Inter-tester reliability of a new diagnostic classification system for patients with non-specific low back pain

Tom Petersen[1,5], Steen Olsen[1], Mark Laslett[2], Hanne Thorsen[3], Claus Manniche[4], Charlotte Ekdahl[5], Soren Jacobsen[6]

[1]*Back Center of Copenhagen, Denmark* [2]*University of Linköping, Sweden* [3]*University of Copenhagen, Denmark* [4]*Ringe Hospital and Odense University, Denmark* [5]*Lund University, Sweden* [6]*Bispebjerg University Hospital, Denmark*

Most patients referred to physiotherapy with low back pain are without a precise medical diagnosis. Identification of subgroups of non-specific low back pain patients may improve clinical outcomes and research efficiency. A pathoanatomic classification system has been developed, classifying patients with non-specific low back pain into 12 different syndromes and three subcategories based on history and physical examination. The purpose of this study was to estimate the inter-tester reliability of clinical tests used as criteria for classifying patients. Ninety patients with chronic low back pain were each examined by two physiotherapists. A total of four physiotherapists conducted the assessments. Examination findings were recorded independently by the two examiners. Percentage of agreement and kappa coefficients were calculated for each category. The overall rate of agreement was 72% and the kappa coefficient was 0.62 for the mutually exclusive syndromes in the classification system. Agreement rates for each of the syndromes ranged from 74% to 100% and kappa coefficients ranged from 0.44 to 1.00. The findings suggest the inter-tester reliability of the system is acceptable. The relatively modest level of total agreement (39%) for the system as a whole might indicate that the utility of the system for general screening purposes is limited, compared with the utility in identification of particular syndromes. Due to low prevalence of positive findings in some of the syndromes, future work should focus on testing reliability on a larger sample of patients, and testing of validity and feasibility of the system. **[Petersen T, Olsen S, Laslett M, Thorsen H, Manniche C, Ekdahl C and Jacobsen S (2004): Inter-tester reliability of a new diagnostic classification system for patients with non-specific low back pain.** *Australian Journal of Physiotherapy* **50: 85–91]**

Key words: Low Back Pain, Classification, Physical Examination, Reliability of Results, Reproducibility of Results

Introduction

International guidelines for the management of low back pain (LBP) recommend an initial classification process, a diagnostic triage that differentiates between possible serious spinal pathology, nerve root problems, and non-specific LBP (Bigos et al 1994, Clinical Standards Advisory Group 1994, Koes et al 2001). It is estimated that 85% of the LBP patients seen in primary care have non-specific LBP (Deyo and Phillips 1996). Thus, the label of non-specific LBP contains little specific therapeutic information, and refers to a large heterogeneous group of patients suffering from a variety of pathological or pathophysiological conditions.

Discussion about diagnostic subclassification of non-specific LBP arises from the assumption that this large heterogeneous group of patients would be treated more effectively if the patients were assigned to more homogeneous subgroups on the basis of valid criteria (Borkan et al 1998, Bouter et al 1998, Leboeuf-Yde et al 1997, Spitzer 1987).

Several classification systems have been proposed for subdividing non-specific LBP patients by means of clinical examination (Petersen et al 1999). In physiotherapy, three of those are of particular interest as they are all sufficiently detailed to have implications for choice of treatment for the individual patient, and have been tested for reliability and validity (Delitto et al 1995, Maluf et al 2000, McKenzie 1981).

The classification system proposed by McKenzie (1981) is based on information from history taking and symptom response to patient- or therapist-generated movements or positions. It has been reported as the most commonly used system by physiotherapists (Battie et al 1994, Foster et al 1999, Gracey et al 2002). When applied by trained examiners, categorisation of the main syndromes (derangement, dysfunction, and postural) has substantial inter-tester agreement with kappa coefficients ranging from 0.6 to 0.7 (Kilpikoski et al 2002, Razmjou 2000). Randomised controlled trials examining validity of the McKenzie system, i.e. whether categorisation enables selection of more effective treatment, have shown mixed results when the outcomes have been compared with that of other treatments (Cherkin et al 1998, Gillan et al 1998, Nwuga and Nwuga 1985, Petersen et al 2002, Stankovic and Johnell 1990 and 1995).

Delitto et al (1995) have developed a classification system for categorisation of patients with acute LBP. The system classifies patients using information from history taking and clinical examination. It has moderate inter-tester agreement (kappa coefficient of 0.56) (Fritz and George 2000). Randomised controlled trials comparing outcomes of a classification-based treatment with other types of treatment in patients with acute non-specific LBP support the validity of some of the six categories of the system (Delitto et al 1993, Erhard et al 1994, Fritz et al 2003).

Sahrman and co-workers have developed a classification system comprising five categories based on assessment of muscular stability, alignment, asymmetry, and flexibility of the lumbar spine, pelvis, and hip joints (Maluf et al 2000). Reliability of the individual tests used in criteria for classification has been shown to vary from fair to almost perfect (kappa coefficients ranging from 0.21 to 1.00) (Van Dillen et al 1998). Percentage of agreement in classifying

patients into the five categories has been claimed to range from 95% to 100% (Van Dillen et al 2003). However, no data or kappa values were reported. The use of the system has been illustrated by a case report (Maluf et al 2000) and preliminary evidence has been published showing that modification of movements and alignments of the spine during testing resulted in a decrease in symptoms for a majority of the LBP patients when measured immediately after the intervention (Van Dillen et al 2003).

Although data on reliability and validity have been published supporting the utility of some of these systems, none has proven its superiority over others by identifying subgroups of patients with better outcomes in response to a specific treatment. Therefore, the existing classification systems do not eliminate the need for development of alternatives.

Recently, the present authors have proposed a new classification system that has a pathoanatomic orientation (Petersen et al 2003). This system aims to overcome some fundamental problems with the existing treatment-oriented systems. These issues are discussed briefly below.

Labels and criteria used for classification in existing systems differ according to the treatment methods preferred by the developers and the result is a variety of competing classification systems. For example, it appears that a LBP patient who responds with an increase in pain intensity following lumbar flexion movements and abolition of pain following extension movements would be classified as having a 'posterior derangement syndrome' in the system proposed by McKenzie (1981), an 'extension syndrome' in the system proposed by Delitto et al (1995), and a 'flexion category' in the system proposed by Sahrman and co-workers (Maluf et al 2000). It has been pointed out that various practitioners have different, but equally acceptable approaches to the management of a particular treatment-oriented category (Binkley et al 1993). In our opinion, rather than making the diagnostic system fit the treatment system preferred by the developers, it should be the other way around. Once a generally accepted diagnostic classification system has been developed, the results of research should determine the most effective treatments for particular categories of patients. Thus, the new classification system is primarily developed for use in clinical research (e.g. outcome studies investigating efficacy of different treatments to homogeneous subgroups of patients with non-specific LBP). When criteria for categorising patients in clinical research are based on physical examination findings, clinicians can recognise patients in the different groupings and implement research results into daily practice.

The classification system may also be useful to researchers within the field of LBP and to health professionals engaged in diagnosis and treatment of patients with LBP in primary care.

The greater part of the new classification system consists of pathoanatomically labelled syndromes assumed to refer to a specific pathological condition. The 12 syndromes and three sub-syndromes included in this system and criteria for categorisation are summarised in the Appendix.

In some syndromes, there is evidence to support the assumption of a specific anatomical source of symptoms, while in others this assumption is hypothesised. The development phase followed three steps (see Box 1). In step one, pathoanatomic categories that could be derived from

---

Box 1. The three steps of the development phase.

Step 1

Disc syndrome (sensitivity 0.94, specificity 0.52) (Donelson et al 1997)
- reducible disc syndrome
- irreducible disc syndrome
- non-mechanical disc syndrome

Nerve root compression syndrome (sensitivity 0.96-0.98, specificity 0.63 to 0.99) (Deyo et al 1994)

Spinal stenosis syndrome (sensitivity 0.88, specificity 0.93) (Fritz et al 1998)

Zygapophysial joint syndrome (sensitivity 0.92, specificity 0.80) (Revel et al 1998)

Sacroiliac joint syndrome (sensitivity 0.91, specificity 0.83) (Laslett et al 2003)

Step 2

Adherent nerve root syndrome

Nerve root entrapment syndrome

Myofascial pain syndrome

Adverse neural tension syndrome

Postural syndrome

Dysfunction syndrome

Step 3

Abnormal pain syndrome

---

evidence were included. In step two, additional categories widely assumed within the physiotherapy profession to be pathoanatomically oriented, or which indicated pain producing connective tissue, although not specific to certain anatomical structures, were included. In step three, a category widely assumed to indicate that patient responses during clinical examination should be re-evaluated was included. The existing evidence for the reliability and validity of the individual criteria used for categorisation is discussed in more detail elsewhere (Petersen et al 2003).

Reaching a consensus regarding usefulness of classification systems requires data of reliability and validity. The purpose of this study was to investigate inter-tester reliability of the new classification system by having different physiotherapists assess and classify the same patient.

Method

Subjects Patients were referred from general practitioners or specialists for assessment and treatment at the Back Center of Copenhagen. Subjects between 18 and 65 years of age who had low back pain with or without leg symptoms were included in the study. Exclusion criteria were: spondylolisthesis, fracture, osteoporosis, history of spinal surgery, pregnancy, inflammatory disease, cancer, current application for pension, or suspected inability to communicate adequately with the examiner because of language problems, psychiatric disorders, alcoholism, or other impediment.

The design and possible risks for the patient were explained to an authorised representative of the local ethics committee, whose advice was that a formal ethical approval was not required.

Table 1. Characteristics of the study sample (N = 90)

Variable	No. of of subjects
Gender	
Male	36 (40%)
Female	54 (60%)
Age (years)	
Mean	38
SD	11.6
Range	20–68
Duration of symptoms (months)	
Mean	13.4
SD	24.9
Range	0–120
Location of symptoms	
Low back only	29 (32%)
Low back/proximal LE	23 (26%)
Low back/distal LE	38 (42%)
Pain rating (1–10)	
Mean	4.7
SD	2.2
Range	1–10
On sick leave	39 (43%)
Previous treatment	
Physiotherapist	24 (26%)
Chiropractor	14 (16%)
Both above	14 (16%)
Hospital	2 (2%)
General practitioner	36 (40%)

Low back: area between T12 and gluteal fold; Proximal LE (lower extremity): area between gluteal fold and knee; Distal LE: area between knee and foot.

Examiner Four physiotherapists (mean age 41.5 years, range 37 to 51), each with several years of experience in examination of LBP patients (mean 14.5 years, range 7 to 27) and a diploma or credentialling certificate in the McKenzie method of mechanical diagnosis and therapy, participated in the study.

Training Training of the examiners involved three steps: procedures and criteria for classification were discussed; a four-day seminar of training was conducted by the developers of the new classification system during which an assessment form was tested on patients and revised; and finally several meetings were held discussing patient cases and refining questioning skills and manual techniques.

Clinical procedure The clinical procedure was as follows:

• First the examiner considered category 1, disc syndrome, and the three sub-syndromes (reducible disc syndrome, irreducible disc syndrome, or non-mechanical disc syndrome). If the patient satisfied minimal criteria for one of the sub-syndromes, the examiner moved on to consider categories 10–12 (myofascial pain syndrome, adverse neural tension syndrome, and abnormal pain syndrome). Categories 10–12 may coexist with each other and with categories 1–9.

• If the patient did not satisfy criteria for category 1, and had dominant pain below the gluteal fold, categories 2–5 (adherent nerve root syndrome, nerve root entrapment syndrome, nerve root compression syndrome, spinal stenosis syndrome) or 'Other' were considered sequentially before moving on to categories 10–12 (myofascial pain syndrome, adverse neural tension, and abnormal pain syndrome).

• If the patient did not satisfy criteria for category 1, and had dominant symptoms above the gluteal fold, categories 6–9 (zgyapophysial joint syndrome, postural syndrome, sacroiliac joint syndrome, dysfunction syndrome) or 'Other' were considered sequentially before moving on to category 10–12 (myofascial pain syndrome, adverse neural tension, and abnormal pain syndrome).

• If a patient was unable to distinguish whether the symptoms were dominant above or below the gluteal fold, the examiner considered all categories 1–9 before moving on to categories 10–12.

The rationale behind this procedure is described in detail elsewhere (Petersen et al 2003). In brief, the rationale is that if the most common structural source of symptoms (discogenic pain) is eliminated there would be an increase in the prevalence of other syndromes in the remaining group of patients. By removing of the greatest potential for false positive tests we have increased the positive predictive value of the subsequent tests.

Design It is a basic requirement for reliability studies that the phenomenon tested is stable during the testing. Therefore, we chose to let two physiotherapists examine the patient simultaneously in the procedures related to criteria for categories 1–7 and 9 which required lasting changes in symptoms to take place. One therapist served as the first examiner. The second therapist acted as an observer and was not allowed to intervene in any way.

In categories 8 and 10–12, we did not assume that the test procedures were at risk of altering the patient's response, so these procedures were repeated twice. First, the procedure was conducted by one examiner, with the second examiner absent from the room. Afterwards, the procedure was repeated by the second examiner with the first examiner absent. Each therapist in a pair recorded examination findings on a separate assessment form blinded from each other.

The examination procedure was supervised by an independent adjudicator. The tasks of the adjudicator were: to ensure that the examiners did not communicate, to decide if a patient should be excluded from the study because of a previously undiscovered exclusion criterion or because of the patient's inability to endure testing procedures, and to collect the assessment forms at the end of each assessment.

A pseudo-random assignment was used to ensure that the four therapists were paired an equal number of times and performed an equal number of examinations in both roles. Each patient was classified by one pair of examiners. Random assignment by coin flip was used to decide which therapists would form a pair and to assign the paired therapists to the first or the second examiner roles. If, however, a therapist in a pair was nearing the quota of total patient examinations and had not performed an equal number of examinations in both roles, this therapist was assigned to a specific role to eliminate the inequity.

Table 2. Number of subjects, kappa values (K) and 95% confidence intervals, percentages of agreement, and values of P_+ and P_- for the categories of the proposed classification system (N = 90)

Category	No of subjects in cells with positive/ negative agreement	K (95% CI)	Percentage of agreement	P_+	P_-
Disc syndrome	43/31	0.64 (0.48 to 0.80)	82	0.84	0.79
Reducible disc syndrome	36/41	0.71 (0.57 to 0.86)	86	0.85	0.86
Irreducible disc syndrome	2/85	0.56 (0.12 to 1.00)	97	0.57	0.98
Non-mechanical disc syndrome	3/83	0.58 (0.20 to 0.95)	96	0.60	0.98
Adherent nerve root syndrome	3/87	—	97	—	—
Nerve root entrapment syndrome	1/89	—	99	—	—
Nerve root compression syndrome	5/84	0.90 (0.72 to 1.00)	99	0.91	0.99
Spinal stenosis syndrome	1/88	0.66 (0.04 to 1.00)	99	0.67	0.99
Zygapophysial joint syndrome	2/88	1.00 (1.00 to 1.00)	100	1.00	1.00
Postural syndrome	1/89	1.00 (1.00 to 1.00)	100	1.00	1.00
Sacroiliac joint syndrome	8/73	0.65 (0.41 to 0.89)	91	0.67	0.95
Dysfunction syndrome	1/88	0.66 (0.04 to 1.00)	100	0.67	0.99
Other	6/68	0.32 (0.07 to 0.58)	82	0.43	0.89
Myofascial pain syndrome	45/21	0.44 (0.25 to 0.64)	74	0.80	0.65
Adverse neural tension syndrome	63/14	0.59 (0.39 to 0.79)	86	0.91	0.68
Abnormal pain syndrome	5/78	0.55 (0.25 to 0.84)	92	0.36	0.96
Inconclusive	1/84	0.26 (-0.19 to 0.71)	94	0.29	0.97

— = calculation of kappa not possible due to lack of observations in one or more cells

Data analysis Percentage agreement and the kappa statistic (Kramer and Feinstein 1981) were used to estimate inter-tester agreement in choice of category. The kappa statistic estimates the degree of agreement corrected for chance agreement. The calculation of the proportion of positive agreement (P_+) and the proportion of negative agreement (P_-) were made because they offer information about the proportions of agreement for the average of the positive and negative findings respectively (Cicchetti and Feinstein 1990). P_+ and P_- indicate the consistency of the two examiners in categories where their judgements are predominantly positive or negative. A kappa coefficient above 0.4 is generally regarded as acceptable (Landis and Koch 1977).

Statistical calculations were performed using the SPSS 9.0 software for Windows[a] or StatXact-3[b].

Results

A total of 102 consecutive patients were included in the study. Four patients did not show up for examination. Eight patients were excluded by judgement of the adjudicator for the following reasons: no symptoms at entry (2 patients), lack of co-operation (2 patients), and major aggravation of pain during assessment precluding examiners from the clinical examination (4 patients). A total of 90 patients were examined in the study. Patient characteristics are presented in Table 1.

Prevalence of syndromes was calculated from the results of the first examiner. Prevalence of the 11 syndromes, that are mutually exclusive in the classification system, was as follows. Reducible disc syndrome was most common (46%) followed by sacroiliac joint syndrome (13%), nerve root compression syndrome (7%), and non-mechanical disc syndrome (6%). Irreducible disc syndrome, adherent nerve root syndrome, nerve root entrapment syndrome, spinal stenosis syndrome, zygapophysial joint syndrome, postural syndrome, and dysfunction syndrome, were less common, with prevalence ranging from 1% to 3%. The therapists classified 14 (16%) of the patients as 'Other'. This latter category is not a distinct category as these patients were classified into one or more of the remaining three syndromes that may coexist, or into the 'Inconclusive' category. In these remaining syndromes, the distribution of patients was: myofascial pain syndrome 65%, adverse neural tension syndrome 78%, and abnormal pain syndrome 7%. Six patients (4%) were categorised in the final 'Inconclusive' category by one or both of the examiners.

Percentages of agreement ranged from 74% to 100% and kappa coefficients ranged from 0.26 to 1.00 (Table 2). The overall percentage of agreement between therapists for the 11 syndromes that are mutually exclusive and the 'Other' category (N = 90) was 72% and the kappa coefficient was 0.62 (95% CI 0.50 to 0.74). An alternative analysis of the 11 syndromes without the 'Other' category (N = 68) showed 86.8% agreement and an estimated kappa coefficient of 0.79. This alternative analysis was made because 'Other' is not a distinct category and as such it is of limited value. Regarding the three syndromes, myofascial pain syndrome, adverse neural tension syndrome, and abnormal pain syndrome, that may coexist, agreement ranged from 74% to 94% and estimated kappa coefficients ranged from 0.44 to 0.59.

Estimation of agreement over all categories was calculated. Agreement over all categories was said to occur when examiners agreed on all categories considered for a patient. Percentage of agreement over all categories was 39%.

Discussion

This study showed that inter-tester reliability of categorisation

of the syndromes was acceptable. Kappa coefficients of syndromes were above 0.4. In syndromes where calculation of kappa coefficients was not possible due to lack of positive findings the percentage agreement was above 90%. In some syndromes, however, the confidence intervals were quite wide, reflecting the low prevalence of patients allocated to those syndromes.

The moderate to excellent levels of reliability in this study might result from features of the study design. We wanted to test the level of variability in different examiners' administration of clinical tests and interpretation of patient responses, rather than the variability arising from changes in the patient's condition over time. We therefore chose to have the first and the second examiner present simultaneously during testing for categories 1–7 and 9. Possible variability due to repeated testing, or variations in therapist application of assessment procedures, was thus eliminated for these categories. To that extent, the test procedures do not fully mimic the demands of clinical practice. Inter-tester agreement was generally higher in syndromes where both the examiners were present during examination compared with syndromes where the two therapists examined patients separately. The procedure was chosen because the examination included repeated movements and sustained positions until a lasting change in the symptoms could be achieved. This made it highly likely that symptom response during the second examination would be affected by prior testing. While it is possible that the procedure might have introduced bias toward higher agreement, we felt that this design was necessary because previous authors have suggested that poor reliability of certain tests may have been a result from altered symptom response with examination (Delitto et al 1992, Fritz et al 2000, Kilby et al 1990, Razmjou et al 2000, Van Dillen et al 1998).

The generalisability of our findings may be limited by at least two factors. First, the examiners in this study had information from only a single assessment on which to base their conclusion. Some LBP patients may require several assessments in the course of days with the response to specific interventions and the application of home-exercises contributing to the final categorisation (Werneke and Hart 2003). Second, the examiners were all experienced physiotherapists with credentials in the McKenzie system, and had substantial training in administration of the classification system. They may, therefore, not be representative of the average therapist. However, we believed training to be necessary to ensure that the first step in the examination procedure (identification of a possible disc syndrome) was performed carefully. The clinical reasoning process in the proposed classification system is quite complex and not easy to perform. The application of the classification system requires some training and experience in the clinical examination, especially in the McKenzie assessment. The rationale behind this assumption is outlined in Petersen et al (2003).

In our study the majority of patients were classified as having reducible disc syndrome, confirming the results of several studies using similar criteria for classification (Donelson et al 1997, Kilpikoski et al 2002, Razmjou et al 2000, Young and Aprill 2000).

Only a few patients were classified as having adherent nerve root syndrome, nerve root entrapment syndrome, postural syndrome, and dysfunction syndrome. These findings are also in concordance with those of recent studies (Kilpikoski et al 2002, Razmjou 2000). In the study by Kilpokoski et al (2002), the examiners, using the McKenzie classification system, classified no patients as having adherent nerve root syndrome, nerve root entrapment syndrome, or postural syndrome, and only one of 39 non-specific LBP patients was classified as having dysfunction syndrome. In a similar study by Razmjou et al 2002 a total of 45 patients were examined and no patients were classified as having nerve root entrapment syndrome, one patient was classified as having adherent nerve root syndrome or postural syndrome, and three patients were classified as having dysfunction syndrome. One interpretation of our results is that the examiners were not able to categorise these less common conditions using our proposed criteria. This finding might also be related to the relatively limited number of patients included in the study as well as to the characteristics of patients. The great majority of patients had chronic low back pain and were referred to a primary health care centre. Therefore, inclusion of a larger sample of patients and patients with a wider range of characteristics might increase the prevalence of these less common conditions. The same might also apply to spinal stenosis syndrome and zygapophysial joint syndrome.

Adherent nerve root syndrome and nerve root entrapment syndrome are recognisable patterns of signs and symptoms that are transferred from the McKenzie system (McKenzie 1981), which has been shown to be used widely within the physiotherapy profession (Battie et al 1994, Gracey et al 2002, Foster et al 1999). However, the findings in the current and previous studies suggest that these two syndromes might be collapsed into some of the other categories without loss of information.

Percentage of total agreement between therapists was 39% for the classification system as a whole. This level of agreement is not unexpected, when categories are allowed to coexist, given the fact that the greater amount of information that the examiner wishes to obtain, the more she or he is at risk of decreasing its reliability (Nelson et al 1979). Further work is needed to reveal whether some categories may be collapsed without reducing validity.

This classification system may prove useful in clinical studies because it allows the use of the same criteria for categorising patients in both research and clinical practice. Signs and symptoms of the patient sample included in outcome studies will be recognisable to the clinician, enabling implementation of results of these studies in clinical practice. Further studies are needed to determine the guidelines for treatment of patients with a particular syndrome and patterns of coexisting syndromes.

The validation of a new instrument is a continuous multi-step process including studies of reliability and validity (McDowell and Newell 1996). A classification system should also be evaluated in randomised controlled trials to compare the efficacy of different interventions for any given category. The current authors are engaged in ongoing studies designed to test aspects of construct validity of the proposed classification system.

References

Altman DG (1991): Practical Statistics for Medical Research. London: Chapman & Hall.

Andersson GB, Deyo RA (1996): History and physical examination in patients with herniated lumbar discs. *Spine* 21 Suppl: 10S–18S.

Battie MC, Cherkin DC, Dunn R, Ciol MA and Wheeler KJ (1994): Managing low back pain: Attitudes and treatment preferences of physical therapists. *Physical Therapy* 74: 219–226.

Bigos SJ, Bowyer ROR, Braen GR et al (1994): Acute Low Back Problems in Adults. Clinical Practice Guidelines No 14. *AHCPR Publication No 95-0642*: 1–160.

Binkley J, Finch E, Hall J, Black T and Gowland C (1993): Diagnostic classification of patients with low back pain: report on a survey of physical therapy experts. *Physical Therapy* 73: 138–150.

Borkan JM, Koes B, Reis S and Cherkin DC (1998): A report from the Second International Forum for Primary Care Research on Low Back Pain. Re-examining priorities. *Spine* 23: 1992–1996.

Bouter LM, van Tulder MW and Koes BW (1998): Methodologic issues in low back pain research in primary care. *Spine* 23: 2014–2020.

Cherkin DC, Battie MC, Deyo RA, Street JH and Barlow W (1998): 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* 339: 1021–1029.

Cicchetti DV, Feinstein AR (1990): High agreement but low kappa: II. Resolving the paradoxes. *Journal of Clinical Epidemiology* 43: 551–558.

Clinical Standards Advisory Group (1994): Report of a CSAG Committee on Back Pain. London, HMSO: 1–88.

Delitto A, Cibulka MT, Erhard RE, Bowling RW and Tenhula JA (1993): Evidence for use of an extension-mobilization category in acute low back pain syndromes: A prescriptive validation pilot study. *Physical Therapy* 73: 216–228.

Delitto A, Erhard RE and Bowling RW (1995): A treatment-based classification approach to low back syndrome: identifying and staging patients for conservative treatment. *Physical Therapy* 75: 470–485.

Delitto A, Shulman AD, Rose SJ, Strube MJ, Erhard RE, Bowling RW and Smith DS (1992): Reliability of a clinical examination to classify patients with low back syndrome. *Physical Therapy Practice* 1: 1–9.

Deyo RA and Phillips WR (1996): Low back pain. A primary care challenge. *Spine* 21: 2826–2832.

Donelson R. Aprill CN, Medcalf R, Grant W (1997): A prospective study of centralization of lumbar and referred pain. A predictor of symptomatic discs and anular competence. *Spine* 22: 1115–1122.

Erhard RE, Delitto A and Cibulka MT (1994): Relative effectiveness of an extension program and a combined program of manipulation and flexion and extension exercises in patients with acute low back syndrome. *Physical Therapy* 74: 1093–1100.

Foster NE, Thompson KA, Baxter GD and Allen JM (1999): Management of nonspecific low back pain by physiotherapists in Britain and Ireland. A descriptive questionnaire of current clinical practice. *Spine* 24: 1332–1342.

Fritz JM, Delitto A, Welch WC, Erhard RE (1998): Lumbar spinal stenosis: A review of current concepts in evaluation, management, and outcome measurements. *Archives of Physical Medicine and Rehabilitation* 79: 700–708.

Fritz JM, Delitto A, Vignovic M and Busse RG (2000): Interrater reliability of judgments of the centralization phenomenon and status change during movement testing in patients with low back pain. *Archives of Physical Medicine and Rehabilitation* 81: 57–61.

Fritz JM and George S (2000): The use of a classification approach to identify subgroups of patients with acute low back pain. Interrater reliability and short-term treatment outcomes. *Spine* 25: 106–114.

Fritz JM, Delitto A and Erhard RE (2003): Comparison of classification-based physical therapy with therapy based on clinical practice guidelines for patients with acute low back pain. A randomized controlled trial. *Spine* 28: 1363–1372.

Gillan MG, Ross JC, McLean IP and Porter RW (1998): The natural history of trunk list, its associated disability and the influence of McKenzie management. *European Spine Journal* 7: 480–483.

Gracey JH, McDonough SM and Baxter GD (2002): Physiotherapy management of low back pain: A survey of current practice in Northern Ireland. *Spine* 27: 406–411.

Kilby J, Stigant M and Roberts A (1990): The reliability of back pain assessment by physiotherapists using a McKenzie algorithm. *Physiotherapy* 76: 579–583.

Kilpikoski S, Airaksinen O, Kankaanpaa M, Leminen P, Videman T and Alen M (2002): Interexaminer reliability of low back pain assessment using the McKenzie method. *Spine* 27: E207–E214.

Koes BW, van Tulder MW, Ostelo R, Kim BA and Waddell G (2001): Clinical guidelines for the management of low back pain in primary care: An international comparison. *Spine* 26: 2504–2513.

Kramer MS and Feinstein AR (1981): Clinical biostatistics. LIV. The biostatistics of concordance. *Clinical Pharmacology Therapeutics* 29: 111–123.

Landis JR and Koch GG (1977): The measurement of observer agreement for categorical data. *Biometrics* 33: 159–174.

Laslett M, Young S, Aprill C, McDonald B (2003): Diagnosing painful sacroiliac joints: A validity study of a McKenzie evaluation and sacroiliac provocation tests. *Australian Journal of Physiotherapy* 49: 89–97.

Leboeuf-Yde C, Lauritsen JM and Lauritzen T (1997): Why has the search for causes of low back pain largely been nonconclusive? *Spine* 22: 877–881.

Maluf KS, Sahrmann SA and Van Dillen LR (2000): Use of a classification system to guide nonsurgical management of a patient with chronic low back pain. *Physical Therapy* 80: 1097–1111.

McDowell I. and Newell C (1996): Measuring Health. A Guide to Rating Scales and Questionnaires (2nd ed). New York: Oxford University Press.

McKenzie RA (1981): The Lumbar Spine. Mechanical Diagnosis and Therapy. Waikanae: Spinal Publications.

Nelson MA, Allen P, Clamp SE and de Dombal FT (1979): Reliability and reproducibility of clinical findings in low-back pain. *Spine* 4: 97–101.

Nwuga G and Nwuga V (1985): Relative therapeutic efficacy of the Williams and McKenzie protocols in back pain management. *Physiotherapy Practice* 1: 99–105.

Petersen T, Kryger P, Ekdahl C, Olsen S and Jacobsen S (2002): 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* 27: 1702–1709.

Petersen T, Laslett M, Thorsen H, Manniche C, Ekdahl C and Jacobsen S (2003): Diagnostic classification of non-specific low back pain. A new system integrating pathoanatomic and clinical categories. *Physiotherapy Theory and Practice* 19: 213–237.

Petersen T, Thorsen H, Manniche C and Ekdahl C (1999): Classification of nonspecific low back pain. A review of the literature on classification systems relevant to physiotherapy. *Physical Therapy Reviews* 4: 265–281.

Razmjou H, Kramer JF and Yamada R (2000): Intertester reliability of the McKenzie evaluation in assessing patients with mechanical low-back pain. *Journal of Orthopaedic and Sports Physical Therapy* 30: 368–383.

Revel M, Poiraudeau S, Auleley GR, Payan C, Denke A, Nguyen M, Chevrot A and Fermanian J (1998): Capacity of the clinical picture to characterize low back pain relieved by facet joint anesthesia. Proposed criteria to identify patients with painful facet joints. *Spine* 23: 1972–1977.

Spitzer WO (1987): Scientific approach to the assessment and management of activity-related spinal disorders. A monograph for clinicians. Report of the Quebec Task Force on Spinal Disorders. *Spine* 12, Suppl: S1–S59.

Stankovic R and Johnell O (1990): Conservative treatment of acute low-back pain. A prospective randomized trial: McKenzie method of treatment versus patient education in mini back school. *Spine* 15: 120–123.

Stankovic R and Johnell O (1995): Conservative treatment of acute low back pain. A 5-year follow-up study of two methods of treatment. *Spine* 20: 469–472.

Van Dillen LR, Sahrmann SA, Norton BJ, Caldwell CA, Fleming DA, McDonnell MK and Woolsey NB (1998): Reliability of physical examination items used for classification of patients with low back pain. *Physical Therapy* 78: 979–988.

Van Dillen LR, Sahrmann SA, Norton BJ, Caldwell CA, McDonnell MK and Bloom N (2003): The effect of modifying patient-preferred spinal movement and alignment during symptom testing in patients with low back pain. A preliminary report. *Archives of Physical Medicine and Rehabilitation* 84: 313–322.

Werneke M and Hart DL (2003): Discriminant validity and relative precision for classifying patients with nonspecific neck and back pain by anatomic pain patterns. *Spine* 28: 161–166.

Young S and Aprill CN (2000): Characteristics of a mechanical assessment for chronic lumbar facet joint pain. *Journal of Manual and Manipulative Therapy* 8: 78–84.

Appendix

Low back pain classification: Definitions and criteria for categorisation*

Glossary

Mechanical loading strategies The performance of repeated trunk movements, sustained positions by the patient, and the application of manual overpressure, mobilisation, and/or manipulation by the therapist.

Produce symptoms During performance of mechanical loading strategies, symptoms appear that were not present prior to the performance.

Increase symptoms During performance of mechanical loading strategies, symptoms that were already present prior to the performance are enhanced.

Decrease symptoms During performance of mechanical loading strategies, symptoms that were already present prior to the performance are diminished.

Abolish symptoms During performance of mechanical loading strategies, symptoms that were already present prior to the performance are eliminated.

Centralisation: The abolition of symptoms in the most distal body component during the performance of mechanical loading strategies. The symptoms remain abolished from that component as a result.

Peripheralisation The production of symptoms in a more distal body component during the performance of mechanical loading strategies. The symptoms remain present in that component as a result.

Worse Symptoms that are produced, increased, or peripheralised as a result of mechanical loading strategies remain produced, increased, or peripheralised as a result.

Better Symptoms that are abolished, decreased, or centralised as a result of mechanical loading strategies remain abolished, decreased, or centralised as a result.

Relevant lateral shift A lateral lumbar deformity that is related to the patient's present symptoms.

1 Disc syndrome

a. Mechanical reducible disc

Definition Low back and/or referred pain assumed to be caused by a displacement of the contents of an intervertebral disc that is reversible by specific mechanical loading strategies.

Minimal criteria

- At least one movement is painfully limited.
- *Either* there are mechanical loading strategies that centralise the symptoms from the most distal component of the pain distribution. *Or* isolated midline low back pain is decreased and remains better by the application of a loading strategy in one direction, and a loading strategy in another direction increases midline pain that remains worse and/or produces a limitation of movement.

b. Mechanical irreducible disc

Definition Low back and/or referred pain assumed to be caused by a displacement of the contents of an intervertebral disc that are not reversible by mechanical loading strategies.

Minimal criteria

- At least one movement is painfully limited.
- There are no mechanical loading strategies that centralise, decrease and/or abolish symptoms so that they remain better as a result.
- *Either* at least one mechanical loading strategy peripheralises the symptoms to a more distal component o*r* the symptoms referred into the foot are increased and remain worse by the application of a loading strategy, and a limitation of movement is produced or increased.

c. Non-mechanical disc

Definition Low back pain with or without referred pain with dominant symptoms above the gluteal fold in which the principal source of nociceptor receptor activity is assumed to be a chemically sensitive intervertebral disc and no evidence for a mechanical disc lesion exists.

Minimal criteria

- The criteria for reducible and irreducible mechanical disc are not satisfied.
- Mechanical loading strategies in any direction increase the symptoms and may remain worse as a result.
- There are no mechanical loading strategies that decrease and/or abolish the symptoms.
- The range of movement remains unaffected by mechanical loading strategies.
- One or more 'other disc characteristics' are present.

Other disc characteristics

- The dominant symptoms are midline or bilateral above S1.
- The dominant symptoms are unilateral above S1 but the zygapophysial joint criteria are not satisfied.
- The symptoms change sides under the influence of unilateral mechanical loading strategies.
- There is a relevant lateral shift.

2 Adherent nerve root syndrome

Definition Dominant symptoms below the gluteal fold with limited nerve root mobility assumed to be caused by fibrosis or scarring involving one or more lumbosacral nerve roots.

Minimal criteria

- History of acute sciatica at least 2 months ago or lumbar spine surgery.

- Flexion in standing is limited and produces the lower limb symptoms at the end of the available movement range that is not rapidly altered by mechanical loading strategies.

- Repeated flexion in standing reproduces the symptoms with each movement but they do not remain worse as a result.

- Extension in standing or lying, and flexion in lying do not produce the symptoms.

3 Nerve root entrapment syndrome

Definition Dominant symptoms below the gluteal fold assumed to be caused by a persistent compression and movement limitation of a lumbar nerve root.

Minimal criteria

- The criteria for reducible and irreducible mechanical disc and adherent nerve root are not satisfied.

- History of acute disc lesion causing nerve root symptoms for at least 2 months.

- Flexion in standing is limited and produces or increases the lower limb symptoms.

- Repeated flexion in standing reproduces or increases the symptoms but they do not remain worse as a result.

- Repeated flexion in standing may cause an increase in movement range but this is temporary and does not remain better as a result.

- There are no mechanical loading strategies that centralise, decrease or abolish the lower limb symptoms so that they remain better as a result.

4 Nerve root compression syndrome

Definition Dominant symptoms below the gluteal fold assumed to be caused by a compression of a nerve root that is not made worse or better by mechanical loading strategies.

Minimal criteria

- The criteria for reducible and irreducible mechanical disc lesion, adherent nerve root, and nerve root entrapment are not satisfied.

- The straight leg raise test is positive (familiar lower limb symptoms are produced below 60 degrees of elevation) and at least one of the following is present in the corresponding myotome/dermatome:

  - hip flexion weakness (L2/L3)
  - knee extension weakness (L3/L4)
  - ankle dorsiflexion weakness (L4–L5)
  - great toe dorsiflexion weakness (L5)
  - hip extension weakness (L4/L5–S1/S2)
  - knee flexion or great toe extension weakness (L5–S1)
  - ankle plantarflexion weakness (S1–S2)
  - patellar tendon reflex weakness (L4)
  - achilles tendon reflex weakness (S1)

5 Spinal stenosis syndrome

Definition Dominant symptoms below the gluteal fold that are assumed to be secondary to a narrowing of the lumbar spinal canal or a lumbar intervertebral foramen.

Minimal criteria

- The criteria for reducible and irreducible mechanical disc, adherent nerve root, nerve root entrapment, and nerve root compression are not satisfied.

- History of standing or walking intolerance.

- Symptoms are improved when seated or there is improved walking tolerance with the spine in flexion.

- Best posture with regard to symptoms is sitting or worst postures with regard to symptoms are standing or walking.

6 Zygapophysial joint syndrome

Definition Low back pain with or without referred pain with dominant symptoms above the gluteal fold in which the principal source of nociceptor receptor activity is assumed to be a zygapophysial joint.

Minimal criteria

- The criteria for disc syndrome are not satisfied.

- Pain well-relieved by lying down and the presence of at least four of the following criteria:

  - age greater than 65 years
  - pain not increased by coughing
  - no pain with flexion in standing
  - pain not increased by rising from flexion
  - pain not increased by extension/rotation
  - pain not increased by extension in standing.

7 Postural syndrome

Definition Low back pain with or without referred pain with dominant symptoms above the gluteal fold assumed to result from mechanical deformation of innervated normal soft tissues by prolonged static end range loading.

Minimum criteria

- Full range of motion in all directions.

- No pain with any movement.

- Repeated dynamic end range loading does not produce the symptoms.

- Sustained end range loading in at least one direction produces the familiar symptoms.

8 Sacroiliac joint syndrome

Definition Low back pain with or without referred pain with dominant symptoms above the gluteal fold in which the principal source of nociceptor receptor activity is assumed to be a sacroiliac joint.

Minimal criteria

- The criteria for disc syndrome, zygapophysial joint syndrome, and postural syndrome are not satisfied.

- Three or more of five sacroiliac joint pain provocation tests are positive:

  - distraction
  - compression
  - thigh thrust (posterior shear)
  - pelvic torsion (Gaenslen's test)
  - sacral thrust.

9 Dysfunction syndrome

Definition Low back pain with or without referred pain with dominant symptoms above the gluteal fold assumed to result from mechanical deformation by end range loading of innervated shortened soft tissues.

Minimal criteria

- The criteria for disc syndrome, zygapophysial joint syndrome, postural syndrome, and sacroiliac joint syndrome are not satisfied.

- At least one movement is limited in range that is not rapidly altered by mechanical loading strategies

- The limited movement produces familiar symptoms only at the end of the available movement range.

- End range loading in the painfully limited direction of motion does not progressively increase or peripheralise the symptoms and the symptoms do not remain worse as a result.

- Repeated or sustained end range loading in the painfully limited direction does not rapidly produce limitation of movement range in any other direction.

10 Myofascial pain syndrome

Definition Low back and/or referred pain with dominant symptoms above or below the gluteal fold assumed to result from a hyperirritable point in a skeletal muscle or fascia that is painful on compression and can give rise to referred pain in a characteristic area.

Minimal criteria

- Firm palpation of a painful point within a taut band in a specific muscle reproduces familiar symptoms.

11 Adverse neural tension syndrome

Definition Low back and/or referred pain assumed to result from abnormal physiological and mechanical responses produced from nervous system structures when their range of movement and stretch capabilities are challenged.

Minimal criteria

- Familiar symptoms are reproduced by at least two stages of neural testing:

  - straight leg raise with cervical flexion or slump test

  - sidelying knee bending test (femoral nerve stretch test).

12 Abnormal pain syndrome

Definition Maladaptive overt illness-related behaviour disproportionate to the underlying physical disease and more readily attributable to associated cognitive and affective disturbances.

Minimal criteria

- At least three of five tests of non-organic signs are positive:

  - widespread superficial or non-anatomic tenderness

  - pain provocation on axial loading or simulated rotation of the back

  - straight leg raise improved at least 30 degrees with distraction

  - regional muscle weakness or sensory disturbances in non-anatomic distribution

  - overreaction during examination.

13 Inconclusive

Definition Non-specific low back pain patients not included in any of the above listed classes.

*Note Categories 1 to 9 are mutually exclusive, however they may coexist with categories 10 to 12.