

Non-specific Low Back Pain Classification and treatment

Tom Petersen



LUND
UNIVERSITY

Lund 2003

From The Department of Physical Therapy
Lund University, Lund, Sweden
and The Back Center of Copenhagen
Copenhagen, Denmark

Layout by Ortonova AB
Printed in Sweden
Studentlitteratur AB
2003

*When a thing ceases to be a subject of controversy,
it ceases to be a subject of interest.*

William Hazlitt (1778–1830),
British essayist.

Contents

Abstract, 2

List of publications, 4

Abbreviations, 5

Introduction, 6

Prevalence of low back pain, 6

Definitions and diagnosis, 6

Treatment, 8

Heterogeneity in outcome studies, 9

Methodological requirements for classification systems, 10

Classification systems in low back pain, 11

Objectives, 14

Designs and Methods, 15

Effects of the McKenzie method of mechanical diagnosis and therapy, 15

Literature review on classification systems of relevance to physiotherapy, 16

Development of the new classification system, 17

Inter-tester reliability of the new classification system, 17

Statistical methods, 17

Summary of Results, 19

McKenzie therapy versus strengthening training, 19

Existing classification systems for non-specific low back pain patients, 21

The rationale behind the new classification system based on existing evidence, 21

Inter-tester reliability of the new classification system, 29

Discussion, 30

Treatment, 30

Development of categories and criteria of the new classification system, 32

Reliability of the new system, 38

Conclusions, 40

Summary in Danish, 41

Acknowledgements, 43

References, 45

Papers I–IV

Abstract

“Non-specific Low Back Pain—Classification and Treatment”

The aims of this thesis were to investigate the effects of the McKenzie method of mechanical diagnosis and therapy compared with that of intensive dynamic strengthening training for patients with non-specific low back pain (NSLBP), and to develop a clinical diagnostic classification system for use in primary care.

The McKenzie method is one of the most common methods for examination and treatment of patients with NSLBP used by physiotherapists in the Western World. Study number one is the first published randomised controlled trial testing the efficacy of the McKenzie method for patients with long-term NSLBP. Results from the 260 patients included showed that the McKenzie method was at least equally effective as strengthening training, which is the generally recommended treatment for these patients. Furthermore, the results support the need for a classification system for this heterogeneous patient group.

In the second study, a systematic review of the existing classification systems of relevance to physiotherapy was conducted. It was concluded, that several systems compete to be generally accepted within the physiotherapy profession. In a few of these, studies were published demonstrating some level of reliability. However, none were able to document their superiority over others regarding validity, i.e. their ability to identify subgroups of patients with NSLBP that would benefit the most from a particular treatment, compared with other treatments. There appears to be a need for further testing of the existing classification systems as well as for the development of alternative ones.

In the third study, a new diagnostic classification system including pathoanatomic and clinical categories, was presented. Selection of categories and criteria for categorisation of patients with NSLBP was based in part on the content of earlier systems, on the existing evidence regarding diagnostic value of criteria, and on the input from a conference of Danish experts. A systematic method was used to critically appraise the reliability, validity, feasibility, and generalizability of criteria for use in primary care. It was concluded that the new system has the potential to overcome some of the fundamental problems inherent in the existing ones, and that the new system may prove itself useful for research purposes. Further studies testing reliability and validity of the new system as a whole is warranted.

In the fourth study, the inter-tester reliability of the new classification system was evaluated. Four physiotherapists examined 90 patients with NSLBP. The results showed that trained examiners were able to obtain an acceptable level of agreement when classifying the same patients. However, the low prevalence of

positive findings in some categories indicates that there is a need for further testing of inter-tester reliability in a larger patient sample. Future studies investigating the validity and utility of the new classification system are required.

List of publications

This thesis is based on the following studies, which are referred to in the text by their roman numerals.

- I.** Petersen T, Kryger P, Ekdahl C, Olsen S, Jacobsen S: The effect of McKenzie therapy as compared with that of intensive strengthening training for the treatment of patients with subacute or chronic low back pain. A randomized controlled trial. *Spine*, 27, 1702–1709, 2002.
- II.** Petersen T, Thorsen H, Manniche C, Ekdahl C: Classification of nonspecific low back pain. A review of the literature on classifications systems relevant to physiotherapy. *Physical Therapy Reviews*, 4, 265–281, 1999.
- III.** Petersen T, Laslett M, Thorsen H, Ekdahl C, Manniche C, Jacobsen S: Diagnostic classification of non specific low back pain. A new system integrating pathoanatomic and clinical categories. Accepted for publication in *Physiotherapy Theory and Practice*, December issue, 2003.
- IV.** Petersen T, Olsen S, Laslett M, Thorsen H, Ekdahl C, Manniche C, Jacobsen S: Inter-tester reliability of a new diagnostic classification system for patients with non-specific low back pain. Submitted to *Australian Journal of Physiotherapy*, 2003.

Some additional data, not previously published, have been included in the results.

Reprints of papers was made with permission from the publishers.

Abbreviations

ANR	Adherent Nerve Root Syndrome
ANT	Adverse Neural Tension Syndrome
AP	Abnormal Pain Syndrome
Dys	Dysfunction Syndrome
IRD	Irreducible Disc Syndrome
LBP	Low Back Pain
MP	Myofascial Pain Syndrome
NMD	Non-mechanical Disc Syndrome
NRC	Nerve Root Compression Syndrome
NRE	Nerve Root Entrapment Syndrome
NSLBP	Non-specific Low back Pain
Post	Postural Syndrome
RD	Reducible Disc Syndrome
SIJ	Sacroiliac Joint Syndrome
SS	Spinal Stenosis Syndrome
ZJ	Zygapophysial Joint Syndrome

Introduction

Pain in the low back area is a common reason for seeing a physiotherapist in primary care. The goal for the therapist managing these patients is to select the appropriate treatment for each patient. The clinical reasoning process required to achieve this goal starts with a diagnostic classification that place the patient into a recognisable group with a particular pattern of signs and symptoms. The medical professions in primary care most commonly classify these patients with patho-anatomically labelled categories. However, there appear to be a wide diversity in the opinion as to the patterns of signs and symptoms that constitute a category.⁵⁵

A mantra of the last ten to twenty years in physiotherapy management of patients with low back pain (LBP), has been the biopsychosocial approach¹²¹ reflecting the fact that disability related to LBP must be viewed as a multifactorial problem. However, there may have been an overemphasis on the psychosocial part of the biopsychosocial model in this period. It appears that the time has come to put more research effort into investigating how psychosocial factors interact with particular physical diagnostic categories.¹²²

This thesis concerns the exploration of the “bio” part of the biopsychosocial model inasmuch as it tries to meet the challenges described above regarding classification and treatment of primary care patients with LBP.

Prevalence

Disability related to LBP is a major problem in the Western World. About 60–65% of the Nordic population is likely to experience LBP during their lifetime and 45–55% of adults will experience pain within a 12 month period.⁶⁴ Studies from a variety of countries investigating the long-term course of LBP show that most patients will improve rapidly.⁸⁸ Further improvement is apparent until about three months. Thereafter levels for pain, disability, and return to work remains almost constant. Six months after an episode, 60–70% of patients will have experienced relapses of pain and 16% will be sick-listed. As much as 62% will still be experiencing pain after 12 months.^{43;88}

Definitions and diagnosis

Low back pain (LBP) is usually defined as pain, aching, or discomfort localised

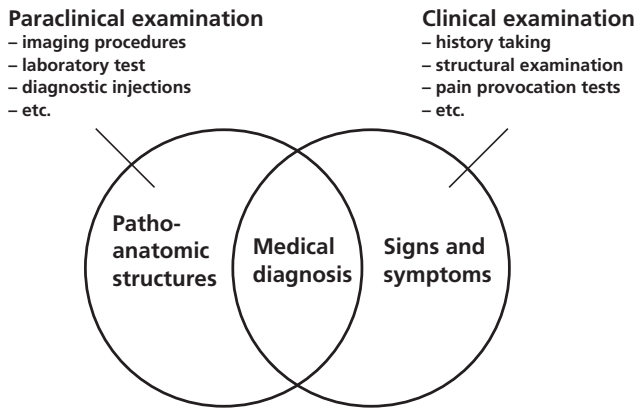


Figure 1. A diagnostic model. Establishing a medical diagnosis requires a consistency between known clinical examination findings and pathological structures or pathophysiological mechanisms identified by paraclinical methods.

in the area between the costal margin and the gluteal folds.^{1;43;59;117} Sciatica is defined as pain radiating from the low back to one or both legs.

Acute LBP can be defined as pain in the low back and/or sciatica of no more than 6 weeks duration. Subacute pain is defined as pain between 6 and 12 weeks in duration. Chronic pain is defined as pain of at least 12 weeks duration.^{84;117;120}

The term LBP refers to a large heterogeneous group of different clinical and etiological entities. Thus, LBP refers to a pattern of symptoms rather than a diagnosis.

Figure 1 illustrates a diagnostic model. Establishing a medical diagnosis requires a consistency between known clinical examination findings and pathological structures identified by paraclinical methods.²⁸ If this is achieved, a medical diagnosis such as herniated disc, spinal stenosis, spondylololsthesis, fracture etc is established. The clinician is then in a position to suggest a treatment strategy and offer a prognosis. If not, there is either a structural abnormality, which has no necessary relation to the patient's complaint, or a clinical pattern that cannot be linked to a specific pathoanatomic structure. In the search for clinically meaningful ways of identifying subgroups of LBP, various methods of labelling patients by syndromes have been attempted, and there have been attempts at identifying painful structures using patterns of signs and symptoms. A syndrome is defined as "The aggregate of signs and symptoms associated with any morbid process that together constitute the picture of a known disease".⁷ Examples of commonly used labels are facet syndrome, sacroiliac syndrome, myofascial pain syndrome, or nerve root compression syndrome.^{14;33;57;81;90}

LBP is difficult to diagnose patho-anatomically.^{28;83;123} One reason for this could be that a precise diagnosis is based on procedures not available in most clinical settings e.g. roentgenogram, computerised axial tomography, or magnetic resonance imaging. In addition, abnormal findings illustrated by advanced imaging techniques are common in asymptomatic individuals.^{50;91;126}

Consequently, the problem of diagnosis is a matter of controversy within the community of LBP researchers. In primary care, it has been estimated that a patho-anatomic cause is found in 15–25% of patients,²⁶ whereas others claim that up to 70% of patients may be diagnosed where discography and other diagnostic injection procedures are available.¹⁵ In an attempt to reach a consensus, most international guidelines for the management of musculoskeletal LBP recommend an initial diagnostic classification process, a diagnostic triade, that differentiates between possible serious spinal pathology, nerve root problems, and non-specific LBP.⁵⁸ The diagnostic label of non-specific 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.

Treatment

The three most recent consensus reports for the treatment of patients with acute non-specific LBP recommend reassurance of the patient, provision of adequate information, advice to stay active, prescription of pain medication (when necessary), consideration of spinal manipulation, and consideration of multidisciplinary treatment programs for workers with sick leave for more than 4–8 weeks.^{2;84;117}

For patients with subacute or chronic non-specific LBP, maintenance or progressive resumption of activities of daily living, and back programs combining strength training, stretching and fitness, is recommended as an effective treatment for the reduction of disability and improvement of physical function.^{2;84} The latest published systematic review from the Cochrane Collaboration concluded that there is strong evidence that exercise therapy is more effective than usual care by general practitioners, and that exercise therapy and conventional physiotherapy (consisting of a combination of hot packs, massage, traction, mobilization, short-wave diathermy, ultrasound, stretching, flexibility and coordination exercises, electrotherapy) are equally effective for the treatment of these patients. However, the authors concluded that it is still unclear whether exercise therapy is more effective than inactive treatment (consisting of semihot packs and light traction, hot packs and rest, detuned ultrasound and detuned short-wave diathermy). It remains unclear whether any specific type of exercises (flexion, extension, or strengthening exercises) is more effective than another.¹¹⁸

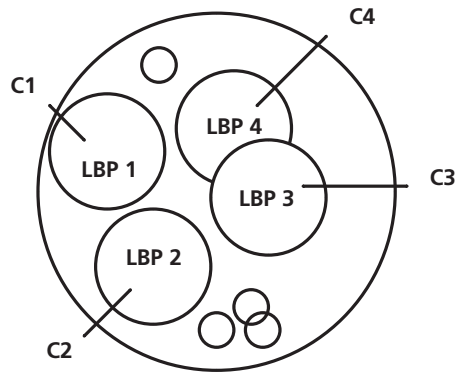


Figure 2. Non-specific LBP (large circle) may consist of a number of largely unidentified sub-entities (smaller circles LBP1, LBP2, etc.), each having its own set of causal mechanisms (C1, C2, etc.)

The McKenzie-method of Mechanical Diagnosis and Therapy (MDT) is widely used within primary care. It has been reported as the most commonly used method by physiotherapists for the management of patients with LBP.^{9;35;41} The evidence supporting the effectiveness of the method is inconclusive. Results of randomised controlled trials including patients with acute non-specific LBP showed better outcome from the McKenzie-method when compared with a mini-back school.^{109;110} However, no differences in outcome were found when the McKenzie-method was compared with chiropractic manipulation or an educational booklet,²⁰ back massage and general back care advice⁴⁰, or usual general practitioner care,¹¹⁴ and poorer outcome was found in comparison with manipulation and general exercises.^{22;32} No randomised controlled trials have been published investigating the effects of the McKenzie method for subacute and chronic LBP patients.

Heterogeneity in outcome studies

Recently, several randomized controlled trials of high methodological quality have shown that physiotherapy has no demonstrable benefits over the natural history in patients with acute non-specific LBP.^{20;48;71;97} Furthermore, a great number of high quality studies have shown little or no differences between various physiotherapy treatments for acute, subacute or chronic cases.^{6;11;18;21;40;42;45;46;78;89;106;113;114} An obvious explanation could be that there actually is no difference. However, several authors have suggested that these results may reflect the heterogeneity of the non-specific LBP group.^{17;62;65;66;96} The “subgroup concept” is illustrated in Figure 2.

It is likely that non-specific LBP consists of several distinct subgroups each with its own causal mechanisms and thus with its own potential set of beneficial treatments. Patients within a subgroup for which a specific treatment is of bene-

fit, may be rendered statistically invisible by the number of patients in subgroups for which actual harm or no demonstrable benefits occur from that treatment.

The problem is recognised in an “Agenda for Future Primary Care Research on Low Back Pain”, which was developed by participants in the multi-disciplinary International Forum for Primary Care Research on Low Back Pain, Washington, 1995.¹⁶ The number one priority for the agenda was to determine whether different varieties, natural courses, or subgroups of LBP could be identified and, if so, to establish criteria that could be used to differentiate them. There was a particular interest in establishing a classification system that would allow practitioners to determine: “1) how subgroups differed in terms of natural course, and 2) whether treatment and management strategies could be tailored to each subgroup”

Methodological requirements for classification systems

The usefulness of classification systems in research, and of a clinical classification system in particular, is dependent on its ability to meet basic requirements regarding appropriateness for purpose, content validity, face validity, feasibility, construct validity, reliability, and generalizability.¹⁹

The purpose of the system, the target population, and the setting appropriate for the use of the system must be clearly specified.

Content validity refers to the extent to which the system covers the domain of interest intended to be described, considering the purpose.

Face validity is generally a matter of the extent to which a measurement appears (in the opinion of experts) to be reflecting the domain of interest. Regarding classification systems, face validity refers to whether the criteria used for categorisation in the system are appropriate in terms of definitions, labelling, and based on evidence, i.e. demonstrated to have reliability and validity.

Feasibility is based on an appraisal of whether the system is simple to understand, easy to perform, relies on clinical examination, and whether special skills or tools are required.

Construct validity is based on a theoretical understanding of the domain of interest. It reflects whether the system discriminates between entities that are thought to be different in a way appropriate for the purpose and whether it performs satisfactorily when compared to other classification systems that classify the same domain.

Reliability is defined as the degree to which a system provides consistent results when classifying the same condition, especially whether intra-observer and inter-observer agreement are satisfactory.

Generalizability is a matter of whether the system has been used in other studies and/or settings.

Classification systems in low back pain

Classification systems are defined as devices for sorting the complex elements of reality into reasonable and logical entities.⁷² Ideally, the objectives of a diagnostic classification are to find a label that indicates the cause of the disease, predicts outcome, predicts responses to specific therapies, and which can be used to describe the disease in communicating experience or research.

The challenge for health professionals in primary care where most LBP patients are managed is that diagnostic imaging procedures often fail to provide a meaningful diagnosis^{10;119} and generally there is no access to discography and diagnostic injections in routine clinical practice.

Several classification systems have been proposed for subdividing non-specific LBP patients by means of clinical examination. In physiotherapy, three 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.^{23;73;79} All three are treatment-oriented systems in that they place patients in categories with the purpose of determining an appropriate intervention.

The classification system proposed by McKenzie⁷⁹ is based on information from history taking, and symptom response to patient or therapist generated loading of the lumbar spine. It has been reported as the most commonly used system by physiotherapists.^{9;41} The system as a whole has been tested for reliability and has substantial inter-tester agreement according to the criteria of Landis and Koch⁶¹ when applied by trained examiners (Kappa coefficients ranging from 0.6 to 0.7)^{56;92}. A number of studies have supported the validity of the system's ability to predict outcome of treatments with McKenzie-therapy or active rehabilitation for patients whose symptoms have centralised, i.e. abolished from their most distal location, following the McKenzie examination procedures.^{30;52;68;111;124;125} Randomised controlled trials investigating treatment-related validity of the McKenzie system, i.e. ability to categorise patients in a way that might result in selection of the most effective treatment, have shown conflicting results.^{20;22;32;40;89;109;110}

Delitto et al.²³ have developed a classification system for patients with acute LBP. The system classifies patients into four main categories using information gathered from history taking and clinical examination. Intertester reliability of single categories has been questioned,^{24;95} however, the system as a whole has been shown to have a moderate intertester reliability (Kappa coefficient 0.56)³⁸. One epidemiological study has shown differences in short term prognosis among a few of the categories in the system when collapsed into four overall treatment categories.³⁸ Two randomised controlled trials have, with regard to choice of treatment, shown validity of one (the extension-mobilization category) of the seven categories of the system.^{22;32} In addition, a recently published trial has shown that treatment based on the classification system as a whole was more

beneficial for acute non-specific LBP patients than treatment based on clinical practice guidelines.³⁶

Sahrman et al.⁷³ have developed a classification system comprising five categories based on testing of muscular stability, alignment, asymmetry, and flexibility of the lumbar spine, pelvis, and hip joints. Of particular interest to the system is the recording of movements and activities in daily functioning that provoke the patients familiar symptoms. 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).¹¹⁵ However, there are no reports on reliability in classification of the patients into the five categories. The use of the system has been illustrated by a case report.⁷³ A recent study showed that the majority of the included patients with symptom provocation during testing experienced a short-term decrease in the symptoms after the provoking movements and alignments had been modified.¹¹⁶ However, no data have been published supporting or refuting the validity of the system concerning its ability to categorise patients in a way that might result in selection of the most effective treatment.

To summarise, although recent data on reliability and validity have been published supporting usefulness of some of these systems, none has clearly 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.

In this author's opinion, there are several fundamental problems concerning the existing treatment oriented classification systems. It appears that the developers of those systems are inclined to search for the patterns of signs and symptoms that they are able to treat. As a consequence, labels and criteria used for classification in those systems differ according to the treatment methods preferred by the developers, communication across systems is hampered, 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 a 'posterior derangement syndrome' in the system proposed by McKenzie,⁷⁹ an 'extension syndrome' in the system proposed by Delitto et al.,²³ and a 'flexion category' in the system proposed by Sahrman et al.⁷³

It has been pointed out that various practitioners may have different, but equally acceptable approaches to the management of a particular treatment-oriented diagnostic category.¹⁴ Therefore, instead of making the diagnostic system fit the treatment system preferred by the developers, in this other's opinion, it should be the other way around. Once a generally accepted diagnostic classification system has been developed, it should be the results of research that determines the most effective treatments for particular categories of patients. A classification system with a pathoanatomic orientation might be the answer to some of the above-mentioned problems.

Table 1. Overview of the arguments for and against a clinical pathoanatomic oriented classification system

Disregarded in the literature because of **theoretical arguments**

- Not necessary for a physical therapist to identify pathology.^{4;25;128}
- No clear correlation between imaging findings and symptoms.^{10;119}
- Focussing on pathology increases fear-avoidance beliefs and sick-role in patients.^{10;26;121 (p.152);128}

Ought to be regarded because

- Patients expect an explanation of an anatomical cause when seeking care.^{80;98}
- Neglecting patients' expectations has implications for patient satisfaction²⁷ as well as for the outcome of treatment.^{8;51;54;69;112}
- Fear-avoidance is only related to uncertainty, not to identification of pathology.^{54;121p.192}
- Awareness of the connection between anatomical cause, response to tests, and disability in daily living empowers and motivates the patient towards an active approach.^{60;86;107}
- A pathoanatomic system is a prerequisite that hypotheses for evaluating prognosis and efficacy of treatments can be supported by anatomically grounded theory.
- A pathoanatomic system enabling researchers to test hypotheses as to whether treatment methods, e.g. McKenzie or orthopaedic medicine, influences the conditions on structural level that they assume, is required.
- When communicating the results of interventions to colleges or other medical professionals, a more universal clinical classification, independent of schools of thought and meaningful to the medical community, is preferable.

Disregarded in the literature because of **data**

- Former pathoanatomic classification systems are not supported by scientific data. On the contrary, data exists that refute validity.^{31;49;70;74;75;99;100;103-105;108}

Ought to be regarded because

- Recent data support validity of several pathoanatomic categories.^{5;29;37;63;93;94}
-

Table 1 summarises the arguments for and against a pathoanatomic oriented classification system that are discussed in more detail in Paper III.

Objectives

The overall aim of this thesis was to evaluate the effects of the McKenzie method for the treatment of patients with LBP, and to develop and evaluate a new LBP classification system of relevance to physiotherapy.

The specific aims were:

- to investigate the effect of McKenzie mechanical diagnosis and therapy compared with that of intensive dynamic strengthening training in a randomised controlled trial including patients with subacute or chronic non-specific LBP,
- to review the literature on reliability and validity of existing classifications systems for non-specific LBP patients,
- to develop a new classification system for patients with non-specific LBP integrating pathoanatomic and clinical categories of relevance to physiotherapy,
- to provide the rationale behind the new system based on existing evidence,
- to evaluate the inter-tester reliability of the new system in a clinical setting.

Designs and Methods

Effects of McKenzie-therapy (Paper I)

A prospective randomised controlled trial comprising 260 patients with subacute or chronic non-specific LBP was carried out. The full protocol was completed by 180 patients. The patients were randomized to either a McKenzie group or a strengthening training group. The treatment period in both groups was 8 weeks at an outpatient clinic followed by 2 months of self-training at home. Treatment results were recorded at the end of the treatment period at the clinic, and 2, 8, and 14 months after. An intention-to-treat analysis of the main effect variables, disability and pain, was performed on all patients included in the study. A supplementary analysis of 180 patients who completed the full treatment program was also performed. Secondary effect variables, i.e. return to work, use of pain medication, the patients own perception of change in back-related quality of life, and number of patients visiting a general practitioner because of LBP during follow-up, were recorded.

Treatment programs

Treatment according to the McKenzie method for Mechanical Diagnosis and Therapy was planned individually following an initial physical assessment according to the principles described by Robin McKenzie.⁷⁹ The essence of the treatment is performance of selfmobilising repeated movements or sustained positions in specific movement directions, and the application of manual overpressure, and/or mobilisation by the therapist.⁷⁹ Initial assessment and treatment lasted one hour, and subsequent treatments typically lasted half an hour. Intervals between treatments were at the discretion of the therapist. All seven physiotherapists performing the treatment had completed the course program A to D arranged by the McKenzie Institute International, and six of the therapists had attended and passed a credentialling examination in the method during the period of this study from August 1996 to December 1998.

The strengthening training was carried out in groups of six patients under the guidance of a physiotherapist. Sessions began with a period of 5 to 10 minutes riding a stationary bicycle. This was followed by low intensity warm up exercises for about 10 minutes comprising ten repetitions of low resistance exercises for the lumbo-pelvic muscles in flexion, extension, and rotation. After that, an intensive dynamic back strengthening training program was performed in flexion and extension. The training program was chosen because it had shown excellent outcome in chronic non-specific LBP patients compared to treatment with inactive physical modalities or low-intensity physical training.⁷⁶ This program

has become a common treatment for patients with subacute and chronic non-specific LBP in Denmark. The program has been described in detail elsewhere.⁷⁷ Four strengthening exercises were done in series of ten repetitions with one minute rest between each series. The number of repetitions were individually graded. The initial dosage was typically a total of 50 repetitions, whereas the dosage during the treatment period of 8 weeks was gradually increased to 100 repetitions. At conclusion of each session, patients performed 10 minutes of stretching exercises for the trunk and hip muscles. Training sessions lasted from 60 to 90 minutes, and were performed twice a week. All physiotherapists guiding the patients' performance of exercises had several years of experience in this type of group strengthening training.

In both treatment groups, patients received a maximum of 15 treatments for a period of 8 weeks. Missed sessions for a maximum of 30% were accepted. Patients were instructed to continue self-administered exercises at home or at a fitness centre for a minimum of 2 months after completion of the treatment at the clinic. Because most of the patients suffered from long lasting LBP we expected this period of self-administered exercises to be necessary for the patients to experience the full effect of the intervention. If a patient was totally free of symptoms some time before the maximum of 15 treatments was reached he/she was allowed to continue with self-administered exercises from that time. Patients were encouraged not to seek any other kind of physical treatment for the 2 months period of self-administered exercises.

Literature review on classification systems of relevance to physiotherapy (Paper II)

To explore the usefulness of existing classification systems, a systematic and critical review of the literature was done. The review was based on articles retrieved from a systematic search of the Medline and Embase databases for the eleven year period, January 1988 to December 1998, using the following keywords: back ache/back pain/low back pain in combination with classification, diagnostic tests, or physical examination. All English and Scandinavian language papers were selected. In addition a search by hand was carried out in five physiotherapy journals (Physical Therapy, Physiotherapy, Australian Journal of Physiotherapy, Physiotherapy Canada, and Physiotherapy - Theory and Practice) for the same eleven-year period.

A critical appraisal of the classification systems was performed using a systematic approach proposed by Buchbinder et al.¹⁹

Development of the new classification system (Paper III)

The development was based on the review of the literature (Paper II) and proposed criteria for categorisation presented in a recently published version of a classification.⁶² An initial version of the new 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. At the conference, minimum criteria for placing patients within particular categories were discussed and comments from the specialists were incorporated in the final version of the system.

The development phase followed three steps. In step one, pathoanatomic categories that could be derived from evidence were included, i.e. Reducible Disc Syndrome (RD), Irreducible Disc Syndrome (IRD), Non-mechanical Disc Syndrome (NMD), Nerve Root Compression Syndrome (NRC), Spinal Stenosis Syndrome (SS), Zygapophysial Joint Syndrome (ZJ), and Sacroiliac Joint Syndrome (SJ). In step two, additional categories widely assumed within the physiotherapy profession to be pathoanatomically oriented, i.e. Adherent Nerve Root Syndrome (ANR), Nerve Root Entrapment Syndrome (NRE), Myofascial Pain Syndrome (MP), and Adverse Neural Tension Syndrome (ANT) or indicated pain producing connective tissue, although not specific to certain anatomical structures, i.e. Postural Syndrome (Post) and Dysfunction Syndrome (Dys), were included. In step three, a category widely assumed to indicate that patient responses during clinical examination should be re-evaluated, i.e. Abnormal Pain Syndrome (AP), was included.

Inter-tester reliability of the new classification system (Paper IV)

Ninety subjects with mainly chronic non-specific LBP were examined by four physiotherapists, and the level of agreement among therapists regarding categorisation was evaluated. The therapists were trained in performing the examination procedures and the clinical reasoning required for categorisation of patients within the classification system. Within a pair, one therapist was the first examiner for half of the subjects and the other therapist was the first examiner for the remaining half of the subjects. Examination findings by the two examiners were recorded independently.

Statistical methods

In Paper I, non-parametric statistical methods, Wilcoxon Sign Test and Mann-

Whitney U test, were used. Chi-square tests were used for proportional data comparison. The level of statistical significance was set at $P \leq 0.01$

In Paper III, percentages of agreement and Kappa coefficients with 95% confidence intervals were calculated for each category and for mutually exclusive categories as a whole to estimate the level of inter-tester reliability of the classification system.

Summary of Results

McKenzie therapy versus strengthening training (Paper I)

The intention-to-treat analysis comprising 260 patients was performed on the main outcome variables, disability and pain. The results are presented in Figures 3 and 4. After adjustment for base-line values, there was a trend toward greater reduction in the disability score in favour of the McKenzie group at 2 months follow-up only ($P=0.04$). Changes in back and leg pain intensity did not differ between the groups after treatment at the clinic ($P=0.38$), at 2 months follow-up ($P=0.41$), or at 8 months follow-up ($P=0.83$).

A supplementary analysis comprising the 180 patients who completed the full intervention was performed on all outcome variables. After adjustment for base-line values, the only statistically significant difference between the groups was changes in back and leg pain. The McKenzie group showed a greater reduction in pain intensity than the strengthening training group at 2 months follow-up ($P=0.01$), but the differences were not significant at end of treatment at the clinic ($P=0.02$) or at 8 months follow-up ($P=0.16$). The supplementary analysis showed that there were no significant differences between groups at any follow-up for changes in disability, global change in back related quality of life, number of patients using pain medication because of LBP, number of patients on sick-leave, or number of patients visiting a general practitioner because of LBP during follow-up.

In both analyses, all trends and statistically significant differences in disability or pain exceeded the minimum criteria for clinical importance predefined in this study as 25 % difference between groups. In the intention-to-treat-analysis, the difference in change of disability at 2 months follow-up between groups was 122%. In the supplementary analysis, the differences in change of pain between groups were 33% at end of treatment and 83% at 2 months follow-up.

Subsequent to publication of Paper I, analysis of results from the 14 month follow-up has been completed. Results regarding group comparisons in main effect variables, disability and pain, are displayed in Figure 3 and 4. A total of 234 patients responded to the disability questionnaire and 233 patients responded to the pain questionnaire at follow-up (dropout rate of 10%). In the main outcome variables, the pairwise comparisons of differences within both of the treatment groups remained statistically significant after 14 months ($P<0.001$). In the McKenzie group the median disability score was 26.7 (10th and 90th percentile 7.1–53.8) and the median pain score was 12 (10th and 90th percentile 1–35). In the strengthening training group, the median disability score was 26.7 (10th and 90th percentile 6.7–60) and the median pain score was 14 (10th and 90th percentile

Disability scores

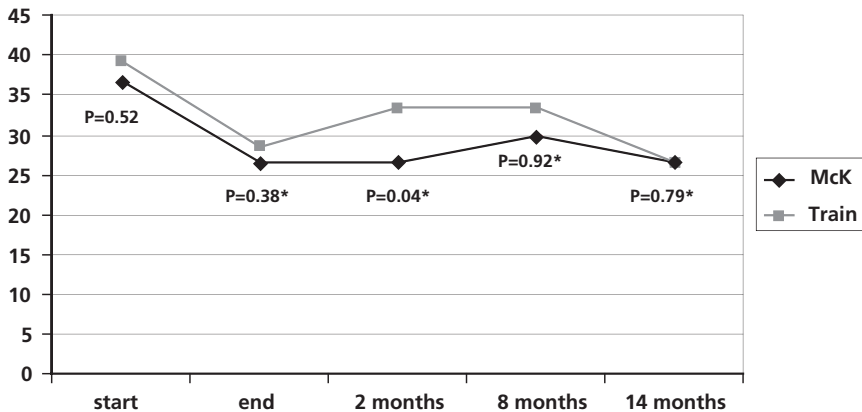


Figure 3. Disability scores. Intention-to-treat analysis.

* P value refers to analysis of between group differences adjusted for baseline scores at start of treatment. N=260 except for 14 months follow up: N=234.

Pain score

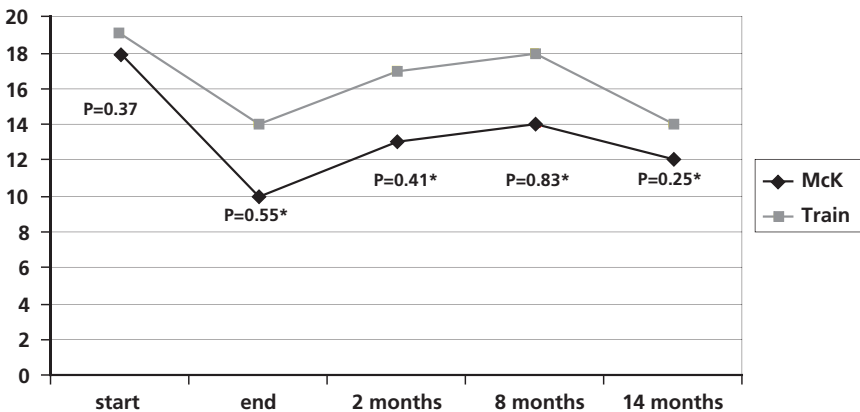


Figure 4. Back and leg pain scores. Intention-to-treat analysis.

* P value refers to analysis of between group differences adjusted for baseline scores at start of treatment. N=260 except for 14 months follow up: N=233.

2–29). After adjustment for baseline scores on disability and pain, no significant differences were found between groups (disability: P=0.79, pain: P=0.25).

Existing classifications systems for non-specific LBP patients (Paper II)

In the systematic review of the literature, eight systems were found to be of relevance to physiotherapy in as much as they subdivided non-specific LBP by means of symptoms and clinical tests. They were all treatment oriented, i.e. developed for the purpose of guiding choice of physiotherapeutic treatments. Following the critical appraisal within the framework proposed by Buchbinder et al.¹⁹, none of the classification system fulfilled all of the requirements and none were considered to have included all relevant categories separated in a way suitable for the purpose. Studies concerning reliability and validity were rarely reported. Generally aspects of validity and reliability were only tested for a few of the criteria used for categorising patients. Construct validity and reliability of the classification systems as a whole were not tested. The most promising treatment-oriented classification systems for non-specific LBP identified in the review were those developed by McKenzie⁷⁹ and Delitto et al.²³. It was recommended that future research should address the usefulness of existing classification systems as well as the development of new classification systems designed using commonly accepted methodological principles.

The rationale behind the new system based on existing evidence (Paper III)

In Paper III, the criteria for classifying patients in the new system were described. The tests included in criteria for the different syndromes in the system are illustrated in Figure 5-10.

The systematic framework proposed by Buchbinder et al.¹⁹ was used again to critically appraise issues of purpose, content validity, face validity, feasibility, construct validity, reliability, and generalizability of the new classification system. Results from studies examining elements of the 12 categories of the classification system were reported to allow the appraisal to be made. An overview of the critical appraisal is presented in Table 2.

For some of the categories, i.e. RD, IRD, NRC, SS, SJ, MP, and ANT, there was evidence suggesting the criteria to have reliability and validity, i.e. that the proposed criteria are reproducible and are able to identify a specific symptomatic structure with an acceptable degree of accuracy. In AP, data suggest that criteria are reproducible and are able to identify patients whose pain behavior is disproportionate to the underlying structural pathology. In the other categories, future evidence regarding reliability and validity has yet to be established.

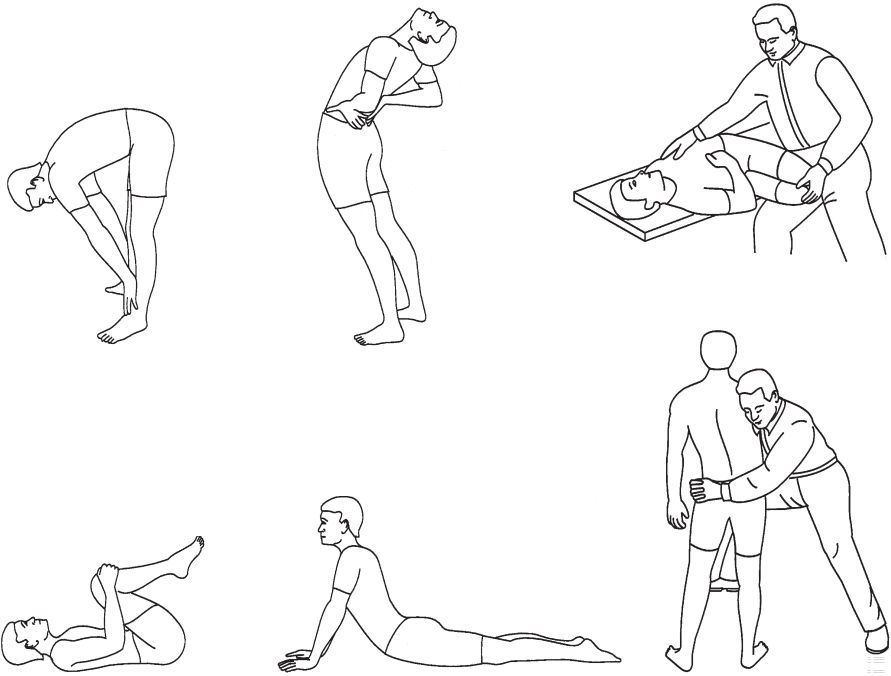
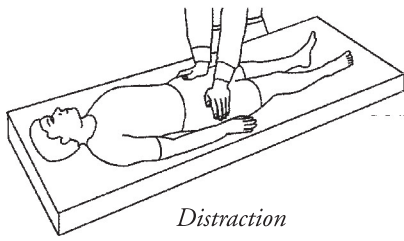


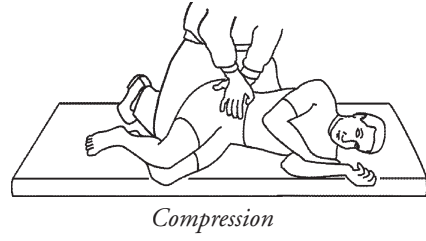
Figure 5. Examples of tests included as criteria for Disc Syndrome, Adherent Nerve Root Syndrome, Nerve Root Entrapment Syndrome, Postural Syndrome, and Dysfunction Syndrome.



Figure 6. Example of sustained end range loading included in criteria for Postural Syndrome.



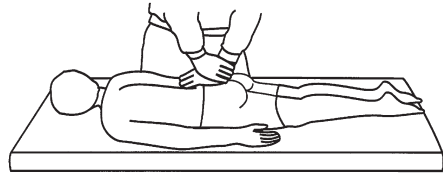
Distraction



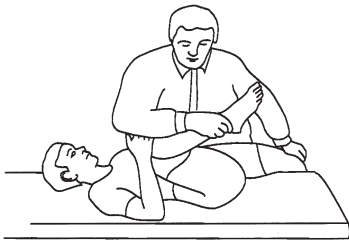
Compression



Thigh thrust (posterior shear)

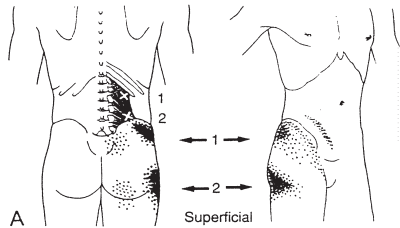


Sacral thrust

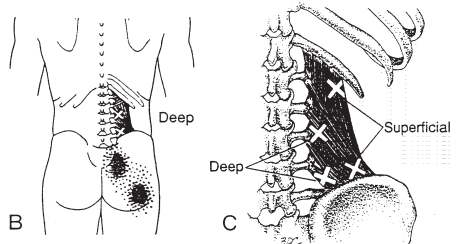


Pelvic torsion (Gaenslen's test)

Figure 7. Tests included in criteria for Sacroiliac Joint Syndrome.



A

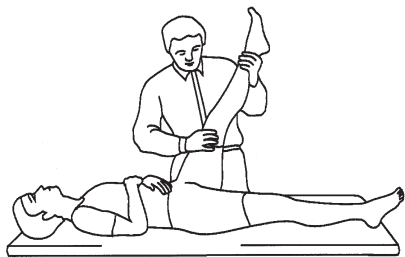


B

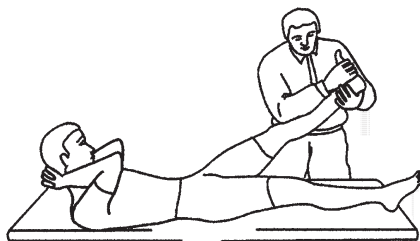
C

Quadratus lumborum

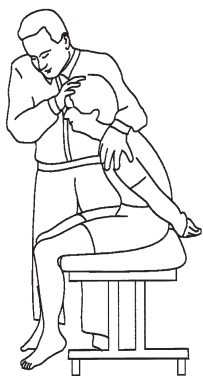
Figure 8. An example of pain referral in a characteristic area on palpation of a painful myofascial trigger point. Included in criteria for Myofascial Pain Syndrome.



Straight leg raise test



Straight leg raise test with neck flexion and ankle dorsal flexion



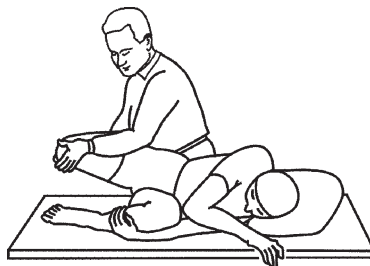
Slump test with lumbar flexion



Slump test with lumbar flexion, neck flexion, knee extension, and ankle dorsal flexion

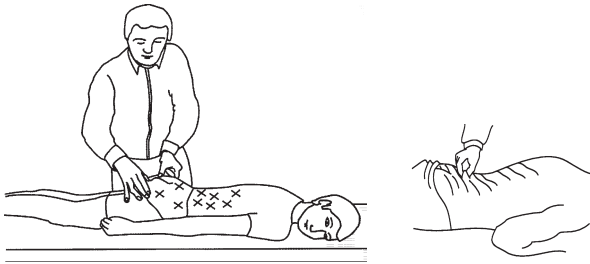


Femoral nerve stretch test

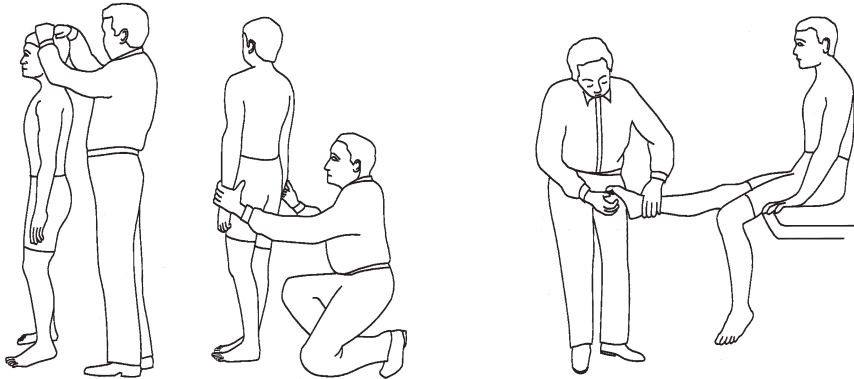


Femoral nerve stretch test with neck flexion

Figure 9. Examples of stages in the testing included as criteria for the Adverse Neural Tension Syndrome.



Widespread superficial or non-anatomic tenderness



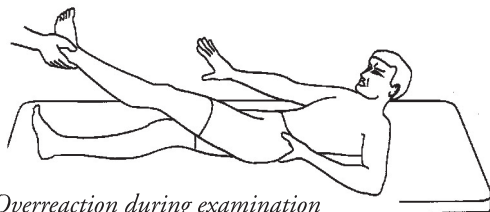
Pain provocation on axial loading or simulated rotation of the back

Straight leg raise improved at least 30 degrees with distraction



Regional muscle weakness or sensory disturbances in non anatomic distribution

Figure 10. Examples of tests included in criteria for Abnormal Pain Syndrome.



Overreaction during examination

Table 2. Critical appraisal of the new classification system.

ITEMS	JUDGEMENTS RELATED TO THE DIFFERENT ITEMS
Purpose	
Is the purpose, population, and setting clearly specified?	Yes. To identify clinically homogeneous subgroups of non-specific LBP patients in primary care according to the assumed symptomatic anatomical structures.
Content validity	
Is the domain and all specific exclusions from this domain clearly specified?	Yes. Non-specific LBP.
Are all relevant categories included?	Yes. Comprises all possible pain producing anatomical structures.
Is the breakdown of categories appropriate, considering the purpose?	Yes. Intervertebral Disc is most common structural source of symptoms. Is tested first in the system in order to remove false positives before testing for other structures. Symptomatic vertebral structures rarely co-exist. Remaining categories widely assumed to refer to anatomical structures.
Are the categories mutually exclusive?	Yes, regarding cat. 1-9. Cat. 10-12 may co-exist with all others.
Was the method of development appropriate?	Yes. Developed by experienced physiotherapists. Based on literature review and conference of experts.
If multiaxial, are criteria for content validity satisfied for each additional axis?	Not relevant. No additional axes.
Face validity	
Is the nomenclature used to label the categories satisfactory?	Yes. Cat. 1-6, 8, and 10-11 are labelled according to supposed origin of pain. Cat. 7 and 9 are labelled according to connective tissue not specified. Cat. 12 are labelled according to inappropriate pain behavior.
Are the terms used based upon empirical (i.e. directly observable) evidence?	Yes. Physical examination.
Are the criteria determining inclusion into each category specified? If yes, do these criteria appear reasonable?	Yes. Clinical tests, specified for each category.
Have the criteria been demonstrated to have validity and/or reliability?	Yes, regarding cat. 1,4,5,8, and 12. Partly, regarding cat. 10 and 11. No, regarding cat. 2, 3, 7, and 9. Undetermined in cat. 6.
Are the definitions of criteria clearly specified?	Yes.
If multiaxial, are criteria of face validity satisfied for each additional axis?	Not relevant. No additional axes.

<p>Feasibility Is classification simple to understand? Is classification easy to perform? Does it rely on clinical examination alone? Are special skills, tools, and/or training required? How long does it take to perform?</p>	<p>No. Quite complex clinical reasoning. No. Training required. Yes. Yes. Training in McKenzie method required. A maximum of one hour.</p>
<p>Construct validity Does it discriminate between entities that are thought to be different in a way appropriate to the purpose? Does it perform satisfactorily when compared to other classification systems which classify the same domain?</p>	<p>Primary purpose: Partly (yes, regarding cat. 1, 6, and 8). Secondary purpose: Mixed results regarding cat. 1-3, 7, and 9; No reports regarding cat. 4, 5, 6, and 8; inconclusive results in cat. 10 and 11; yes, regarding cat. 12. Not tested yet.</p>
<p>Reliability Does the classification system provide consistent results when classifying the same conditions (test-retest)? Is the intra-observer and inter-observer reliability satisfactory?</p>	<p>Not reported. Partly (data suggest acceptable reliability for cat. 1, 2, 3, 7, and 9).</p>
<p>Generalizability Has it been used in other studies and/or settings? Cat.: category. See text in Paper III for references.</p>	<p>Partly (yes, regarding cat. 1, 2, 3, 7, and 9).</p>

Table 3. Agreement between therapists regarding categorisation of patients in mutually exclusive categories

Syndrome	RD	IRD	NMD	ANR	NRE	NRC	SS	ZJ	Post	SIJ	Dys	Other	Total
RD	36	1	0	0	0	0	0	0	0	1	0	3	41
IRD	0	2	0	0	0	0	0	0	0	0	0	0	2
NMD	1	0	3	0	0	0	0	0	0	0	0	1	5
ANR	2	0	1	0	0	0	0	0	0	0	0	0	3
NRE	1	0	0	0	0	0	0	0	0	0	0	0	1
NRC	0	0	0	0	0	5	0	0	0	0	0	1	6
SS	0	0	0	0	0	0	1	0	0	0	0	1	2
ZJ	0	0	0	0	0	0	0	2	0	0	0	0	2
Post	0	0	0	0	0	0	0	0	1	0	0	0	1
SIJ	2	0	0	0	0	0	0	0	0	8	0	2	12
Dys	0	0	0	0	0	0	0	0	0	0	1	0	1
Other	2	2	1	0	0	0	0	0	0	2	1	6	14
Total	44	5	5	0	0	5	1	2	1	11	2	14	90

RD = Reducible Disc, IRD = Irreducible Disc, NMD = Non-mechanical Disc, ANR = Adherent Nerve Root, NRE = Nerve Root Entrapment, NRC = Nerve Root Compression, SS = Spinal Stenosis, ZJ = Zygapophysial Joint, Post = Postural, SIJ = Sacroiliac Joint, Dys = Dysfunction.

Table 4. Agreement between therapists regarding categorisation of patients in other categories

Examiner 1		Examiner 2		Total
		+	-	
MP	+	45	10	55
	-	13	21	34
	Total	58	31	89
ANT	+	63	5	68
	-	8	14	22
	Total	71	19	90
AP	+	5	5	10
	-	2	78	80
	Total	7	83	90
Inconclusive	+	1	3	4
	-	2	84	86
	Total	3	87	90

MP = Myofascial Pain Syndrome, ANT = Adverse Neural Tension Syndrome, AP = Abnormal Pain Syndrome.

Inter-tester reliability of the new classification system (Paper IV)

The physiotherapists participating in the study ranged in age from 37 to 51 years (mean 41.5), and their years of clinical experience in management of LBP ranged from 7 to 27 years (mean 14.5). Table 3 summarises the descriptive statistics for the agreement between therapists regarding the 11 syndromes that are mutually exclusive in the classification system (RD, IRD, NMD, ANR, NRE, NRC, SS, ZJ, Post, SIJ, and Dys). Table 4 summarises the descriptive statistics for the categories that each can coexist with the mutually exclusive categories (MP, ANT, and AP).

Among mutually exclusive syndromes, RD was the most common (46%) followed by SIJ (13%), NRC (7%) and NMD (6%). IRD, Dys, ANR, NRE, SS, ZJ, and Post were less common, each with prevalence ranging between 1 and 3%. The therapists classified 14 of the patients (16%) in the 'Other' category. Most of these patients were later classified into one or more of the remaining three syndromes (MP, ANT, and AP), leaving six patients in the final 'Inconclusive' category by one or both of the examiners.

Percentage of agreement ranged from 74% to 100% and Kappa coefficients ranged from 0.26 to 1.00. The overall percentage of agreement between examiners for the 11 mutually exclusive syndromes and the 'Other' category (N=90), was 72% and the Kappa coefficient was 0.62 (95% CI 0.50–0.74). An alternative analysis of the 11 syndromes without the 'Other' category (N=68) showed a percentage of agreement of 86.8% and a Kappa coefficient of 0.79 (95% CI 0.66–0.92). Regarding the three syndromes, MP, ANT, and AP, percentage of agreement ranged from 74% to 94% and Kappa coefficients ranged from 0.44 to 0.59.

Discussion

The aims of this thesis were to investigate the effects of the McKenzie method for the treatment of patients with LBP, and to develop and evaluate a new LBP classification system of relevance to physiotherapy.

One part of this thesis showed that for patients with long-term non-specific LBP, the McKenzie method was at least as effective as strengthening training. It was suggested that designs of randomised controlled trials might be improved if groups of patients that may benefit from different treatment methods were identified in a classification process.

In another part of this thesis, the studies suggested that different pathoanatomic oriented categories of non-specific LBP patients could be distinguished and that the inter-examiner reliability of a new classification system was acceptable.

Overall, the results present new knowledge of relevance to researchers and clinicians dealing with LBP patients in primary care.

Treatment (Paper I)

The main limitation in the design of this randomised controlled trial is the lack of a no-treatment control group. Therefore, it was not possible to estimate the overall influence of the regression-to-the-mean phenomenon or the contribution of natural healing to the outcome. All patients were referred by general practitioners to a hospital clinic for examination and treatment of a persistent LBP problem and therefore it was unacceptable to both the referring doctors and the patients to be randomised to a no-treatment group. Considering that almost all of the patients included had a chronic problem, we would not expect these factors to be of major importance.

A methodological discussion is included in Paper I and will not be reiterated here. The focus in this discussion will be on alternative explanations for the results. There are at least six possible explanations for the lack of a marked difference in efficacy between of the two types of treatment.

First, there actually might be no difference in reality. LBP of some duration is a fluctuating condition and possible differences between treatments may be overshadowed by natural variation. Consequently, there is a high probability of recurrence of the LBP problem sooner or later. Thus, the general picture will be that many different treatment methods might have a certain effect but none will be able to cure the problem. In the present study, the lack of a no-treatment

group makes it impossible to assess the natural course. However, this explanation is supported by the wave-like shape of the curve in Figures 3 and 4.

Second, any type of intensive exercise program that gives patients the experience of expanded limits to their physical functioning, may provide them with a method that increases the feeling of control over the pain, thus inhibiting negative pain behaviour related to the LBP problem. This explanation is supported by recent studies providing evidence that treatment programs containing active exercises are equally effective for patients with subacute or chronic LBP, irrespective of the type of exercises that have been compared.^{11;12;53;67;78} A thorough physical examination, patient education in coping strategies, and the provision of information encouraging the patient to increase activity and reducing fear has been shown to be effective as the sole intervention for these patients.^{47;48} The challenge for the physiotherapy profession in the future is to prove that specific treatments can provide additional therapeutic value over and above these interventions.

Third, the possibility exists, that our findings are a result of a Type II error. In the pre-study calculation of sample size we ran the risk of making a Type II error at a level of 10%. Thus, although the risk of Type II error is low in this study, it cannot be excluded.

Fourth, contamination of the intervention might have happened if treatments were not performed as intended. To minimise this possible confounding influence we chose therapists experienced in the type of treatments that they performed, excluded patients whose proportion of missed visits exceeded 30%, and encouraged patients not to seek other treatments for the two months period of self-training.

Fifth, outcome measurements might not have been sensitive enough to detect a difference. Post hoc analysis of the data showed that patients indicating an outcome of "better" or "much better" on a five point Likert scale, scored an average change of at least 10% on the disability and pain scales. Thus, the scales measuring the main outcome variables, disability and pain, appeared to be sufficiently sensitive to change.

Finally, an explanation for the findings might be that the patients included were selected from a heterogeneous group of non-specific LBP. In heterogeneous samples of patients, subgroups for which a specific treatment, i.e. the McKenzie method or strengthening training, has been of benefit may be masked by subgroups for which no demonstrable benefits occur. It would be a logical assumption that subgroups of patients with different pathoanatomic causes for their symptoms e.g. intervertebral discs, nerve roots, or sacroiliac joints, might need different types of treatment to achieve an optimal outcome. Therefore, if these subgroups could be identified by a pathoanatomic oriented classification system, it would increase the researchers ability to distinguish patient characteristics predictive of effects of different types of treatments, and to distinguish clinically

relevant subgroups for testing the hypotheses of treatment effectiveness.

Our results need to be replicated by new studies with these limitations in mind.

Development of categories and criteria of the new classification system (Paper II and III)

Since the completion of the systematic review in Paper II, additional studies have been published investigating aspects of reliability and validity of the classification systems proposed by McKenzie and Delitto et al. Furthermore, data on the system recently proposed by Sahrman et al. has been published. The recent information, which is summarised in the Introduction section of this thesis, does not substantially alter the conclusions regarding the need for further investigation of these three promising treatment oriented systems as well as the need for development of new ones that might overcome some of the fundamental limitations inherent in the existing ones.

The strengths of the new classification system are outlined in the discussion section of Paper III. To summarise:

- There are reasons to believe that the new classification system has the potential to become generally accepted within the physiotherapy profession inasmuch as it comprises all anatomical structures known to be potential pain generators. It is based on best evidence (where available) or widely recognised assumptions (where evidence is not available). Data indicate that currently, pathoanatomic categories are most commonly used by the medical professions involved in the management of patients with LBP in primary care.⁵⁵ Apparently, the ability to determine whether treatment should be directed to the discs, sacroiliac joints, muscles, nerve tissues etc. has meaning to all medical professions regardless of the school of thought.
- In research, hypotheses concerning efficacy of treatments supported by anatomically grounded theory can be tested on homogeneous groups of patients. When criteria for categorising patients in clinical research are based on physical examination findings, the clinician can recognise patients in the different groupings and implement research results into daily practice.

The development of the new system is in the initial phase compared to the existing treatment oriented ones, and a long line of future testing is required to substantiate or refute the utility of this pathoanatomic approach. Basically, the judgement as to whether a classification system is useful, requires data on reliability and validity.

The issue of reliability of the new system is addressed in the discussion section concerning Paper IV below. Regarding validity, there is evidence supporting the criteria's ability to identify relevant entities considering the purpose in some

of the categories, Reducible Disc Syndrome (RD), Irreducible Disc Syndrome (IRD), Non-mechanical Disc Syndrome (NMD), Nerve Root Compression Syndrome (NRC), Spinal Stenosis Syndrome (SS), Sacroiliac Joint Syndrome (SIJ), and Abnormal Pain Syndrome (AP), in others, Adherent Nerve Root Syndrome (ANR), Nerve Root Entrapment Syndrome (NRE), Zygapophysial Joint Syndrome (ZJ), Postural Syndrome (Post), Dysfunction Syndrome (Dys), Myofascial Pain Syndrome (MP), and Adverse Neural Tension Syndrome (ANT), the validity is mainly hypothesised. The latter categories are included initially in the system and future studies will reveal if they should remain. Such studies should evaluate the diagnostic test in comparison with a generally accepted reference standard for identification of symptomatic anatomical structures, such as surgery findings, response to diagnostic injections, imaging findings, or results of electrodiagnostic procedures.²⁸

Several critical issues in the new classification system deserve particular attention. Namely the clinical decision process, the identification of symptomatic neural tissues, a possible expansion of the system, and the relevance of the system for improving treatment outcomes. These issues are discussed in detail below.

The clinical decision process

The clinical decision process in the new classification system deserves some discussion. The order of the decision making is described in detail in the Content Validity section of Paper III.

Important terms used to describe the comparison between a diagnostic test and a reference standard diagnosis are sensitivity and specificity. The term sensitivity denotes the ability of a test to correctly identify those patients with the condition. The ability of a negative test to correctly identify the absence of a condition is called specificity. The frequencies are expressed as percentages. If a test has high sensitivity, a negative test result is useful for ruling out the condition. If a test has high specificity, a positive test result is useful for ruling the condition in.³⁹ Although the generalisation frequently is made that sensitivity and specificity are stable, i.e. not influenced by prevalence of the condition, sensitivity of a test is greater among patients with a more common condition than among patients with a less common condition.²⁸

Often, a new diagnostic test is tested on a group of patients who obviously have the condition and are contrasted with persons who are completely normal. In the clinical setting, however, a test is most valuable when it can distinguish patients with the condition from patients with other signs and symptoms that mimic that condition clinically.

Of main interest in the decision process of the new classification system is initially to identify the patients that meet the criteria for a Disc Syndrome. Data suggest that intervertebral disc pathology is the most common structural source

of symptoms in non-specific LBP^{29;103} and that the co-existence of symptomatic discs and other symptomatic anatomical structures is uncommon^{34;102;127}. Therefore, if discogenic pain is eliminated, it follows that there is an increase in the prevalence of other syndromes in the remaining group of patients. By the removal of the syndrome with the highest prevalence, and thus the removal of the greatest potential for false positive tests, we have increased the predictive value of the subsequent tests in identifying the miscellaneous mutually exclusive syndromes (category 2–9) among the remaining patients. A thorough discussion of the relations between sensitivity/specificity and predictive values of tests is beyond the limits of this thesis. The reader is referred to Fritz and Wainner.³⁹

This approach is useful when particular clinical tests have documented an acceptable level of accuracy in identifying a symptomatic structure in the overall group of non-specific LBP patients. For this purpose, the specificity of the criteria is of particular importance. Specificity may be improved by tightening up criteria required for a positive test response.³⁹ This will increase the utility of the test for identifying a condition when the test is positive (fewer false positive tests). 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 negative tests). For our purpose, it is more important to make sure that patients included in the Disc Syndrome category have high probabilities of actually having discogenic pain, rather than overlooking a patient with this condition, 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 strict criterion for a positive test of centralisation or peripheralisation proposed by Laslett and van Wijmen⁶² i.e. that the symptoms are abolished or produced in a most distal body component (see definitions in Paper III, Table 1).

Nevertheless, the method has certain limitations that can be illustrated by a closer look at the criteria chosen for the ZJ. Because the criteria for a symptomatic zygapophysial joint proposed by Revel et al.^{93;94} are predominantly negative, i.e. absence of a pain provocation by various movements (See Table 1 in Paper III), the accuracy of criteria might be inflated because they are valuable for excluding other conditions and not necessarily because they are valuable for identification of symptomatic zygapophysial joints. Some of the ZJ criteria are likely to exclude symptomatic discs (no pain by lumbar flexion in standing and pain not increased by rising from flexion) and thus by increasing the prevalence of the condition, the likelihood by chance of identifying a symptomatic zygapophysial joint is increased. The studies by Revel et al. showed a sensitivity ranging from 82% to 92% and a specificity ranging from 78% to 80%.^{93;94} The validity of the criteria has been questioned in a study by Manchicanti et al.⁷⁴ who found a sensitivity of 13% and a specificity of 84%. The great variation in sensitivity estimates might be explained by at least two factors. First, there might have been

an uneven distribution of particular symptomatic structures, i.e. intervertebral discs or zygapophysial joints, in the different study samples. In a heterogeneous sample of patients with non-specific LBP, the results of diagnostic studies would be expected to vary greatly, unless the study samples included identical proportions of subgroups.⁶⁶ A second possible factor is the influence of differences in reference standards used in the studies. Double anaesthetic blocks, which were used as the reference standard in the study by Manchicanti et al., is regarded as having superior diagnostic accuracy compared to single blocks that were used in the studies by Revel et al. A false positive rate of 37–38% from the use of single blocks compared to the use of double blocks has been reported.^{74;101}

In summary, although some data support the proposed criteria for the diagnosis of ZJ, the ability of the criteria to identify symptomatic zygapophysial joints is not clear. The issue needs further investigation, for example in studies testing the ability of the criteria to discriminate between symptomatic zygapophysial joints and symptomatic intervertebral discs, by means of discography and double anaesthetic blocks as the reference standards.

Identification of symptomatic neural tissues

Mutual exclusiveness in the conclusion of the examination procedure concerning categories 1–9 does not mean that categories necessarily are mutually exclusive in identifying different symptomatic anatomical structures. The fact that the classification system comprises several syndromes (ANR, NRE, and NRC), which all are assumed to indicate a symptomatic nerve root, is problematic. The reader might be confused if focusing on our pathoanatomical assumptions which are plausible but not critical to the classification system. In many cases a patient with a symptomatic nerve root (caused by a chemically irritated sciatic nerve root, nerve root oedema, inflammatory reaction in the dura membranes etc.) might be placed in all of these syndromes, if tested for all. These syndromes are presumably a mixture of different pathological conditions (lateral recess stenosis, nerve root compression following a herniated disc, intraneural fibrotic scar tissue etc.) By following the decision order described in the Content Validity section of Paper III, the examiner will be able to differentiate between them. The order of these three syndromes is made 1) to differentiate between different specific types of nerve root pathology claimed by other systems to be identifiable by signs and symptoms and 2) to make it possible to identify syndromes where a notion of a different prognosis and a different treatment has been presented and thereby made testable in further studies. However, it is possible that NRE and ANR are clinically irrelevant variations.

The ANR and NRE (which have been transferred from the McKenzie classification system) may be difficult to identify in patients with non-specific LBP, and to separate them from other nerve root syndromes. It is under consideration to collapse NRE into one of the other syndromes when data from future studies

investigating reliability and validity of the classification system are analysed. This collapse is supported by reliability studies by Kilpikoski et al.⁵⁶ and Razmjou et al.⁹² as well as the results presented in this thesis (Table 3). No patients were classified as ANR or NRE in the study by Kilpikoski et al., one and zero patients were classified as ANR or NRE respectively in the study by Razmjou et al., and three and one patients were classified as ANR or NRE respectively in the study presented in this thesis. There was no agreement among examiners in any of these positive cases. Further support for the collapse would be provided if either the previous results are reproduced in larger samples of LBP patients, or if no verification of an actual fibrosis or compression of a nerve root can be made during surgery in patients classified as ANR or NRE.

Similarly, collapsing ANT into other categories would be premature at this stage of the development process. The theoretical hypothesis behind ANT is that 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. In future studies, it is intended to test the hypothesis that patients not classified in one or more of the syndromes ANR, NRE, or NRC, may still be classified as ANT. This clinical pattern would presumably occur in cases where only a pathological condition in peripheral nervous structures is present (and a nerve root involvement is not).

If the purpose of classifying patients in a future study is other than to test hypotheses of treatment efficacy, it might be relevant to use a different order. For example, in studies of prognostic value of the identification of subcategories of a symptomatic nerve root, it may be initially useful to identify patients fulfilling the criteria for NRC followed by a subdivision into the categories 1–3, 5 or 9.

Expansion of the classification system

A thorough discussion of the possibility of including an “Instability Syndrome” is provided in Paper III. The purpose of the inclusion of such a syndrome would be to identify patients with a reduced muscular ability to control or stabilise the neutral position of the spine in daily activities. Fundamentally, the identification of such a syndrome is beyond the purpose of the new classification system inasmuch as the deep stabilising trunk muscles are not the pain producing structures themselves. The concept of instability refers to a muscular imbalance mechanism, i.e. reduced muscular ability to control the neutral position of the spine, or a loss of structural integrity in the lumbopelvic osteoligamentous system.^{13;87} The resulting development of symptoms in the low back appears either to be the consequence of a painful anatomical structure, i.e. an intervertebral disc, in which symptoms are maintained by the resultant instability, or the result of an uneven distribution of stress in the low back tissues that may trigger pain from

several pain producing anatomical structures.

Recent studies have showed promising results supporting the reliability and validity of tests based on the modification of symptom provoking movements and alignments of the spine.¹¹⁶ Furthermore, there is convincing evidence showing the effectiveness of specific stabilising exercises in the treatment of patients with non-specific LBP.^{44;82;116} In relation to our system, an “Instability Syndrome” will presumably be included as a subgroup to several of our proposed categories in the future process of developing treatment oriented subcategories in order to assist in selecting the most appropriate treatment for an individual patient. For example, a hypothesis to be tested in this phase might be that patients fulfilling criteria for both RD and Instability Syndrome would achieve better outcomes when treated with McKenzie therapy in combination with stabilising exercises compared with that of McKenzie therapy alone.

Likewise, future studies are needed to address the topic as to whether particular syndromes ought to be subdivided further into subgroups to assist in selecting the most appropriate treatment for the individual patient. For example, a SJ might be subdivided into unstable or inflammatory sub-syndromes, a RD into subgroups related to the patient’s direction of preference (flexion, extension, side-gliding, or rotation), and an ANT into subcategories dependent on presence/absence of a hip-rotation-sensitizing component.

Classification and treatment outcome

Inasmuch as our classification system is primarily developed for use in clinical research (e.g. outcome studies) the question of prognosis and treatment of the different categories is beyond the main purpose of this thesis. Future studies are needed to investigate how the new system could be further developed into a treatment oriented one.

Currently, orthopaedic physiotherapy has treatment systems that assume some influence or effect of the structural origin of symptoms. A pathoanatomic classification system is necessary to enable researchers to test these hypothetical influences. That is whether treatment methods, such as the McKenzie method, manipulation, ANT, or neuromuscular stabilisation therapy, actually alter the conditions on a structural level that are assumed by their theoretical basis. For example, a consensus within the physiotherapy profession regarding the clinical criteria for symptomatic sacroiliac joint pathology might facilitate testing the efficacy of various commonly used treatments directed at the structures around these joints, i.e. training of the stabilising muscles, manual therapy techniques, mobilising exercises, manual and electrical pain inhibition methods, or patient education in self care. Hypotheses regarding efficacy of these methods on particular pain producing structures would be verified if pain provocation tests included in criteria were positive at baseline and negative at end of treatment.

For evaluation of the new classification system as a whole, randomised con-

trolled trials of patient outcomes is necessary. This may be achieved by allocating non-specific LBP patients into a group that is classified and a control group not classified. Only if outcomes were significantly better in the group that was treated according to classification than in the control group receiving a non-specific treatment, could it be concluded that an effective therapy for a specific subgroup has been identified.

In conclusion, the proposed new classification appears to represent a step forward from the present situation in primary care which is characterised by widespread use of pathoanatomically labelled diagnostic categories without agreement on which clinical criteria that should be used to identify a category. The next task is to provide data on the degree of reliability and validity of the system.

Reliability of the classification system (Paper IV)

Results of the study presented in Paper IV showed that the inter-tester reliability of the syndromes was acceptable for trained examiners. The Kappa coefficients were above 0.4, which is considered to reflect acceptable reliability in a clinical context.⁶¹ In syndromes where calculation of Kappa coefficients was not possible due to insufficient numbers of positive findings the percentage agreement was above 90%.

The levels of reliability in this study might have been inflated by the fact that both the first and the second examiner were present simultaneously during testing for category 1–7, and 9. Possible variability due to repeated testing, was thus eliminated for these categories. This issue is addressed in more detail in the Discussion section of Paper IV.

The Kappa coefficient is the recommended statistic for analysis of agreement between two examiners on nominal data because of its correction for the amount of agreement that can be expected by chance. However, the value of Kappa depends upon the proportion (prevalence) of subjects in positive and negative cells in the contingency table.³ The greater the difference between the counts in the positive and negative cell is, the lower the Kappa value generally is. Another problem is that the Kappa value depends on the number of categories. With more categories, the Kappa value is generally lower. The present study is influenced by both of these problems (see Tables 3 and 4). The overall Kappa value of 0.62 for the 11 mutually exclusive categories might indicate an underestimation of the actual agreement between examiners expressed by the percentage of agreement in these categories.

The percentage of a 'perfect match' between therapists was 39% for the classification system as a whole. This level of agreement is not unexpected when categories are allowed to co-exist given the fact that the more information that the researcher or clinician wishes to obtain the greater he is at risk of decreasing

its reliability.⁸⁵ This relatively modest level of total agreement might indicate that the utility of the system for general screening purposes is limited compared with the utility in identification of particular syndromes.

The inter-tester reliability of the system was found to be acceptable for the use of testing hypotheses of treatment efficacy for particular categories. If the tests included in the system were to be used as outcome measures, further investigation of the tests' ability to reliably measure changes over time is warranted.

Conclusions

- The McKenzie method for mechanical diagnosis and therapy was as effective as strengthening training for patients with subacute or chronic non-specific LBP. There seems to be a need for a generally accepted classification system in order to identify subgroups of patients that may benefit from different treatment methods.
- Three treatment oriented classification systems for patients with non-specific LBP appear promising for use within the physiotherapy profession. However, none of those have demonstrated superior reliability or validity to justify its selection above others. There is a need for further investigation of those three systems as well as the need for development of new ones that might overcome some of the fundamental limitations inherent in the existing ones.
- A new pathoanatomic oriented classification system of relevance to physiotherapy was developed based on clinical experience, review of the literature and comments from a conference of experts. It was proposed that the new system has the potential to become generally accepted inasmuch as it might overcome some of the limitations of the treatment oriented ones. For some of the categories of the system there was existing evidence supporting the reliability and validity of the criteria used for categorisation. For others, these were hypothesised. The rationale for all categories was provided. Further testing of the validity of categories for identification of symptomatic anatomical structures in the low back is needed.
- Inter-tester reliability of the new classification system was shown to be acceptable for use in future testing of treatment efficacy for subgroups of patients with non-specific LBP.

Summary in Danish

“Uspecifik lænderygbesvær—klassifikation og behandling”

Afhandlingens emne er måling af effekt af McKenzie behandlingsmetoden til patienter med uspecifik lænderygbesvær og udvikling af et diagnostisk klassifikationssystem til brug i primærsektoren.

McKenzie metoden er en af de mest anvendte af fysioterapeuter i en række vestlige lande til undersøgelse og behandling af patienter med rygbesvær. Delarbejde I er den første publicerede randomiserede kliniske undersøgelse, som tester effekten af McKenzie metoden til undersøgelse og behandling af patienter med længerevarende lænderygbesvær. Resultaterne for de inkluderede 260 patienter viste, at McKenzie-metoden er mindst lige så effektiv til patienter med subakut og kronisk uspecifik lænderygbesvær som intensiv dynamisk udholdenhedstræning, hvilket er den almindeligvis anbefalede behandling. Endvidere understøtter resultaterne behovet for en klassifikation af denne heterogene gruppe af patienter, som ofte er fremført af rygforskere i primærsektoren.

I delarbejde II gennemførtes en systematisk litteraturgennemgang af tidligere foreslåede klassifikationssystemer med relevans for fysioterapi. Gennemgangen konkluderede, at flere forskellige systemer konkurrerer om at blive generelt accepterede i fysioterapifaget. I få tilfælde var undersøgelser publiceret, som kunne vise en vis grad af reproducerbarhed. Men ingen af disse systemer var i stand til at dokumentere deres overlegenhed frem for andre med hensyn til validitet (d.v.s. deres evne til at identificere subgrupper af patienter med bedre effekt af en bestemt behandling i sammenligning med andre). Der var således behov for dels fortsat testning af de eksisterende systemer, og dels udvikling af nye alternative systemer.

I delarbejde III præsenteredes et nyt diagnostisk klassifikationssystem baseret på patoanatomiske og kliniske kategorier af patienter med uspecifik lænderygbesvær. Valg af kategorier og kriterier for kategorisering byggede på indhold i tidligere klassifikationssystemer, resultatet af en dansk ekspertkonference samt eksisterende evidens vedrørende diagnostisk præcision af kriterier. En systematisk metode blev anvendt til at vurdere kriteriernes reproducerbarhed, validitet, anvendelighed og generaliserbarhed i primærsektoren. Der argumenteredes for, at det nye klassifikationssystem er i stand til at overvinde adskillige af begrænsningerne i tidligere systemer, og at det nye system har potentiel nytteværdi til forskningsformål. Der var dog behov for yderligere undersøgelser af reproducerbarhed og validitet af systemet som helhed.

Delarbejde IV testede den nye klassifikation for reproducerbarhed i form af inter-observatør overensstemmelse. Fire undersøgere klassificerede 90 patienter

med længerevarende uspecifik lænderygsbesvær. Resultaterne viste, at trænede undersøgere var istand til at opnå enighed med en acceptable grad af pålidelighed ved brug af klassifikationen. Dog bevirkede en lav prevalens af positive fund i flere af kategorierne at der er behov for yderligere test af inter-observatør overensstemmelse i en større patientpopulation. Endvidere er fremtidige undersøgelser af validitet og anvendelighed af systemet nødvendig.

Acknowledgements

Firstly, I would like to thank Mark Laslett who has been a great inspiration and energetic support to me since his instant reply to my initial e-mail to him back in 1999. It was a mind-blowing experience to meet a colleague on the other side of the world whose thinking and reasoning in classification matters were so alike my own. He has opened his home in New Zealand to me and I am indebted for his guidance and friendship. Whenever we meet there is great fun in the air. Our collaboration has been fruitful and I hope it will continue for many years.

A huge thanks to Hanne Thorsen and Søren Jacobsen for guidance, creative criticism and linguistic advice whenever I needed it.

To Charlotte Ekdahl for her guidance, support, and belief in me from the very beginning of it all in 1992 when I started to study for my Bachelor's degree under her supervision in Ö. Grevie.

I want to thank my co-authors (who I have not already mentioned above), Claus Manniche, Steen Olsen, and Peter Kryger, who all have contributed to improve the quality of my research and to make my writing clinically relevant. A special thank to Steen Olsen for never-ending intellectual combats, joyful discussions about clinical decision making, and beer drinking.

Thanks to my network of physiotherapy researchers, especially Hanne Albert, Per Kjær, Annette Winkel and Hans Lund for charing many scientific congresses, laughs, parties, and in-depth discussions with me.

Thanks to all my dear colleagues, first at the Clinic for Rheumatology and Physiotherapy, Valdemarsgade, and later at the Back Center of Copenhagen, for their support and interest in developing their profession. Several of you have willingly given your time to participate in the studies.

I am indebted to my boss through many years, Pia Hansen, for always supporting me, for recognising the importance of doing research in our profession, and for giving me a position that made it possible to make a contribution.

Thanks to Paula van Wijmen, Jette Toldam, Lone Frandsen, Jette Wos, Britt Pedersen, Inger Skjærbæk, Asbjörn Midttun, Carl Terje Røervik, Charlotte Krog, and Tom Bendix for clinical expert advice.

Most importantly I thank my family. First of all Ingelise, my wife, for her love and for always being the main source of joy in my life. To my children, Morten, Tim, and Lasse for just being there as the wonderful persons that they are.

Financial support for these studies was gratefully received from the Danish Physiotherapy Organization, The Danish Physical Therapist Association for Musculoskeletal Therapy, the Health Insurance Fund, the Danish Rheumatology Association, Madsen's Fund, and the Fund for Promotion of Physiotherapeutic Research and Education.

Reprint permissions

Figure 2 was reprinted from Leboeuf-Yde C et al.: Commentary. *J Manipulative Physiol Ther* 2003; 24(1): 63-66 by permission of Elsevier, National University of Health Sciences.

Figure 8 was reprinted from Travell J G and Simons D G: *Myofascial Pain and Dysfunction. The Triggerpoint Manual*. 1982. Williams and Wilkins. Baltimore, USA, by permission of Lippincott Williams & Wilkins.

References

1. Low back pain: frequency, management and prevention from a HTA perspective. Danish Institute for Health Technology Assessment, Series B 1999; 1.
2. Abenhaim L, Rossignol M, Valat JP et al. The role of activity in the therapeutic management of back pain. Report of the International Paris Task Force on Back Pain. *Spine* 2000; 25: 1S-33S.
3. Altman DG. *Practical statistics for medical research*. London: Chapman & Hall, 1991.
4. American Physical Therapy Association. *Guide to Physical Therapist Practice*. Second Edition. *Phys Ther* 2001; 81: 9-746.
5. Andersson GB and Deyo RA. History and physical examination in patients with herniated lumbar discs. *Spine* 1996; 21: 10S-8S.
6. Andersson GB, Lucente T, Davis AM, Kappler RE, Lipton JA, and Leurgans S. A comparison of osteopathic spinal manipulation with standard care for patients with low back pain. *N Engl J Med* 1999; 341: 1426-31.
7. Appendix I. A glossary of operational definitions in physical therapy. *Phys Ther* 75(8), 749-756. 1995.
8. Bass MJ, Buck C, Turner L, Dickie G, Pratt G, and Robinson HC. The physician's actions and the outcome of illness in family practice. *J Fam Pract* 1986; 23: 43-7.
9. Battie MC, Cherkin DC, Dunn R, Ciol MA, and Wheeler KJ. Managing low back pain: attitudes and treatment preferences of physical therapists. *Phys Ther* 1994; 74: 219-26.
10. Beattie P. The relationship between symptoms and abnormal magnetic resonance images of lumbar intervertebral disks. *Phys Ther* 1996; 76: 601-8.
11. Bendix T, Bendix A, Labriola M, Hastrup C, and Ebbelohj N. Functional restoration versus outpatient physical training in chronic low back pain: A randomized comparative study. *Spine* 2000; 25: 2494-500.
12. Bentsen H, Lindgarde F, and Manthorpe R. The effect of dynamic strength back exercise and/or a home training program in 57-year-old women with chronic low back pain. Results of a prospective randomized study with a 3-year follow-up period. *Spine* 1997; 22: 1494-500.
13. Bergmark A. Stability of the lumbar spine. A study in mechanical engineering. *Acta Orthop Scand* 1989; Suppl 230: 1-54.
14. Binkley J, Finch E, Hall J, Black T, and Gowland C. Diagnostic classification of patients with low back pain: report on a survey of physical therapy experts. *Phys Ther* 1993; 73: 138-50.
15. Bogduk N. The anatomical basis for spinal pain syndromes. *J Manipulative Physiol Ther* 1995; 18: 603-5.

16. Borkan JM and Cherkin DC. An agenda for primary care research on low back pain. *Spine* 1996; 21: 2880-4.
17. Bouter LM, van Tulder MW, and Koes BW. Methodologic issues in low back pain research in primary care. *Spine* 1998; 23: 2014-20.
18. Bronfort G, Goldsmith CH, Nelson CF, Boline PD, and Anderson AV. Trunk exercise combined with spinal manipulative or NSAID therapy for chronic low back pain: a randomized, observer-blinded clinical trial. *J Manipulative Physiol Ther* 1996; 19: 570-82.
19. Buchbinder R, Goel V, Bombardier C, and Hogg-Johnson S. Classification systems of soft tissue disorders of the neck and upper limb: do they satisfy methodological guidelines? *J Clin Epidemiol* 1996; 49: 141-9.
20. Cherkin DC, Battie MC, Deyo RA, Street JH, and Barlow W. A comparison of physical therapy, chiropractic manipulation, and provision of an educational booklet for the treatment of patients with low back pain. *N Engl J Med* 1998; 339: 1021-9.
21. Chok B, Lee R, Latimer J, and Tan SB. Endurance training of the trunk extensor muscles in people with subacute low back pain. *Phys Ther* 1999; 79: 1032-42.
22. Delitto A, Cibulka MT, Erhard RE, Bowling RW, and Tenhula JA. Evidence for use of an extension-mobilization category in acute low back pain syndromes: A prescriptive validation pilot study. *Phys Ther* 1993; 73: 216-28.
23. Delitto A, Erhard RE, and Bowling RW. A treatment-based classification approach to low back syndrome: identifying and staging patients for conservative treatment. *Phys Ther* 1995; 75: 470-85.
24. Delitto A, Shulman AD, Rose SJ et al. Reliability of a clinical examination to classify patients with low back syndrome. *Physical Therapy Practice* 1992; 1: 1-9.
25. Delitto A and Snyder Mackler L. The diagnostic process: examples in orthopedic physical therapy. *Phys Ther* 1995; 75: 203-11.
26. Deyo RA. Diagnostic evaluation of LBP: reaching a specific diagnosis is often impossible. *Arch Intern Med* 2002; 162: 1444-7.
27. Deyo RA and Diehl AK. Patient satisfaction with medical care for low-back pain. *Spine* 1986; 11: 28-30.
28. Deyo RA, Haselkorn J, Hoffman R, and Kent DL. Designing studies of diagnostic tests for low back pain or radiculopathy. *Spine* 1994; 19: 2057S-65S.
29. Donelson R, Aprill CN, Medcalf R, and Grant W. A prospective study of centralization of lumbar and referred pain. A predictor of symptomatic discs and anular competence. *Spine* 1997; 22: 1115-22.
30. Donelson R, Silva G, and Murphy K. Centralization phenomenon. Its usefulness in evaluating and treating referred pain. *Spine* 1990; 15: 211-3.
31. Dreyfuss P, Michaelsen M, Pauza K, McLarty J, and Bogduk N. The value of medical history and physical examination in diagnosing sacroiliac joint pain. *Spine* 1996; 21: 2594-602.

32. Erhard RE, Delitto A, and Cibulka MT. Relative effectiveness of an extension program and a combined program of manipulation and flexion and extension exercises in patients with acute low back syndrome. *Phys Ther* 1994; 74: 1093-100.
33. Fast A. Low back disorders: conservative management. *Arch Phys Med Rehabil* 1988; 69: 880-91.
34. Fortin JD, Aprill CN, Ponthieux B, and Pier J. Sacroiliac joint: pain referral maps upon applying a new injection/arthrography technique. Part II: Clinical evaluation. *Spine* 1994; 19: 1483-9.
35. Foster NE, Thompson KA, Baxter GD, and Allen JM. Management of nonspecific low back pain by physiotherapists in Britain and Ireland. A descriptive questionnaire of current clinical practice. *Spine* 1999; 24: 1332-42.
36. Fritz JM, Delitto A, and Erhard RE. Comparison of classification-based physical therapy with therapy based on clinical practice guidelines for patients with acute low back pain: a randomized clinical trial. *Spine* 2003; 28: 1363-71.
37. Fritz JM, Delitto A, Welch WC, and Erhard RE. Lumbar spinal stenosis: a review of current concepts in evaluation, management, and outcome measurements. *Arch Phys Med Rehabil* 1998; 79: 700-8.
38. Fritz JM and George S. The use of a classification approach to identify subgroups of patients with acute low back pain. Interrater reliability and short-term treatment outcomes. *Spine* 2000; 25: 106-14.
39. Fritz JM and Wainner RS. Examining diagnostic tests: an evidence-based perspective. *Phys Ther* 2001; 81: 1546-64.
40. Gillan MG, Ross JC, McLean IP, and Porter RW. The natural history of trunk list, its associated disability and the influence of McKenzie management. *Eur Spine J* 1998; 7: 480-3.
41. Gracey JH, McDonough SM, and Baxter GD. Physiotherapy management of low back pain: a survey of current practice in northern ireland. *Spine* 2002; 27: 406-11.
42. Hemmila HM, Keinanen-Kiukaanniemi SM, Levoska S, and Puska P. Does folk medicine work? A randomized clinical trial on patients with prolonged back pain. *Arch Phys Med Rehabil* 1997; 78: 571-7.
43. Hestbaek L, Leboeuf-Yde C, and Manniche C. Low back pain: What is the long term course? A review of studies of general patient populations. *Eur Spine J* 2003; 12: 149-65.
44. Hides JA, Jull GA, and Richardson CA. Long-term effects of specific stabilizing exercises for first-episode low back pain. *Spine* 2001; 26: E243-E248.
45. Hsieh CY, Adams AH, Tobis J et al. Effectiveness of four conservative treatments for subacute low back pain: a randomized clinical trial. *Spine* 2002; 27: 1142-8.
46. Hurwitz EL, Morgenstern H, Harber P et al. A randomized trial of medical care with and without physical therapy and chiropractic care with and without physical modalities for patients with low back pain: 6-month follow-up outcomes from the UCLA low back pain study. *Spine* 2002; 27: 2193-204.

47. Indahl A, Haldorsen EH, Holm S, Reikeras O, and Ursin H. Five-year follow-up study of a controlled clinical trial using light mobilization and an informative approach to low back pain. *Spine* 1998; 23: 2625-30.
48. Indahl A, Velund L, and Reikeraas O. Good prognosis for low back pain when left untampered. A randomized clinical trial. *Spine* 1995; 20: 473-7.
49. Jackson RP. The facet syndrome. Myth or reality? *Clin Orthop* 1992; 110-21.
50. Jensen MC, Brant Zawadzki MN, Obuchowski N, Modic MT, Malkasian D, and Ross JS. Magnetic resonance imaging of the lumbar spine in people without back pain. *N Engl J Med* 1994; 331: 69-73.
51. Kalauokalani D, Cherkin DC, Sherman KJ, Koepsell TD, and Deyo RA. Lessons from a trial of acupuncture and massage for low back pain: patient expectations and treatment effects. *Spine* 2001; 26: 1418-24.
52. Karas R, McIntosh G, Hall H, Wilson L, and Melles T. The relationship between nonorganic signs and centralization of symptoms in the prediction of return to work for patients with low back pain. *Phys Ther* 1997; 77: 354-60.
53. Keel PJ, Wittig R, Deutschmann R et al. Effectiveness of in-patient rehabilitation for sub-chronic and chronic low back pain by an integrative group treatment program (Swiss Multicentre Study). *Scand J Rehabil Med* 1998; 30: 211-9.
54. Kendrick D, Fielding K, Bentley E, Kerslake R, Miller P, and Pringle M. Radiography of the lumbar spine in primary care patients with low back pain: randomised controlled trial. *BMJ* 2001; 322: 400-5.
55. Kent, P. and Keating, J. L. The sub-grouping and classification of non-specific low back pain in primary care. Proceedings from the 14th International Congress of the World Confederation for Physical Therapy, Barcelona 2003: RR-PL-0386.
56. Kilpikoski S, Airaksinen O, Kankaanpaa M, Leminen P, Videman T, and Alen M. Interexaminer reliability of low back pain assessment using the McKenzie method. *Spine* 2002; 27: E207-E214.
57. Kirkaldy Willis WH and Hill RJ. A more precise diagnosis for low-back pain. *Spine* 1979; 4: 102-9.
58. Koes BW, van Tulder MW, Ostelo R, Kim BA, and Waddell G. Clinical guidelines for the management of low back pain in primary care: an international comparison. *Spine* 2001; 26: 2504-13.
59. Kuorinka I, Jonsson B, Kilbom A et al. Standardised nordic questionnaires for the analysis of musculoskeletal symptoms. *Applied ergonomics* 1987; 18: 233-7.
60. Lacroix JM, Powell J, Lloyd GJ, Doxey NC, Mitson GL, and Aldam CF. Low-back pain. Factors of value in predicting outcome. *Spine* 1990; 15: 495-9.
61. Landis JR and Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159-74.
62. Laslett M and van Wijmen PM. Low back and referred pain: Diagnosis and a proposed new system of classification. *New Zealand J Physiother* 1999; 27: 5-14.

63. Laslett M, Young S, Aprill C, and McDonald B. Diagnosing painful sacroiliac joints. A validity study of a McKenzie evaluation and sacroiliac provocation tests. *Aust J Physiother* 2003; 49: 89-97.
64. Leboeuf-Yde C, Klougart N, and Lauritzen T. How common is low back pain in the Nordic population? *Spine* 2003; 21: 1518-25.
65. Leboeuf-Yde C, Lauritsen JM, and Lauritzen T. Why has the search for causes of low back pain largely been nonconclusive? *Spine* 1997; 22: 877-81.
66. Leboeuf-Yde C and Manniche C. Low back pain: Time to get off the treadmill (commentary). *J Manipulative Physiol Ther* 2001; 24: 63-6.
67. Ljunggren AE, Weber H, Kogstad O, Thom E, and Kirkesola G. Effect of exercise on sick leave due to low back pain. A randomized, comparative, long-term study. *Spine* 1997; 22: 1610-6.
68. Long AL. The centralization phenomenon. Its usefulness as a predictor of outcome in conservative treatment of chronic low back pain (A pilot study). *Spine* 1995; 20: 2513-21.
69. Lutz GK, Butzlaff ME, Atlas SJ, Keller RB, Singer DE, and Deyo RA. The relation between expectations and outcomes in surgery for sciatica. *J Gen Intern Med* 1999; 14: 740-4.
70. Maigne JY, Aivaliklis A, and Pfefer F. Results of sacroiliac joint double block and value of sacroiliac pain provocation tests in 54 patients with low back pain. *Spine* 1996; 21: 1889-92.
71. Malmivaara A, Hakkinen U, Aro T et al. The treatment of acute low back pain--bed rest, exercises, or ordinary activity? *N Engl J Med* 1995; 332: 351-5.
72. Malterud K and Hollnagel H. The magic influence of classification systems in clinical practice. *Scand J Prim Health Care* 1997; 15: 5-6.
73. Maluf KS, Sahrman SA, and Van Dillen LR. Use of a classification system to guide nonsurgical management of a patient with chronic low back pain. *Phys Ther* 2000; 80: 1097-111.
74. Manchikanti L, Pampati V, Fellows B, and Ghafoor Baha A. The inability of the clinical picture to characterize pain from facet joints. *Pain Physician* 2001; 3: 158-66.
75. Manchikanti L, Pampati V, Fellows B, and Bakhit CE. The diagnostic validity and therapeutic value of lumbar facet joint nerve blocks with or without adjuvant agents. *Curr Rev Pain* 2000; 4: 337-44.
76. Manniche C, Hesselsoe G, Bentzen L, Christensen I, and Lundberg E. Clinical trial of intensive muscle training for chronic low back pain. *Lancet* 1988; 2: 1473-6.
77. Manniche C, Lundberg E, Christensen I, Bentzen L, and Hesselsoe G. Intensive dynamic back exercises for chronic low back pain: a clinical trial. *Pain* 1991; 47: 53-63.
78. Mannion AF, Muntener M, Taimela S, and Dvorak J. A randomized clinical trial of three active therapies for chronic low back pain. *Spine* 1999; 24: 2435-48.

79. McKenzie RA. The lumbar spine. Mechanical diagnosis and therapy. Waikanae. New Zealand.: Spinal Publications, 1981.
80. McPhillips-Tangum CA, Cherkin DC, Rhodes LA, and Markham C. Reasons for repeated medical visits among patients with chronic back pain. *J Gen Intern Med* 1998; 13: 289-95.
81. Mooney V. The syndromes of low back disease. *Orthop Clin North Am* 1983; 14: 505-15.
82. Moseley L. Combined physiotherapy and education is efficacious for chronic low back pain. *Aust J Physiother* 2002; 48: 297-302.
83. Nachemson AL. Newest knowledge of low back pain. A critical look. *Clin Orthop* 1992; 8-20.
84. Nachemson AL, et al. [Back Pain, Neck Pain. An Evidence Based Review]. *Ondt i ryggen, ondt i nacken. En evidensbaseret kunskapssammanställning. The Swedish Council on Technology Assessment in Health Care. Report Number 145 ed. Stockholm, Sweden.: 2000.*
85. Nelson MA, Allen P, Clamp SE, and de Dombal FT. Reliability and reproducibility of clinical findings in low-back pain. *Spine* 1979; 4: 97-101.
86. Novack DH. Therapeutic aspects of the clinical encounter. *J Gen Intern Med* 1987; 2: 346-55.
87. Panjabi MM. The stabilizing system of the spine. Part II. Neutral zone and instability hypothesis. *J Spinal Disord* 1992; 5: 390-6.
88. Pengel HM, Maher CG, and Refshauge KM. Systematic review of conservative interventions for subacute low back pain. *Clin Rehabil* 2002; 16: 811-20.
89. Petersen T, Kryger P, Ekdahl C, Olsen S, and Jacobsen S. The Effect of McKenzie Therapy as Compared With That of Intensive Strengthening Training for the Treatment of Patients With Subacute or Chronic Low Back Pain: A Randomized Controlled Trial. *Spine* 2002; 27: 1702-9.
90. Porter RW. Mechanical disorders of the lumbar spine. *Ann Med* 1989; 21: 361-6.
91. Powell MC, Wilson M, Szypryt P, Symonds EM, and Worthington BS. Prevalence of lumbar disc degeneration observed by magnetic resonance in symptomless women. *Lancet* 1986; 2: 1366-7.
92. Razmjou H, Kramer JF, and Yamada R. Intertester reliability of the McKenzie evaluation in assessing patients with mechanical low-back pain. *J Orthop Sports Phys Ther* 2000; 30: 368-83.
93. Revel M, Listrat VM, Chevalier XJ et al. Facet joint block for low back pain: identifying predictors of a good response. *Arch Phys Med Rehabil* 1992; 73: 824-8.
94. Revel M, Poiraudau S, Auleley GR et al. 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* 1998; 23: 1972-7.
95. Riddle DL and Freburger JK. Evaluation of the presence of sacroiliac joint region dysfunction using a combination of tests: a multicenter intertester reliability study. *Phys Ther* 2002; 82: 772-81.

96. Rose SJ. Physical therapy diagnosis: role and function. *Phys Ther* 1989; 69: 535-7.
97. Scheel IB, Hagen KB, Herrin J, Carling C, and Oxman AD. Blind faith? The effects of promoting active sick leave for back pain patients: a cluster-randomized controlled trial. *Spine* 2002; 27: 2734-40.
98. Schers H, Wensing M, Huijsmans Z, van Tulder M, and Grol R. Implementation barriers for general practice guidelines on low back pain a qualitative study. *Spine* 2001; 26: E348-E353.
99. Schwarzer AC, Aprill CN, and Bogduk N. The sacroiliac joint in chronic low back pain. *Spine* 1995; 20: 31-7.
100. Schwarzer AC, Aprill CN, Derby R, Fortin J, Kine G, and Bogduk N. Clinical features of patients with pain stemming from the lumbar zygapophysial joints. Is the lumbar facet syndrome a clinical entity? *Spine* 1994; 19: 1132-7.
101. Schwarzer AC, Aprill CN, Derby R, Fortin J, Kine G, and Bogduk N. The false-positive rate of uncontrolled diagnostic blocks of the lumbar zygapophysial joints. *Pain* 1994; 58: 195-200.
102. Schwarzer AC, Aprill CN, Derby R, Fortin J, Kine G, and Bogduk N. The relative contributions of the disc and zygapophyseal joint in chronic low back pain. *Spine* 1994; 19: 801-6.
103. Schwarzer AC, Aprill CN, Derby R, Fortin J, Kine G, and Bogduk N. The prevalence and clinical features of internal disc disruption in patients with chronic low back pain. *Spine* 1995; 20: 1878-83.
104. Schwarzer AC, Derby R, Aprill CN, Fortin J, Kine G, and Bogduk N. Pain from the lumbar zygapophysial joints: a test of two models. *J Spinal Disord* 1994; 7: 331-6.
105. Schwarzer AC, Derby R, Aprill CN, Fortin J, Kine G, and Bogduk N. The value of the provocation response in lumbar zygapophyseal joint injections. *Clin J Pain* 1994; 10: 309-13.
106. Seferlis T, Nemeth G, Carlsson AM, and Gillstrom P. Conservative treatment in patients sick-listed for acute low- back pain: a prospective randomised study with 12 months' follow- up. *Eur Spine J* 1998; 7: 461-70.
107. Skelton AM, Murphy EA, Murphy RJ, and O'Dowd TC. Patients' views of low back pain and its management in general practice. *Br J Gen Pract* 1996; 46: 153-6.
108. Slipman CW, Sterenfeld EB, Chou LH, Herzog R, and Vresilovic E. The predictive value of provocative sacroiliac joint stress maneuvers in the diagnosis of sacroiliac joint syndrome. *Arch Phys Med Rehabil* 1998; 79: 288-92.
109. Stankovic R and Johnell O. Conservative treatment of acute low-back pain. A prospective randomized trial: McKenzie method of treatment versus patient education in »mini back school«. *Spine* 1990; 15: 120-3.
110. Stankovic R and Johnell O. Conservative treatment of acute low back pain. A 5-year follow- up study of two methods of treatment. *Spine* 1995; 20: 469-72.

111. Sufka A, Hauger B, Trenary M et al. Centralization of low back pain and perceived functional outcome. *J Orthop Sports Phys Ther* 1998; 27: 205-12.
112. Thomas KB. General practice consultations: is there any point in being positive? *Br Med J (Clin Res Ed)*. 1987; 294: 1200-2.
113. Torstensen TA, Ljunggren AE, Meen HD, Odland E, Mowinckel P, and Geijerstam S. Efficiency and costs of medical exercise therapy, conventional physiotherapy, and self-exercise in patients with chronic low back pain. A pragmatic, randomized, single-blinded, controlled trial with 1- year follow-up. *Spine* 1998; 23: 2616-24.
114. Underwood MR and Morgan J. The use of a back class teaching extension exercises in the treatment of acute low back pain in primary care. *Fam Pract* 1998; 15: 9-15.
115. Van Dillen LR, Sahrman SA, Norton BJ et al. Reliability of physical examination items used for classification of patients with low back pain. *Phys Ther* 1998; 78: 979-88.
116. Van Dillen LR, Sahrman SA, Norton BJ, Caldwell CA, McDonnell MK, and Bloom N. The effect of modifying patient-preferred spinal movement and alignment during symptom testing in patients with low back pain: a preliminary report. *Arch Phys Med Rehabil*.2003; 84: 313-22.
117. van Tulder, M. European guidelines for the management of acute nonspecific low back pain in primary care. Preliminary draft. Guidelines for its Management, Cost Action B13. www.backpaineurope.org. 2003: 1-52.
118. van Tulder M, Malmivaara A, Esmail R, and Koes B. Exercise therapy for low back pain: A systematic review within the framework of the cochrane collaboration back review group. *Spine* 2000; 25: 2784-96.
119. van Tulder MW, Assendelft WJ, Koes BW, and Bouter LM. Spinal radiographic findings and nonspecific low back pain. A systematic review of observational studies. *Spine* 1997; 22: 427-34.
120. van Tulder MW, Koes BW, and Bouter LM. Conservative treatment of acute and chronic nonspecific low back pain. A systematic review of randomized controlled trials of the most common interventions. *Spine* 1997; 22: 2128-56.
121. Waddell G. *The Back Pain Revolution*. 1 ed. Edinburgh: Churchill Livingstone, 1998.
122. Waddell G. (personal communication).
123. Waddell G, Somerville D, Henderson I, and Newton M. Objective clinical evaluation of physical impairment in chronic low back pain. *Spine* 1992; 17: 617-28.
124. Werneke M and Hart DL. Centralization phenomenon as a prognostic factor for chronic low back pain and disability. *Spine* 2001; 26: 758-64.
125. Werneke M, Hart DL, and Cook D. A descriptive study of the centralization phenomenon. A prospective analysis. *Spine* 1999; 24: 676-83.
126. Wiesel SW, Tsourmas N, Feffer HL, Citrin CM, and Patronas N. A study of computer-assisted tomography. I. The incidence of positive CAT scans in an asymptomatic group of patients. *Spine* 1984; 9: 549-51.

127. Young S and Aprill CN. Characteristics of a mechanical assessment for chronic lumbar facet joint pain. *J Man Manipul Ther* 2000; 8: 78-84.
128. Zusman M. Structure-orientated beliefs and disability due to back pain. *Aust J Physiother* 1998; 44: 13-20.

Paper I–IV

