Inter-examiner reliability of diplomats in the mechanical diagnosis and therapy system in assessing patients with shoulder pain

Afshin Heidar Abady[1], Richard Rosedale[2], Tom J. Overend[1], Bert M. Chesworth[1,3], Michael A. Rotondi[4]

[1]School of Physical Therapy, Western University, London, Canada, [2]London Health Sciences Centre, London, Canada, [3]Department of Epidemiology and Biostatistics, Schulich School of Medicine and Dentistry, Western University, London, Canada, [4]School of Kinesiology and Health Sciences, York University, Toronto, Canada

Objective: To investigate the inter-examiner reliability of Mechanical Diagnosis and Therapy (MDT)-trained diplomats in classifying patients with shoulder disorders. The MDT system has demonstrated acceptable reliability when used in patients with spinal disorders; however, little is known about its utility when used for appendicular conditions.

Methods: Fifty-four clinical scenarios were created by a group of 11 MDT diploma holders based on their clinical experience with patients with shoulder pain. The vignettes were made anonymous, and their clinical diagnoses sections were left blank. The vignettes were sent to a second group of six international McKenzie Institute diploma holders who were asked to classify each vignette according to the MDT categories for upper extremity. Inter-examiner agreement was evaluated with kappa statistics.

Results: There was 'very good' agreement among the six MDT diplomats for classifying the McKenzie syndromes in patients with shoulder pain (kappa=0.90, SE=0.018). The raw overall level of multi-rater agreement among the six clinicians in classifying the vignettes was 96%. After accounting for the actual MDT category for each vignette, kappa and the raw overall level of agreement decreased negligibly (0.89 and 95%, respectively).

Discussion: Using clinical vignettes, the McKenzie system of MDT has very good reliability in classifying patients with shoulder pain. As an alternative, future reliability studies could use real patients instead of written vignettes.

Keywords: Inter-examiner reliability, McKenzie, Mechanical diagnosis and therapy, Shoulder, Clinical vignette

Introduction

It is accepted that an accurate diagnosis is an important prerequisite for developing an effective treatment strategy.[1] Interventions are ideally targeted to a specific diagnosis; hence, an incorrect diagnosis may well lead to inappropriate management of a pathological condition and an increased likelihood for a poor treatment outcome. If the procedures and tests used in an examination are not reliable and valid, an incorrect diagnosis is the likely sequela.[2] A key to accurate diagnosis is the reliability of the diagnostic tests being used by the clinician. Inter-rater reliability has been defined as 'the extent to which examiners, using the same test on the same patients, agree on the results of the test'.[3]

The literature has highlighted the fact that establishing an accurate diagnosis in patients with shoulder pain is problematic.[4–8] Many commonly used examination procedures and orthopedic special tests for the shoulder lack reliability[2,8] and validity.[4,9,10] Additionally, there is a growing body of evidence suggesting that the findings from imaging tests, such as US, CT, or MRI, should not be relied upon entirely for clinical decision making, as the incidence of pathological findings in clinically asymptomatic shoulders is significant.[11–14] This clearly compromises the clinician's ability to make an accurate patho-anatomical diagnosis. As a result, there have been calls for[6,8] and the development of[7,8,15,16] non-patho-anatomic shoulder subgroups so that interventions can be more accurately matched to the patients who are classified within a given subgroup.

One widely used non-patho-anatomical classification scheme is the Mechanical Diagnosis and Therapy (MDT) system, which was initially introduced by Robin McKenzie in 1981 as a new approach to the classification and management of patients with low back pain.[17] He later described application of this system to the cervical and thoracic spines.[18] The MDT system classifies patient presentations based on

Correspondence to: Afshin Heidar Abady, School of Physical Therapy, Western University, Room 1588 Elborn colleg, 1201 Western Road, London, Ont. N6G 1H1, Canada. Email: aheidara@uwo.ca

analyzing the symptomatic and mechanical effect of different loading strategies, positions, and postures.[19] Each MDT syndrome requires its own particular management approach.

A series of systematic reviews support the efficacy of the MDT system in the management of acute and chronic low back pain.[20–27] The MDT system for patients with spinal disorders has also demonstrated acceptable reliability,[28–34] as well as diagnostic and prognostic validity,[35–45] among experienced physiotherapists. McKenzie proposed that this system of diagnosis and treatment could also be applied to extremity disorders.[17] McKenzie's book on the application of MDT to human extremities[46] contains a detailed explanation of its clinical application to patients with peripheral joint disorders.

According to McKenzie, patients with extremity disorders can be classified into the following four syndromes.[46]

- Derangement syndrome: identified by a rapid response to a direction-specific loading strategy, known as the directional preference. A lasting improvement in symptoms, range of motion, and enhanced function will be achieved once the directional preference has been established and utilized.
- Articular dysfunction: distinguished by intermittent and consistent pain only produced at a diminished end range with a slower response to specific tissue loading strategy.
- Contractile dysfunction: distinguished by intermittent pain consistently produced, but this time only when the musculo-tendinous unit is loaded, for instance, with an isometric contraction against resistance.
- Postural syndrome: intermittent pain only produced by prolonged postures that, once avoided, result in a return to a normal pain-free state. The remainder of the physical examination is normal.
- Other: patients who cannot be classified under any of the mechanical syndromes. Examples include trauma, articular structurally compromised, recent surgery, and chronic pain syndrome (Appendix).

These categories allow for the full spectrum of musculoskeletal presentations to be classified within the MDT system.

The use of MDT in the extremities has not been investigated to the same extent as it has been in the spine. Currently, the scientific literature in this area has been limited to individual case studies, which generally reveal a very good treatment response.[47–54] One survey of the prevalence, classification, and preferred loading strategies for the use of the MDT system in the extremities has also been published; demonstrating that 30 participating therapists were able to use the system to successfully classify all patients with an extremity problem.[16] A more recent pilot RCT study conducted on patients with rotator cuff tendinopathy revealed comparable treatment outcomes in these patients using the MDT-based, self-managed, loaded exercise program versus the usual physiotherapy program.[55]

The MDT classification system, when used on patients with spinal disorders, has demonstrated acceptable inter-examiner reliability among trained physiotherapists.[28–34] In the extremities, Kelly *et al.*[56] conducted a pilot study with 11 patient vignettes and three MDT-trained practitioners, including two credentialed and one diploma therapists. May and Ross[19] continued with a follow-up study using 25 patient vignettes and 93 MDT diploma therapists. However, the inter-examiner reliability of the MDT classification system for the extremities has not been investigated in any samples comprised exclusively of patients with shoulder disorders. The previous two studies included patients with variety of extremity joint disorders, with no secondary analysis exploring inter-examiner reliability of the MDT system in any individual joint such as the shoulder. Only 7 out of 25 vignettes of the larger reliability study[19] were shoulder cases (correspondence from study author). The aim of our study was to investigate the inter-examiner reliability of MDT-trained diploma therapists when classifying patients with shoulder disorders.

Methods
Design and procedure
This was a two-phase study. In phase 1, a convenience sample of 11 MDT diploma holders were recruited from a publicly available list of MDT practitioners registered with the McKenzie Institute International who practice in Canada or the United States. They were asked to create 54 anonymous written clinical vignettes based upon findings from the initial assessment of previously treated patients with shoulder disorders. They were directed to document the patients' age in years, but 'not transfer' any identifying information regarding their patients including their name, address, telephone, and date of birth in order to maintain anonymity of the patients. The number of vignettes created for each sub-classification was 11 derangements, 11 articular dysfunctions, 11 contractile dysfunctions, 11 'spinal' category, which represents patients with shoulder pain deemed to be originating from the cervical spine, and 10 'other' MDT categories. Due to a very low incidence of 'postural syndrome' in patients with extremity disorders[16] a 'spinal' category was used as the fifth MDT subgroup for this study and the 'postural' subgroup was assigned to the 'other' categories. The 'spinal' category included patients with complaints of shoulder pain who were determined to have pain originating from the neck; this is commonly seen clinically and has been extensively reported in the literature.[46,52]

The standard McKenzie extremity assessment form routinely utilized by MDT practitioners was used to structure the clinical findings of the vignettes. In the

Table 1 Demographic information of the participating practitioners

Variables	Distribution
Number of raters	6
Age, mean (SD) (years)	51 (8.6)
Gender	Female: 2
	Male: 4
Years in practice, mean (SD)	25.7 (8)
Years since MDT diploma, mean (SD)	16 (4)
Proportion of extremity patients in caseload (n)	<25%: 2
	25%–50%: 4
Practice setting (n)	Private: 4
	Hospital outpatient: 1
	Specialty clinic: 1

MDT: Mechanical Diagnosis and Therapy, SD: standard deviation.

event that a clinician did not have any recent patients that would fit one specific MDT sub-classification, the vignette was created based on the presentation of patients in that subgroup from the past. A representative group of five vignettes are uploaded to the JMMT website. Ethical approval for the study was obtained from the Health Sciences Research Ethics Board of Western University.

In phase two, the 54 vignettes from phase 1 were used to examine inter-rater reliability. These vignettes were sent to six MDT diploma holders who practice in Canada and the United States who had no involvement with the first phase of the study. They were also recruited from the publicly available list of MDT practitioners registered with the McKenzie Institute International. Following informed consent, an explanation of the study was provided and the clinicians were asked to review each vignette and identify the MDT classification for each vignette from the following five subgroups: derangement, articular dysfunction, contractile dysfunction, spinal and other. All six clinicians were blinded to the MDT classification represented by each vignette.

Sample size

A confidence interval (CI) approach for sample size estimation of kappa was used.[57] This method allows researchers to design their inter-examiner agreement study with any number of outcomes and any number of examiners using a pre-specified level of precision in the estimation of kappa.[57] Assuming a preliminary estimate of kappa=0.7, with a 95% CI of 0.2, we

determined that 54 vignettes were needed for six examiners (MDT diploma holders).

Analysis

The kappa coefficient, standard error (SE), and raw percentage of agreement were calculated across the six participating physiotherapists. Data were analyzed using the MAGREE macro in Statistical Analysis System (SAS) version 9.3 for Windows. Kappa values were interpreted using the traditional thresholds of: less than 0.40=poor; 0.41–0.60=moderate; 0.61–0.80=good; and 0.81–1.00=very good.[58]

Results

Five physical therapists and one chiropractor who solely apply the MDT method when treating their patients with extremity disorders were recruited to classify the clinical vignettes. Demographic information provided by the participating practitioners is shown in Table 1. Distribution of the MDT classification ratings of the clinicians, in addition to the true classification of the vignettes is shown in Table 2.

There was consensus among all six raters on the vignettes' classification in 78% of the vignettes (42 out of 54). The raw overall level of multi-rater agreement among the six clinicians was 96%. The corresponding kappa value was 0.90 (SE=0.018). The highest level of chance-adjusted agreement was for the spinal category with kappa=0.96; the lowest level was for the 'other' category with kappa=0.80. By factoring in the true diagnoses of the vignettes in our analysis, the raw agreement and kappa were 95% and 0.89, respectively. Values of agreement for each one of the MDT classifications are shown in Tables 3 and 4.

Discussion

To our knowledge, this study is the first to address inter-examiner reliability of the MDT system exclusively in patients with shoulder pain. The results support the findings of previous reliability studies on application of the MDT in extremities.[50,51] The principal findings of our study suggest that experienced McKenzie practitioners have a 'very good' level of inter-examiner agreement when classifying patients with shoulder pain using the MDT system. The highest level of agreement was for the 'spinal' category with kappa=0.96, and the lowest

Table 2 Frequency (%) of vignette classification by rater

MDT classification	Actual classification (%)	Rater					
		1 (%)	2 (%)	3 (%)	4 (%)	5 (%)	6 (%)
Derangement	11 (20)	14 (26)	12 (22)	13 (24)	11 (20)	13 (24)	13 (24)
Articular dysfunction	11 (20)	11 (20)	9 (16)	10 (19)	11 (20)	10 (19)	10 (19)
Contractile dysfunction	11 (20)	11 (20)	11 (20)	11 (20)	10 (20)	12 (22)	11 (20)
Spinal	11 (20)	12 (22)	12 (22)	12 (22)	11 (20)	12 (22)	11 (20)
Other	10 (20)	6 (12)	10 (18)	8 (15)	11 (20)	7 (13)	9 (17)
Total	54 (100)	54 (100)	54 (100)	54 (100)	54 (100)	54 (100)	54 (100)

MDT: Mechanical Diagnosis and Therapy.

Table 3 Agreement findings by MDT classification across raters

MDT classification	Raw agreement (%)	Kappa	SE
Derangement	95	0.90	0.035
Articular dysfunction	97	0.90	0.035
Contractile dysfunction	97	0.92	0.035
Spinal	97	0.96	0.035
Other	94	0.80	0.035
Overall agreement	96	0.90	0.018

MDT: Mechanical Diagnosis and Therapy; SE: standard error.

level of agreement was for the 'other' MDT categories with kappa=0.80. The relatively lower level of agreement for the 'other' category was anticipated because multiple subcategories are included in this MDT classification. This makes diagnosis more challenging particularly when the decision is solely based on information collected in the initial assessment. A relatively higher level of agreement for the 'spinal' category may be due to the presence of more identifying symptoms, such as paresthesia, reported in some of the vignettes, and also the presence of, in some cases, a relatively quick response in the shoulder pain level of these patients by addressing their cervical spine. By including the actual classification of the vignettes in our analysis, as shown in Table 4, there is only a slight decline in both percent agreement and the kappa value. This slight decline could be due to the presence of insufficient clinical information provided in the vignettes, as these were based only on the clinical information gathered in the initial assessment session.

The results of our study on the shoulder generally reinforce the findings of previous reliability studies in the spine and the extremities, suggesting that the MDT system is a reliable method to classify patients with musculoskeletal shoulder disorders. Multiple studies have been conducted on inter-examiner reliability of the MDT system in patients with spinal disorders demonstrating an acceptable level of reliability among MDT practitioners in classifying their patients.[28–34] For instance, Razmjou *et al.*[28] and Kilpikoski *et al.*[30] reported good inter-examiner reliability between two MDT-trained therapists in classifying patients with low back pain into MDT classifications (kappa=0.7). In another type of study using video and written clinical vignettes, Werneke *et al.*[34] reported substantial to almost perfect inter-rater agreement in identifying

treatment approaches for neck and low back disorders among MDT-trained therapists. There are only two studies addressing inter-examiner reliability of the MDT system for patients with extremity disorders.[19,56] These two studies included a pilot study with 11 clinical vignettes[56] and three therapists, and a follow-up study with 25 clinical vignettes and 93 MDT diploma holders.[19] The pilot study showed 'good' agreement with kappa a value of 0.7, and the follow-up study revealed 'very good' agreement with kappa value of 0.83 (95% CI, 0.68–0.98). The clinical vignettes used for these studies were based on patients with both upper and lower extremity disorders. There was little difference between the reliability in upper (kappa=0.85) and lower extremity (kappa=0.80) cases.[19]

The major limitation of the current study was that only practitioners with an MDT diploma, the highest level of MDT training, were included. This limits the generalizability of the findings of this study, as the inter-rater agreement among clinicians without this level of training may not be as high. Therefore, this study is a first step toward evaluating the reliability of using the MDT system to classify patients with shoulder pain. Future studies should include practitioners with different levels of training and experience so that the agreement findings are generalizable to a broader group of practitioners. Another limitation of this study was using written vignettes instead of having actual patients. The major concern in this regard, as stated by Werneke *et al*,[34] is the purification of the intervention being expressed in the vignettes, which may not represent all aspects of clinical practice, making the diagnosis easier for the raters and inflating the calculated kappa value. One strength of using written vignettes is that this approach eliminates the potential error created by inconsistent patient presentations between raters. As an alternative, future studies could consider the use of real patients instead of written vignettes in order to further establish reliability of the MDT in extremities.

Table 4 Agreement by MDT classification across raters and the actual MDT vignette classification

MDT classification	Raw agreement (%)	Kappa	SE
Derangement	93	0.88	0.030
Articular dysfunction	96	0.87	0.030
Contractile dysfunction	97	0.93	0.030
Spinal	96	0.96	0.030
Other	93	0.77	0.030
Overall agreement	95	0.89	0.015

MDT: Mechanical Diagnosis and Therapy; SE: standard error.

Appendix: Alternative diagnostic sub-classifications in the Mechanical Diagnosis and Therapy system comprising the 'Other' category

Category	Definitions	Criteria - Essential (Common)	Examples (where necessary)
Trauma/recovering trauma	Recent trauma associated with onset of symptoms	Recent trauma associated with onset of constant symptoms/recent trauma associated with onset of symptoms in previous 6 weeks now intermittent and improving	
Red flags	Fracture to bone Malignant tumor	History of significant trauma Loss of function All movement make worse (Age>55) (History of cancer) (Unexplained weight loss) Progressive, non-mechanical pain, not relieved by rest	
Inflammatory	Inflammatory arthropathy	Constant Excessive movements exacerbate symptoms	RA, some stages of OA
Chronic pain syndrome	Pain-generating mechanism influenced by psychosocial factors or neurophysiological changes	Persistent widespread pain Aggravation with all activity Disproportionate pain response to mechanical stimuli Inappropriate beliefs and attitudes about pain	Regional pain syndromes
Post-surgery	Presentation relates to recent surgery	Recent surgery (Local post-surgery protocols may apply)	
Mechanically inconclusive	Unknown joint pathology	Inconsistent response to loading strategies (Inconsistent pattern of obstruction to movement)	
Peripheral nerve entrapment	Peripheral nerve entrapment	No spinal symptoms Local paresthesia/anesthesia (local muscle weakness)	Carpal tunnel syndrome, meralgia paraesthetica
Articular structurally compromised	Soft tissue and/or bony changes compromising joint integrity	Mechanical symptoms (ROM restricted, clunking, locking, catching) (Sensation of instability) Long history of symptoms or history of trauma Irreversible with conservative care	Late stage OA, dislocation, labral tear, cruciate ligament rupture, irreducible meniscal tear
Soft tissue disease process	A fibroblastic or degenerative disease process affecting inert soft tissue with unknown or disputed etiology	Each disease process has a unique clinical presentation, natural history, and varying degrees of efficacy to a variety of interventions	Frozen shoulder, Dupuytren's, plantar fascia syndrome
Vascular	Symptoms induced by poor blood supply due to pressure increase in a closed anatomical space	Poorly localized severe ache (Commonly induced by exercise or trauma) (Paresthesia in field of local cutaneous nerve) (Muscle feels tight or full)	Compartment syndrome

Source: May S, Rosedale R. A survey of the McKenzie classification system in the extremities: prevalence of mechanical syndromes and preferred loading strategies. Phys. Ther. 2012 Jul 26 [Epub ahead of print].

References

1 Dessaur WA, Magarey ME. Diagnostic accuracy of clinical tests for superior labral anterior posterior lesions: a systematic review. J Orthop Sports Phys Ther. 2008;38(6):341–52.

2 De Winter AF, Jans MP, Scholten RJPM, Deville W, van Schaardenburg D, Bouter LM. Diagnostic classification of shoulder disorders: interobserver agreement and determinants of disagreement. Ann Rheum Dis. 1999;58(5):272–7.

3 Lucas N, Macaskill P, Irwig L, Moran R, Bogduk N. Reliability of physical examination for diagnosis of myofascial trigger point. Clin J Pain. 2009;25:80–9.

4 Hegedus EJ, Goode A, Campbell S, Morin A, Tamaddoni M, Moorman CT. Physical examination tests of the shoulder: a

systematic review with meta-analysis of individual tests. Br J Sports Med. 2008;42:80–92.

5 Hughes P, Taylor NF, Green RA. Most clinical tests cannot accurately diagnose rotator cuff pathology: a systematic review. Aust J Physiother. 2008;54:159–70.

6 Schellingerhout JM, Verhagen AP, Thomas S, Koes BW. Lack of uniformity in diagnostic labeling of shoulder pain: time for a different approach. Man Ther. 2008;13:478–83.

7 Lewis JS. Rotator cuff tendinopathy: a model for the continuum of pathology and related management. Br J Sports Med. 2010;44(13):918–23.

8 May S, Chance-Larsen K, Littlewood C, Lomans D, Saad M. Reliability of physical examination tests used in the assessment of patients with shoulder problems: a systematic review. Physiotherapy. 2010;96:179–90.

9 Munro W, Healy R. The validity and accuracy of clinical tests used to detect labral pathology of the shoulder – a systematic review. Man Ther. 2009;14:119–30.

10 Tucker S, Taylor NF, Green RA. Anatomical validity of the Hawkins-Kennedy test – A pilot study. Man Ther. 2011;16:399–402.

11 Schibany N, Zehetgruber H, Kainberger F, Wurnig C, Ba-Sslamah A, Herneth AM, et al. Rotator cuff tears in asymptomatic individuals: a clinical and ultrasonographic screening study. Eur J Radiol. 2004;51(3):263–8.

12 Reilley P, MacLeod I, MacFarlane R, Windley J, Emery RJH. Dead men and radiologists don't lie: a review of cadaveric and radiological studies of rotator cuff tear prevalence. Ann R Coll Surg Engl. 2006;88:116–21.

13 Yamaguchi K, Ditsios K, Middleton WD, Hildebolt CF, Galatz LM, Teefey SA. A comparison of asymptomatic and symptomatic shoulders the demographic and morphological features of rotator cuff disease. J Bone Joint Surg Am. 2006;88(8):1699–704.

14 Girish G, Lobo LG, Jacobson JA, Morag Y, Miller B, Jamadar DA. Ultrasound of the shoulder: asymptomatic findings in men. AJR Am J Roentgenol. 2011;197(4):W713–9.

15 Carter T, Hall H, McIntosh G, Murphy J, MacDougall J, Boyle C. Intertester reliability of a classification system for shoulder pain. Physiotherapy. 2012;98(1):40–6.

16 May S, Rosedale R. A survey of the McKenzie classification system in the extremities: prevalence of mechanical syndromes and preferred loading strategies. Phys Ther. 2012;92(9):1175–86.

17 McKenzie RA. The lumbar spine: mechanical diagnosis and therapy. Waikanae, New Zealand: Spinal Publications; 1981.

18 McKenzie RA. The cervical and thoracic spine: mechanical diagnosis and therapy. Waikanae, New Zealand: Spinal Publications; 1990.

19 May S, Ross J. The McKenzie classification system in extremities: a reliability study using McKenzie assessment forms and experienced clinicians. J Manipulative Physiol Ther. 2009;32(7):556–63.

20 Clare HA, Adams R, Maher CG. A systematic review of efficacy of McKenzie therapy for spinal pain. Aust J Physiother. 2004;50:209–16.

21 Cook C, Hegedus EJ, Ramey K. Physical therapy exercise intervention based on classification using the patient response method: a systematic review of the literature. J Man Manip Ther. 2005;13:152–62.

22 Machado LAC, de Souza MVS, Ferreira PH, Ferreira ML. The McKenzie method for low back pain. A systematic review of the literature with a meta-analysis approach. Spine. 2006;31:E254–62.

23 Hettinga DM, Jackson A, Klaber Moffett J, May S, Mercer C, Woby SR. A systematic review and synthesis of higher quality evidence of the effectiveness of exercise interventions for nonspecific low back pain of at least 6 weeks duration. Phys Ther Rev. 2007;12:221–32.

24 Slade SC, Keating J. Unloaded movement facilitation exercise compared to no exercise or alternative therapy on outcomes for people with non-specific chronic low back pain: a systematic review. J Man Manip Ther. 2007;30:301–11.

25 May S, Donelson R. Evidence-informed management of chronic low back pain with the McKenzie method. Spine J. 2008;8:134–41.

26 Aina A, May S, Clare H. The centralization phenomenon of spinal symptoms – a systematic review. Man Ther. 2004;9:134–43.

27 May S, Littlewood C, Bishop A. Reliability of procedures used in the physical examination of non-specific low back pain: a systematic review. Aust J Physiother. 2006;52:91–102.

28 Razmjou H, Kramer JF, Yamada R. Intertester reliability of the McKenzie evaluation in assessing patients with mechanical low back pain. J Orthop Sports Phys Ther. 2000;30:368–89.

29 Fritz JM, Delitto A, Vignovic M, Busse RG. Interrater reliability of judgments of the centralisation phenomenon and status change during movement testing in patients with low back pain. Arch Phys Med Rehabil. 2000;81:57–61.

30 Kilpikoski S, Airaksinen O, Kankaanpaa M, Leminen P, Videman T, Alen M. Interexaminer reliability of low back pain assessment using the McKenzie method. Spine. 2002;27:207–14.

31 Clare HA, Adams R, Maher CG. Reliability of the McKenzie spinal pain classification using patient assessment forms. Physiotherapy. 2004;90:114–9.

32 Clare HA, Adams R, Maher CG. Reliability of McKenzie classification of patients with cervical or lumbar pain. J Manipulative Physiol Ther. 2005;28:122–7.

33 Dionne CP, Bybee RF, Tomaka J. Inter-rater reliability of McKenzie assessment in patients with neck pain. Physiotherapy. 2006;92:75–82.

34 Werneke MW, Hart DL, Deutscher D, Stratford PW. Clinician's ability to identify neck and low back interventions: an inter-rater chance corrected agreement pilot study. J Man Manip Ther. 2011;19:172–81.

35 Dionne CP, Bybee RF, Tomaka J. Correspondence of diagnosis to initial treatment for neck pain. Physiotherapy. 2007;93:62–8.

36 Long A. The centralisation phenomenon. Its usefulness as a predictor of outcome in conservative treatment of chronic low back pain. Spine. 1995;20:2513–21.

37 Sufka A, Hauger B, Trenary M, Bishop B, Hagen A, Lozon R, et al. Centralisation of low back pain and perceived functional outcome. J Orthop Sports Phys Ther. 1998;27:205–12.

38 Werneke M, Hart DL, Cook D. A descriptive study of the centralization phenomenon. A prospective analysis. Spine. 1999;24:676–83.

39 Werneke M, Hart DL. Centralisation phenomenon as a prognostic factor for chronic pain or disability. Spine. 2001;26:758–65.

40 Werneke M, Hart DL. Discriminant validity and relative precision for classifying patients with nonspecific neck and back pain by anatomic pain patterns. Spine. 2003;28:161–6.

41 Wetzel FT, Donelson R. The role of repeated end-range/pain response assessment in the management of symptomatic lumbar discs. Spine J. 2003;3:146–54.

42 Berthelot JM, Delecrin J, Maugars Y, Passuti N. Contribution of centralization phenomenon to the diagnosis, prognosis, and treatment of diskogenic low back pain. Joint Bone Spine. 2007;74:319–23.

43 Long A, May S, Fung T. The comparative prognostic value of directional preference and centralization: a useful tool for front-line clinicians? J Man Manip Ther. 2008;16:248–54.

44 Werneke M, Hart DL, Resnik L, Stratford PW, Reyes A. Centralization: prevalence and effect on treatment outcomes using a standardized operational definition and measurement method. J Orthop Sports Phys Ther. 2008;38(3):116–25.

45 Werneke M, Hart DL, Cutrone G, Oliver D, McGill M, Weinberg J, et al. Association between directional preference and centralization in patients with low back pain. J Orthop Sports Phys Ther. 2011;41(1):22–31.

46 McKenzie RA, May S. The human extremities: mechanical diagnosis and therapy. Waikanae, New Zealand: Spinal Publications; 2000.

47 Aina A, May S. A shoulder derangement. Man Ther. 2005; 10:159–63.

48 Littlewood C, May S. A contractile dysfunction of the shoulder. Man Ther. 2007;12:80–3.

49 Shouta K, Hiroshi T, May S. Application of mechanical diagnosis and therapy to a patient diagnosed with a de Quervain's disease: a case study. J Hand Ther. 2009;22:278–84.

50 Littlewood C. Contractile dysfunction of the shoulder (rotator cuff tendinopathy): an overview. J Man Manip Ther. 2012;20(4):209–13.

51 Krog C, May S. Derangement of the temporomandibular joint; a case study using mechanical diagnosis and therapy. Man Ther. 2012;17(5):483–6.

52 Menon A, May S. Shoulder pain: differential diagnosis with mechanical diagnosis and therapy extremity assessment – a case report. Man Ther. 2013;18(4):354–7.

53 Lynch G, May S. Directional preference at the knee: a case report using mechanical diagnosis and therapy. J Man Manip Ther. 2013;21(1):60–6.

54 Kidd J. Treatment of shoulder pain utilizing mechanical diagnosis and therapy principles. J Man Manip Ther. 2013;21(3):168–73.

55 Littlewood C, Malliaras P, Mawson S, May S, Walters SJ. Self-managed loaded exercise versus usual physiotherapy treatment for rotator cuff tendinopathy: a pilot randomised controlled trial. Physiotherapy. 2014;100(1):56–60

56 Kelly E, May S, Ross J. The reliability of the McKenzie classification system using extremity McKenzie assessment forms. Int J Mech Diagn Ther. 2008;3:3–6.

57 Rotondi MA, Donner A. A confidence interval approach to sample size estimation for studies of interobserver agreement with multiple raters and outcome. J Clin Epidemiol. 2012;65:778–84.

58 Streiner DL, Norman GR, Health measurement scales. A practical guide to their development and use, 3rd edn. Oxford: Oxford University Press; 2003.