Interexaminer Reliability of the Palpation of Trigger Points in the Trunk and Lower Limb Muscles

Chang-Yu J. Hsieh, DC, PT, Chang-Zern Hong, MD, Alan H. Adams, DC, Katherine J. Platt, DC, Clark D. Danielson, MPA, Fred K. Hoehler, PhD, Jerome S. Tobis, MD


Objectives: To determine the interexaminer reliability of palpation of three characteristics of trigger points (taut band, local twitch response, and referred pain) in patients with subacute low back pain, to determine whether training in palpation would improve reliability, and whether there was a difference between the physiatric and chiropractic physicians.

Design: Reliability study.

Setting: Whittier Health Campus, Los Angeles College of Chiropractic.

Participants: Twenty-six nonsymptomatic individuals and 26 individuals with subacute low back pain.

Intervention: Twenty muscles per individual were first palpated by an expert and then randomly by four physician examiners.

Main Outcome Measures: Palpation findings.

Results: Kappa scores for palpation of taut bands, local twitch responses, and referred pain were .108, .050, .118, and .326, respectively, between the expert and the trained examiners, and .001, .022, and .320, respectively, among the nonexpert, trained examiners, and −.019, .022, and .200, respectively, among the nonexpert, untrained examiners.

Conclusions: Among nonexpert physicians, physiatric or chiropractic, trigger point palpation is not reliable for detecting taut band and local twitch response, and only marginally reliable for referred pain after training.

Key Words: Myofascial pain syndrome; Myofascial trigger point; Inter-rater reliability; Low back pain; Rehabilitation.

© 2000 by the American Congress of Rehabilitation Medicine and the American Academy of Physical Medicine and Rehabilitation

When there are no measurable structural changes in the vertebra or the disc and no detectable neurophysiologic abnormalities, the diagnosis of low back pain can become confusing. Various specialty fields use different terminology in the literature when describing such conditions. Recent literature in rheumatology, physiatry, orthopedics, psychology, and chiropractic has recognized three distinct musculoskeletal dysfunctions as common causes of lower back pain syndrome without organic changes: (1) myofascial pain syndrome (MPS) caused by trigger points, (2) fibromyalgia syndrome, and (3) articular dysfunction.

Myofascial pain syndrome is one of the most frequent diagnoses made in relation to low back pain. It may be caused by or associated with many conditions, including acute or chronic repetitive trauma (such as strain, sprain, contusion, poor posture, muscle overloading), lesions involving various structures (such as tendinitis, bursitis, synovitis, arthritis, intervertebral disc lesion), or emotional stress. Conditions believed to perpetuate or exacerbate the severity of MPS include mechanical stress, nutritional inadequacies, metabolic and endocrine inadequacies, chronic infections, and psychological factors.

The myofascial trigger point (MTrP) is a highly localized, painful or sensitive spot in a palpable taut band of skeletal muscle fibers. Pain from an active MTrP can occur spontaneously or in response to movement. A latent MTrP is defined as a sensitive spot from which pain or discomfort is elicited by compression only. The characteristics of an MTrP include (1) a discriminable tender spot (MTrP) in a palpable taut band (also referred to as a “nodule”); (2) a consistent and characteristic referred pain pattern on compression; (3) a local twitch response (LTR), a brief and rapid contraction of muscle fibers around the MTrP elicited in some muscles by snapping palpation on MTrPs or, in almost all cases, by needling of MTrPs; (4) a restricted range of stretch caused by shortening of muscle fibers in a taut band; (5) muscle weakness with no remarkable atrophy resulting from a waxing and waning phenomena of MTrPs; (6) the spread of pain to other parts of the body in severe cases of MPS; and (7) associated referred autonomic phenomena including vasovagal reaction, coldness, sweating, pilomotor response, piosis, and hypersecretion.

The diagnosis of MPS, manifested with one or more MTrPs, is usually based on the patient’s subjective symptoms and the presence of some of the above characteristics. However, the clinical usefulness of the MTrP examination has been questioned because of reported poor reliability (low kappa scores). Table 1, modified from Simons, lists the results of four different studies on the interexaminer reliability of MTrP examination. Experienced examiners showed unsatisfactory kappa values when no training sessions were given before the studies. Njoo and Van der Does used inexperienced examiners who were given training. They reported kappa values ranging from .09 to .70. Recently, satisfactory kappa values were demonstrated by Gerwin and coworkers when experienced examiners were trained beforehand. It appears that...
both experience and training are important in attaining reliable information from the MTTrP examination.

To date, there has been no literature reporting comparisons between chiropractic and allopathic physicians in reliability of MTTrP palpation. Because chiropractors treat many patients with low back pain, and frequently diagnose and treat MPS, it is important to determine whether there are differences in the reliability of MTTrP palpation between chiropractors and allopathic physicians. It is especially important to determine whether training sessions in MTTrP palpation improve the reliability of MTTrP palpation of nonexpert physicians. Therefore, this study was designed to assess (1) the interexaminer reliability of palpation of three characteristics of trigger points (taut band, local twitch response, and referred pain) for nonexpert physicians; (2) whether training would improve the reliability; (3) whether there was a difference in reliability between the physiatric and chiropractic physicians.

METHODS

Subjects

From January 19, 1995 to June 25, 1995, 775 respondents were initially screened by phone interviews, after being recruited by public radio and cable television news announcements, radio interviews, flyers, news releases in newspapers, and paid advertisements in local newspapers and publications. A total of 27 media presentations were made. From the 775 telephone respondents, 52 were included in the study. Twenty-six individuals (14 men, 12 women) had low back pain, and 26 individuals (15 men, 11 women) were asymptomatic. Table 2 compares the demographics and subject characteristics of the patients with back pain with those of the control subjects. The only significant differences between the two groups were the visual analogue pain score, the Roland-Morris activity scale score, and the presence of referred pain. The visual analogue pain score is based on a 10-cm horizontal line, with the left end being no pain and the right end being the worst imaginable pain (modified from the original work of Huskisson27). Patients were asked to mark on the line to indicate the severity of their pain in the previous week. The Roland Morris Activity scale contains 24 items of activities designed specifically for low back pain.28 Patients were asked to mark whether the pain affected these functional activities on the day of assessment. Only the “yes” items were counted for the score. Therefore, a higher score indicates a higher functional disability. Because these measures are indicators of back pain, such differences are expected between the two groups.

Individuals with low back pain met the following inclusion criteria: they were at least 18yrs of age and had experienced low back pain for more than 3 weeks and less than 6 months. The exclusion criteria were (1) pregnancy; (2) serious medical problems (eg, advanced cancer, heart failure); (3) definable neurologic abnormalities in the lower extremities (eg, peripheral neuropathy, multiple sclerosis, hemiplegia, myelopathy); (4) spinal disorders with bony lesions (eg, osteoporosis, fracture, unstable spondylolisthesis, multiple myeloma; x-rays were taken as clinically indicated); (5) significant mental disorders as determined by phone inquiry and clinical interview (eg, psychosis, mania, major depression); (6) obesity, using Davenport body mass index; (7) radiocinal pain by clinical examination (if further tests were clinically indicated, eg, CT, MRI, electrodiagnostic tests, blood chemistries, the patients were referred for testing at their own expense); (7) litigation, auto injuries, or work injuries; and (8) inappropriate illness behavior (positive Wadell’s sign).25

Healthy volunteers met the following inclusion criteria: they were: at least 18yrs of age and had no history of low back pain during the previous 6 months.

Informed Consent and Approval of Institutional Review Board

The research protocol was reviewed and approved by the institutional review boards of both the Los Angeles College of Chiropractic and the University of California, Irvine. Signed informed consent was obtained from all participants.

Recruitment of Examiners

Four licensed chiropractors representing the Los Angeles College of Chiropractic and four licensed physiatric residents from the University of California, Irvine, participated in the study. Recruited through recommendations of the research team or the college faculty members, they were divided into trained and untrained groups. The two chiropractic residents with 3 and 5yrs of licensure, one second-year physiatric resident (an osteopathic physician) with 3yrs of licensure, and one third-year physiatric resident with 4yrs of licensure were assigned to the group to be trained. A senior faculty chiropractic clinician with 15yrs of licensure, an experienced field chiropractic practitioner with 15yrs of licensure, and two second-year physiatric residents, both with 2yrs of licensure, were assigned to the group that received no training. All examiners were considered as nonexpert physicians for MTTrP palpation.

Examiners’ Educational and Training Programs

The training given to the trained examiners included three 2-hour classroom lectures on the overview of myofascial pain
research and instruction of localization of the trigger points under investigation. Handouts of the trigger point locations from the Travell and Simons' textbook were used for instruction. To standardize the manual diagnostic procedures for palpation of taut bands, local twitch responses, and referred pain, the trained examiners participated in three 2-hour, hands-on practice sessions using both normal and symptomatic subjects. Referred pain was defined as any pain the individual felt away from the area under palpation. The trained examiners were also trained to use a standardized amount of pressure in palpation, 3kg/cm² of pressure over the superficial muscles and 7 to 8kg/cm² of pressure over the deep muscles. This was done applying the appropriate pressure to an algometer placed against a hard surface and approximating that pressure when palpating muscles.

The untrained examiners were told to identify whether there was a taut band, a local twitch response, or referred pain in the muscles palpated. They were given a handout showing trigger point locations. No other specific training was given.

**Assessment of Reliability of MTrP Examination**

Subjects were arranged in groups of three to seven per session. They were individually placed in different examination rooms. The investigator (CZH) first examined them to confirm the existence of a taut band, muscle twitch response, and/or referred pain pattern of the low back muscles. Four examiners (one trained physiatric resident, one trained chiropractic resident, one untrained physiatric resident, and one untrained chiropractor) then examined subjects in a predetermined random order.

Before the palpation sessions, the trained examiners were first checked to ensure that the pressure they would apply over the trigger points approximated the standard pressures to be used in the study. They were also reviewed for the locations of the trigger points by CZH in the earlier sessions. All examiners were blinded to the expert examiner’s findings, each other’s findings, and whether the individual was asymptomatic or had low back pain.

All subjects were placed in a prone position with the low back and lower extremities exposed. Each examiner palpated the muscles (10 on each side) to determine whether a taut band, any local twitch response, and any referred pain pattern were present. The muscles examined included the following: iliocostalis lumborum (L1 level), rectus femoris, rectus abdominis (lowest TrP), tensor fascia latae, quadratus lumborum (distal superficial TrP), gluteus maximus (TrP2), gluteus medius (TrP2), gluteus minimus (anterior), piriformis (TrP1), and soleus (TrP1) muscles. The data were recorded as “yes” or “no” responses. Each examiner had an assistant to record the results of palpation.

**Statistical Analysis**

Three aspects of trigger point diagnosis (taut band, local twitch response, and referred pain) were assessed in this study and were separately analyzed. Analyses of agreement were performed using simple and modified kappa coefficients. To test the extent of agreement with the expert (CZH), assessments for all examiners of all muscle groups were paired with the expert’s assessments of all muscle groups, and each assessment was treated as a separate case. Simple kappa coefficients were calculated, with standard errors estimated using the method of Fleiss and coworkers. A second method of calculation treated examiners as separate strata as described by Fleiss. A third method was based on calculations of individual kappa scores for each examiner. Groups of examiners can then be analyzed using Student’s t test.

The first method treats all pairs of assessments as independent. Examiners, subjects, and muscle groups are treated as potential strata that are not used for statistical adjustment.

The second and third methods attempted to treat examiners as separate strata or cases. The third method, although less statistically sophisticated, seems preferable because it allows the use of robust parametric and nonparametric methods for comparing groups of examiners.

To test the extent of agreement among examiners, the method described by Fleiss for multiple examiners was used. As before, each muscle group in each subject was treated as independent.

Comparisons of groups used Fisher’s exact test for categorical data. Statistical significance was not routinely calculated for measures of agreement. With the numbers of cases used here, highly significant p values can usually be obtained even when the level of agreement is far from clinically significant.

**RESULTS**

**Trigger Point Assessment**

Figure 1 shows the proportions of taut band findings for each of the 10 muscle groups assessed in this study. The expert examiner reported a taut band in almost all the assessments, and no obvious differences between muscle groups were found. He found that 90% of all assessments in patients with low back pain and 70% in healthy subjects showed taut bands. Reports of this finding were lower in the trained examiners and lowest of all in the untrained examiners. The mean ± SD proportions were 73% ± 33% for the patients with low back pain and 62% ± 37% for the healthy subjects as reported by the trained examiners. Similarly, 35% ± 36% for patients with low back pain and 33% ± 32% for healthy subjects were reported by the untrained examiners.

Local twitch response findings are shown in figure 2. The proportion of muscles with this finding was much lower than the proportion of muscles with a taut band, and the difference between the expert and the examiners was not large. The expert examiner found 25% of all assessments in patients with low back pain and 15% in healthy subjects had findings of local twitch response. The mean ± SD proportions were 18% ± 27% for the patients with low back pain and 8% ± 20% for the healthy subjects as reported by the trained examiners. Similarly, 19% ± 29% for patients with low back pain and 13% ± 24% for healthy subjects were reported by the untrained examiners.

Figure 3 shows proportions of muscles with referred pain. As
expected, such findings were much more likely in the back-pain group than in the normal controls. The expert examiner found 19% of all assessments in patients with low back pain and 1% in healthy subjects had findings of referred pain phenomenon. The mean (±SD) proportions were 19% (±33%) for the patients with low back pain and 2% (±10%) for the healthy subjects as reported by the trained examiners. Similarly, 13% (±28%) for patients with low back pain and 1% (±8%) for healthy subjects were reported by the untrained examiners.

Agreement With Expert

Table 3 shows the proportions of agreements and disagreements for taut band, local twitch response, and referred pain, along with calculated kappa scores and their standard errors. The proportions of agreements and the overall kappa score are based on the treatment of each assessment as an individual case. The stratified kappa score uses examiners as strata, and the mean kappa score is the mean of individual kappa scores calculated for each examiner. Although most of these kappa scores reach conventional levels of statistical significance, they are not high. The highest kappa value was .316 for referred pain. This is below the level of .40 that is generally considered to be the lowest kappa score that indicates at least “fair” agreement.27

Figure 4 shows the proportions of agreements and kappa coefficients for taut band, twitch response, and referred pain. Kappa is somewhat higher for the trained examiners but does not reach .40. The greatest difference between trained and untrained examiners is in the assessment of taut bands. Trained examiners had significantly higher kappa values ($p < .001$). However, kappa scores were not high, and the effect of training could be described as raising the level of agreement from nonexistent (kappa = .042) to poor (kappa = .238).

Table 4 shows overall kappa scores for taut band, local twitch response, and referred pain in several subgroups. Physiatric and chiropractic examiners did not differ in the palpation findings and were equally unreliable. Subgroups based on the clinical condition of the subject produce the apparently paradoxical effect that each subgroup has a lower kappa coefficient than the entire population. Such effects are not unusual in calculation of kappa.28 Some of the muscle groups showed higher kappa coefficients, with gluteus maximus and quadratus lumborum exceeding the .40 threshold. Estimation of the reliability of this
effect and interpretation of its significance will require further research.

Agreement Among Evaluators

Figure 5 shows proportions of agreements and kappa coefficients for taut band, local twitch response, and referred pain with trained and untrained examiners. Kappa coefficients are somewhat higher. Kappa for referred pain in the trained group is .435. Kappa coefficients for the entire subject population and for subgroups are shown in Table 5. Results were similar to those seen for agreement with the expert.

DISCUSSION

This study shows how difficult it is to obtain good interexaminer reliability in clinical research on palpatory findings. For palpation of taut band and local twitch response, the examiners’ findings were not reliable and not different from chance. Palpation of referred pain was marginally reliable among the nonexpert physicians with extensive training (kappa = .435), and not reliable without training (kappa = .32). However, both the trained and untrained nonexpert physicians did not show acceptable reliability when results were compared with those of our expert, with a kappa value of .342 for the trained physicians and .326 for the untrained physicians. Our finding on the interexaminer reliability among trained, nonexpert physicians (kappa = .435) was similar to that reported by Njoo and Van der Does21 (kappa = .41), who also used nonexpert examiners and gave training. When expert physicians were not trained, the interexaminer kappa values (.4019 and .3820) were similar to those of nonexpert physicians with training. However, a kappa value of .6922 was reported when expert physicians were trained before the study. We believed that training which attempts to standardize the palpatory pressure for palpating referred pain would help to achieve a marginal interexaminer reliability for the nonexpert physicians, and a higher reliability for the expert physicians.

Clinically, physicians may have different findings when palpating MTrPs. These variations may arise from several different sources. For example, physicians may not palpate the same sites in palpating for the MTrP within a specific muscle, may not interpret the findings of a taut band the same way in terms of size or degree of tautness of the band, or may not snap the taut band the same way to produce local twitch responses. Therefore, differences in the amount of pressure used, in the angle of pressure or snapping applied, or in the tension of the taut band resulting from different positions would all contribute to the variability. As a result, these variations might have caused the poor interexaminer reliability for palpation of taut band and local twitch response.

Although we chose to have only one expert physician to serve as the “gold standard” in the study, it is understandable that the “gold standard” would change if more than one expert was used. If several experts were used, the kappa scores for the agreement between the nonexpert and the expert examiners would be less than what we reported (Table 4). Thus, palpation for taut band, local twitch response, and referred pain would still remain unreliable. However, the kappa scores among the nonexpert examiners themselves should not change (Table 5). This would imply that it is extremely difficult to obtain a gold standard for clinical palpation of MTrPs.

The kappa values on the different aspects of MTrP palpation, based on the findings of four previous studies,19–22 are summa-
rized in table 1. Local twitch response appears to show the lowest interexaminer reliability. Spot tenderness (local pressure required to produce pain) has higher interexaminer reliability than does referred pain. Jump sign (patient movement response to local pressure), pain recognition (application of pressure reproduces or aggravates patient’s pain complaint), and taut band (cordlike band of palpable muscle fibers) have either insufficient or conflicting results. Higher kappa values are assumed to result from less difficult examinations and lower values to result from more difficult examinations. It was therefore considered that palpation for local twitch response was the most difficult examination, palpation for spot tenderness and jump sign was the easiest examination, and palpation for pain recognition was easier than that for taut band and referred pain.

Simons estimated the relative diagnostic value of each examination based on the following considerations: Spot tenderness is an easy test to perform and an essential finding when examining for an MTrP. Spot tenderness is not, however, a specific sign of MTrP because it is also a defining characteristic of tender points of fibromyalgia, which are often located where MTrPs are likely to occur, and it is equally characteristic of enthesopathy. The jump sign is a more objective indicator of the degree of spot tenderness.

Pain recognition is a relatively easy test to perform and probably the most useful single diagnostic test for MTrP. Although good reliability has been reported, it is not specific to MTrP because it can also be found in patients with fibromyalgia.

Palpable taut band is a required sign for the diagnosis of an MTrP, although it may also occur in normal muscles without other clinical evidence of an MTrP. Palpating for the taut band, however, requires skill. Palpation of a taut band in a deep muscle is especially difficult for a nonexpert examiner. Spot tenderness of an MTrP is always found in a taut band of an accessible muscle.

Referred pain is thought to be a nonspecific finding and therefore of limited diagnostic value unless observed in combination with other findings. The location of the referred pain is useful as an initial guide as to which muscles may be harboring active MTrPs.

Recently, Simons suggested that the presence of spot tenderness or a jump sign in a palpable taut band is highly indicative of a latent or active trigger point and, only with additional confirmation of pain recognition, is diagnostic of an active trigger point. The added findings of referred pain, local twitch response, or both further confirms the presence of a trigger point, particularly during MTrP injection.

In our study, we were unable to obtain satisfactory reliability on MTrP examinations. Subgroup analyses of all nonexpert physicians, whether physiatric or chiropractic, did not show different results. Furthermore, short-term training of physicians who did not have extensive clinical experience in MTrP examination did not improve the reliability of palpation for taut band and local twitch response. On the contrary, Gerwin and coworkers reported kappa values of .85 for taut band and .44 for local twitch response among four expert examiners (two physiatrists and two neurologists) who took 3 hours of training to obtain uniform findings before they started the experiment. Based on our study and that of Gerwin and coworkers, it appears that extensive clinical experience in MTrP examination, coupled with training and standardization procedures, may be necessary to obtain acceptable interexaminer reliability for palpation of taut band, and local twitch response. The content of training sessions will also require a component of pilot study whereby all the trained examiners are able to achieve uniform definitions of terminology, examination procedures, and most importantly, the same results of findings in a trial sample. Our failure to show acceptable reliability could have come from lack of uniform results between the expert and nonexpert examiners before we started the experiment, even though we had done an intensive training on the nonexpert examiners. In fact, some of our patients reported that our expert actually gave stronger pressures than most other examiners.

There are some limitations to the study: (1) the findings are limited to patients with subacute mechanical low back pain and may not be applied to either acute or chronic low back pain conditions; (2) the findings are limited to one expert (CZH) and may not be generalized to other experts; (3) the findings may not be applied to other patients with subacute low back pain who have conditions that were excluded in our study; and (4) the findings are limited to three components of MTrP characteristics and shall not be generalized to MTrP palpation as a whole.

Research is needed to identify training methods to achieve interexaminer reliability for nonexpert physicians in identifying local twitch responses.

CONCLUSION

Among nonexpert physicians, physiatric or chiropractic, trigger point palpation is not reliable for identifying taut bands and local twitch responses. Palpation for referred pain was only marginally reliable if intensive training was provided. We suspect that extensive clinical experience in MTrP examination may play a more important role than the training factor in obtaining reliability.

Acknowledgments: We thank the following examiners for their participation: Jeffrey Ameen, DC; Ward Henry, DC; Calvin Okey, DO; Sanjay Patel, MD; Curtis Rigney, DC; Brad Smith, DC; David Sweeney, MD; Joseph Tarazzino, MD; and Kathleen Yang, DC. We also thank Sharon Davis, DC, who served as a research assistant. Finally, we thank the student volunteers from the Los Angeles College of Chiropractic who recorded the data.

References


Supplier
a. Activator Methods, Inc., 3714 East Indian School Road, Phoenix, AZ 85018.