PAIN

Pain 69 (1997) 65–73

Interrater reliability in myofascial trigger point examination

Robert D. Gerwin[a,b,*], Steven Shannon[c,d], Chang-Zern Hong[e,f], David Hubbard[g,h], Richard Gevirtz[i]

[a]*Department of Neurology, Johns Hopkins University, Baltimore, MD, USA*
[b]*Pain and Rehabilitation Medicine, Bethesda, MD, USA*
[c]*Physical Medicine and Rehabilitation Service, Fitzsimmons Army Medical Center, Aurora, CO, USA*
[d]*Department of Physical Medicine and Rehabilitation, Walter Reed Army Hospital, Washington, DC, USA*
[e]*Department of Physical Medicine and Rehabilitation, University of California Irvine, Irvine, CA, USA*
[f]*National Cheng Kung University Hospital, Taipei, Taiwan, ROC*
[g]*Department of Neurology, University of California San Diego, San Diego, CA, USA*
[h]*Pain Rehabilitation Service, Sharp HealthCare, San Diego, CA, USA*
[i]*California School of Professional Psychology, San Diego, CA, USA*

Received 17 May 1996; revised version received 30 July 1996; accepted 13 August 1996

Abstract

The myofascial trigger point (MTrP) is the hallmark physical finding of the myofascial pain syndrome (MPS). The MTrP itself is characterized by distinctive physical features that include a tender point in a taut band of muscle, a local twitch response (LTR) to mechanical stimulation, a pain referral pattern characteristic of trigger points of specific areas in each muscle, and the reproduction of the patient's usual pain. No prior study has demonstrated that these physical features are reproducible among different examiners, thereby establishing the reliability of the physical examination in the diagnosis of the MPS. This paper reports an initial attempt to establish the interrater reliability of the trigger point examination that failed, and a second study by the same examiners that included a training period and that successfully established interrater reliability in the diagnosis of the MTrP. The study also showed that the interrater reliability of different features varies, the LTR being the most difficult, and that the interrater reliability of the identification of MTrP features among different muscles also varies.

Keywords: Myofascial pain syndrome; Myofascial trigger point; Muscle pain; Interrater reliability; Physical examination

1. Introduction

Myofascial pain syndrome (MPS) is a common cause of pain in clinical practice (Sola et al., 1955; Fishbain et al., 1986; Skootsky et al., 1989; Fricton, 1990). It is characterized by the myofascial trigger point (MTrP; Travell and Simons, 1983) (Table 1), an exquisitely tender point (TeP) in a taut band (TB) of muscle. When the MTrP is mechanically stimulated, the TB contracts producing a local twitch (LTR) response. Mechanical stimulation of the MTrP by palpation or by needling the trigger point (TrP) also produces the phenomenon of referred pain (Ref P) that is felt at a distance from the point of stimulation. The zone where Ref P is perceived can be local in the muscle or adjacent sites, or may be distant. Manual stimulation of the MTrP reproduces or aggravates the spontaneously occurring pain of the active TrP. Range of motion is restricted due to the TB and pain of the TrP, weakness without muscle atrophy

Table 1

Clinical features of the myofascial trigger point

1. Point tenderness on a taut muscle band
2. Local twitch response
3. Referred pain
4. Reproduction of usual pain
5. Restricted range of motion
6. Weakness without atrophy
7. Autonomic symptoms

* Corresponding author. Pain and Rehabilitation Medicine, 7830 Old Georgetown Road, Suite C-15, Bethesda, MD 20814, USA. Tel.: +1 301 6560220; fax: +1 301 6540333.

PII S0304-3959(96)03248-4

occurs, and autonomic phenomena occur with TrP stimulation of certain muscles such as the sternocleidomastoid muscle. These physical features of the MTrP are the clinical signs by which the MTrP is identified. Identification of the MTrP is essential for diagnosis and subsequent treatment of the MPS. The easiest and most common way to identify the MTrP clinically is by manual palpation of the muscle.

Despite the clear definition of the MTrP as stated by Travell and Simons (1983), its identification by palpation requires skill in the physical examination of muscle. Moreover, the TrP has never been established as a clinical sign in a controlled study (Tunks, 1993). Consequently, questions have been raised about the MTrP and its interrater reliability (Bohr, 1995).

Reliability refers to the accuracy, consistency, stability and reproducibility of the examination technique (Hobart et al., 1996). Interrater reliability measures the agreement between two or more examiners. While two examiners constitute a valid sampling, studies involving more examiners than two increase the ability to generalize from the results. High interrater reliability also infers that there will be a low error rate associated with the MTrP examination. Three previous studies (Nice et al., 1992; Wolfe et al., 1992; Njoo and Van der Does, 1994) have examined this problem, and none of them could establish the reliability of MTrP examination in all of its major manifestations.

Wolfe et al. (1992) in a preliminary study reported good concordance among rheumatologists when identifying TePs, but poor correlation for identification of the MTrP among examiners skilled in evaluating MPS. This study was useful as a pilot study, but suffered from several problems that we tried to avoid in the present study. The examiners met for about two hours prior to the actual subject examination phase of the study. During this time the rheumatologists were instructed in MTrP examination techniques, and the non-rheumatologists were instructed in TeP examination. Instruction in palpation of MTrPs included too many muscles of both the upper and lower extremities to permit agreement on identification characteristics of the MTrP. The study design allowed fifteen minutes to evaluate muscles in bilateral upper and lower body. This was clearly too short a time period for the myofascial exam, and the number of muscles to be examined per subject was reduced by half midway in the study.

Nice et al. (1992) likewise reported poor agreement among examiners attempting to identify TrPs by eliciting tenderness (TE) that reproduced pain and that referred pain to the area of pain complaint, without attempting to identify TBs and LTRs. The identification of the TrP in this study was a global assessment based on three features: TE, reproduction of pain (Rep P), and pain in the referral zone (i.e., pain in the area of the subject's complaint). The latter two criteria were combined into one, namely, pain in the zone of reference for that particular TrP. Eleven percent of the responses were excluded because elicited pain was referred outside of the area of the subject's low back pain. Nevertheless, these points may still have fulfilled the definition of a TrP. It is not clear from their examination method if the areas palpated were confined to the anatomic sites marked A, B and C in their Fig. 1, corresponding to the X marks on the diagrams in the Travell and Simons (1983) text. If so, that restriction would decrease the likelihood of finding a relevant TrP. Despite these limitations, the percent agreement for the three sites ranged from 76 to 79%. They assessed the significance of agreement by use of the kappa coefficient, which we discuss in detail later. The low kappa values that they obtained suggested to them a high likelihood of chance agreement, although inadequate split of target choices could also lead to low kappa values. This latter possibility cannot be evaluated on the basis of the data presented in their paper.

Njoo and Van der Does (1994) found TE, recognition or reproduction of the usual pain, and a palpable (taut) band to be reliable signs among different examiners when studying the quadratus lumborum and gluteus medius muscles. Insufficient numbers of positive responses limited the evaluation of LTRs and Ref P. The examiners in this study were either an experienced general practitioner or medical students at the end of their formal training, who spent three months working with the practitioner. The implication is that the medical students were trained by the practitioner in myofascial examination techniques. No comment is made regarding when the examinations included in the study were made relative to the beginning or end of the three-month period, which would determine the extent of their training and experience in myofascial trigger point examination. The subjects selected were patients with low back pain, and the muscles selected were considered appropriate for this complaint. Hence, the study combined the objectives of evaluating the presence of TrPs in two specific muscles (quadratus lumborum and gluteus medius) pertinent to low back pain and evaluating the reliability of identifying clinical features of MTrPs. Only the latter objective is made explicit in the study. The muscles selected can be difficult to examine for all of the features of the TrP. The quadratus lumborum is a thin, deep muscle that presents special problems in examination, especially for the identification of TBs and LTRs (Travell and Simons, 1992, pp. 64–68). The gluteus medius muscle was examined only for the presence of its anterior TrP, which has the advantage that the muscle is not covered by the gluteus maximus nor overlies the gluteus minimus muscle. TE and TBs, along with Rep P and Ref P, may be identified in this area, but LTRs can be difficult to elicit. The examination was performed in the prone position, not the lateral decubitus position described by Travell and Simons (1992, pp. 157–158). A static examination evaluating the patient in only one position increases the difficulty of TrP identification. The authors picked a kappa value of 0.5 or higher as

an indication of significant reliability. TE (including the jump sign) and recognition (or reproduction) of pain were reliable signs in both muscles. Identification of a TB was reliably made in the gluteus medius, and approached reliability in the quadratus lumborum muscle (kappa = 0.47). The data presented in this paper describes the number of times a sign was identified by one or other of two examiners, but it is not stated whether the figures cited refer to the agreement in pairs, or to the total number of times the sign was seen in the overall study population. A particular sign may have been seen in different subjects by the two examiners, but still may have added up to a similar incidence among the study population (e.g., the 29 and 25 subjects identified by two examiners respectively as having localized TE in the gluteus medius muscle out of 61 subjects examined may not have been the same 25 subjects in whom each examiner found the sign).

Attempts have been made to make or confirm the identification of the MTrP by more objective means than palpation. Fricton and Schiffman (1986) developed a craniomandibular index that included a palpation index to provide a standardized measure of problems in mandibular movement. The palpation index included TE and range of motion and was highly reliable between examiners and for repeat examinations. Fischer (1988) utilized a handheld pressure threshold meter to measure the minimum force which induced pain. He combined this with measurement of soft tissue compliance (Fischer, 1987) to document TBs and other changes in muscle tissue consistency. Finally, he utilized thermographic measurement of heat emission from the skin projection of the underlying trigger point to document its presence (Fischer and Chang, 1986). These techniques have been used to quantify changes in the TrP, as a response to treatment, for example (Jaeger and Reeves, 1986), and have been very effective in clinical studies and in documenting physical findings (Ohrbach and Gale, 1989). Nevertheless, they only identify two features of the MTrP, TE and the TB. Moreover, they are difficult to use in a general pain clinic for routine examination of patients in a limited period of time, although if performed by trained personnel like physical therapists their use can be routine.

Jensen (1990) evaluated manual palpation and pressure algometry for the quantification of muscle TE, but not for other aspects relevant to the MTrP. Tunks et al. (1995) examined TeP in subjects with fibromyalgia (FM), MPS pain, and normal controls, using pressure dolorimetry and palpation. Ratings of TE by these two methods showed good interrater reliability and consistency. The point TE did not correlate well with the location of the pain, a finding consistent with Ref P phenomena. TE alone did not permit segregation of the MPS patients from FM subjects. A new instrument called a 'palpometer' has been proposed for use in the study of tender myofascial tissues (Bendtsen et al., 1994). This instrument quantitates the amount of finger pressure an examiner uses to elicit TE in the target muscle. This approach does not address the issue of identifying the other characteristic features of the MTrP that distinguish it from other causes of muscle pain, and which are best identified by manual palpation.

A new finding of spontaneous electrical activity appears to have high specificity for MTrPs (Hubbard and Berkoff, 1993). The activity consists of two components, a constant low level of approximately 50 μV amplitude, and a superimposed irregular higher amplitude activity of 500–1000 μV. The origin of this activity is currently being investigated, (Simons et al., 1995a,bc; Simons, 1996) but it nevertheless serves as an electrodiagnostic signature for the trigger point, and may have utility for identification of the trigger point in research studies as well as in certain clinical studies. In order to perform credible epidemiological and multicenter studies of MPS, the identification of the patient population must be certain (Yunus, 1993). Unless and until the electromyographic identification of the trigger point becomes the defining characteristic of the MTrP (and it is inconvenient for many clinical studies where more than the identification of a single MTrP is necessary), the best method of diagnosis remains the physical examination.

We therefore revisited the problem of interrater reliability, to see if the major criteria defining the TrP could be identified by different examiners. The criteria evaluated were (a) TE, (b) presence of a TB, (c) presence of Ref P, (d) LTR, and (e) reproduction of the subject's symptomatic pain, and finally, (f) a global assessment was made regarding presence of a TrP.

2. Methods

2.1. Phase I of the study

Four physicians (two physiatrists and two neurologists) experienced in the diagnosis and treatment of MPS sequentially examined twenty-five subjects drawn from the personnel at Fitzsimmons AMC, Denver, CO. Subjects ranged from 27 to 75 years old, (mean age 50 years old). Thirteen were female, 12 were male. Inclusion criteria was age over 18 and willingness to be examined in the upper half of the body, excluding the breasts. This study did not attempt to relate the presence of MTrPs to any particular clinical diagnosis or condition, but addressed only the problem of identification of a physical feature. Therefore, no entry criteria were used to select subjects with FM, MPS, or no pain, as the presence or absence of these conditions would make no difference in the determination of the presence or absence of the physical findings. Subjects were randomized according to a latin square, and examined in three cohorts of 8–10 subjects each. Subjects were coded by number for purposes of data analysis. Prior to examination, each subject completed a questionnaire about their pain, medications, and pertinent job activity. Ten paired muscles were examined for TE, TBs, Ref P, LTR, Rep P,

and a global assessment of whether or not a TrP was present. Muscles examined were the sternocleidomastoid, trapezius (upper and lower segments), anterior scalene, levator scapulae, infraspinatus, latissimus dorsi, teres minor, triceps and extensor digitorum. Each subject was examined in approximately 15 min. Subjects were instructed to only answer questions about their response to examination (e.g., presence of Ref P.) and not to discuss their diagnosis or treatment. Findings were graded on a ± (present/absent) basis.

The physicians held a discussion session the night before the study to clarify the muscles to be studied, and to review the nature of the physical findings to be determined.

2.2. Phase II of the study

The second phase was structured similar to the phase I study, but with some significant changes. The same four physicians participated in the study. Subjects meeting the same criteria were selected from the private practice of one of the participants (CZH) and included both patients with pain and pain-free friends or spouses of patients. Seven subjects were female, three were male. Their ages ranged from 30 to 57 years old, with a median and mean age both of 42 years. Six subjects had cervical spine injury, spondylosis, or radiculopathy, one had a piriformis syndrome, but no cervical or upper back complaints, and three subjects had no pain complaints. Subjects were again randomized for examination sequence using a latin square distribution. Each subject was examined for fifteen minutes. Five muscle pairs were examined in order to assess reliability in different muscles, for a total of ten muscles per subject. Hence, the study was based on a comparison of one hundred muscle examinations per physician. The entire muscle was examined, except the trapezius and latissimus dorsi muscles where the upper trapezius and the axillary portion of the latissimus were examined. The five pairs of muscles examined were the sternocleidomastoid, upper trapezius, infraspinatus, latissimus dorsi, and extensor digitorum. Subjects were instructed only to answer questions about their response to the examination, and not to discuss their diagnosis or treatment.

The physicians met for three hours immediately prior to the study for a training session. Definitions of the features of the TrP were reviewed to be certain that all examiners were interpreting physical findings similarly. Each clinical sign was reviewed on a live subject (a physician or a volunteer who was not a subject in the subsequent study) to eliminate any variation in the interpretation of physical findings. Each physician then examined each of the other three examiners following the study protocol and the results were reviewed. Discrepancies among the findings were evaluated and reassessed until the examiners could all elicit the same signs and were in agreement about the physical findings. The TB was defined as a linear, or in some cases a nodular, firmness distinguishable from the surrounding muscle, and parallel to the orientation of the muscle fibers when linear. Palpation was performed by pincer or flat palpation or both, depending on the muscle being examined. The sternocleidomastoid and latissimus dorsi muscles are best examined by pincer palpation, rolling the muscle between the fingers. The infraspinatus and extensor digitorum muscles are only accessible to flat palpation. The trapezius muscle can be palpated well by both techniques. Thus, this study was able to assess the reliability of TrP palpation using the commonly available mix of manual palpation techniques (Travell and Simons, 1983, pp. 59–63). The pressure of palpation used to identify a TB or to elicit an LTR or Ref P was varied to meet the needs of examining a specific subject, since the pressure required to elicit TE, to feel a TB, or elicit an LTR response was not the same for each subject or for each muscle even in the same subject. We agreed to use such force as we judged would not be uncomfortable in normal muscle. Pressure was used that was sufficient to elicit the features sought or to satisfy us that they were not present. The individual examiner had to make the final judgment as to how much force was required in each specific case. Some of the examiners had identified the LTR largely by feel rather than by sight. We agreed that either means was acceptable, as long as the observation was definite. We agreed on a uniform way of asking about Ref P and reproduction of the subject's pain problem (if any) that did not use leading questions. Finally, we decided to specify a TrP as absent, latent or active. An active TrP had to reproduce the subject's pain. A latent TrP had to include a TeP and a TB. TE is not specific to the MTrP, but a TB distinguishes a TrP from other causes of TE. The presence of an LTR or Ref P make the identification even more certain.

Phase I of the study was approved by the Institutional Research Review Board at Fitzsimmons Army Medical Center, Department of the Army. Phase II of the study was approved by the Human Subjects Review Committee, University of California Irvine.

2.3. Statistical methods

To assess the agreement between judges an overall measure was used which represents a general observer-agreement statistic called the S_{av} (O'Donnell and Dobson, 1984). This statistic is a generalized version of the Cohen's kappa which reports pairwise judge agreement corrected for chance agreement (Cohen, 1960). Reported is the S_{av} statistic itself (ranging from 0 to 1.0), its significance level (the likelihood that it is not really 0) and the percent agreement among all the judges. Note that when almost perfect agreement occurs, but without an adequate split of targets, kappa is not applicable. This is an important feature of the kappa statistic, which is a measure of the likelihood that the results are not simply a chance occurrence. The kappa statistic is dependent on the presence of two or more

choices and measures the strength of the relation between the variables rather than the strength of agreement itself. When all subjects have a given finding (e.g., all extensor digitorum muscles have TBs), there is an insufficient range of actual choices, and the kappa value will be low even when agreement among examiners is high (Hobart et al., 1996). In such cases the percentage agreement is reported and noted as such in the results section.

3. Results

3.1. Phase I study

Results for the phase one study were analyzed for each of the twenty muscles examined in each subject for each of the five criteria studied, and for the global assessment that a TrP was or was not present. No distinction was made in the phase I study between active and latent TrPs.

For the TE rating, 14 of 20 muscles had an S_{av} greater than zero, but most were low (average S_{av} was 0.22) and showed only slight agreement. The greatest agreement was for the left extensor digitorum which had an S_{av} of 0.55, which is considered to be of moderate agreement. The left upper trapezius had an S_{av} of 0.52. No other muscle approached significance for TE. S_{av} did not approach 0.5 for any other feature of the TrP in any muscle studied. In several instances agreement was 70% or greater, rarely over 80%, but the S_{av} was low because there was not an adequate distribution of positive and negative findings.

In summary, the phase I study failed to establish high degree of agreement among the examiners for any of the features of the MTrP or for the presence or absence of the TrP itself, except for TE in the trapezius and extensor digitorum muscles.

3.2. Phase II study

Individual muscle evaluation of the five pairs of muscles examined in each of the subjects showed a considerable variation in the number of positive and negative responses. A finding that is positive in a particular muscle in every subject cannot be evaluated using a kappa coefficient. Therefore, the actual number of positive and negative findings is shown in Table 2. The split in positive (present) and negative (absent) observations ranges from 53 negative/27 positive for Ref P from the trapezius muscles to 80 positive/0 negative, in the extensor digitorum, to an even split of 40/40 for Ref P in the infraspinatus muscles. There are a number of specific features in certain muscles where the feature was found in nearly all subjects. Only the percentage of agreement is given as an indication of agreement among examiners. These features are TE in the sternocleidomastoid muscle, TBs in the sternocleidomastoid, trapezius, and in the extensor digitorum muscles, and LTR in the extensor digitorum muscle.

The percent agreement among the examiners is reported for each muscle and for each clinical sign. The significance of the agreement is expressed by the kappa coefficient S_{av} for each muscle and for each clinical sign (Table 3). A kappa of 1.0 indicates perfect agreement with an adequate split of targets. Kappa values between 0.4 and 0.6 are considered in moderate agreement. Values above 0.6 are considered in substantial agreement. Values of 0.80 or above are considered in excellent agreement (O'Donnell and Dobson, 1984). A total of thirty categories were rated, six features each in five pairs of muscles. Agreement was perfect or almost perfect in five of the categories where the kappa coefficient was not applicable. Table 3 shows that agreement was substantial or almost perfect in about 2/3 of the remaining categories, and agreement was moderate in three categories, and less than moderate in only six categories.

3.3. Assessment by muscle

Agreement for the presence or absence of features of the

Table 2
Number of myofascial trigger point physical features present or absent in each muscle in phase II

Muscle		TE	TB	LTR	Ref P	Rep P	TrP		
							0	L	A
Stcm	+	78	79	72	37	52	1	26	53
	–	2	1	7	43	28			
Trap	+	69	76	45	27	45	13	24	43
	–	11	4	33	53	35			
Infsp	+	68	70	23	40	43	12	29	39
	–	12	10	56	40	37			
Lats	+	72	72	55	30	46	11	31	38
	–	8	8	22	50	34			
Exdg	+	72	80	78	52	48	3	33	44
	–	8	0	2	28	32			

TE, tenderness; TB, taut band; LTR, local twitch response; Ref P, referred pain; Rep P, reproduced pain; TrP, trigger point; Stcm, sternocleidomastoid muscle; Trap, trapezius muscle; Infsp, infraspinatus muscle; Lats, latissimus dorsi muscle; Exdg, extensor digitorum muscle; 0, no trigger point; L, latent trigger point; A, active trigger point.

Table 3
Kappa coefficients (S_{av}) for each muscle and each attribute of the MTrP in phase II

Muscle	TE	TB	LTR	Ref P	Rep P	TrP
Stcm	[a]	[a]	0.11	0.57	0.89	0.84
Trap	0.61	[a]	0.36	0.65	0.84	0.66
Infra	0.48[b]	0.40	0.17	0.84	0.79	0.65
Lats	1.0	0.46[b]	0.57	0.71	0.90	0.79
Extdig	0.51	[a]	[a]	0.67	1.0	0.95

Stcm, sternocleidomastoid, Trap, trapezius; Infra, infraspinatus; Lats, latissimus dorsi; Extdig, extensor digitorum.
[a]Indicates that there was almost complete agreement among the examiners, but without an adequate split of targets, so that the kappa coefficient is not applicable.
[b]Indicates that the kappa is derived from values of only one side (right or left) because the opposite side had an inadequate distribution of options and kappa did not apply. Other abbreviations as in Table 2.

TrP in the sternocleidomastoid muscle was high, except for the LTR where agreement was not significant (Fig. 1). Agreement was perfect or almost perfect for the features TE, TB, Rep P and for the global determination of the presence or absence of a TrP. The P value for the kappa coefficient (S_{av}) was <0.001, except for the LTR where the low rate of agreement was not significant.

Agreement among examiners for the features of the TrP in the trapezius muscle was substantial for all features except for TB where agreement was almost perfect, and for LTR, where agreement was slight. The S_{av} P value was <0.001 for all features of the TrP except the LTR where its significance was $P < 0.01$. The strength of the agreement was substantial where it was not almost perfect, except for the LTR where it was slight.

Agreement for TrP characteristics in the infraspinatus muscle was substantial or almost perfect for all features except the LTR response where agreement was slight. The significance of the S_{av} was $P < 0.001$ except for the LTR where $P < 0.01$.

Agreement for the latissimus dorsi muscle was high for all TrP features, $P < 0.001$ for each appropriate S_{av}. The strength of agreement was substantial, including that for the LTR. Agreement was almost perfect for TE and for Rep P.

Agreement among the examiners for TrP features in the extensor digitorum muscle was high. It was perfect or almost perfect for identification of TB, LTR, Rep P and for the global determination of presence or absence of an active or latent TrP.

3.4. Assessment by clinical sign

Evaluation of TE showed a high percentage of agreement in all muscles (Fig. 1). Agreement was over 90% in every muscle except the infraspinatus, where there was 83% agreement. Agreement was substantial, almost perfect or perfect (100% agreement) for each muscle.

TB identification was made with a high percentage of agreement among the examiners, ranging from 83% to 100%. the P (S_{av}) was <0.05 for the infraspinatus and latissimus dorsi muscles, but agreement was almost perfect for the sternocleidomastoid, trapezius, and extensor digitorum muscles. Agreement was fair to moderate for the infraspinatus and latissimus dorsi muscles.

The LTR was less reliable than any other feature of the TrP. Nonetheless, it was found with almost perfect agreement in the extensor digitorum muscle. Agreement was low for the infraspinatus muscles at 59% and at 68% for the trapezius muscles, but was higher for the other muscles (sternocleidomastoid: 82%; latissimus dorsi: 85%; extensor digitorum: 95%). The P (S_{av}) was not significant for the sternocleidomastoid. Agreement was slight for the trapezius and infraspinatus muscles, was substantial for the latissimus dorsi, and was almost perfect for the extensor digitorum.

Ref P showed good agreement for all muscles, with a kappa coefficient S_{av} of either $P < 0.001$ or almost perfect for each muscle, and a strength of agreement that was

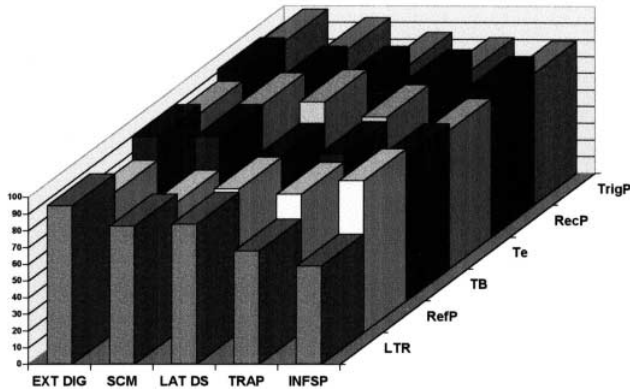

Fig. 1. Percent interrater agreement. The percentage of agreement among the examiners is shown for each muscle and for each trigger point feature. The percentage of agreement was high for most of the features studied. The local twitch response showed the greatest variability in agreement and was the most difficult feature to reproduce among different examiners. Percent agreement among the different muscles studied varied for different trigger point features. Some muscles are more easily examined than others, and agreement is higher in these muscles. See the text for full discussion. Ext Dig, extensor digitorum; INFSP, infraspinatus; LAT DS, latissimus dorsi; LTR, local twitch response; RecP, recall spontaneous pain; RefP, referred pain; SCM, sternocleidomastoid; TB, taut band; Te, tenderness; TraP, trapezius; TrigP, trigger point.

moderate for the sternocleidomastoid, but which was substantial or almost perfect for all other muscles tested.

Agreement for ReP P was high (sternocleidomastoid: 95%; trapezius: 93%; infraspinatus: 89%; latissimus dorsi: 95%; extensor digitorum: 100%). The kappa coefficient S_{av} was high for all muscles and the strength of agreement was substantial or almost perfect for each muscle.

Agreement for the presence or absence of a latent or active TrP varied among the five muscles from 98% for the extensor digitorum to 79% for the infraspinatus, with the other three muscles falling between (sternocleidomastoid: 93%; trapezius: 81%; latissimuss dorsi: 84%). The kappa coefficient S_{av} was significant for all muscles at $P < 0.001$. The strength of agreement was substantial or almost perfect for each muscle examined.

4. Discussion

The present study shows that four examiners can achieve statistically significant agreement, at times almost perfect agreement, about the presence or absence of five major features of the MTrP and on the presence or absence of the TrP, whether it be latent or active. This establishes the MTrP as a reliable clinical sign. The present study also shows that these features are identified with greater or lesser reliability depending on the specific feature and the specific muscle being examined. The LTR was less reliable than other features like the TB. A training period was found to be essential in order to achieve these results.

This study differs from previous ones in that it focuses on the identification of TrP features, rather than on the presence of certain TrP signs in clinical conditions like low back pain, FM or MPS. Thus, it makes no difference if the subject had a pain complaint or not for the purposes of this study. All of the major features of the TrP were sought, rather than just a limited array of signs. The examinations were dynamic in that each examiner was free to reposition the subject and to use flat or pincer palpation techniques in order to facilitate the examination. Finally, a training session was incorporated into the protocol when it became clear after the first attempt to perform this study that the examiners were not all defining the features of the TrP similarly and were not all identifying the same muscles with equal ease (e.g., the anterior scalene and teres minor muscles in the first study phase). The phase II study included fewer muscles in order to allow more time for examination of each muscle.

The initial attempt to assess the agreement among four examiners evaluating subjects for five features of the MTrP showed poor interrater agreement for almost every muscle and every feature. Our clinical experience suggested to us that the poor outcome might be the study design, rather than a failure of our examination technique or a problem inherent in the nature of the TrP. Discussion among the four examiners indicated that criteria that we took for granted for the identification of certain features of the MTrP differed among the examiners. These included differences for identification of the TB and the LTR. Different examination techniques were also used; e.g., differing pressure applied to a possible MTrP, and different ways of attempting to elicit an LTR. Some examiners had not previously examined certain muscles used in the first study routinely, and identification of these muscles was uncertain in specific instances. The first study also required that twenty muscles be examined in each subject in fifteen minutes, less than one minute per muscle, including time for positioning the subject and recording the results. Difficulty in identifying the MTrP arose because no distinction was made between active and latent TrPs. The authors agreed to redesign the study to address these issues.

Phase II of the study was designed to answer the same question as phase I. Limiting the number of muscles to be examined, and using muscles that were both accessible and familiar to the examiner eliminated the problems of inadequate examination time and of uncertain muscle identification. A most important change in the study was the addition of a three-hour training period in which the four examiners clearly defined how each of the MTrP features was to be identified. The feature of TE presented a problem because there is no uniform force or pressure that can be applied to all muscles at all times. In contrast to the examination of the fibromyalgia TeP where 4-kg pressure is recommended (Wolfe et al., 1990), differences in tissue compliance, size of the TB, and depth of the muscle require different degrees of force applied to the tissue to determine the presence or absence of the TB and its associated TeP.

Travell (1990) distinguished between TrPs that spontaneously cause pain continuously, even at rest, those that cause pain with movement or activity, and those that cause referred TE, stiffness, and restricted range of motion, but not pain. These distinctions have utility in clinical assessment and management of the patient. However, we did not think that they were necessary distinctions for our study.

The studies of Wolfe et al. (1992) attempted to relate the physical findings of TeP for FM and those of the MTrP for MPS to classification of subjects into these two disease categories. Thus, their study did not focus on the reliability of the identification of the physical finding alone. That study did not take into account overlap syndromes or possible misclassification of the subjects when it reported failure to identify MPS by physical findings. In our study we looked only at the ability of examiners to agree on the presence of a physical sign in a particular muscle. It made no difference whether the subject had active TrPs or not for this purpose. Likewise, because we were comparing examiners findings, it made no difference that one physician may have previously examined a particular patient, nor what the patient's diagnosis was.

We found that certain characteristics of the MTrP were very common in some muscles. This was especially true of

TB (100%) and LTR (98%) in the extensor digitorum, and of TE (98%) and TBs (99%) in the sternocleidomastoid. This was not expected prior to the study, at least not to the degree that we found. Given the high degree of agreement for these features and the significant agreement for most of the other features of the MTrP where there was an adequate split of present and absent TrP features, we think that these results are also significant in demonstrating that different examiners can agree on the presence of these features.

The study also showed the importance of clarifying the definitions of the features to be identified, and the need to be certain that each of the examiners was interpreting the physical findings similarly. All of the examiners were familiar with the criteria for the identification of the MTrP from prior experience in the clinic, but a training program was still necessary in order to introduce a level of uniformity to the examination and the definitions used. We believe that the same uniformity can be achieved through a formal training program of novices, though no credible study of the effects of such a training program has been done as yet.

The pressure used to elicit TE, an LTR, Ref P, or to reproduce spontaneous pain, and to identify a TB varies with the muscle being examined, and with subject characteristics. More superficial muscle needs less force to palpate. Deeper muscles need a more penetrating technique.

Muscles with prominent TB and easily obtained LTR need less palpation force than less active TrPs. We did not quantitate TE (e.g., on a scale of 1 + to 4 +), but noted it only as present or absent, since a TrP index has not been used in MPS research in the way that a TeP index has been used in the study of FM. The force of palpation was necessarily less in persons with very active TrPs (very tender muscle) than for persons with no TrPs in the muscle examined. We believe, therefore, that there is no standard force of palpation that applies to all TrPs.

The reliability of identification of certain features of the TrP varies among muscles. One should not expect that a certain feature such as a TB will be as reliable in one muscle as another. Identification of the TB in the infraspinatus muscle was less certain than in the sternocleidomastoid or extensor digitorum. Likewise, the LTR was more difficult to identify reliably in all muscles except the extensor digitorum, showing that this feature differed from the other features in reliability across most muscles. The design of studies of MPS and of the MTrP must take into account the difference in reliability of MTrP features in different muscles. Examination of a muscle in which a specific feature is less reliable may lead to an inconclusive or misleading result.

The TB and TE are the most reliable of the TrP features to identify, and the minimal criteria by which to make such an identification. Rep P denotes a TrP as active or latent. Ref P and the LTR are most useful as confirmatory signs of the TrP. However, although not addressed in this study, our experience is that Ref P, LTR and Rep P is often elicited when placing a needle into the MTrP, even when these features are not found by manual palpation.

One implication of this study is that researchers studying MPS or the MTrP need to define the TrP for the purposes of their study. The criteria by which a TrP is identified, or the diagnosis of MPS made, needs to be clearly stated in order to properly interpret the reliability of the study. Many past studies merely state that the subjects met the criteria established by Travell and Simons (1983). Investigators should state whether TE alone is used, which does not distinguish MPS from FM, or whether a TB, Ref P, an LTR or a variable subset of features is being used to define the clinical condition being studied. Clearly, the identification of the MTrP will be more specific when more features are assessed and included in the criteria.

Multicenter studies of MPS using physical signs as a means of identifying the TrP must demonstrate that the persons performing the physical examination can agree on the presence or absence of its physical features. Otherwise, there can be no confidence that the centers are evaluating the same populations or looking at the same condition.

References

Bendtsen, L., Jensen, R., Jensen, N.K. and Olesen, J., Muscle palpation with controlled finger pressure: new equipment for the study of tender myofascial tissues, Pain, 59 (1994) 235–239.

Bohr, T.W., Fibromyalgia and myofascial pain syndrome: do they exist? Neurologic Clin. North Am., 13 (1995) 365–381.

Cohen, J., A coefficient of agreement for nominal scales, Educ. Psychol. Meas. 20 (1960) 37–46.

Fischer, A.A., Tissue compliance meter for objective, quantitative documentation of soft tissue consistency and pathology, Arch. Phys. Med. Rehabil., 68 (1987) 122–125.

Fischer, A.A., Documentation of myofascial trigger points, Arch. Phys. Med. Rehabil., 69 (1988) 286–291.

Fischer, A.A. and Chang, C.H., Temperature and pressure threshold measurements in trigger points, Thermology, 1 (1986) 212–215.

Fishbain, D.A., Goldberg, M., Meagher, B., Steele, R. and Rosomoff, H., Male and female chronic pain patients categorized by DSM-III psychiatric diagnostic criteria, Pain, 26 (1986) 181–197.

Fricton, J.R., Myofascial pain syndrome: characteristics and epidemiology. In: J.R. Fricton and E.A. Awad (Eds.), Myofascial Pain and Fibromyalgia, Advances in Pain Research and Therapy, Vol. 17, Raven Press, New York, 1990, pp. 107–128.

Fricton, J.R. and Schiffman, E.L., Reliability of a craniomandibular index, J. Dent. Res., 65 (1986) 1359–1364.

Hobart, J.C., Lamping, D.D., and Thompson, A.J., Evaluating neurologic outcome measures: the bare essentials (editorial), J. Neurol. Neurosurg. Psychiatry, 60 (1996) 127–130.

Hubbard, D.R. and Berkoff, G.M., Myofascial trigger points show spontaneous needle EMG activity, Spine, 18 (1993) 1803–1807.

Jaeger, B. and Reeves, J.L., Quantification of changes in myofascial trigger point sensitivity with the pressure algometer following passive stretch, Pain, 27 (1986) 203–210.

Jensen, K., Quantification of tenderness by palpation and use of pressure algometers. In: J.R. Fricton and E.A. Awad (Eds.), Myofascial Pain and Fibromyalgia, Advances in Pain Research and Therapy, Vol. 17, Raven Press, New York, 1990, pp. 165–181.

Nice, D.A., Riddle, D.L., Lamb, R.L., Mayhew, T.P. and Rucker, K., Intertester reliability of judgments of the presence of trigger points in patients with low back pain, Arch. Phys. Med. Rehabil., 73 (1992) 893–898.

Njoo, K.H. and Van der Does, E., The occurrence and inter-rater reliability of myofascial trigger points in the quadratus lumborum and gluteus medius: a prospective study in non-specific low back pain patients and controls in general practice, Pain, 58 (1994) 317–323.

O'Donnell, D. and Dobson, A., General observer-agreement measures on individual subjects and groups of subjects, Biometric, 40 (1984) 973–983.

Ohrbach, R. and Gale, E.N., Pressure pain thresholds, clinical assessment and differential diagnosis: reliability and validity in patients with myogenic pain, Pain, 39 (1989) 157–170.

Simons, D.G., Hong, C-Z. and Simons, L.S., Prevalence of spontaneous electrical activity at trigger spots and at control sites in rabbit skeletal muscle, J. Musculoskeletal Pain, 3 (1995a) 35–48.

Simons, D.G., Hong, C-Z. and Simons, L.S., Spontaneous electrical activity of trigger points (abstract), J. Musculoskeletal Pain, 3 (Suppl. 1) (1995b) 124.

Simons, D.G., Hong, C-Z. and Simons, L.S., Spike activity in trigger points (abstract), J. Musculoskeletal Pain, 3 (Suppl. 1) (1995c) 125.

Simons, D.G., Clinical and etiological update of myofascial pain from trigger points, J. Musculoskeletal Pain, (1996) in press.

Skootsky, S.A., Jaeger, B. and Oye, R.K., Prevalence of myofascial pain in general internal medicine practice, West. J. Med., 151 (1989) 157–160.

Sola, A.E., Rodenberger, M.L. and Gettys, B.B., Incidence of hypersensitive areas in posterior shoulder muscles, Am. J. Phys. Med., 34 (1955) 585–590.

Travell, J.G., Chronic myofascial pain syndromes: mysteries of the history. In: J.R. Fricton and E. Awad (Eds.), Myofacial Pain and Fibromyalgia, Advances in Pain Research and Therapy, Vol. 17, Raven Press, New York, 1990, pp. 129–137.

Travell, J.G. and Simons, D.G., Myofascial Pain and Dysfunction: The Trigger Point Manual, Vol. 1, Williams and Wilkins, Baltimore, MD, 1983, pp. 5–19.

Travell, J.G. and Simons, D.G., Myofascial Pain and Dysfunction: The Trigger Point Manual, Vol. 2, Williams and Wilkins, Baltimore, MD, 1992.

Tunks, E., Clinical experimental investigations in fibromyalgia and myofascial pain. In: H. Voeroy and H. Merskey (Eds.), Progress in Fibromyalgia and Myofascial Pain, Elsevier, Amsterdam, 1993, pp. 161–172.

Tunks, E., McCain, G.A., Hart, L.E., Teasell, R.W., Goldsmith, C.H., Rollman, G.B., McDermid, A.L. and DeShane, J., The reliability of examination for tenderness in patients with myofascial pain, chronic fibromyalgia and controls, J. Rheumatol., 22 (1995) 944–952.

Wolfe, F. et al., The American College of Rheumatology 1990 criteria for the classification of fibromyalgia: report of the multicenter criteria committee, Arthritis Rheum., 33 (1990) 160–172.

Wolfe, F., Simons, D.G., Friction, J.R., Bennett, R.M., Goldenberg, D.L., Gerwin, R.D., Hathaway, D.E., McCain, G.A., Russell, I.J., Sanders, H. and Skootsky, S.A., The fibromyalgia and myofascial pain syndromes: a preliminary study of tender points and trigger points in persons with fibromyalgia, myofascial pain syndrome and no disease, J. Rheumatol., 19 (1992) 944–951.

Yunus, M.B., Research in fibromyalgia and myofascial pain syndromes: current status, problems and future directions, J. Musculoskeletal Pain, 1 (1993), 23–41.