Introducing relevant subgroups of nonspecific low back pain patients may improve research efficiency and clinical outcomes. This article presents available data supporting the rationale behind a new pathoanatomic oriented classification system for low back pain. A systematic framework was used to assess issues of validity, reliability, feasibility, and generalizability. Results from studies examining elements of the 12 categories of the classification system are presented to allow the assessment to be made. Particular emphasis has been given to the evidence of diagnostic accuracy of the criteria based on data from history taking and clinical tests. The classification system was found to be potentially useful for research purposes. Further testing of the classification system is required.

**INTRODUCTION**

It is generally acknowledged that most patients with low back pain (LBP) are without a specific medical diagnosis due to the lack of an identified pathology that matches the patient’s symptoms (Spitzer, 1987). Discussion about diagnostic subclassification of LBP arises from the assumption that this large heterogeneous group of patients would be treated more effectively if the patients could be assigned to more homogeneous subgroups on the basis of valid criteria (Borkan et al, 1998; Bouter, van Tulder, and Koes, 1998; Leboeuf-Yde, Lauritsen, and Lauritzen, 1997; Spitzer, 1987).

International guidelines for the management of low back pain recommend an initial diagnostic classification process, a diagnostic triad, that differentiates among possible serious spinal pathology, nerve root problems, and nonspecific LBP (Bigos et al, 1994; Clinical Standards Advisory Group, 1994; Koes et al,
It is estimated that 85% of LBP patients seen in primary care have nonspecific LBP (Deyo and Phillips, 1996). Thus, the diagnostic label of nonspecific LBP does not contain specific therapeutic information, and refers to a large heterogeneous group of patients suffering from a variety of different pathological or patho-physiological conditions.

Several classification systems have been proposed for subdividing nonspecific LBP patients by means of clinical examination (Petersen et al, 1999). In physiotherapy, three of those are of particular interest inasmuch as they (1) are sufficiently detailed to have implications for choice of treatment for the individual patient and (2) have been tested for reliability and validity (Delitto, Erhard, and Bowling, 1995; Maluf, Sahrmann, and Van Dillen, 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 loading of the spine. It has been reported as the most commonly used system by physical therapists (Battie et al., 1994; Gracey, McDonough, and Baxter, 2002). The system has been tested for reliability and has substantial intertester agreement (Kappa coefficients ranging from 0.6 to 0.7) (Kilpikoski et al, 2002; Razmjou, Kramer, and Yamada, 2000). Randomized controlled trials examining the validity of the McKenzie system, ie, its ability to categorize patients in a way that might result in selection of the most effective treatment, have shown conflicting results (Cherkin et al, 1998; Gillan et al, 1998; Nwuga and Nwuga, 1985; Petersen et al, 2002; Stankovic and Johnell, 1990, 1995).

Delitto et al (1985) have developed a classification system for categorization of patients with acute LBP by means of history taking and clinical examination. The system as a whole has shown a moderate intertester agreement (Kappa coefficient 0.56) (Fritz and George, 2000). Two randomized controlled trials have, with regard to choice of treatment, shown validity of one (extension-mobilization category) of the seven categories of the system (Delitto et al, 1993; Erhard, Delitto, and Cibulka, 1994). In addition, one epidemiological study has shown differences in short-term prognosis for a few of the categories in the system when they were collapsed into four overall treatment categories (Fritz and George, 2000).

Maluf, Sahrmann and Van Dillen (2000) 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. The recording of movements and activities in daily functioning that provoke the patient's familiar symptoms is of particular interest in the system. Reliability of the individual tests used as criteria for classification varies from fair to perfect (Kappa coefficients ranging from 0.21 to 1.00) (Van Dillen et al, 1998). However, there are no reports on the general reliability of classifying patients into the five categories. The use of the system has been illustrated by a case report (Maluf, Sahrmann, and Van Dillen, 2000), but no data have been published on the validity of the system.

To summarize, although data have been published supporting reliability of some of these systems, none has proven its superiority over others in identifying subgroups of patients with better outcomes from a specific treatment compared to others. Therefore, existing classification systems do not eliminate the need for development of alternative ones.

The present authors propose a new classification system that has a different, pathoanatomic orientation. The development of this system is in the early stages compared to the existing treatment-based ones, and a long line of future testing is required to substantiate or refute the validity and utility of this pathoanatomic approach.

This system aims at overcoming some of the fundamental problems concerning the existing treatment-oriented approaches. Labels and criteria used for classification in current 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 the abolition of pain following extension movements would be classified as a “posterior derangement syndrome” in the system proposed by McKenzie (1981) and “extension syndrome” in the system proposed by Delitto et al (1995), and in the “flexion category” in the system proposed by Maluf, Sahrmann, and Van Dillen (2000).

It has been pointed out that various practitioners have different, but equally acceptable, approaches to the management of a particular treatment-oriented diagnostic category (Binkley et al, 1993). Instead of 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, it should be the results of clinical trials that determine the most effective treatments for particular categories of patients.

**THE PROPOSED CLASSIFICATION SYSTEM**

The 12 syndromes and 3 subsyndromes included in this system and the proposed criteria for categorization are summarized in Table 1.

In the following text, the classification system is presented within the methodological framework to evaluate classification systems developed by Buchbinder et al (1996). Where it has been possible to identify studies that examined particular elements of the different categories, results from these studies have been included as a basis for the evaluation. The components of the framework are summarized in Table 2.

**RESULTS OF THE EVALUATION**

**Purpose of the classification system**

The purpose of the classification system is to identify clinically homogeneous subgroups of patients with nonspecific LBP according to the assumed symptomatic structures. The system is primarily developed for use in outcome studies in primary care settings where the efficacy of different treatments to a particular diagnostic category is tested. The system takes into account that access to imaging and advanced diagnostic procedures is limited in primary care. The secondary purpose of the classification system is to assist the therapist in selecting the most effective treatment for the individual patient based on the results of outcome studies.

**Content validity**

The domain of the classification system is nonspecific LBP patients without a specific medical diagnosis. Further exclusions are patients with symptoms known as “red flags” (indicating possible serious spinal pathology) (Bigos et al, 1994), hip disorders, or suspected referred pain from the viscera.

In order to include all relevant categories, the proposed classification system is made up of 12 categories that refer to the known possible pain-producing structures in the low back, ie, discs, zygapophysial joints, sacroiliac joints, nerves, and muscles (Fairbank and Hall, 1990) (Table 1). In addition, category 12, “abnormal pain syndrome,” includes patients that are characterised by abnormal illness behaviour (Waddell, Pilowsky, and Bond, 1989).

Categories 1–9 are mutually exclusive inasmuch as each category refers to structures included in a lumbosacral motion segment. However, they can each coexist with categories 10–12.

Categories 10 and 11 refer to other structures containing nociceptor receptors and are therefore able to produce pain in addition to a symptomatic vertebral problem.

Category 12, although not referring to a specific pathoanatomic condition, is included to identify patients for which the examiner cannot trust pain related responses or pain provocation/abolition tests for other categories. Positive nonorganic signs should lead the examiner to evaluate what these signs might mean in the individual patient, eg, neural hyperirritability might mean that the
<table>
<thead>
<tr>
<th>Syndromes</th>
<th>Definition and localization of symptoms</th>
<th>Key examination findings constituting minimal criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a. Disc syndrome: Reducible disc</td>
<td>Low back and/or referred pain with dominant symptoms above or below the gluteal fold assumed to be caused by a displacement of the contents of an intervertebral disc that are reversible by specific mechanical loading strategies.</td>
<td>At least one movement is painfully limited; Centralization of symptoms</td>
</tr>
<tr>
<td>1b. Disc syndrome: Irreducible disc</td>
<td>Low back and/or referred pain with dominant symptoms above or below the gluteal fold assumed to be caused by a displacement of the contents of an intervertebral disc that are not reversible by specific mechanical loading strategies.</td>
<td>At least one movement is painfully limited; No centralization, decrease and/or abolishment of symptoms so that they get better; Peripheralization of symptoms</td>
</tr>
<tr>
<td>1c. Disc syndrome: Nonmechanical disc</td>
<td>Low back and/or 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.</td>
<td>Mechanical loading strategies in any direction increase the symptoms, which may get no worse or worse as a result; No decrease and/or abolishment of symptoms; Range of movement remains unaffected; One or more “other disc characteristics” are present: dominant symptoms are midline or bilateral above S1; dominant symptoms are unilateral above S1 but the zygapophysial joint criteria are not satisfied; symptoms change sides; relevant lateral shift;</td>
</tr>
<tr>
<td>2. Adherent nerve root</td>
<td>Dominant symptoms below the gluteal fold with limited nerve root mobility assumed to be caused by fibrosis or scarring involving one or more lumbo-sacral nerve roots.</td>
<td>The criteria for disc syndrome are not satisfied; 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; Repeated flexion in standing reproduces the symptoms and they do not increase or get worse as a result; Extension in standing or lying and flexion in lying do not produce the symptoms</td>
</tr>
<tr>
<td>3. Nerve root entrapment</td>
<td>Dominant symptoms below the gluteal fold assumed to be caused by a persistent compression and movement limitation of a lumbar nerve root.</td>
<td>The criteria for disc and adherent nerve root syndrome are not satisfied; History of acute nerve root symptoms at least 2 months ago; 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 get worse as a result; Repeated flexion in standing may cause an increase in movement range but this does not get better as a result;</td>
</tr>
</tbody>
</table>
4. Nerve-root compression 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.

The criteria for disc, adherent nerve root, and nerve root entrapment syndrome are not satisfied; The straight leg raise test is positive and reflex and/or muscle weakness are present in the corresponding myotome/dermatome.

5. Spinal stenosis 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.

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 improved by lying down and the presence of at least one of the following criteria:

- Pain well-relieved by lying down and the presence of at least one of the following:
  - Lying on the side in the supine position
  - Sitting
  - Lying on the back with the knees flexed
  - Lying on the back with the knees extended

6. Zygapophysial joint 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.

The criteria for disc syndrome are not satisfied; At least one of the following criteria:

- Pain relieved by lying down
- No pain with any movement
- Pain well-relieved by sitting or standing
- Pain relieved by a physician's suggestion
- Pain relieved by lying on the back with the knees flexed
- Pain relieved by lying on the back with the knees extended
- Pain relieved by sitting
- Pain relieved by standing
- Pain relieved by sitting or standing
- Pain relieved by movement
- Pain relieved by a physician's suggestion
- Pain relieved by lying down
- No pain with any movement
- Pain well-relieved by sitting or standing
- Pain relieved by a physician's suggestion
- Pain relieved by lying on the back with the knees flexed
- Pain relieved by lying on the back with the knees extended
- Pain relieved by sitting
- Pain relieved by standing
- Pain relieved by sitting or standing
- Pain relieved by a physician's suggestion

7. Postural 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.

The criteria for all other directions are not satisfied; At least one of the following 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 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.

The criteria for disc, zygapophysial joint, and postural syndrome are not satisfied; At least one of the following criteria:

- Distraction
- Compression
- Thigh thrust (posterior shear)
- Pelvic torsion (Gaenslen's test)
- Sacral thrust

9. Dysfunction 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.

The criteria for disc, zygapophysial joint, postural, and sacroiliac joint syndrome are not satisfied; At least one movement is limited in range that is not rapidly altered; 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 peripheralize the symptoms and the symptoms do not get worse as a result; End range loading in the painfully limited direction does not rapidly produce limitation of the movement range in any other direction.

(Continued)
<table>
<thead>
<tr>
<th>Syndromes</th>
<th>Definition and localization of symptoms</th>
<th>Key examination findings constituting minimal criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>10. Myofascial pain</td>
<td>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.</td>
<td>Firm palpation of a painful point within a taut band in a specific muscle reproduces familiar symptoms</td>
</tr>
</tbody>
</table>
| 11. Adverse neural tension| 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. | Familiar symptoms are reproduced by at least two stages of neural testing:  
  - straight leg raise with cervical flexion or slump test  
  - side-lying knee bending test |
| 12. Abnormal pain         | Maladaptive overt illness related behaviour disproportionate to the underlying physical disease and more readily attributable to associated cognitive and affective disturbances. | At least three of five tests of nonorganic 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 disturbance in non-anatomic distribution  
  - overreaction during examination |
| 13. Inconclusive          | Nonspecific low back pain patients not included in any of the above listed classes.                      |                                                                                                                      |

Mechanical loading strategies: The performance of repeated trunk movements, sustained positions by the patient, and the application of manual overpressure, mobilization, 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.
Centralization: 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.
Peripheralization: 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 peripheralized as a result of mechanical loading strategies remain produced, increased or peripheralized as a result.
Better: Symptoms that are abolished, decreased, or centralized as a result of mechanical loading strategies remain abolished, decreased, or centralized as a result.
Relevant lateral shift: A lateral lumbar deformity that is related to the patient’s present symptoms.
patient might have a certain propensity toward or purpose in overreacting to tests, a fear of movement, or is malingering.

**Breakdown of categories**

In the breakdown of categories, the proposed clinical procedure is as follows:

- First, the examiner considers category I, disc syndrome, and the three subsyndromes (reducible disc, irreducible disc, or non-mechanical disc) (see Table 1). If the patient fulfills the minimal criteria for one of the subsyndromes, the examiner moves on to consider categories 10–12 (myofascial pain syndrome, adverse neural tension syndrome, and abnormal pain syndrome).
- If the patient does not fulfill criteria for category I, and has dominant pain below the gluteal fold, categories 2–5 (adherent nerve root syndrome, nerve root entrapment syndrome, nerve root compression syndrome, or spinal stenosis syndrome) are considered sequentially before moving on to categories 10–12.
- If the patient does not fulfill criteria for category 1, and has dominant symptoms above the gluteal fold, categories 6–9 (zygapophysial joint syndrome, postural syndrome, sacroiliac joint syndrome, or...
dysfunction syndrome) are considered sequentially before moving on to categories 10–12.

- If a patient is unable to distinguish whether the symptoms are dominant above or below the gluteal fold, the examiner considers all categories, 1–9, before moving on to categories 10–12.

The sequential ordering of tests means that once the patient is classified in a category from 1–9, further testing with respect to the remaining mutually exclusive categories is not to be performed.

The order of this decision process has been chosen for several reasons:

1. Data suggest that intervertebral disc pathology is the most common structural source of symptoms in nonspecific LBP. Studies using discography as a diagnostic test found that the prevalence of discogenic pain ranges from 39% (Schwarzer et al., 1995) to 57% (Donelson et al., 1997) in chronic LBP samples. These studies have been conducted with patients who have failed to respond to previous treatments. The prevalence of discogenic pain may be different in primary care settings, but this has not yet been estimated. The estimated prevalence of symptomatic zygapophysial joints based on double blocks varies from 9% to 42% (Schwarzer et al., 1994a, 1994b, 1994c; Manchikanti et al., 2000; Manchikanti et al., 2001) in chronic LBP patients. Using single injections as a diagnostic test, the prevalence of symptomatic sacroiliac joints has been estimated to range from 13% to 30% (Maigne, Aivaliklis, and Pfefer, 1996; Schwarzer, April, and Bogduk, 1995; Fortin et al., 1994). Assuming an examiner is able to identify the symptomatic disc category with an acceptable degree of accuracy, it is reasonable to believe that the likelihood of identifying the remaining syndromes increases after excluding discogenic pain. Coexistence of symptomatic discs and zygapophysial joints was found in only 3% of one sample of patients (Schwarzer et al., 1994a) and the combination of symptomatic discs and sacroiliac joints was not found at all (Fortin et al., 1994) in another sample. Studies have shown that symptomatic zygapophysial and sacroiliac joints either do not coexist (Fortin et al., 1994), or do so with a low prevalence of 2% (Schwarzer, Aprill, and Bogduk, 1995) in chronic nonspecific LBP patients. The prevalence of combinations of other symptomatic anatomical structures is not known.

2. Most tests used as criteria for placing a patient within a specific syndrome category are likely to stress several lumbar structures simultaneously (eg discs, zygapophysial joints, sacroiliac joints, and neural tissues) (Di Fabio, 2001; Manchikanti et al., 2001) thereby generating false positive responses. Therefore, it seems logical to consider the category with the highest prevalence first so that the tests with the greatest potential for false positives are removed before the remaining categories are considered.

3. Certain syndromes, ie, dysfunction syndrome and postural syndrome, are assumed to indicate a pain producing origin in low back connective tissue, although the examiner is not able to specify the source further. In cases where the minimal criteria would mean a possible overlap of syndromes, ie, zygapophysial or sacroiliac joint vs. dysfunction, classification of the patient in a more specific pathoanatomic category before dysfunction is considered. The same decision rule applies to adherent nerve root or nerve root entrapment syndrome vs. nerve root compression syndrome. The mutual exclusiveness in the conclusion of the examination procedure does not mean that categories necessarily are mutually exclusive in identifying different structures. The order of these three syndrome was made, first, to differentiate among different specific types of nerve root pathology that are identified by other systems by signs and symptoms and, second, to make it possible to identify syndromes where a different prognosis and a different treatment have been presented and thereby made testable. No data is available to
support the location in the decision-making procedure of the syndromes listed above.

**Method of development**

The proposed classification system was developed by experienced physical therapists and was based on both criteria for categorization that was presented in a recently published version of a new system of classification (Laslett and van Wijmen, 1999) and results from an extensive literature search (Petersen et al, 1999).

An initial version of the classification system was evaluated by a conference of five Danish back specialists. These specialists were appointed by the chairmen of the Danish branch of the International Federation of Orthopaedic Manipulative Therapists, the McKenzie Institute Denmark, and the Danish Institute for Health Technology Assessment committee dealing with LBP.

**Face validity**

In most categories of the classification system, ie, categories 1–6, 8, and 10–11, a nomenclature is used to label the categories according to the supposed origin of the pain (Table 1). The system does not attempt to establish a medical diagnosis, which may require the use of paraclinical methods (eg, imaging procedures, discography, and diagnostic injections) that are not available in routine physiotherapy practice. However, while the categories do comprise specific syndromes (entities based on signs and symptoms) that refer to low back structures, at this point of time, most of them can only be assumed to be the origin of the patient’s symptoms. Categories 7 and 9 are assumed to indicate a pain producing origin in low back connective tissue, not specifically identified. Category 12 is labelled according to a distinguishable clinical pattern believed to identify (mainly chronic) LBP patients whose pain behaviour is disproportionate to the underlying physical disorder.

The terms used are based on empirical information obtained from clinical examination alone and the definitions of the criteria that determine inclusion into each category (see Table 1).

For some of the categories, there is evidence suggesting the criteria to have validity, ie, that the proposed criteria are able to identify a specific symptomatic structure with an acceptable degree of accuracy. In the other categories, this ability is hypothesized and future evidence has to be established. A summary of the face validity and reliability of the criteria used for categorization is presented in Table 3.

**CATEGORY 1**

The validity of the criteria for placing the patient in category 1, disc syndrome, has been examined in several studies. Centralization and peripheralization of symptoms following repeated or sustained lumbar loading in different directions, used as a diagnostic tool, has been examined by Donelson et al (1997). A symptomatic intervertebral disc lesion was confirmed by discography in 72% of cases. Further analysis of this study estimated 95% sensitivity and 52% specificity (Delaney and Hubka, 1999). Several studies have indicated the centralization phenomenon to be a useful clinical sign for discriminating among patients with different prognoses. Identification of patients as “centralizers” (equivalent to the reducible mechanical disc subcategory), has shown prognostic value in predicting good outcome from treatments by the McKenzie method (Donelson, Silva, and Murphy, 1990; Sufka et al, 1998; Werneke and Hart, 2001; Werneke, Hart, and Cook, 1999), a combination of various exercises (Karas et al, 1997), or a work hardening program (Long, 1995). Provocation of pain in several directions that mechanically stress the affected disc has been hypothesized to indicate pain from chemical irritation associated with internal disc disruption (Bogduk, 1997). Such a pattern has prognostic value in identifying acute LBP patients with a five-fold increased risk of developing chronic pain (Hellings, Linton, and Kalvemark, 1994).

In order to incorporate the knowledge of prognostic value in the proposed classification
Table 3
Summary table on the evidence of validity and reliability of the criteria and categories used in the classification system

<table>
<thead>
<tr>
<th>Syndromes and criteria</th>
<th>Evidence of validity of criteria</th>
<th>Kappa values and/or % agreement of criteria when included in the McK system as a whole</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Disc Centralization/Peripheralization</td>
<td>Sens. 0.94. Spec. 0.52</td>
<td>—</td>
</tr>
<tr>
<td>Centralization</td>
<td>—</td>
<td>0.51–0.79</td>
</tr>
<tr>
<td>2. Adherent nerve root</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>3. Nerve root entrapment</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>4. Nerve root compression</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Leg pain</td>
<td>Sens. 0.98. Spec. 0.88</td>
<td>0.53–0.96</td>
</tr>
<tr>
<td>SLR</td>
<td>Sens. 0.96</td>
<td>0.44–0.86</td>
</tr>
<tr>
<td>Reflexes</td>
<td>Spec. 0.63–0.97</td>
<td>0.39–0.50</td>
</tr>
<tr>
<td>Motor weakness</td>
<td>Spec. 0.71–0.99</td>
<td>0.65–1.00</td>
</tr>
<tr>
<td>5. Spinal stenosis</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Leg pain</td>
<td>Sens. 0.88. Spec. 0.34</td>
<td>—</td>
</tr>
<tr>
<td>Best seated</td>
<td>Sens. 0.89</td>
<td>0.70</td>
</tr>
<tr>
<td>Worse walking/standing</td>
<td>Sens. 0.89</td>
<td>0.82/0.83</td>
</tr>
<tr>
<td>No pain seated</td>
<td>Spec. 0.93</td>
<td>0.70</td>
</tr>
<tr>
<td>Improved walk tolerance</td>
<td>Spec. 0.91</td>
<td>—</td>
</tr>
<tr>
<td>6. Zygaphophysial joint</td>
<td>Contradictory</td>
<td>—</td>
</tr>
<tr>
<td>7. Postural</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>8. Sacroiliac joint</td>
<td>Sens. 0.91. Spec. 0.83*</td>
<td>—</td>
</tr>
<tr>
<td>Distraction</td>
<td>Sens. 0.72–0.97. Spec. 0.11–0.66</td>
<td>0.44–0.86</td>
</tr>
<tr>
<td>Compression</td>
<td>0.73</td>
<td>88%</td>
</tr>
<tr>
<td>Thigh thrust</td>
<td>0.88</td>
<td>94%</td>
</tr>
<tr>
<td>Pelvic torsion</td>
<td>0.75</td>
<td>88%</td>
</tr>
<tr>
<td>Sacral thrust</td>
<td>0.52</td>
<td>78%</td>
</tr>
<tr>
<td>9. Dysfunction</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>10. Myofascial pain Reproduct. of fam. pain</td>
<td>Diff between occurrence in LBP patients and healthy controls</td>
<td>0.57–0.58</td>
</tr>
<tr>
<td>11. Adverse neural tension</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>SLR</td>
<td>Sens. 0.72–0.97. Spec. 0.11–0.66</td>
<td>0.44–0.86</td>
</tr>
<tr>
<td>Slump test</td>
<td>Inconclusive</td>
<td>0.89</td>
</tr>
<tr>
<td>Femoral nerve test</td>
<td>Identification of symptomatic nerve root in 95% of patients with sciatica from L4–5 level</td>
<td>0.27–0.57</td>
</tr>
<tr>
<td>12. Abnormal pain</td>
<td>Closely related to other measures</td>
<td>86%*</td>
</tr>
<tr>
<td>Tenderness of distress or affective</td>
<td>—</td>
<td>80%</td>
</tr>
<tr>
<td>Simulation</td>
<td>—</td>
<td>78%</td>
</tr>
<tr>
<td>Distraction</td>
<td>Identification of delayed return</td>
<td>84–86%</td>
</tr>
<tr>
<td>Regional disturbances</td>
<td>—</td>
<td>82%</td>
</tr>
<tr>
<td>Overreaction</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

system, the disc syndrome category is subdivided into three subcategories dependent upon whether the patient has a reducible, irrevocable, or nonmechanical response to pain provocation testing.

Three studies support the intertester reliability of criteria used in category 1. In two of these studies, acceptable kappa values of agreement in assessment of pain centralization was reported (k: 0.51–0.79) (Fritz et al, 2000; Kilby, Stigant, and Roberts, 1990). However, kappa values might have been artificially inflated, inasmuch as the examiners in the study by Fritz et al (2000) were watching videotapes of patients performing the test movements and in the study by Kilby, Stigant, and Roberts (1990) one of the two examiners was performing the tests with the other examiner observing. Thus, a potential source of error arising from differences in the examiners interaction with the patient was eliminated. This limitation was addressed by a recent study in which the examiners performed the examination and categorized the patients independently (Kilpikoski et al, 2002). In this latter study, interobserver agreement between examiners for the assessment of centralization was substantial (k: 0.7).

CATEGORIES 2, 3, 7, AND 9

The validity of the criteria for placing the patient in categories 2, 3, 7, and 9, adherent nerve root syndrome, nerve root entrapment syndrome, postural syndrome, and dysfunction syndrome, have not been established. These categories have been transferred from the McKenzie classification system in order to include distinguishable patterns of signs and symptoms widely used in physiotherapy (Battie et al, 1994; Foster et al, 1999). The pain of dysfunction syndrome is thought to be caused by the tension loading of sensitive, adaptively shortened structures. Adherent nerve root syndrome is considered to be a subset of dysfunction related to nerve tissue. Nerve root entrapment is thought to be the result of reduced nerve root mobility caused by fibrotic herniated disc material. Postural syndrome is thought to be caused by static end-range loading of otherwise normal structures in certain habitual postures (McKenzie, 1990; Razmjou, Kramer, and Yamada, 2000). Testing is needed to reveal if these assumptions are valid. These categories are of value in establishing prognosis and choice of treatment (McKenzie, 1990).

The reliability of criteria for category 2, 3, 7, and 9 has not been reported in the form of kappa-coefficients, either because the authors had restricted the purpose of their study to examine the overall intertester reliability of the McKenzie classification system (Riddle and Rothstein, 1993), or because of low prevalence of these four categories in the study sample (Kilby, Stigant, and Roberts, 1990; Kilpikoski et al, 2002; Razmjou, Karmer, and Yamada, 2000). Intertester percentage of agreement in these categories has shown great variability, ranging from 33% (Kilby, Stigant, and Roberts, 1990) to 98% (Razmjou, Kramer, and Yamada, 2000) in adherent nerve root syndrome; 100% in nerve root entrapment syndrome (Razmjou, Kramer, and Yamada, 2000); from 44% (Riddle and Rothstein, 1993) to 100% (Kilby, Stigant, and Roberts, 1990; Razmjou, Kramer, and Yamada, 2000) in postural syndrome; and 0% (Kilpikoski et al, 2002), 51% (Riddle and Rothstein, 1993), 67% (Kilby, Stigant, and Roberts, 1990), or 98% (Razmjou, Kramer, and Yamada, 2000) in dysfunction syndrome. An explanation for this great variability might be that agreement between examiners increases with experience (Razmjou, Kramer, and Yamada, 2000), inasmuch as the highest percentage of agreement was found in the only study where examiners had passed an examination for minimal competence in the McKenzie classification system (Razmjou, Kramer, and Yamada, 2000). This explanation, however, is challenged by a recent study, in which the lowest percentage of agreement in identification of dysfunction syndrome was between examiners with the highest competence in the McKenzie classification system (Kilpikoski et al, 2002).
CATEGORY 4

Combining highly sensitive tests with highly specific tests should lead to a set of criteria with an acceptable degree of diagnostic power (Deyo et al, 1994). This approach has been used when developing the criteria for categories 4–5.

The validity of criteria used for placing the patient into category 4, nerve root compression syndrome, has been reviewed by Andersson and Deyo (1996). Presence of leg pain was found to be a good indicator of nerve root compression (sensitivity 0.98 and specificity 0.88). The pain of nerve root compression has been described to be predominantly leg pain, although back pain may be present (McCall, 2000). A positive straight leg raise test (SLR) had a sensitivity of 0.96 (Andersson and Deyo, 1996). A combination of SLR with one or more positive high specificity tests within a corresponding dermatome, ought to give an acceptable degree of diagnostic power. Such tests would be quadriceps weakness (specificity 0.99), ankle dorsiflexor weakness (specificity 0.82), diminished patellar tendon reflex (specificity 0.97), great toe extensor weakness (specificity 0.71), ankle plantar flexor weakness (specificity 0.95), and diminished achilles tendon reflex (specificity 0.63) (Andersson and Deyo, 1996).

Intertester reliability has been shown to range from moderate to almost perfect for presence of leg pain (k: 0.53–0.96) (McCombe et al, 1989), SLR (k: 0.44–0.86) (McCombe et al, 1989; Strender et al, 1997; Vroomen, de Krom, and Knottnerus, 2000), ankle dorsiflexor weakness (k: 1.0), great toe extensor weakness (k: 0.65), and diminished achilles tendon reflex (k: 0.39–0.50) (Deyo, Rainville, and Kent, 1992). Estimates of intertester reliability are not known for the other tests included in the criteria.

CATEGORY 5

Criteria for placing the patient into category 5, spinal stenosis syndrome, were derived from a recent review summarizing the validity expressed as the diagnostic value of signs and symptoms in lumbar spinal stenosis (Fritz et al, 1998). Sensitive clinical features were: severe leg pain (sensitivity 0.88), symptoms best when seated (sensitivity 0.89), and symptoms worst when walking/standing (sensitivity 0.89). Findings with high specificity were: no pain when seated (specificity 0.93) and improved walking tolerance with the spine in flexion (specificity of 0.91).

Data exist to support the reliability of these criteria inasmuch as intertester agreement of patients’ reports of pain response to activity and position has been found to be moderate to almost perfect (k: 0.46–0.89) (Roach et al, 1997).

CATEGORY 6

The validity of criteria for category 6, zygapophysial joint syndrome, has been reported by Revel et al (1992, 1998). The criteria were able to identify symptomatic zygapophysial joints, discriminating 92% of patients who would respond to a single anaesthetic injection, and 80% of those who would not. However, the authors cautioned against using these criteria for diagnostic purposes because of higher false-positive rate from the use of single anaesthetic blocks compared to the use of double blocks (Schwazer et al, 1994d). The findings by Revel et al are largely consistent with those reported by others (Jackson, 1992), but a recent study by Manchikanti et al (2001) did not confirm the results when the double anaesthetic block method was used as the criterion standard. In this matter, we agree with Bogduk (1998) who, in a comment on the controversy, has stated: “there is a need for some form of screening, before diagnostic blocks are performed essentially arbitrarily on all patients with back pain, just in case they have zygapophysial joint pain.” We suggest that LBP patients suspected of having discogenic pain be excluded, as a subgroup, prior to interpretation of Revel’s criteria. This will increase the prevalence of facetogenic pain in the remaining
group of patients. Consequently, the risk of producing false positive diagnoses of symptomatic zygapophysial joints in cases having discogenic pain is reduced. In a study of 61 chronic LBP patients (Young and Aprill, 2000), none of 11 patients diagnosed as having symptomatic zygapophysial joints (by a single anaesthetic block), fulfilled the criteria for a disc syndrome. Studies examining pain distribution arising from zygapophysial joint stimulation by injections have shown that the pain is dominantly situated in lumbar area above the gluteal fold (Fukui et al, 1997; McCall, Park, and O’Brien, 1979; Young and Aprill, 2000). Based upon these findings, the classification system excludes patients with dominant symptoms below the gluteal fold before evaluation for facetogenic pain is considered.

Although some data support the proposed criteria for the diagnosis of zygapophysial joint syndrome, the ability of the criteria to identify symptomatic zygapophysial joints remains hypothetical.

Intertester reliability of the criteria has not been reported.

**CATEGORY 8**

The validity of the criteria for placing the patient in category 8, sacroiliac joint syndrome, has been examined in a study by Laslett et al (2003). They found that the examination process was able to identify symptomatic sacroiliac joints confirmed by double anaesthetic blocks (sensitivity 0.91, specificity 0.83).

Intertester reliability of the sacroiliac pain provocation tests included in the criteria for this category has been found to be acceptable (k: 0.52–0.88) (Laslett and Williams, 1994).

**CATEGORY 10**

The criteria for placing the patient in category 10, myofascial pain syndrome, is a simplification of the set of criteria originally recommended by Travell and Simons (1982). Our simplification is based on studies examining validity and reliability of the original criteria. The validity of these criteria has been examined by Wolfe et al (1992). They concluded that the findings of a palpable taut band in the muscle or a local twitch response was not able to separate LBP patients with a myofascial pain syndrome from healthy persons. Conversely, acceptable validity was achieved when a simple set of criteria was applied including only the pain provocation tests of Travell and Simons (focal muscle tenderness elicited on palpation, and production of referred pain). The results were confirmed by Njoo and Van der Does (1994) who examined the recommended criteria separately and found only the pain provocation tests (reproduction of the patients’ familiar pain and “jump sign” on palpation) to have discriminative value in separating chronic non-specific LBP patients from persons without LBP. Studies examining reliability of the criteria have shown moderate to substantial intertester reliability for reproduction of familiar local or referred pain and the “jump sign” (k: >0.5) (Gerwin et al, 1997; Njoo and Van der Does, 1994). In contrast, the originally proposed set of criteria by Travell and Simons have fair to moderate intertester reliability (k: 0.39–0.46) (Nice et al, 1992). Consequently, only the pain provocation tests have been included as criteria in this proposed classification system.

**CATEGORY 11**

The criteria for placing the patient in the category 11, adverse neural tension syndrome, is a simplification of the set of criteria originally proposed by Butler (1991). Examining the validity of this syndrome is complicated. Various conditions such as bulging or herniated discs, swelling or fibrotic tissue adjacent to or within a nerve root, or a pathological relationship with a number of peripheral structures to which the peripheral nerve is attached, may cause an abnormal mechanical sensitivity of the nervous system (Butler, 1991).
The validity of some of the proposed criteria has been reported. In a review of studies testing the straight leg raise test (SLR), sensitivity was found to be high, ranging from 0.72 to 0.97, but specificity was low, ranging from 0.11 to 0.66 in identifying nerve root tension caused by a herniated intervertebral disc (Andersson and Deyo, 1996). Specificity may be improved by tightening up criteria required for a positive test response (Fritz and Wainner, 2001). This will increase the utility of the test for including a condition when the test is positive (fewer false positives). On the other hand, tighter criteria reduces sensitivity, and may compromise the utility of the test because it will decrease the ability of a test to exclude a condition when the test is negative (more false negatives). For our purpose, it is more important to make sure that every patient included in a specific syndrome actually has the assumed pathologic condition, rather than overlooking a patient with a specific syndrome (in which case the patient remains in a non-specific condition category, labelled “inconclusive”). Therefore, it seemed reasonable in the proposed classification system to apply the more strict criterion for a positive SLR test proposed by Butler (1991) ie, that two steps in the neural testing procedure must reproduce the patients’ known pain. Reproduction of pain appears to be an essential component of the criteria. Assessing onset of tissue resistance during SLR has been shown to have no validity in identifying an abnormal mechanical sensitivity of the nervous system (Hall, Zusman, and Elvey, 1998).

The validity of the femoral nerve stretch test has been examined in one study (Christodoulides, 1989). The test was able to identify nerve root involvement caused by L4/5 disc herniation in LBP patients with sciatica in 95% of cases (sensitivity and specificity were not reported). A general problem in establishing the validity of neural tension manoeuvres is that other pain producing structures besides the neural tissues are stressed simultaneously (Di Fabio, 2001; Johnson and Chiarello, 1997). Further studies are needed to determine whether these manoeuvres can distinguish painful neural tissues from pain produced by other structures.

Most studies testing reliability of the SLR and the slump test have shown acceptable degrees of intertester agreement of both tests (k: 0.44–0.89) in nonspecific LBP patients (McCombe et al, 1989; Philip, Lew, and Matyas, 1989; Strender et al, 1997; Vroomen, de Krom, and Knottnerus, 2000).

The femoral nerve stretch test has been shown to have from fair to moderate intertester agreement (k: 0.27–0.57) (McCombe et al, 1989). In an other study showing 99% agreement, kappa values were incalculable because of a low prevalence of positive findings (Strender et al, 1997).

**CATEGORY 12**

The criteria for placing the patient into category 12, abnormal pain syndrome, have been introduced and validated by Waddell et al (1980). They found the occurrence of at least three out of five positive nonorganic signs to be the most clinically useful tool in identifying magnified illness behaviour in chronic LBP patients. Nonorganic signs may be used in several ways: (1) Identification of patients characterized by pain behavior dominated by emotional distress and excessive preoccupation with physical illness (Waddell, Pilowsky, and Bond, 1989), (2) Identification of patients whose pain is not primarily a symptom of tissue injury, but of a hypersensitivity of the nervous system, making the patient’s pain experience the main problem (Feuerstein and Beattie, 1995), and (3) Identification of patients where the examiner is unable to reach to a conclusion from the patient’s response to pain provocation tests (Waddell et al, 1984). The validity of nonorganic signs has been documented in studies showing the existence of positive nonorganic signs to be closely related to other measures of distress and affective disturbances (Novy et al, 1998; Ransford, Cairns, and Mooney, 1976; Waddell et al, 1980; Waddell, Pilowsky, and Bond, 1989). The prevalence of
abnormal pain syndrome (using the criteria of at least three of five positive tests) is 19.5% to 25% in chronic nonspecific LBP patients (Chan et al, 1993; Hirsch et al, 1991), 30% in problem patients referred to special clinics (Waddell et al, 1980), and 66% in patients who are receiving or anticipating financial gain due to their chronic LBP problem (Hayes et al, 1993). Three or more positive tests can discriminate between patients receiving or anticipating financial gain from their chronic LBP problem and patients who are not (Hayes et al, 1993).

Although kappa statistics were not reported, percentage of agreement between examiners has been found to be 86% for at least three out of five positive tests, and 78–86% for the individual tests included in the criteria, when examiners had been trained in the assessment (Waddell et al, 1980). A study using examiners with limited training found poor to moderate intertester agreement (k: –0.03–0.48) (McCombe et al, 1989).

Feasibility

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.

Our own experience from the application of the proposed classification system indicates that the full assessment takes a maximum of one hour to complete.

Construct validity

**Discriminative ability of the classification system**

In relation to our primary purpose, some evidence suggests that three categories—disc syndrome, zygapophysial joint syndrome, and sacroiliac joint syndrome—discriminate different sources of pain at the structural level. In a recent study, 48 chronic nonspecific LBP patients were screened for symptomatic sacroiliac joints by double diagnostic blocks (Laslett et al, 2003). Eleven patients were diagnosed with a symptomatic sacroiliac joint. Of these, ten were correctly identified by the proposed criteria for sacroiliac joint syndrome (sensitivity 0.91, specificity 0.83). Another study (Young and Aprill, 2000), examined whether specific symptom characteristics could distinguish between different symptomatic structures diagnosed by discography, and single block injections of the sacroiliac joints or zygapophysial joints. Of 93 chronic nonspecific LBP patients, no diagnostic conclusion could be reached in 32 cases. Four patients had primary bony pathology (retrolisthesis or lateral stenosis) or hip pathology. Of the remaining 57 patients, 25 were diagnosed as having discogenic pain, 21 with sacroiliac joint pain, and 11 with zygapophysial joint pain. Following the physical examination, the signs that had the highest discriminative ability were centralization (48% of the category, and 0% of the zygapophysial and sacroiliac joint categories) and peripheralization (56% of the disc category, 0% of the zygapophysial joint category, and 5% of the sacroiliac joint category). These findings are compatible with the results reported by Donelson et al (1997). However, confirmation in future studies using double blocks in the diagnosis of symptomatic zygapophysial and sacroiliac joints is needed. No other studies have been conducted testing the assumption that initial screening for disc syndromes will increase the diagnostic accuracy of testing for other syndromes.

In relation to our secondary purpose, the relevant question is are there any outcome studies demonstrating that the proposed classification system is able to assist the therapist in selecting treatments that are more effective than others for patients in specific categories? For the classification system as a whole, no data are available answering this question. For some categories, however, effectiveness and prognostic value have been evaluated.

In categories 1–3, 7, and 9, disc syndrome, adherent nerve root syndrome, nerve root entrapment syndrome, postural syndrome, and dysfunction syndrome, which are based on the McKenzie examination system, randomized controlled trials have examined the efficacy of
treatment according to the findings from the initial mechanical examination. These studies deal with the McKenzie classification system as a whole, without reports of the distribution of patients across different syndromes. Stankovic and Johnell (1990) studied 100 acute nonspecific LBP patients and found that the McKenzie treatment resulted in more pain reduction, earlier return to work, and fewer recurrences in the following year compared to patients treated with education in a mini back school. However, these results are likely to be biased inasmuch as the McKenzie group received more treatments (average of 5.5 treatments) compared to the 45 minutes of mini back school the other patients received. Furthermore, there was no blinding of the investigator at follow up assessment. A five-year follow up investigation assessed 89 of the patients included in the original study, and showed that the McKenzie treatment group had less recurrences and less episodes of sick leave caused by LBP than the control group (Stankovic and Johnell, 1995).

Another study included 62 acute LBP patients with the diagnosis of prolapsed intervertebral disc (Nwuga and Nwuga, 1985). The results showed that patients treated with the McKenzie method had less pain and were able to sit for a longer period of time without symptoms six weeks after the start of treatment, compared to patients treated with Williams flexion exercises. However, the randomization procedure was not described and all patients were treated by the same therapist, so the possibility of intervention bias cannot be excluded. Gillan et al (1998) included 40 acute nonspecific LBP patients with a lateral shift deformity and pain referred into the leg. At the end of treatment and at three-months follow up, results showed no difference in reduction of disability between patients treated with the McKenzie method compared to patients receiving general back care advice and massage. The credibility of the results is weakened by a drop-out rate of 38% after randomization.

Cherkin et al (1998) allocated 321 acute or subacute nonspecific LBP patients to three groups: (1) treatment with the McKenzie method, (2) chiropractic treatment, or (3) provision of an educational booklet. At the end of the treatment, there was a nonsignificant trend towards greater pain reduction and satisfaction in the patients treated with the McKenzie method or chiropractic treatment with regard to disability, recurrence rates, or use of health care services. Petersen et al (2002) tested the effectiveness of the McKenzie method as compared with that of intensive strengthening training for patients with subacute or chronic LBP. At the end of treatment and at two-months follow up there was a trend towards greater reduction of pain and disability in the McKenzie group. No differences between groups were found at eight-months follow up.

To summarize, randomized controlled trials designed to test efficacy of the McKenzie classification system as a whole, in terms of outcome of the treatments assigned to the different categories, have produced mixed results.

For categories 4, 5, 6, and 8, nerve root compression syndrome, spinal stenosis syndrome, zygapophysial joint syndrome, and sacroiliac joint syndrome, no reports of testing for aspects of construct validity have been published.

In category 10, myofascial pain syndrome, recent reviews of the literature show that studies examining the efficacy of physiotherapy treatments or needling therapies provide contradictory results and contain methodological errors (Cummings and White, 2001; Hey and Helewa, 1994). Therefore, no conclusions can be reached concerning aspects of construct validity for this category.

In category 11, adverse neural tension syndrome, clinical usefulness of the identification of this syndrome is only anecdotally presented in a few case reports describing efficacy of examination and treatment (Klingman, 1999; Koury and Scarpelli, 1994). A randomized controlled trial found no additional effect of neural mobilization supplementary to a postoperative standard exercise program for patients undergoing lumbar surgery (Scrimshaw and Maher, 2001). However, it is not clear if the patients exactly fulfilled the
criteria of adverse neural tension syndrome in the proposed classification system. Therefore, the evidence must be regarded as inconclusive.

In category 12, abnormal pain syndrome, studies have documented the ability of nonorganic signs to identify chronic non-specific LBP patients who will perform poorly on biomechanical testing (Hirsch et al, 1991), or who are without objective findings (Vallfors, 1985). A study examining prognostic value of nonorganic signs in LBP patients with varying symptom duration has demonstrated that an initial high score predicted low likelihood of return to work after physiotherapy treatment (Karas et al, 1997). In acute LBP patients, findings are contradictory. In two studies, nonorganic signs were associated with delayed return to work (Gaines and Hegmann, 1999; McIntosh et al, 2000) following physiotherapy treatment, whereas another study found no such influence (Fritz, Wainner, and Hicks, 2000). Lack of reduction of nonorganic signs during physiotherapy predicted which patients would not return to work (Werneke, Harris, and Lichter, 1993). However, such an association was not seen following an interdisciplinary functional restoration program (Polatin et al, 1997). It is suggested that the findings of nonorganic signs have implications for treatment, ie, focus on management of illness behaviour, establishment of self-control over pain, and reduction of disease conviction (Feuerstein and Beattie, 1995).

Comparisons with other classification systems

At present, no studies have made such comparisons. Studies are currently in progress by one of the authors of the present paper to test how well some of the categories of this proposed classification system correlates with a reference standard identification of pathology (based on discography, diagnostic injections, and imaging techniques).

Reliability

The proposed classification system has not been tested as a whole for intertester reliability. However, various studies provide data that suggest the degree of reliability to expect for categories 1, 2, 3, 7, and 9 that are based on the McKenzie classification system. A summary is presented in Table 3.

Overall intertester reliability of the McKenzie system has been examined in four studies. In a study by Kilby, Stigant, and Roberts (1990), 41 nonspecific LBP patients were categorized by two therapists with various levels of training in the McKenzie examination techniques. The therapists were unable to classify 13 of the 41 patients for reasons not described. In the remaining 28 patients, the therapists agreed on syndrome categorization in only 58% of the cases. In the classification of patients in seven subcategories of derangement syndrome, the therapists agreed in only 57% of the cases (kappa-coefficients not reported). The results were weakened by an insufficient sample size and by the fact that only one of the two examiners was performing the tests while the other examiner observed. Both of these limitations were addressed in a study by Riddle and Rothstein (1993) examining 363 nonspecific LBP patients, tested independently by 49 therapists with no or limited training in the McKenzie techniques. The study showed only 39% of agreement between therapists in syndrome categorization (k: 0.15). Razmjou, Kramer, and Yamada (2000) evaluated 45 nonspecific LBP patients using two examiners who had passed an examination of minimal competence in the McKenzie techniques. The study showed only 39% of agreement between therapists in syndrome categorization (k: 0.15). Razmjou, Kramer, and Yamada (2000) evaluated 45 nonspecific LBP patients using two examiners who had passed an examination of minimal competence in the McKenzie techniques. Their results indicated substantial intertester reliability (k: 0.7) for selection of the McKenzie syndromes. Again, a drawback of this study is that one of the examiners performed the tests while the other examiner observed. Kilpikoski et al (2002) evaluated 39 nonspecific LBP patients examined independently by two examiners who held the highest degree of competence in the McKenzie techniques. Their results indicated good intertester reliability (k: 0.6) for selection of the McKenzie syndromes. In the latter two studies, however, the use of only two examiners limits generalizability of these results.
To summarize, the overall intertester reliability of the McKenzie classification system appears to be good when the examiners have at least a minimal level of competence in the McKenzie assessment technique. It is not known to what degree these previously reported results can be generalised to categories 1, 2, 3, 7 and 9 in the proposed classification system. Furthermore, the examiners in the studies reporting acceptable intertester reliability, were all experienced physical therapists, who held credentials in the McKenzie system, and had a substantial amount of training in administration of this system. They may therefore not be representative of the average therapist.

**Generalizability**

Generalizability of this proposed classification system as a whole has yet to be evaluated. Generalizability of the categories transferred from the McKenzie classification system is expected to be satisfactory inasmuch as review of the literature reveals that it has been used by numerous researchers in different settings. Furthermore, reports from the United States and Great Britain indicate that examination and therapy by the McKenzie method is the most commonly used method by physical therapists for the management of LBP (Battie et al, 1994; Foster et al, 1999).

**DISCUSSION**

We have provided the available data supporting or refuting the rationale behind the new classification system oriented at pathoanatomic subgroupings for LBP relevant to physiotherapy. In our view, the strength of this classification system is three-fold.

First, it is based on the best available evidence for a clinical pathoanatomic oriented diagnostic system. Therefore, we expect it to be able to identify syndromes characteristic of symptomatic low back structures with an acceptable degree of accuracy.

Second, hypotheses for evaluating prognosis and efficacy of treatments can be supported by anatomically grounded theory. For example, efficacy of McKenzie treatment might be recommended for reducible mechanical disc syndrome based on an understanding of disc mechanics with flexion and extension movements (Adams et al, 2000; Adams and Hutton, 1982; Beattie et al, 1994; Brault et al, 1997; Fennell, Jones, and Hukins, 1996; Magnusson et al, 1996; Schnebel et al, 1988; Schnebel, Watkins, and Dillín, 1989; Simunic, Broom, and Robertson, 2001).

Finally, structure-oriented diagnostic classifications systems have been held responsible for increasing fear avoidance beliefs, use of ineffective treatments, and poor prognosis in LBP patients (Waddell, 1998; Zusman, 1998). In our opinion, this may be true in cases, where results of imaging procedures are not interpreted together with clinical findings (Deyo et al, 1994). However, we have reasons to believe that the contrary is true for a pathology-oriented classification system based on signs and symptoms. These issues are discussed below.

Identification of the cause of the symptoms is important for LBP patients and the main reason for seeing a general practitioner (McPhillips-Tangum et al, 1998; Schers et al, 2001). Neglecting patients’ expectations has implications for patient satisfaction (Deyo and Diehl, 1986) as well as for other outcomes of treatment (Base et al, 1986; Kendrick et al, 2001; Thomas, 1987; Kalauokalani et al, 2001; Lutz et al, 1999). As stated by Waddell (1998), fear avoidance beliefs seem to be more related to the uncertainty of diagnosis than to the severity of the physical problem. This notion is supported by a recent study (Kendrick et al, 2001) that found increased reported disability at follow-up in nonspecific LBP patients who had x-rays taken, compared to patients not having x-rays taken. However, no differences were found between patients with normal x-rays and patients with x-rays showing anatomical abnormalities. Thus, a possible explanation of these results is that, given the inability to relate radiography findings to the patients’ symptoms, the explanations provided by the doctor contribute to a sense of uncertainty in most cases and fail to
provide the patient with options to handle the situation. If patients’ sick-roles were reinforced merely by the identification of structural pathology, an association between pathological findings and poor prognosis would have been expected.

Medical sociologists point out that uncertainty and lack of consensus in relation to examination and treatment of nonspecific LBP problems give rise to a diversity of explanations among professionals treating LBP, leaving the patient in a conceptual vacuum. This may result in the development of myths and personal misconceptions regarding aetiology and treatment (Skelton et al, 1996) and continued search for more diagnostic tests (Deyo and Diehl, 1986; McPhillips-Tangum et al, 1998; Mechanic, 1977).

In the light of the fact that diagnostic imaging procedures often fail to provide a meaningful diagnosis in patients with nonspecific LBP (Beattie, 1996; van Tulder et al, 1997), the need for alternative ways to develop the potentially healing visualization of the correspondence between objective findings and subjective pain has been emphasized (Rhodes et al, 1999). If we succeed in illustrating a connection among the functional limitations in daily life, the symptom response to clinical tests, and the symptomatic structure to which we direct treatment, then we, together with the patient, would have created a conceptual model that may reduce uncertainty and fear and motivate the patient toward an active approach to self-care (Lacroix et al, 1990; Novack, 1987; Skelton et al, 1996).

There are other reasons for developing a pathoanatomic oriented classification system. Within orthopaedic physiotherapy and the McKenzie system it is hypothesized that symptomatic anatomical structures are identified and treated. For example, sacroiliac joint pathology is assumed in “sacroiliac strain” (Ombregt et al, 1995, p. 706) and disc pathology is assumed in “derangement syndrome” (McKenzie, 1981). If hypotheses regarding the effectiveness of these treatment techniques are to be tested, it requires a diagnostic classification system that is (1) able to separate pathologic conditions regarded as being suitable for the specific treatment from those not suitable, and (2) based on clinical tests commonly used in daily practice. A prerequisite for implementation of research in daily practice, is that patients with particular pathological conditions are readily identifiable by the practitioner.

Furthermore, when communicating the results of one’s intervention to other medical professionals, a more universally accepted clinical classification, meaningful to the medical community, is preferable.

The primary limitations of the classification system is that not all of the criteria and categories have been tested for validity. For instance, in categories 2, 3, 7, 9, and 11, the claimed sources of symptoms have not been identified. These categories are labelled by recognizable patterns of signs and symptoms that are widely claimed within the physiotherapy profession to indicate possible tissues known to be able to produce pain and/or to lead specific treatment strategies. We therefore find it reasonable, at least initially, to include these categories in the classification system. Future studies will reveal if these categories should remain.

All the professionals who have taken part in the development of the present classification have discussed the possibility of including a separate “instability” category. However, there was consensus that, at the present, an instability category would not satisfy the purpose or criteria we have used, namely the identification of pain producing anatomical structures by means of clearly defined signs and symptoms.

Biomechanical studies have documented the importance of the deep stabilizing muscles, ie, transversus abdominis and multifidus, in reducing the shear stress forces influencing the lumbar spine (Bergmark, 1989; Panjabi, 1992). Theories concerning control of the lumbar neutral zone and the therapeutic use of specific stabilizing exercises has been further developed and refined (Moseley, Hodges, and Gandevia, 2002; O’Sullivan, 2000; Richardson et al, 1999). Several studies have been published that report the effectiveness of specific stabilizing exercises in the treatment of LBP
A great number of different clinical tests have been proposed that might contribute to the identification of functional instability (Hamilton 1995; Maluf, Sahrmann, and Van Dillen 2000; O’Sullivan et al, 2002; Richardson et al, 1999; Van Dillen et al, 2001). Assessment criteria involve visual evaluation of control of lumbar spine motion, ability to control the neutral zone, muscle activity, effort in performance, reactions to verbal instructions, and manual facilitation. When examined for interobserver reliability, a number of these tests came out with excellent kappa values when including pain response and substantially lower values when based on visual evaluation (Van Dillen et al, 1998). Although the deep stabilizing muscles are not the pain producing structures themselves, they appear to be involved in a muscular imbalance mechanism that may trigger pain from potential pain producing structures, such as discs, zygapophysial (Z)-joints and sacroiliac (SI)-joints. As such, an “instability syndrome” would be complementary to several of the existing categories, and we plan to include such a supplementary syndrome in the classification system when it has been possible to establish a clear set of minimal criteria.

The order of the clinical examination procedure described in the content validity section is worthy of discussion. It might be argued that spinal stenosis syndrome should be regarded as a specific type of nerve root pathology, and logically it should be tested for before nerve root compression syndrome. However, at this stage we do not regard the available evidence (Fritz et al, 1998; Jenis and An, 2000) as supporting the need for inclusion of a neurological deficit within the essential criteria for clinical diagnosis of spinal stenosis.

If the purpose of classifying patients is not to test hypotheses of treatment efficacy, the order might be different. For example, in studies of the prognostic value of the identification of nerve root compression, it might be suitable initially to identify patients fulfilling the criteria for nerve root compression syndrome eventually followed by a subdivision into the categories 1–3, 5, or 9.

It is not known to which degree the reported evidence of intertester reliability of the McKenzie classification system can be transferred to the proposed classification system. For example, there is a slight difference in the criteria for placing the patient in the disc syndrome compared to the criteria for placing the patient in the derangement syndrome in the McKenzie system. We have adopted a stricter definition of centralization and peripheralization, i.e., the requirement that symptoms have to move from one body component to another and remain there as a result of mechanical loading strategies. In the studies examining reliability of the McKenzie system, alterations in intensity of symptoms or small movements of symptoms within a particular body component appear to be sufficient to place the patient in the derangement category. The reason for this difference in definitive criteria is that the purpose of the present classification system is to identify a symptomatic disc. Therefore, we have adopted the criteria used in the study by Donelson et al, (1997). In contrast, the purpose of the McKenzie system is to guide the therapist in choosing treatment and in this process the recording of even slight changes in the symptom response is valuable. The impact of this discrepancy in definitions is addressed in ongoing studies.

Likewise, studies are needed to address the topic as to whether or not categories should be subdivided further into subgroups in order to assist in selecting the most appropriate treatment for the individual patient. For example, a sacroiliac joint syndrome might be subdivided into unstable or inflammatory subsyndromes, a reducible disc syndrome into subgroups related to the patient’s direction of preference (flexion, extension, side-gliding, or rotation), and an adverse neural tension syndrome into subcategories dependent on presence/absence of a hip-rotation-sensitizing component.

The present paper does not pretend to provide answers to all questions raised concerning usefulness of the proposed classification system. It is but one step in a continuous
scientific process. The authors are engaged in ongoing studies designed to test reliability and aspects of construct validity of the classification system.

Acknowledgments

The preparation of this article was, in part, supported by grants from the Danish Physiotherapy Organization, the Health Insurance Fund, and the Fund for Promotion of Physiotherapeutic Research and Education.

The authors want to thank Paula van Wijmen for methodological comments, and Steen Olsen, Charlotte Engman, Lone Frandsen, Irene Stilling, Jette Toldam, Jette Wos, Charlotte Krog, Inger Skjaerbaek, Carl Terje Roervik, Asbjoern Midttun, and Tom Bendix for clinical advice.

References


Butler DS 1991 Mobilisation of the nervous system. Churchill Livingstone, Melbourne


Christodoulides AN 1989 Ipsilateral sciatica on femoral nerve stretch test is pathognomonic of an L4/5 disc protrusion. Journal of Bone and Joint Surgery 71-B:88–89


Fennell AJ, Jones AP, Hukins DW 1996 Migration of the nucleus pulposus within the intervertebral disc during flexion and extension of the spine. Spine 21:2753–2757
Gaines WG, Jr, Hegmann KT 1999 Effectiveness of Waddell's nonorganic signs in predicting a delayed return to regular work in patients experiencing acute occupational low back pain. Spine 24:396–400
pain in primary care: An international comparison. Spine 26:2904–2913
Landis JR, Koch GG, 1977 The measurement of observer agreement for categorical data, Biometrics 33:159–174
Laslett M, Van Wijmen PM 1999 Low back and referred pain: Diagnosis and a proposed new system of classification. New Zealand Journal of Physiotherapy 27:5–14
Leboeuf-Yde C, Lauritsen JM, Lauritzen T 1997 Why has the search for causes of low back pain largely been nonconclusive? Spine 22:877–881
McCall IW, Park WM, O'Brien JP 1979 Induced pain referral from posterior lumbar elements in normal subjects. Spine 4:441–446
Nwuga G, Nwuga V 1985 Relative therapeutic efficacy of the active straight-leg-raise test. Spine 27:E1
CLASSIFICATION OF NON-SPECIFIC LBP 235
Travell JG, Simons DG 1982 Myofascial pain and dysfunction. The triggerpoint manual. Williams and Wilkins, Baltimore, Maryland, USA


Waddell G 1998 The back pain revolution, 1 edn, Churchill Livingstone, Edinburgh


