Manual Therapy xxx (2008) 1–7

Contents lists available at ScienceDirect

Manual Therapy

journal homepage: www.elsevier.com/locate/math

Original article

Inter-examiner reliability of a classification system for patients with non-specific low back pain

K. Vibe Fersum [a,*], P.B. O'Sullivan [b], A. Kvåle [a], J.S. Skouen [a,c]

[a] Section for Physiotherapy Science, Department of Public Health and Primary Health Care, University of Bergen, Kalfarveien 31, 5018 Bergen, Norway
[b] School of Physiotherapy, Curtin University, Bentley 6102, WA, Australia
[c] The Outpatient Spine Clinic, Department of Physical Medicine and Rehabilitation, Haukeland University Hospital, Bergen, Norway

ARTICLE INFO

Article history:
Received 20 February 2008
Received in revised form 10 July 2008
Accepted 1 August 2008

Keywords:
Agreement
Classification
Low back pain
Reliability

ABSTRACT

There is a lack of studies examining whether mechanism-based classification systems (CS) acknowledging biological, psychological and social dimensions of long-lasting low back pain (LBP) disorders can be performed in a reliable manner. The purpose of this paper was to examine the inter-tester reliability of clinicians' ability to independently classify patients with non-specific LBP (NSLBP), utilising a mechanism-based classification method. Twenty-six patients with NSLBP underwent an interview and full physical examination by four different physiotherapists. Percentage agreement and Kappa coefficients were calculated for six different levels of decision making. For levels 1–4, percentage agreement had a mean of 96% (range 75–100%). For the primary direction of provocation Kappa and percentage agreement had a mean between the four testers of 0.82 (range 0.66–0.90) and 86% (range 73–92%) respectively. At the final decision making level, the scores for detecting psychosocial influence gave a mean Kappa coefficient of 0.65 (range 0.57–0.74) and 87% (range 85–92%). The findings suggest that the inter-tester reliability of the system is moderate to substantial for a range of patients within the NSLBP population in line with previous research.

1. Introduction

LBP represents a common and very costly health problem and a definite diagnosis is difficult to achieve in most cases (85%) (Waddell, 2004). As a result, uncertainty regarding treatment of this group of patients is common (Cherkin et al., 1998).

A number of studies have shown little or no difference between various physiotherapy treatments for chronic LBP (Delitto et al., 1995; Petersen et al., 1999; Ferreira et al., 2007). Several authors have suggested that these results may reflect the heterogeneity of the NSLBP group, with several distinct subgroups, including psychosocial problems, each with its own potential set of beneficial treatments (O'Sullivan, 2000; Petersen et al., 2002; O'Sullivan, 2005; Dankaerts et al., 2006b). There is growing evidence suggesting that sub-classifying patients and offering them tailored interventions matching their disorder improves patient outcome (Frymoyer et al., 1985; Main and Watson, 1996; O'Sullivan, 1997; Nachemson, 1999; Linton, 2000; Skouen et al., 2002; Fritz et al., 2003; Stuge et al., 2004). It has been proposed that a classification system (CS) for NSLBP should identify the underlying mechanisms driving the disorder within a bio-psycho-social framework, enabling specific therapies to be applied so as to favourably influence the outcome of the disorder (O'Sullivan, 2005).

A number of CS have been proposed (McKenzie, 1981; Spitzer, 1987; Maluf et al., 2000; Sahrmann, 2001). However, only a few are found sufficiently reliable and valid (Petersen et al., 1999), and even fewer consider the disorder from a bio-psycho-social perspective (Petersen et al., 1999; Ford et al., 2003; McCarthy et al., 2004; O'Sullivan, 2005; Dankaerts et al., 2006b).

The Quebec Task Force system was designed to classify all LBP patients to help with clinical decision making, establishing prognosis and evaluating treatment effectiveness (Spitzer, 1987). However, it has not been tested for reliability and does not consider the underlying mechanism (Dankaerts et al., 2006b), except for differentiating somatic from radicular pain. Within this system there is no subgrouping of NSLBP except on the basis of pain area, and no specific treatment is advocated for this large group of patients other than general exercise, therefore limiting its use for physiotherapy assessment and treatment (Padfield et al., 2002).

The McKenzie (1981) system is based on information from history taking, and symptom response to generated loading of the lumbar spine. The system has been tested for reliability, and has substantial inter-tester agreement when applied by trained examiners (Kappa coefficients ranging from 0.6 to 0.7) (Kilpikoski et al., 2002).

* Corresponding author: Tel.: +47 55586711.
E-mail address: kjartan.fersum@isf.uib.no (K. Vibe Fersum).

1356-689X/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.math.2008.08.003

Petersen and co-workers (2004) have proposed a McKenzie-based CS with good inter-tester reliability, but it has a patho-anatomical orientation and lacks clear guidelines for management.

Sahrmann and co-workers have developed another CS, comprising five categories based on testing of muscular stability, alignment, asymmetry, flexibility of the lumbar spine, pelvis, and hip (Maluf et al., 2000). Reliability of the individual tests used for classification has been shown to vary from fair to almost perfect (Van Dillen et al., 1998, 2003). However, there are no reports on reliability in classification of the patients into the five categories, nor does this system consider patho-anatomical or psychosocial dimensions.

Since 1997 Peter O'Sullivan has developed a novel system, based on the Quebeck Task Force, incorporating multiple dimensions in the classification of patients into subgroups based on proposed underlying pain mechanisms. Initially, this mainly targeted a subgroup of patients with localised NSLBP where provocative movement behaviours and positions of the spine, associated with a loss of spinal control, represent a mechanism for ongoing pain. These patients are classified as LBP patients with motor control impairment (MCI). The evidence validating this subgroup is growing (O'Sullivan et al., 1997, 2005; O'Sullivan, 1997, 2000, 2003; Dankaerts et al., 2006a) and the reliability of clinicians to identify these different subgroups has been established (Dankaerts et al., 2006b). Lately, this approach has also incorporated classification of patients with lumbo-pelvic pain and a wider range of pain mechanisms linked to their disorder (O'Sullivan, 2005; O'Sullivan and Beales, 2007a). This system differentiates between specific LBP versus NSLBP. NSLBP is further split into subgroups based on the proposed driving mechanism behind the disorder (Fig. 1). The classification is based on a systematic examination process (subjective history, objective examination and available medical information). Within this system psychosocial factors are accounted for, acknowledging their potential to amplify pain and drive disability. To date the ability of clinicians to agree on this broad classification process has not been formally tested.

Validating the system has been a multi-step process, in which establishing inter-tester reliability is crucial. The aim of this study was therefore to examine the inter-tester reliability of clinicians' ability to independently classify a wide range of patients with NSLBP, utilising an extended mechanism-based classification method lately developed by O'Sullivan.

2. Methods

The study was conducted from March 2006 to June 2006, and was approved by the regional ethics committee of medical research in western Norway.

2.1. Patients

Patients were recruited consecutively from physiotherapy clinics around Bergen and from The Outpatient Multidisciplinary Spine Clinic, Haukeland University Hospital. After recruitment, a telephone screening was performed, and the first 30 patients that fit the inclusion criteria, were tested (Table 1). Since the patients were tested twice on each occasion, a 0–10 pain numerical rating scale was conducted prior to each testing. If a patient's pain score changed $\geq$2 levels between two examinations on the same day, this was considered to be a threat to the classification validity and the patient would then be excluded. Four patients were excluded after further examination: three did not fulfil the inclusion criteria and one reported a two-level change in pain between examinations on the given day.

This left 26 patients participating in the study. See Table 2 for the patients' characteristics. Prior to the study, design and possible risks were fully explained to each subject, and all signed a consent form.

2.2. Examiners

There were four physiotherapists, each with several years of experience in examination and treatment of LBP patients (mean 12 years, range 7–20 years). Three of the four examiners were physiotherapists with a masters degree in manual therapy. One was the developer of the system.

2.3. Training

All the examiners had been educated in the CS during several workshops with the developer, and were using it in their clinical practice. Prior to the study, O'Sullivan explained procedures and classifications were discussed using a series of case studies. The examiners also underwent a pilot training period where O'Sullivan examined and classified six patients, while the three others observed. The aim was to refine the specific criteria for assessment, as well as making testers more familiar with the system. The estimated training time for each therapist ranged from 69 to 140 h, the average being 106.3 h (workshops and pilot study included).

2.4. Clinical procedure

A test–retest design was utilised. A classification manual was developed by O'Sullivan prior to the study. The patients underwent a comprehensive interview and full physical examination by each of the four physiotherapists. Rather than assess the reliability of individual tests, this system involved making a disorder classification based on compilation of subjective and physical examination findings in relation to other medical tests and radiological imaging. The subjective assessment included pain area (pain drawing), intensity and nature, pain behaviour (aggravating/easing movements), identification of primary impairments, disability levels, avoidance behaviours, pain coping and pain beliefs. The examination involved assessment of spinal range of movement, analysis of the patient's primary physical impairments (pain provocative and easing postures, movements and functional tasks). Specific muscle and movement tests were performed to identify the relationship between the control of the lumbo-pelvic region and the pain disorders (O'Sullivan, 2000), as well as specific articular tests for the lumbar spine and pelvic region as indicated to identify the structural source of pain and the presence of movement impairments (MI). These are important elements in the classification of the pain disorder and in determining whether the habitual movements or postures are provocative or protective (O'Sullivan, 2000, 2005; O'Sullivan and Beales, 2007a,b). The process consists of several stages before reaching a classification (Fig. 1):

1 The first part involves screening; determining if the condition is specific LBP or NSLBP (O'Sullivan, 2005).
2 The second stage considers whether specific LBP disorders have an adaptive or maladaptive response to the disorder (O'Sullivan, 2005). If the disorder is classified as non-specific, then consideration of whether the disorder is predominantly centrally or peripherally mediated is made. The presence of localised and anatomically defined pain, associated with specific and consistent mechanical aggravating and easing factors, suggests that physical/mechanical factors are likely to dominate the disorder resulting in a peripheral nociceptive drive. Constant, non-remitting widespread pain, not influenced by mechanical factors, could on the other hand indicate inflammatory or centrally driven pain (O'Sullivan, 2005).

Please cite this article in press as: Vibe Fersum K et al., Inter-examiner reliability of a classification system for patients with non-specific low back pain, Manual Therapy (2008), doi:10.1016/j.math.2008.08.003

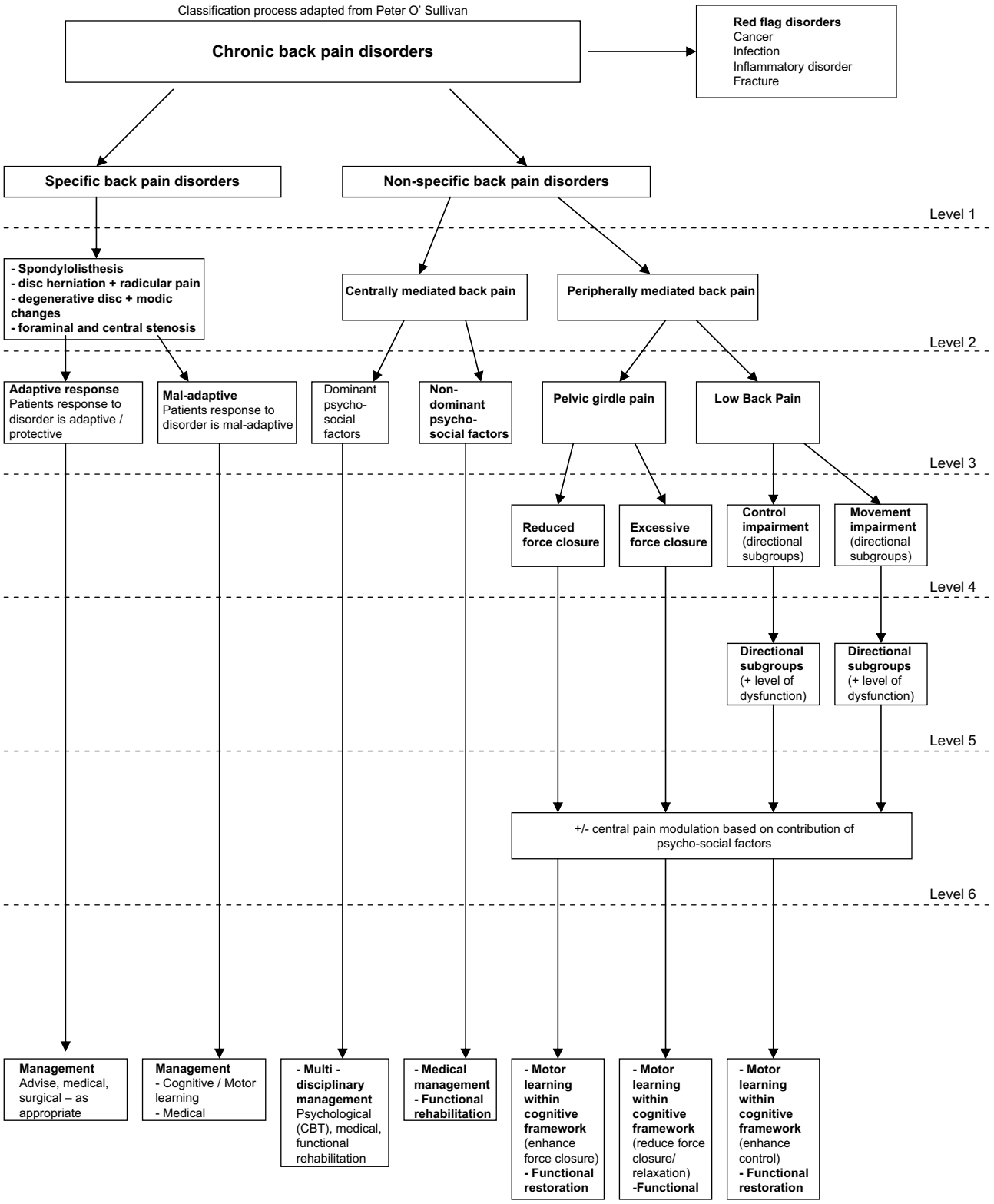

Fig. 1. Classification process adapted from Peter O'Sullivan (O'Sullivan, 2005; O'Sullivan and Beales, 2007a,b).

3 Centrally mediated pain can then be further sub-classified into the presence of non-dominant or dominant psychosocial factors. Peripherally mediated disorders are sub-classified into either LBP or a pelvic girdle pain disorders.
4 Peripherally mediated lumbar spine pain disorders are divided into MI or MCI disorders and peripherally mediated pelvic girdle pain into excessive or deficit of force closure. Both these classifications have been described in detail elsewhere (O'Sullivan, 2005; O'Sullivan and Beales, 2007a,b).
5 If the lumbar spine is the source of pain, the primary directional provocation bias as well as the symptomatic spinal level is noted.

Please cite this article in press as: Vibe Fersum K et al., Inter-examiner reliability of a classification system for patients with non-specific low back pain, Manual Therapy (2008), doi:10.1016/j.math.2008.08.003

Table 1
Inclusion/exclusion criteria

Inclusion criteria	Exclusion criteria
Patients with non-specific LBP (NSLBP) (≥6 weeks)	Sick-listed for more than 4 months continuous duration during last year
Male or female	Acute exacerbation of LBP
Age between 18 and 65 years	Radicular pain. Positive neural tissue provocation tests (primary peripheral symptoms)
Localised LBP: primarily in the area from T12 to gluteal folds	Any low limb surgery on the last 3 months
Moderate ongoing LBP, VAS > 2/10 and Oswestry > 14%	Surgery involving the lumbar spine (fusion)
Mechanical provocation of pain: postures, movement and activities	Pregnancy
	Psychiatric disorders
	Widespread non-specific pain disorder (no primary LBP focus)
	Specific diagnoses: active rheumatologic disease, progressive neurological disease, serious cardiac or other internal medical disease

6 The final decision is to indicate if significant psychosocial factors are associated with the disorder, based on all information from the examination process. The evaluation of psychosocial factors considers the presence of underlying fear avoidance behaviour, as well as psychological and social drivers considered to contribute to the pain disorder. Within this reasoning process, consideration is given to whether the patient has adapted in a positive (confrontation, active coping and minimal avoidance behaviours) or negative manner (passive coping and fear avoidance).

Each testing took about 1 h. The patient was examined independently twice on two days, within a 1-week period. Each therapist filled out a classification form (see Supplementary Appendix 1) and put it in a sealed opaque envelope after their patient assessment. After examination the patient completed several questionnaires to formally assess their disorder. This included a pain drawing, a functional assessment chart from the Dartmouth Primary Care Cooperative Information Project (COOP/WONCA), Oswestry Disability Index (ODI), Hopkins Symptoms Check List (HSCL), Fear Avoidance Beliefs Questionnaire (FABQ) and Ørebro Musculoskeletal Pain Screening Questionnaire (Ørebro MSPSQ).

2.5. Analysis

After completed examinations, the results were compared and logged. The developer's classification of each patient was used as the gold standard to which the other results were compared. Kappa coefficients and percentage of agreement were calculated using SPSS 13.0 for Windows. Cohen's Kappa statistic was used to calculate inter-tester reliability and Landis and Koch's (1977) values for interpretation of the reliability scores were used. Kappa values <0.20 indicate poor agreement, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 substantial, and 0.81–1.00 indicate almost perfect agreement. The data was analysed based on agreement of overall classification (specific LBP vs NSLBP), centrally or peripherally mediated, adaptive or maladaptive movement disorders, and whether it was considered to be a pelvic girdle pain or LBP disorder. Kappa agreement of the primary directional pain provocation, the spinal level of pain provocation and the presence of psychosocial influence on their LBP disorder was calculated.

Table 2
Patients' characteristics

Number of patients	26
Female	11
Male	15
Mean age (years)	32.4
Mean pain intensity	6/10
Mean duration (years)	4.9
Mean Oswestry	21.2/100
Mean HSCL	1.53/4
Mean Ørebro score	87.5/210

3. Results

In the first part of the classification process, all patients were classified with NSLBP with 98% agreement for this level. All patients in the study had pain arising from a peripheral pain source, with 99% agreement for this. One patient was classified by all four testers as having pelvic girdle pain (100% agreement); the rest were classified as LBP disorders (99% agreement). The fourth level considered increased or decreased force closure for pelvic pain (one patient, 100% agreement), MCI (24 patients, 99% agreement) or MI (one patient, 75% agreement) for low back. In the fifth level, Kappa agreement could be calculated, deciding the directional pattern of provocation (Fig. 2). For the primary direction of provocation, Kappa (K) and percentage agreement had a mean between the four testers of 0.82 (range 0.66–0.90) and 86% (range 73–92%) respectively. Increased familiarity with the system also increased the reliability results (<100 h $K = 0.66$, >100 h $K = 0.90$). In the final level of decision making, the mean Kappa coefficient for detecting psychosocial influence was 0.65 (range 0.57–0.74) and the mean agreement 87% (range 85–92%).

4. Discussion

The principal finding of our study suggests that therapists with substantial training in this CS (O'Sullivan, 2005) demonstrated fair to excellent agreement (Landis and Koch, 1977) in primary classification of the disorder, identification of directional patterns of provocation and the presence of psychosocial factors associated with the disorder, when applied to a range of NSLBP patients. Our findings are in accordance with a recent study (Dankaerts et al., 2006b), who also found moderate to excellent agreement between testers examining patients within the MCI subgroup. Their study consisted of two separate parts. The first part demonstrated almost perfect agreement between two expert clinicians when classifying 35 patients with MCI identified from a clinical case load, into the various directional patterns ($K = 0.96$, agreement 97%). In the second part, 25 out of 35 patients with MCI in the first study were randomly selected. These were videotaped and classified into directional groups by 13 other therapists based on the video and subjective complaints of the patients. The agreement between the 13 different raters was moderate to excellent (mean Kappa 0.61, agreement 70%).

As in Dankaerts et al.'s study (2006b), familiarity with the CS also influenced the reliability results, demonstrating higher agreement among raters with more CS training. These findings are in line with Strender's study (1997), concluding that reliability of clinical tests requires sufficient time for examination and conformity of performance, definitions and evaluations. The protocol of our study followed a similar examination procedure as the first part of Dankaerts et al.'s (2006b) study. By including any patient with localised low back pain in our study, we anticipated a more heterogenic NSLBP population with the inclusion of patients with back pain associated with MI as well as pelvic girdle pain disorders. However, 24 out of the 26 patients were classified as having MCI, which is in line with the findings of Dankaerts et al. (2006b). Furthermore, the current study involved four therapists examining the patients versus two in the first part of Dankaerts study (2006b).

Please cite this article in press as: Vibe Fersum K et al., Inter-examiner reliability of a classification system for patients with non-specific low back pain, Manual Therapy (2008), doi:10.1016/j.math.2008.08.003

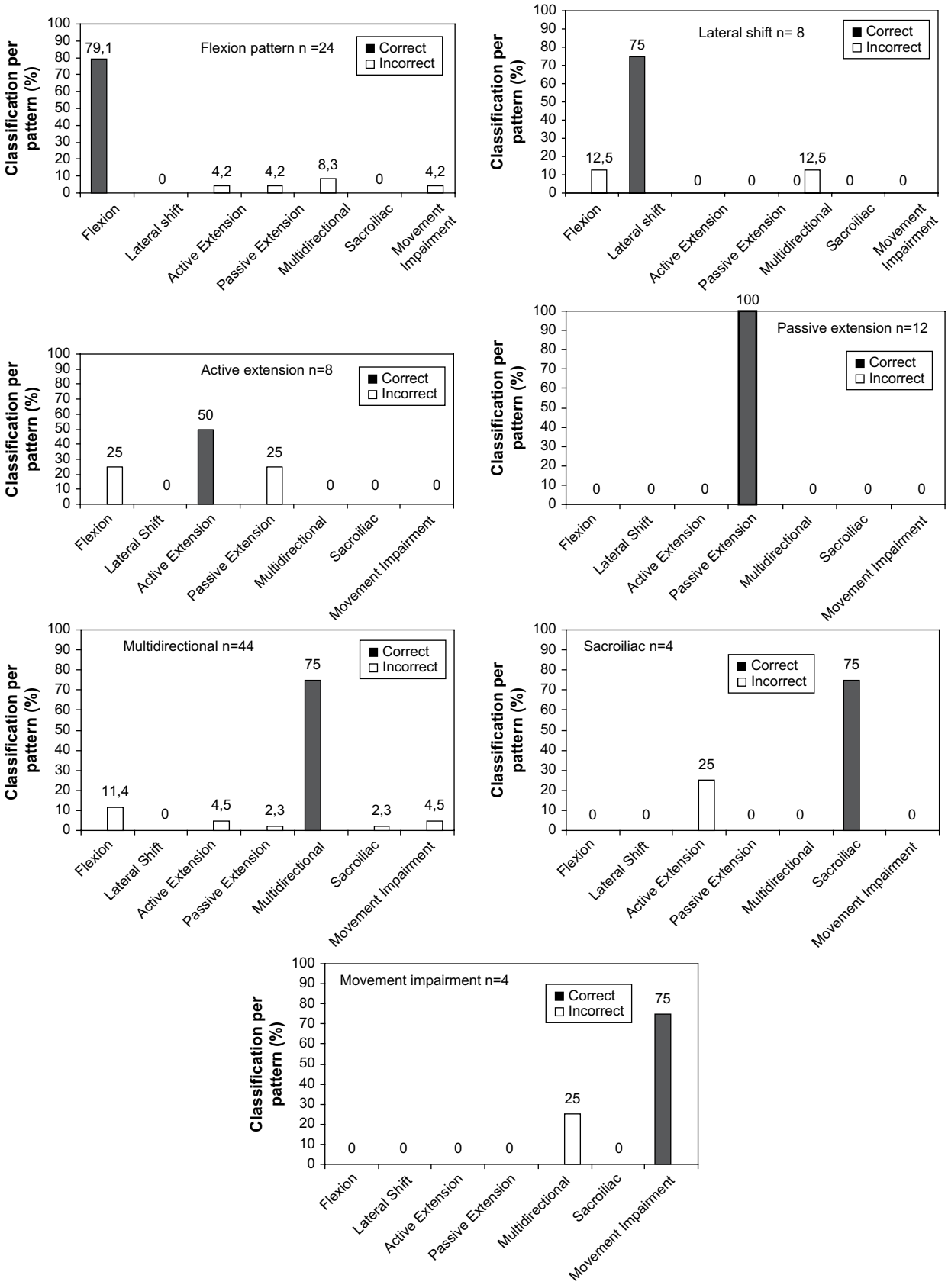

Fig. 2. Classification per different pattern (in %) by all examiners; n = total number of that specific pattern included × 4 (total number of examiners).

This may explain the greater reliability in this aspect of Dankaerts study, in comparison to ours. With regards to the second part of Dankaerts et al.'s study (2006b), it was acknowledged that the use of previously collected information (both subjective and video) represented a bias for the 13 clinicians. In our study, the testers did not have any prior information regarding the patient's disorder as this could influence the classification reliability, as different raters may gather information from patients in different ways.

Eight subjects in our study out of 26 with disorders classified as peripherally mediated NSLBP were also identified as having moderate, but significant psychosocial factors contributing to their disorders. Analysis of the questionnaire data collected after all assessments, confirmed that these eight patients scored significantly higher on HSCL and Ørebro MSPSQ ($p < 0.05$). Linton and Hallden (1998) identified potential psychosocial risk factors associated with future sick absenteeism in a study, using the Ørebro as the screening instrument. High total scores were related to outcome and to cut-off points that correctly identified the prognosis of nearly 80% of the patients. Psychosocial factors can modulate pain behaviour, which then can increase disability via fear avoidance, as well as promoting pain levels via central mechanisms (Vlaeyen and Linton, 2000). However there is little evidence to date that physiotherapists can identify these subjects at risk, based on subjective examination.

It has been emphasised (Dankaerts et al., 2006b) that the development of a multi-dimensional mechanism-based CS based on a bio-psycho-social framework should be seen as a critical development of a CS. The Quebeck Task Force has been considered by many as the first CS that included biomedical, psychological and social considerations in the classification process (McCarthy et al., 2004). The system used in our study, developed by O'Sullivan, utilises the Quebeck Task Force as an underlying framework, by classifying specific LBP versus NSLBP, the stage of the disorder, and the presence of red and dominant yellow flags. However, patients are sub-classified further, identifying the primary direction of provocation and the proposed underlying mechanism of the disorder. Furthermore, very specific interventions are indicated for the different classifications (O'Sullivan, 2005; O'Sullivan and Beales, 2007a,b).

In contrast, the McKenzie CS is a bio-system that lacks validity within a chronic LBP population, as only about 40% of patients have a directional pain preference (Donelson et al., 1990). Consistent with our findings, 45% of the subjects were classified as having MCI with multi-directional pain provocation, suggesting that a uni-directional preference was not present. This lack of uni-directional preference limits the use of directional treatment methods as advocated by McKenzie.

Interestingly, 25 of the patients in our study had MCI, and only one had MI. This finding is consistent with reports that impairments of range of motion are often not present in chronic low back pain disorders (Nattrass et al., 1999). However the lack of subjects with MI disorders in this study limits the ability to confirm the reliability of physiotherapists when identifying this subgroup.

The Sahrmann CS for NSLBP proposes a single mechanism for LBP (movement dysfunction), but does not consider specific diagnosis of LBP, CNS mediated pain, MIs or psychosocial factors, limiting its application within a chronic LBP setting. Petersen et al. (2004) in contrast proposed a system that demonstrated substantial reliability, but it lacked clear guidelines for management.

Reliability can be influenced by many different factors. The participants seemed representative of the population normally seen in primary health care, but the small sample may not be representative of the chronic LBP population. The first part of the classification process in this study was to determine whether the patient's condition was specific or non-specific. Secondly, an assessment was made to classify the source of the underlying mechanism as being centrally or peripherally driven. Our study's inclusion criteria were aimed at subjects with localised NSLBP that was mechanically provoked, making it more likely that they had a peripheral pain disorder. None of our subjects were classified with neurogenic pain. This fits with Bogduk's study (1995), which concluded that most NSLBP disorders are peripherally mediated, having a pain source that most likely is discogenic or from the facet joint and less commonly from the sacroiliac joint.

It can be argued that the Kappa scores could have been higher if all the testing procedures had been standardised. However, the study's intention was to evaluate the reliability as a result of the whole examination as performed in clinical practice, and standardising the examination for this heterogenic group of patients could have influenced the validity.

5. Conclusion

The findings provide evidence that the inter-tester reliability of O'Sullivan's CS is substantial for a range of patients within the NSLBP population in line with previous research. Using a mechanism-based CS has implications in terms of treatment being directed towards identified subgroups. The use of the CS is currently being evaluated in a randomised controlled trial in order to compare the efficacy of different interventions for any given category.

Appendix A. Supplemental material

Supplementary information for this manuscript can be downloaded at doi: 10.1016/j.math.2008.08.003.

References

Bogduk N. The anatomical basis for spinal pain syndromes. Journal of Manipulative and Physiological Therapeutics 1995;18(9):603–5.

Cherkin D, Deyo R, Battie M, Street J, Barlow W. A comparison of physical therapy, chiropractic manipulation, and provision of an educational booklet for the treatment of patients with low back pain. New England Journal of Medicine 1998;339:1021–9.

Dankaerts W, O'Sullivan PB, Burnett AF, Straker LM. Differences in sitting posture are associated with non-specific chronic low back pain disorders when patients are sub-classified. Spine 2006a;31(6):698–704.

Dankaerts W, O'Sullivan PB, Straker LM, Burnett AF, Skouen JS. The inter-examiner reliability of a classification method for non-specific chronic low back pain patients with motor control impairment. Manual Therapy 2006b;11(1):28–39.

Delitto A, Erhard RE, Bowling RW. A treatment-based classification approach to low back syndrome: identifying and staging patients for conservative treatment. Physical Therapy 1995;75(6):470–85.

Donelson R, Silva G, Murphy K. Centralization phenomenon. Its usefulness in evaluating and treating referred pain. Spine 1990;15(3):211–3.

Ferreira ML, Ferreira PH, Latimer J, Herbert RD, Hodges PW, Jennings MD, et al. Comparison of general exercise, motor control exercise and spinal manipulative therapy for chronic low back pain: a randomized trial. Pain 2007;131(1-2):31–7.

Ford J, Story I, McKeenen J. A systematic review on methodology of classification system research for low back pain. Musculoskeletal Physiotherapy Australia 13th Biennial Conference, Sydney, Australia, 2003.

Fritz JM, Delitto A, Erhard RE. Comparison of classification-based physical therapy with therapy based on clinical practice guidelines for patients with acute low back pain – a randomized clinical trial. Spine 2003;28(13):1363–71.

Frymoyer J, Rosen J, Clements J, Pope M. Psychological factors in low back pain disability. Clinical Orthopaedics and Related Research 1985;May;(195):178–84.

Kilpikoski S, Airaksinen O, Kankaanpaa M, Leminien P, Viderman T, Alen M. Inter-examiner reliability of low back pain assessment using the Mckenzie method. Spine 2002;27(8):207–14.

Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33(1):159–74.

Linton SJ. A review of psychological risk factors in back and neck pain. Spine 2000;25(9):1148–56.

Linton SJ, Hallden K. Can we screen for problematic back pain? A screening questionnaire for predicting outcome in acute and subacute back pain. Clinical Journal of Pain 1998;14(3):209–15.

Main C, Watson P. Guarded movements: development of chronicity. Journal of Musculoskeletal Pain 1996;4(4):163–70.

Maluf KS, Sahrmann SA, Van Dillen LR. Use of a classification system to guide nonsurgical management of a patient with chronic low back pain. Physical Therapy 2000;80(11):1097–111.

Please cite this article in press as: Vibe Fersum K et al., Inter-examiner reliability of a classification system for patients with non-specific low back pain, Manual Therapy (2008), doi:10.1016/j.math.2008.08.003

McCarthy C, Arnall F, Strimpakos N, Freemont A, Oldham J. The biopsychosocial classification of non-specific low back pain: a systematic review. Physical Therapy Reviews 2004;9:17–30.
McKenzie R. The lumbar spine, mechanical diagnosis and treatment. Waikanae, New Zealand: Spinal Publications Ltd; 1981.
Nachemson A. Back pain; delimiting the problem in the next millennium. International Journal of Law Psychiatry 1999;22(5-6):473–80.
Nattrass CL, Nitsche JE, Disler PB, Chou MJ, Ooi KT. Lumbar spine range of motion as a measure of physical and functional impairment: an investigation of validity. Clinical Rehabilitation 1999;13:211–8.
O'Sullivan PB. Evaluation of specific stabilizing exercise in the treatment of chronic low back pain with radiologic diagnosis of spondylolysis or spondylolisthesis. Spine 1997;22(24):2959–67.
O'Sullivan PB. Lumbar segmental 'instability': clinical presentation and specific stabilizing exercise management. Manual Therapy 2000;5(1):2–12.
O'Sullivan PB. Lumbar repositioning deficit in a specific low back pain population. Spine 2003;28(10):1074–9.
O'Sullivan P. Diagnosis and classification of chronic low back pain disorders: maladaptive movement and motor control impairments as underlying mechanism. Manual Therapy 2005;10(4):242–55.
O'Sullivan PB, Beales DJ. Diagnosis and classification of pelvic girdle pain disorders – Part 1: a mechanism based approach within a biopsychosocial framework. Manual Therapy 2007a;12(2):86–97.
O'Sullivan PB, Beales DJ. Diagnosis and classification of pelvic girdle pain disorders – Part 2: illustration of the utility of a classification system via case studies. Manual Therapy 2007b;12(2):1–12.
O'Sullivan P, Twomey L, Allison G, Sinclair J, Miller K, Knox J. Altered patterns of abdominal muscle activation in patients with chronic back pain. Australian Journal of Physiotherapy 1997;43(2):91–8.
Padfield B, Chesworth B, Butler R. Use of an outcome measurement system to answer a clinical question: is the Quebec task force classification system useful in an outpatient setting? Physiotherapy Canada 2002:254–60.
Petersen T, Kryger P, Ekdahl C, Olsen S, Jacobsen S. The effect of McKenzie therapy as compared with that of intensive strengthening training for the treatment of patients with subacute or chronic low back pain: a randomized controlled trial. Spine 2002;27(16):1702–9.
Petersen T, Olsen S, Laslett M, Thorsen H, Manniche C, Ekdahl C, et al. Intertester reliability of a new diagnostic classification system for patients with non-specific low back pain. Australian Journal of Physiotherapy 2004;50(2):85–94.
Petersen T, Thorsen H, Manniche C, Ekhdahl C. Classification of non-specific low back pain: a review of the literature on classification systems relevant to physiotherapy. Physical Therapy Reviews 1999;4:265–81.
Sahrmann SA. Diagnosis and treatment of movement impairment syndromes. Mosby: St Louis; 2001.
Skouen JS, Grasdal AL, Haldorsen EM, Ursin H. Relative cost-effectiveness of extensive and light multidisciplinary treatment programs versus treatment as usual for patients with chronic low back pain on long-term sick leave: randomized controlled study. Spine 2002;27(9):901–9.
Spitzer WO. Scientific approach to the assessment and management of activity-related spinal disorders. Spine 1987;7S:S1–55.
Strender LE, Sjoblom A, Sundell K, Ludwig R, Taube A. Interexaminer reliability in physical examination of patients with low back pain. Spine 1997;22(7):814–20.
Stuge B, Laerum E, Kirkesola G, Vollestad N. The efficacy of a treatment program focusing on specific stabilizing exercises for pelvic girdle pain after pregnancy: a randomized controlled trial. Spine 2004;29(4):351–9.
Van Dillen LR, Sahrmann SA, Norton BJ, Caldwell CA, Fleming DA, McDonnell MK, et al. Reliability of physical examination items used for classification of patients with low back pain. Physical Therapy 1998;78(9):979–88.
Van Dillen LR, Sahrmann SA, Norton BJ, Caldwell CA, McDonnell MK, Bloom NJ. Movement system impairment-based categories for low back pain: stage 1 validation. Journal of Orthopaedic and Sports Physical Therapy 2003;33(3):126–42.
Vlaeyen JW, Linton SJ. Fear-avoidance and its consequences in chronic musculoskeletal pain: a state of the art. Pain 2000;85(3):317–32.
Waddell G. The back pain revolution. 2nd ed. Edinburgh: Churchill Livingstone; 2004.

Please cite this article in press as: Vibe Fersum K et al., Inter-examiner reliability of a classification system for patients with non-specific low back pain, Manual Therapy (2008), doi:10.1016/j.math.2008.08.003