REVIEW ARTICLE

A Systematic, Critical Review of Manual Palpation for Identifying Myofascial Trigger Points: Evidence and Clinical Significance

Corrie Myburgh, PhD, Anders Holsgaard Larsen, PhD, Jan Hartvigsen, PhD

ABSTRACT. Myburgh C, Larsen AH, Hartvigsen J. A systematic, critical review of manual palpation for identifying myofascial trigger points: evidence and clinical significance. Arch Phys Med Rehabil 2008;89:1169-76.

Objective: To determine the reproducibility of manual palpation in identifying trigger points based on a systematic review of available literature.

Data Sources: Medline (1965–2007), CINHAL (1982–2007), ISI Web of Science (1945–2007), and MANTIS (1966–2007) databases and reference lists of articles.

Study Selection: Reproducibility studies relating to identification and diagnosis of trigger points through palpation. Acceptable studies were required to specifically consider either inter- or intrarater reliability of trigger point identification through manual palpation and include κ statistics as part of their statistical assessment.

Data Extraction: Three independent reviewers considered the studies for inclusion and rated their methodologic quality based on the Standards for Reporting of Diagnostic Accuracy guidelines for the reporting of diagnostic studies.

Data Synthesis: Eleven studies were initially included; however, 5 were subsequently excluded based on the inclusion and exclusion criteria. Only 2 studies were judged to be of high quality, and the level of evidence criteria suggested that, at best, moderate evidence could be found from which to make pronouncements on the literature. Only local tenderness of the trapezius (κ range, .15–.62) and pain referral of the gluteus medius (κ range, .298–.487) and quadratus lumborum (κ range, .36–.501) were found to be reproducible.

Conclusions: The methodologic quality of the majority of studies for the purpose of establishing trigger point reproducibility is generally poor. More high-quality studies are needed to comment on this procedure. Clinicians and scientists are urged to move toward simpler, global assessments of patient status.

Key Words: Diagnosis; Myofascial pain syndromes; Palpation; Rehabilitation; Review [publication type].

AS AN ENTITY CENTRAL to myofascial pain phenomena, the so-called myofascial trigger point, or simply trigger point, continues to gain acceptance as a contributor to musculoskeletal pain.[1,2] The use of innovative technologies to quantify trigger point characteristics and establish a diagnostic criterion standard test is encouraging.[3,4] However, manual palpation still stands central to trigger point identification during the diagnostic process.[5,6]

The rationale for determining the presence of a trigger point by judging the presence or absence of a fairly unique set of criteria (ie, taut band, local tenderness, patient pain recognition, pain referral, a local twitch response [LTR], a jump sign) was initially popularized by Simons et al.[7] However, as research has progressed, these criteria have proven problematic. For instance, the authors of a recent systematic review[8] reported that trigger point criteria application has varied over time and that fewer criteria are now considered diagnostically relevant. Furthermore, investigations into the reproducibility of palpation exhibit both heterogonous methodologies and conflicting conclusions, making it difficult to assess whether this approach has indeed shown the necessary reproducibility.[6,9-18] However, there do not appear to be any systematic, critical reviews available, presenting a synthesis of available material concerning the latter issue of intra- and interexaminer reliability.

This situation makes it difficult to apply evidence-driven criteria in the clinical setting and to advance the scientific study of myofascial pain in an orderly and rigorous manner. We therefore decided to conduct a systematic review to critically assess the design and statistical methodology of the literature pertaining to this subject by means of standardized criteria for judging diagnostic studies.

Specifically, this study was aimed at determining the reproducibility of manual palpation in identifying the trigger point based on available literature and to apply our study outcomes to the clinical and research settings to inform scientific discourse around this topic.

METHODS

Definitions and Terms Relating to Trigger Point Criteria

Variations in trigger point terminology[7,8,19] warranted specific definitions in the context of our study. Taut bands are ropelike indurations palpated in the muscle fiber. Exquisite sensitivity within the taut band evoked during palpation is termed local tenderness. Pain referral is a distinct pain pattern perceived at a distance from the trigger point site (either spontaneously present or induced during palpation of the trigger point site). In instances in which the patient recognizes local tenderness or pain referral as their pain, patient pain recognition is noted. The LTR is a momentary contraction (fasciculation) of the taut band when stimulated mechanically. Last, the jump sign is patient vocalization or withdrawal from palpation when exquisite local tenderness is perceived.

Global assessment was defined as the process whereby the clinician judges trigger points as present/absent based on the

From the Institute of Sports Science and Clinical Biomechanics, University of Southern Denmark (Myburgh, Larsen, Hartvigsen).

Supported by the University of Southern Denmark.

No commercial party having a direct financial interest in the results of the research supporting this article has or will confer a benefit upon the authors or upon any organization with which the authors are associated.

Reprint requests to Corrie Myburgh, PhD, Institute of Sport Science and Clinical Biomechanics, Campusvej 55, Odense, Denmark 5230, e-mail: *cmyburgh@health.sdu.dk.*

0003-9993/08/8906-00849$34.00/0
doi:10.1016/j.apmr.2007.12.033

presence of 2 or more diagnostic findings and, consequently, whether they are of clinical significance.[17,20]

Study Selection

An expert research librarian assisted us in conducting searches in Medline (1965–2007), CINHAL (1982–2007), ISI Web of Science (1945–2007), and MANTIS (1966–2007) databases. In Medline, we initially attempted to identify any systematic reviews related to myofascial pain by using *trigger point** AND *systematic [sb]* as search terms. This yielded 25 articles, however, none relating to our topic of investigation. The search strategy was then broadened to include the search terms *myofascial pain* or *trigger point* or *trigger points* or *trigger area* or *trigger areas.* This resulted in 2916 hits. A separate key word search related to diagnostic testing using *sensitivity* or *specificity* or *criteria* or *validity* or *reliability* or *assessment* or *assessments* or *judgment* or *judgments* or *classification* or *classifications* or *reproducibility* or *clinical precision* or *manual palpation* resulted in 2,291,121 hits. These 2 searches were then combined, resulting in 612 hits, constituting the final search strategy output.

References were subsequently scanned for appropriate material. Once identified, we also conducted a related articles search for possible search-strategy exclusions. The search strategy was then logged into PubMed to retrieve automatic updates of new articles relevant to our topic. This process was run up to the date of submission for review (October 2007). In the independent ISI Web of Science, MANTIS, and CINHAL search engines, the same search rubrics were used and resulted in 687, 231, and 305 hits, respectively. The collective search strategy yielded 11 articles (table 1).

To be included, studies were required to be original published and peer-reviewed articles investigating the reproducibility of myofascial trigger point palpation in human subjects. Anecdotal, speculative, or editorial articles were excluded as well as studies not explicitly reporting on the reproducibility of trigger points. In this regard, statistical assessment of agreement, for example, κ or interclass correlation coefficient (ICC), was considered essential.

Data Extraction

Using a checklist based on the Standards for the Reporting of Diagnostic Accuracy guidelines (a Cochrane collaboration initiative),[21,22] descriptive data from included studies were extracted and recorded independently by all 3 authors. These were subsequently compared and discordances resolved through discussion to form a summarized composite of each study.

Table 1: Studies Identified Through Search Strategy and Final Inclusion

No.	Study Reference	Reason(s) for Exclusion
1	Ohrbach and Gale[9]	Objectives not applicable, no reliability statistics
2	Rasmussen et al[10]	Objectives not applicable
3	Wolfe et al[11]	NA
4	Nice et al[12]	No reliability statistics
5	Levoska et al[13]	NA
6	Njoo and Van der Does[14]	NA
7	Tunks et al[15]	Objectives not applicable, no reliability statistics
8	Lew et al[16]	No reliability statistics
9	Gerwin et al[17]	NA
10	Hsieh et al[18]	NA
11	Al-Shenqiti and Oldham[6]	NA

Abbreviation: NA, not applicable (study included).

Quality Assessment of Studies

A preexisting score sheet, devised for a similar use in the area of spinal palpation, was adapted for the purposes of this study.[20] The tool, constructed to be sensitive to 3 key dimensions in reproducibility (study population, design, statistical analysis), provides a quality criteria rating, which contributes to the determination of overall level of evidence. Two adaptations were made to the score sheets to render them appropriate to this study context. First, a "clear operational definition of trigger point findings" item was added. This decision was based on the literature expressing concerns over the variation in diagnostic criteria of studies relating to trigger point phenomena.[8,19] Second, we decided at the outset not to use publications that did not present κ or ICC values.[23,24]

The resultant quality assessment criteria are presented in figure 1. In the case of intraexaminer studies, the maximum score attainable was 4 because criteria 1 and 5 were not applicable. Studies with a methodologic score of 4 out of 6 (interexaminer) or 3 out of 4 (intraexaminer) or higher were considered to be of higher quality.

Synthesis

After completing the evaluation process, we constructed tables to view and consider the data. The first was descriptive in nature, reporting all the outcomes across the studies (table 2), and the second presented statistical data, muscle groups, body regions, and the strength of evidence supporting reproducibility (table 3).

Weighting of evidence and strength of association. We were aware that level of evidence (LOE) criteria could influence the results of this systematic review.[25] Given the variability in trigger point reporting, we opted for what Ferreira et al[25] termed a moderate LOE strategy.[20] Strong evidence required generally consistent findings in more than 1 high-quality study, moderate evidence required generally consistent findings in 1 high-quality study and 1 or more low-quality studies, or in multiple low-quality studies and insufficient evidence when only 1 study was available or inconsistent findings were observed in multiple studies.

Furthermore, given the similarity to the context of spinal palpation,[20] clinically acceptable reproducibility was set at the cut-point threshold of 0.4. Thus, a 2-step process was followed as previously used in similar studies.[20,26]

Sensitivity analysis. In keeping with systematic review methodology,[27] we tested the robustness of our LOE weighting system by subjecting the reviewed studies to cut-point value variation of $\pm 25\%$ (± 1.5 points) in quality score and ± 0.1 in reproducibility.[20]

RESULTS

Data Extraction

Eleven studies were initially identified as being potentially appropriate; however, 5 were excluded based on our inclusion and exclusion criteria. As indicated by table 1, the second-level filtering was caused by inappropriate objectives and/or the lack of reproducibility statistics. Three studies dealt with interexaminer reliability, 1 with both inter- and intraexaminer reliability, and 1 was purely an intraexaminer study (see table 2).