Published in final edited form as:

J Oral Rehabil. 2010 October ; 37(10): 784–798. doi:10.1111/j.1365-2842.2010.02144.x.

Assessment and Further Development of RDC/TMD Axis II Biobehavioral Instruments: A Research Program Progress Report

Richard Ohrbach

Abstract

A symposium was held in Toronto, 2008, in which research progress regarding the biobehavioral dimension of the Research Diagnostic Criteria for Temporomandibular Disorders (RDC/TMD) was presented. An extended workshop was held April 2009 in which further recommendations were made from an expert panel, using the 2008 symposium material as a base. This paper is a summary of the 2008 symposium proceedings with elaborations based on further developments. Seven studies were conducted between 2001 and 2008, in which the following were investigated: (a) basic properties of Axis II instruments, (b) reliability and criterion validity of Axis II instruments, (c) expansion of predictors, (d) metric equivalence of the depression and non-specific physical symptoms subscales in the RDC/TMD, (e) laboratory investigation of oral behaviors, (f) field data collection of oral behaviors, and (g) functional limitation of the jaw. Methods and results for each of these studies are described. Based on the results of these studies that have been published, as well as the direction of interim results from the few studies that await completion and publication, the biobehavioral domain of the RDC/TMD, as published in 1992, is reliable and valid. These results also provide strong evidence supporting the future growth of the biobehavioral domain as the RDC/TMD matures into subsequent protocols for both clinical and research applications.

Keywords

biobehavioral; TMD; screening; validity; reliability; Research Diagnostic Criteria

Background

This report is a description of progress within the biobehavioral part of a research program, the RDC/TMD Validation Project (which started in 2001), and this report builds on a formal presentation given at a symposium, Validation Studies of the RDC/TMD: Progress Towards Version 2, presented at the International Association for Dental Research, Toronto, Canada, on July 2, 2008. The purpose of this symposium was to provide a first overall report of the major progress made by this Project. Since that symposium, a number of further developments have occurred, and our biobehavioral research activities are updated in this report.

Author contact: Richard Ohrbach, DDS PhD, Department of Oral Diagnostic Sciences, University at Buffalo, Buffalo, NY 14214, ohrbach@buffalo.edu.

The development of the Research Diagnostic Criteria for Temporomandibular Disorders (RDC/TMD; (1)) was guided by two templates which remained a foundation for these biobehavioral studies as well. One template, based on how psychiatry has moved substantially forward in the last 60 years, was methodological: when very little is known about a condition, use reliable operational definitions within a boot-strap process that intentionally includes continual updating of those definitions as findings accrue. Such definitions and boot-strapping promote discovery, parallel investigations, and good theory development (2). For example, psychiatry, now planning the 5th edition of the Diagnostic and Statistical Manual, has created revisions of its diagnostic manual at roughly 15-year intervals.

The other template was focused on scope – the biopsychosocial model – and intent of Axis II – the biobehavioral domain. The biopsychosocial model indicates that disease is not just a biological phenomenon but rather disease and its course are the result of an intersection of psychological and social factors with biology. The initial perspective underlying the content of Axis II was that markers were available for detecting clinically important biopsychosocial dysregulation relevant to pain conditions – that is, characteristics that could contribute to pain and suffering or could interfere with expected responses to usual treatment. The assessment of the person within Axis II of the RDC/TMD was intended to be brief for ready use as a screener in clinical settings. In parallel with the disablement model developed by the World Health Organization (WHO) (3), psychological status and functional status were identified as the initial markers for person assessment within the RDC/TMD. Like the model used in psychiatry, multiple and interacting levels of analysis capture different levels of complexity, and this becomes part of our map for further development of a robust and empirically supported assessment framework of the person with TMD.

The biopsychosocial model and the dimensions identified via the WHO disablement model helped outline the range of measures that were selected for the projects in this report. Pain research – specifically that which focused on behavior, mood, and cognition – had progressed substantially since 1992 and it provided a set of constructs and measures which were to be tested for their relevance to TMD. An explicit goal in the study design was to provide a tractable process of moving from the content of the RDC/TMD to a new assessment model and which would also permit, if possible, legacy data to remain interpretable and relevant. One primary consideration was to ensure that any improvements in the Axis II assessment protocol, as demonstrated by our data, were to be practical with regards to any future clinical implementation. Consequently, accepted knowledge that had accumulated over the very scientifically productive 10-year period of pain research following the completion of the RDC/TMD manuscript and deemed relevant to a future RDC/TMD was carefully appraised, and measures for this project were defined. Some additional areas of interest to the investigators were considered but not incorporated into the study design either because the current knowledge was still in transition or because the clinical implementation would perhaps be too cumbersome for incorporation into the planned extensions of the RDC/TMD.

Therefore, the goals of the Validation Project's investigation of the RDC/TMD, with regards to the biobehavioral domain, were to (a) conduct assessments of fundamental psychometric

properties of the instruments comprising Axis II, (b) assess the reliability and validity of the current Axis II instruments with respect to a reference standard, (c) assess whether the scope of Axis II could be improved by adding measures representing other constructs, (d) develop a new measure for assessment of oral-related behavior, and (e) ultimately improve the screening function of Axis II by utilizing measurement models that involve item response characteristics as well as effect indicator approaches and create a shorter but better screener for the relevant biobehavioral dimensions. Note that a core principle informing the goals of the Axis II portion of the Validation Project was that Axis II was to continue to serve as a screening method and not to provide a mental health diagnostic classification regarding person status. The latter was regarded as the proper place for pain psychology or behavioral medicine evaluations.

In this report, seven studies which address these goals are summarized. Emerging directly from the symposium in Toronto, a subsequent workshop (4, 5) was held April 2009 in order to obtain, for the Axis II part of this project, consensus guidelines that would assist in the completion of these goals.

Core constructs and measures in the RDC/TMD Axis II

As the entry point for all of these studies, the RDC/TMD Axis II is comprised of measures that assess the following constructs:

- Depression: assesses core symptoms known to affect pain modulation and coping as well as indicating potential morbidity and suicidal ideation (6, 7), and measured with a depression scale adapted from the SCL-90 (8);
- Non-specific physical symptoms: assesses symptoms currently considered to represent functional somatic syndromes, other pain conditions, or comorbid conditions, and measured with the somatization scale from the SCL-90 (8);
- Pain-related disability: provides a hierarchical pain-related disability classification based on extent of life interference due to pain, and measured with the Graded Chronic Pain Scale (9);
- Characteristic pain intensity: average of current pain, and past 6-month average pain and past 6-month worst pain, derived as a sub-scale from the Graded Chronic Pain Scale (9); and
- Limitation of the jaw: assesses a wide range of potential jaw functions affected by pain, and measured with an ad hoc checklist created by the RDC/TMD authors using only face validity.

With the exception of the functional limitation checklist, the other 4 measures were previously validated independent of their incorporation into the RDC/TMD. These measures corresponded to the following domains within the World Health Organization disablement model (3): functional limitation (limitation checklist), psychosocial disability (depression, non-specific physical limitations), and pain-related disability (graded chronic pain).

Overview of Axis II studies

Seven studies pertaining to the further development of Axis II are described here.

Study 1, Basic Properties of Axis II Instruments (10), was conducted in parallel with the full Axis I study (11); the goal of Study 1 was to extend prior findings (12) by administering the Axis II instruments in a larger multi-site study sample comprised of a sufficient representation of all Axis I diagnoses, conduct standard psychometric tests of internal consistency, assess construct validity by comparing scale scores against appropriate other measures, and provide a base from which additional instruments could be assessed.

Study 2, Reliability and Criterion Validity of Axis II Instruments (10), was comprised of 3 sub-studies in order to assess: temporal stability of the pain-relevant measures, temporal stability of the depression and non-specific physical symptoms measures, and utility of the depression and non-specific physical symptoms screeners.

Study 3, Expansion of Predictors, targeted the third major goal of the Validation Project's Axis II component: could the existing Axis II be improved by incorporation of additional measures assessing other pain-relevant constructs, which would provide a more comprehensive person assessment for research settings and tertiary care clinical settings. The fifth major goal was to then create from this larger identified set of constructs a more refined and shorter screener for clinical use that would function effectively to identify patients, at the outset of evaluation, who were at risk for continued chronicity and not responding adequately to usual treatments.

Study 4, Metric Equivalence of Isolated Subscales used in RDC/TMD (13), emerged in response to critical questions regarding the psychometric properties of the depression and non-specific physical symptoms scales in the RDC/TMD. Those scales had been extracted from the original SCL-90 (8); while the sample data set used to develop interpretation norms for these extracted scales was larger and more comprehensive (i.e., based on a full population study) than that which was used for the original SCL-90, the possible non-commensurate measurement between the RDC/TMD scales and the original scales (as administered in the SCL-90 or SCL-90R (14)) warranted examination of this relationship in order to either confirm generalization of findings or to identify limitations in interpretation across methods of administration.

Study 5, Oral Behaviors in the Laboratory (15, 16), began with the study team developing a new self-report instrument in response to a specific request by our NIH advisory panel who asked us to consider one of the major hypotheses regarding TMD onset – the putative causal role of parafunctional behaviors. A draft instrument was constructed and that instrument was implemented for comprehensive data collection in Study 3. It was not clear whether some of the purported behaviors should be split or combined as single items. Moreover, with the identification of the different behaviors emerged a fundamental question of whether individuals had a common frame of reference for understanding what those terms meant behaviorally. Consequently, a study was developed which would evaluate the actual behaviors associated with the parafunctional terms.

Study 6, Oral Behaviors in the Field, followed Study 5 and evaluated the occurrence of the behaviors in the natural environment in order to test the validity of a self-report instrument that seeks to identify and quantify behaviors that normally occur outside conscious awareness.

Study 7, Functional Limitation of the Jaw, was another laboratory study. A newly developed instrument, the Jaw Functional Limitation Scale (17, 18) which had not yet been published at the time of beginning the Validation Project, had emerged directly from earlier research based on the functional checklist items within the RDC/TMD. Prior analyses had shown that “limitation” in using the jaw was poorly correlated with depression, anxiety, non-specific physical symptoms, and examination-based palpation scores and mobility, and was only moderately associated with pain and jaw symptoms. The Validation Project afforded an opportunity to more closely examine what “limitation” means to patients in real-time, given the importance of the functional limitation construct to a sufficiently comprehensive Axis II; that relationship of limitation to Axis II has been further expanded in a recent review (19). The new instrument used a response scaling hierarchy as computed from a logistic-based item-response model (20), and the aim of this laboratory study was to assess the concept of “limitation” as reported on a 100mm visual analog scale and in response to actual performance.

Methods and Results of Axis II Studies

Figure 1 provides an overview of recruitment sources, sample size, and flow of participants across the different studies.

Study 1: Basic Psychometric Properties of Axis II Instruments

The core RDC/TMD Axis II constructs were assessed with the depression and non-specific physical symptoms scales and the Graded Chronic Pain Scale (GCPS). These instruments were administered to 724 individuals comprised of n=626 TMD cases, representing all Axis I diagnostic sub-groups, and of n=98 non-TMD controls who were not part of these analyses given that we wanted the results to generalize to TMD cases. For purposes of testing construct validation, the following additional instruments were administered: Center for Epidemiologic Studies – Depression (21); General Health Questionnaire-28, which is comprised of a dominant general scale and four subscales: depression, anxiety, physical symptoms, and social functioning (22); Multidimensional Pain Inventory, of which the interference and dysfunctional scales as derived from item response models were of particular value (23, 24); and SF12v2, which is usefully subdivided into two subscales: physical component summary (PCS) and mental component summary (MCS) (25). The data from the RDC/TMD Axis II instruments were analyzed for internal consistency and construct validity. Complete methods and results are presented elsewhere (10).

Table 1 presents the summary statistics for internal consistency. For depression and non-specific physical symptoms, the values ranged from good to excellent. For pain intensity, internal consistency was good, while for pain-interference, internal consistency was excellent. The nature of these two pain-relevant measures perhaps explains the difference in their respective reliability statistics. The 3 items comprising the pain intensity scale assess

quite different aspects and different time domains of pain, and the reliability statistic appears consistent with those differences in time domain; in contrast, the pain-interference items assess a single concept – how much does pain interfere with each of these 3 major activity domains? – over a single time-period, and the high reliability indicates an excellent measurement tool relating to the construct.

The evaluation of construct validity was based on the standard method of computing the convergent and discriminant validity statistics, as shown in Table 2. Validity measures were identified from the additional instruments administered in the study, and specific associations were hypothesized for expected high correlations (convergent validity) and lower correlations (discriminant validity) with each of the core constructs. For this interpretation of results, a strong correlation was considered to be in the range of 0.4 – 0.9, and most of them were in the 0.5–0.6 range. For depression, the 3 targeted correlations from the other 3 depression measures for convergent validity were the strongest. For non-specific physical symptoms, the 1 targeted correlation with the other measure of just somatic symptoms was moderate within a mix of 6 other correlations of roughly the same magnitude and assessing associations with depression, distress, pain severity, pain interference, dysfunctional coping style, and mental health. For characteristic pain intensity, the 2 targeted correlations with other measures of pain severity were the strongest. For pain interference, the 3 targeted correlations comprised 2 that were the strongest, but 1 other construct, pain severity, was also strong. For graded chronic pain classification, the 3 targeted associations were clearly the strongest.

In sum, the existing measures in the RDC/TMD Axis II exhibit sufficient internal consistency. Each of the measures for depression, pain interference, pain intensity, and pain-related disability exhibit appropriate relationships with other similar and dissimilar measures, indicating good construct validity. In contrast, the measure for non-specific physical symptoms clearly assesses a phenomenon that is related to multiple constructs.

Study 2: Reliability and Criterion Validity of Axis II Instruments

A sub-set of participants (n=65 cases) from Study 1 was recruited and the Graded Chronic Pain Scale was re-administered with a target interval of 3 days, accepting 2–7 days interval. A second sub-set of participants (n=60) was selected using another instrument (General Health Questionnaire-28 (22)) in order to screen for the presence of at least mild distress in the participants. Requiring mild distress as an inclusion would ensure a range of responses in the target instruments of depression and non-specific physical symptoms, leading to more realistic reliability statistics. The instruments were administered again with a target interval of 14 days later, accepting 7–27 days interval. These two studies assessed the temporal stability of the respective constructs, using time intervals selected to exhibit minimal change in the participant's state. A third sub-set of participants (n=170) was recruited at a 2:1 ratio of cases vs controls. "Cases" were determined on the basis of having a high likelihood of a psychiatric disorder, while "controls" were determined on the basis of a low likelihood. The determination of high vs low likelihood was based on the use of the General Health Questionnaire-28 (22) as a separate instrument for selecting participants in order to avoid circularity between predictor (RDC/TMD scales) and the outcome (psychiatric

classification). These participants received a structured psychiatric interview (26, 27) in order to use DSM-IV (28) psychiatric classifications as a measure of overall functioning within the domains of the constructs of interest, depression and somatization diagnoses. The psychiatric classifications served as reference standard diagnoses. The data from Studies 2a and 2b were analyzed for temporal stability and from Study 2c for diagnostic utility. Complete methods and results are presented elsewhere (10).

As shown in Table 3, the temporal stability for depression and non-specific physical symptoms is in the acceptable range at 0.7–0.8; it appears that over the time interval of 7–27 days, as allowed for this study, the symptom reporting changes, and it is inferred that that is because these symptoms do fluctuate. Pain intensity and interference are stable, and the classification by the Graded Chronic Pain Scale is also stable.

For assessing utility of depression and non-specific physical symptoms as screening instruments for the identification of individuals at risk for substantial distress and symptom reporting, the depression score and the non-specific physical symptoms score were tested, separately, in logistic regression models controlling for study site, and validity statistics of sensitivity and specificity were then computed. The criterion measures were diagnoses of major depression and dysthymia over the past 1 year for the depression screener, and the Somatic Symptom Index (29–33) for the non-specific physical symptoms screener. The Somatic Symptom Index is considered positive for 4+ symptoms in males or 6+ symptoms in females.

Table 4 provides the statistics using different cut-points, as established in the RDC/TMD from population studies, in the depression and non-specific physical symptoms screeners. The cut-points that separate normal, moderate, and severe symptom levels were reduced to binary categories by collapsing moderate with normal or by collapsing moderate with severe. Considering normal vs moderate/severe depressive symptoms, screening efficiency is quite good at 87% sensitivity, while considering normal/moderate vs severe, efficiency is good at 91% specificity. In other words, setting the threshold lower enhances the sensitivity, leading to better use to rule out, while setting the threshold higher enhances the specificity, leading to better use to rule in (34). The implication of this finding is that a moderate level of depression is hard to interpret, and the ROC analysis is consistent: there is no single cut-point between the existing cut-points that differentiate moderate from normal or from severe that is better. Nevertheless, the area under the curve for the depression screener is 0.81, which is considered good for a screening instrument, and consequently, the use of dual cut-points, depending on the use of the instrument in a given situation, may be optimal. Considering normal vs moderate/severe physical symptoms based on the Somatic Symptom Index, screening is acceptable at 86% sensitivity, but specificity at even the most optimal cut-point of normal/moderate vs severe symptoms is not clinically useful.

In sum, the RDC/TMD depression scale has good psychometric properties for its intended purpose: screening for psychological distress. The Graded Chronic Pain Scale also has good psychometric properties for assessing pain intensity and pain-related disability. The non-specific physical symptoms scale, while not useful as a screener for psychiatric morbidity, remains useful for identifying the clinically salient characteristic of pain amplification (35),

and its corresponding use in clinical decision-making clearly warrants further investigation given the significance of that phenotypic characteristic with a potentially identifiable genetic basis (36).

Study 3: Expansion of Predictors

In order to fully address the third and fifth goals, the Axis II component of the Validation Project included the following additional constructs and associated instruments:

- Functional limitation of the jaw: Jaw Functional Limitation Scale (17, 18), and Mandibular Functional Impairment Questionnaire (37);
- Overuse behaviors of the jaw: Oral Behaviors Checklist (see Study 4);
- General dimensions of distress: General Health Questionnaire-28 (22), and Symptom Checklist 90R (14);
- Depression: Center for Epidemiologic Studies – Depression (21);
- Anxiety: State-Trait Anxiety Inventory (38);
- Stress: Perceived Stress Scale-10 (39);
- Pain quality: McGill Pain Questionnaire (40);
- Sleep quality: Pittsburgh Sleep Quality Index (41);
- Health-related quality of life: Short Form-12 version 2 (25);
- Pain-disability: Multidimensional Pain Inventory (23, 24), and Graded Chronic Pain Scale for 1 month (adaptation of the existing scale, requires validation);
- Explanation of illness: Explanatory Model Scale (unpublished);
- Personality: Structured Clinical Interview for DSM: Personality Disorders (42–44);
- Mental status: Neurobehavioral Cognitive Status Examination (45, 46).

These instruments have been tested in conjunction with the depression, non-specific physical symptoms, and graded chronic pain measures, against suitable reference standards, in order to expand the scope of Axis II. Preliminary results were presented at a subsequent workshop, and based on other published findings as well as these results recommendations were made for expanding Axis II (5, 19). This study is in the final stage of completion. A goal of the Validation Project was to create, based on the results from Study 3, a short screener for Axis II – replacing the present depression and non-specific symptoms scales. However, during the period of this study, a general mental health screener, the PHQ-9 (47), with appropriate content validity for use in a pain population, excellent clinical utility, and many language translations has emerged. The PHQ-9 instrument is also appropriate for use as a general mental health and functional status screener in a general dental setting and has been so recommended (48). Further expansion of Study 3 will consequently address whether a more specialized secondary screener for more pain-relevant constructs can be derived from the primary Study 3 results, and which would perhaps serve in conjunction with the PHQ-9.

Study 4: Metric Equivalence of Isolated Subscales used in RDC/TMD

Subjects (n=103) were recruited from TMD private practice, TMD teaching clinic, and regular dental patients identified as at risk for psychosocial complications to routine dental care. At-risk status in the dental patients was based on any problems such as with transportation, income, access to care, distress, or other complex medical conditions, as assessed by a social support services group at the School. The goal in recruitment was to identify two groups of individuals: those with TMD symptoms in order to provide generalizability to the RDC/TMD, and those specifically with distress not related to TMD pain in order to insure generalizability to the psychometric literature where these instruments are more commonly used.

Two instruments were administered to all subjects: the full SCL-90R and only the depression and non-specific physical symptoms scales as published in the RDC/TMD. The order of administration of the two instruments was counter-balanced across subjects as they entered the study; one instrument was administered to each subject at the initial entry into this study, and the other instrument was administered within 3–7 days. Administrations for each subject occurred at home and at the clinic, replicating the manner in which these instruments are typically administered. Scale scores were created using the customary scoring rules; the depression and non-specific symptom scores derived from the full SCL-90R were termed the “full-instrument” scores, while those scores derived from the RDC/TMD-based instrument were termed the “modified-instrument” scores. The scores for each construct were compared for internal consistency (did administration format affect how the underlying latent variable informed the respondent’s approach to the items as a whole, regardless of a change in item position within the instrument as administered?) and for metric equivalence (did administration format affect the total score?). The full report is available elsewhere (13).

Internal consistency was not affected by administration format; coefficient alpha, a statistic describing how well each of the items correlates to the total score of all of the items, was 0.95 for both the full-instrument depression measure and the modified-instrument depression measure, and it was 0.87 for both the full-instrument non-specific physical symptoms measure and for the modified-instrument non-specific physical symptoms measure. The concordance correlation coefficient (CCC) (49, 50) was used to assess agreement in total score for each measure as administered via the two formats. For assessing consistency in measures between instruments, a reliability coefficient that takes into account not only a necessary linear relationship but also close proximity to the line of unity (that is, a commensurate measure) is required (51, 52), and the intraclass correlation coefficient (ICC) is commonly used for this purpose. The CCC, compared to the ICC, has a reported advantage of being less affected by range of scores and thus, in principle, provides a statistic that is less influenced by sample characteristics. The CCC for depression was 0.96 and for non-specific physical symptoms was 0.91.

In conclusion, the depression and non-specific physical symptoms scales as implemented in the RDC/TMD are reliable and valid methods for assessing the respective constructs and the scores can be interpreted in the same manner as scores for the respective scales as obtained from administering the full SCL-90R.

Study 5: Oral Behaviors in the Laboratory

The Oral Behaviors Checklist is comprised of 21 items, with 2 items referring to behaviors that occur during the sleep stage and the remainder referring to behaviors that occur during the waking hours. Some of the original items, as administered in the Validation Project, contained a blend of concepts (e.g., “clench or press the teeth together”) which needed to be tested. A subset of behaviors, most of them visually unobservable, from the OBC was identified and tested in the laboratory using multichannel EMG. The first goal was to assess reliability of the behavior as understood by the subject, and the second goal was to assess how distinctive the behaviors were based on the pattern of associated EMG activity as determined by a vector multiple muscles. The data indicated that individuals understand each concept equally well, with correlation coefficients of at least 0.8 for most behaviors, indicating excellent reliability in performance. These reliability statistics were not systematically affected by the reported frequency, based on using the checklist, of these behaviors outside the laboratory. The full findings are available (16).

The right side and left side channels from each of the masseter and temporalis muscles were collapsed into single variables. A multivariate ANOVA, using the masseter, temporalis, and suprahyoid areas as dependent variables and the set of induced behaviors as the independent variable, was fit from the data. This analysis tests whether the variance defined by the 3-dimensional space as determined by the three dependent variables is less than the variance defined by the set of behaviors, and if significant, then the implication is that the specific behaviors, as identified by post-hoc testing, are distinguishable from each other with respect to how the behavior is actually produced, as measured by the EMG. The outcome of that analysis, as shown in Figure 2, displays a significant MANOVA with all behaviors distinguishable from each other except for ‘pressing the tongue’ and ‘holding the jaw to the side or forward’ which cannot be distinguished from each other based on the measured EMG. Not shown are the behavioral profiles from cases and from controls; this pair of profiles was visually indistinguishable and the formal statistical testing failed to demonstrate a difference in the behavioral profiles produced by cases vs those produced by controls. The full findings are available (15).

In sum, individuals reporting a high frequency of these behaviors or a low frequency of these behaviors utilize similar strategies for producing each behavior, as determined by magnitudes of EMG activity and pattern in producing each behavior. Each of the behaviors, with one exception, is associated with distinct patterns of EMG.

The final instrument will be constructed based on data from the Validation Project, from collaboration with the ACTA group in Amsterdam, and from the Orofacial Pain: Prospective Evaluation and Risk Assessment (OPPERA) study; collectively, these sources provide an extensive amount of data (n = approximately 5000) for developing a measurement scale that points to unconscious behaviors. The OPFERA data have only been recently cleaned, and now this study can proceed.

Study 6: Oral Behaviors in the Field

Further studies supported by this project included the collection of field data based on using an Ecological Momentary Assessment protocol (53). Individuals identified as those with high frequency of these behaviors (n=18) and individuals identified as those with low frequency (n=5), based on the use of the Oral Behaviors Checklist, were recruited; each individual carried a pocket computer for 14 hours/day for 7 days. At pseudorandom intervals, the computer prompted the individual to answer a short series of questions aimed at identifying the presence/absence of a sub-set of the behaviors contained within the checklist. The standard protocol of lockout if no response within a short designated time, coupled with reimbursement tied to responding to the prompts, resulted in high compliance with a mean response rate of 84% to the initial prompt. Initial analyses of these data were presented at the annual meeting of the International RDC/TMD Consortium, 2007. Further analyses of these data await completion of the instrument from the multiple datasets, as described under Study 5.

Study 7: Functional Limitation of the Jaw

The RDC/TMD contains a non-validated checklist of oral functions commonly believed to be affected by jaw pain, and data collected using this checklist served as the starting point for development of a formal instrument for assessing limitation of the masticatory system. From those data and using an item response model (20), the prototype of the Jaw Functional Limitation Scale (JFLS) was developed (17), and that prototype was compared to the RDC/TMD checklist and the Mandibular Functional Impairment Questionnaire (37). Based on concepts of limitation and disability as discussed elsewhere (3, 19), further instrument development proceeded using contemporary measurement methods (54). These methods are distinguished by an improvement in statistics regarding the performance of the individual item, commensurate measurement of both items and persons providing data, and explicitly incorporating a model-building approach to instrument development that links to the previously established methods of exploratory and confirmatory factor analysis. Particularly relevant for Study 7, these methods also provide numeric difficulty ratings for the selected items. Using a known groups design for assessing utility and validity, established diagnostic groups were assembled and a final instrument, the Jaw Functional Limitation Scale, was created. The subscales assessing mastication and jaw mobility have excellent measurement properties including cross-cultural equivalence across two settings (US, Sweden) (18). For the subscale assessing verbal and non-verbal communication, the measurement properties remain to be assessed in larger samples due to the problems identified with cross-cultural equivalence between the English and Swedish versions (55, 56).

With this level of instrument development as a starting point and noting that the mastication scale, as a result of the measurement methods used, is based on a series of foods that range in texture, Study 7 utilized a set of foods as suggested by the selected item content in the mastication scale of the JFLS. Each of the foods used in the JFLS mastication scale has a difficulty rating as determined by the Rasch statistical model used for instrument development; such ratings are a fundamental component of the model (20), and the goals of Study 7 were to assess whether ratings of difficulty in mastication, as reported directly by the subject in this experimental context, were consistent with the difficulty level determined

by the measurement model and to explore potential covariates underlying the subjects' ratings.

Subjects (TMD clinic cases, TMD community cases, and controls) chewed a food according to their own rhythm and extent, swallowed, and rated pain and limitation following the swallow. Preliminary analyses of real-time difficulty ratings reported by the subjects were consistent with the statistically-determined difficulty ratings associated with the different foods. This study has been expanded to include additional subject groups and further analyses will commence in the near future when the study is completed.

Summary

The Axis II component of the RDC/TMD Validation Project has determined that the original measures for depression and non-specific physical symptoms, the measure for typical pain intensity (characteristic pain intensity) and the measure for pain-related disability classification (Graded Chronic Pain Scale) are reliable and valid with respect to selected reference standards and for use in this patient population. The depression and non-specific physical symptoms measures as used in the RDC/TMD yield data that are comparable to that obtained by using the full SCL-90R instrument. The depression measure has good to excellent clinical utility as a screener for psychiatric morbidity while non-specific physical symptoms is not useful as a screener with regards to psychiatric morbidity; the latter instrument requires further investigation in terms of better clarifying its potential utility as a screener. However, prior studies of measures related to the non-specific physical symptoms construct (33, 35, 57) as well as the emerging focus on multi-symptom presentation (58–62) clearly support the original concepts behind the inclusion of the particular non-specific physical symptoms measure into the RDC/TMD. These findings indicate that Axis II legacy data collected using the RDC/TMD protocol remain valid and interpretable. Selection of reference standards for evaluating Axis II instruments is far from trivial, and other investigators could easily emerge with different findings via selecting a different reference standard; it is here that further conceptual development regarding the fundamental phenomenology of “being a person in pain” is needed.

In terms of new instruments to add to the Axis II protocol, the JFLS is a logical choice as a replacement instrument for the limitation checklist that was part of the RDC/TMD, and data to date have been strong in supporting its reliability, validity, sensitivity to change, translatability, and cross-cultural equivalence for mastication and mobility. A further area of exploration lies in the impact of pain, more broadly, and TMD, more specifically, on verbal and non-verbal communication from the perspective of how those vital functions are viewed from a cross-cultural perspective; early data clearly indicate that individuals in different cultures experience this aspect of the masticatory system differently.

The research team was requested by the Project's External Advisory Panel to develop an instrument to assess oral behaviors; preliminary analyses of sample data and in particular more detailed data emerging from laboratory and field studies indicate that the concepts initially placed on paper were good and that, moreover, a paper-and-pencil instrument for the assessment of this construct exhibits the desirable characteristics of standardized

instruments: better measurement. As a result, it may be possible to explore in a more systematic manner and with less measurement error the role that overuse behaviors has in TMD onset and maintenance.

Finally, the scope of instruments administered in the Validation Project provides an opportunity to improve on the comprehensiveness of assessing Axis II beyond the role of just screening, leading to further improvements in measuring the biobehavioral dimension of the RDC/TMD. The RDC/TMD was considered a model system at the time of its development (63); we are in hopes that revisions made to the RDC/TMD based on both this project as well as all of the methods studies conducted over the past 20 years will again lead to an assessment protocol for the clinic and research setting that will be considered exemplary. Subsequent to this workshop, a consensus workshop was held in which further recommendations were made for expanding Axis II (5), and after that, another workshop expanded these recommendations for application of the biobehavioral principles into general dental settings (48). Collectively, these workshops are contributing substantially to the progress in improving the biobehavioral assessment domain of the RDC/TMD.

The foundations for many of the studies described above were provided by the substantial contributions from other research groups, and this report of progress in a research program would be remiss if that seminal work in the TMD field was not acknowledged (64–71).

Next steps

Our next steps, based on data already available, will consist of the following:

- Improve effectiveness and efficiency of a focused screener for detecting distress and other correlates of factors that interfere with expected treatment response.
- Expand Axis II with other constructs for consistent use of the RDC/TMD in theoretical research and research on the RDC/TMD.
- Augment Axis II via incorporation of standardized outcomes measures for consistent usage in clinical trials, based on the recommendations from IMMPACT (72, 73).

And directions for future studies, based on the above considerations, include:

- Identify biobehavioral risk factors for progression or change in Axis I diagnosis over time.
- Evolve new self-report instruments for assessing correlates of basic CNS functions: sensory, affective, cognitive, autonomic, motor – in order to augment the current assessment of mood and somatic symptom experience.
- Address complexity associated with phenotypic characterization based on the Axis II measures.
- Pursue the interaction between somatic diagnosis and person status.

- Expand the axial framework to include additional dimensions of CNS function, pain sensitivity and response by the pain transmission and modulatory systems, and genetic influences on pain processing.
- Improve the structure of how clinical interview, evaluation through use of standardized instruments, and targeting outcomes assessment can lead to better tailored evaluation for clinical use.

Acknowledgments

The entire investigation group comprised of individuals from three universities have contributed to the data and ideas represented in this report. Specific individuals have contributed substantially to the development of specific ideas and data analyses as reported here. The author wishes to thank Sam D work in and Jeff Sherman for their continued involvement and help during the design and data collection phases of this project, Judy Turner for inspired and critical reactions to the project at the manuscript stage, Lloyd Mancil for invaluable guidance regarding analytic issues, and Thomas List for providing helpful comments regarding the major Axis II manuscript. Sharon Michalovic provided outstanding management of the additional studies conducted at the University at Buffalo. Marylee van der Meulen has graciously provided the ACTA data described under Study 5. Support by NIDCR (U01-DE013331) is recognized as pivotal for the conduct of this project and its extensions.

Reference List

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. Platt JR. Strong inference. *Science.* 1964; 146:347–353. [PubMed: 17739513]
3. World Health Organization International Classification of Impairments, Disabilities, and Handicaps. Geneva: World Health Organization; 1980.
4. Ohrbach R, List T, Goulet J-P, Svensson P. International Consensus Workshop: Convergence on an Orofacial Pain Taxonomy: Workshop Recommendations. 2010 <http://www.rdc-tmdinternational.org/Default.aspx?tabid=98>.
5. Ohrbach R, List T, Goulet J-P, Svensson P. Recommendations from the International Consensus Workshop: Convergence on an Orofacial Pain Taxonomy. *J Oral Rehabil.* 2010 In press.
6. Turner JA, Dworkin SF. Screening for psychosocial risk factors in patients with chronic orofacial pain: recent advances. *JADA.* 2004 Aug; 135(8):1119–1125. [PubMed: 15387050]
7. Fishbain DA. The association of chronic pain and suicide. *Seminars in Clinical Neuropsychiatry.* 1999; 4:221–227. [PubMed: 10498789]
8. Derogatis LR, Lipman RS, Covi L. SCL-90: an outpatient psychiatric rating scale--preliminary report. *Psychopharmacology.* 1973; 9:13–28.
9. Von Korff M, Ormel J, Keefe FJ, Dworkin SF. Grading the severity of chronic pain. *Pain.* 1992; 50:133–149. [PubMed: 1408309]
10. Ohrbach R, Turner JA, Sherman JJ, Mancil LA, Truelove EL, Schiffman EL, et al. Research Diagnostic Criteria for Temporomandibular Disorders: Evaluation of Psychometric Properties of the Axis II Measures. *J Orofacial Pain.* 2010 In press.
11. Schiffman EL, Truelove EL, Ohrbach R, Anderson GC, John MT, List T, et al. Assessment of the Validity of the Research Diagnostic Criteria for Temporomandibular Disorders: I: Overview and Methodology. *Journal of Orofacial Pain.* 2010; 24:7–24. [PubMed: 20213028]
12. Dworkin SF, Sherman JJ, Mancil L, Ohrbach R, LeResche L, Truelove E. 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]
13. Ohrbach R, Sherman J, Beneduce C, Zittel-Palamara K, Pak Y. Extraction of RDC/TMD subscales from the Symptom Checklist-90: Does context alter respondent behavior? *J Orofacial Pain.* 2008; 22:331–339. [PubMed: 19090406]

14. Derogatis, LR. SCL-90-R: Symptom Checklist-90-R. Administration, Scoring and Procedures Manual. Minneapolis, MN: National Computer Systems; 1994.
15. Ohrbach R, Markiewicz MR, McCall WD Jr. Waking-state oral parafunctional behaviors: specificity and validity as assessed by electromyography. *European Journal of Oral Sciences*. 2008; 116:438–444. [PubMed: 18821986]
16. Markiewicz MR, Ohrbach R, McCall WD Jr. Oral Behaviors Checklist: Reliability of Performance in Targeted Waking-state Behaviors. *J Orofacial Pain*. 2006; 20:306–316. [PubMed: 17190029]
17. Ohrbach R, Granger CV, List T, Dworkin SF. 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.
18. 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. [PubMed: 18780535]
19. Ohrbach R. Disability assessment in temporomandibular disorders and masticatory system rehabilitation. *J Oral Rehabil*. 2010 In press.
20. Andrich, D. Rasch Models for Measurement. Newbury Park, CA: Sage Publications; 1988.
21. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. *App Psych Measurement*. 1977; 1:385–401.
22. Goldberg, D.; Williams, P. A User's Guide to the General Health Questionnaire. Windsor, Berkshire, England: nfer-Nelson Publishing Company; 1988.
23. Kerns RD, Turk DC, Rudy TE. The West Haven-Yale Multidimensional Pain Inventory (WHYMPI). *Pain*. 1985; 23:345–356. [PubMed: 4088697]
24. Rudy, TE. Multidimensional Pain Inventory Version 3.0 User's Guide. Pittsburgh, PA: University of Pittsburgh; 2005. www.pain.pitt.edu/mpi
25. Ware, JE., Jr; Kosinski, M.; Turner-Bowker, DM.; Gandek, B. How to Score Version 2 of the SF-12 Health Survey. Lincoln, RI: Quality Metric Incorporated; 2002.
26. Robins LN, Helzer JE, Ratcliff KS, Seyfried W. Validity of the diagnostic interview schedule, version II: DSM-III diagnoses. *Psychol Med*. 1982 Nov; 12(4):855–870. [PubMed: 7156256]
27. Robins, LN.; Cottler, LB.; Bucholz, KK.; Compton, WM.; North, CS.; Rourke, KM. Diagnostic Interview Schedule for the DSM-IV (DIS-IV). St Louis, MO: Washington University School of Medicine; 2000.
28. American Psychiatric Association Diagnostic and Statistical Manual of Mental Disorders - IV. 4. Washington, D.C: American Psychiatric Association; 1994.
29. Bucholz KK, Robins LN, Shayka JJ, Przybeck TR, Helzer JE, Goldring E, et al. Performance of two forms of a computer psychiatric screening interview: version I of the DISSI. *J psychiat Res*. 1991; 25(3):117–129. [PubMed: 1941708]
30. Escobar JI, Manu P, Matthews D, Lane T, Swartz M, Canino G. Medically unexplained physical symptoms, somatization disorder and abridged somatization: studies with the Diagnostic Interview Schedule. *Psychiat Dev*. 1989; 7(3):235–245.
31. Escobar JI, Rubio-Stipec M, Canino G, Karno M. 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 Mar; 177(3):140–146. [PubMed: 2918297]
32. Escobar JI, Golding JM, Hough RL, Karno M, Burnam MA, Wells KB. Somatization in the community: relationship to disability and use of services. *Am J Public Health*. 1987 Jul; 77(7): 837–840. [PubMed: 3592038]
33. Katon W, Line E, Von Korff M, Russo J, Lipscomb P, Bush T. Somatization: a spectrum of severity. *Am J Psychiat*. 1991; 148:34–40. [PubMed: 1984704]
34. Sackett, DL.; Haynes, RB.; Guyatt, GH.; Tugwell, P. Clinical Epidemiology: A Basic Science for Clinical Medicine. 2. Boston: Little, Brown, and Company; 1991.
35. Dworkin SF, Von Korff MR, LeResche L. Multiple pains and psychiatric disturbance: An epidemiologic investigation. *Arch Gen Psychiat*. 1990; 47:239–244. [PubMed: 2306165]
36. Diatchenko L, Nackley AG, Slade GD, Fillingim RB, Maixner W. Idiopathic pain disorders - pathways of vulnerability. *Pain*. 2006; 123:226–230. [PubMed: 16777329]

37. Stegenga B, de Bont LGM, de Leeuw R, Boering G. Assessment of mandibular function impairment associated with temporomandibular joint osteoarthritis and internal derangement. *J Orofacial Pain*. 1993; 7:183–195. [PubMed: 8358365]
38. Spielberger CD, Vagg PR. Psychometric properties of the STAI: a reply to Ramanaiah, Franzen, and Schill. *J Person Assess*. 1984; 48:95–97.
39. Cohen S, Kamarck T, Mermelstein RA. global measure of perceived stress. *J Health Human Behav*. 1983; 24:385–396.
40. Melzack R. The McGill Pain Questionnaire: major properties and scoring methods. *Pain*. 1975; 1:277–299. [PubMed: 1235985]
41. Buysse DJ, Reynolds CF III, Monk TH, Hoch CC, Yeager AL, Kupfer DJ. Quantification of subjective sleep quality in healthy elderly men and women using the Pittsburgh Sleep Quality Index (PSQI). *Sleep*. 1991 Aug; 14(4):331–338. [PubMed: 1947597]
42. First, MB.; Gibbon, M.; Williams, JBW.; Spitzer, RL.; Benjamin, LS. Computer-Assisted SCID II Expert System (CAS II ES). Tonawanda, NY: Multi-Health Systems; 2004.
43. First MB, Spitzer RL, Gibbon M, Williams JBW. The Structured Clinical Interview for DSM-III-R Personality Disorders (SCID-II): II. Multi-site test-retest reliability study. *Journal of Personality Disorders*. 1995; 9(2):92–104.
44. First MB, Spitzer RL, Gibbon M, Williams JBW. The Structured Clinical Interview for DSM-III-R Personality Disorders (SCID-II): I. Description. *Journal of Personality Disorders*. 1995; 9(2):83–91.
45. Engelhart C, Eisenstein N, Johnson V, Wolf J, Williamson J, Steitz D, et al. Factor structure of the Neurobehavioral Cognitive Status Exam (COGNISTAT) in healthy, and psychiatrically and neurologically impaired, elderly adults. *Clinical Neuropsychologist*. 1999 Feb; 13(1):109–111. [PubMed: 10937653]
46. Kiernan RJ, Mueller J, Langston JW, Van DC. The Neurobehavioral Cognitive Status Examination: a brief but quantitative approach to cognitive assessment. *Ann Intern Med*. 1987 Oct; 107(4):481–485. [PubMed: 3631786]
47. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine*. 2001 Sep; 16(9):606–613. [PubMed: 11556941]
48. Cairns B, List T, Michelotti A, Ohrbach R, Svensson P. Workgroup recommendations from the Siena JOR CORE. *J Oral Rehabil*. 2010 In press.
49. Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*. 1989; 45:255–268. [PubMed: 2720055]
50. Lin LI. A note on the concordance correlation coefficient. *Biometrics*. 2000; 56:324–323.
51. Bartko JJ, Carpenter WT Jr. On the methods and theory of reliability. *J Nerv Ment Dis*. 1976; 163:307–317. [PubMed: 978187]
52. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986; 1:307–310. [PubMed: 2868172]
53. Stone, AA.; Turkkan, JS.; Bachrach, CA.; Jobe, JB.; Kurzman, HS.; Cain, VS. The Science of Self-Report: Implications for Research and Practice. Mahwah, NJ: Lawrence Erlbaum; 2000.
54. Hays RD, Morales LS, Reise SP. Item response theory and health outcomes measurement in the 21st century. *Medical Care*. 2000 Sep; 38(9 Suppl):II28–II42. [PubMed: 10982088]
55. Beaton, D.; Bombardier, C.; Guillemin, F.; Ferraz, MB. Recommendations for the cross-cultural adaptation of health status measures. Toronto: Institute of Work and Health, and American Academy of Orthopaedic Surgeons; 2002.
56. Ohrbach R, Bjorner J, Jezewski MA, John MT, Lobbezoo F. Guidelines for Establishing Cultural Equivalency of Instruments. 2007 Available at URL www.rdc-tmdinternational.org.
57. Macfarlane GJ, Croft PR, Schollum J, Silman AJ. Widespread pain: is an improved classification possible? *J Rheumatol*. 1996 Sep; 23(9):1628–1632. [PubMed: 8877936]
58. Allen LA, Escobar JI, Lehrer PM, Gara MA, Woolfolk RL. Psychosocial treatments for multiple unexplained physical symptoms: A review of the literature. *Psychosom Med*. 2002; 64:939–950. [PubMed: 12461199]

59. Rief W, Hessel A, Braehler E. Somatization symptoms and hypochondriacal features in the general population. *Psychosom Med*. 2001; 63:595–602. [PubMed: 11485113]
60. Rantala MA, Ahlberg J, Suvinen TI, Nissinen M, Lindholm H, Savolainen A, et al. Temporomandibular joint related painless symptoms, orofacial pain, neck pain, headache, and psychosocial factors among non-patients. *Acta Odontologica Scandinavica*. 2003 Aug; 61(4):217–222. [PubMed: 14582589]
61. 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 Apr; 256(3):187–196. [PubMed: 16328107]
62. McBeth J, Chiu YH, Silman AJ, Ray D, Morriss R, Dickens C, et al. Hypothalamic-pituitary-adrenal stress axis function and the relationship with chronic widespread pain and its antecedents. *Arthritis Research & Therapy*. 2005; 7(5):R992–R1000. [PubMed: 16207340]
63. Garofalo JP, Wesley L. Research Diagnostic Criteria for Temporomandibular Disorders: Reflection of the Physical-Psychological Interface. *APS Bulletin*. 1997 May-Jun; 4–8.
64. Dworkin SF. Illness behavior and dysfunction: review of concepts and application to chronic pain. *Can J Physiol Pharmacol*. 1991; 69:662–671. [PubMed: 1863918]
65. Dworkin SF, Von Korff MR, LeResche L. Epidemiologic studies of chronic pain: a dynamic-ecologic perspective. *Ann Behav Med*. 1992; 14:3–11.
66. Gatchel RJ, Garofalo JP, Ellis E, Holt C. Major psychological disorders in acute and chronic TMD: an initial examination. *JADA*. 1996; 127:1365–1374. [PubMed: 8854613]
67. Epker J, Gatchel RJ, Ellis EI. A model for predicting chronic TMD: Practical application in clinical settings. *JADA*. 1999; 130:1470–1475. [PubMed: 10570591]
68. Kight M, Gatchel RJ, Wesley L. Temporomandibular disorders: evidence for significant overlap with psychopathology. *Health Psychol*. 1999 Mar; 18(2):177–182. [PubMed: 10194053]
69. Turk DC, Flor H. Pain > pain behaviors: the utility and limitations of the pain behavior construct. *Pain*. 1987; 31:277–298. [PubMed: 3320882]
70. Turk DC, Rudy TE. The robustness of an empirically derived taxonomy of chronic pain patients. *Pain*. 1990; 43:27–35. [PubMed: 2148974]
71. Turk DC, Rudy TE. Cognitive factors and persistent pain: a glimpse into Pandora's Box. *Cog Ther Res*. 1992; 16:99–122.
72. Turk DC, Dworkin RH, Allen RR, Bellamy N, Brandenburg N, Carr DB, et al. Core outcome domains for chronic pain clinical trials: IMMPACT recommendations. *Pain*. 2003; 106:337–345. [PubMed: 14659516]
73. Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. *Pain*. 2005; 113:9–19. [PubMed: 15621359]

Validation Project Axis II Studies: Overview

Figure 1.

Flow of participants from recruitment sources and across studies. Study site for the different stages of the Axis II studies are also indicated. Heavy lines with arrows signify pathway for the core Validation Project by which participants were recruited for the main Axis II study (Study 1) and the primary studies for examining specific psychometric properties (Studies 2a, 2b, 2c) or utilizing additional data (Study 3). Solid thin lines with arrows represent recruitment of psychologically symptomatic patients or TMD patients (Study 4), TMD clinic patients (Studies 5, 6, 7), or community cases (optional for Studies 5 or 6; mandatory for

Study 7). Dashed lines with arrows signify recruitment of healthy controls into Studies 5, 6, and 7. Study 1 and Study 3 were conducted at all three study sites: University at Buffalo (UB), University of Minnesota, and University of Washington (UW). Study 2 was conducted at only UB and UW, and Studies 4, 5, 6, and 7 were conducted at only UB.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 2.

Plot from profile analysis of log EMG activity from each muscle by task demonstrating significant differences across tasks based on 3-dimensional variate formed from measured electromyographic (EMG) activity of masseter, temporalis, and suprahyoid muscle units (Wilks lambda=.07034, $F(27, 20)=9.7906$, $p=.00000$). Experimental groups (cases, controls) are collapsed together. Y-axis refers to natural log-transformed values of EMG activity in microvolts; vertical bars denote 0.95 confidence intervals. See text for further description of each task.

Table 1

Reliability and consistency of core RDC/TMD Axis II measures.

	Internal consistency		Temporal stability	
	Cronbach's alpha	Lower-bound 95% CI	Lin's CCC	95% CIs
Depression	0.91	0.90	0.78	0.69 – 0.88
Non-specific physical symptoms	0.84	0.82	0.72	0.59 – 0.84
Characteristic pain intensity	0.84	0.82	0.91	0.87 – 0.95
Pain interference	0.95	0.94	0.89	0.83 – 0.94
			Kappa	
Chronic pain grade	N/A	N/A	0.87	0.77 – 0.98

Table 2

Convergent and discriminant validity: Associations 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 [/] (-0.10, 0.07)	-0.08 [/] (-0.16, 0.00)	-0.07 [/] (-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 [/] (-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, -0.65)	-0.43 (-0.49, -0.36)	-0.40 (-0.47, -0.33)	-0.09 [/] (-0.18, -0.00)	-0.20 (-0.30, -0.12)	-0.12 (-0.21, -0.03)

Convergent validity is shown in the shaded boxes which indicate pairs expected to show higher associations; discriminant validity is shown in the nonshaded boxes which indicate pairs expected to show lower associations. Values shown are the association statistics and the 95% CIs in parentheses. The Spearman rank correlation coefficient was used for associations with Chronic Pain Grade, and Lin's CCC was used for associations between all other Axis II measures. All parameter estimates were adjusted for study site. See text for further information about the measures.

[/]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 3

Temporal Stability of Axis II Measures

Measure	Lin's CCC or Kappa ^I
Depression	0.78
Nonspecific Physical Symptoms, with pain items	0.72
Nonspecific Physical Symptoms, without pain items	0.63
Characteristic Pain Intensity	0.91
Interference	0.89
Number of Disability Days	0.74
Chronic Pain Grade	0.87

^I Kappa used for Chronic Pain Grade reliability; CCC was used for all other measures.

Table 4

Axis II Screening Instruments: Sensitivity and Specificity in Identifying Current-year Psychiatric Diagnoses

Axis II measures and cut-points	Criterion Psychiatric Status: Current Year	
	Sensitivity	Specificity
Depression		
Normal vs moderate-severe	87%	53%
Normal-moderate vs severe	56%	91%
Non-specific physical symptoms		
Normal vs moderate-severe	86%	31%
Normal-moderate vs severe	68%	68%