Development of a brief and effective temporomandibular disorder pain screening questionnaire:

Reliability and validity

Dr. Yoly M. Gonzalez, DDS, MS, MPH,
Assistant professor, Department of Oral Diagnostic Sciences, School of Dental Medicine, University at Buffalo, 3435 Main St., 355 Squire Hall, Buffalo, N.Y. 14214

Dr. Eric Schiffman, DDS, MS,
Associate professor, Department of Diagnostic and Biological Sciences, School of Dentistry, University of Minnesota, Minneapolis

Dr. Sharon M. Gordon, DDS, MPH, PhD,
Associate professor, Department of Oral-Maxillofacial Surgery, School of Dentistry, University of Maryland, Baltimore

Dr. Bradley Seago, DDS,
Dental student, Department of Oral Diagnostic Sciences, School of Dental Medicine, University at Buffalo, N.Y., when this article was written. He now is a dentist, Colorado Community Dental, Grand Junction, Colo

Dr. Edmond L. Truelove, DDS, MDS,
Professor, Department of Oral Medicine, School of Dentistry, University of Washington, Seattle

Dr. Gary Slade, DDS, PhD, and
Professor, Department of Dental Ecology, School of Dentistry, University of North Carolina at Chapel Hill

Dr. Richard Ohrbach, DDS, PhD
Associate professor, Department of Oral Diagnostic Sciences, School of Dental Medicine, University at Buffalo, N.Y

Abstract

Background—Available screening instruments for identifying temporomandibular disorders (TMDs) exhibit methodological or logistic limitations. The authors conducted a study to develop and assess the validity of a self-report instrument in screening patients for pain-related TMDs.

Methods—By using psychometric methods for item selection, the authors developed short (three-item) and long (six-item) versions of the questionnaire and evaluated them for validity among 504 participants.

Address reprint requests to: Dr. Gonzalez, ymg@buffalo.edu.

Results—Internal reliability was excellent, with coefficient α values of 0.87 and 0.93 for the short and long versions, respectively. When the authors dichotomized instrument scores at optimal thresholds, both versions had a sensitivity of 99 percent and a specificity of 97 percent for correct classification of the presence or absence of TMD. The specificity was at least 95 percent in the correct identification of people with nonpainful TMJ disorders or headahce without TMD pain.

Conclusions—With use of appropriate psychometric methodology, the selected items exhibited excellent content validity. The excellent levels of reliability, sensitivity and specificity demonstrate the validity and usefulness of this instrument.

Clinical Implications—Using this instrument will allow clinicians to identify more readily—and cost-effectively—most patients with painful TMD conditions for whom early and reliable identification would have a significant effect on diagnosis, treatment and prognosis.

Keywords

Temporomandibular disorders; temporomandibular joint; pain; screening; diagnostic validity

Clinicians increasingly are using disease screening instruments, or "screeners," in clinical settings for triage. Primary care providers can obtain patients' self-reported symptoms efficiently with use of a screener and, if the results are positive, then more comprehensively assess signs and symptoms. Screeners must have adequate reliability and validity. Validity is best judged by identifying true- and false-negative results and true- and false-positive results relative to a reference standard diagnosis, thereby yielding diagnostic validity statistics such as sensitivity, specificity, predictive value and likelihood ratios. In addition, a screening instrument must be evaluated for its utility to determine whether it will demonstrate clear benefits in terms of either community or individual health.

The assessment of temporomandibular disorders (TMDs) is recognized as integral to the practice of dentistry.[1,2] Consequently, investigators have developed a variety of instruments to accomplish the first step of evaluation. TMD screeners could help provide answers to a variety of questions important for patient care in general dental practice, such as whether a patient should be referred for diagnosis and care by a health care provider with training in orofacial pain and whether a patient is a likely candidate for intervention. A valid and reliable TMD screener also could be useful in general population settings. In some special populations, a TMD screener might make a valuable contribution to determining a patient's fitness for forthcoming tasks—such as military personnel who are to be deployed to areas in which treatment of an acute episode of pain is difficult logistically. In each of these settings, the relative importance of false-negative and false-positive results justifiably would vary according to the triage setting. In screening a patient for fitness, the clinician may find false-negative results more relevant than false-positive results, whereas false-negative results may be more desirable than false-positive results in making referral decisions.

Findings from previous studies have demonstrated the feasibility of TMD screening,[1,3–8] although there are several notable problems with the instruments used (Table 1). Reference standards varied considerably in quality, and none used all three Standards for Reporting of Diagnostic Accuracy–recommended parameters of assessment: operationalized criteria,

examiners using a calibrated technique and consensus diagnosis.[9] Some instruments combined items for both pain-related TMD and mechanical symptoms associated with the temporomandibular joint (TMJ), but they did not discriminate between these two major subtypes of TMD. This discrimination is critical because treatments and prognosis for the two types of problems differ. In summary, each of the existing screening instruments exhibits limitations in at least one of the following areas: psychometric properties and diagnostic statistics, application in representative populations, multidimensionality of items and use of a reference standard diagnosis.

These shortcomings motivated us to develop our screening instrument, which emerged from several related studies of participants who were phenotyped extensively. Our design goals for this instrument were to

- restrict content domain to items that are relevant for pain-related TMD diagnoses only;

- use comprehensive psychometric methods for item selection and development;

- keep the instrument short and easy to score;

- account for the context-specific aspects associated with reporting pain symptoms;

- identify additional symptoms relevant for a TMD pain diagnosis;

- assess the instrument's usefulness in distinguishing related diagnoses of nonpainful TMJ disorders and headache.

Because TMD has clinical characteristics that overlap with those of most other orofacial pain conditions, we included a group of participants with odontalgia to identify a source of potential false-positive responses to our screening instrument.

METHODS

Participants

We obtained data from two convenience samples (Figure 1, page 1186). The first sample consisted of 732 adults at three academic health centers in the Research Diagnostic Criteria for TMD (RDC-TMD) Validation Project whom we characterized by using the expanded RDC-TMD assessment protocol.[10] Case status was based on consensus by two dentists at each site (among them Y.M.G. and R.O. at the University at Buffalo, N.Y., E.S. at the University of Minnesota, Minneapolis, and E.L.T. at the University of Washington, Seattle) using calibrated technique.[10] We recruited putative control participants on the basis of absence of pain in the facial area during the preceding six months. At the time of participants' enrollment, we evaluated them according to history, clinical examination and panoramic radiographic findings to exclude people with any possibility of odontogenic pain. We used this first sample for initial item development. Among the participants with TMD, we included 65 in a reliability assessment, with an interval of two to seven days between survey administrations to evaluate temporal stability.[11]

For validity testing, we divided the 732 participants into four groups. We defined the target group, those having pain-related TMD, as those having a diagnosis of pain-related TMD

(that is, myofascial pain, arthralgia or both) (as described by Schiffman and colleagues[10]). We identified two comparison groups without pain. One of them consisted of healthy control participants, defined as not meeting criteria for a diagnosis of TMD; exclusion criteria at enrollment permitted only low-severity headaches (per International Classification of Headache Disorders, second edition [ICHD-II], criteria[12]) that were not affected by masticatory function. The second comparison group consisted of those with a nonpainful TMJ disorder, defined as a TMJ disorder (such as disk displacement or osteoarthrosis) identified via magnetic resonance imaging or computed tomography, and this group served as a comparison for reporting of masticatory system symptoms. Participants in this latter group may have had jaw pain symptoms, but we required that those symptoms be insufficient to meet criteria for a diagnosis of TMD pain. We identified a third comparison group—those with headache in the temple region—by means of an algorithm from the ICHD-II criteria.[12] This group was a subset of the healthy control participants and those with nonpainful TMJ disorders, and we selected it on the basis of the absence of a diagnosis of TMD pain. Consequently, these participants represented those with regional headache without TMD pain and served as a comparison for pain symptom reporting. To create groups with similar sample sizes, we randomly selected a subset of the participants with pain-related TMD and retained it for analyses.

Another pain group, that with odontalgia, consisted of 80 participants whose chief complaint was toothache and odontogenic disease confirmed by means of clinical examination and radiographs. We did not determine the presence or absence of TMD in this group owing to logistic limitations, so we used these data for secondary analyses to determine the false-positive rate associated with a competitive pain condition.

Initial instrument development

The investigators in the Validation Project[10] developed a comprehensive self-administered symptom instrument that contained 49 items assessing masticatory system pain. These items inquired into pain symptoms on each side of the face or in each of the designated areas (jaw, temple, jaw joint, ear) with a range of response options. Many of the items were conditional on response to a prior filter item inquiring about pain in the preceding month. If a person responded "no" to the filter items, the instrument directed him or her to skip to the next major section; however, many responded "no" to the filter item but completed the subsequent pain items within the skip range as though their response to the filter question had been "yes." This type of seemingly contradictory behavior, according to the results of participant interviews, typically occurred because of the relation of the "no" response to the simplicity of the filter question, whereas "yes" responses to the subsequent items within the skip range occurred because the greater level of detail regarding pain elicited recognition of their applicability.

For this first part of item analysis, we accepted responses "as is" in the data set to assess the maximal information provided by the items and ignore any possible misclassification of response related to following or ignoring a skip pattern. To assess each item, we used TestGraf (McGill University, Montreal),[13] a nonparametric approach for discovering response behaviors among people. TestGraf makes few assumptions about the

dimensionality of the item, and the response curves are constructed on the basis of the presumed severity of the symptoms experienced by the people providing responses. The goal was to simplify the items to produce new items that clearly assessed one construct and that were reliable in terms of item-response probability curves.[14–16] We explored each of the 49 items and constructed an interim 18-item instrument. Importantly, we could not reliably distinguish reports of pain arising from only the TMJ versus those of pain arising from only the masseter area, and so we combined these anatomical areas, which led to better reliability. Furthermore, we reduced the response options to no more than two beyond the null response.

We administered this interim instrument to the participants with odontalgia to assess responses to these items on the basis of a competing pain diagnosis.

Final instrument development and statistical analysis

We considered both item-response and factor-analytic approaches appropriate techniques for exploration of further item reduction of the interim 18-item instrument. We used Rasch analysis[17,18] to determine if a response hierarchy was present among the items but found no discernible hierarchical pattern of severity among the items. We then used exploratory-factor analysis (EFA) to determine if an indicator variable model would identify a latent construct (a non–directly measurable concept such as TMD pain). We performed EFA by using a polychoric correlation matrix (because of the binary or ordinal response options) and weighted least squares estimator; for simplified scoring, a single factor solution was desirable, with an eigenvalue of greater than 1.0 required for factor identification.[19]

We constructed short and long versions of the instruments and tested them across the different groups. We defined "case participants" as those with pain-related TMD, whereas the comparison control groups were healthy control, nonpainful TMJ and headache. By using traditional 2×2 tables, we computed sensitivity as correct identification of case participants and specificity as correct identification of control participants. Additional statistics included likelihood ratios and area under the curve (AUC) from the receiver operating characteristic (ROC) analysis. We computed the response rate of above-threshold scores by the odontalgia group. Among the software we used were Test-Graf 98[13]; Winsteps, Version 3.68.0[20]; Mplus, Version 5.0[21]; and Stata, Version 11.[22] We compared demographic characteristics by using robust errors (as described by Wilcox[23]) for age, owing to unequal variances, and by using the proportion test for sex.[24]

RESULTS

Table 2 describes group statistics for sample sizes, age and sex distribution. A two-factor EFA using all 18 items explained 60.6 percent of the variance. We found substantial overlap in factor structure assessing modification of pain in the temporalis and masseter/TMJ area. Therefore, we combined the items pertaining to these three anatomical areas, resulting in 11 items. An EFA of the 11 items resulted in a single factor solution with 66.7 percent variance explained. Six of the retained items assessed core symptoms (Figure 2) on the basis of emerging definitions of TMD pain diagnosis,[25] which requires each of two findings: pain of sufficient frequency across a recent period and modification of the pain by function. Table 3

(page 1189) shows their contribution. We explored an additional item—pain modified by resting the jaw—but its inclusion did not improve the model. Therefore, we selected a single-factor solution based on six items for subsequent testing. It explained 76.8 percent of the variance, which is considered an excellent fit. We also created a shorter form (Figure 2), and we found that these three items explained 83.0 percent of the variance within a single-factor solution.

We based scores for the three-item and six-item instruments on simple sums (Figure 2 shows the scoring rules), and we then evaluated the instruments for other characteristics. We computed internal consistency using the composite data set, and coefficient α was 0.87 and 0.93 for the short and long versions, respectively. When we assessed temporal stability, we found the reliability of the individual items (κ) ranged from 0.52 to 0.78, indicating fair to excellent agreement, whereas the summary scores of the short and long versions exhibited acceptable intraclass correlation values of 0.83 and 0.79, respectively.

Table 4 (pages 1190–1191) presents information regarding sensitivity and specificity of the short and long versions of the screener. We established threshold values for a positive score as 2 for the short version and 3 for the long version. These threshold values provided for each of the versions, respectively, the best balance between sensitivity and specificity with use of ROC analysis. Both versions exhibited excellent validity in correct identification of participants with pain-related TMD (sensitivity, 99 percent) and healthy control participants (specificity, 97 percent). The validity of the short screener was excellent in correct identification of participants with competing symptom conditions of non-painful TMJ disorder (specificity, 95 percent) and headache not related to TMD (specificity, 96 percent). The validity of the long screener was slightly better in identification of people with nonpainful TMJ and with headache. The response rate of the odontalgia group was 29 percent and 26 percent for the short and long screeners, respectively.

Represented by the AUC, the probability that a test result from a randomly selected pair of participants with pain-related TMD, versus results from healthy control participants or participants with nonpainful TMJ disorder, would correctly identify the true-positive and true-negative participants was 98 to 99 percent. The AUC values remained high for the comparison with headache. The positive likelihood ratio obtained ranged from 19.2 to 44.6, varying according to the comparison control group, whereas the negative likelihood ratio was 0.01 throughout. Both positive and negative ratio findings exceeded the accepted benchmarks of 10 or more and 0.1 or less for the positive and negative likelihood ratios, respectively.

DISCUSSION

In this study, we synthesized information from responses to a large number of questions and tested the questions' performance in relevant diagnostic and control groups to derive two screening questionnaires for pain-related TMD, one with three items and one with six items. Both versions had excellent validity when judged against a reference standard diagnosis with use of the expanded RDC-TMD protocol. Our primary intent in developing the three-item TMD screener was to use it in a cost- and time-effective manner in population-based

studies and other research applications in which aggregate group statistics are an acceptable outcome. The intent with the six-item TMD screener was to enhance its internal reliability (and thereby increase its precision) for use in clinical settings. Also, the three additional items in the instrument could be used as part of the formal diagnosis of a pain-related TMD consistent with emerging diagnostic criteria for TMD,[25] which require positive modifying factors for a TMD pain diagnosis. In population settings, one could use the six-item version if more certainty about the likely diagnosis is desired. With the use of either version of the screener, we recognized that the clinician must perform a comprehensive pain assessment, including interview and clinical examination, if the goal is to provide a diagnosis.

Similarities in pain characteristics across different pain-related disorders that share the same anatomical region and, hence, involve some of the same functional structures challenge the process of differential diagnosis. Moreover, these similarities equally challenge the development of screening instruments capable of detecting a pain condition of interest, simultaneous with the ability to discriminate among different pain conditions that often coexist. Consequently, absolute detection of pain-related TMD and absolute nondetection of odontogenic pain should not be expected from an inquiry into a small number of symptoms. The 29 percent response rate among the participants with odontogenic pain could represent some mix of splinting pain,[2] perhaps as a consequence of odontogenic nociception, or it could represent a possible source of false-positive responses to the screening instrument. Nevertheless, the present result does not detract from the performance of this instrument in the detection of pain-related TMD. In addition, the large separation in threshold response rates, as based on symptom reporting by participants with TMD versus that by participants with odontogenic pain, should be regarded as encouraging.

The low sensitivity among participants who reported having temple headache but no pain-related TMD diagnosis would appear to represent an ideal outcome for use of this instrument in a setting in which the user wants to focus on detecting TMDs. One could expect much higher detection of headache related to a TMD, especially on the basis of the expected endorsement of the items pertaining to modifying factors—but, of course, this is an empirical question subject to confirmation.

The initial patient history questionnaire in the RDC/TMD[26] contains a single item for assessing the presence of regional pain during the preceding month. Through systematic examination of this instrument during the Validation Project,[10] we discovered that using a single item as a basis to assess pain, despite its being quite logical in structure, does not capture completely the nuances of pain in people whose pain experience is ambiguous—that is, sits on the borderline between clearly present and clearly absent. Although published data demonstrate that this single item can have excellent sensitivity and specificity (96 percent and 95 percent, respectively),[27] and although we were able to replicate such findings in our data set, we were concerned that those values of sensitivity and specificity in our data were biased by our recruitment method, in which we asked essentially the same question to determine inclusion for people with pain and exclusion for control participants. Furthermore, we noticed that the reference standard Reissmann and colleagues[27] used also ultimately was based on this same item for those who did or did not receive a diagnosis, leading to circularity in the findings. More generally, the findings of Reissmann and colleagues may

illustrate the known problems with reliability associated with the use of a single item for a complex construct. Consequently, this instrument does not use this single item as a filter question. Alternatively, the indicator variable model for measurement would appear to be a better approach for developing an instrument whose purpose is to identify the presence of pain, given that its recognition by the person often is based on access to semantic memory,[28,29] which is enhanced by using a variety of cues within the questionnaire structure. An additional advantage of this approach is that in many settings, the presence or absence of pain across the preceding 30 days is used frequently as either a critical inclusion criterion (for research) or as a critical criterion for a pain diagnosis.

Given that risk factors for TMD onset or a symptom profile of early-stage TMD have not been established yet, a screening instrument for TMD is, at this time, best constructed on the basis of symptoms. An alternative approach to questionnaire-based screening instruments has been the use of a screening examination, as proposed by Schiffman and colleagues,[30] although this is in the context of first determining who has symptoms on the basis of the presence of pain. The primary challenge, however, in using an examination as part of a screening process is that clinical parameters such as joint sounds, jaw opening pattern and pain on palpation may represent a variation of normality or simply may be poor descriptors of the phenomena under study. One additional challenge in the use of clinical examination procedures for screening purposes is that the examiner must be sufficiently reliable. In contrast, the virtue of a short self-administered screening instrument is that it can be more reliable and can be used by the patient before a clinic visit or in field studies in which examiners are not available.

If a screening instrument is used for early identification of a condition in a clinical setting, there must be a suitable consequence if the finding is positive. In the case of TMD, this means that a positive finding with the screener must be followed by a more comprehensive evaluation to establish a diagnosis. Diagnosis must be followed by appropriate intervention at the early stage of identification to minimize chronicity.[31]

Limitations

This study has several limitations. First, our study samples were based on convenience and are not intended necessarily to represent population-based demographics. However, the demographic distributions of our samples were similar to typical findings for corresponding clinical patient groups. A second limitation is that the final instrument item content is based on collapsing specific items. Although the data from the odontogenic pain group provide some support for how participants would respond to the items as currently worded, further research with the final instrument is needed to replicate the findings. A third limitation is that we did not assess the odontogenic pain group for TMD owing to logistic restrictions in data collection.

CONCLUSION

The validity of the TMD-pain screening instrument we developed and tested is demonstrated by the instrument's excellent levels of sensitivity and specificity. The intended uses of this instrument are to identify patients who require further clinical evaluation and to provide a

standardized screening instrument for research purposes. For clinical use, the responses from the screener then can be used as part of the diagnostic process for a pain-related TMD diagnosis. The brevity of this instrument allows for its routine use in clinical and research settings for better assessment of patients who may have pain-related TMD.

Acknowledgments

ABBREVIATION KEY

ICHD-II	International Classification of Headache Disorders, second edition
RDC-TMD	Research Diagnostic Criteria for Temporomandibular Disorders
TMD	Temporomandibular disorder
TMJ	Temporomandibular joint

References

1. McNeill C, Mohl ND, Rugh JD, Tanaka TT. Temporomandibular disorders: diagnosis, management, education, and research. JADA. 1990; 120(3):253, 255, 257. [PubMed: 2179355]

2. de Leeuw, R., editor. American Academy of Orofacial Pain. Orofacial Pain: Guidelines for Assessment, Diagnosis, and Management. 4. Hanover Park, Ill: Quintessence; 2008. p. 25-48.

3. Lundeen TF, Levitt SR, McKinney MW. Discriminative ability of the TMJ scale: age and gender differences. J Prosthet Dent. 1986; 56(1):84–92. [PubMed: 3459881]

4. Kleinknecht RA, Mahoney ER, Alexander LD, Dworkin SF. Correspondence between subjective report of temporomandibular disorder symptoms and clinical findings. JADA. 1986; 113(2):257–261. [PubMed: 3462230]

5. Locker D, Slade G. Association of symptoms and signs of TM disorders in an adult population. Community Dent Oral Epidemiol. 1989; 17(3):150–153. [PubMed: 2736896]

6. Nielsen L, Terp S. Screening for functional disorders of the masticatory system among teenagers. Community Dent Oral Epidemiol. 1990; 18(6):281–287. [PubMed: 2090378]

7. Gerstner GE, Clark GT, Goulet JP. Validity of a brief questionnaire in screening asymptomatic subjects from subjects with tension-type headaches or temporomandibular disorders. Community Dent Oral Epidemiol. 1994; 22(4):235–242. [PubMed: 7924238]

8. Nilsson IM, List T, Drangsholt M. The reliability and validity of self-reported temporomandibular disorder pain in adolescents. J Orofac Pain. 2006; 20(2):138–144. [PubMed: 16708831]

9. Bossuyt PM, Reitsma JB, Bruns DE, et al. Standards for Reporting of Diagnostic Accuracy. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. Clin Chem. 2003; 49(1):7–18. [PubMed: 12507954]

10. Schiffman EL, Truelove EL, Ohrbach R, et al. The Research Diagnostic Criteria for Temporomandibular Disorders, part I: overview and methodology for assessment of validity. J Orofac Pain. 2010; 24(1):7–24. [PubMed: 20213028]

11. Ohrbach R, Turner JA, Sherman JJ, et al. The Research Diagnostic Criteria for Temporomandibular Disorders, part IV: evaluation of psychometric properties of the Axis II measures. J Orofac Pain. 2010; 24(1):48–62. [PubMed: 20213031]

12. Headache Classification Subcommittee of the International Headache Society. The International Classification of Headache Disorders, 2nd edition. Cephalalgia. 2004; 24(supplement 1):37–43. [PubMed: 14687011]

13. Ramsay, JO. TestGraf: A Program for the Graphical Analysis of Multiple Choice Test and Questionnaire Data. Montreal: McGill University; 1995.

14. Lange R, Irwin HJ, Houran J. Top-down purification of Tobacyk's Revised Paranormal Belief Scale. Pers Individ Dif. 2000; 29(1):131–156.

15. Ramsay J. Kernel smoothing approaches to nonparametric item characteristic curve estimation. Psychometrika. 1991; 56(4):611–630.

16. Santor DA, Ramsay JO. Progress in the technology of measurement: applications of item response models. Psychol Assess. 1998; 10(4):345–359.

17. Hambleton RK. Emergence of item response modeling in instrument development and data analysis. Med Care. 2000; 38(9 suppl):II60–II65. [PubMed: 10982090]

18. Andrich, D. Rasch Models for Measurement. Newbury Park, Calif: Sage; 1988.

19. Tabachnick, BG.; Fidell, LS. Using Multivariate Statistics. 4. Boston: Allyn and Bacon; 2000. p. 582-652.

20. Linacre, JM. A User's Guide to Winsteps: Rasch-Model Computer Programs—Program Manual 3.68.0. Chicago: MESA Press; 2009.

21. Muthén, L.; Muthén, BO. Mplus User's Guide. 3. Los Angeles: Muthén & Muthén; 2004. p. 39-42.

22. Stata User's Guide, Release 11. College Station, Texas: Stata; 2009.

23. Wilcox RR. New designs in analysis of variance. Annu Rev Psychol. 1987; 38:29–60.

24. Wang D. Confidence intervals for the ratio of two binomial proportions by Koopman's method. Stata Tech Bull Reprints. 2001; 10:244–247.

25. Ohrbach R, List T, Goulet JP, Svensson P. Recommendations from the International Consensus Workshop: convergence on an orofacial pain taxonomy. J Oral Rehabil. 2010; 37(10):807–812. [PubMed: 20374436]

26. Dworkin SF, LeResche L. Research diagnostic criteria for temporomandibular disorders: review, criteria, examinations and specifications, critique. J Craniomandib Disord. 1992; 6(4):301–355. [PubMed: 1298767]

27. Reissmann DR, John MT, Schierz O, Hirsch C. An abbreviated version of RDC/TMD (in German). Schmerz. 2009; 23(6):618–627. [PubMed: 19838739]

28. Erskine A, Morley S, Pearce P. Memory for pain: a review. Pain. 1990; 41(3):255–265. [PubMed: 1697054]

29. Terry R, Brodie EE, Niven CA. Exploring the phenomenology of memory for pain: is previously experienced acute pain consciously remembered or simply known? J Pain. 2007; 8(6):467–475. [PubMed: 17306627]

30. Schiffman E, Haley D, Baker C, Lindgren B. Diagnostic criteria for screening headache patients for temporomandibular disorders. Headache. 1995; 35(3):121–124. [PubMed: 7721570]

31. Gatchel RJ, Polatin PB, Kinney RK. Predicting outcome of chronic back pain using clinical predictors of psychopathology: a prospective analysis. Health Psychol. 1995; 14(5):415–420. [PubMed: 7498112]
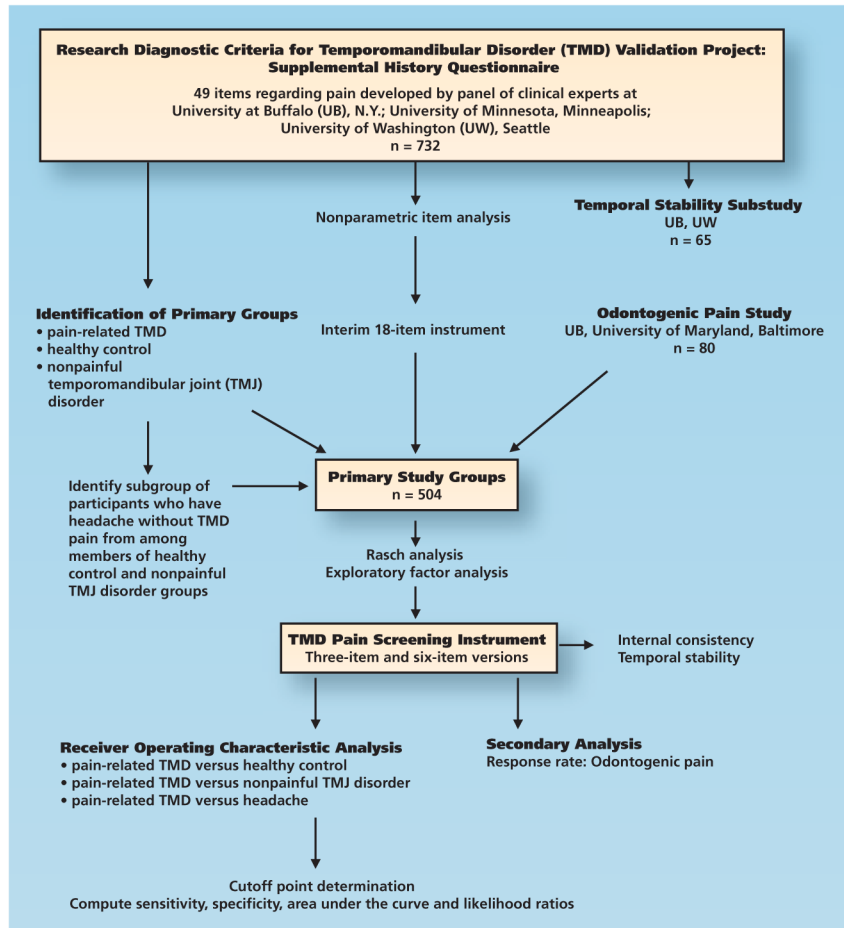
Figure 1.
Flowchart of study design showing participant sources, analyses and participant groups.

Temporomandibular pain disorder screening instrument.

1. In the last 30 days, on average, how long did any pain in your jaw or temple area on either side last?

    a. No pain

    b. From very brief to more than a week, but it does stop

    c. Continuous

2. In the last 30 days, have you had pain or stiffness in your jaw on awakening?

    a. No

    b. Yes

3. In the last 30 days, did the following activities change any pain (that is, make it better or make it worse) in your jaw or temple area on either side?

A. Chewing hard or tough food

    a. No

    b. Yes

B. Opening your mouth or moving your jaw forward or to the side

    a. No

    b. Yes

C. Jaw habits such as holding teeth together, clenching, grinding or chewing gum

    a. No

    b. Yes

D. Other jaw activities such as talking, kissing or yawning

    a. No

    b. Yes

Items 1 through 3A constitute the short version of the screening instrument, and items 1 through 3D constitute the long version. An a response receives 0 points, a b response 1 point and a c response 2 points.

Figure 2.
Final instrument and scoring rules.

TABLE 1

Screening instruments for temporomandibular disorders (TMDs).

SOURCE	SAMPLE	METHODS	INSTRUMENT	DIAGNOSTIC PERFORMANCE	LIMITATIONS
Lundeen and Colleagues,[3] **1986**	153 patients with symptomatic temporomandibular joints (TMJs); estimated 73 percent female; age groups < 30 to > 50 years	Participants' self-administration of 97-item inventory of symptoms and self-examination	97 items assessing palpation pain, pain report, perceived malocclusion, jaw limitation and joint dysfunction	t test scores presented for each subscale comparing symptomatic and symptom-free participants; only total score is reliable	Lack of reference standard diagnosis; instrument length too long for routine screener; includes unreliable self-examination; total score represents multidimensional constructs
Kleinknecht and Colleagues,[4] **1986**	65 nonpatients; 75 percent female; ages 22 to 67 years	Reference standard clinical classification provided by four dentists using a calibrated technique	14 dichotomous questions assessing TMD symptoms; no instrument provided	Interexaminer reliability (percentage agreement) 50 to 92 percent	Lack of reference standard diagnosis; no psychometric properties presented
Locker and Slade,[5] **1989**	148 participants; 61 percent female; ages 18 to 82 years	Telephone interview; standard protocol for diagnosis	Nine items assessing TMD symptoms including pain, function, perceived mechanical interference and occlusal changes	Sensitivity, 81 percent; specificity, 48 percent	Lack of reference standard diagnosis; Helkimo dysfunction index used as reference standard; total score represents multidimensional constructs; high false-positive rate
McNeill and Colleagues,[1] **1990**	None	Self-report, followed by conditional clinical examination	Nine items assessing pain, TMJ sounds, and jaw function; suggested cutoff to determine need for examination; items have face validity only	Not based on data; psychometric properties undetermined	No data available regarding diagnostic usefulness
Nielsen and Terp,[6] **1990**	706 participants; ages 14 to 16 years	Reference standard clinical classification provided by a single clinical examiner	26 items assessing dysfunction, head and face pain, habits, dental symptoms and general health	Reliability (π): 0.52 to 0.70; eight of 26 items selected for final screener; sensitivity, 9 to 65 percent; specificity, 60 to 99 percent	Total score represents multidimensional constructs; no cutoff provided; optimal balance in each of sensitivity and specificity
Gerstner and Colleagues,[7] **1994**	307 participants; ages 21 to 42 years; participants divided into headache, TMD and pain-free control groups	Self-report	13 ordinal severity items assessing jaw pain and jaw function; no scoring rules provided	Sensitivity 95 percent, specificity 100 percent in comparing participants with TMD versus control participants on basis of consideration of multiple cutoffs; performance with headache group not reported	Lack of reference standard diagnosis; total score represents multidimensional constructs
Nilsson and Colleagues,[8] **2006**	120 participants; ages 12 to 19 years; case and control participants matched for age and sex	Self-report provided a dichotomous classification; RDC-TMD* used as reference standard	Two questions with hierarchical scoring; administered by clinician after clinical examination; clinician provided anatomic instruction for pain localization	Reliability (κ) 0.83; sensitivity, 96 percent; specificity 83 percent	Self-report biased by clinician's guidance of pain localization; pain report confounded by clinical examination; sensitivity and specificity statistics biased by virtue of criterion measure having been obtained first

*RDC-TMD: Research Diagnostic Criteria for Temporomandibular Disorders (Dworkin and LeResche[26]).

TABLE 2

Demographic characteristics of the study population.

DIAGNOSTIC GROUP	SAMPLE SIZE	MEAN (SD) AGE, IN YEARS	FEMALE SEX (%)
Healthy Control	96	36.2 (13.2)	62.5
Nonpainful Temporomandibular Joint (TMJ) Disorder	116	39.9 (13.1)	76.7
Pain-Related Temporomandibular Disorder (TMD)	212	36.9 (12.6)	91.0
Headache Without TMD Pain[†]	45	37.7 (11.7)	80.0
Odontalgia	80	37.4 (18.3)	53.8

[*]Across the primary groups, age did not differ; significant Bartlett's test for equal variances ($\chi^2_3 = 18.6$, $P = .000$) was followed by a W test with robust errors ($W = 1.81$ [3, 213.9], $P = .15$). Female sex was predominant overall ($P = .000$) and was allocated unequally across groups ($\chi^2_3 = 58.2$, $P = .000$). Across the two subsets, age did not differ ($F = 0.08$ [1, 195], $P = .78$); female sex was predominant ($P = .000$) but the proportion was different across the two groups ($\chi^2_1 = 3.3$, $P = .07$).

[†]This group is a subset of the group with nonpainful TMJ disorder and the healthy control group.

TABLE 3

Interitem correlation matrix. *†

VARIABLE	Stiffness or Pain on Awakening	Pain Frequency	Chewing Affects Pain	Movement Affects Pain	Parafunction Affects Pain
Pain Frequency	0.72				
Chewing Affects Pain	0.89	0.78			
Movement Affects Pain	0.89	0.74	0.92		
Parafunction Affects Pain	0.90	0.78	0.96	0.91	
Other Activities Affect Pain	0.83	0.65	0.90	0.85	0.90

* Correlations with "pain frequency" are based on Kendall τ-b rank correlation; all other correlations are tetrachoric with use of a maximum likelihood estimator.

† Empty cells are redundant with other cells in the matrix that contain the value.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

TABLE 4

Validity of temporomandibular disorder (TMD) pain screener questionnaire according to diagnostic groups.[*]

DIAGNOSTIC GROUP[*]	CELL FREQUENCIES (N)				SENSITIVITY (%)[†]	SPECIFICITY (%)	AREA UNDER THE CURVE		LIKELIHOOD RATIO	
	True Positive	False Negative	False Positive	True Negative			Estimate	95% Confidence Interval	Positive	Negative
Short Screener (Threshold Value: 2)										
Pain-related TMD versus healthy control	210	2	3	93	99.1	96.9	98.6	96.8, 100	31.7	0.01
Pain-related TMD versus nonpainful temporomandibular joint (TMJ)	210	2	6	110	99.1	94.8	99.0	97.8, 100	19.2	0.01
Pain-related TMD versus headache	210	2	2	43	99.1	95.6	99.7	99.3, 100	22.3	0.01
Long Screener (Threshold Value: 3)										
Pain-related TMD versus healthy control	210	2	3	93	99.1	96.9	99.3	98.3, 100	31.7	0.01
Pain-related TMD versus nonpainful TMJ	210	2	3	113	99.1	97.4	99.1	97.6, 100	38.3	0.01
Pain-related TMD versus headache	210	2	1	44	99.1	97.8	99.9	99.8, 100	44.6	0.01

[*] Pain-related TMD: Participants with clinical diagnoses of myofascial pain, arthralgia or both. Healthy controls: Participants who did not meet criteria for either myofascial pain or arthralgia diagnoses. Non-painful TMJ: Participants with temporomandibular joint (TMJ) disk displacements, osteoarthrosis or both, not meeting criteria for a pain-related TMD diagnosis. Headache: Participants with tension-type headaches in the temple region (according to International Classification of Headache Disorders, second edition, criteria), formed as a subset of healthy control participants and participants in the nonpainful TMJ disorder group.

[†] Note that sensitivity remains the same when the same case group (pain-related TMD) is used in a pairing.