Comparison of Two Scales in the Assessment of Muscle and Joint Palpation Tenderness in Chronic Temporomandibular Disorders

Michael J. Atwood, DMD
Resident
General Dentistry

D. Carl Dixon, DDS, MS
Chief
TMD Clinic

G. Wayne Talcott, PhD
Health Psychologist

Wilford Hall USAF Medical Center
San Antonio, Texas

Alan L. Peterson, PhD
Health Psychologist
Keesler USAF Medical Center
Biloxi, Mississippi

Correspondence to:
Dr D. Carl Dixon
11727 Millsway Drive
San Antonio, Texas 78253

Following the guidelines of the Craniomandibular Index, 23 chronic patients had 42 muscle and temporomandibular joint sites palpated. Two- and four-point scales were generated in response to the patient's reactions. After 6 weeks of treatment, patients measured with the 2-point scale showed 6.39% improvement in muscle tenderness and patients measured with the 4-point scale showed a 14.99% improvement. These changes were significantly different. Correlation between scales was r = .88 *originally and increased to* r = .91 *after 6 weeks. Results showed that the 4-point scale is significantly more able to detect clinical changes in muscle and joint tenderness than is the 2-point scale.*
J OROFACIAL PAIN 1993;7:403–407.

The ability to dependably measure clinical signs and symptoms is a minimum requirement for arriving at rational diagnoses and tracking treatment progress for any health problem. To accomplish these goals, the measurement process must be valid, reliable, and sensitive enough to differentiate between relatively small levels of disease severity. Increased sensitivity and precision facilitates rating the efficacies of various treatments. Despite significant effort directed at designing an accurate and sensitive measuring instrument, precise quantification of pain and dysfunction remains, at best, an unattained goal.

The Craniomandibular Index (CMI)[1,2] was developed to provide a standardized method of assessing disease severity. Fricton and Schiffman[1,2] use content, construct, and criterion validity methods to document the reliability and validity of the CMI.

However, some question regarding the CMI's ability to detect small to moderate changes in severity of muscle and joint palpation tenderness remains. The CMI uses a 2-point categoric scale (subject's report of the presence or absence of pain at each palpation site) that may limit its sensitivity to severity changes. Some data indicate that increasing the number of options or categories within a scale increases its sensitivity to severity changes.[3,4] A disadvantage of multiple categories is the possibility of introducing variability and error as the rater must subjectively assign a categoric value commensurate with the subject's response to palpation pressure. The 2-point scale would appear to be less susceptible to variability, as it is entirely subject generated.

The objective of this study was to compare the 2-point categoric scale (currently used in the CMI) with a 4-point categoric scale (expansion of CMI categories) relative to their abilities to respond to changes in muscle and joint tenderness. A secondary objective

was to assess the increased variability that may be introduced by increasing the number of categories within a scale.

Materials and Methods

Patients

Individuals (1 man, 22 women; age range 23 to 67 years; mean age 44.8 years) were selected from the patient population at the Wilford Hall Medical Center TMD Clinic, Lackland AFB, San Antonio, Texas. All complained of recurring orofacial pain (auricular, preauricular, temporal, zygomatic, ramus areas) of greater than 6 months' duration. All patients reported tenderness in 3 or more of the 36 muscle sites palpated (mean number of tender sites = 14.7; range 3 to 28). Some patients (n = 13) reported tenderness in 1 or more of the 6 joint sites palpated. Joint noise (clicking or crepitus) was found in 15 patients. Partial results of the clinical examinations are shown in Table 1. The patients were categorized on the basis of their histories and clinical exams as having primary muscle etiology (11 patients), a combination of joint and muscle etiology (11 patients), or primary joint etiology (1 patient). Patients were excluded from the study if they had received splint treatment within the previous 6 months, had a history of temporomandibular joint (TMJ) surgery, were taking psychoactive drugs, or if trauma could be associated with the onset of their symptoms.

Procedure

Each TMD patient had 36 muscle and 6 joint sites palpated in accordance with the CMI examination protocol (Table 2). Three pounds of force were applied extraorally and one pound of force was applied to all intraoral palpation sites. A single examiner performed all the clinical exams after practicing extensively to standardize his technique before beginning the study. The examiner used a

Table 1 Exam Findings and Palpation Data

Pt No.	Tnd Mus	Tnd jnt	Clk	Crp	Pain Func	DX	TX	2 Pt	4 Pt	VAS	2 Pt (6)	4 Pt (6)	VAS (6)
1	12	2	x		x	C	3	14	28	293	17	31	275
2	20	4	x		x	C	2	24	56	1952	26	46	1425
3	18				x	M	3	21	22	330	10	10	16
4	7	2		x		C	3	9	14	80	8	8	30
5	8	1			x	M	1	9	16	102	14	27	253
6	17	3	x		x	C	1	20	25	339	19	19	92
7	12	1	x	x	x	M	3	17	31	412	13	18	98
8	9	1				M	1	9	13	45	10	10	49
9	5		x			C	2	5	7	81	8	8	182
10	27	4			x	C	3	31	52	1156	11	16	137
11	12	4	x			C	1	18	23	409	19	23	414
12	11	1				M	3	13	17	339	18	25	239
13	18		x		x	M	3	19	24	171	35	37	582
14	27	2			x	M	2	31	33	698	27	30	122
15	15	2		x	x	C	3	17	23	352	17	20	84
16	25	4	x		x	C	3	30	44	1299	35	54	2078
17	28	5			x	C	2	33	58	418	19	30	62
18	11	1	x			M	1	12	17	67	19	21	194
19	6	1	x			M	1	7	7	47	9	10	43
20	17	3	x		x	C	2	20	20	412	16	22	270
21	12	2		x	x	C	2	24	27	157	18	21	234
22	19	1			x	M	1	20	26	512	8	8	56
23	3	2	x			J	1	4	4	52	5	5	44

Tnd Mus = no. tender muscle palpation sites at baseline, Tnd Jnt = no. tender joint muscle palpation sites at baseline, Clk = presence of clicking in at least one joint at baseline, Crp = presence of crepitus in at least one joint at baseline, Pain Func = pain on movement of mandible at baseline, DX = patient diagnosis (M = primary muscle etiology, J = primary joint etiology, C = combination etiology), TX = treatment group (1 = occlusal splint only, 2 = behavioral treatment only, 3 = both treatments), 2 Pt = 2-point palpation score at baseline, 4 Pt = 4-point palpation score at baseline, VAS = sum of home VAS scales for week prior to baseline, 2 Pt(6) = 2-point palpation score after 6 weeks' treatment, 4 Pt(6) = 4-point palpation score after 6 weeks' treatment, VAS (6) = sum of home VAS scales for week prior to 6-week examination.

Table 2 Palpation Sites

Muscle: extraoral	Muscle: neck	Muscle: intraoral	TMJ
Anterior temporalis	Superior SCM	Lateral pterygoid	Lateral capsule
Mid temporalis	Mid SCM	Medial pterygoid	Posterior capsule
Posterior temporalis	Inferior SCM	Temporalis insert	Superior capsule
Deep masseter	Upper trapezius		
Anterior masseter	Splenius capitus		
Inferior masseter			
Posterior digastric			
Medial pterygoid			
Vertex			

counterbalance scale to calibrate his palpation force at the one and three pound force levels prior to each examining session.

Outcome Measures

Three assessments of disease severity were made: a 2-point categoric scale palpation index, a 4-point categoric scale palpation index, and a subjective rating of pain (daily pain diary).

Two-point categoric scale palpation indexes (0 = no reported pain and 1 = reported pain) were generated at each CMI muscle and joint palpation site at baseline and again after 6 weeks. The values assigned to each muscle and joint site palpated were added together to obtain one overall score for muscle and joint tenderness.

Four-point categoric scale palpation indexes were generated simultaneously with the 2-point indexes for each patient at baseline and again after 6 weeks of treatment. The examiner graded each patient's response on a scale of 0 to 3: 0 = no pain, 1 = verbal report of discomfort, 2 = facial movement such as palpebral reflex along with report of pain, and 3 = retreating of head in anticipation of palpation along with report of considerable pain. The sums of the positive responses were added together to obtain one score for muscle and TMJ tenderness.

Subjective assessments of symptoms were made on a visual analog scale (VAS)[5] by each patient 3 times a day (morning, afternoon, and evening) for 7 days at baseline and again for 7 days immediately prior to the 6-week follow-up appointment. The VASs used in this study consisted of standard 100-mm-long lines anchored at the left end by the phrase "No Pain" and on the right end by the phrase "Most Intense Pain Imaginable." Patients marked the 100-mm line at the point that best quantified their perceived pain. The distance from the "No Pain" end of the scale to the patient's mark was measured to the nearest mm on each scale. The measurements were summed to provide a subjective index of each patient's pain over the 7-day intervals.

This report is part of a large study in which patients received one of three treatments: *(1)* an occlusal splint (eight patients), *(2)* intensive instruction in relaxation and stress management (six patients), and *(3)* both an occlusal splint and intensive instruction in stress management and relaxation (nine patients).

Results

The degree of relationship between the two scales was assessed by correlations at both baseline and after 6 weeks of treatment. The correlation between the 2-point scale and the 4-point scale at baseline was $r = .88$ and at 6 weeks was $r = .91$ ($P < .001$). These results suggest that both scales are measuring the same content.

Using analysis of variance, the pretreatment and posttreatment scores for the 4-point scale were significantly greater than the differences for the 2-point scale ($P < .038$).

Mean improvement in the 2-point scale scores between the baseline and 6-week examination was 6.39%, while improvement in the 4-point scale scores was 14.99%. Another measure used to determine improvement was the home VAS which showed a mean improvement of 28.22% between baseline and the 6-week examinations, although this difference was not significantly different.

Discussion

The 4-point categoric scale reflected a significantly greater decrease in muscle and joint tenderness at the 6-week follow-up examination than did the 2-point categoric scale (Table 3). This implies that

Table 3 Results Summary

	Mean (range) scale score Wk 0	Mean (range) scale score Wk 6	Improvement in scale scores (%) Wk 0 - Wk 6	Coefficient variation Wk 0	Coefficient variation Wk 6
2-Point Scale	17.70 (4–33)	16.57 (5–35)	6.39	48.30	49.07
4-Point Scale	25.52 (4–58)	21.70 (5–54)	14.99*	58.20	57.66
VAS	422.74 (45–1952)	303.44 (16–2078)	28.22	110.40	160.37

*Significantly different from 6.39% ($P < .038$).

the 4-point categoric scale, as used with the CMI, is significantly more sensitive to changes in muscle and joint tenderness; the larger improvement in this scale more closely paralleled the improvement subjectively reported by the patients than did the 2-point scale.

The high degree of correlation between the 2-point and the 4-point categoric scales ($r = .88$ at baseline and $r = .91$ at 6 weeks) suggests that both scales are measuring the same entity and that the 2-point scale and the 4-point scale are similarly valid.

Conceptually, the 2-point scale is simple and minimizes the opportunity for interpretation errors to occur. The 4-point scale requires the examiner to interpret the patient's response to palpation, which could increase variability in outcome scores. However, in this study, the high correlation between the two scales and the similar coefficients of variation (Table 3) for the 2- and 4-point scales indicate that examiner-induced confounding factors had a minimal effect on increasing variability in the 4-point scales. By the same reasoning, it is apparent that a potential rise in variability that could have resulted from the increased degrees of freedom in the 4-point scale did not occur.

In this study all patients were examined by the same individual who was previously trained in the examination technique. This may account for some of the consistency found, as interexaminer variability was eliminated. Inconsistencies in the amount of pressure, the palpation technique, the size of the distal phalanx, and the specific anatomic area palpated will introduce variability, particularly between different raters. Dworkin et al[8] show that untrained examiners had much lower levels of reliability, while Fricton and Schiffman[1] report that for trained examiners the intrarater and interrater reliability were high and comparable. This reinforces the importance of standardization and calibration of examination technique.

However, a standard palpation technique has not been universally practiced, making comparisons across studies impossible. Three pounds of palpation force was used in this study as a middle ground between previous studies. Gross and Gale[7] use 3 pounds of palpation force in a large epidemiologic study and were among the first to recognize the necessity of standardizing palpation technique. The examiners in that study self-calibrated themselves by using a postage scale prior to, and periodically during, the study. Fricton and Schiffman[1] recommend the use of 1 pound per square inch palpation pressure. The palpation force required to produce 1 pound per square inch of pressure over the area of the finger tip (approximately 0.2 in^2/1 cm^2) is 0.15 pounds per cm^2, a palpation force likely to result in a high rate of false-negative findings. Dworkin et al[8] use 2 pounds of palpation force (2 pounds per cm^2 of pressure), stating that 3 pounds produced a high false-positive rate. Goulet and Clark,[9] in a study using a pressure algometer, find that higher pressures (4.4 to 5.72 pounds per cm^2) produce more consistent results. That study implies that the palpation force used in the present study may be too low. During a pilot study and efforts to standardize examination technique, a low rate of positive palpation findings was found when 2 pounds of palpation force was used. Others[10,11] propose the use of pressure pain thresholds, defined as the force level at which a slowly increasing force applied by a pressure algometer results in a change of sensation from pressure to pain. Obviously, a need exists to reach a consensus on a standardized palpation technique that can be used across research designs, rather than the multitude of techniques currently in use.

Subjects in this study may be considered to be their own controls in that the two indexes were applied to each subject at the same point in time. Therefore, the differences found between the indexes are intrapatient differences and are free from error that could occur from comparing changes between different subjects or between patients and nonpatient controls.

Conceivably, different outcome scores between the three treatment groups included in this study

could introduce confounding error. However, there was no statistically significant difference in palpation score outcome during the course of the study between the three groups (ANOVA, time × treatment, $P = .525$).

Similarly, the inclusion of different diagnostic subgroups in the study population could potentially confound the results. Again, when the palpation scores of the primary muscle etiology group were compared to those of the combination muscle and joint etiology group, no significant difference in palpation scores was found over the course of the study (ANOVA, time × diagnostic group, $P = .949$).

The results of this study suggest that increasing the number of choices within a categoric scale enhances its sensitivity to detecting changes in muscle and joint tenderness. Increasing the number of categories beyond four may additionally enhance sensitivity. However, there is likely an upper limit beyond which increasing the number of categories will indeed induce unacceptable variability and error.

Since this study dealt primarily with quantifying muscle and joint palpation tenderness in general, it is likely that its findings are not specific to the CMI and that 4-point scales would be more sensitive than 2-point scales when used with palpation indexes other than the CMI.

Conclusion

In chronic TMD patients, the 4-point categoric scale is significantly more sensitive to changes in muscle and joint tenderness than is the 2-point categoric scale. In future treatment comparison studies, the use of a 4-point categoric scale may allow identification of smaller differences in treatment efficacies than scales with fewer categories.

References

1. Fricton JR, Schiffman EL. Reliability of a craniomandibular index. J Dent Res 1986;65:1359–1364.
2. Fricton JR, Schiffman EL. The craniomandibular index: Validity. J Prosthet Dent 1987:58:222–228.
3. Downie WW, Leathan PA, Rhind VM, Wright V, Branco JA, Anderson JA. Studies with pain rating scales. Annals Rheum Dis 1978;37:378–381.
4. Hart FD, Huskisson EC. Measurement in rheumatoid arthritis. Lancet 1972:Jan 1:28–30.
5. Huskisson EC. Visual analog scales. In: Melzack R (ed). Pain Measurement and Assessment. New York: Raven Press, 1983:33–37.
6. Dworkin SF, LeResche L, DeRouen T. Reliability of clinical measurement in temporomandibular disorders. Clin J Pain 1988;4:89–99.
7. Gross A, Gale EN. A prevalence study of the clinical signs associated with mandibular dysfunction. J Am Dent Assoc 1983;107:932–936.
8. Dworkin SF, Huggins KH, LeResch L, Von Korff M, Howard J, Truelove EL, et al. Epidemiology of signs and symptoms in temporomandibular disorders: Clinical signs in cases and controls. J Am Dent Assoc 1990;120:273–281.
9. Goulet JP, Clark GT. Clinical TMJ examination methods. J Calif Dent Assoc 1990;18:25–33.
10. Jensen K, Anderson HO, Olesen J, Lindblom U. Pressure pain threshold in human temporal region. Evaluation of a new pressure algometer. Pain 1986;25:313–323.
11. Orbach R, Gale EN. Pressure pain thresholds in normal muscles: Reliability, measurement effects and topographic differences. Pain 1989;37:257–263.

Resumen

Comparación de Dos Escalas en la Evaluación de la Sensibilidad a la Palpación de los Músculos y las Articulaciones en los Desórdenes Temporomandibulares Crónicos

Se realizó la palpación de 42 sitios (músculos y articulaciones temporomandibulares) en veintitrés pacientes crónicos, siguiendo las guías del Indice Craneomandibular. En cada sitio, se generó una escala de 2- y 4- puntos como respuesta a las reacciones del paciente. Después de 6 semanas, la escala de 2-puntos reveló una mejoría del 6,39% en cuanto a la sensibilidad muscular y la escala de 4- puntos revelo una mejoria del 14,99%. Estos cambios fueron significativamente diferentes. La correlación entre las escalas fué de 0,88 originalmente y aumentó a 0,91 a las 6 semanas. Los resultados indican que la escala de 4- puntos está significativamente en mejor capacidad para detectar cambios clínicos en cuanto a la sensibilidad muscular y articular, en comparación a la escala de 2- puntos.

Zusammenfassung

Vergleich von zwei Skalen zur Einteilung von Muskel- und Gelenkpalpationsempfindlichkeit bei Patienten mit chronischen Myoarthropathien des Kausystems (MAP)

Bei 23 Patienten mit chronischen MAP wurden 42 Muskel- und Kiefergelenkstellen gemäss den Richtlinien des Craniomandibular Index palpiert. Jede Stelle wurde auf einer 2- respektive 4-Punkte-Skala gemäss den Reaktionen des Patienten bewertet. Nach 6 Wochen zeigte die 2-Punkte-Skala eine Verbesserung der Muskelpalpationsempfindlichkeit von 6.39%, die 4-Punkte-Skala eine solche von 14.99%. Die Veränderungen unterschieden sich signifikant. Die Korrelation zwischen den Skalen war am Anfang 0.88, nach 6 Wochen 0.91. Die Resultate zeigten, dass die 4-Punkte-Skala signifikant besser in der Lage ist, klinische Veränderungen in der Muskel- und Gelenkpalpationsempfindlichkeit sichtbar zu machen als die 2-Punkte-Skala.