Validity of a brief questionnaire in screening asymptomatic subjects from subjects with tension-type headaches or temporomandibular disorders

Geoffrey E. Gerstner[1], Glenn T. Clark[2], and Jean-Paul Goulet[3]

[1]University of Michigan School of Dentistry, Ann Arbor, MI, [2]Dental Research Institute, UCLA School of Dentistry, Los Angeles, CA, U.S.A. and [3]Ecole de Medecine Dentaire, Université Laval, Quebec, Canada

Gerstner GE, Clark GT, Goulet J-P. Validity of a brief questionnaire in screening asymptomatic subjects from subjects with tension-type headaches or temporomandibular disorders. Community Dent Oral Epidemiol 1994; 22: 235–42. © Munksgaard, 1994

Abstract – Clinical investigations of temporomandibular disorders require objective, repeatable methods for screening diseased subjects from non-diseased control subjects. This study evaluated whether information gathered from a short, public domain questionnaire was useful in distinguishing temporomandibular disorder subjects ($n=216$) from non-temporomandibular disorder controls ($n=69$) and tension-type headache subjects ($n=22$). The questionnaire consisted of eight questions relating to jaw pain (i.e., location of pain, precipitating factors, and temporal pattern of pain) and five questions relating to jaw function (i.e., joint noises, locking, and difficulty in opening). There were five possible answers to each question which ranged from 0 (no symptoms) to 4 (unbearable or constant symptoms). The total scores for the eight pain questions and the five jaw function questions were used to determine the questionnaire's sensitivity and specificity in each group, and ROC curves were plotted to identify the best cutoff point for disease presence or absence. Results showed that the questionnaire reliably distinguished between the control group and temporomandibular disorder group with 90.3%–97.7% sensitivity and 95.7%–100% specificity at cutoff values between 5 and 9. These results support the use of the questionnaire as a primary screening tool for general practice and as a supplementary screening tool for clinical temporomandibular disorder studies. However, results also showed that the questionnaire was unable to distinguish easily between TMD subjects and temporalis region tension-type headache subjects.

Subject selection and categorization is a difficult task when attempting clinical or epidemiological research on temporomandibular disorders (TMD). These difficulties exist because there are no universally acceptable, validated diagnostic criteria for TMD. By the same token, there is no agreement on what should be considered evidence of a normal or non-diseased status. Researchers dealing with

TMD have used several diagnostic tools in subject selection procedures. Tools that have been used, often with equivocal success, have been subject self-report, clinical findings uncovered during a standardized screening examination, electronic diagnostic tests, or a combination of these procedures. DWORKIN & LERESCHE (1) have recently provided an excellent review of such TMD diagnostic

criteria. Each of these tools has advantages and disadvantages. A major disadvantage of most is that they have not been tested for validity and reliability.

HELKIMO (2–4) proposed an index for TMD and performed an epidemiological survey examination for TMD on Scandinavian subjects. He found that only 12% were free of all signs and symptoms, and that as many as 22% had at least two

severe TMD symptoms. At that time, he stated that the clinical dysfunction index developed to classify subjects in his study had not been tested for validity, and that it suffered from both a lack of standardization and from "subjective effects" (3).

Other indexes have since been proposed, but to date none has ever been evaluated for diagnostic validity using a blind-to-subject status research method (5, 6). In contrast to these clinical examination based methods, COOPER & RABUZZI (7) advocated the use of electronically-defined signs derived by measuring jaw positions as a means of diagnosing TMD. They concluded that TMD prevalence in an asymptomatic and otherwise healthy subject population was as high as 81%. Both HELKIMO's (3) and COOPER & RABUZZI's (7) criteria severely overestimate the disease prevalence. We argue that the physical signs they evaluated do not alone define TMD.

Several attempts have been made to assess the reliability of proprietary diagnostic questionnaires (8, 9), clinical examinations, patient self-report methods (10), and various electronic diagnostic instruments (11, 12) for clinical research purposes. The diagnostic validity of imaging techniques has also been evaluated (13, 14). The recent MRI data regarding TM joint disk position variability in asymptomatic subjects (15) and previous cadaver-based histological research (16, 17) suggest that substantial TM joint hard tissue (i.e., "deviation in form") and soft tissue (i.e., disc displacement) changes occur in a large percentage of the population by the third to fourth decade of life. These data suggest that in some situations such anatomic changes are not specifically associated with TMD symptoms and are probably within the range of normal adaptive or aging processes. Thus, it is clear that TM joint imaging techniques are better at documenting joint tissue form than at diagnosing the presence of current disease.

The practicing clinician's "gold standard" for TMD is patient self-report in combination with a confirmatory clinical examination. This combination is necessary, because TMD have no reliable histopathologic markers, more than half of the population has at least one of the characterizing signs or minor symptoms, and the condition is usually diagnosed

when it has a substantial impact on the patient's life.

A self-report was the only method attempted by DWORKIN et al. (18) when they used the single criterion of "'facial ache or pain in the jaw muscles, the joint in front of the ear, or inside the ear (other than infection)' during the previous 6 months" to estimate the prevalence of TMD-related pain. They found that 12.1% of patients in a Seattle-based HMO responded positively to the question. As a sole criterion for the presence of a TMD, this single question is likely to identify most TMD patients with pain. However, it would exclude a small group of patients with non-painful but clearly abnormal TM function (e.g., jaw opening restrictions, joint noises, open locking), and it would not necessarily differentiate diseases causing sinus pain, eye pain, salivary gland pain, non-infectious ear pain, and headache and facial pain disorders that do not involve the TM apparatus (e.g., muscles and joints).

Considering the above limitations of a single question self-report method, it would seem logical to explore a multiple item questionnaire as a means of identifying or ruling out TMD. LEVITT et al. (9) developed such a questionnaire. Unfortunately, their questionnaire is only available commercially, which involves added costs to clinicians. Furthermore, their questionnaire contains multiple psychological profile questions, which must be scored by computer. Hence, results cannot be interpreted immediately after a subject has finished the questionnaire, and thus unavoidable delays occur in evaluating a potential subject's disease status. The current study was, therefore, undertaken to evaluate the validity of a public domain, 13-item screening questionnaire for TMD, which can be scored by the clinician on-site.

Materials and methods

The questionnaire – It was decided to use an existing brief questionnaire developed and refined for the specific purpose of TMD. The development of the questionnaire occurred over a 15-yr period and represents a distillation of items from several questionnaires that have been in use in TMJ clinics around the country.

The questionnaire development process involved careful screening and anecdotal evaluation of each item by clini-

cians practicing in the private care facility of the UCLA TMJ and Facial Pain Clinic. The questionnaire consisted of 13 questions, eight concerning jaw pain and five concerning jaw function (see sample questionnaire, Fig. 1). There were five possible answers to each question, ranging in a graded order from no signs or symptoms to extreme signs or symptoms. Subjects were instructed to select the most appropriate response to each question.

Subject groups – The three test groups of subjects were the following (Table 1): 1) non-TMD, non-headache subjects (Controls); 2) subjects with tension-type headache symptoms (HA); and 3) subjects who were either attending a temporomandibular disorder clinic for treatment or were recruited as TMD research subjects by advertisement (TMD). General statistical information on subjects within each test group appear in Table 2.

Subjects were recruited from one of three sources (Table 1). One source of subjects ("Clinic", Table 1) was the UCLA TMJ and Orofacial Pain Clinic. This included 180 new patients who were attending for private treatment of their temporomandibular disorder and consented to participate in a research study that involved gathering pre-treatment psychosocial data. These 180 patients were sequentially enrolled in the study as they presented to the clinic for care. The only exception to sequential enrollment in the study occurred if patients presented to the clinic without a TMD or if they were only seeking a second opinion from the clinic.

A second source of subjects ("Tension study", Table 1) was the UCLA Clinical Research Center. These included 55 subjects, 22 with headache symptoms (as described below) and 33 without symptoms, who were part of a study on tension-type headaches. These subjects replied to newspaper advertisements soliciting both tension headache sufferers and control subjects. Respondents were sent a set of screening forms which asked them to describe the intensity, location, and duration of signs and symptoms of headaches. These forms did not use questions that were a part of the questionnaire being investigated in the current study.

The third source of subjects ("TMD Study", Table 1) came from respondents to newspaper advertisements soliciting

JAW SYMPTOM & ORAL HABIT QUESTIONNAIRE - Form #4

Name: _____   Date: _____

INSTRUCTIONS: Please check the appropriate answer to the following questions.

A. Jaw Pain Questions

	Doesn't Hurt At All	Hurts A Little	Hurts A Lot	Almost Unbearable	Unbearable Pain Without Relief
1. Does it hurt when you open wide or yawn?					
2. Does it hurt when you chew, or use the jaws?					
3. Does it hurt when you are not chewing or using the jaws?					
4. Is your pain worse on waking?					
5. Do you have pain in front of the ears or earaches?					
6. Do you have jaw muscle (cheek) pain?					
7. Do you have pain in the temples?					
8. Do you have pain or soreness in the teeth?					

B. Jaw Function Questions

	No	Maybe A Little	Quite A Lot	Almost All The Time	All The Time Without Stopping
9. Do your jaw joints make noise so that it bothers you or others?					
10. Do you find it difficult to open your mouth wide?					
11. Does your jaw ever lock closed so you cannot open it?					
12. Does your jaw ever lock open so you cannot close it?					
13. Do you have a problem with your bite being uncomfortable?					

Fig. 1. The questionnaire used in the study.

both control and TMD-symptomatic subjects. Final subject selection from the 3 sources described was based on specific exclusion and inclusion criteria discussed in the next section.

Subject inclusion and exclusion criteria – Data used to establish inclusion and exclusion criteria for subjects were gathered during a preliminary examination at which time the subjects were required to fill out forms as well as participate in a clinical examination performed by calibrated examiners. The clinical examination involved acquiring a thorough subject history as well as performing an intra-oral and extra-oral evaluation by the calibrated examiners. The examiners performed muscle and joint palpation for tenderness and pain, mandibular movement, and joint sound assessments. Examiner calibration techniques have been described elsewhere (19). It is important to emphasize that these methods were performed in a rigorously standardized fashion. It is also important to note that the questionnaire being investigated in the current study was not used during this preliminary examination.

Inclusion criteria for the TMD group ($n = 216$ subjects, Table 1) included subject awareness of continuous or intermittent jaw region pain or dysfunction of ≥ 4 months duration, occurring $\geq$ three times a week, with a usual pain level ≥ 10 mm on a 100 mm visual analog scale. Also, one or more of the following signs had to be discovered during the preliminary examination by the calibrated examiners: painful TMJ noises, limited jaw opening of recent onset (< 38 mm including overbite), or pain in the masticatory system replicated by palpation of the TM joint, masseter or temporalis muscles. Methods used by the examiners to gather these data have been reported elsewhere (19, 20).

Inclusion criteria for the headache subjects ($n = 22$, see "HA" column, Table 1) were a positive complaint of a frequent, dull aching temporal region pain of a protracted nature. The frequency of pain had to be $\geq$ three times per week, lasting ≥ 1 hour in duration, with a usual intensity level ≥ 20 mm on a 100 mm visual analog scale. Calibrated examiners verified that each subject's headache disorder was primarily a temporalis region headache, and subjects were excluded from the study if the clinical examination revealed that this was not the case. Furthermore, headache subjects included in the study had to be free of moderate to severe craniocervical pain or dysfunction, painful TM joint clicking, crepitus, or jaw locking problems as determined from subject self-report and by methods used by the calibrated examiners (19, 20).

Control group subjects ($n = 69$, Table 1) were selected that matched the ages, genders, and races of the headache subjects and the TMD subjects. Inclusion criteria for control subjects were the absence of a TMD or headache, and this was determined to be the case by subject self-report and by muscle and joint palpation techniques used by the examiners (19, 20). Potential control group subjects

Table 1. Test group sample sizes by source of patient

Source	Controls	HA	TMD
Clinic			180
Tension study	33	22	
TMD study	36		36
Total	69	22	216

Table 2. Distributions by gender and age of subjects in each test group

Test group*	Gender (M:F)	Age ± 1 SD
Controls	1:5.90	27.6 ± 7.7
HA	1:6.33	25.0 ± 7.1
TMD	1:5.35	31.6 ± 10.6

* Test groups are: asymptomatic subjects ("Controls", $n = 69$), subjects with headache symptoms only ("HA", $n = 22$), and subjects with temporomandibular symptoms only ("TMD", $n = 216$).

who reported no TMD, but who on clinical examination reported occasional joint sounds, were included in the study if they never had any associated pain or jaw movement limitation, and did not think they had a TMD problem for which they would seek treatment.

No control, headache or TMD subjects were included in the study if they had signs or symptoms of sinus problems or disease, a dental infection, a recurrent ear disorder or earaches, systemic inflammatory polyarthritis, moderate to severe cervical pain or dysfunction. Also subjects with any ongoing dental restorative or orthodontic treatment were not included.

Data analysis – As previously stated, the questionnaire (Fig. 1) consisted of 13 questions with five possible answers to each question. For analysis purposes, answers were converted to ordinal rankings from 0 (i.e., no disease signs or symptoms) to 4 (i.e., the most intense signs or symptoms). Occasionally, subjects selected more than one answer to a specific question. In such instances, the average of all marked responses, rounded to the next whole number was used as the specific question's score. For instance, if a subject marked the first and second category for a specific question (i.e., ordinal rankings of 0 and 1, respectively), the score was considered to be 1; if a subject marked the fourth and fifth category (i.e., ranked scores of 3 and 4, respectively), the question's score was rounded to 4.

The purpose of the current study was to test the questionnaire's ability to discriminate TMD subjects from non-TMD pain-free subjects, and to determine how well the questionnaire separated subjects who were claiming the presence of a temporal region tension-type headache problem from the other two groups. This separation was assessed as follows. First, the total scores from all 13 questions, the eight jaw pain questions only, and the five jaw function questions only, were used to determine sensitivity and specificity for comparisons among the three test groups. Sensitivity and specificity were determined at various cutoff points from 0 up to the maximum reported questionnaire score. Receiver operator characteristic curves were utilized to determine the best combination of sensitivity and specificity values for each group. Secondly, these sensitivity and specificity measures

were used to determine the predictive value of a positive and negative test result relative to the presence and absence of TMD for different prevalence rates (21).

Results

The distribution of questionnaire total scores for each of the test groups is shown in Fig. 2. The X axis is divided into bins. Numbers below each bin define the total scores included in the bin. The Y axis depicts the number of questionnaire total scores, expressed as a percentage of the total group size, found within each bin. This figure indicates the degree of overlap in total questionnaire scores between the three test groups. A Kruskal-Wallis one-way analysis of variance showed that the questionnaire score results shown in Fig. 2 differed significantly among the three groups (H = 164.68, 2 df, P < 0.001).

Receiver operator characteristic (ROC) curves of three pairwise comparisons among the three test groups can be seen in Fig. 3. Total scores of all 13 questions were used in the construction of Fig. 3A. Total scores from the five jaw function questions only were used in Fig. 3B. Total scores from the eight jaw pain questions only were used in Fig. 3C.

Fig. 3A compares the Control group to the TMD group. The ROC curve indicates that the best cutoff point for distinguishing Control from TMD subjects occurred between 5 and 7. With a cutoff point of 6, sensitivity was 95.8% and specificity was 100%. Sensitivity was ≥70% and specificity was ≥95% for cutoff points between 5 and 12.

Figure 3B compares the HA group with the TMD group. For sensitivity and specificity calculations, the HA group served as the "non-diseased" group. Since both of these groups consisted of subjects complaining of head region pain, it was believed that the five jaw function questions would be best at distinguishing between the two groups. Therefore, this ROC curve (Fig. 3B) shows results from the five jaw function questions only. This ROC curve shows a change in slope at the cutoff point of 5. At this point, sensitivity was 68.5% and specificity was 86.4%. When total scores for all 13 questions were used in the comparison between the HA group and the TMD group, sensitivity and specificity were not improved.

The graph in Fig. 3C compares the Control group with the HA group. Since both of these groups were free of TMJ functional problems, only answers to the eight jaw pain questions were used in this comparison. The best sensitivity and specificity occurred at either a cutoff point of 3, where sensitivity was 100% and specificity was 97.1%, or a cutoff of 4, where sensitivity was 90.9% and
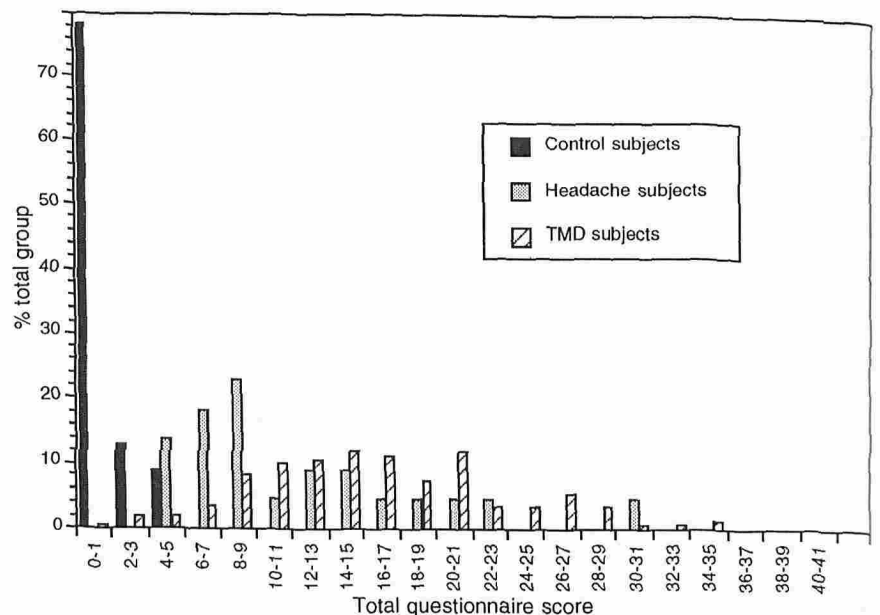

Fig. 2. Total questionnaire score distributions by test group expressed as a percentage of the total, individual group. Headache subjects = tension-type headache sufferer group; TMD subjects = group with only TMD signs and symptoms.
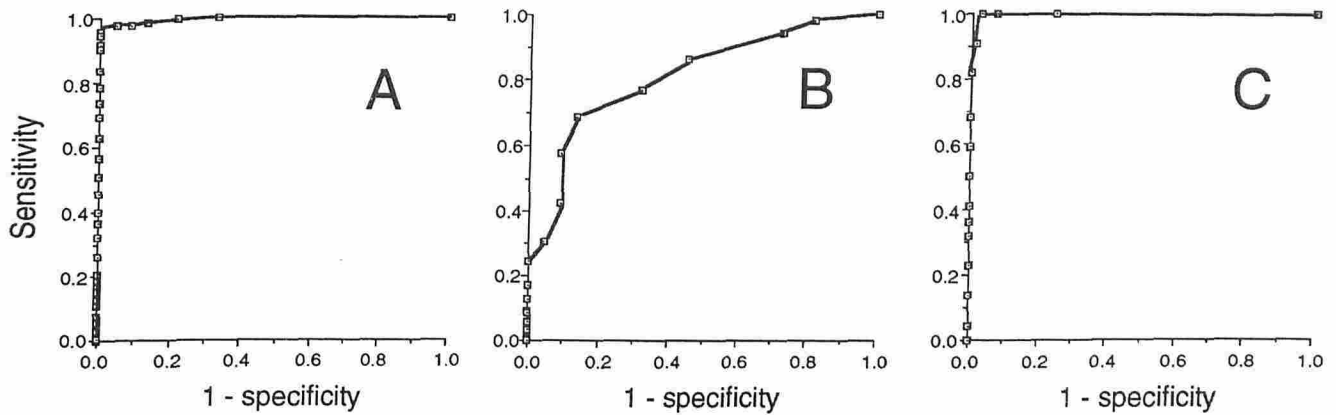
Fig. 3. Receiver operator characteristic (ROC) curves for pairwise comparisons between the three test groups. Various cutoff values can be determined by counting sequentially along the plotted curve, starting from 0 which is the upper righthand-most point in each graph. A: The Control group is compared with the TMD group; total scores from all 13 questions were used. B: The headache group is compared with the TMD group; total scores from the five jaw function questions only were used. C: The Control group is compared with the headache group; total scores from the eight jaw pain questions only were used.

specificity was 98.6%. When total scores for all 13 questions were used in the comparison between the Control group and the HA group, the best cutoff point occurred either at 4 where sensitivity was 100% and specificity was 91.3%, or at 5 where sensitivity was 95.5% and specificity was 95.7%.

Table 3 shows positive and negative predictive values and incremental gains expected from using total questionnaire scores to confirm or exclude TMD presence based on different disease prevalence rates. For general information, the positive predictive value estimates the probability that the disease is present in a given subject or patient whose test result is positive and the negative predictive value estimates the probability that the disease is not present in a given subject or patient whose test result is negative. Predictive values are good indicators of the overall clinical usefulness of a test, like the questionnaire, in different subject or patient populations.

Table 3 was constructed using the best sensitivity and specificity scores for the case where the questionnaire was used to discriminate between Control and TMD subjects (see above and footnote under table). In this case the best sensitivity and specificity occurred using a total questionnaire score of 6 as the cutoff point. Thus, in Table 3 total questionnaire scores ≥ 6 are considered positive tests (i.e., such subjects are assumed to have TMD), and total questionnaire scores < 6 are considered negative tests (i.e., such subjects are assumed not to have TMD).

In Table 3, column 1 shows selected prior estimates of disease prevalence. For instance, row 1 of column 1 indicates the case where the estimated disease prevalence rate within a hypothetical population = 90%. Column 2 shows the positive predictive value of total questionnaire scores that are ≥ 6. The equation used to calculate the positive predictive value (PPV) is shown under the table. Note that questionnaire scores ≥ 6 will always result in the positive predictive value = 100% (column 2). This is because specificity = 100% for such questionnaire scores. Thus, the incremental ruling-in gain (column 3) is greatly improved for cases where estimated disease prevalence is low. For instance, suppose there exists a disease prevalence rate = 5% (second to last row, column 1). At this prevalence rate, for a questionnaire score ≥ 6, the likelihood that the subject would have the disease = 100% (column 2). Thus, the examiner's ability to rule-in the presence of a disease in this subject would have increased by 95% (column 3), i.e., 100% (column 2) − 5% (column 1) = 95%.

The columns 4 and 5 in Table 3 show

Table 3. Predictive values and incremental gains expected from the screening questionnaire in confirming or excluding TMD presence based on various prevalence rates and reported sensitivity and specificity scores*

Prior estimate of disease prevalence (PEDP) (%)	Positive test (score ≥ 6)		Negative test (score < 6)		Prior estimated prevalence of no disease (100 − PEDP) (%)
	Positive predictive value (PPV)	Incremental gain	Negative predictive value (NPV)	Incremental gain	
90	100	10	73	63	10
80	100	20	86	66	20
70	100	30	91	61	30
60	100	40	94	54	40
50	100	50	96	46	50
40	100	60	97	37	60
30	100	70	98	28	70
20	100	80	99	19	80
10	100	90	100	10	90
5	100	95	100	5	95
1	100	99	100	1	99

* Sensitivity = 95.8%; Specificity = 100% (see text)

$$PPV = \frac{Sensitivity \times PEDP}{(Sensitivity \times PEDP) + [(100 - specificity) \times (100 - PEDP)]} \times 100$$

$$NPV = \frac{Specificity \times (100 - PEDP)}{[(Specificity \times (100 - PEDP)] + [(100 - sensitivity) \times PEDP]} \times 100$$

Incremental gain for positive test = PPV − PEDP; incremental gain for negative test = NPV − (100 − PEDP)

the negative predictive value and incremental gain respectively for total questionnaire scores that are considered negative test results, i.e., scores < 6. The equation used to calculate the negative predictive value (NPV) is shown under the table. Column 6 shows the prior estimated prevalence of no disease. For a hypothetical population where the estimated prevalence of no disease is 95% (second to last row, column 6), the negative predictive value of a questionnaire score that is < 6 is 100% (column 4). In other words, in a population where the estimated prevalence of no disease = 95%, if a subject is sampled from this population and his/her questionnaire score is < 6, the likelihood that s/he does not have the given disease = 100%. The incremental gain from having used the questionnaire score to rule out the presence of disease in the subject = 5% (column 5), i.e., 100% (column 4) − 95% (column 6) = 5%. As will be discussed in the next section, Table 3 indicates that the screening questionnaire provides important information, which can augment the clinician's or investigator's ability to confirm the presence or absence of disease in a given subject.

Discussion

Sensitivity estimates the probability that a test will be positive in a population of subjects known to have the disease, whereas specificity estimates the probability that a test will be negative among a non-diseased population. In other words, sensitivity estimates a diagnostic tool's ability to detect diseased subjects correctly; specificity estimates a diagnostic tool's ability to detect non-diseased subjects correctly. Receiver operator characteristic (ROC) curves are standard graphs designed to show the relationship between sensitivity and specificity scores at various cutoff values (22). An ideal diagnostic tool (i.e., one with 100% sensitivity and 100% specificity) will have an ROC curve that "reaches the upper left corner of the graph ... [whereas] ... a worthless test follows the diagonal from the lower left to the upper right corner ..." (22). Furthermore, BROWNER et al. (22) define the "best" cutoff as being where the curve "turns the corner" (i.e., the point at which the largest change in slope occurs). This definition of "best" was used throughout the paper.

With this in mind, the sensitivity–specificity values reported for the brief questionnaire show that, for the tested samples, a total questionnaire score of 6 was the "best" cutoff for distinguishing between subjects in the Control group and subjects with TMD symptoms (Fig. 3A). The questionnaire was also successful at distinguishing between the Control group and headache group (Fig. 3C) with a total score of 3 for the eight jaw pain questions being the "best" cutoff point. However, the questionnaire was less successful at differentiating between HA group subjects and TMD group subjects (Fig. 3B). We did not test the questionnaire for its ability to distinguish between subgroups of TMD disorders (e.g., TMJ locking, myalgia, or TMJ clicking), as the utility of such a distinction is not obvious.

Weighting the relative importance and implications of false positive and false negative test results, DWORKIN & LERESCHE (1, p. 319) suggested a sensitivity of 70% and a specificity of 95% as the minimum acceptable levels for clinically useful TMD diagnostic tests. The test characteristics of the questionnaire under investigation in the current study were beyond these standards and therefore acceptably distinguished between Control and TMD subjects for cutoff points between 5 and 12 (Fig. 3A). It also acceptably distinguished between Control subjects and HA subjects for cutoff points between 3 and 5 for the 8 jaw pain question answers (Fig. 3C) and for cutoff points between 5 and 7 for the 13 questions combined (not shown in figure format). On the other hand, the questionnaire would not be acceptable for distinguishing between the HA group and the TMD group (Fig. 3B).

Figs. 3A and 3C showed that the questionnaire was acceptable at discriminating Control subjects from either HA or TMD subjects; however, the reported sensitivity and specificity values are probably somewhat inflated, because the "best" cutoff points were chosen post hoc; therefore, the questionnaire needs to be tested on other subject populations to determine its robustness. However, the fact that sensitivity and specificity were extremely high over a wide range of cutoff values (see Results) suggests that the questionnaire will be useful as a screening tool regardless of the sampled test group.

There are two possible reasons why

the questionnaire did not adequately differentiate between the HA group and the TMD group (Fig. 3B). First, the HA group was constructed from subjects who complained of moderate to severe chronic daily tension-type headache symptoms which were located in the temporalis region. Such subjects are usually classified as tension-type headache sufferers, since they often complain of cervical region pain problems as well. Such a group of tension-type headache subjects is undoubtedly more difficult to distinguish from TMD subjects than subjects with other types of headaches (e.g., migraines, cluster headaches, etc.).

Second, it is likely that both HA and TMD subjects not only had similar pain complaints, as we had anticipated, but also had similar function complaints with respect to the questionnaire's jaw function questions. For instance, we expected that many headache sufferers in the current study would report pain in the temples and therefore would respond positively to question #7 (Fig. 1). In fact, 82% (18/22) of the HA subjects responded positively to question #7. Furthermore, we believed that HA subjects might answer question #3 positively if they assumed that the question did not relate specifically to jaw pain only. In fact, 45% (10/22) of the HA subjects responded positively to question #3. Hence, a positive response to question #3 did not necessarily indicate the presence of TMD-related jaw pain.

In anticipation of these problems, we hoped that the five jaw function questions would provide the necessary weighting to differentiate the HA group from the TMD groups. However, as Fig. 3B demonstrates, this was not the case. This suggests that jaw function problems, as addressed by the five jaw function questions (Fig. 1), were sufficiently prevalent in the headache subjects to bring the false positive rate to an unacceptable level. Indeed, of the 18 HA subjects responding positively to question #7, 16 had positive responses to one or more of the five jaw function questions.

The conclusion from these results is that the questionnaire may benefit from careful rewording or even deletion of existing questions, and the addition of a few items. Thereafter, the questionnaire may be better equipped to distinguish true TMD subjects from tension-type headache subjects. The important point

is that the five jaw function questions, as currently written, were not useful in separating HA subjects from TMD subjects.

These shortcomings do not invalidate use of the questionnaire for several important reasons. First, a more objective and critical appreciation of the questionnaire's utility is gained by carefully examining Table 3. This table shows the degree to which the questionnaire can correctly confirm the presence or absence of TMD in an individual subject or patient, given specific prevalence rates of the disease (Table 3). Subtracting the prior estimate of disease prevalence from the positive predictive value gives the incremental gain expected from using the screening questionnaire to confirm TMD presence or absence (Table 3). For instance, suppose there exists a TMD prevalence rate of 5% in the general population, as reported by Rugh & Solberg (23). At this prevalence rate, when a given questionnaire score is positive for TMD (i.e., total score ≥6), then the likelihood that the subject has TMD has increased from 5% to 100% (Table 3). In other words, the use of the questionnaire on the general population where TMD prevalence is known to be low results in a large incremental ruling-in gain, in the current example an incremental gain of 95% (positive predictive value minus prevalence rate of TMD = 100% − 5%, see Table 3). By the same token, the probability of no disease present in a subject whose total questionnaire score < 6 would be 100%, an incremental ruling-out gain of 5% (negative predictive value minus prevalence rate of no TMD = 100% − 95%, see Table 3). Thus, for a general practitioner or clinical scientist, the questionnaire can serve as an excellent screening test to rule out TMD in patient populations or subject samples, because it is generally believed that the TMD prevalence rate is low in such groups (23). However, a positive test should arouse suspicions to the point that a more thorough clinical examination should be considered in order to verify the presence or absence of TMD.

The questionnaire could also be useful when recruiting subjects for TMD studies from referral sources or from those who arrive for an initial visit in a TMD clinic. The prevalence estimate of TMD is often much higher in such cases, because many of these subjects or patients have already been pre-screened by themselves or health professionals. Therefore, a higher proportion of subjects will actually have TMD, and on the basis of the positive predictive values, the use of the screening questionnaire substantially increases the probability of TMD and reduces the chance of a false result. For instance, at a prevalence rate of 80%, the probability that a TMD is present in a subject whose total questionnaire score is ≥6 increases to 100% (Table 3). Thus, positive scores in such subject populations indicate with a high probability that TMD is likely to be present.

In conjunction with the above, we believe that no single diagnostic test, especially a patient report questionnaire, can or should be considered a gold standard or be capable of providing more information than is possible from a thorough clinical examination and a comprehensive medical history evaluation (20). With this in mind, we believe that this short questionnaire can be useful for two purposes. First, it may be useful as a screening tool for defining research subject groups in TMD studies. Confirmation of the subject groupings can be done with subsequent, thorough clinical examination and comprehensive medical history. The second purpose is to use this questionnaire, which has the advantage of immediate scoring, in the clinical setting as part of an initial diagnostic test. It is recommended that diagnostic interpretation should be based on both the questionnaire scores and findings from a careful and thorough clinical examination. Further research is required to create a unified test battery for use in TMD diagnosis. We feel that, ultimately, a questionnaire similar to the one presented in this study will play a significant and permanent role in screening for TM disorders.

References

1. Dworkin SF, LeResche L. Research diagnostic criteria for temporomandibular disorders: review, criteria, examinations and specifications, critique. *J Craniomand Dis* 1992; *6:* 301–55.

2. Helkimo M. Studies on function and dysfunction of the masticatory system. I. An epidemiological investigation of symptoms of dysfunction in Lapps in the north of Finland. *Proc Finn Dent Soc* 1974; *70:* 37–40.

3. Helkimo M. Studies on function and dysfunction of the masticatory system. II. Index for anamnestic and clinical dysfunction and occlusal state. *Swed Dent J* 1974; *67:* 101–21.

4. Helkimo M. Studies on function and dysfunction of the masticatory system. III. Analysis of anamnestic and clinical recordings of dysfunction with the aid of indices. *Swed Dent J* 1974; *67:* 165–82.

5. Fricton J, Schiffman EL. Reliability of a craniomandibular index. *J Dent Res* 1986; *65:* 1359–64.

6. Fricton J, Schiffman E. The craniomandibular index: validity. *J Prosthet Dent* 1987; *58:* 221–8.

7. Cooper BC, Rabuzzi DD. Myofacial [sic] pain dysfunction syndrome: a clinical study of asymptomatic subjects. *Laryngoscope* 1984; *94:* 68–75.

8. Lundeen TF, Levitt SR, McKinney MW. Discriminative ability of the TMJ scale: age and gender differences. *J Prosthet Dent* 1986; *56:* 84–92.

9. Levitt SR, McKinney MW, Lundeen TF. The TMJ scale: cross-validation and reliability studies. *J Craniomand Disorders* 1988; *6:* 17–25.

10. Deardorff WW, Butterworth JC. Patient self-report vs. examiner assessment of TMJ symptoms: stability of correspondence rates during treatment. *J Craniomand Pract* 1988; *6:* 316–9.

11. Feine FS, Hutchins MO, Lund JP. An evaluation of the criteria used to diagnose mandibular dysfunction with the mandibular kinesiograph. *J Prosthet Dent* 1988; *60:* 374–80.

12. Widmer CG, Lund JP, Feine JS. Evaluation of diagnostic tests for TMD. *J Calif Dent Assoc* 1990; *18:* 53–60.

13. Bibb CA, Pullinger AG, Baldioceda F, Murakami KI, Ross JB. Temporomandibular joint comparative imaging: diagnostic efficacy of arthroscopy compared to tomography and arthrography. *Oral Surg Oral Med Oral Pathol* 1989; *68:* 352–9.

14. Fava C, Gatti G, Cardesi E, Parchetti R, Rocca G, Preti G. Possibilities and limits in identifying the TMJ articular meniscus with the CT scanner: a comparative anatomoradiological study. *J Craniomandib Disord Facial Oral Pain* 1988; *2:* 141–7.

15. Hatala MP, Westesson PL, Tallents RH, Katzberg RW. TMJ disc displacement in asymptomatic volunteers detected by MRI imaging. *J Dent Res* 1991; *70* (Spec Iss): 278.

16. Oeberg T, Carlsson GE, Fajers CM. The temporomandibular joint: a morphologic study on human autopsy material. *Acta Odontol Scand* 1971; *29:* 349–84.

17. Solberg WK, Hansson TL, Nordstroem B. The temporomandibular joint: in young adults at autopsy: a morphologic classification and evaluation. *J Oral Rehabil* 1985; *12:* 303–21.

18. DWORKIN SF, HUGGINS KH, LERESCHE L, et al. Epidemiology of signs and symptoms in temporomandibular disorders: clinical signs in cases and controls. *J Am Dent Assoc* 1990; *120:* 273–81.

19. GOULET JP, CLARK GT, FLACK VF. Reproducibility of examiner performance for muscle and joint palpation in the temporomandibular system following training and calibration. *Community Dent Oral Epidemiol* 1993;*21:* 72–7.

20. GOULET JP, CLARK GT. Clinical TMJ examination methods. *J Calif Dent Assoc* 1990; *18:* 25–33.

21. GRINER PF, MAYEWSKY RJ, MUSHLIN AI, GRENNLAND P. Selection and interpretation of diagnostic tests and procedures. *Ann Int Med* 1981; *94:* 553–600.

22. BROWNER WS, NEWMAN TB, CUMMINGS SR. Designing a new study: III. Diagnostic tests. In: HULLEY SG, CUMMINGS SR, eds. *Designing clinical research.* Baltimore: Williams and Wilkins, 1988; 87–97.

23. RUGH JD, SOLBERG WK. Oral health status in the United States: temporomandibular disorders. *J Dent Educ* 1985; *49:* 398–405.