The Reproducibility of Muscle and Joint Tenderness Detection Methods and Maximum Mandibular Movement Measurement for the Temporomandibular System

Jean-Paul Goulet, DDS, MSD
Professor
Département de Stomatologie
Faculté de Médecine Dentaire
Université Laval
Cité Universitaire
Québec, Canada

Glenn T. Clark, DDS, MS
Professor
UCLA School of Dentistry
Section of Diagnostic Sciences and Orofacial Pain

Virginia F. Flack, JD, PhD
Associate Professor
UCLA School of Public Health
Department of Biostatistics

Changriu Liu, DDS
Lecturer
UCLA School of Dentistry

University of California Los Angeles
Los Angeles, California

Correspondence to:
Dr Jean-Paul Goulet
Département de Stomatologie
Faculté de Médecine Dentaire
Université Laval
Cité Universitaire
Québec, Canada G1K 7P4

The purpose of this study was (1) to evaluate the reproducibility of two masticatory muscle and joint tenderness detection methods; (2) to evaluate the reproducibility of maximum mandibular movement measurements; and (3) to investigate factors influencing examiner agreement. The tenderness assessment procedures involved application of a standard pressure for 2 seconds over four anatomically defined masticatory muscle sites, one control forehead site, and two temporomandibular joint sites on each side of the face. One technique utilized a pressure algometer (PAP), while the other technique required that a trained examiner apply pressure with the index fingertip (FPP). Seventy-two subjects (36 patients and 36 controls) were evaluated in a single-blind study design. Control subjects were matched for age, gender, and race with temporomandibular disorder subjects. Each subject was examined twice with each of the described methods in a randomized, fully balanced sequence by calibrated examiners. Tenderness levels were determined by the subject via self-report of pain upon pressure using a standard set of verbal descriptors. Maximum pain-free, active, and passive opening, and maximum active right and left lateral movements were measured using a millimeter ruler. Intraclass correlation coefficients (ICC) for the tenderness assessment methods ranged from 0.220 to 0.739 for the FPP method and from 0.391 to 0.880 for the PAP method. ICCs for mandibular movement measurement were much less variable, ranging from 0.59 to 0.68 for lateral movement and from 0.78 to 0.93 for opening movement. These results indicate good to excellent agreement between calibrated examiners for mandibular movement measurement and for tenderness assessment methods at two masseter (ie, superficial and deep) and the anterior temporalis sites. Only fair agreement was found for the middle temporalis and lateral TMJ capsule sites using these methods.

J OROFACIAL PAIN 1998;12:17–26.

key words: interexaminer reliability, muscle tenderness assessment, mandibular movement measurements

Jaw pain is the main reason patients with temporomandibular disorders (TMD) seek treatment.[1] Documenting the source of the pain, whether it originates from the masticatory muscles and/or the temporomandibular joints (TMJs), remains an important component of the clinical examination for diagnosis and treatment decisions. A wide variety of testing devices are available to assess musculoskeletal structures. The most popular in the TMD field are electronic devices aimed at documenting joint sounds, electromyographic activity, and range and pattern of mandibular movements.[2,3] The assessment of muscle and joint tenderness, on the

other hand, is still accomplished through more conventional techniques. Recently published TMD examination guidelines[4] recommend that masticatory muscles and TMJs be palpated with the index fingertip, using a specific amount of pressure. Also recommended is the use of a millimeter ruler for recording the amplitude of all vertical and horizontal mandibular movements. Simple and straightforward techniques like these are also subject to reliability and validity standards.

Over the years, these issues have been studied for quantification of signs and symptoms of TMD. Most studies report relatively good intraexaminer and interexaminer agreement for assessment of mandibular range of motion with a millimeter ruler.[5–12] This has not been the case, however, for muscle and joint tenderness assessment using finger palpation. The technique is by itself highly dependent on the amount of pressure exerted by the examiner and on the subject's self-report of pain. A number of studies claim acceptable to fairly good intraexaminer and interexaminer agreement,[6,8,9,13] while others have conceded to marginal reproducibility when the study design complied with the basic requirements needed to look at reliability issues.[5,7,10,11]

Pressure algometry was developed and refined in an attempt to better quantify muscle tenderness through measurement of pressure pain threshold (PPT) and pressure tolerance, in healthy subjects and in patients suffering a variety of musculoskeletal disorders.[14–25] Determination of PPT with a manual pressure algometer is influenced by several factors, such as the rate of pressure application, the reaction time of both the subject and the examiner as the pain arises, and the subjectivity of the pain report.[15,26] In addition, the range of PPT varies between genders, muscle groups, and individuals.[18,19,22,23,25] Although determination of PPT is subject to variation, reliability studies have generally shown good-to-excellent intraexaminer and interexaminer concordance.[15,16,19,21,22]

Whether one chooses finger palpation or PPT determination for muscle and joint tenderness assessment, because they rely merely on the examiner's abilities and on the subject's self-report of pain, both procedures have shortcomings. Today, manual palpation is still considered the gold standard[3,4,27–29] and remains the most widely used technique: it allows the examiner to feel the muscle sites and to have better control over the force vector when applying pressure, and it is, overall, a cost-effective procedure. Any improvement aimed at increasing the reliability of muscle and joint tenderness assessment methods would certainly be beneficial for patients, clinicians, and researchers. One step in this direction is the application of a specific amount of pressure over predetermined anatomic sites known for their diagnostic value. More reliable clinical assessment methods could eventually help improve case definition of TMD subtypes whenever pain is the predominant feature.

The present reliability study was designed to compare the consistency of two muscle and joint tenderness assessment procedures while attempting to determine whether the observed variability was a result of the subject's self-report, the method of examination, or the transient nature of the symptoms. Knowledge of the factors influencing the reliability of these methods can provide a basis for developing better standardized calibration protocols and cost-effective examination procedures, and thereby enhance the validity of future controlled clinical studies.

The specific purposes of this study were *(1)* to measure and compare the reliability of finger palpation and pressure algometry for the assessment of muscle and joint tenderness; *(2)* to test the interexaminer consistency of one method of mandibular movement measurement; and *(3)* to investigate factors influencing examiner agreement consistency and patient reliability. Interexaminer consistency refers to the proportion of agreement among examiners when making observations of the same variable. Patient reliability is defined as the intrapatient consistency to report the presence or absence of a symptom, usually pain or discomfort, whenever repeated examinations are performed under controlled conditions.

Materials and Methods

Study Sample

A total of 72 subjects (62 women and 10 men) participated in this study; 36 patients were matched for sex, age (within 3 years), race, and education level with an equal number of controls. The mean age (± one standard deviation) was 29.0 (± 6.0) years. The TMD subjects were recruited from the University of California at Los Angeles TMJ Clinic and from newspaper advertisements. Potential candidates were asked to participate in the study if they had a history of continuous or intermittent pain in the face, temple, or jaw region of at least 4 months' duration occurring a minimum of three times a week; if they had a usual pain level rated 10 or more on a 100-mm visual analogue scale (VAS); and/or if they stated they would seek or were seeking treatment for their jaw pain and/or dysfunction. Controls were selected from subjects responding to an advertisement posted in the UCLA Dental School and in the University

Daily Journal. They were recruited if they had a negative history when questioned about jaw pain and/or dysfunction problems. Potential control subjects reporting occasional joint sounds were selected if they never had any joint pain or limitation and did not think they had a problem for which they would seek treatment. Exclusion criteria for both groups included the presence of sinus disease, dental infection, ear disorder, moderate to severe episodic headaches, systemic inflammatory polyarthritis, or moderate to severe cervical pain or dysfunction.

Examination Procedures

All of the examination methods under investigation were performed twice on each participant. A total of five dentist examiners were involved. One examiner (the project coordinator) performed each examination using the pressure algometer for muscle and joint pain assessment (Method A). Two different examiners, selected randomly from a pool of four, performed the finger palpation examinations (Method B) and the mandibular movement recordings. The latter group of examiners were blinded to the status of the subject (ie, control versus patient). Three examination sequences were used in a random fashion, allowing for investigation of an order effect of each exam (A-B-A-B; A-B-B-A; B-A-B-A).

All participants in the study gave their full consent and signed the approved Human Subject Form. At the beginning of the experimental session, subjects were advised by the project coordinator (PC) to follow the examiner's instructions carefully, to report their level of pain as accurately as possible, and to suppress information regarding their clinical status at all times. The PC was present during all examination procedures to ensure that these instructions were adhered to and that no indication or clue as to a subject's status was left in sight. Before the first examination, the PC used self-adhesive paper dots to mark the five palpation sites (forehead, superficial masseter, deep masseter, anterior temporalis, and lateral capsule of the TMJ), referred to as "fixed-reference points," on the randomly selected side of the subject's face to ensure perfect site replicability during the four examinations. For practicality, the middle temporalis site was not marked. All unmarked sites on the contralateral side were referred to as the "reference calibration points." The PC then introduced the first examiner to the subject, who underwent the first examination. Each subject was examined according to the predetermined sequence, once by the two examiners from the pool and twice by the PC. The level of pain intensity induced by the pressure at each palpation site was rated by the subject via self-report using a verbal descriptor scale (VDS), described below. Each examination was followed by a 15-minute rest, during which no eating or gum chewing was allowed.

Pressure Algometer Pain Assessment

Six bilaterally predetermined muscle and joint sites (marked and unmarked) were assessed for the presence of pain with the pressure algometer pain (PAP) assessment technique. The pressure algometer consisted of a metal plunger with a rubber tip 1 cm in diameter mounted on a gauge calibrated in kilograms. The plunger tip was placed parallel to the skin surface before the selected pressure was applied for 2 seconds. The predetermined standard pressures were as follows: 1.8 kg/cm^2 to the forehead (control site), the superficial and deep masseter, and the anterior temporalis; 2.5 kg/cm^2 to the middle temporalis; and 0.8 kg/cm^2 to the lateral capsule. The amount of pressure for the muscle sites was chosen from receiver operator characteristic curves (ROC) determined in a previous pilot study assessing PPT in healthy and TMD subjects.[27] As for the lateral capsule, the pressure selected was empirically chosen since no data were then available on PPT in small joints. During the procedure, the examiner braced the subject's head with the open palm of the opposite hand to prevent any displacement while applying pressure. Subjects kept their teeth together so they touched slightly during muscle site testing. They were asked to bite gently on a tongue blade placed between the premolars during the lateral capsule palpation.

Fingertip Pressure Pain Assessment

The fingertip pressure pain (FPP) method consisted of applying a standard pressure to each muscle and joint site with the palm of the index finger. In addition to the sites already mentioned for the PAP assessment method, the dorsal capsule was palpated through the external auditory meatus on both sides. The examiners were trained and calibrated according to the protocol described in a previous paper.[30] When the study started, all had shown the ability to apply a standardized pressure for 2 seconds within the range of 1.5 to 2.1 kg/cm^2 for the muscle sites, and 0.5 to 1.1 kg/cm^2 for the TMJ.

Palpation Sites

On each side of the face, the palpation sequence was the forehead, superficial masseter, deep masseter, anterior temporalis, middle temporalis, lateral capsule, and dorsal capsule. The examiners were

trained to localize the palpation sites using the following guidelines:

A. Forehead control site (F-head): 15 mm above the orbital rim along a vertical line drawn perpendicular to a line joining both canthus of the eye at a point 15 mm lateral to the inner canthus
B. Superficial masseter site (S mass): 20 mm from the angle along a line joining the angle of the mandible and the alar of the nose
C. Deep masseter site (D mass): In the depression at the level anterior to the upper half of the earlobe, just behind the posterior border of the superficial masseter, located after the subject was asked to clench and then relax the jaw
D. Anterior temporalis site (A temp): A point 30 mm posterior to the most lateral part on the orbital rim and 15 mm above the upper edge of the zygomatic arch
E. Middle temporalis (M temp): A point 60 mm along a vertical line drawn perpendicular to the tragus-canthus line from the middle of the external auditory meatus
F. Lateral capsule (L caps): The lateral pole of the condyle located just anterior to the tragus while the subject bites on a tongue blade in the premolar region
G. Dorsal capsule (D caps): Through the external auditory meatus while facing the subject after the subject was instructed to open the mouth about one finger width and then close the teeth together, at which point the finger pressure was exerted forward against the posterior aspect of the condyle with the teeth in intercuspal position

Pain Word Descriptors

Following each PAP and FPP testing, subjects were asked if they felt any pain (yes or no) and, if so, to select the appropriate word from the VDS that best described its intensity. The anchor words from the Present Pain Intensity (PPI) index of the McGill Pain Questionnaire were selected for the VDS.[31] This list and the corresponding scores later used to compute the results were: none = 0, mild = 1, discomforting = 2, distressing = 3, horrible = 4, excruciating = 5.

Mandibular Movement Measurements

Subjects' maximum pain-free opening (PFO), maximum active opening (MAO), maximum passive opening (MPO), maximum right and left lateral excursion (MRL and MLL), overjet (OJ), and overbite (OB) were measured using a millimeter ruler. The amplitudes of the three vertical openings were recorded by measuring the distance between the midincisal edges of the right maxillary and its corresponding tooth on the opposing arch. Following specific standardized instructions, the PFO was measured first, followed by the MAO and then the MPO. The latter measurement was made when the mouth was pried open until resistance was felt with the thumb and the index finger placed over the incisal edge of the incisors. For the maximum lateral excursions, a lead pencil mark corresponding to the midline between the mandibular incisors was drawn on the labial surface of the opposite upper incisor. The subject was then asked to move the mandible sideways as far as possible. The examiner measured and recorded the distance from the midline between the lower central incisors and the line drawn on the upper central incisor. The overjet was measured by placing the edge of the ruler on the labial surface of the right mandibular incisor and recording the horizontal distance to the corresponding labial surfaces of the right maxillary incisor with the teeth in maximum intercuspal position. Finally, for the overbite, a horizontal line was traced on the labial surface of the mandibular incisor at the level of the incisal edge of the upper incisor in intercuspal position; the overbite was determined by measuring the distance between the line and the incisal edge of the lower incisor.

Examiner Calibration

Prior to the study, each examiner received detailed guidelines regarding the examination protocol, and each was trained and calibrated by the PC according to these guidelines. For mandibular movement measurements, the performance criteria for successful calibration included: *(1)* accurate positioning of the millimeter ruler for the vertical mouth opening measurements; *(2)* correct use of the verbal instruction script to the subject; and *(3)* proper tooth marking and ruler positioning for measurement of the lateral excursions, overjet, and overbite. Successful calibration for muscle palpation was achieved when examiners consistently: *(1)* localized palpation sites within 7 mm of the anatomic locations described earlier; and *(2)* applied fingertip pressure within the range of 1.5 to 2.1 kg/cm^2 and 0.5 to 1.1 kg/cm^2 in at least 8 trials out of 10. Examiners were recalibrated for finger pressure on a weekly basis during the course of the study to control for the passage of time.

Data Reduction and Analysis Procedures

The examiner reliability measures were computed using the data of two clinical observations per patient. For the mandibular movement measure-

Table 1 Pain Word Scores (range 0–5) for the Masticatory Muscle and TMJ Sites Using the Pressure Algometer (PAP) and the Finger Pressure (FPP) Method (Mean and One Standard Deviation)

	PAP method				FPP method			
	1st exam*		2nd exam*		1st exam*		2nd exam*	
Site	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
S mass	1.49	(1.15)	1.70	(1.26)	0.98	(1.03)	1.08	(1.25)
D mass	1.37	(1.08)	1.49	(1.23)	1.03	(1.02)	1.12	(1.26)
A temp	0.71	(0.91)	0.73	(0.99)	0.62	(0.91)	0.56	(0.94)
M temp	0.60	(0.77)	0.66	(0.93)	0.35	(0.54)	0.37	(0.66)
L cap	0.23	(0.39)	0.34	(0.63)	0.42	(0.79)	0.39	(0.70)
D cap	NA		NA		0.30	(0.65)	0.35	(0.72)
F-head	0.60	(0.83)	0.63	(0.81)	0.26	(0.52)	0.21	(0.45)

*Average value for right and left sides derived from both the control and TMD subjects (n = 72).
S mass = superficial masseter; D mass = deep masseter; A temp = anterior temporalis; M temp = middle temporalis; L cap = lateral capsule; D cap = dorsal capsule; F-head = forehead control.

Table 2 Mean Slope (and Two-Tailed *P* Value) Calculated for Repeat Masticatory Muscle and TMJ Sites Examination Data From 72 Subjects*

	Right		Left	
Site	Mean slope	*P*	Mean slope	*P*
S mass	.120	.01	.076	.11
D mass	−.076	.54	.105	.04
A temp	.000	1.00	.027	.27
M temp	.049	.25	.006	.90
L cap	.063	.09	.001	.99
D cap	NA	NA	NA	NA
F-head	.012	.77	.028	.42

*Statistiically significant if $P \leq .0042$.

ments and the FPP assessment methods, the sample design for the 36 patients and matched controls was a balanced, incomplete-block experiment, with exactly two of the four trained examiners seeing each subject. Balancing for the frequency of the possible examiner pairs was performed. The data from repeated measures for mandibular movements, FPP, and PAP at marked and unmarked palpation sites were compared across sessions and analyzed for level of intraexaminer and interexaminer agreement using the Intraclass Correlation Coefficients (ICC) analysis (see discussion below). Both the subject and the examiner effects (two-way analysis of variance [ANOVA]) were accounted for when computing the ICC values for interexaminer reliability (mandibular movements and FPP data).[32] The ICC values computation for the intraexaminer reliability (PAP data) was based on a one-way ANOVA model with a subject effects factor.[33] The range of ICC values defining each level of examiner agreement were: 0.00 to 0.39 = unacceptable, 0.40 to 0.59 = moderate, 0.60 to 0.79 = good, 0.80 to 1.00 = excellent.

Results

The distribution of the first and second examination scores for the PAP and FPP muscle tenderness assessment method conducted, respectively, by the same and pairs of examiners are summarized in Table 1. To test for the presence of an examination order effect, the data were analyzed using a linear fit of exam score versus exam order per subject. The mean slopes of the lines were tested against the hypothesis of average slope equals zero using a two-tailed *t* test. The results revealed that the right superficial and the left deep masseter had the steepest slopes across the four examinations (Table 2), with increments of VDS scores of 0.32 and 0.36, respectively, after four examinations. After adjusting the level of significance for the 12 comparisons being performed, it was concluded that there were no statistically significant order effects.[34]

The ICC values for the PAP scores at marked and unmarked sites revealed good to excellent intraexaminer agreement at four of five sites (Table 3). Only

Table 3 Intraclass Correlation Coefficient Values (and Level of Agreement) From Repeat Examination at Marked and Unmarked Masticatory Muscle and TMJ Sites (n = 72 Subjects)

	Marked sites		Unmarked sites	
Site	PAP	FPP	PAP	FPP
S mass	0.73 (G)	0.70 (G)	0.81 (E)	0.69 (G)
D mass	0.61 (G)	0.70 (G)	0.87 (E)	0.67 (G)
A temp	0.85 (E)	0.62 (G)	0.85 (E)	0.71 (G)
M temp	NA	NA	NA	NA
L cap	0.41 (M)	0.55 (M)	0.53 (M)	0.57 (M)
D cap	NA	NA	NA	NA
F-head	0.67 (G)	0.31 (U)	0.64 (G)	0.24 (U)

PAP = pressure algometer; FPP = finger pressure; E = excellent; G = good; M = moderate; U = unacceptable.

Table 4 Intraclass Correlation Coefficient (and Level of Agreement) From Repeat Examination for Right and Left Masticatory Muscle and TMJ Sites (n = 72 Subjects)

	PAP method			FPP method		
	Right	Left	Mean	Right	Left	Mean
S mass	0.79 (G)	0.76 (G)	(0.77) (G)	0.72 (G)	0.66 (G)	(0.69) (G)
D mass	0.77 (G)	0.71 (G)	(0.74) (G)	0.66 (G)	0.72 (G)	(0.69) (G)
A temp	0.82 (E)	0.88 (E)	(0.85) (E)	0.61 (G)	0.74 (G)	(0.68) (G)
M temp	0.58 (M)	0.56 (M)	(0.57) (M)	0.30 (U)	0.58 (M)	(0.44) (M)
L cap	0.39 (U)	0.57 (M)	(0.48) (M)	0.66 (G)	0.45 (M)	(0.55) (M)
D cap	NA	NA	NA	0.70 (G)	0.22 (U)	(0.46) (M)
F-head	0.61 (G)	0.70 (G)	(0.66) (G)	0.23 (U)	0.30 (U)	(0.27) (U)

PAP = pressure algometer; FPP = finger pressure; E = excellent; G = good; M = moderate; U = unacceptable.

the lateral capsule site showed moderate reliability. Over the superficial and deep masseter sites, better reliability was observed at unmarked than at marked sites even if accurate site and pressure replication could be achieved. The interexaminer agreement at marked and unmarked sites with the FPP technique was comparable despite the fact that finger palpation may not have been done precisely on the same spot at unmarked sites. The replicability was good for the anterior temporalis, superficial, and deep masseter, moderate for the lateral capsule, and unacceptable for the control forehead site.

The right and left ICC values for the repeat PAP and FPP assessment are presented in Table 4. Compared to the FPP method, the PAP showed higher ICC values at 8 of 12 sites. Both methods showed similar levels of agreement in terms of category in half of the sites. The agreement levels were one and two categories apart, respectively, at four and two palpation sites. The two masseter (superficial and deep) and the anterior temporalis muscle sites showed good to excellent reproducibility (ICC range: 0.61 to 0.88), with little difference between the right and left side in either method. Differences in the levels of agreement between the right and left sides were more frequently observed with the FPP (three sites) than the PAP method (one site). Except for the anterior temporalis and the forehead sites, the differences in the mean ICC values at each site between both methods were not enough to change the overall level of agreement in terms of category. The forehead control site showed the greatest difference in ICC values between both methods, clearly indicating an unacceptable level of agreement using the FPP method.

The means and one standard deviation of the mandibular movement measurements derived from the two subject samples (36 controls and 36 TMD cases) are presented in Table 5. As shown in Table 6, ICC values for the MAO and MPO were excellent (ICC >

Table 5 Average Value (and SD) Derived From Repeat Examination for Mandibular Position Measurements

	Controls (n = 36)		TMDs (n = 36)	
	Mean (mm)	(SD)	Mean (mm)	(SD)
PFO + OB	49.1	6.6	39.8	10.2
MAO + OB	52.6	6.3	50.9	6.6
MPO + OB	54.7	6.1	53.1	6.2
MRL	10.1	1.8	9.5	2.0
MLL	9.9	2.2	9.7	1.8

PFO = pain-free opening; MAO = maximum active opening; MPO = maximum passive opening; OB = overbite; MRL = maximum right lateral; MLL = maximum left lateral.

Table 6 Intraclass Correlation Coefficient Values (and Level of Agreement) for Repeat Mandibular Position Measurements (n = 72 Subjects)

Mandibular position measurements	ICC value
Pain free opening	0.78 (G)
Maximum active opening	0.87 (E)
Maximum passive opening	0.93 (E)
Maximum right lateral	0.59 (M)
Maximum left lateral	0.68 (G)
Overbite	0.72 (G)
Overjet	0.73 (G)

ICC = intraclass correlation coefficient; E = excellent; G = good; M = moderate; U = unacceptable.

0.8), a level of agreement not quite achieved with the PFO recording, which remained the most variable measurement of the three vertical openings. Except for the recording of the maximum right lateral excursion, the reliability of all other jaw position measurements was good, with ICC values ranging from 0.68 to 0.73. The right lateral excursion almost achieved this level of agreement, with an ICC value of 0.59.

Discussion

In this study, the authors tried to fulfill most of the basic requirements cited by Chilton[35] as necessary for a reliability study. All examiners conducting the FPP assessment were blinded to the status of the subjects. They were carefully trained and calibrated for both the examination techniques and the instructions given to the subject. A sufficient number of patients with clear signs and symptoms of TMD were included, and the choice for the statistical analysis took into account type of variables, agreement by chance, and correction for systematic error. Cutoff points for range of ICC values defining excellent, good, moderate, and unacceptable agreement were predetermined. Overall, the study design attempted to achieve better understanding of the source of the variability by looking at specific factors possibly accounting for the observed interexaminer inconsistency. For example, is the patient's self-report shifted on repeated examination because the pressure was slightly different or not applied precisely over the same spot?

Regarding the statistical analysis, the rationale for choosing ICC rather than Kappa statistics was twofold. First, consideration had to be given to the fact that a pair of observations in adjacent categories indicated less disagreement than a pair of observations two or more categories apart. One alternative is to use weighted Kappa, but the results are greatly influenced by the relative magnitude of the weight given to each category. According to Cohen,[36] when a standard weight corresponding to the square of the deviation of the pair of observations from exact agreement is given in each instance of disagreement, the weighted Kappa approximately equals the product-moment correlation coefficient. Maclure and Willett[37] find that a logical alternative for polychotomous data is to use the ICC analysis. Second, we asked each subject to rate, on a 100-mm VAS at the beginning and at the end of the study, the pain intensity that best corresponded to each verbal pain descriptor. Computation of the data supported the linearity assumption for the VDS categories derived from the McGill Pain Questionnaire.

Previous reliability studies have generally not reported on the replicability of muscle tenderness at specific anatomically defined sites when examiners were blinded to the subject's status. In most studies, the palpation method was poorly described in terms of amount of pressure and technique. The present results reveal that it is possible, with rigorous examiner calibration, to have reproducible muscle tenderness scores (see the results for the superficial and deep masseter and the anterior temporalis) over specific sites with either method. When compared to the pressure algometer method, the nearly equal performance of the traditional fingertip technique strongly underscores the effectiveness of the finger pressure calibration procedure at reducing the inherent variability of the technique. This performance is somewhat remarkable considering the fact that the same examiner performed the repeat PAP measurements, and thus higher intraexaminer than interexaminer

reliability is expected. One advantage of the fingertip method is the tactile feedback the examiner receives, while the main disadvantage is the uncertainty regarding the loading forces being used.

All sites evaluated in this study were not of equal reproducibility. Looking at the mean right and left ICC values, the middle temporalis and both TMJ sites had only moderate reproducibility, a trend seen with both methods (see Table 4). This suggests that, whatever the technique, the reproducibility is likely site- or tissue-dependent. Our attempt to decrease the variability by marking palpation sites did not lead to a consistent distribution of better ICC values over marked versus unmarked sites; the opposite was somehow more frequent with the PAP method (see Table 3). These results are interpreted to mean that perfect or near-perfect replicability of muscle tenderness from examination to examination is probably impossible, perhaps due merely to the inherent variation in the patient's response more than to variation in other methodologic factors when there is proper calibration for pressure level and site location. A ceiling effect should therefore be expected for the replicability of pressure tenderness assessment methods. Even so, the ICC scores are high enough at some sites to be used as potentially good outcome variables in longitudinal studies, assuming there is ongoing examiner calibration. In fact, the ICC values for the two masseter and the anterior temporalis pain scores are slightly less than the values obtained for the mandibular opening measurements, which are considered to be the most reliable clinical measures in TMD patients.

It should be noted that the mean pain word scores were slightly higher for the PAP than for the FPP method on all muscle sites (see Table 1). In fact, the score for the middle temporalis sites should have been higher with the pressure algometer because more pressure was used with this method by design. The clear elevations across almost all palpation sites suggest that the pressure algometer device may have been perceived as more threatening than the examiner's fingertip. The harder and less resilient flat rubber tip may have provided increased loading level per square centimeter (especially at its edges if it was not perfectly parallel to the surface of the skin) when compared to the examiner's fingertip pad, or the actual pressure applied with the fingertip may have been systematically lower in spite of calibration. This last assumption is very likely based on subjects' self-written reports collected at the end of the study. When asked about the method eliciting more pressure and more pain, 88% and 80% of the subjects, respectively, said that the PAP method induced more pressure and pain than the FPP method. On the other hand, the fingertip score was clearly rated as more painful than the pressure algometer at the lateral capsule site. One explanation is that it may have been difficult for the examiners to lower their finger palpation pressure as required by the protocol when they had to palpate the TMJ sites. Regardless of the reasons, these data need to be replicated to determine if variations are systematic or random in their occurrence.

The interexaminer agreement data for the mandibular movement measurements (active opening, right lateral, and left lateral) in this study support the results of prior experiments, evidence that it is a highly reliable method. Our results indicate excellent reproducibility for MAO as well as for MPO, both showing more consistency than the PFO. The greater variability of the PFO was likely related to the status of the subjects, whether jaw pain was present or absent, and not so much associated with the measurement technique itself. In fact, the higher standard deviation to the mean for the PFO measurement among TMD subjects strongly supports this assumption (SD: 10.2 mm for TMD versus 6.2 mm for control). Unpublished data from a pilot study leading to this project indicate that the order of measurements and the verbal instructions may also influence the PFO measurement, especially in symptomatic TMD subjects.

Two previous studies by Dworkin et al[10,11] report ICC values for PFO, MAO, and MPO that are slightly better (ie, 0.90, 0.96, and 0.98, respectively) than those observed in the present study (ie, 0.78, 0.87, and 0.93). Although the ICC values differ, the results are certainly not at odds because five of six measurements showed excellent interexaminer agreement. The discrepancy for the PFO measurements is likely related to differences between the study samples and, therefore, inherent to the subjects' performance. A higher ratio of controls to TMD subjects and the fact that TMD subjects were currently receiving treatment in one study[10] may account for most of the differences between these studies. Examination of asymptomatic controls and TMD subjects with improved symptoms (ie, decreased level of pain during function) is likely to result in better consistency for PFO measurement. In the present study, we had balanced groups of controls and TMD subjects, none of the TMD subjects were currently receiving treatment, and no retraining or recalibration took place after the beginning of the study for mandibular movement measurements. The ICC values reported here represent a fair estimate of the level of consistency one would expect if no retraining to control for the passage of time is done for mandibular movement measurements.

Conclusion

The present data on examiners' ability to reproduce masticatory muscle tenderness measurements using finger pressure indicate it can be a reasonably reliable method when methodologic issues are properly addressed and the procedure is well-controlled. Good reliability can be achieved at specific muscle sites with proper training and calibration. As to the marginal reproducibility of muscle tenderness assessment at certain anatomic sites, it is more likely the result of nonspecific patient response variations than of specific factors such as a slightly different palpation site or finger palpation pressure. One interesting point is that the three muscle sites (ie, superficial and deep masseter and anterior temporalis) exhibiting reasonably good reproducibility had clearly different mean tenderness values (unpublished data) for the TMD subjects and the controls at the pressure selected for this study. These three masticatory muscle sites may potentially be considered better discriminators between disease and nondisease states. Finally, the present data confirm the results of prior research reporting good to excellent reproducibility for repeated measurements of mandibular movement using a millimeter ruler.

Acknowledgment

This research was supported by NIDR Grant #DE07618.

References

1. Dworkin SF, Huggins KH, LeResche L, Von Korff M, Howard J, Truelove E, et al. Epidemiology of signs and symptoms in temporomandibular disorders: Clinical signs in cases and controls. J Am Dent Assoc 1990;120: 273–281.
2. Mohl ND, McCall WD Jr, Lund JP, Plesh O. Devices for the diagnosis and treatment of temporomandibular disorders. Part I: Introduction, scientific evidence, and jaw tracking. J Prosthet Dent 1990;63:198–201.
3. Mohl ND, Lund JP, Widmer CG, McCall WD Jr. Devices for the diagnosis and treatment of temporomandibular disorders. Part II: Electromyography and sonography. J Prosthet Dent 1990;63:332–336.
4. Dworkin SF, LeResche L. Research diagnostic criteria for temporomandibular disorders. J Craniomandib Disord Facial Oral Pain 1992;6:301–355.
5. Smith JP. Observer variation in the clinical diagnosis of mandibular pain dysfunction syndrome. Comm Dent Oral Epidemiol 1977;5:91–93.
6. Carlsson GE, Egermark-Eriksson I, Magnusson T. Intra and interobserver variation in functional examination of the masticatory system. Swed Dent J 1980;4:187–194.
7. Kopp S, Wenneberg B. Intra and interobserver variability in the assessment of the signs of disorder in the stomatognathic system. Swed Dent J 1983;7:239–246.
8. Fricton JR, Schiffman EL. Reliability of a craniomandibular index. J Dent Res 1986;65:1359–1364.
9. Duinkerke AS, Luteijn F, Bouman TK, De Jong HP. Reproducibility of a palpation test for the stomatognathic system. Comm Dent Oral Epidemiol 1986;14: 80–85.
10. Dworkin SF, LeResche L, De Rouen T. Reliability of clinical measurement in temporomandibular disorders. Clin J Pain 1988;4:89–99.
11. Dworkin SF, LeResche L, De Rouen T, Von Korff M. Assessing clinical signs of temporomandibular disorders: Reliability of clinical examiners. J Prosthet Dent 1990;63: 574–579.
12. de Wijer A, Lobbezoo-Scholte AM, Steenks MH, Bosman F. Reliability of clinical findings in temporomandibular disorders. J Orofacial Pain 1995;9:181–191.
13. Stockstill JW, Gross AJ, McCall WD. Interrater reliability in masticatory muscle palpation. J Craniomandib Disord Facial Oral Pain 1989;3:143–146.
14. Campbell SM, Clark S, Tindall EA, Forehand ME, Bennett RM. Clinical characteristics of fibrositis. Arthritis Rheum 1983;26:817–824.
15. Jensen K, Anderson HO, Oleson J, Lindblom U. Pressure pain threshold in human temporal region: Evaluation of a new pressure algometer. Pain 1986;25:313–323.
16. Reeves JL, Jaeger B, Graff-Radford SB. Reliability of the pressure algometer as a measure of myofascial trigger point sensitivity. Pain 1986;24:313–321.
17. Fischer AA. Pressure Threshold Meter: Its use for quantification of tender spots. Arch Phys Med Rehabil 1986;67: 836–838.
18. Fischer AA. Pressure tolerance over muscles and bones in normal subjects. Arch Phys Med Rehabil 1986;67:406–409.
19. Fischer AA. Pressure algometry over normal muscles: Standard values, validity and reproducibility of pressure threshold. Pain 1987;30:115–126.
20. Fischer AA. Pressure threshold measurements for diagnosis of myofascial pain and evaluation of treatment results. Clin J Pain 1987;2:207–214.
21. Shiffman E, Fricton J, Haley D, Tylka D. A pressure algometer for myofascial pain syndrome: Reliability and validity testing. Dubner, Gebhart, Bond (eds). Proceedings of the Vth World Congress on Pain. IASP Press, 1988:407–413.
22. List T, Helkimo M, Falk G. Reliability and validity of a threshold meter in recording tenderness in the masseter muscle and the anterior temporalis muscle. J Craniomandib Pract 1989;7:223–229.
23. Ohrbach R, Gale EN. Pressure pain threshold in normal muscles: Reliability measurement effects, and topographic differences. Pain 1989;37:257–263.
24. Ohrbach R, Gale EN. Pressure pain thresholds, clinical assessment and differential diagnosis: Reliability and validity in patients with myogenic pain. Pain 1989;39:157–169.
25. Jensen R, Rasmussen BK, Pedersen B, Lous I, Oleson J. Cephalic muscle tenderness and pressure pain threshold in a general population. Pain 1992;48:197–203.
26. Lavigne GJ, Thu Thon M, Rompré R, Goulet J-P. Sensory descriptors from touch to pain and experimental variables: A psychophysiological study in human on pressure pain threshold. Proceedings of the VII World Congress on pain. IASP Press, 1994:831–842.

27. Goulet JP, Clark GT. Clinical TMJ examination methods. J Calif Dent Assoc 1990;18:25–33.
28. Stohler CS. Clinical perspective on masticatory and related muscle disorders. In: Sessle B, Bryant, Dionne (eds). Temporomandibular Disorders and Related Conditions, Progress in Pain Research and Management. Vol. 4. IASP Press, 1995:3–29.
29. Widmer CG, Lund JP, Feine JS. Evaluation of diagnostic tests for TMD. J Calif Dent Assoc 1990;18:53–60.
30. Goulet JP, Clark GT, Flack VF. Reproducibility of examiner performance for muscle and joint palpation in the temporomandibular system following training and calibration. Comm Dent Oral Epidemiol 1993;21:72–77.
31. Melzack R, Torgenson WS. On the language of pain. Anesthesiology 1971;34:50–59.
32. Fleiss JL. The Design and Analysis of Clinical Experiments. New York: John Wiley and Sons, 1986.
33. Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. Psychol Bull 1979;86:420–428.
34. Dunn OJ, Clark VA. Applied Statistics: Analysis of Variance and Regression, 2nd ed. New York: John Wiley and Sons, 1987.
35. Chilton NW. Reliability studies. In: Design and Analysis in Dental and Oral Research. 2nd ed. New York: Praeger, 1982:421.
36. Cohen J. Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull 1968;70:213–220.
37. Maclure M, Willett WC. Misinterpretation and misuse of the Kappa statistic. Am J Epidemiol 1987;126:161–169.

Resumen

La reproducibilidad de los métodos para la detección de la sensibilidad de las articulaciones y los músculos y las medidas del movimiento mandibular máximo en el sistema temporomandibular

El propósito de este estudio fue el de: *(1)* Evaluar la reproducibilidad de dos métodos para la detección de la sensibilidad de las articulaciones y los músculos masticatorios; *(2)* evaluar la reproducibilidad de las medidas del movimiento mandibular máximo; y *(3)* investigar los factores que influencian el acuerdo del examinador. Los procedimientos para la evaluación de la sensibilidad incluyeron la aplicación de una presión estándar por dos segundos sobre cuatro sitios de los músculos masticatorios, definidos anatómicamente; un sitio de control en la frente, y dos sitios en las articulaciones temporomandibulares a cada lado de la cara. Una técnica utilizó un algómetro de presión, mientras que la otra técnica exigió que un examinador entrenado aplicara presión con la punta del dedo índice. Se evaluaron a 72 personas (36 pacientes experimentales y 36 pacientes de control) durante un estudio en ciego simple. Los pacientes de control fueron acoplados de acuerdo a su edad, género y raza con los pacientes que sufrían de desórdenes temporomandibulares (DTM). Cada persona fue examinada dos veces con cada uno de los métodos descritos en una secuencia completamente balanceada y al azar, por examinadores calibrados. Los niveles de sensibilidad fueron determinados por el paciente por medio de un auto-reporte del dolor al aplicar presión utilizando un conjunto estándar de descriptores verbales. Se midieron las aperturas máximas sin dolor, activa y pasiva; además de los movimientos laterales izquierdos y derechos activos máximos, utilizando una regla milimetrada. Los coeficientes de correlación entre las clases (CCE) en cuanto a los métodos para evaluar la sensibilidad variaron entre 0,220 a 0,739 para el método que utilizaba la punta del dedo índice y entre 0,391 a 0,880 para el método que utilizaba el algómetro de presión. Los CCE para la medida del movimiento mandibular eran mucho menos variables, y variaban entre 0,59 y 0,68 en el caso del movimiento lateral y de 0,78 a 0,93 para el movimiento de apertura. Estos resultados indican que existe un acuerdo de bueno a excelente entre los examinadores calibrados para la medida del movimiento mandibular y para los métodos que evalúan la sensibilidad en dos sitios del músculo masetero (o sea superficial y profundo) y el del temporal anterior. Se encontró solamente un acuerdo regular en cuanto a los sitios del temporal medio y la cápsula lateral de la articulación temporomandibular, utilizando estos métodos.

Zusammenfassung

Die Reproduzierbarkeit von Methoden zur Bestimmung der Gelenk- und Muskelempfindlichkeit und zur Messung der maximalen Unterkieferbewegungen für das temporomandibuläre System

Der Zweck dieser Studie bestand *(1)* in der Berechnung der Reproduzierbarkeit von zwei Methoden zur Bestimmung der Empfindlichkeit der Kaumuskulatur und des Gelenkes; *(2)* in der Berechnung der Reproduzierbarkeit von Messungen der maximalen Unterkieferbewegungen; *(3)* sowie in der Untersuchung von Faktoren, welche die Uebereinstimmung der Untersucher beeinflussen. Die Verfahren zur Empfindlichkeitsbeurteilung enthielten eine Anwendung eines standardisierten Druckes für zwei Sekunden über vier anatomisch definierten Stellen der Kaumuskulatur, eine Kontrollstelle an der Stirn und zwei Stellen des Kiefergelenkes auf jeder Gesichtsseite. Die eine Technik verwendete ein Druckalgometer (PAP), während die andere Technik verlangte, dass ein geschulter Untersucher mit der Zeigefingerspitze (FPP) Druck anwendet. Zweiundsiebzig Personen (36 Patienten und 36 Kontrollen) wurden mit einem einfach-blinden Studiendesign ausgewertet. Die Kontrollpersonen wurden bezüglich Alter, Geschlecht und Rasse den Personen mit temporomandibulären Erkrankungen (TMD) angepasst. Jede Person wurde zweimal mit jeder der beschriebenen Methoden in einer zufälligen, völlig ausgeglichenen Reihenfolge von kalibrierten Prüfern untersucht. Die Empfindlichkeitsstufen wurden durch die Probanden mittels Selbstreport auf Schmerz bei Druck bestimmt, wobei eine Standardauswahl von verbalen Ausdrücken verwendet wurde. Die maximale schmerzlose, aktive und passive Oeffnung, sowie die maximale aktiven Lateralbewegungen nach rechts und links wurden mit einem Millimetermassstab gemessen. Die Korrelationskoeffizienten innerhalb der Klassen (ICC) für die Methoden der Empfindlichkeitsbeurteilung reichten von 0.220 bis 0.739 für die FPP-Methode und von 0.391 bis 0.880 für die PAP-Methode. Die ICCs für die Messungen der Unterkieferbeweglichkeit waren viel weniger variabel, sie waren im Bereich von 0.59 bis 0.68 für die Lateralbewegungen und von 0.78 bis 0.93 für die Oeffnungsbewegung. Diese Ergebnisse deuten auf eine gute bis exzellente Uebereinstimmung zwischen den kalibrierten Untersuchern für die Messungen der Unterkieferbewegungen und für die Methoden zur Empfindlichkeitsbeurteilung an zwei Stellen des Masseters (d.h. oberflächlich und tief) und des vorderen Temporalis hin. Nur eine mässig gute Uebereinstimmung wurde für die mittlere Temporalis- und die laterale TMJ-Kapselstelle mittels dieser Methoden gefunden.

Copyright of Journal of Orofacial Pain is the property of Quintessence Publishing Company Inc. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.