Inter-tester Reliability in Classifying Acute and Subacute Low Back Pain Patients into Clinical Subgroups: A Comparison of Specialists and Non-Specialists. A Pilot Study

MARKKU PAATELMA, MSc, PT¹; EIRA KARVONEN, MSc, PT²; ARI HEINONEN, PhD, PT³

Low back pain (LBP) has a lifetime prevalence of approximately 60 to 80%¹ and is recognized internationally as a major health, social, and economic burden². Despite this fact, the current literature has not conclusively demonstrated a single, specific, most effective treatment method for LBP³. Explanations for this failure to identify a single effective treatment likely involve differing etiological factors and variations in the pathoanatomical/pathophysiological/tissue origins of LBP.

Identification of subgroups of LBP has been a focus of major research. Several authors have suggested that because

non-specific low back pain (NSLBP) is a benign problem, emphasis should be on clinical tests and assessments; NSLBP should not be viewed as a homogenous condition, and treatment outcomes can be improved when sub-grouping is used to guide treatment decision-making⁴⁻⁷. In contrast, others have believed that sub-grouping is only one of a number of possible explanations for the manifestations of NSLBP⁸.

Although a number of LBP classification systems have been proposed, such as a pathoanatomical/pathophysiological classification system⁹, the McKenzie classification¹⁰, treatment-based classifica-

tion¹¹, and the movement-impairment classification¹², what is still unclear is which clinical tests between two assessing clinicians are sufficiently reliable to allow subgroup categorization¹³. The reliability and validity of the overall classification systems has been tested¹⁴⁻²¹ and has been reported as moderate or good. Nonetheless, reliability has only been shown to be effective in clinicians who receive advanced training. For example, the McKenzie system has been shown to be reliable in LBP sub-grouping classification only by suitably trained examiners and not by minimally trained or untrained assessors^{16,17}. In contrast, in the movement-impairment classification studies by Dankaerts et al²⁰, patients were independently assessed by two "experts" and 13 physicians or physiotherapists participating in workshops or postgraduate training under supervision of the developer of the classification system. Neither length of training nor differences between experts in either comparison were described.

For comprehensive use in a number of clinical settings by multiple healthcare providers, the accuracy and repeatability of subgroup classification should be similar, not only between specialists but also between specialists and non-specialists working with musculoskeletal disorders. With the largest percentage of physiotherapists²² and physicians²³ using a general pathoanatomical/pathophysiological

ABSTRACT: Many systems have been suggested for classifying low back pain (LBP); the most commonly used among physiotherapists involves a pathoanatomical/pathophysiological tissue classification system. Few studies have examined whether this form of classification of LBP disorders can be performed in a reliable manner between specialists with advanced training, or between specialists with advanced training and non-specialists who lack advanced training. The purpose of this paper was to examine the inter-tester reliability of two specialists, and the ability of a specialist and non-specialist to independently classify patients with LBP, utilizing clinical tests and history-based classification methods after a short educational course on the classification system. Subjects were acute or sub-acute patients with LBP who visited their occupational healthcare or municipal healthcare center. Inter-tester reliability between the specialist and non-specialists was at almost the same level: overall Kappa 0.60 (95%CI; 0.40 to 0.85), overall agreement 70%, as between the two specialists: overall Kappa 0.65 (95%CI; 0.33-0.86), overall agreement 77%. The findings suggest that a short educational course can provide rather reliable examination tools to allow non-specialized physiotherapists to classify patients according to tissue origination.

KEYWORDS: Inter-tester Reliability, Low Back Pain Classification, Orthopedic Manual Therapy

¹OMT Instructor, University of Jyväskylä, Jyväskylä, Finland

²OMT, University of Jyväskylä, Jyväskylä, Finland

³Professor, University of Jyväskylä, Department of Health Sciences, PO Box Fin-40014 Finland

Address all correspondence and requests for reprints to: Markku Paatelma, MSc, PT, markku.paatelma@auron.fi

cal classification system for LBP, this area is appropriate for investigation. Although the validity of the many proposed general pathoanatomical/pathophysiological classification systems has not yet been convincingly confirmed by objective methods, clinical experience suggests that even an idea of the origin of the symptoms may aid the therapist as to the best choice of treatment methods^{24,25}. Therefore, the aim of this pilot study was to evaluate the inter-tester reliability of pathoanatomical/pathophysiological classification within a group of acute and sub-acute non-specific LBP patients. Reliability was examined between the patient findings of 1) general practitioners in primary care physiotherapy compared to findings classified by a specialist in orthopedic manual therapy (OMT); and 2) the findings of a physiotherapist specialist of OMT with multiple years of training compared to a physiotherapist with short-term experience as a specialist in OMT.

Methods

This pilot study was conducted in two parts. In Part 1, we compared the inter-tester reliability of LBP subgroup classification between a specialist in OMT and non-specialists, and in Part 2 between a physiotherapist specialist of OMT with multiple years of training compared to a physiotherapist with short-term experience as a specialist in OMT (Figure 1).

Participants

For the first part of the study, 21 eligible, consecutive, and voluntary patients with LBP with ages ranging from 18 to 56 years were recruited by an invitation letter from four municipal healthcare centers in central Finland. Additionally, 30 patients were recruited from a private occupational healthcare center (Medivire) in the city of Jyväskylä, Finland (Table 1, Figure 1). All patients who had visited their municipal or occupational healthcare center because of low back pain that had lasted less than 3 months were recruited. The inclusion criteria were 18- to 65-year-old individuals with current low back pain with or without

radiating pain to one or both lower legs. The back pain episode could be the first or recurrent with the last episode lasting less than 3 months.

Exclusion criteria were pregnancy, use of psychogenic medications, and diagnosed osteoporosis. The history of present and former LBP episodes and of physical work was asked by a standardized questionnaire. Pain and symptom location were identified with a body chart completed by the patient. All patients showed interest in participating in this study and no one refused. The subjects provided written informed consent before the study and the local ethics committee approved the study protocol.

Procedure

In Part 1, all 21 patients at the healthcare centers were examined by one physiotherapist with 15 years of experience as an OMT and also by four physiotherapists who lacked an OMT specialization, each in his or her own municipal health care center, with the physiotherapists examining six, six, four, and five patients between each. These four non-specialist physiotherapists had a range of 4 to 12 years of clinical experience in physiotherapy and had completed brief post-graduate courses in musculo-skeletal physiotherapy. These physiotherapists were taught by the OMT specialist (EK) to perform the specific clinical tests in order to identify the LBP subgroup classification based on OMT practice. The physiotherapists were blinded to the results of the OMT specialist, who was also blinded to the results of the four non-specialists. A neutral observer supervised the study.

Before initiating the study, all PTs participated in five half-day training sessions, during which they performed the tests used for the classification. This included training on both patients and healthy subjects to standardize the tests by performing the tests according to written instructions and under guidance of the OMT specialist.

In Part 2, the 30 patients at Medivire were examined by two physiotherapists who specialized in OMT, one with 20 years of experience in OMT (MP) and one with 2 years of experience in OMT

(JR). MP had five years of clinical experience in the field of musculoskeletal physiotherapy prior to specialization in OMT, and JR had six. Examiners were blinded to each other's results and the tests were supervised by a neutral observer. OMT specialization in Finland requires 3.5 years of training and includes a post-graduate examination supervised by the Finnish Association of Physiotherapists.

In Parts 1 and 2 of the study, clinical assessments included 50 different tests (Table 2). Each patient assessment session lasted 30 minutes and the data were collected on a lumbar spine-assessment form. This assessment comprised an LBP history, observation of the posture of the low back and lower extremities, function of the lower back and lower extremities, stability tests for lumbar spine and pelvic girdle, specific pain provocation and alleviation tests, mobility tests of the lower back and sacroiliac joints, neurological and neurodynamic tests, and tests for muscle tightness. In all the tests, the decision was binary: the test was either negative (normal finding) or positive (pathological finding). Based on the tests and a defined clinical reasoning process, the patients were classified into one of five mutually exclusive clinical LBP pathoanatomical/pathophysiological subgroups:

1. Discogenic pain
2. Clinical lumbar instability
3. Clinical lumbar spinal stenosis
4. Segmental dysfunction/facet pain
5. Sacroiliac joint pain/dysfunction

One patient with a post-operative spine could not be classified by either of two examiners into any of the subgroups and was therefore excluded from statistical analysis. Thus, the number of analyzed patient groups was 20 in Part 1 and 30 in Part 2. The inter- and intra-tester reliability, sensitivity, and specificity of the tests used were evaluated before the testing²⁶.

Subgroup Classification

The clinical reasoning process by which the physiotherapist reached a classification involved a dedicated deductive

FIGURE 1. Flow chart of subjects during study.

TABLE 1. Characteristics of subjects in subgroup classification between a specialist in OMT and non-specialists, and between two specialists in OMT (an expert vs. a novice).

	OMT specialist vs. non-specialists ¹ (N=20)	OMT expert vs. OMT novice ² (N=30)
Age (years, mean, SD)	40.0 (11.5)	37.9 (4.5)
Gender (female/male)	65/35	74/26
History of present LBP episode		
Acute: < 6 weeks	35	27
Subacute: > 6 weeks <12 weeks	65	73
Chronic: >12 weeks	0	0
On sick-leave because of LBP	0	0
Pain and symptom location		
VAS from 0 to 10 over last 24 hrs (mean, SD):	3 (3.5)	3 (2.9)
Signs of pain drawing in low back	42	45
Signs of pain drawing in lower leg	58	55
Physical work		
Light (%)	65	67
Heavy (%)	35	33

¹Municipal Health Care (Part 1)

²Occupational Health Care (Part 2)

Note: Values are percentages unless stated otherwise.

process. Briefly, *discogenic pain without nerve root irritation* was the diagnosis when a patient's pain (local or referred) could be provoked in modified slump test and when movement into extension was less painful or alleviated the same pain (centralized). *Discogenic pain with nerve root irritation* was recorded when radiating pain was provoked by nerve tension tests (SLR) and by other neurodynamic tests.

The construct for *clinical lumbar instability* involved assessment of three interdependent components: the passive, the active, and the neuromuscular subsystem²⁷. Clinical lumbar instability was recorded when the patient reported low back pain or fatigue or both during prolonged sitting/standing/lying down, and when pain during extension was relieved and movement increased with traction. In addition, the classification was made if there were difficulties in a one-leg stance or active straight-leg raise

TABLE 2. Clinical Tests and Findings used in Classification.

Clinical Assessment	Finding
Functions	Normal walking, heel walking, and walking on toes, undressing, squat and rise
Inspection	posture of spine in standing, knees, and feet length difference of lower limbs
Mobility	lumbar spine flexion, extension, lateral flexion (right and left) specific Posterior-Anterior mobility of T12-S1 hip rotation left and right
Pain Provocation	lumbar spine extension with traction from max extension left to max flexion right from max extension right to max flexion left posterior pelvic pain provocation right and left interspinal ligament provocation Kibler's skin rolling sacroiliac joint provocation right and left L4, L5 rotation provocation right and left
Muscle tightness	hamstrings right and left piriformis right and left gluteus med/min right and left iliopsoas right and left
Stability	one-leg standing right and left active SLR right and left isometric lumbar extension transverse abdominis activity
Neurology and neurodynamics	SLUMP in sitting SLR right and left patella reflex right and left Achilles reflex right and left Ely's test right and left

(ASLR) or both, or inability to activate either transverse abdominus or lumbar multifidi combined with local interspinal pain, or a combination of these problems.

Clinical central spinal stenosis was recorded when the patient reported a clear pattern of intermittent claudication provoked by extension, which was relieved by sitting or flexed spinal posture. Symptoms and signs could be combined with tightness of hip flexors or positive sciatic or a femoral nerve tension test or several. *Clinical lateral spinal stenosis* was recorded when the patient reported radiating pain with nerve tension tests and during extension/lateral flexion toward the symptomatic side or during transverse process provocation (passive foramina approach), or both.

Segmental dysfunction/facet pain was recorded when pain and movement restrictions were identified during physiological movements in standing and painful hypomobility while lying prone, whereas sacroiliac joint pain/dysfunction was recorded if the patient's pain was provoked while standing on one leg and relieved with a sacroiliac joint belt, or provoked with sacral thrust and/or during posterior pelvic pain provocation (PPPP) or both, or if pain and difficulties occurred during an ASLR.

Examination Techniques

To improve reliability in the examination, all tests were standardized with operational definitions for their use and interpretation, and was taught compre-

hensively and trained by an OMT specialist prior to the study. In our study, the examination techniques of concern were based on provocation and alleviation techniques²⁸, sacroiliac joint provocation²⁹⁻³², and neurodynamic tests^{33,34}.

Statistical Analysis

Statistical analyses were performed with SPSS software, version 14.0. Percentage agreement and the kappa statistic served to test inter-tester agreement in the choice of subgroup classification. Clinical agreement was recorded when the two subgroup classifications were exactly the same.

Overall inter-tester agreement and overall kappa coefficient were calculated first using a 2x2 contingency table. Then the prevalence of positive observations and percentages of agreement for all categories and kappa in discogenic pain and clinical instability were calculated.

The kappa statistic estimates the degree of agreement corrected for chance agreement³⁵. There is general agreement that for physical therapists, kappa a preferred statistic for estimating the accuracy of nominal and ordinal data in clinical research³⁶. The percentage agreement does not take into account agreement due solely to chance³⁷.

The classification system proposed by Landis and Koch³⁸ allowed determination of the level of kappa as follows: poor: smaller than zero; slight: zero to 0.20; fair: 0.21 to 0.40; moderate: 0.41 to 0.60; substantial: 0.61 to 0.80; almost perfect: 0.81 to 1.00. Clinical relevance was considered in this study as a kappa of 0.41³⁸ and the percentage of agreement at 70%³⁹.

Results

For Part 1, which involved a municipal healthcare center, LBP subgroup prevalence was as follows: clinical instability and discogenic pain were the most common (35% and 30%), followed by segmental dysfunction/facet pain (18%), sacroiliac joint pain/dysfunction (10%), and clinical spinal stenosis (7%). In Part 2, which involved an occupational healthcare setting, clinical instability and discogenic pain were also the most

frequent subgroups (43% and 37%), followed by segmental dysfunction/facet pain (7%), sacroiliac joint pain/dysfunction (7%), and clinical spinal stenosis (6%) (Tables 3 and 4).

Percentage of agreement ranged from 75% to 100% between the experienced physiotherapist with OMT specialization and the four physiotherapists without OMT specialization. Overall inter-tester agreement was 70% and the overall kappa coefficient was 0.60 (95% CI; 0.40 to 0.85) (Table 3). The prevalence of positive observations and percentages of agreement for all categories and the kappa in discogenic pain and clinical instability are presented in Table 5.

Overall agreement between these OMT specialists for Part 2 was 77% with an overall kappa of 0.65 (95% CI; 0.33 to 0.86) (Table 4). Table 6 gives the prevalence of positive observations and percentages of agreement for all the categories and the kappa in discogenic pain and clinical instability. Percentages of agreement ranged from 75% to 100% (Table 6).

Discussion

This pilot study demonstrated that inter-tester reliability of categorization of LBP subgroups between an experienced

physiotherapist in OMT and four physiotherapists without OMT specialization and then between two clinicians with OMT training, one experienced and one inexperienced, was acceptable. Comparison with other inter-tester reliability studies is difficult because in the majority of comparative studies, the study group consisted of chronic LBP patients^{13,14,19} or the length of LBP was not reported^{16-18,20}. However, at least two studies support our findings that a short training period in the workplace is effective. Fritz et al⁴⁰ found acceptable overall agreement on classification of a decision-making algorithm using physical therapists with varying levels of experience, identifying no significant differences based on level of experience. When clinicians were newly trained in a classification system for acute LBP during a one-day course, that classification showed moderate reliability⁴¹.

Worth noting is that the reliability of discogenic and sacroiliac joint pain/dysfunction clinical tests varies from fair to good, but the reliability of tests for segmental dysfunction/facet pain is poor⁴². This may also explain, at least partly, the low prevalence of this subgroup. The tests for clinical lumbar instability also vary from poor to good^{43,44}. At present, there are no reliability studies that exist for detecting clinical cen-

tral or lateral stenosis. Only a self-reported history questionnaire has proven successful as a diagnostic tool for lumbar spinal stenosis⁴⁵.

In the present pilot study, the number of patients was small, which thus reduces the generalizability of the findings. Prevalence of some of the subgroups was also low, for example, sacroiliac joint pain/dysfunction and clinical spinal stenosis, and thus calculation of kappa was not possible. Despite this limitation, a notable strength was that the patients were recruited from routine referrals, and because the care was provided for free, cost containment did not limit physiotherapy examination. Although the time-limited examination was 30 minutes, some LBP patients may require longer assessment or even several assessments, with their responses to specific interventions and applications of home-exercises contributing to the final subgroup categorization⁴⁶.

Concepts of diagnosis and classification of LBP have a long history in medicine, whereas formal schemes of diagnostic classification in physical therapy are relatively new⁴⁷. Because of this short history, some questions concerning the validity of the LBP classification are still prevalent. In addition, the few high-quality studies that exist demonstrate either conflicting evidence or

TABLE 3. Cross-tabulation of diagnostic groups between an experienced physiotherapist (15 years) in OMT shown vertically, and horizontally figures by four physiotherapists without OMT specialization. Cases (%). Number of subjects 20.

Experienced OMT-physiotherapist diagnosis	Four physiotherapists without OMT specialization					Total
	Discogenic pain	Clinical instability	Clinical spinal/lateral stenosis	Segmental dysfunction/facet pain	Sacroiliac & pelvic pain	
Discogenic pain	5 (25)	1 (5)	0 (0)	1 (5)	0 (0.0)	7 (35)
Clinical instability	0 (0)	5 (24)	0 (0)	1 (5)	0 (0.0)	6 (30)
Clinical spinal/lateral stenosis	0 (0)	0 (0)	1 (5)	1 (5)	0 (0.0)	2 (10)
Segmental dysfunction/facet pain	0 (0)	2 (10)	0 (0)	1 (5)	0 (0.0)	3 (15)
Sacroiliac & pelvic pain	0 (0)	0 (0)	0 (0)	0 (0)	2 (10)	2 (7)
Total	5 (25)	8 (40)	1 (5)	4 (20)	2 (10)	20 (100)
Agreement %	70					
Kappa (95% CI)	0.60 (0.40 to 0.85)					

Bold represents counts for agreement.

TABLE 4. Cross-tabulation of diagnostic groups between two physiotherapists specialized in OMT: one experienced (20 years) in OMT shown vertically, and a physiotherapist with less experience (shown horizontally) (2 years). Cases (%). Number of subjects 30.

Experienced OMT- physiotherapist	Novice physiotherapist in OMT					Total
	Discogenic pain	Clinical instability	Clinical spinal/ lateral stenosis	Segmental/ dysfunction facet pain	Sacroiliac & pelvic pain	
Discogenic pain	9 (30)	0 (0)	0 (0)	1 (3)	0 (0)	10 (33)
Clinical instability	3 (10)	10 (33)	0 (0)	0 (0)	0 (0)	13 (43)
Clinical spinal/lateral stenosis	0 (0)	0 (0)	2 (7)	0 (0)	0 (0)	2 (7)
Segmental dysfunction /facet pain	0 (0)	3 (10)	0 (0)	0 (0)	0 (0)	3 (10)
Sacroiliac & pelvic pain	0 (0)	0 (0)	0 (0)	0 (0)	2 (7)	2 (7)
Total	12 (40)	13 (43)	2 (7)	1 (3)	2 (7)	30 (100)
Agreement %	77					
Kappa (95 % CI)	0.65 (0.33 to 0.86)					

Bold represents counts for agreement.

TABLE 5. Inter-examiner reliability in assessment of low back pain subgroup classification between a physiotherapist specialized in OMT and by four physiotherapists without OMT specialization. Number of subjects 20.

LBP classification	Number of positive observations by examiner		Agree ¹	Disagree ²	Agreement %	Kappa (95 % CI)
	Experienced Examiner	Inexperienced Examiners				
Discogenic pain	7	5	18	2	90	0.76 (0.35 to 1.00)
Clinical instability	6	8	16	4	80	0.57 (0.14 to 0.90)
Clinical spinal/ lateral stenosis	2	1	19	1	95	Φ
Segmental dysfunction/ facet pain	3	4	15	5	75	Φ
Sacroiliac & pelvic pain	2	2	20	0	100	Φ

¹Examiners agreed that a given number of patients had the classification.

²Examiners disagreed whether patients had the classification.

Φ=calculation of Kappa impossible, due to low number of subjects in subgroups.

TABLE 6. Inter-examiner reliability in assessment of low back pain subgroup classification between two physiotherapists specialized in OMT (expert vs. novice). Number of subjects 30.

LBP classification	Number of positive observations by examiner		Agree ¹	Disagree ²	Agreement %	Kappa (95 % CI)
	Expert	Novice				
Discogenic pain	10	12	26	4	87	0.71 (0.40 to 0.93)
Clinical instability	13	13	24	6	80	0.59 (0.28 to 0.86)
Clinical spinal/ lateral stenosis	2	2	30	0	100	Φ
Segmental dysfunction/ facet pain	3	1	26	4	75	Φ
Sacroiliac & pelvic pain	2	2	30	0	100	Φ

¹Examiners agreed that a given number of patients had the classification.

²Examiners disagreed whether patients had the classification.

Φ=calculation of Kappa impossible, due to low number of subjects in subgroups.

only moderate evidence of LBP classification reliability⁴⁸.

In the present study, the inter-tester reliability was good, particularly in the clinical instability and discogenic pain subgroups in which most of our patients were classified. Although instability is fraught with conceptual and terminological difficulties without widely used accepted clinical diagnostic criteria, it is the most common subgroup in clinical physiotherapy praxis⁴⁹. Discogenic pain was also common, in accordance with many findings⁵⁰⁻⁵².

In addition to the low prevalence of given subgroups, the low mean age of the patients may explain why only 7% were classified into a clinical spinal stenosis subgroup, a finding in line with the study of Leinonen et al⁵³. Furthermore, although research has shown that pain arising from the facet joint is difficult to characterize by clinical examination variables^{9,54}, we concluded that 11% of patients were classified into this subgroup. This agrees with a primary subgroup classification, that is, with sacroiliac joint pain/dysfunction appearing in only 8%. In the literature, prevalence of sacroiliac joint pain varies but is routinely considered to be around 9% to 20%^{27,30,55}.

The type of pathoanatomical/pathophysiological/tissue origin classification might be very useful in cases of acute and subacute pain, because the examination is non-invasive and widely available. If no major trauma or suspected malignancy exists, invasive diagnostic techniques are not recommended in acute and subacute cases, techniques that may differ greatly among chronic cases. Classification in chronic LBP is often even more complicated, with its high levels of distress and disability. Psychological factors consistent with fear-avoidance models are associated with the development of chronic LBP. In addition, supplementing behavioral treatment options by treatment-based classification (TBC) physical therapy intervention for acute and sub-acute LBP patients was shown to be ineffective for improving important outcomes related to development of chronic LBP⁵⁶.

European clinical guidelines for LBP recommend early referral of appro-

priate patients to health services such as physiotherapy. Casserley-Feeny et al showed significantly higher percentages of acute LBP patients in the private setting than in the public setting; they also found longer wait times and a higher number and longer duration of physiotherapy treatments in the public setting, suggesting the need to develop primary healthcare in the aim of preventing acute LBP from becoming chronic⁵⁷. Our classification was used previously by Paatelma et al in a randomized controlled trial that compared OMT and the McKenzie Method with advice only for LBP treatment⁵⁸, and included patients from a full spectrum of chronicity. Whether this kind of classification used in our study would improve the efficacy of LBP pain treatment compared to a classification based on time-duration of pain⁵⁹⁻⁶¹ or diagnostic triage⁶² is unknown. Still, although empirical clinical evidence has shown the efficacy of physiotherapy for acute first-episode low back disorders, it has not shown the same efficacy in patients with chronic low back disorders⁶³. Whether classification might enhance the efficacy in all subgroups is also unknown.

Conclusions

The results of this study suggest that inter-tester reliability in classifying patients with LBP into the clinical subgroups seems to be moderate or high regardless of experience level of the OMT-specialist physiotherapist or whether the physiotherapist had a background in OMT principles. This clinically relevant and clearly defined pain pattern system uses key elements of the history and examination to classify patients with low back pain. However, larger trials using those tests in every subgroup, including those which have high odd ratios, are necessary before we can make general statements about the reliability of subgroup classification in the early stage of low back pain.

REFERENCES

1. Torstensen TA, Ljunggren AE, Meen HD, Odland E, Mowinckel P, Geijerstam S. Efficiency and costs of medical ex-

ercise therapy, conventional physiotherapy, and self-exercise in patients with chronic low back pain: A pragmatic, randomized, single-blinded, controlled trial with 1-year follow-up. *Spine* 1998;23:2616-2624.

2. Hemmälä HM. Quality of life and cost of care of back pain patients in Finnish general practice. *Spine* 2002;27:647-653.
3. Cohen SP, Argoff CE, Carragee EJ. Management of low back pain. *BMJ* 2008;337:a2718 DOI: 10.1136/bmj.a2718.
4. McKenzie R, May S. *Mechanical Diagnosis and Therapy: The Lumbar Spine*. Waikanae, New Zealand: Spinal Publications, 2003.
5. O'Sullivan P. Diagnosis and classification of chronic low back pain disorders: Maladaptive movement and motor control impairments as underlying mechanism. *Man Ther* 2005;10:242-255.
6. Fritz JM, Lindsay W, Matheson JW, et al. Is there a subgroup of patients with low back pain likely to benefit from mechanical traction? Results of a randomized clinical trial and subgrouping analysis. *Spine* 2007;32:E793-E800.
7. Brennan GP, Fritz JM, Hunter SJ, Thackeray A, Delitto A, Erhard RE. Identifying subgroups of patients with acute/subacute "nonspecific" low back pain: Results of a randomized clinical trial. *Spine* 2006;31:623-631.
8. Wand BM, O'Connell NE. Chronic nonspecific low back pain: Sub-groups or a single mechanism? *BMC Musculoskelet Disord* 2008;9:11.
9. Coste J, Paolaggi JB, Spira A. Classification of nonspecific low back pain. II. Clinical diversity of organic forms. *Spine*. 1992;17:1038-1042.
10. McKenzie R. *The Lumbar Spine: Mechanical Diagnosis and Therapy*. Waikanae, New Zealand: Spinal Publications, 1981.
11. Delitto A, Erhard RE, Bowling RW. A treatment-based classification approach to low-back syndrome: Identifying and staging patients for conservative treatment. *Phys Ther* 1995;75:470-485.
12. Sahrman S. *Diagnosis and Treatment of Movement Impairment Syndromes*. St. Louis, MO: Mosby, 2002.
13. McCarthy CJ, Gittins M, Roberts C, Oldham JA. The reliability of the clinical tests and questions recommended in international guidelines for low back pain. *Spine* 2007;32:921-926.
14. Petersen T, Olsen S, Laslett M, et al. Inter-tester reliability of a new diagnostic classifi-

- cation system for patients with non-specific low back pain. *Aus J Physiother* 2004;50:85–94.
15. Laslett M, McDonald B, Tropp H, Aprill C, Oberg B. Agreement between diagnoses reached by clinical examination and available reference standards: A prospective study of 216 patients with lumbopelvic pain. *BMC Musculoskelet Disord* 2005;6:28.
16. George S, Delitto A. Clinical examination variables discriminate among treatment-based classification groups: A study of construct validity in patients with acute low back pain. *Phys Ther* 2005;85:306–314.
17. Kilpikoski S, Airaksinen O, Kankaanpää M, Leminen P, Videman T, Alen M. Interexaminer reliability of low back pain assessment using the McKenzie method. *Spine* 2002;8:E207–E214.
18. Clare HA, Adams R, Maher CG. Reliability of McKenzie classification of patients with cervical or lumbar pain. *J Manipulative Physiol Ther* 2005;28:122–127.
19. Van Dillen L, Sahrman S, Norton B, Caldwell C, McDonnell M, Bloom N. Movement system impairment-based categories for low back pain: Stage 1 validation. *J Orthop Sports Phys Ther* 2003;33:126–142.
20. Dankaerts W, O'Sullivan P, Straker L, Burnett A, Skouen J. The inter-examiner reliability of a classification method for non-specific chronic low back pain patients with motor control impairment. *Man Ther* 2006;11:28–39.
21. Luomajoki H, Kool J, de Bruin ED, Airaksinen O. Reliability of movement control tests in the lumbar spine. *BMC Musculoskelet Disord* 2007;8:90.
22. Spoto MM, Collins J. Physiotherapy diagnosis in clinical practice: A survey of orthopaedic certified specialists in the USA. *Physiother Res Int* 2008;13:31–41.
23. Billis EV, McCarthy CJ, Oldham JA. Subclassification of low back pain: A cross-country comparison. *Eur Spine J* 2007;16:865–879.
24. Brennan GP, Fritz JM, Hunter SJ, Thackeray A, Delitto A, Erhard RE. Identifying subgroups of patients with acute/subacute “nonspecific” low back pain: Results of a randomized clinical trial. *Spine* 2006;31:623–631.
25. Rothwell PM. Treating individuals 2. Subgroup analysis in randomised controlled trials: Importance, indications, and interpretation. *Lancet* 2005;365:176–186.
26. Paatelma M, Karvonen E, Heiskanen J. How do clinical test results differentiate chronic and subacute low back pain patients from “non-patients”? *J Man Manip Ther* 2009;17:11–19.
27. Demoulin C, Distrée V, Tomasella M, Crie-laard JM, Vanderthommen M. Lumbar functional instability: A critical appraisal of the literature. *Ann Readapt Med Phys* 2007;50:677–684, 669–676.
28. Kaltenborn F, Evjenth O, Kaltenborn T, Morgan D, Vollowitz E. *The Spine. Manual Mobilization of the Joints*. Vol II. Oslo, Norway: Norli, 2003.
29. Laslett M, Williams M. The reliability of selected pain provocation tests for sacroiliac joint pathology. *Spine* 1994;19:1243–1249.
30. Laslett M. Evidence-based diagnosis and treatment of the painful sacroiliac joint. *J Man Manip Ther* 2008;16:142–152.
31. Mens J, Vleeming A, Snijders C, Koes B, Stam H. Validity of active straight leg raise test for measuring disease severity in patients with posterior pelvic pain after pregnancy. *Spine* 2002;27:196–200.
32. van der Wurff P, Buijs E, Groen G. A multitest regimen of pain provocation tests as an aid to reduce unnecessary minimally invasive sacroiliac joint procedures. *Arch Phys Med Rehabil* 2006;87:10–14.
33. Butler D. *Mobilisation of the Nervous System*. 1st ed. Edinburgh, UK: Churchill Livingstone, 1991.
34. Shacklock M. *Clinical Neurodynamics: A New System of Musculoskeletal Treatment*. 1st ed. Edinburgh, UK: Elsevier, 2005.
35. Kramer M, Feinstein A. Clinical biostatistics LIV: The biostatistics of concordance. *Clin Pharmacol Ther* 1981;29:111–123.
36. Haley S, Osberg J. Kappa coefficient calculation using multiple ratings per subject: A special communication. *Phys Ther* 1989;69:970–974.
37. Hunt R. Percentage agreement, Pearson's correlation, and Kappa as measures of inter-examiner reliability. *J Dent Res* 1986;65:128–130.
38. Landis J, Koch G. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–174.
39. Potter N, Rothstein J. Intertester reliability for selected clinical tests of the sacroiliac joint. *Phys Ther* 1985;65:1671–1675.
40. Fritz J, Brennan G, Clifford S, Hunter S, Thackeray A. An examination of the reliability of a classification algorithm for subgrouping patients with low back pain. *Spine* 2006;31:77–82.
41. Heiss D, Fitch D, Fritz J, Sanchez W, Roberts K, Buford J. The interrater reliability among physical therapists newly trained in a classification system for acute low back pain. *J Orthop Sports Phys Ther* 2004;34:430–439.
42. Hancock MJ, Maher CG, Latimer J, et al. Systematic review of tests to identify the disc, SIJ or facet joint as the source of low back pain. *Eur Spine J* 2007;16:1539–1550.
43. Abbott JH, McCane B, Herbison P, Moginie G, Chapple C, Hogarty T. Lumbar segmental instability: A criterion-related validity study of manual therapy assessment. *BMC Musculoskelet Disord* 2005;6:56.
44. Kasai Y, Morishita K, Kawakita E, Kondo T, Uchida A. A new evaluation method for lumbar spinal instability: Passive lumbar extension test. *Phys Ther* 2006;86:1661–1667.
45. Konno S, Kikuchi S, Tanaka Y, et al. A diagnostic support tool for lumbar spinal stenosis: A self-administered, self-reported history questionnaire. *BMC Musculoskelet Disord* 2007;30;8:102.
46. Werneke M, Hart D. Categorizing patients with occupational low back pain by use of the Quebec Task Force Classification system versus pain pattern classification procedures: Discriminant and predictive validity. *Phys Ther* 2004;84:243–254.
47. Zimny N. Diagnostic classification and orthopaedic physical therapy practice: What we can learn from medicine. *J Orthop Sports Phys Ther* 2004;34:105–109.
48. 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 Physiot* 2006;52:91–102.
49. Barr K, Griggs M, Cadby T. Lumbar stabilization: Core concepts and current literature, Part 1. *Am J Phys Med Rehabil* 2005;84:473–480.
50. Derby R, Lee S, Kim B, Chen Y, Aprill C, Bogduk N. Pressure-controlled lumbar discography in volunteers without low back symptoms. *Pain Med* 2005;6:213–221.
51. Schwarzer A, Aprill C, Derby R, Fortin J, Kine G, Bogduk N. The prevalence and clinical features of internal disc disruption in patients with chronic low back pain. *Spine* 1995;20:1878–1883.
52. Freemont AJ, Peacock TE, Goupille P, Hoyland JA, O'Brien J, Jayson MI. Nerve ingrowth into diseased intervertebral disc in chronic back pain. *Lancet* 1997;350:178–181.
53. Leinonen V, Maatta S, Taimela S, et al. Im-

- paired lumbar movement perception in association with postural stability and motor- and somatosensory-evoked potentials in lumbar spinal stenosis. *Spine* 2002;27:975–983.
54. Manchikanti L, Manchukonda R, Pampati V, Damron K, McManus C. Prevalence of facet joint pain in chronic low back pain in postsurgical patients by controlled comparative local anesthetic blocks. *Arch Phys Med Rehabil* 2007;88:449–455.
55. Riddle D, Freburger J. Evaluation of the presence of sacroiliac joint region dysfunction using a combination tests: A multicenter intertester reliability study. *Phys Ther* 2002;82:772–781.
56. George SZ, Zeppieri G Jr, Cere AL, et al. A randomized trial of behavioral physical therapy interventions for acute and subacute low back pain. *Pain* 2008;140:145–157.
57. Casserley-Feeny S, Bury G, Daly L, Hurley D. Physiotherapy for low back pain: Differences between public and private healthcare sectors in Ireland. A retrospective survey. *Man Ther* 2008;13:441–449.
58. Paatelma M, Kilpikoski S, Simonen R, Heinenen A, Alen M, Videman T. Orthopaedic manual therapy, McKenzie method or advice only for low back pain in working adults: A randomized controlled trial with one year follow-up. *J Rehabil Med* 2008;40:858–863.
59. Spitzer WO. Low back pain in the workplace: Attainable benefits not attained. *Br J Ind Med* 1993;50:385–388.
60. Merskey H, Bogduk N. *Classification of Chronic Pain*. Seattle, WA: IASP Press, 1994.
61. Bogduk N, McGuirk B. *Medical Management of Acute and Chronic Low Back Pain: An Evidence-based Approach*. Edinburgh, UK: Elsevier, 2002.
62. Waddell G. Subgroups within “nonspecific” low back pain. *J Rheumatol* 2005;32:395–396.
63. Hides J, Jull G, Richardson C. Long-term effects of specific stabilizing exercises for first-episode low back pain. *Spine* 2001;26:E243–248.