



Published in final edited form as:

J Orofac Pain. 2010 ; 24(1): 48–62.

Research Diagnostic Criteria for Temporomandibular Disorders: Evaluation of Psychometric Properties of the Axis II Measures

Richard Ohrbach, DDS, PhD[Associate Professor],

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

Judith A. Turner, PhD[Professor],

Department of Psychiatry and Behavioral Sciences and Department of Rehabilitation Medicine, Box 356560, University of Washington School of Medicine, Seattle, WA 98195

Jeffrey J. Sherman, PhD[Clinical Assistant Professor],

Departments of Rehabilitation Medicine and Oral Medicine, Box 356370, University of Washington, Seattle, WA 98195

Lloyd A. Mancl, PhD[Research Associate Professor],

Dental Public Health Sciences, Health Sciences, D-583, Box 357475, University of Washington, Seattle, WA 98195

Edmond L. Truelove, DDS, MSD[Professor and Chair],

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

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

Division of TMD and Orofacial Pain, University of Minnesota School of Dentistry, 6-320 Moos Tower, 515 Delaware Street SE, Minneapolis, MN 55455

Samuel F. Dworkin, DDS, PhD[Professor Emeritus]

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

Abstract

AIMS—To evaluate the psychometric properties of the Research Diagnostic Criteria for Temporomandibular Disorders (RDC/TMD) biobehavioral (Axis II) screening instruments.

METHODS—Participants with Axis I TMD diagnoses ($n=626$) completed the Axis II instruments (Depression, Nonspecific Physical Symptoms, Graded Chronic Pain) and other instruments assessing psychological distress, pain, and disability at three study sites. Internal consistency, temporal stability, and convergent/discriminant validity of the Axis II measures were assessed. To assess criterion validity of Depression and Nonspecific Physical Symptoms instruments as screeners, 170 participants completed a structured psychiatric diagnostic interview.

RESULTS—The Axis II instruments showed very good-excellent internal consistency (Cronbach's $\alpha = 0.80 - 0.95$). Their convergent (correlation range 0.3–0.9) and discriminant (range 0.0–0.6) validity were generally supported, although Nonspecific Physical Symptoms was more strongly associated with depressive than with somatic symptoms. Temporal stability was high for characteristic pain intensity (Lin's correlation concordance coefficient [CCC] = 0.91), interference (CCC = 0.89), and chronic pain grade (weighted kappa = 0.87), and fair-good for

Depression and Nonspecific Physical Symptoms (CCC = 0.63 – 0.78). The Depression instrument normal vs moderate-severe cut-point was good at identifying current-year DSM-IV depression and dysthymia diagnoses (sensitivity 87%, specificity 53%). Nonspecific Physical Symptoms did not have high utility for detecting psychiatric disorders (sensitivity 86%, specificity 31%).

CONCLUSION—The Axis-II Depression and Graded Chronic Pain instruments have clinically relevant and acceptable psychometric properties for reliability and validity and utility as instruments for identifying TMD patients with high levels of distress, pain, and disability that can interfere with treatment response and course of Axis I disorders.

Keywords

RDC/TMD; biobehavioral; screening; sensitivity; specificity

Introduction

When first published in 1992, the Research Diagnostic Criteria for Temporomandibular Disorders (RDC/TMD) ¹ represented a paradigm shift in the evaluation and diagnosis of patients with TMD, a heterogeneous group of disorders with orofacial pain as the most salient symptom. In contrast to previous TMD diagnostic systems, which emphasized classification focused on physical findings, the RDC/TMD includes assessment of both clinical signs and symptoms (Axis I) and the biobehavioral domain (Axis II). Consistent with the biopsychosocial model, research findings indicate that clinical diagnosis alone is often insufficient to explain observed levels of pain and disability.^{2–5} Accordingly, Axis II instruments were selected to screen patients for psychological status (depression and nonspecific physical symptoms) and to classify patients into a “chronic pain grade” based on characteristic pain and activity interference levels from the Graded Chronic Pain Scale (GCPS) ⁶. These Axis II measures were intended to serve as screening instruments for the constructs of depression, somatization, and disability, given their relevance as risk factors for poor clinical outcomes.^{7–9} Patients with these characteristics could then be referred for psychological assessment and interventions to address psychosocial barriers to TMD recovery.

At the time that the criteria were established, Dworkin and LeResche ¹ emphasized the need for further research to evaluate the reliability, validity, and clinical utility of the RDC/TMD. A subsequent prospective cohort study supported the clinical utility of Axis II: patients with acute TMD who had persistent pain at a 6-month follow-up, as compared with those who did not have pain at 6 months, had higher initial scores on the Axis II instruments.¹⁰ Additional research demonstrated good internal consistency for the Depression, Nonspecific Physical Symptoms, and Chronic Pain Grade instruments, and concurrent validity for the Depression and Chronic Pain Grade instruments.¹¹ Dworkin et al.¹¹ also presented evidence for the clinical utility of each instrument, but again noted the need for further research on the reliability, validity, and clinical utility of the instruments, using different samples. A fourth Axis II instrument, a jaw functional status checklist, has already been evaluated psychometrically and revised ^{12,13} and will not be discussed here.

Psychiatric disorders such as major depression and generalized anxiety, as well as psychological distress, are common among patients who seek treatment for chronic TMD pain and may interfere with response to pain treatments.^{7,14} Although the RDC/TMD measures of depression and nonspecific physical symptoms presumably relate to mood and somatization disorders in the Diagnostic and Statistical Manual-IV (DSM-IV),¹⁵ research is needed to determine the ability of the Depression and Nonspecific Physical Symptoms instruments to identify TMD pain patients who have these psychiatric disorders. Because

these instruments might also be nonspecific indicators of any psychiatric disorder, it is also of interest to assess their ability to detect the presence of any psychiatric diagnosis. Psychiatric disorders have the potential to affect patient pain-related problems and response to treatment for many types of chronic pain.^{2,16} Although a current psychiatric disorder would be more directly relevant to a patient's current TMD problem, a history of psychiatric disorder could be associated with increased vulnerability to maladaptive cognitive, affective, and behavioral responses to pain.^{17–19}

The aim of this study was to comprehensively evaluate the psychometric properties of the Axis II instruments used to screen for psychological status and disability in TMD patients. Specifically, the study assessed the internal consistency, temporal stability, and convergent and discriminant validity of each Axis II instrument. For the Depression and Nonspecific Physical Symptoms instruments, the primary criterion validity analyses focused on the association of each instrument with its corresponding psychiatric diagnosis, as derived from a standardized interview, but associations with other psychiatric diagnoses were also examined. Current-year diagnoses were the primary criterion variables, but lifetime diagnoses were also examined, given their potential as a risk factor for poor outcomes.

Materials and Methods

Study Sample and Procedures

Overview—Data collections were carried out at three sites, University at Buffalo (UB), University of Minnesota (UM), and University of Washington (UW), as part of the RDC/TMD Validation Project to assess the reliability and validity of the RDC/TMD Axis I and II taxonomic system. This report focuses on the Axis II instruments completed by study participants who were determined on the basis of a reference-standard clinical evaluation to have an Axis I TMD diagnosis. Sub-study 1 was part of the larger study conducted at all three centers, whereas sub-studies 2 and 3 were conducted only at UB and UW because only those study sites, by initial design, included licensed clinical psychologists. For the larger study, participants aged 18–70 years were recruited via advertisements in clinical and community settings. The inclusion and exclusion criteria for participation in the larger study are reported by Schiffman et al.²⁰ Figure 1 illustrates the flow of participants through the three studies.

Study 1 (internal consistency and convergent and discriminant validity)—Study 1 participants (n = 626; 533 female [two of the original 628 participants were not included in the analysis sample due to missing SCL90R data]) at UB, UM, and UW completed the Axis II measures and other measures used to assess convergent and discriminant validity described below. In addition, the General Health Questionnaire-28 (GHQ-28)²¹ was administered in order to select participants for Studies 2 and 3, as described below. The GHQ-28 has demonstrated validity in screening for psychiatric disorders.^{21,22} It consists of four groups of seven questions assessing somatic symptoms, anxiety and insomnia, social dysfunction, and severe depression.

Study 2a and Study 2b (temporal stability)—The targeted sample sizes for Study 2a and 2b were 75 each. These sample sizes were determined to be sufficient if the assumption is true that the test-retest reliability of each measure is at least 0.80 and that an acceptable lower-bound of the 95% CI is at least 0.70. All Study 1 participants during a designated recruitment period at UB and UW with GHQ-28 scores ≥ 10 were invited to complete the Depression and Nonspecific Physical Symptoms instruments a second time two weeks after the first completion for Study 2a, and 74 participants provided data. The GHQ-28 cut-point ≥ 10 was selected to restrict the sample to individuals representative of the population to

which the study findings are intended to apply; that is, limiting the number of participants likely to have very low scores on the Depression and Nonspecific Physical Symptom instruments. This restriction also results in a conservative estimate of the temporal stability of the instruments because individuals with very low scores are more likely to have perfect agreement at both administrations.

To assess the temporal stability of the GCPS measures (Study 2b: pain intensity, activity interference, and chronic pain grade), all Study 1 participants during another designated recruitment period at UB and UW and regardless of GHQ-28 score were invited to complete the GCPS again three days after the first administration, and 74 participants provided data. The GCPS was embedded in the RDC/TMD Patient History which, along with the Supplemental Axis I Patient Questionnaires, was administered in Study 2b but not reported here.

The between-tests intervals of 14 days for Study 2a and 3 days for Study 2b are consistent with intervals used in comparable reliability studies and with periods of time over which the target states (e.g., pain, depressed mood) would not be expected to change (pain intensity can fluctuate substantially over brief periods of time whereas depressive symptoms tend to be more stable).^{23–26}

Study 3 (criterion validity)—Study 1 participants at UB and UW were invited to participate in Study 3 if they had a score on the GHQ-28 that indicated either low likelihood (scores ≤ 10 ; $n=79$) or high likelihood (scores > 17 ; $n=187$) of having a psychiatric disorder. To increase the number of participants meeting criteria for at least one of the two psychiatric diagnoses of primary interest (depressive disorder or somatization disorder), we attempted to enroll two study participants with high scores for every participant with a low score. Study 3 participants ($n = 170$) completed a structured psychiatric interview that yielded DSM-IV¹⁵ diagnostic information used as the criterion to assess the validity of the Depression and Nonspecific Physical Symptoms instruments. This strategy of using the GHQ-28 score for selective subject recruitment was chosen to limit the number of psychiatric interviews.

The three study protocols were approved by the Institutional Review Board overseeing each study site. All participants provided informed consent. Participants were compensated \$200 for Study 1, \$25 for Study 2a or 2b, and \$75 for Study 3.

Measures

RDC/TMD Axis II—The RDC/TMD Axis II includes measures from the GCPS^{6,27} and the Symptom CheckList-90 (SCL-90).²⁸ The GCPS assesses pain intensity and interference with daily activities. It has been validated and has exhibited good psychometric properties in a large population survey and in large samples of primary care patients with pain.^{6,27} On the GCPS, study participants rated on scales from 0 = “no pain” to 10 = “pain as bad as could be” their current pain and average and worst facial pain in the past six months. The mean of these three ratings, multiplied by 10, is the *characteristic pain intensity (CPI)* score.^{6,25,27} Participants also rated on scales from 0 = “no interference” to 10 = “unable to carry on any activities” the degree of facial pain interference with daily activities, recreational/social/family activities, and work/housework activities in the past six months. The mean of these three ratings, multiplied by 10, is the *pain-related activity interference score*.²⁷ The GCPS also assesses the number of days of significant activity limitation due to pain in the past six months. Based on all three variables, the GCPS can be used to classify individuals into chronic pain grades: 0 = no pain, I = low pain intensity and low pain-related disability, II = high pain intensity and low pain-related disability, III = moderate pain-related disability, and IV = severe pain-related disability.

The Axis II Depression instrument includes the 13 SCL-90 depression scale items plus 7 SCL-90 “additional items” intended to assess vegetative symptoms of depression. The 7 additional items were included in the Axis II Depression instrument due to their content validity as part of the DSM construct of depression.^{1,15} The Nonspecific Physical Symptoms instrument (see Discussion for rationale for renaming this instrument) consists of the 12 items in the SCL-90 somatization scale. These SCL-90 items for both Depression and Nonspecific Physical Symptoms are identical to the corresponding items in the SCL-90-Revised (SCL-90R).²⁹

Validity measures—The Center for Epidemiologic Studies-Depression instrument (CES-D)³⁰ was administered to assess the validity of the Axis II Depression instrument. The CES-D has demonstrated good internal consistency (0.85–0.90), temporal stability (4 weeks, $r = 0.67$), and validity (sensitivity 0.85, specificity 0.64, for predicting DSM depression in elderly adults).^{30–32} Among patients with chronic pain, the CES-D has been demonstrated to have good ability to identify those with depression diagnoses (e.g., sensitivity of 0.82, specificity of 0.73 for detecting DSM-IV major depression diagnosis from semi-structured interview), indicating that the somatic symptoms of depression and of pain do not confound the assessment of depressed mood^{33,34}; for example, the CES-D has a sensitivity of 0.98 for detecting DSM-IV major depression from semi-structured interview.²¹

In addition to the use of the GHQ for recruitment into Study 3, the 7-item Somatic Symptoms instrument from the GHQ-28³⁵ was used to assess the convergent validity of the Axis II Nonspecific Physical Symptoms instrument. Symptoms (e.g., pain in the head, tightness or pressure in the head, hot and cold spells, feeling ill, feeling run down and out of sorts) are rated in terms of the respondent’s experiences in the past few weeks relative to “how one usually feels.”

The Multidimensional Pain Inventory (MPI)^{36,37} is a 60-item questionnaire designed to assess pain patients’ cognitive, behavioral, and affective responses to pain. Internal consistency of the instrument subscales ranges from 0.73 to 0.90, while test-retest reliability ranges from 0.68 to 0.83.³⁸ The utility of the MPI has been demonstrated in samples of patients with various chronic pain syndromes.^{36,39–41} The utility of the MPI has been demonstrated in samples of patients with various chronic pain syndromes.⁴² The MPI affective distress, pain severity, general activity, and interference scales (as well as the MPI dysfunctional score, which is a composite index of pain and interference) were used to assess the convergent and discriminant validity of the Axis II measures.

The SF-12 version 2 (SF-12v2)⁴²: The SF-12 version 2 (SF-12v2) is a widely used health-related quality of life measure, with internal consistency of 0.77–0.80 and temporal stability of 0.76–0.89.^{43,44} The SF-12v2 includes physical component summary (PCS) and mental component summary (MCS) scales, which have a mean of 50 and a standard deviation of 10 in the general U.S. population. Low scores indicate poor health and high scores reflect well-being. At the outset of the study, 32 participants completed the SF-12; their scores were then converted to SF-12v2 scores using algorithms provided by the instrument developer.

The Computer version of the Diagnostic Interview Schedule-IV (CDIS-IV) was administered to assess the criterion validity of the Axis II Depression and Nonspecific Physical Symptoms instruments. The original National Institute of Mental Health Diagnostic Interview Schedule (DIS)⁴⁵ was a structured interview designed to be administered by lay interviewers to obtain reliable Diagnostic and Statistical Manual-III (DSM-III) psychiatric diagnoses.⁴⁶ Diagnoses obtained using a computerized version of the DIS compared to those obtained using the traditional interviewer-administered (non-computerized) DIS exhibit up to 100% sensitivity and 95% specificity, depending on the particular diagnosis.⁴⁷ The CDIS-

IV yields DSM-IV¹⁵ psychiatric diagnoses. The CDIS-IV interviews were conducted by a trained psychometrist under the supervision of a clinical psychologist and were conducted blind to the responses on the self-report instruments. The C-DIS presents one question at a time, and the verbal response from the subject is interpreted by the psychometrist and entered into the software program, which then determines the next question based on DSM criteria as implemented by the DIS structure. In order to assess the reliability of interpretation by the psychometrists and following previous procedures,⁴⁸ 45-minute samples from audiotapes of four interviews from UW site and five interviews from UB site were independently coded by the psychometrist at the other site. Percent agreement between the two psychometrists was 99.6% for the diagnostic items; a Kappa statistic could not be computed due to the format of the data as obtained from the interview structure.

psychiatric diagnoses: The CDIS-IV interviews were conducted by a trained psychometrist at each site, under the supervision of a clinical psychologist, and were conducted blind to the responses on the study instruments. Following previous procedures,⁴⁷ 45-minute samples from audiotapes of nine interviews (four from UW site and five from UB) were independently coded by the psychometrist at the other site. Percent agreement between the two psychometrists was 99.6% for the diagnostic items.

The CDIS-IV interviews were scored^{47,50,51} for diagnoses of somatization disorder, panic attack, agoraphobia, generalized anxiety disorder, posttraumatic stress disorder, major depressive episode, dysthymic disorder, manic episode, hypomanic episode, obsessive disorder, and compulsive disorder. The diagnoses were divided into those for which criteria were met in the current year (i.e., in the 12-month period prior to interview) and those for which criteria were met prior to that period (“lifetime”). For individuals with a “current-year” diagnosis, the C-DIS algorithm does not indicate whether they also met criteria for the diagnosis prior to the past year. Following published procedures,⁴⁷ we grouped diagnoses of a major depressive episode or dysthymic disorder into a single category we labeled as “depression” for purposes of assessing the criterion validity of the Axis II Depression instrument. Only two participants met criteria for a somatization disorder, consistent with the rarity of this disorder. To assess the criterion validity of the Axis II Nonspecific Physical Symptoms instrument, like previous investigators, we used a lower number of symptoms than required for a DSM-IV diagnosis. We used cut-offs of four or more symptoms for males and six or more symptoms for females out of the 38 DIS somatization disorder items [Somatic Symptom Index (SSI)]; individuals scoring above these cut-offs are more likely to seek medical care for physical problems and to report recent sick leave or restricted activity.^{51,52} Work by Katon et al.⁵³ also supports the utility of using these cut-offs, noting that many clinical and behavioral features of somatization (e.g., lifetime diagnoses of panic disorder and major depression, disability, medical utilization) are common in patients scoring above these cut-offs. Physicians view these patients as more frustrating than patients with lower levels of symptoms.

Statistical Procedures

Study 1 (internal consistency and convergent and discriminant validity)—

Inspection of the GCPS measures indicated n=111 participants with no current pain (report of no jaw pain in the past 30 days and “0” on the GCPS rating of current pain) who were excluded from Study 1 analyses involving CPI, interference, and chronic pain grade; excluding such individuals results in a more conservative estimate of the statistics. Number of disability days was also assessed for temporal stability, but because it is a single item it was not further assessed. Cronbach’s alpha coefficients⁵⁴ were calculated to assess the internal consistency of the Axis II Depression, Nonspecific Physical Symptoms, and CPI and activity interference scales. Convergent and discriminant validity were assessed by

examining associations of each Axis II measure with the validity measures, using Spearman rank correlation coefficient for chronic pain grade and Lin's correlation concordance coefficient (CCC)⁵⁵ for all other Axis II measures. The CCC, combines measures of precision and accuracy in its estimates. The CCC is scaled to have a range of -1 to 1, is often similar to the intraclass coefficient coefficient (ICC), but is typically closer to zero than is Pearson's correlation coefficient. To compute the CCC, raw scores were converted to z-scores (computed as the difference between the observed value and its mean divided by the standard deviation), which were adjusted for study site. Measures of similar and dissimilar constructs were specified a priori to examine convergent and discriminant validity; for example, we expected another depression measures to agree highly (convergent validity) and a physical activity measure to not agree (discriminant validity) with the Axis II Depression.

Study 2 (temporal stability)—Temporal stability was examined for the Axis II depression, nonspecific physical symptoms, CPI, and interference measures using Lin's CCC, and for chronic pain grade by weighted kappa analysis. All participants were retained in the analyses for CPI and chronic pain grade but individuals reporting no pain were excluded from analyses for interference because the presence of pain is implicit in the "interference" assessment. Some study participants completed the second assessment at longer intervals than requested. We selected upper limits for the interval length based on attempts to minimize subject loss while also limiting the impact of additional interval length on reliability estimates. Analyses were conducted only for the 60 Study 2a participants with test-retest intervals of 7–27 days for the Depression and Nonspecific Physical Symptoms instruments (39/60 participants completed the second assessment exactly at the requested 14 days) and for the 65 Study 2b participants with intervals of 2–7 days for the GCPS measures. This loss of subjects from the desired sample size of 75 for each study resulted in a decrease in the acceptable lower-bound 95% CI from 0.70 to approximately 0.67. See Table I for complete description of the study samples.

Study 3 (criterion validity)—We used the published RDC/TMD raw scale score cutpoints for categorizing study participants into "normal," "moderate," and "severe" groups on the Depression (normal: <0.535, moderate: 0.535 to <1.105, and severe: ≥1.105) and Nonspecific Physical Symptoms (normal: <0.5, moderate: 0.5 to <1.0, and severe: ≥ 1.0) instruments.¹ Sensitivity and specificity of these Depression and Nonspecific Physical Symptoms cut-points in identifying current-year and lifetime depression diagnosis or elevated SSI, respectively, were computed, adjusting for study site to control for any differences in study samples across sites. We used the same approach to examine the ability of these instruments to identify presence of any current-year and of any lifetime psychiatric diagnosis. Because the criterion validity of the Depression instrument was acceptable, two further properties were investigated. The first was to calculate the area under the receiver operating characteristic (ROC) curve (AUC) for the Depression instrument predicting current-year depression. The second was to assess the clinical utility of the Depression instrument by calculating the positive predictive value (PPV; the probability that an individual scoring above the cut-point will actually have the diagnosis) and the negative predictive value (NPV; the probability that an individual with a score below the cut-point will not have the diagnosis), using the normal versus moderate/severe cut-point and published prevalence estimates.

Level of statistical significance was set at $P < .05$ for all tests. Stata 9.2 software (Stata Corp, College Station, TX) was used for all analyses.

Results

Study Participants: Demographic and Clinical Characteristics

Across sites and studies, most participants were female and White, had completed at least some college, and met criteria for at least one RDC/TMD Axis I pain diagnosis (Table 1). The Study 1 sample differed significantly ($P < .05$) across sites in age, education, and proportion with an Axis I TMD pain diagnosis. Most Study 1 participants had Axis I diagnoses of temporomandibular joint (TMJ) and muscle pain, disc displacement, or both (Table 2). The Study 1 sample also differed ($P < .05$) across sites with respect to most of the other study measures as well (see Table 3). The differences across sites largely reflected a deliberate attempt at UM to enroll more young, healthy subjects so that they could also be enrolled in a subsequent sub-study, a greater number of clinical referrals at UW, and differing educational levels in the larger communities in the three cities. Moreover, the younger and healthier cohort at UM exhibited lower scores on most of the comparison measures, not surprisingly. Data were combined across sites for subsequent analyses; the convergent and discriminant validity analyses and the criterion validity analyses were adjusted for study site. Internal reliability analyses were compared individually for each site, and there were no marked differences in the coefficient alpha statistics.

Overall, the Study 1 sample had fairly low mean Depression and Nonspecific Physical Symptoms scores but with a suitable range (Table 3). Scores on each Study 1 measure ranged from the best to worst possible. Mean SF-12v2 PCS and MCS scores were similar to those in the general U.S. population.⁴² On average among Study 1 participants reporting current pain, there was a moderate level of characteristic pain intensity and a low level of interference (as reflected in the classification of only 13% as chronic pain grade III or IV).

Study 1 (Internal Consistency and Convergent and Discriminant Validity)

Internal consistency—Internal consistency for the measures ranged from very good (nonspecific physical symptoms, with or without the pain items; characteristic pain intensity) to excellent (depression, interference) (Table 4).

Convergent and discriminant validity—In general, the expected pattern of higher (convergent validity, shaded cells) and lower (discriminant validity, unshaded cells) associations between the Axis II measures and the validation measures was observed (Table 5). For example, the Axis II Depression instrument was highly correlated with the CES-D (CCC = 0.85). Its association with the SF-12v2 MCS was somewhat lower but still strong (CCC = -0.70), as would be expected given that the MCS reflects both depression and other types of psychological distress (e.g., anxiety). Discriminant validity of the Axis II Depression instrument was evidenced by substantially lower associations with measures of constructs other than depression (e.g., somatic symptoms, pain severity, MPI activity and interference, SF-12 PCS).

Convergent and discriminant validity of characteristic pain intensity was supported by a substantial association with MPI pain severity (CCC = 0.65) and smaller associations with measures of constructs other than pain (e.g., depression [CES-D], somatic symptoms). Such validity was supported for interference by substantial associations with the MPI interference (CCC = 0.52) and dysfunctional measures (CCC = 0.51), and by smaller associations with measures of other constructs (e.g., the SF-12 MCS, CES-D). Associations of the chronic pain grade with the convergent validity measures were lower (Spearman rho = 0.35 – 0.44) than those seen with the other Axis II measures, but still higher than the associations in the discriminant validity tests. The Nonspecific Physical Symptoms instrument, both with and without pain items, showed only a moderate association with the GHQ-28 Somatic

Symptoms instrument (CCC approximately 0.45) and a stronger association with the CES-D (CCC = 0.56). The 95% CIs, listed in Table 5, demonstrate appropriately narrow intervals for this sample size, and do not readily modify the pattern of higher associations for convergent measures and lower measures for discriminant measures.

Study 2 (Temporal Stability)

The temporal stability (2–7 days) was high for characteristic pain intensity (CCC = 0.91), interference (CCC = 0.89), and chronic pain grade (weighted kappa = 0.87) (Table 6). The temporal stability (7–27 days) was fair to good for depression and nonspecific physical symptoms (CCC = 0.63 – 0.78). The lower-bound 95% CI of 0.69 for Depression was on the margin of the minimally accepted value. Otherwise, the observed lower-bound 95% CI was greater than our expected minimal value for all measures except for Nonspecific Physical Symptoms and, not surprisingly given that it is a single item, Number of Disability Days.

Study 3 (Criterion Validity of Depression and Nonspecific Physical Symptoms)

The GHQ-28 screener, relative to the psychiatric classification, proved to have a high false-positive and false-negative rate (results not presented), reducing efficiency for our recruitment process but enhancing the robustness of our findings by eliminating selection bias for Study 3. Among the 170 participants who completed a structured psychiatric interview using the C-DIS, 29% met criteria for a DSM-IV diagnosis of depression/dysthymia in the current year (meaning the past 12 months), 12% met criteria for being positive on the SSI in the current year, 38% met criteria for at least one psychiatric diagnosis in the current year, and 68% met criteria for a lifetime psychiatric diagnosis. The sensitivity and specificity of various Axis II measure groupings in predicting psychiatric diagnoses were examined (Table 7). In predicting a current-year depression diagnosis, the low cut-point for the Depression instrument had moderately high sensitivity (87%) and low specificity (53%), whereas the high cut-point resulted in low sensitivity (56%) and high specificity (91%). The ability of the Depression instrument to discriminate between those with vs without any psychiatric diagnosis was as good as it was for discriminating those with only a depression diagnosis vs those without any DSM diagnosis.

Using a cut-point of normal versus moderate or severe scores, the Nonspecific Physical Symptoms instrument (with pain items) had 86% sensitivity and 31% specificity in discriminating between participants with SSI scores above versus below the criterion. A cut-point of normal or moderate versus severe had 68% sensitivity and 68% specificity. The Depression instrument showed good ability to discriminate between individuals with and without current-year depression diagnoses (AUC = 0.81; Figure 2). As shown in the Figure, the region between arrows 1 and 2 is relatively flat, indicating that selecting an in-between cut-point would not provide substantial improvements in both sensitivity and specificity.

Clinical utility—Affective disorder prevalence has been estimated at 11.8% for acute TMD and 34% for chronic TMD.⁷ Using those estimates, the normal versus moderate-severe cut-point of the Depression instrument had a PPV of 19% (NPV = 97%) for detecting depression in patients with acute TMD and 48% (NPV = 88%) for detecting depression in patients with chronic TMD.

Discussion

The purpose of this study was to validate the existing RDC/TMD Axis II measures of depression, non-specific physical symptoms, and grade of chronic pain. Results of these studies indicate that, overall, the RDC/TMD Axis II measures have good to excellent psychometric properties. Internal consistency was very good to excellent, replicating

previous findings for Depression and Nonspecific Physical Symptoms instruments using TMD clinic and community samples¹¹ and extending examination to the other Axis II measures. Temporal stability was excellent for the pain and interference measures over a short period, and fair to good for the Depression and Nonspecific Physical Symptoms instruments over a two-week period. Perfect test-retest reliability would not be expected, as pain, interference, depressive symptoms, and nonspecific physical symptoms naturally vary over even short periods of time. With the exception of the Nonspecific Physical Symptoms instrument, convergent validity for all measures tested was demonstrated by moderate to high correlations with other established measures of similar constructs and discriminant validity was demonstrated by lower correlations with measures of less similar constructs. Criterion validity was demonstrated for the Depression instrument by its adequate sensitivity, using the normal versus moderate-severe cut-point, for identifying individuals with a psychiatric diagnosis of depression. However, the specificity of this cut-point was only 53%. Although the number of subjects with scores in the highest range on the study measures was limited, the diversity of the sample in age, education, and scores on the measures helps to increase the generalizability of the results. For example, the proportions of Study 1 participants categorized as chronic pain grade III and IV (8% and 5%, respectively) were comparable to rates found in a population-based study of patients seeking treatment for TMD (11% and 5%, respectively).⁶

The RDC/TMD Axis II measures were designed to assess the extent of a patient's psychosocial disability (e.g., disruption in performance of customary activities), because one of the most deleterious consequences of chronic pain is its impact on the ability to engage in daily activities and remain productive at work, home, or school. For the GCPS, the present data regarding reliability, temporal stability, and convergent and discriminant validity, as well as its ease of use, support its clinical utility for identifying TMD patients likely to have high levels of disability in performing customary activities.

The Axis II Depression instrument was not intended to yield a psychiatric diagnosis, but rather to screen for significant psychosocial distress that may or may not be present concurrently with a formal psychiatric disorder.¹¹ The present data support the use of the Depression instrument for this purpose. The instrument showed reasonably good ability to discriminate between study participants who did versus those who did not have a current-year psychiatric diagnosis of depression. The published RDC/TMD cut-point between low versus moderate-severe scores, which were determined from a population sample, had a sensitivity of 87% and specificity of 53% in identifying patients with this diagnosis in this study. Using the cut-point between moderate and severe increased the specificity at the cost of decreasing sensitivity, as would be expected.

Based on the only available prevalence data, the findings suggest that if a TMD patient scores in the moderate-severe range on the Depression instrument, there is approximately a 19% probability the patient will meet criteria for a depression or dysthymia diagnosis if the TMD problem is acute and a 48% chance if the problem is chronic. If a TMD patient scores in the normal range, there is approximately a 97% chance the patient does not have a depression or dysthymia diagnosis if the TMD is acute and an 88% chance if the TMD is chronic. These probabilities could differ in settings in which the prevalence of depression is higher (e.g., in academic-center specialty clinics) or lower (e.g., in a general dentist's practice). Nonetheless, as a screener for a depression diagnosis, our results indicate that the Depression instrument is most useful if the patient scores in the normal range; such patients are unlikely to have a diagnosis of depression. The Depression instrument discriminated between patients with and without any psychiatric diagnoses as well as it did for depression diagnoses, supporting the clinical utility of this instrument as a screener for psychosocial distress more generally.

Clinicians may wish to choose which Depression instrument cut-point to use for purposes of selecting patients for referral to mental health professionals based on the unique aspects of their setting (e.g., availability of mental health resources) and other characteristics of the patient (e.g., high scores on the other Axis II measures would support a referral for a patient with a lower score on the Depression instrument). For patients with scores below the moderate cut-point and no other indicators of significant psychosocial problems, it would seem appropriate to direct treatment primarily, if not solely, to Axis I conditions. Regardless of total score on the instrument, it is important to examine the item concerning suicidal ideation; positive responses on this item require inquiry and consideration of referral of the patient to a qualified mental health professional for further evaluation.

The Nonspecific Physical Symptoms instrument was not originally intended to serve the purpose of making a psychiatric diagnosis. When used to identify patients with psychiatric diagnoses in this study, the Nonspecific Physical Symptoms instrument had a low specificity when sensitivity was adequate; these results underscore that it should not be used for this purpose. The Nonspecific Physical Symptoms instrument, labeled “somatization” in the SCL-90, was renamed by Dworkin and LeResche¹ to describe the instrument content without etiological inference. As Dworkin et al.¹¹ note, these symptoms may be associated with an underlying disease, the effects of the pain condition, and/or psychological distress. Thus, one would not necessarily expect a strong association between the instrument and a formal diagnosis of somatization disorder. Another reason for not expecting to see a strong association with a diagnosis of somatization disorder is the rarity of this disorder (around 0.2% in the general population).¹⁵ In fact, only two study participants (0.3%) had this diagnosis.

The Nonspecific Physical Symptoms instrument showed only a moderate association with the GHQ-28 Somatic Symptoms instrument. This may be a function of differences in both the item content and the response choices. Four of the seven GHQ-28 items assess general feelings of poor health rather than specific symptoms, and two assess head pain/pressure/tightness, which study participants might frequently endorse. For each GHQ item, respondents are asked to indicate how often they felt that way recently relative to “usual.” Individuals with symptoms of long duration may respond “same as usual” even when experiencing the symptom frequently. In contrast, the Nonspecific Physical Symptoms instrument asks how much respondents were distressed by various somatic symptoms (e.g., headaches, faintness/dizziness, chest pain, low back pain, heart pounding or racing, nausea), rather than general malaise, in the past week. Consequently, even though both instruments purportedly assess physical symptoms, the SCL90-based instrument may have content validity more appropriate to its intended RDC/TMD Axis II purpose.

Correlations of the Nonspecific Physical Symptoms instrument were higher (but still in a moderate range) with the MPI dysfunctional score (a composite measure of pain and interference) and the CES-D than they were with the GHQ-28 instrument. With pain items removed from the Nonspecific Physical Symptoms instrument, its correlation with the CES-D did not meaningfully change, but correlations decreased somewhat with GHQ Somatic Symptoms, MPI affective distress, pain severity, interference, and dysfunctional scales as would be expected. These results suggest some overlap (but a substantial amount of independence as well) between endorsement of nonspecific physical symptoms on this instrument and endorsement of depressive symptoms. Previous studies have also found that endorsement of multiple somatic symptoms is associated with affective and anxiety disorders.⁵⁶

As a psychiatric diagnosis, somatoform disorders in general and somatization disorder in particular have been subject to criticism and debate within the psychiatric community.

Somatization disorder has a very low prevalence and there is substantial overlap with anxiety and depression. The diagnosis can only be made when symptoms are viewed as “psychogenic,” but it can be very difficult to decide whether a symptom is medically explained or not. Somatoform diagnoses are also problematic in that they lack acceptability to patients and they do not guide treatment. A number of experts have called for substantial revisions to the somatoform disorders section and the somatization disorder diagnosis in the next version of the DSM.^{57–59}

Despite these problems, there may be utility in both the RDC/TMD Nonspecific Physical Symptoms instrument and the construct of somatization. The measure detected individuals bothered by multiple somatic symptoms that have not been found to relate to a medical diagnosis. Although somatization has been defined in a number of different ways, the core element is presence of multiple somatic symptoms that cannot be explained adequately by biomedical findings and that patients find distressing. Patients identified as meeting criteria for the presence of multiple nonspecific physical symptoms but not meeting formal DSM-IV criteria for somatization disorder consistently have been shown to have increased health care utilization and disability.^{51,53,60,61} Furthermore, Dworkin et al.¹¹ found significant associations between Nonspecific Physical Symptoms instrument scores and the number of muscles painful to palpation on RDC/TMD examination, and Wilson et al.⁶² found a strong association between scores and number of placebo sites reported as painful during the RDC/TMD examination. Dworkin et al.¹¹ proposed that high Nonspecific Physical Symptoms instrument scores may reflect heightened vigilance for noticing and heightened tendency to be disturbed by somatic symptoms, and that this tendency might affect response to clinical examination.

The prevalence and importance of multiple nonspecific or “medically unexplained” symptoms in various patient populations has led to increased attention in the literature.^{63–66} Health care providers view patients with multiple medically unexplained symptoms as more frustrating than patients with more localized symptoms, and such patients commonly receive extensive clinical investigations.^{64,67} While these medically unexplained symptoms do not necessarily qualify a patient for a formal DSM IV Somatoform Disorder diagnosis, they nevertheless indicate a poor prognosis for the outcome of any chronic condition, including chronic pain problems in general and TMDs more specifically.^{66,68} Awareness that a patient is bothered by multiple nonspecific somatic symptoms may help a clinician interpret clinical findings in the context of the “whole patient” and prompt inquiry about prior somatic problems for which the patient has sought care.

Several limitations of this study warrant noting. Because the aim of the study was to evaluate the psychometric properties of the existing screener instruments in Axis II of the RDC/TMD, the study analyses were guided by classical test theory (CTT), which includes convergent/discriminant correlation matrices as well as reliability and diagnostic utility statistics. Convergent/discriminant matrices, as demonstrated in Table 7, often result in interesting mosaics that are not always readily interpretable as demonstrating uniformly high consistency with existing similar measures or uniformly low consistency with existing dissimilar measures. The limitations of this approach versus item response theory (IRT) are extensively described elsewhere,^{69–71} but because we are not developing new measures or even modifying existing measures, we relied solely on CTT for this paper. A second limitation, already discussed, is the lack of a “gold standard” criterion against which the screening utility of the Non-Specific Physical Symptoms scale can be judged.

A third limitation relates to the complexity of recruiting the sufficient mix of individuals representing the different Axis I diagnoses and its impact on the distribution of Axis II variables. The study recruitment plan was oriented around Axis I needs, and we let the

subject characteristics, as measured by our Axis II instruments, be dependent on the Axis I diagnostic needs. The resultant Axis II dataset reflects a mix of individuals with TMD. While the means of many of the demographic variables as well as comparison variables differed across study sites, those differences were not substantial in terms of clinical meaningfulness and the pattern of those different means was accounted for by varying underlying populations that each study site recruited from as part of the study design. The net effect of those differences on the reliability and validity coefficients was negligible, as demonstrated by secondary analyses summarized in the Results. Moreover, for the more critical sub-studies (2a, 2b, and 3), the UB and UW samples were very similar on most of the measures, and consequently the diagnostic validity coefficients were not affected by site differences. In contrast, a limited number of subjects at the more severe end of the instruments' measurement scales may have resulted in an underestimate for the test-retest reliability and convergent validity of the measures' true psychometric properties, due to restriction of range.

Axis I validity is essential for confirming that the RDC/TMD can yield a diagnosis that reflects the best understanding of what is wrong with the patient's pathophysiology that current information allows. The major significance of such a diagnosis – any diagnosis – is its implications for differential treatment: if all diagnoses led to the same treatment, diagnosis would be merely an intellectual exercise. By contrast, Axis II validity has no such parallel relationship to diagnosing whether or not the patient suffers a diagnosable psychiatric condition. Instead, the Axis II domains of pain, psychological status, and psychosocial disability reflect the self-report of dysphoric subjective symptoms related to those domains. Validity of the depression and nonspecific physical symptom measures of the RDC/TMD, based on how well those measures perform compared to psychiatric diagnostic criteria, is useful only because it demonstrates that one can have confidence that the patient's subjective experience bears resemblance to known phenomenological states associated with significant distress and psychosocial morbidity. While these phenomenological states were identified in this study in terms of formal psychiatric disorders, specifically, depression and somatization, burden of establishing validity for the Axis II measures is conceptually and clinically different as compared to that required for Axis I because the RDC/TMD deliberately avoids any implication of making a psychiatric diagnosis.

To summarize, Axis II serves two critical functions. The first is the clinical function of alerting health care providers that: (a) in the domain of depression, the patient may be experiencing a dysphoric mood state and/or may be at risk for a depressive disorder; (b) in the domain of non-specific physical symptoms, the patient may be experiencing widespread pain problems and/or a more generalized somatic dysregulation as reflected by multiple physical complaints in diverse organ systems, and/or (c) in the domain of disability, the patient may be exhibiting significant illness behaviors with corresponding limitations in activities. The second critical function of Axis II is its utility for research into how personal and psychosocial levels of function may or may not be related to peripheral and/or central pathophysiologic processes as well as cultural and ethnic influences affecting pain perception. In summary, abundant evidence has established that the natural history and clinical course of TMD and other chronic pain conditions is adversely affected by elevations in symptoms of depression and/or widespread pain or medically unexplained physical symptoms, hence the clinical utility in assessing Axis II levels of personal functioning.

How might RDC/TMD Axis II measures be developed further? First, Axis II does not include measures of anxiety or fear of pain and consequent avoidance of activities, constructs that have received considerable attention recently in the literature on chronic pain. Anxiety, as a state or trait characteristic of the person, has a close association with anxiety

disorders that may affect TMD patient psychosocial functioning and response to treatment (as well as close associations with depressive disorders).^{72–74} We plan in future work to report on the value of including a measure of anxiety in RDC/TMD Axis II. Second, since the publication of the RDC/TMD, other screening measures for depression have been published and validated. The Patient Health Questionnaire-9 (PHQ-9)⁷⁵ and even shorter PHQ-2⁷⁶ are two screening measures that have good psychometric properties and have been demonstrated to be useful for screening purposes in medical populations. Unlike the SCL-90 measure, the PHQ was designed to assess the specific symptoms that constitute the core criteria for a DSM-IV depression diagnosis; thus, it might be a better choice than the SCL-90 depression scale if the goal is to assess presence of this diagnosis. It might be of interest in further work to compare different depression measures in terms of clinical utility in TMD populations. Other directions for further work include developing shorter measures (e.g., using IRT techniques and computer adaptive testing) of psychological distress and nonspecific symptoms, and exploring the utility of adding measures of other problems commonly experienced by patients with TMD, such as sleep disturbance or psychosocial stress.

Acknowledgments

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

References

1. Dworkin SF, LeResche L. Research Diagnostic Criteria for Temporomandibular Disorders: Review, Criteria, Examinations and Specifications, Critique. *J Craniomandib Disord Facial Oral Pain.* 1992; 6:301–355.
2. Gatchel RJ, Peng YB, Peters M, et al. The biopsychosocial approach to chronic pain: scientific advances and future directions. *Psychol Bull.* 2007; 133:581–624. [PubMed: 17592957]
3. Waddell G. Preventing incapacity in people with musculoskeletal disorders. *Br Med J.* 2006; 78:55–69.
4. Brage S, Sandanga I, Nygard JF. Emotional distress as a predictor for low back disability. *Spine.* 2007; 32:269–274. [PubMed: 17224825]
5. Turner JA, Franklin G, Turk D. Predictors of chronic disability in injured workers: a systematic literature synthesis. *American Journal of Industrial Medicine.* 2000; 38:707–722. [PubMed: 11071692]
6. Von Korff M, Ormell J, Keefe FJ, et al. Grading the severity of chronic pain. *Pain.* 1992; 50:133–149. [PubMed: 1408309]
7. Gatchel RJ, Garofalo JP, Ellis E, et al. Major psychological disorders in acute and chronic TMD: an initial examination. *JADA.* 1996; 127:1365–1374. [PubMed: 8854613]
8. Kinney RK, Gatchel RJ, Ellis E, et al. Major psychological disorders in chronic TMD patients: implications for successful management. *JADA.* 1992; 123:49–54. [PubMed: 1401593]
9. Wright AR, Gatchel RJ, Willdenstein L, et al. Biopsychosocial differences between high-risk and low-risk patients with acute TMD-related pain. *JADA.* 2004; 135:474–83. [PubMed: 15127871]
10. Garofalo JP, Gatchel RJ, Wesley AL, et al. Predicting chronicity in acute temporomandibular joint disorders using the Research Diagnostic Criteria. *JADA.* 2007; 129:438–447. [PubMed: 9573694]
11. Dworkin SF, Sherman JJ, Mancl L, et al. Reliability, validity, and clinical utility of RDC/TMD Axis II scales: Depression, non-specific physical symptoms, and graded chronic pain. *J Orofacial Pain.* 2002; 16:207–220. [PubMed: 12221737]

12. Ohrbach R, Granger CV, List T, et al. Pain-related Functional Limitation of the Jaw: Preliminary Development and Validation of the Jaw Functional Limitation Scale. *Comm Dent Oral Epidem.* 2008; 36:228–236.
13. Ohrbach R, Larsson P, List T. The Jaw Functional Limitation Scale: Development, reliability, and validity of 8-item and 20-item versions. *Journal of Orofacial Pain.* 2008
14. Kight M, Gatchel RJ, Wesley L. Temporomandibular disorders: evidence for significant overlap with psychopathology. *Health Psychol.* 1999; 18:177–182. [PubMed: 10194053]
15. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders - IV.* Washington, D.C: American Psychiatric Association; 1994.
16. Scott VM, Von Korff M, Alonso J, et al. Mental-physical co-morbidity and its relationship to disability: results from the World Mental Health Surveys. *Psychol Med.* 2009; 39:33–43. [PubMed: 18366819]
17. Conner TS, Tennen H, Zautra AJ, et al. Coping with rheumatoid arthritis pain in daily life: within-person analyses reveal hidden vulnerability for the formerly depressed. *Pain.* 2006; 126:198–209. [PubMed: 16904829]
18. Tennen H, Affleck G, Zautra A. Depression history and coping with chronic pain: a daily process analysis. *Health Psychol.* 2006; 25:370–379. [PubMed: 16719609]
19. Banks SM, Kerns RD. Explaining high rates of depression in chronic pain: a diathesis-stress framework. *Psychol Bull.* 1996; 119:95–110.
20. Schiffman EL, Truelove EL, Ohrbach R, et al. Assessment of the Validity of the Research Diagnostic Criteria for Temporomandibular Disorders: Overview and Methodology. *Journal of Orofacial Pain.* 2009
21. Goldberg DP, Gater R, Sartorius N, 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. [PubMed: 9122299]
22. Benjamin S, Lennon S, Gardner G, et al. The validity of the General Health Questionnaire for first-stage screening for mental illness in pain clinic patients. *Pain.* 1991; 47:197–202. [PubMed: 1762815]
23. Cardenas DD, Turner JA, Warms CA, et al. Classification of chronic pain associated with spinal cord injuries. *Archives of Physical Medicine & Rehabilitation.* 2002; 83:1708–1714. [PubMed: 12474174]
24. Wright KD, Asmundson GJ, McCreary DR. Factorial validity of the short-form McGill pain questionnaire (SF-MPQ). *European Journal of Pain.* 2001; 5:279–284. [PubMed: 11558983]
25. Dworkin SF, Von Korff M, Whitney CW, et al. Measurement of characteristic pain intensity in field research. *Pain.* 1990; (Suppl 5):S290.
26. Visser M, Leentjens AF, Marinus J, et al. Reliability and validity of the Beck depression inventory in patients with Parkinson's disease. *Movement Disorders.* 2006; 21:668–672. [PubMed: 16450355]
27. Von, Korff M. Epidemiologic and survey methods: chronic pain assessment. In: Turk, DC.; Melzack, R., editors. *Handbook of Pain Assessment.* New York: Guilford Press; 1992. p. 389-406.
28. Derogatis LR, Lipman RS, Covi L. SCL-90: an outpatient psychiatric rating scale--preliminary report. *Psychopharmacology.* 1973; 9:13–28.
29. Derogatis, LR. *SCL-90-R: Administration, Scoring and Procedures Manual-II, for the Revised Version.* Towson, MD: Clinical Psychometric Research; 1983.
30. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. *App Psych Measurement.* 1977; 1:385–401.
31. Knight RG, Williams S, McGee R, et al. Psychometric properties of the Centre for Epidemiologic Studies Depression Scale (CES-D) in a sample of women in middle life. *Behaviour Research & Therapy.* 1997; 35:373–380. [PubMed: 9134792]
32. Haringsma R, Engels GI, Beekman ATF, et al. The criterion validity of the Center for Epidemiological Studies Depression Scale (CES-D) in a sample of self-referred elders with depressive symptomatology. *International Journal of Geriatric Psychiatry.* 2004; 19:558–563. [PubMed: 15211536]

33. Geisser ME, Roth RS, Robinson ME, et al. Assessing depression among persons with chronic pain using the Center for Epidemiological Studies-Depression Scale and the Beck Depression Inventory: a comparative analysis. *Clin J Pain*. 1997; 13:163–170. [PubMed: 9186024]
34. Turk DC, Okifuji A. Detecting depression in chronic pain patients: adequacy of self-reports. *Behaviour Research & Therapy*. 1994; 32:9–16. [PubMed: 8135727]
35. Werneke U, Goldberg DP, Yalcin I, et al. The stability of the factor structure of the General Health Questionnaire. *Psychol Med*. 2000; 30:823–829. [PubMed: 11037090]
36. Rudy TE, Turk DC, Zaki HS, et al. An empirical taxometric alternative to traditional classification of temporomandibular disorders. *Pain*. 1989; 36:311–320. [PubMed: 2710560]
37. Flor H, Turk DC. Chronic back pain and rheumatoid arthritis: predicting pain and disability from cognitive variables. *J Behav Med*. 1988; 11:251–265. [PubMed: 2971810]
38. Kerns RD, Turk DC, Rudy TE. The West Haven-Yale Multidimensional Pain Inventory (WHYMPI). *Pain*. 1985; 23:345–356. [PubMed: 4088697]
39. Turk DC, Rudy TE. Toward an empirically derived taxonomy of chronic pain patients: Integration of psychological assessment data. *J Consult Clin Psych*. 1988; 56:233–238.
40. Turk DC, Rudy TE. The robustness of an empirically derived taxonomy of chronic pain patients. *Pain*. 1990; 43:27–35. [PubMed: 2148974]
41. Rudy, TE. Multidimensional Pain Inventory Version 3.0 User's Guide. Pittsburgh, PA: University of Pittsburgh; 2005. (www.pain.pitt.edu/mpi)
42. Ware, JE., Jr; Kosinski, M.; Turner-Bowker, DM., et al. How to Score Version 2 of the SF-12 Health Survey. Lincoln, RI: Quality Metric Incorporated; 2002.
43. Ware JE Jr, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Medical Care*. 1996; 34:220–233. [PubMed: 8628042]
44. Luo X, Lynn George M, Kakouras I, et al. Reliability, validity, and responsiveness of the short form 12-item survey (SF-12) in patients with back pain. *Spine*. 2003; 28:1739–1745. [PubMed: 12897502]
45. Robins LN, Helzer JE, Croughan J, et al. National Institute of Mental Health Diagnostic Interview Schedule. Its history, characteristics, and validity. *Arch Gen Psychiat*. 1981; 38:381–389. [PubMed: 6260053]
46. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders - III. Washington, D.C: American Psychiatric Association; 1980.
47. Bucholz KK, Robins LN, Shayka JJ, et al. Performance of two forms of a computer psychiatric screening interview: version I of the DISSI. *J psychiat Res*. 1991; 25:117–129. [PubMed: 1941708]
48. Williams JB, Gibbon M, First MB, et al. The Structured Clinical Interview for DSM-III-R (SCID). II. Multisite test-retest reliability. *Arch Gen Psychiat*. 1992; 49:630–636. [PubMed: 1637253]
49. Robins, LN.; Cottler, LB.; Bucholz, KK., et al. Diagnostic Interview Schedule for the DSM-IV (DIS-IV). St Louis, MO: Washington University School of Medicine; 2000.
50. Escobar JI, Rubio-Stipec M, Canino G, et al. Somatic Symptom Index (SSI): A new and abridged somatization construct. Prevalence and epidemiological correlates in two large community samples. *J Nerv Ment Dis*. 1989; 177:140–146. [PubMed: 2918297]
51. Escobar JI, Golding JM, Hough RL, et al. Somatization in the community: relationship to disability and use of services. *Am J Public Health*. 1987; 77:837–840. [PubMed: 3592038]
52. Escobar JI, Rubio-Stipec M, Canino G, et al. Somatic symptom index (SSI): a new and abridged somatization construct. Prevalence and epidemiological correlates in two large community samples. *Journal of Nervous & Mental Disease*. 1989; 177:140–146. [PubMed: 2918297]
53. Katon W, Line E, Von Korff M, et al. Somatization: a spectrum of severity. *Am J Psychiat*. 1991; 148:34–40. [PubMed: 1984704]
54. DeVellis, RF. Scale Development: Theory and Application. Thousand Oaks: Sage Publications; 2003.
55. Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*. 1989; 45:255–268. [PubMed: 2720055]

56. Escobar JI, Manu P, Matthews D, et al. Medically unexplained physical symptoms, somatization disorder and abridged somatization: studies with the Diagnostic Interview Schedule. *Psychiat Dev.* 1989; 7:235–245.
57. Fava GA, Wise TN. Issues for DSM-V: psychological factors affecting either identified or feared medical conditions: a solution for somatoform disorders. *Am J Psychiatry.* 2007; 164:1002–1003. [PubMed: 17606648]
58. Kroenke K, Sharpe M, Sykes R. Revising the classification of somatoform disorders: key questions and preliminary recommendations. *Psychosomatics.* 2007; 48:277–285. [PubMed: 17600162]
59. Mayou R, Kirmayer LJ, Simon G, et al. Somatoform disorders: time for a new approach in DSM-V. *Am J Psychiatry.* 2005; 162:847–855. [PubMed: 15863783]
60. Fink P. Somatization--beyond symptom count. *J Psychosom Res.* 1996; 40:7–10. [PubMed: 8730638]
61. Simon GE, Von Korff M, Piccinelli M, et al. An international study of the relation between somatic symptoms and depression. *N Engl J Med.* 1999; 341:1329–1335. [PubMed: 10536124]
62. Wilson L, Dworkin SF, Whitney C, et al. Somatization and pain dispersion in chronic temporomandibular disorder pain. *Pain.* 1994; 57:55–61. [PubMed: 8065797]
63. Nettleton S. I'd just want permission to be ill: Towards a sociology of medically unexplained symptoms. *Social Science & Medicine.* 2006; 62:1167–1178. [PubMed: 16135395]
64. Brown RJ. Introduction to the special issue on medically unexplained symptoms: Background and future directions. *Clin Psych Rev.* 2007; 27:769–780.
65. Houtveen JH, van Doornen LJP. Medically unexplained symptoms and between-group differences in 24-h ambulatory recording of stress physiology. *Biol Psychol.* 2007; 76:239–249. [PubMed: 17900786]
66. olde Hartman TC, Borghuis MS, Lucassen PLBJ, et al. Medically unexplained symptoms, somatisation disorder, and hypochondriasis: Course and prognosis. A systematic review. *J Psychosom Res.* 2009 In Press, Corrected Proof.
67. Salmon P. Conflict, collusion or collaboration in consultations about medically unexplained symptoms: The need for a curriculum of medical explanation. *Patient Education and Counseling.* 2007; 67:246–254. [PubMed: 17428634]
68. Frohlich C, Jacobi F, Wittchen HU. DSM-IV pain disorder in the general population. An exploration of the structure and threshold of medically unexplained pain symptoms. *European Archives of Psychiatry & Clinical Neuroscience.* 2006; 256:187–196. [PubMed: 16328107]
69. Cella D, Chang CH. A discussion of item response theory and its applications in health status assessment. *Medical Care.* 2000; 38:II66–II72. [PubMed: 10982091]
70. Hambleton RK. Emergence of item response modeling in instrument development and data analysis. *Medical Care.* 2000; 38:II60–II65. [PubMed: 10982090]
71. Embretson, SE.; Reise, SP. *Item Response Theory for Psychologists.* Mahwah, NJ: Lawrence Erlbaum; 2000.
72. Dura E, Andreu Y, Galdon MJ, et al. Psychological assessment of patients with temporomandibular disorders: confirmatory analysis of the dimensional structure of the Brief Symptoms Inventory 18. *J Psychosom Res.* 2006; 60:365–370. [PubMed: 16581360]
73. De LR, Bertoli E, Schmidt JE, et al. Prevalence of traumatic stressors in patients with temporomandibular disorders. *Journal of Oral & Maxillofacial Surgery.* 2005; 63:42–50. [PubMed: 15635556]
74. Turner JA, Dworkin SF. Screening for psychosocial risk factors in patients with chronic orofacial pain: recent advances. *JADA.* 2004; 135:1119–1125. [PubMed: 15387050]
75. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine.* 2001; 16:606–613. [PubMed: 11556941]
76. Kroenke K, Spitzer RL, Williams JB. The Patient Health Questionnaire-2: validity of a two-item depression screener. *Medical Care.* 2003; 41:1284–1292. [PubMed: 14583691]

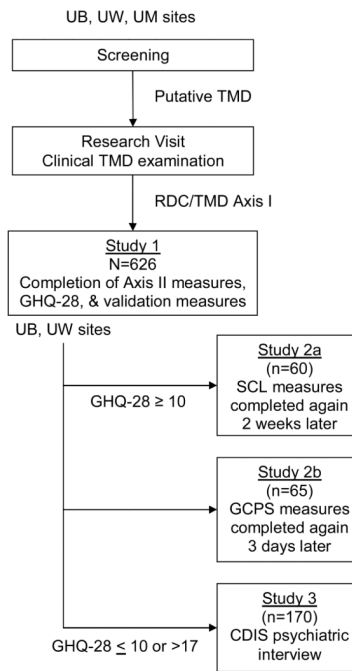


Figure 1. Flowchart of studies

Flow path, as part of the overall study of the RDC/TMD, for subjects entering into the 3 studies examining the reliability and validity of Axis II.

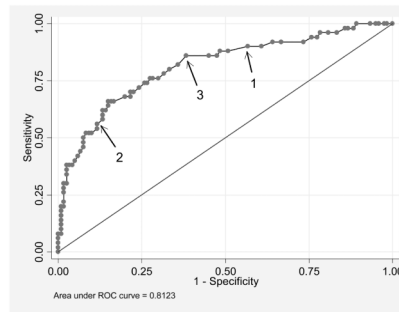


Figure 2. ROC Curve of Depression Instrument Predicting Current-year Depressive Disorders
Receiver operating characteristic (ROC) curve demonstrating relationship of the true-positive rate (sensitivity) versus false-positive rate (1 – specificity) for the RDC/TMD Axis II Depression instrument in predicting current-year DSM major depression or dysthymia. The utility of the published cut-points at the moderate (0.505, arrow 1) and at the severe (1.105, arrow 2) level are shown, with the flat part of the curve between them demonstrating that there is no empirically determined better cut-point in terms of overall utility. Modification of the moderate cut-point to 0.60 (arrow 3) does improve the cut-point visually, but with only slight improvement in the false-positive rate with a slight decrease in true-positive rate. The line of unity references a test with 50% for each of true-positive versus false-positive rates.

Table 1

Study Population Demographic and Clinical Characteristics

Demographics	Study ¹			
	1	2a	2b	3
n	626 ²	60 ³	65 ⁴	170 ⁵
Age (years), mean (SD)	37.5 (13.1)	39.8 (11.3)	40.8 (13.3)	39.9 (12.2)
Female, n (%)	533 (85%)	49 (82%)	57 (88%)	149 (88%)
White, n (%)	573 (92%)	57 (95%)	60 (92%)	155 (91%)
Education, n (%)				
High school or less	99 (16%)	11 (18%)	10 (15%)	33 (19%)
Some college	239 (38%)	23 (38%)	28 (43%)	62 (37%)
College graduate	287 (46%)	26 (44%)	27 (42%)	75 (44%)
RDC/TMD Axis I pain diagnosis, ⁶ n (%)	510 (81%)	27 (45%)	37 (57%)	153 (90%)

¹ Study 1 = Internal Consistency and Convergent and Discriminant Validity; Study 2a = Temporal Stability: Depression and Nonspecific Physical Symptoms instruments; Study 2b = Temporal Stability: Graded Chronic Pain Scale measures; Study 3 = Criterion Validity.

² Of 628 participants with an RDC/TMD diagnosis, two had missing SCL-90R data and were dropped from all analyses.

³ 74 Study 1 participants were asked to complete the Depression and Nonspecific Physical Symptoms instruments on two occasions. Of these participants, 10 were not TMD cases, 3 had a retest interval that was less than the desired interval of 7–27 days, and 1 had missing data.

⁴ 74 Study 1 participants were asked to complete the GCPS measures on two occasions. Of these participants, 4 were not TMD cases, 1 had a retest interval less than the desired interval of 2–7 days, 2 had a retest interval greater than 7 days, and 2 participants had missing data. Of this n=65 analysis sample, all provided complete CPI data, 4 had missing interference data, and chronic pain grade could not be determined for one.

⁵ 170 Study 1 participants were recruited and completed the interviews

⁶ Diagnosis = myofascial pain, arthralgia, or both, and with or without other Group II or Group III diagnoses.

SD = standard deviation.

Table 2

Number of Study 1 (n = 626) Participants with each Combination of the Three Major RDC/TMD Diagnostic Group Diagnoses

Axis I Diagnostic Combinations	n
Myofascial or TMJ pain	82
Disc displacements only	75
Arthritis or arthrosis only	11
Myofascial or TMJ pain, and disc displacements	221
Myofascial or TMJ pain, and arthritis or arthrosis	9
Disk displacements, and arthritis or arthrosis	30
Myofascial or TMJ pain, disc displacements, and arthritis/arthrosis	198

Table 3

Scores on Study 1 Measures

RDC/TMD Measures	Mean (SD)		Range
<i>Entire sample:</i>			
Depression ¹	0.50*	(0.52)	0 – 3.5
Nonspecific Physical Symptoms, with pain items ¹	0.55*	(0.51)	0 – 3.1
Nonspecific Physical Symptoms, without pain items ¹	0.33*	(0.48)	0 – 3.1
<i>Only participants who reported current pain:</i>			
Characteristic pain intensity ²	51.3*	(20.1)	6.7 – 100
Interference ²	19.7*	(22.4)	0 – 100
Number of Disability Days ²	8.6*	(27.4)	0 – 180
Chronic pain grade ²	N (%)		
I	217*	(42%)	N/A
II	226*	(45%)	N/A
III	39*	(8%)	N/A
IV	27*	(5%)	N/A
Validity Measures	Mean (SD)		Range
CES-D ³	8.8	(8.6)	0 – 54
GHQ-28 Somatic Symptoms ¹	5.2*	(3.4)	0 – 18
MPI Affective Distress ¹	40.4	(14.1)	0 – 85
MPI Pain Severity ¹	33.1*	(20.7)	0 – 100
MPI General Activity ¹	56.9*	(7.2)	0 – 88
MPI Interference ¹	27.7*	(19.0)	0 – 77
MPI Dysfunctional ¹	34.2*	(11.3)	5 – 72
SF-12v2 PCS ⁴	50.5	(4.7)	31 – 66
SF-12v2 MCS ⁴	52.2	(9.7)	12 – 67

* Values across the study sites differ significantly ($P < .05$ using ANOVA for continuous measures, chi-square for chronic pain grade).

¹Total n=626; UB n=224, UW n=220, UM n=182.

²Total n=509 limited to participants who reported jaw pain in past 30 days, had characteristic pain intensity scores >0, and had no missing data on these variables; UB n=178, UW n=191, UM n=140.

³Total n=444; UB n=224, UW n=220; not administered at UM.

⁴Total n=566 (less than full sample of 626 due to administrative errors); UB n=201, UW n=190, UM n=175.

SD = standard deviation; N/A = not applicable.

Table 4

Internal Consistency of Axis II Measures

Measure	Cronbach's alpha	Lower-bound 95% CI
Depression ¹	0.91	0.903
Nonspecific Physical Symptoms, with pain items ¹	0.84	0.821
Nonspecific Physical Symptoms, without pain items ¹	0.80	0.778
Characteristic Pain Intensity ²	0.84	0.815
Interference ²	0.95	0.939

¹ n = 626 for Depression and Nonspecific Physical Symptoms instruments.

² n = 515 for characteristic pain intensity and interference scales, which were calculated using only participants who reported pain in the prior 30 days and had characteristic pain intensity scores > 0.

Table 5
 Convergent and Discriminant Validity¹: Associations² between RDC/TMD Axis II Measures and Validity Measures

Validity Measure	Axis II Measure					
	Depression	Nonspecific Physical Symptoms, with pain items	Nonspecific Physical Symptoms, without pain items	Characteristic Pain Intensity	Interference	Chronic Pain Grade
CES-D	0.85 (0.82, 0.87)	0.56 (0.50, 0.63)	0.56 (0.50, 0.62)	0.20 (0.10, 0.30)	0.30 (0.21, 0.40)	0.21 (0.11, 0.31)
GHQ-28 Somatic Symptoms	0.39 (0.32, 0.46)	0.47 (0.41, 0.53)	0.42 (0.36, 0.49)	0.24 (0.16, 0.32)	0.29 (0.22, 0.37)	0.19 (0.10, 0.27)
MPI: Affective Distress	0.59 (0.54, 0.64)	0.41 (0.34, 0.48)	0.35 (0.28, 0.42)	0.13 (0.05, 0.21)	0.20 (0.11, 0.28)	0.15 (0.06, 0.23)
MPI: Pain Severity	0.32 (0.24, 0.39)	0.46 (0.40, 0.52)	0.34 (0.27, 0.41)	0.64 (0.60, 0.69)	0.49 (0.43, 0.55)	0.37 (0.29, 0.44)
MPI: General Activity	-0.15 (-0.23, -0.07)	-0.11 (-0.19, -0.03)	-0.10 (-0.17, -0.02)	-0.02 ³ (-0.10, 0.07)	-0.08 ³ (-0.16, 0.00)	-0.07 ³ (-0.16, 0.02)
MPI: Interference	0.33 (0.26, 0.40)	0.42 (0.36, 0.49)	0.33 (0.26, 0.40)	0.43 (0.37, 0.50)	0.53 (0.47, 0.59)	0.44 (0.37, 0.51)
MPI: Dysfunctional	0.59 (0.54, 0.64)	0.55 (0.49, 0.60)	0.445 (0.39, 0.51)	0.45 (0.38, 0.52)	0.51 (0.45, 0.57)	0.35 (0.27, 0.42)
SF-12v2: PCS	0.01 ³ (-0.07, 0.10)	-0.29 (-0.37, -0.22)	-0.25 (-0.33, -0.17)	-0.23 (-0.32, -0.15)	-0.33 (-0.42, -0.25)	-0.26 (-0.35, -0.17)
SF-12v2: MCS	-0.70 (-0.74, -0.66)	-0.43 (-0.49, -0.36)	-0.40 (-0.47, -0.33)	-0.09 ³ (-0.18, -0.00)	-0.20 (-0.30, -0.12)	-0.12 (-0.21, -0.03)

¹ Shaded boxes indicate pairs expected to show higher associations (convergent validity); nonshaded boxes indicate pairs expected to show lower associations (discriminant validity). Values shown are the association statistics and the 95% CIs in parentheses.

² Spearman rank correlation coefficient for associations with Chronic Pain Grade, Lin's CCC for associations with all other Axis II measures; all associations are adjusted for study site.

³ Not significantly different from 0 (alpha = 0.05); all other reported correlations are significantly different from 0 (p<0.05).

Note: n = 626 for all measures except for characteristic pain intensity and interference, which were analyzed only for participants who reported jaw pain in the past month and had characteristic pain intensity >0 (n = 515).

Table 6

Temporal Stability of Axis II Measures

Measure	Lin's CCC or Kappa ^I	95% CIs
Depression	0.78	0.687, 0.879
Nonspecific Physical Symptoms, with pain items	0.72	0.591, 0.840
Nonspecific Physical Symptoms, without pain items	0.63	0.481, 0.782
Characteristic Pain Intensity	0.91	0.867, 0.952
Interference	0.89	0.832, 0.941
Number of Disability Days	0.74	0.629, 0.842
Chronic Pain Grade	0.87	0.765, 0.976

^ILin's CCC for all measures except for chronic pain grade, which was weighted kappa (percent agreement for chronic pain grade = 99%).

Note: n = 60 for the depression and nonspecific physical symptoms scales (one participant of the original 61 participants was excluded due to missing data), n = 65 for characteristic pain intensity, n = 61 for interference (four had missing data), and n = 64 for chronic pain grade (grade could not be determined for one).

Table 7
Sensitivity and Specificity (and 95% CIs) of the Axis II Depression and Nonspecific Physical Symptom Instruments in Identifying Current-Year (in Past Year) and Lifetime (Prior to Past Year) Psychiatric Diagnoses

Axis II Measures and Cut-points	Criterion Psychiatric Status, ¹ Current Year		Any Psychiatric Diagnoses, ² Current Year		Any Psychiatric Diagnoses, ³ Lifetime	
	Sens ⁵	Spec	Sens	Spec	Sens	Spec
Depression						
Normal vs moderate-severe	87% (73, 94)	53% (42, 61)	84% (71, 91)	57% (45, 65)	68% (59, 76)	60% (44, 71)
Normal-moderate vs severe ⁴	56% (41, 70)	91% (84, 95)	49% (37, 62)	93% (87, 97)	34% (25, 43)	98% (0, 100)
Nonspecific Physical Symptoms						
Normal vs moderate-severe	86% (64, 97)	31% (24, 39)	82% (70, 90)	36% (26, 45)	74% (65, 82)	36% (22, 49)
Normal-moderate vs severe	68% (43, 85)	68% (59, 75)	56% (43, 68)	76% (65, 82)	45% (36, 55)	82% (67, 90)

¹ Diagnosis of major depressive episode or dysthymia for comparisons with the Depression instrument, and Somatic Symptom Index with the Nonspecific Physical Symptoms instrument (with pain items).

² Diagnostic grouping for “any diagnosis” includes all assessed diagnoses except for DSM-IV pain disorder.

³ Criterion diagnosis “lifetime” includes symptoms that met diagnostic criteria prior to the current year.

⁴ For lifetime diagnostic values, 100% specificity was estimated from the raw data, requiring that one additional observation, who had a positive screener but without diagnosis, be added to estimate the study site-adjusted statistics.

⁵ Sensitivity and specificity values are adjusted for study sites, while the 95% CIs are approximate estimates based on the crude unadjusted relationships.

Note: n = 170.