement with consensus (PA)	CE agreement with consensus (kappa)	CE agreement with consensus (PA)
Group Ia: MFP	51	0.85 (0.68)	92	0.82 (0.64)	91	0.98 (0.91)	99
Group Ib: MFP with limited opening	27	0.81 (0.45)	92	0.87 (0.50)	95	0.98 (0.95)	99
Group IIa: DD with reduction	31	0.83 (0.58)	92	0.86 (0.64)	94	0.96 (0.92)	98
Group IIb: DD without reduction, with limited opening	10	0.84 (−0.04)	97	0.86 (0.10)	97	0.97 (0.75)	99
Group IIc: DD without reduction, without limited opening	33	0.86 (0.66)	94	0.88 (0.65)	95	0.98 (0.88)	99
Group IIIa: Arthralgia	29	0.91 (0.57)	96	0.94 (0.64)	97	0.95 (0.91)	98
Group IIIb: Osteoarthritis	32	0.82 (0.55)	92	0.90 (0.65)	96	0.95 (0.84)	99
Group IIIc: Osteoarthrosis	14	0.59 (0.11)	88	0.53 (−0.01)	90	0.96 (0.79)	99
Average PA	N/A	N/A	93.5	N/A	94.4	N/A	98.9
Absolute error	N/A	N/A	< 7%	N/A	< 6%	N/A	< 2%

With multiple diagnoses present in some participants, the percent of study participants with any given diagnosis is indicated in the second column.
In columns 3, 5, and 7, the kappa coefficient is followed in parenthesis by its lower 95% confidence interval. All upper confidence intervals were ≥0.99.
PA = the computation of percent agreement shown in columns 4, 6, and 8; absolute error is the complement of percent agreement (100 − PA);
MFP = myofascial pain; DD = disc displacement.

The percent agreement ranged from 88% to 97%, with an average percent agreement of 93.5 and an absolute error of less than 7% among the three CEs (Table 6). Absolute error, or percent disagreement, is the complement of percent agreement (PA), that is, 100% − PA.

The overall criterion examination agreement by the three examiners with the consensus diagnosis was excellent, with a range in kappa from 0.82 to 0.94, except for the diagnosis of osteoarthrosis (κ = 0.53) (Table 6). Given a sample size of just 26 participants, the study sample prevalence for osteoarthrosis was very low at 14%. The absolute error associated with a single examination is estimated as the average error for the three examiners relative to the consensus diagnoses, and was observed to be less than 6%. These data indicate that the findings of a single criterion examination agreed with the consensus rendering more than 94% of the time (Table 6).

Intrasite agreement between the second criterion examination and the consensus (n = 724) was very high, with a range of k from 0.95 to 0.98. Percent agreement was 98-99%, with an average of 98.9 and an absolute error at less than 2% (Table 6).

Radiologist reliability. Results reported here are overall agreement computed over the four different calibrations that were done during the study. The radiologists' interrater reliability for reading the CT-depicted hard tissues (osteoarthritis/osteoarthrosis) and MRI-depicted soft tissue (disc position) was good to excellent (κ = 0.71 and 0.84, respectively), and is reported separately.[43]

Test-retest reliability of diagnostic questions. For the published RDC/TMD History Questionnaire,[1] the test-retest reliability for Questions number 3, number 14a, and number 14b was excellent (κ = 0.84, 0.76, and 0.75, respectively).

Discussion

To improve reporting and comparisons between studies, standardized methodology was used for assessing diagnostic accuracy in conformance with STARD recommendations.[26] Testing diagnostic accuracy requires a credible reference standard to assess criterion validity. The credibility of the criterion examination protocol derives initially from the fact that it parallels what is done for comprehensive examinations in clinical practice. It also has content validity because experts in the field used the current knowledge base to develop it.

Reliability of CEs and Radiologists

The results in Table 6 provide further support for the credibility of the criterion examination protocol. The protocol was associated with high interexaminer agreement for the criterion examination and high agreement when the individual criterion diagnoses were compared with the reference standard for Axis I TMD clinical diagnoses. To the authors' knowledge, there is no comparison in the TMD literature between a CE and a reference standard. The two kappas for osteoarthrosis (0.53, 0.59) that were less than 0.75 (the level considered to be excellent agreement) were the result of the low sample prevalence of 14%. It has previously been shown that the magnitude of the reliability coefficients depends on the prevalence of the disorder.[58,59] The reliability of the radiologists' interpretation of the images at each site was assessed at four different times over the course of this project and, overall, was shown to be good to excellent for CT (hard tissue) and MRI (soft tissue), respectively. A detailed description of the results of these reliability studies is reported separately.[43]

Reference Standard for Pain Built on Established Procedures

The reference standard for pain used in the present project was established using what is known about TMD, in addition to paralleling what is done to diagnose other chronic pain problems. The diagnosis of arthralgia and myofascial pain included both the original test items (provocation tests) specified in the RDC/TMD as well as additional test items. These latter tests, vetted by the project's AP, are tests currently used in research and clinical practice.[31–38] If any of the provocation tests elicited a complaint of pain from the participant, the participant was requested to report whether the pain was familiar, that is, similar to or like the pain they experienced from the target condition. This methodology has been used successfully to establish reference standards for assessment of pain in other medical classification schemes.[60–68] The requirement of familiar pain endorsement helps to minimize false positive diagnoses for cases where the pain endorsement is more the result of the provocation test than related to a true pain disorder. It is well understood that provocation tests can provoke pain in controls as well as not previously experienced pain in cases. Finally, the use of two independent CEs for establishing the reference standard parallels what has been done to develop diagnostic criteria for fibromyalgia.[69] In that study,

the reference standard used for fibromyalgia, a musculoskeletal disorder, was based on a consensus diagnosis between two rheumatologists who independently assessed each participant with all available clinical data including a semistructured history and examination.

Reference Standard for Intra-articular Disorders Built on Validated Procedures

Establishing the reference standard for assessing the presence of intra-articular disorders is less complex than for that of pain, given the availability of sophisticated, noninvasive imaging techniques that do not alter the structure being examined. For assessment of soft and hard tissue intra-articular anatomy, MRI and CT, respectively, are standard clinical imaging techniques. The images in this project were obtained using protocols that were standardized between sites, with multiple views of the participant's TMJ for both MRI and CT. All images were also reviewed by both CEs. If there were a question with regard to the radiologist's findings, the two CEs and the radiologist reviewed the images together, with the radiologist rendering the final decision with regard to the interpretation of the images. This methodology was designed to minimize diagnostic misclassification.

Generalizability of the Estimates of Reliability and Validity

The study was designed to include a diverse participant population with a full spectrum and severity of TMD signs and symptoms, and Axis II characteristics that were consistent with literature reports of population-based,[70–75] and clinical studies.[5,76–84] In addition, controls were recruited with no lifetime history of TMD symptoms, or with a prior history of TMD symptoms dating 6 months or more before their examination, but with no current symptoms. This recruitment strategy allowed again for a *spectrum* of participants ranging from "supercontrols" with no lifetime history of TMD to controls with some past history of TMD-like pain. In the absence of well-defined criteria for normalcy in terms of TMD conditions, this approach for defining TMD controls is consistent with literature reports that used the absence of any RDC/TMD diagnosis[1] or the absence of any signs and symptoms included in the Helkimo Indices[85] to define a control.[86]

The authors have three reasons to believe sampling bias that could affect the study's estimates of diagnostic accuracy was minimal. First, sensitivity

and specificity estimates are theoretically independent of prevalence of the target conditions.[87] Second, the cases and the controls covered the spectrum of signs and symptoms observed with the presence or absence of TMD conditions. Third, sensitivity and specificity for diagnosing TMD pain or intra-articular disorders would not likely vary significantly based on the past history of the disorder, presence of comorbid conditions or other exclusion criteria. The authors also believe that the study sample of target conditions is likely to be representative of participants to whom the test will be applied in future research and clinical settings, which is the fundamental requirement of studies investigating diagnostic test accuracy.[88] This study was, however, limited to study population specifications recommended by STARD[89] as a first step for the validity testing of a diagnostic instrument and, as such, was not designed to provide sensitivity and specificity estimates in patients with comorbid conditions or other exclusions specified for this study.

Methods to Minimize Circularity in Validity Assessment

A critical issue in establishing a reference standard is to identify and address any potential for circularity. Circularity occurs when cases and controls are intentionally selected based on characteristics that the test protocol is specifically designed to detect. It also occurs if the reference standard too closely resembles the test protocol. If either of these conditions exists, the estimate of validity may be spuriously inflated. These issues were addressed in the present project by *(1)* inclusion of participants as cases that would not meet criteria for an RDC/TMD diagnosis; *(2)* a CE assessment protocol that contained all items stipulated by the RDC/TMD with the addition of independent diagnostic tests composed of additional history taking, exam procedures, and imaging including TMJ MRI and CT; *(3)* independent examination of participants by two examiners who then established consensus diagnoses as the reference standards; and *(4)* the use of an expanded reference standard taxonomy that was independent of the RDC/TMD and included disorders not specified by the RDC/TMD.

Limitations of the Study

The Axis I reference standards for this project could be in error for several reasons due to either the inherent variability in the clinical phenomena,

or systematic error in the examiners' measurements. Pain to palpation of the TMJ capsule is inherently variable, and this measurement is critical for determining a diagnostic subgroup. Systematic error can occur if the examiner knows the participant's questionnaire responses resulting in a diagnostic suspicion bias that can influence the outcome of the diagnostic process.[90] Finally, all provocation tests can potentially result in pain, even in pain-free controls. Thus, there was a clear need to verify the clinical relevance of examination-induced pain by determining if it was familiar to the participant as the pain complaint and could be verified by the two CEs.

Conclusions

Advancement in our understanding of the prevalence, etiologies, natural progression, and treatment of TMD is dependent on having reliable and valid diagnostic criteria. In studies of diagnostic accuracy, a reference standard is required to differentiate cases with the target condition from controls, and to assess the criterion validity of the index test. The primary goal of this article was to describe in detail the project's methods used for establishing reference standard diagnoses for assessing the validity of Axis I measures of the RDC/TMD. The Axis I criterion procedures that were developed have content validity and acceptable reliability. It is concluded that this methodology constituted a credible reference standard for assessment of Axis I diagnostic validity, and for revision of the published RDC/TMD Axis I diagnostic scheme. Furthermore, the study participants' demographics and clinical characteristics were appropriate for assessing the validity of the RDC/TMD. Finally, for RDC/TMD Axis II biobehavioral instruments, assessment of criterion, convergence, and concurrent validity was performed using previously validated reference standards.

Acknowledgments

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. Dworkin SF, von Korff MR, LeResche L. Epidemiologic studies of chronic pain: A dynamic-ecologic perspective. Ann Behav Med 1992;14:3–11.

3. Proceedings of the National Institutes of Health Technology Assessment Conference on Management of Temporomandibular Disorders, Bethesda, Maryland, April 29–May 1, 1996. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 1992;83:49–183.

4. The International RDC-TMD Consortium Network website. http://rdc-tmdinternational.org. Accessed 2007.

5. Dworkin SF, Sherman J, Mancl L, Ohrbach R, LeResche L, Truelove E. Reliability, validity, and clinical utility of the research diagnostic criteria for temporomandibular disorders Axis II scales: Depression, nonspecific physical symptoms, and graded chronic pain. J Orofac Pain 2002;16:207–220.

6. 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.

7. 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.

8. 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.

9. Shaefer JR, Jackson DL, Schiffman EL, Anderson QN. Pressure-pain thresholds and MRI effusions in TMJ arthralgia. J Dent Res 2001;80:1935–1939.

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. 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.

12. 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.

13. Huddleston Slater JJ, Lobbezoo F, Naeije M. Mandibular movement characteristics of an anterior disc displacement with reduction. J Orofac Pain 2002;16:135–142.

14. 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.

15. Manfredini D, Guarda-Nardini L. Agreement between research diagnostic criteria for temporomandibular disorders and magentic resonance diagnoses of temporomandibular disc displacement in a patient population. Int J Oral Maxillofac Surg 2008;37:612–616.

16. 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.

17. 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.

18. 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.

19. 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.

20. Schmitter M, Kress B, Leckel M, Henschel V, Ohlmann B, Rammelsberg P. Validity of temporomandibular disorder examination procedures for assessment of temporomandibular joint status. Am J Ortho Dentofac Orthopedics 2008;133:796–803.

21. Look JO, John MT, 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.

22. Truelove E, Pan W, Look J, et al. The research diagnostic criteria for temporomandibular disorders. III: Validity of axis I diagnoses. J Orofac Pain 2010;24:35–47.

23. Ohrbach R, Turner JA, Sherman JJ, et al. The research diagnostic criteria for temporomandibular disorders. IV: Evaluation of psychometric properties of the axis II measure. J Orofac Pain 2010;24:48–62.

24. Schiffman EL, Ohrbach R, Truelove EL, et al. The research diagnostic criteria for temporomandibular disorders. V. Methods used to establish and validate revised axis I diagnostic algorithms. J Orofac Pain 2010;24:63–78.

25. Anderson GC, Gonzalez YM, Ohrbach R, et al. The research diagnostic criteria for temporomandibular disordes. VI. Future directions. J Orofac Pain 2010;24:79–88.

26. Bossuyt PM, Reitsma JB, Bruns DE, et al. Standards for reporting of diagnostic accuracy. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. Clin Chem 2003;49:1–6.

27. VonKorff M, Ormel J, Keefe FJ, Dworkin SF. Grading the severity of chronic pain. Pain 1992;50:133–149.

28. Derogatis L. SCL-90-R: Symptom checklist-90-R. Administration, scoring and procedures manual. Psychopharmacol Bull 1994;9:12–28.

29. Derogatis LR, Lipman RS, Covi L. SCL-90: An outpatient psychiatric rating scale. Preliminary report. 1973;9:13–28.

30. 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.

31. Steenks, MH, deWijer A, Lobbezoo-Scholte AM, Bosman F. Orthopedic diagnostic tests for temporomandibular and cervical spine disorders. In: Fricton J, Dubner R (eds). Advances in Pain Research and Therapy: Orofacial Pain and Temporomandibular Disorders. New York: Raven Press, 1995.

32. 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.

33. 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.

34. 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.

35. Okeson JP. History and examination for temporomandibular disorders. In: Management of Temporomandibular Disorders and Occlusion. St. Louis, MO: Mosby, 1993.

36. 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.

37. Howard J. Clinical diagnosis of temporomandibular joint derangements. In: Moffett BC (ed). Diagnosis of Internal Derangements of the Temporomandibular Joint. Seattle: University of Washington, 1984.

38. Wright EF. Manual of Temporomandibular Disorders. Ames, Iowa: Blackwell Munksgaard, 2005.

39. Fricton J, Kroening R, Hathaway KM. TMJ and Craniofacial Pain: Diagnosis and Management. St. Louis, MO: Ishiyaku EuroAmerica, 1988.

40. Schiffman E, Fricton J, Haley DP. The relationship of occlusion, parafunctional habits and recent life events to mandibular dysfunction in a non-patient population. J Oral Rehab 1992;19:201–223.

41. Anderson GC, Schulte JK, Aeppli DM. Reliability of the evaluation of occlusal contacts in the intercuspal position. J Prosthet Dent 1993;70:320–323.

42. Dawson PE. Determining centric relation. In: Dawson PE (ed). Functional Occlusion From TMJ to Smile Design. St Louis: Mosby Elsevier, 2007.

43. Ahmad M, Hollender L, Anderson Q, et al. Research diagnostic criteria for temporomandibular disorders (RDC/TMD): Development of image analysis criteria and examiner reliability for image analysis. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 2009;107:844–860.

44. Williamson JM, Lipsitz SR, Manatunga AK. Modeling kappa for measuring dependent categorical agreement data. Biostatistics 2000;1:191–202.

45. Stata Statistical Software: Release 10. College Station, TX, 2007.

46. Buysse DJ, Reynolds CFI, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research. Psychiatry Res 1989;28:385–396.

47. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Human Behav 1983;24:385–396.

48. Goldberg DP, Williams P. A User's Guide to the General Health Questionnaire. Windsor, England: Nelson Publishing ,1988.

49. Goldberg DP, Gater R, Sartorius NU, et al. The validity of two versions of the GHQ in the WHO study of mental illness in general health care. Psychol Med 1997;27:191–197.

50. Kerns RD, Turk DC, Rudy TE. The West Haven-Yale multidimensional pain inventory (WHYMPI). Pain 1985;23:345–356.

51. Melzack R. The McGill pain questionnaire: Major properties and scoring methods. Pain 1975;1:277–299.

52. Ohrbach R, Granger C, List T, Dworkin S. Pain-related functional limitation of the jaw: Preliminary development and validation of the jaw functional limitation scale. Community Dent Oral Epidemiol 2008;36:228–236.

53. Ohrbach R, Larsson P, List T. The jaw functional limitation scale: Development, reliability, and validity of 8-item and 20-item versions. J Orofacial Pain 2008;22:219–230.

54. Radloff LS. The CES-D scale: A self-report depression scale for research in the general population. App Psych Measurement 1977;1:385–401.

55. MPI Computer Program. Pittsburgh, PA: Rudy, 2005.

56. Stegenga B, Bont LG, deLeeuw R, Boering G. Assessment of mandibular functional impairment associated with temporomandibular joint osteoarthritis and internal derangement. J Orofac Pain 1993;7:183–195.

57. Ware JE, Kosinski M, Turner-Bowker DM, Gandek B. How to Score Version Two of the SF-12 Health Survey. Lincoln, RI: QulaityMetric, 2002.

58. Cicchetti DV, Feinstein AR. High agreement but low kappa: II. Resolving the paradoxes. J Clin Epidemiol 1990;43:551–558.

59. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol 1990;43:543–549.

60. 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.

61. 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.

62. McFadden JW. The stress lumbar discogram. Spine 1988;13:931–933.

63. 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.

64. 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.

65. 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.

66. Couppé C, Torelli P, Fuglsang-Frederiksen A, Andersen KV, Jensen R. Myofascial trigger points are very prevalent in patients with chronic tension-type headache: A double-blinded controlled study. Clin J Pain 2007;23:23–27.

67. 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.

68. 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.

69. 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.

70. Plesh O, Sinisi SE, Crawford PB, Gansky SA. Diagnoses based on the research diagnostic criteria for temporomandibular disorders in a biracial population of young women. J Orofac Pain 2005;19:65–75.

71. Rantala MA, Ahlberg J, Suvinen T, Savolainen A, Kononen M. Chronic myofascial pain, disk displacement with reduction and psychosocial factors in Finnish non-patients. Acta Odontol Scand 2004;62:293–297.

72. Storm C, Wanman A. A two-year follow-up study of temporomandibular disorders in a female Sami population: Validation of cases and controls as predicted by questionnaire. Acta Odontol Scand 2007;65:341–347.

73. Schmitter M, Balke Z, Hassel A, Ohlmann B, Rammelsberg P. The prevalence of myofascial pain and its association with occlusal factors in a threshold country non-patient population. Clin Oral Invest 2007;11:277–281.

74. Nilsson IM. Reliability, validity, incidence and impact of temporomandibular pain disorders in adolescents. Swed Dent J 2007;(183):7–86.

75. List T, Wahlund K, Wenneberg B, Dworkin SF. TMD in children and adolescents: Prevalence of pain, gender differences, and perceived treatment need. J Orofac Pain 1999;13:9–20.

76. Glaros A, Urban D, Locke J. Headache and temporomandibular disorders: Evidence for diagnostic and behavioural overlap. Cephalalgia 2007;27:542–549.

77. Manfredini D, Chiappe G, Bosco M. Research diagnostic criteria for temporomandibular disorders (RDC/TMD) axis I diagnoses in an Italian patient population. J Oral Rehabil 2006;33:551–558.

78. Reiter S, Eli I, Gavish A, Winocur E. Ethnic differences in temporomandibular disorders between Jewish and Arab populations in Israel according to RDC/TMD evaluation. J Orofac Pain 2006;20:36–42.

79. John MT, Reissmann D, Schierz O, Wassell RW. Oral health-related quality of life in patients with temporomandibular disorders. J Orofac Pain 2007;21:183–195.

80. List T, Dworkin SF. Comparing TMD diagnoses and clinical findings at Swedish and US TMD centers using research diagnostic criteria for temporomandibular disorders. J Orofac Pain 1996;10:240–253.

81. Yap AU, Dworkin SF, Chua EK, List T, Tan KB, Tan HH. Prevalence of temporomandibular disorders subtypes, psychologic distress, and psychosocial dysfunction in Asian patients. J Orofac Pain 2003;17:21–28.

82. Lee LT, Yeung RW, Wong MC, McMillan AS. Diagnostic sub-types, psychological distress and psychosocial dysfunction in southern Chinese people with temporomandibular disorders. J Oral Rehabil 2008;35:184–190.

83. Truelove E, Huggins KH, Mancl L, Dworkin SF. The efficacy of traditional, low-cost and nonsplint therapies for temporomandibular disorder: A randomized controlled trial. J Am Dent Assoc 2006;137:1099–1107.

84. Dworkin SF, Huggins KH, Wilson L, et al. A randomized clinical trial using research diagnostic criteria for temporomandibular disorders-axis II to target clinic cases for a tailored self-care TMD treatment program. J Orofac Pain 2002;16:48–63.

85. Helkimo M. Studies on function and dysfunction of the masticatory system. II. Index for anamnestic and clinical dysfunction and occlusal state. Swed Dent J 1974;67:101–121.

86. Reissmann DR, John MT, Schierz O, Wassell RW. Functional and psychosocial impact related to specific temporomandibular disorder diagnoses. J Dent 2007;35:643–650.

87. Fletcher RH, Fletcher SW, Wagner EH. Clinical Epidemiology: The Essentials. Baltimore: Wilkins & Wilkins, 1996.

88. Guyatt G, Sackett D, Haynes B. Evaluating diagnostic tests. In: Hynes BR, Sacket DL, Guyatt GH, Tugwell P (eds). Clinical Epidemiology: How to Do Clinical Practice Research. Philadelphia: Lippincott, Williams & Wilkins, 2006.

89. Bossuyt PM, Reitsma JB, Bruns DE, et al. The STARD statement for reporting studies of diagnostic accuracy: Explanation and elaboration. Ann Intern Med 2003;138:W1–W12.

90. Sackett DL. Bias in Analytic Research. J Chron Dis 1979;32:51–63.