

From the Department of Orthopedics, Spine Centre,
Research School Caphri, University of Maastricht,
Maastricht University Medical Centre,
The Netherlands

DECISION MAKING IN SURGICAL TREATMENT OF CHRONIC LOW BACK PAIN

**The performance of prognostic tests to select patients for
lumbar spinal fusion**

Paul Willems

ACTA ORTHOPAEDICA SUPPLEMENTUM NO. 349, VOL. 84, 2013

informa
healthcare

Contact address

Paul Willems, MD, PhD
Department of Orthopaedics
Maastricht University Medical Centre
P.O. Box 5800
6202 AZ Maastricht
The Netherlands
Fax: 00-31 (0) 43-3874893
Tel: 00-31 (0) 43-3875038
p.willems@mumc.nl

Copyright © Informa Healthcare Ltd 2013. ISSN 1745–3674. Printed in England – all rights reserved
This supplement is based on the Doctoral thesis by Paul WILLEMS, defended 2011
DOI 10.3109/17453674.2012.753565

Printed in England
Latimer Trend & Company Limited
2013

Contents

LIST OF PAPERS, 2

ABSTRACT, 3

INTRODUCTION, 5

Spinal fusion for chronic low back pain, 5

Prognostic tests to aid in clinical decision making, 6

Aims and outline of the present thesis, 7

METHODOLOGY, 9

Survey Dutch spine surgeons (study I), 9

The effect of an orthosis on lumbosacral motion (study II), 9

Cohort study: Value of an orthosis to predict clinical outcome (study III), 9

Value of adjacent level discography to predict the outcome of fusion (study IV), 10

Postdiscography discitis (study V), 11

Value of temporary external fixation to predict clinical outcome (study VI), 12

Systematic review of the literature (study VII), 13

FINDINGS AND DISCUSSION, 15

Survey Dutch spine surgeons (study I), 15

The effect of an orthosis on lumbosacral motion (study II), 17

Cohort study: Value of an orthosis to predict clinical outcome (study III), 18

Value of adjacent level discography to predict the outcome of fusion (study IV), 20

Postdiscography discitis (study V), 22

Value of temporary external fixation to predict clinical outcome (study VI), 24

Systematic review of the literature (study VII), 24

GENERAL DISCUSSION, 29

REFERENCES, 32

List of papers

This thesis was based on the following studies:

- I Paul Willems, Rob de Bie, Cumhur Öner, René Castelein, Marinus de Kleuver.
Spinal fusion for chronic low back pain: No consensus in clinical decision making. Results of a nationwide survey among spine surgeons in the Netherlands.
BMJ Open 2011; 1: e000391.
- II Willems P C, Nienhuis B, Sietsma M, van der Schaaf D B, Pavlov P W.
The effect of a plaster cast on lumbosacral joint motion. An in vivo assessment with precision motion analysis system.
Spine 1997; 22: 1229–34.
- III Willems P C, Elmans L, Anderson P G, Jacobs W C H, van der Schaaf D B, de Kleuver M.
The value of a pantaloone cast test in surgical decision-making for chronic low back pain patients: A systematic review of the literature supplemented with a prospective cohort study.
Eur Spine J 2006; 15: 1487–94.
- IV Willems P C, Elmans L, Anderson P G, van der Schaaf D B, de Kleuver M.
Provocative discography and lumbar fusion: Is preoperative assessment of adjacent discs useful?
Spine 2007; 32: 1094–9.
- V Willems P C, Jacobs W C H, Duinkerke E S, de Kleuver M.
Lumbar discography: Should we use prophylactic antibiotics? A study of 435 consecutive discograms and a systematic review of the literature.
J Spinal Disord Tech 2004; 17: 243–7.
- VI Elmans L, Willems P C, Anderson P G, van Limbeek J, de Kleuver M, van der Schaaf D B.
Temporary external transpedicular fixation of the lumbosacral spine: A prospective, longitudinal study in 330 patients.
Spine 2005; 30: 2813–6.
- VII Willems P C, Staal J B, Walenkamp G H I M, de Bie R A.
Spinal fusion for low back pain: Systematic review on the accuracy of prognostic tests.
Spine J, In press.

Abstract

Chronic low back pain (CLBP) is one of the main causes of disability in the western world with a huge economic burden to society. As yet, no specific underlying anatomic cause has been identified for CLBP. Imaging often reveals degenerative findings of the disc or facet joints of one or more lumbar motion segments. These findings, however, can also be observed in asymptomatic people. It has been suggested that pain in degenerated discs may be caused by the ingrowth of nerve fibers into tears or clefts of the annulus fibrosus or nucleus pulposus, and by reported high levels of pro-inflammatory mediators. As this so-called discogenic pain is often exacerbated by mechanical loading, the concept of relieving pain by spinal fusion to stabilise a painful spinal segment, has been developed. For some patients lumbar spinal fusion indeed is beneficial, but its results are highly variable and hard to predict for the individual patient. To identify those CLBP patients who will benefit from fusion, many surgeons rely on tests that are assumed to predict the outcome of spinal fusion. The three most commonly used prognostic tests in daily practice are immobilization in a lumbosacral orthosis, provocative discography and trial immobilization by temporary external transpedicular fixation.

Aiming for consensus on the indications for lumbar fusion and in order to improve its results by better patient selection, it is essential to know the role and value of these prognostic tests for CLBP patients in clinical practice.

The overall aims of the present thesis were:

- 1) to evaluate whether there is consensus among spine surgeons regarding the use and appreciation of prognostic tests for lumbar spinal fusion;
- 2) to verify whether a thoracolumbosacral orthosis (TLSO) truly minimises lumbosacral motion;
- 3) to verify whether a TLSO can predict the clinical outcome of fusion for CLBP;
- 4) to assess whether provocative discography of adjacent segments actually predicts the long-term clinical outcome fusion;
- 5) to determine the incidence of postdiscography discitis, and whether there is a need for routine antibiotic prophylaxis;
- 6) to assess whether temporary external transpedicular fixation (TETF) can help to predict the outcome of spinal fusion;
- 7) to determine the prognostic accuracy of the most commonly used tests in clinical practice to predict the outcome of fusion for CLBP.

The results of a national survey among spine surgeons in the Netherlands were presented in Study I. The surgeons were questioned about their opinion on prognostic factors and about the use of predictive tests for lumbar fusion in CLBP patients.

The comments were compared with findings from the prevailing literature. The survey revealed a considerable lack of uniformity in the use and appreciation of predictive tests. Prognostic factors known from the literature were not consistently incorporated in the surgeons' decision making process either. This heterogeneity in strategy is most probably caused by the lack of sound scientific evidence for current predictive tests and it was concluded that currently there is not enough consensus among spine surgeons in the Netherlands to create national guidelines for surgical decision making in CLBP.

In Study II, the hypothesized working mechanism of a pantaloons cast (i.e., minimisation of lumbosacral joint mobility) was studied. In patients who were admitted for a temporary external transpedicular fixation test (TETF), infrared light markers were rigidly attached to the protruding ends of Steinman pins that were fixed in two spinal levels. In this way three-dimensional motion between these levels could be analysed opto-electronically. During dynamic test conditions such as walking, a plaster cast, either with or without unilateral hip fixation, did not significantly decrease lumbosacral joint motion.

Although not substantiated by sound scientific support, lumbosacral orthoses or pantaloons casts are often used in everyday practice as a predictor for the outcome of fusion. A systematic review of the literature supplemented with a prospective cohort study was performed (Study III) in order to assess the value of a pantaloons cast in surgical decision-making. It appeared that only in CLBP patients with no prior spine surgery, a pantaloons cast test with substantial pain relief suggests a favorable outcome of lumbar fusion compared to conservative treatment. In patients with prior spine surgery the test is of no value.

It is believed by many spine surgeons that provocative discography, unlike plain radiographs or magnetic resonance imaging, is a physiologic test that can truly determine whether a disc is painful and relevant in a patient's pain syndrome, irrespective of the morphology of the disc. It has been suggested that in order to achieve a successful clinical outcome of lumbar fusion, suspect discs should be painful and adjacent control discs should elicit no pain on provocative discography. For this reason, a cohort of patients in whom the decision to perform lumbar fusion was based on an external fixation (TETF) trial, was analysed retrospectively in Study IV. The results of preoperative discography of solely the levels adjacent to the fusion were compared with the clinical results after spinal fusion. It appeared that in this select group of patients the discographic status of discs adjacent to a lumbar fusion did not have any effect on the clinical outcome.

The most feared complication of lumbar discography is discitis. Although low in incidence, this is a serious complication for a diagnostic procedure and prevention by the use of prophylactic antibiotics has been advocated. In search for clinical guidelines, the risk of postdiscography discitis was assessed in Study V by means of a systematic literature review and a cohort of 200 consecutive patients. Without the use of prophylactic antibiotics, an overall incidence of postdiscography discitis of 0.25% was found. To prove that antibiotics would actually prevent discitis, a randomised trial of 9,000 patients would be needed to reach significance. Given the possible adverse effects of antibiotics, it was concluded that the routine use of prophylactic antibiotics in lumbar discography is not indicated.

In Study VI, the middle- and long-term results of external fixation (TETF) as a test to predict the clinical outcome of lumbar fusion were studied in a group of back pain patients for whom there was doubt about the indication for surgery. The test included a placebo trial, in which the patients were unaware whether the lumbar segmental levels were fixed or dynamised. Using strict and objective criteria of pain reduction on a visual analogue scale, the TETF test failed to predict clinical outcome of fusion in this select group of patients. Pin track infection and nerve root irritation were registered as complications of this invasive test. It was concluded that in chronic low back pain patients with a doubtful indication for fusion, TETF is not recommended as a supplemental tool for surgical decision-making.

In Study VII, a systematic literature review was performed regarding the prognostic accuracy of tests that are currently used in clinical practice and that are presumed to predict the outcome of lumbar spinal fusion for CLBP. The tests of interest were magnetic resonance imaging (MRI), TLSO immobilisation, TETF, provocative discography and facet joint infiltration. Only 10 studies reporting on three different index

tests (discography, TLSO immobilisation and TETF) that truly reported on test qualifiers, such as sensitivity, specificity and likelihood ratios, could be selected. It appeared that the accuracy of all prognostic tests was low, which confirmed that in many clinical practices patients are scheduled for fusion on the basis of tests, of which the accuracy is insufficient or at best unknown. As the overall methodological quality of included studies was poor, higher quality trials that include negatively tested as well as positively tested patients for fusion, will be needed. It was concluded that at present, best evidence does not support the use of any prognostic test in clinical practice. No subset of patients with low back pain could be identified, for whom spinal fusion is a reliable and effective treatment.

In literature, several studies have reported that cognitive behavioural therapy or intensive exercise programs have treatment results similar to those of spinal fusion, but with considerably less complications, morbidity and costs. As the findings of the present thesis show that the currently used tests do not improve the results of fusion by better patient selection, these tests should not be recommended for surgical decision making in standard care. Moreover, spinal fusion should not be proposed as a standard treatment for chronic low back pain. Causality of nonspecific spinal pain is complex and CLBP should not be regarded as a diagnosis, but rather as a symptom in patients with different stages of impairment and disability. Patients should be evaluated in a multidisciplinary setting or Spine Centre according to the so-called biopsychosocial model, which aims to identify underlying psychosocial factors as well as biological factors. Treatment should occur in a stepwise fashion starting with the least invasive treatment. The current approach of CLBP, in which emphasis is laid on self-management and empowerment of patients to take an active course of treatment in order to prevent long-term disability and chronicity, is recommended.

Introduction

Spinal fusion for chronic low back pain

Low back pain has become a major health problem in the western world with 1-year prevalences ranging from 22 to 65% and life-time prevalences of up to 84% (Walker 2000). About 20% of those who suffer from low back pain seek medical attention. The economic burden to society is huge and can be divided into direct costs of health care utilisation (hospitalisation, medication, tests and therapies) and indirect costs of lost productivity due to work absenteeism and early retirement. In a relatively small Western-European country, such as the Netherlands, the total costs of back pain in 2003 were estimated at 4.4 billion euros (Slobbe 2006), which are mainly employment-related costs (van Tulder et al. 1995). The total annual costs of back pain in the USA exceed 100 billion dollars (Katz 2006).

Waddell (Waddell 1996) suggested a simple and practical classification, which divides low back pain into three categories: (1) pain caused by specific spinal pathology, e.g. tumor, infection or trauma, (2) nerve root or radicular pain and (3) nonspecific low back pain, which constitutes a large heterogeneous group of patients (about 85% of total cases (Deyo and Bass 1989)). In these patients imaging often reveals signs of degeneration of one or more intervertebral discs, such as disc space narrowing, vertebral endplate changes (Weishaupt et al. 2001), annular disruption (Aprill and Bogduk 1992) and/or facet joint arthropathy. These degenerative findings, however, can also be observed in asymptomatic subjects (Jensen et al. 1994) and thus, are nonspecific for low back pain.

The present thesis concerns patients who suffer for at least three months from nonspecific pain or discomfort in the lumbar region, with or without referred leg pain (Airaksinen et al. 2006), and who all have signs of degeneration of the lumbar spine on plain radiographs or magnetic resonance imaging (MRI), further to be mentioned as chronic low back pain (CLBP) patients.

CLBP can result in serious physical and social restrictions and has a substantial impact on the life style of those affected. Several socioeconomic risk factors for the onset and persistence of CLBP have been identified (Bigos et al. 1991; McIntosh et al. 2000): Job dissatisfaction, physically strenuous work, low education and Workers' Compensation insurance are all associated with CLBP and resulting disability (Katz 2006). Patients out of work for 6 months due to low back pain have a 50% chance of returning to their previous job, whereas those who have been out of work for two years or more are unlikely to get reemployed again (Waddell 1996). In addition, factors such as smoking, obesity, psychological

distress, depressive mood and to a lesser extent somatisation, have been reported to result in increased risk of persisting symptoms and disability (Pincus et al. 2002). Treating this large heterogeneous group of CLBP patients with signs of degeneration of the intervertebral disc and/or facet joints on imaging, is both challenging and frustrating, as clinicians cannot rely on a true consensus among the peers for a rational therapeutic approach.

Lumbar spinal fusion for CLBP

The exact etiology and pathophysiology of intervertebral disc degeneration is unknown and believed to result from multiple complex interactions of biologic and biomechanical factors. Genetic predisposition appears to play an important role and several gene forms associated with disc degeneration have been identified (Cheung et al. 2006; Battie et al. 2008). Histologically, degenerative lumbar discs reveal abnormal ingrowth of sensory nerve fibers in the endplates and the nucleus pulposus (Coppes et al. 1997). Radial fissures in the annulus fibrosus have been linked to the ingrowth of free nerve endings (so-called nociceptors) and blood vessels with a granulation zone (Peng et al. 2005). This ingrowth of nerve endings has been suggested to correlate to the dull chronic back ache, often referred to as "discogenic pain", that is experienced by CLBP patients, and which is exacerbated by mechanical loading of the spine. Additionally, degenerative herniated discs contain high levels of pro-inflammatory mediators and cytokines (Kang et al. 1997; Burke et al. 2002), such as interleukin-6, nitric oxide, prostaglandin E2 (Kang et al. 1996) and phospholipase A2 (Franson et al. 1992). All these findings have strengthened the idea that disc degeneration, with or without secondary osteoarthritis of the synovial facet joints, could be a major pathway for CLBP. The assumption that the pain is largely associated with continued motion at the affected disc level has led to the concept of stabilisation of a painful degenerative motion segment (i.e., intervertebral disc and facet joints) by spinal fusion to alleviate pain and reduce disability.

Spinal fusion was first described in 1891 for stabilisation of vertebral segments in spinal tuberculosis (Hadra 1891) and has a well-established role in the treatment of spinal fractures and deformities. Its role in the treatment of CLBP, however, remains controversial (Turner et al. 1992; Krismer 2002). Epidemiological research reveals large variation in the amount of spinal fusion operations between countries (Cherkin et al. 1994) and even between different regions within the same country up to 20-fold ranges in fusion rate have been reported (Katz 1995; Weinstein et al. 2006). Although socio-economic factors and the availability of spine surgery in a region or

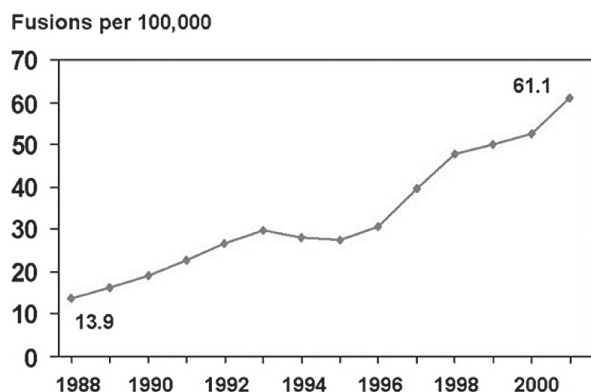


Figure 1. The rise of spinal fusion procedures for degenerative conditions in the USA between 1988 and 2001. Note the rapid increase after 1996, when fusion cages were approved. (Courtesy from Mr Richard A Deyo, J Am Board Fam Med 2009; 22: 62–68, reproduced with permission).

country may differ, such wide geographic variations suggest a poor level of professional consensus on the indications for spinal fusion in CLBP patients.

In a Cochrane review in 1999 (Gibson et al. 1999), no evidence was found on the effectiveness of fusion for lumbar degenerative disc disease (DDD) or low back pain, as compared to natural history, placebo, or conservative treatment. Despite the absence of concurrent reports of clarified indications or improved efficacy, there was a 220% increase in the rate of lumbar spine fusion surgery from 1990 to 2001 in the USA (Figure 1, (Deyo et al. 2005)). This rise coincided with the promotion and FDA-approval of the fusion cages (interbody implants to enhance stability and fusion) in 1996. As a result, Medicare expenses for lumbar fusion increased from 75 million dollars in 1992 to 482 million dollars in 2003, which was about half of total USA spending for spine surgery (Weinstein et al. 2006).

In 2001, a Swedish randomised controlled trial (RCT) (Fritzell et al. 2001) did show a better outcome in patients treated with spinal fusion compared to patients who received standard conservative care, although at longer follow-up this beneficial effect attenuated (Fritzell et al. 2004). Two more recent RCTs, which compared fusion surgery to cognitive behavioural based exercise therapy (Brox et al. 2003) or an intensive rehabilitation program (Fairbank et al. 2005), showed similar results for surgery and nonsurgery at 1 year, and at 2 years follow-up, respectively. In the latter trial, fusion had a higher complication rate (Wilson-MacDonald et al. 2008) and appeared to be less cost-effective than intensive rehabilitation (Rivero-Arias et al. 2005). From these trials one might conclude that a considerable amount of patients do not improve after surgery and that spinal fusion is a questionable treatment for the entire heterogeneous group of CLBP patients. In a recently published survey (Haefeli et al. 2008), renowned spine surgeons did not consider complications or bad surgery, but bad patient selection, wrong diagnosis and disproportionate expectations of

patients to be the major factors for poor outcome in spinal surgery. Therefore, it would be of great value to know whether the results of fusion surgery can be improved by the identification of subgroups of CLBP patients who truly benefit from spinal fusion.

At present, all routine diagnostic tools to select the right patient for lumbar fusion appear to be inadequate. History taking and physical examination may reveal psychological distress or social factors which can lead to chronicity of pain and need treatment (Nelson et al. 1979; Deyo et al. 1992). There are, however, no specific physical findings to predict the outcome of lumbar fusion (Carragee 2005). The association of CLBP with findings on imaging is weak (van Tulder et al. 1997). Degenerative signs on plain radiographs do not appear to be correlated with persisting symptoms of CLBP (Kumar et al. 2001a) and magnetic resonance imaging (MRI), which has been recommended as the imaging study of choice for the initial evaluation of patients with CLBP (Resnick et al. 2005), shows high false-positive rates of degenerative findings in asymptomatic people, especially in the elderly (Boden et al. 1990; Jensen et al. 1994).

Prognostic tests to aid in clinical decision making

To identify those patients who will benefit from fusion surgery, many surgeons rely on tests that are assumed to predict the outcome of spinal fusion. In order to improve the results of lumbar fusion for CLBP and aiming for consensus on the indications for fusion, it is essential to know the role and value of these prognostic tests in clinical practice.

In the present thesis, we will focus on the most commonly used prognostic tests:

Immobilisation in a thoracolumbosacral orthosis (TLSO)

In order to mimic the immobilising effect of a lumbar fusion, a standard manufactured brace or corset can be prescribed or a plaster cast applied on the standing patient usually from just below the shoulder blades to the middle of the sacrum (Figure 2). For immobilisation of level L5–S1, one hip is fixed in 10° of flexion with the thigh-piece ending 10 cm proximal to the patella. Patients are expected to wear the TLSO or cast continuously for at least 2 to 4 weeks and are encouraged to perform as much daily life activities as possible. In case of significant pain relief while wearing the TLSO, the test is considered positive.

Identification of painful segments by provocative discography

This procedure can be performed on an outpatient basis. Under sterile conditions and with the patient in lateral decubitus position, a stilette needle is advanced into the intended



Figure 2. Immobilisation test by a thoracolumbosacral orthosis, e.g., a plaster cast. Extension of the cast with a thigh piece is optional and assumed to better immobilise the lumbosacral joint.

disc space (Figure 3). When correct placement of the needle tip in the center of the disc is verified by biplanar fluoroscopy, a contrast agent is injected under pressure. If this injection provokes pain similar to the patient's usual pain, (i.e., concordant pain) and, preferably, if 1 or 2 discs adjacent to the suspect disc do not elicit concordant pain on provocation (i.e., control levels), the test can be considered positive. In addition, plain radiographs or CT-scans can be obtained to determine the extent of degeneration of the injected disc(s).

Trial immobilisation by temporary external transpedicular fixation (TETF)

For this test, patients are hospitalised for 7 to 10 days. Under general anesthesia, antibiotic prophylaxis and fluoroscopic control, two screws are inserted percutaneously through the pedicles into the vertebra above, and two screws into the vertebra below the suspect disc(s), respectively. Postoperatively, the protruding screw ends are fixed externally with two vertical bars, which immobilises the disc(s) of interest, thus mimicking a lumbar fusion (Figure 4). In case of adequate pain relief, the test is considered positive. Optionally, immobilisation can be discontinued - without knowledge of the patient - by fixing the bars horizontally, which, of course, should annul pain relief (i.e., placebo test).

Aims and outline of the present thesis

The aims of this thesis were to assess the value of three commonly used prognostic tests (i.e., immobilization by a TLSO, provocative discography, trial immobilisation by TETF) for the clinical outcome of lumbar spinal fusion in CLBP patients, and to evaluate the use and appreciation of these tests by spine surgeons in clinical practice.

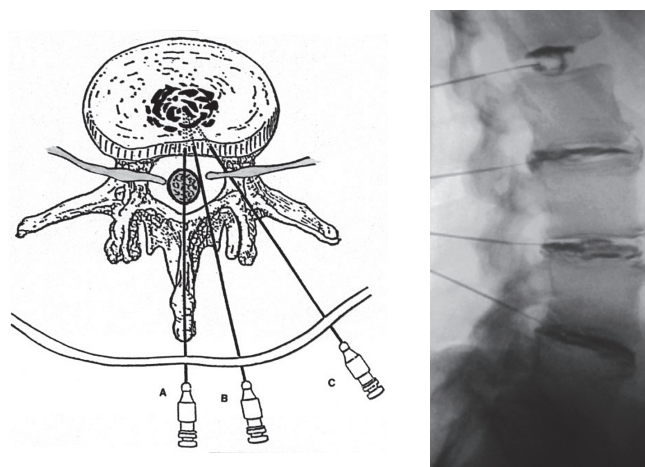


Figure 3. Provocative discography is performed under sterile conditions and with biplanar fluoroscopic control. The intended disc space is injected with contrast agent to provoke usual back pain. In addition, the amount of degeneration can be visualized on plain radiographs or CT scan.

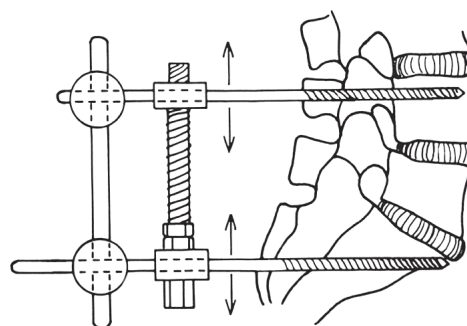


Figure 4. Temporary external transpedicular fixation trial: By externally fixing two vertebrae, one or more vertebral motion segments can be immobilised, thus simulating the effect of a spinal fusion.

Aim 1: Is there consensus among spine surgeons regarding the use of prognostic tests for lumbar spinal fusion in CLBP patients?

In study I the results of a nationwide survey among spine surgeons in the Netherlands were presented. The surgeons were questioned about their opinion on prognostic factors and the use and appreciation of prognostic tests for lumbar spinal fusion in CLBP patients.

Aim 2: Does immobilisation by a pantaloons cast truly minimise lumbosacral motion?

In study II the hypothesized working mechanism of a pantaloons cast, i.e. mimicking a lumbar fusion by minimisation of lumbosacral mobility, was tested. In patients, admitted for a temporary external transpedicular fixation test (TETF), infrared light markers were fixed to the protruding pins of two spinal levels. In this way, three-dimensional motion between these levels could be analysed opto-electronically during dynamic test conditions. Measurements were performed before and after the application of a pantaloons cast.

Aim 3: Can immobilisation by a thoracolumbosacral orthosis (TLSO) predict the clinical outcome of spinal fusion for CLBP?

In study III a systematic review of the literature supplemented with the results of a prospective patient cohort study was presented in order to assess the value of a TLSO in predicting the long-term clinical outcome of fusion for CLBP.

Aim 4: Does provocative discography of adjacent segments predict the long-term clinical outcome of spinal fusion for CLBP?

In study IV the results of provocative discography of levels adjacent to an intended lumbar fusion were presented in a cohort of patients for whom the decision to perform fusion has been based on a TETF trial. The relation between pre-operative discography results and long-term clinical outcome was assessed.

Aim 5: What is the incidence of postdiscography discitis and is there a need for routine antibiotic prophylaxis?

Although low in incidence, intervertebral discitis is the most feared complication of provocative discography. In study V the risk of postdiscography discitis and the need for routine prophylactic antibiotics were studied in a cohort of 200 consecutive patients, and by means of a systematic review of the literature.

Aim 6: Can temporary external transpedicular fixation (TETF) help to predict the clinical outcome of spinal fusion in CLBP patients with an equivocal indication for surgery?

In study VI the middle- and long-term results of TETF as a test to predict the clinical outcome of lumbar fusion, were presented in a group of CLBP patients with a doubtful indication for surgery. Suspect lumbar motion segments were temporarily stabilized by external fixation. In case of adequate pain relief, lumbar fusion was indicated. The test included a placebo trial, in which the patients were unaware whether the lumbar segmental levels were fixed or dynamised.

Aim 7: What is the prognostic accuracy of the most commonly used tests in clinical practice to predict the outcome of spinal fusion for CLBP?

Study VII presented the results of a systematic literature review concerning the prognostic accuracy of tests that are currently used in clinical practice to predict the outcome of lumbar spinal fusion for CLBP. The tests of interest were magnetic resonance imaging (MRI), immobilisation by a thoracolumbosacral orthosis (TLSO), TETF, provocative discography and facet joint infiltration.

Methods

Survey Dutch spine surgeons (study I)

A 25-question survey was sent by mail to all surgeon members of the Dutch Spine Society, by Memic, a Centre for Data and Information Management, University of Maastricht, the Netherlands (www.memic.unimaas.nl). The questionnaire concerned the selection for spinal fusion of patients with low back pain caused by degenerative lumbar disc disease without signs of neurological deficit, spinal stenosis, deformity or spondylolisthesis and in the absence of trauma, tumor or infections. This group was further referred to as chronic low back pain patients. Most questions could be answered according to a 5-point Likert scale. Surgeon specific factors (e.g., discipline, clinical experience), the influence of patient factors (prognostic factors as reported in literature), and the use of tests for patient selection (e.g., provocative discography), were addressed. The respondents were specifically asked to rely on their own individual opinion and management in practice.

Data were entered into Excel™ (Microsoft, Corp., Redmond, WA) and all inconsistencies were resolved. Unanswered questions were coded as missing. Descriptive statistics was used in which all frequencies were based on the number of valid responders.

For analysis the answers on the 5-point Likert scale were merged into one intermediate option ("neutral") and 2 opposite categories ("always/almost always" versus "never/almost never" and "fully/globally agree" versus "globally/fully disagree"). The data were processed with Statistical Package for the Social Sciences software (SPSS, Inc., Chicago, IL). Pearson's chi-square test was used to evaluate whether surgeon specific factors were associated with clinical decision making. Uniformity of opinion was defined to be present if 70% or more of the respondents answered similarly. In other words, there was no consensus if the proportion of the largest category was statistically significantly lower than 70% (Pearson's chi-square test). Differences in mean values rating the impact of factors on decision making, were tested by Independent t-test for equality of means. The level of significance was set at $p < 0.05$.

The effect of an orthosis on lumbosacral motion (study II)

Ten consecutive patients with intractable low back pain volunteered for this study. They were admitted for a temporary external fixation test (Magerl 1984; Olerud et al. 1986), for

which two Schanz screws were inserted transpedicularly into the fourth or fifth lumbar vertebra, and two screws into the sacrum (Figure 4). For the experiment the external fixator was removed and a plaster cast was applied on the standing patient from just below the shoulder blades to the middle of the sacrum. For unilateral hip immobilisation, the left hip was fixed in 10° of flexion with the thigh piece 10 cm proximal to the patella. In both casts a hole was created around the protruding screw ends to permit free range of motion.

Measurements were performed with PRIMAS, an optoelectronic measuring system using infrared light (Hermann 1990). Two video cameras were placed behind the patient under an angle of 80°. Two frames, each containing three reflective spherical markers were attached to the blunt outer end of a proximal and of a sacral protruding Schanz screw, respectively. Motion of the proximal frame relative to the sacral frame was expressed in translations (accuracy ± 1 cm) and rotations (accuracy ± 0.3 degrees) around three axes. Continuous registration with a sampling frequency of 100 Hz were acquired with the patient in neutral standing position, maximal spinal flexion and extension, maximal pelvic tilt, and while walking. All excursions were performed three times and the same sequence was repeated with the patient wearing a plaster cast with and without hip extension, respectively.

For analysis median values of the ranges of rotation and translation during the exercises were calculated. For each step during walking, the ranges between maxima and minima were determined. In each patient the difference (delta) between the ranges without support and those obtained in both casts were calculated. Statistical analysis by means of a two-way analysis of variance (ANOVA), taking a probability of 5% as the level of significance.

Cohort study: Value of an orthosis to predict clinical outcome (study III)

Between April 1990 and October 1999, 278 patients with more than six months of incapacitating LBP without objective neurologic motor deficit underwent a temporary external transpedicular fixation trial (Esses et al. 1989) as a final diagnostic test. For all patients the indication for lumbar fusion had remained unclear after routine diagnostic tests, including standard radiographs, either MRI or discography, and a pantaloone cast test that was not considered to be a decisive test in surgical decision-making. Thus, the decision to perform lumbar fusion or to continue nonsurgical management was based on the results of the TETF trial. All patients gave

Table 1. Orthosis study: Patient variables for both treatment groups

	Lumbar fusion (n = 107)	Conservative management (n = 150)
Age, mean years (SD)	40 (8.8)	41 (8.5)
Gender, female	68 (64%)	97 (65%)
Baseline VAS, mean (SD)	75 (17)	74 (16)
Degenerative disc disease (DDD)	34 (32%)	51 (34%)
Spondylolysis /-olisthesis	3 (3%)	11 (7%)
Prior spine surgery	70 (65%)	88 (59%)
– Postdiscectomy syndrome	29 (27%)	30 (20%)
– Pseudarthrosis	7 (6%)	9 (6%)
– Combination DDD and prior operation	31 (29%)	46 (30%)
– Postchemonucleolysis	2 (2%)	2 (1%)
– Post-trauma	1 (1%)	1 (1%)

informed consent for the TETF procedure and analysis of the patients was approved by an institutional review board.

For the pantaloon cast test, a plaster cast was applied on the standing patient from the nipples to the waist, extending over one leg to the knee (Figure 2). Patients were expected to wear this corset for 6 weeks, with an absolute minimum of three weeks for inclusion in this study. They were encouraged to perform as many daily life activities as possible. Patients were instructed to return earlier for cast removal in case of unbearable pain. At a follow-up visit, generally two weeks after cast removal, the patients were asked whether they had experienced substantial pain relief when wearing the cast. If no significant relief or even a worsening of symptoms had been experienced, a negative test result was noted; in case of significant pain relief, the test was considered positive.

Before the allocation to treatment by the TETF trial, the patients were asked to rate the average pain they had experienced during the previous month on a horizontal Visual Analogue Scale (VAS) from 0 (no pain) to 100 mm (maximal, unbearable pain). All patients were admitted to hospital for the TETF trial. In case of a significant pain reduction in the fixed position of the TETF compared to the nonfixed position (placebo trial), the result was positive with the advice to perform lumbar fusion (van der Schaaf et al. 1999). If pain reduction was absent or not substantial, conservative management was to be continued. At a median follow-up of 76 (15–144) months, patient satisfaction (global self-rating in excellent, good, fair or poor) and pain on the VAS were assessed by a questionnaire that was sent by L.E., who had no previous contact with any of the patients. For analysis, the pain score on VAS at follow-up was subtracted from the initial score before the TETF trial. We defined a decrease of 30% or more from the initial pain to be clinically relevant and scored this as a successful outcome. Patients with less than 30% pain reduction were considered failures. For two of the 278 patients the initial VAS scores were lacking and 19 patients could not be traced for follow-up. Thus, 257 patients remained for analysis (Table 1).

Of these 257 patients, 158 had undergone prior spine surgery while 99 patients had no history of spine surgery. A model was made using a logistic regression to determine which variables contributed to the clinically relevant pain decrease of 30% or more on the VAS score. The scalar variables included were age, duration of symptoms and length of follow-up. The dichotomized variables included were gender, prior spine surgery, pantaloon cast test result and treatment (lumbar fusion or conservative management). Variables with a p-value less than 0.10 were considered to be possible contributors. Two interaction terms were included: prior spine surgery versus treatment and pantaloon cast test result versus treatment. Pearson's Chi-Square test was used to calculate statistical significance of differences between groups and correlations were tested with Spearman's Rank Correlation test. A p-value of 0.05 was considered significant.

Value of adjacent level discography to predict the outcome of fusion (study IV)

From April 1990 to October 1999, 209 patients for whom the indication for a lumbar fusion remained uncertain after routine diagnostic tests underwent a temporary external transpedicular fixation trial (Esses et al. 1989) as the final, decisive test. All patients had suffered from incapacitating low back pain for more than one year and had no objective neurologic motor deficit. Routine diagnostic tests comprised standard radiographs and provocative discography that included the assessment of discs adjacent to the level(s) intended for fusion. The eventual decision whether to perform lumbar fusion or to continue nonsurgical management was determined by the TETF trial for which all patients gave informed consent.

Lumbar discography was performed in the Radiology Department on an outpatient basis by orthopaedic residents affiliated with the spine unit (Figure 3). No standard sedatives or prophylactic antibiotics were used. With the patient in the left lateral decubitus position, the technique described by McCulloch and Waddell (McCulloch and Waddell 1978) was used. The skin was prepared with Betadine and draped in the routine manner. After infiltration of the skin with local anaesthetic (Lidocain 1%), a stiletted double needle (18-gauge outer and 22-gauge inner solid needle) was advanced into the intended disc space (Fraser et al. 1987). Correct placement of the needle tip into the centre of the disc was verified by biplanar fluoroscopy. In case the disc centre could not be reached properly, an open, more flexible 22-gauge needle was advanced through the outer needle to reach the centre. Subsequently, Isovist®, a water-soluble nonionic contrast agent (Schering AG, Berlin, Germany) was injected. Next to the disc(s) that was suspected to be the source of pain in the patient based on plain radiographs or MRI, generally two adjacent discs (1 proximal and 1 distal level or 2 proximal levels in case L5–S1 was suspect) were tested. Attention was

paid to whether the patient experienced pain during injection. The patient was asked about the extent and location of the pain and whether this pain was similar to his/her usual pain (i.e. concordant pain). The injection endpoint was defined by manual resistance. The morphological pattern of the disc was described according to the criteria of Adams (Adams et al. 1986) in which the discogram types I and II were scored as normal, whereas the types III–V were considered to be degenerative (Wetzel et al. 1994). For each disc separate needles were used. One hour after the procedure, patients were sent home with the instruction to return immediately in case of aggravated back pain or fever within the next few weeks (Willemis et al. 2004). Routine follow-up was at 6 to 12 weeks after the procedure.

All patients were admitted to hospital for the TETF trial and were asked to rate their pain on a horizontal visual analogue scale (VAS) ranging from 0 (no pain) to 100 mm (maximal, unbearable pain). The levels considered for fusion were immobilized by TETF. In case of a significant pain reduction in the fixed position compared to the nonfixed state (dynamisation) (van der Schaaf et al. 1999), the result of the TETF trial was considered to be positive and lumbar fusion of the intended level(s) was performed. If pain reduction was absent or not substantial, conservative management with commonly practiced measures, such as physical therapy, manipulations, rehabilitation programs, analgesics, lumbar orthoses or nerve root blocks was to be continued. A standardised nonoperative treatment was not advocated because we feel that commonly used conservative methods tailored to the individual patient portray the real life situation for the average chronic LBP patient (Fritzell et al. 2001). At follow-up, a questionnaire was sent to the patients by L.E. who had no previous contact with any of the patients. Pain on the VAS and patient satisfaction (global self-rating of outcome in excellent, good, fair or poor) were thus assessed in both the fused and conservatively treated patients. In every patient the pain score at follow-up was subtracted from the initial pain score. We considered an individual decrease in pain of 30% or more of the initial pain to be a clinically relevant improvement and defined this as a successful outcome. Patients with less than 30% pain reduction were considered failures. For analysis, the discographic findings of all patients with at least one injected disc adjacent to the levels that had been immobilised by TETF, were reviewed and scored as follows: If provocation of these adjacent disc(s) had produced concordant pain, the discography result was considered negative (which would mean that fusion is not recommendable). If injection into these disc(s) had provoked unfamiliar pain or no pain at all, the discography result was positive. From one of the original 209 patients the initial data were lost and 11 cases could not be traced for follow-up: Two patients with symptomatic adjacent discs and eight patients with asymptomatic discs who had received conservative treatment and the only fused lost-to-follow-up case was a patient with asymptomatic discs at discography.

Table 2. Adjacent level discography: Diagnosis of all analysed patients (numbers)

	Fusion and symptomatic adjacent discs (n=22)	Fusion and asymptomatic adjacent discs (n=60)	Conservative management (n=115)
Degen. disc disease (DDD)	4	20	39
Spondylolysis /-olisthesis	2	3	7
Prior spine surgery	16	37	69
– Postdiscectomy	5	18	16
– Pseudarthrosis	1	6	7
– Combination DDD and prior operation	7	12	43
– Post-chemonucleolysis	3	0	2
– Post-trauma	0	1	1

Thus, 197 patients (63 male, 134 female) with a mean age of 40 (SD 8.5) years remained for analysis at a mean follow-up of 80 (35) months (range 15–144 months; Table 2). Based on the TETF results, 115 patients were managed conservatively and 82 patients underwent a lumbar fusion: 24 patients had an anterior interbody fusion (stand-alone or circumferential) and 58 patients received an instrumented posterolateral intertransverse process fusion. For statistical analysis, the variables to be tested (either dichotomized or ordinally scaled) were age, gender, previous spine surgery, length of follow-up, preoperative VAS pain score, provocative discography result of adjacent discs, treatment (lumbar fusion or conservative), and clinical outcome. A logistic regression model was made to determine which variables contributed significantly to a clinically successful outcome. Pearson's Chi-Square test was used to calculate statistical significance of differences between groups. For insight into the magnitude of effect, odds ratios (OR) with confidence intervals (CI) were calculated for any parameter with a p-value of 0.05 or less. Correlations were tested with Spearman's Rank Correlation test. A p-value of 0.05 was considered to be significant.

Postdiscography discitis (study V)

During 1 year the records of all consecutive patients undergoing lumbar discography as part of the work-up for degenerative lumbar disc disease were investigated. All procedures were carried out in the Radiology Department by orthopaedic residents affiliated with the spine unit. No standard preoperative sedation was used, and no prophylactic antibiotics were administered. The patients were placed in the left lateral decubitus position (McCulloch and Waddell 1978). The skin was prepared with Betadine and draped in a routine manner. An image intensifier was used to provide both lateral and anteroposterior views of the spine. After superficial infiltration of local anesthetic, a solid stiletted double needle (18-gauge outer and 22-gauge inner needle) was advanced to

centre of the intended disc space. After fluoroscopic confirmation, the inner solid needle was removed, and Isovist, a water-soluble, non-ionic contrast dye (Schering AG, Berlin, Germany) was injected slowly. If the disc centre could not be approached properly, an open, more flexible 22-gauge needle was advanced through the outer needle to reach the centre. Volume and pressure of the dye injected, as well as its distribution pattern in the disc were scored according to the criteria of (Adams et al. 1986), and the patient's response was registered. For each disc level, separate needles were used. One hour after the procedure, patients were sent home with the instruction to return immediately in case of aggravated back pain or fever. Routine control followed at 6–12 weeks after discography. In case of pain, discomfort, or fever, laboratory and radiologic controls were performed. Discitis was diagnosed when erythrocyte sedimentation rate (ESR) and C-reactive protein (CRP) were raised and magnetic resonance imaging (MRI) revealed the typical increased bony signal in the vertebral bodies with destruction of the endplates and paravertebral signal enhancement (Schulitz and Assheuer 1994). In those cases where the records did not reveal the postdiscography status or follow-up had been less than 12 weeks, patients were approached by telephone and questioned for the above-mentioned symptoms during the first 3 months after the procedure.

Literature search strategy and selection criteria

To obtain all relevant literature, the most common databases of published literature were used: Medline (1966 to January 2003), Current contents (1996 to February 2003) and the Cochrane database of randomised controlled trials (2003, Issue I). No restrictions were made on language or date. On the basis of the abstracts, articles were selected by two independent reviewers. The references from the selected articles were checked in the same way to ensure no relevant articles were missed. Articles were finally included when they met the following criteria:

- Lumbar discography should be performed at all levels with an adequately described two-needle technique under strict aseptic conditions.
- The presence of complications, such as postdiscography discitis must have been specifically looked for at follow-up and should be noted as such.
- The exact number of patients and those lost to follow-up should be noted.

Preferably a minimum follow-up of 3 months was used. If both reviewers could not reach consensus, a third reviewer was consulted.

Value of temporary external fixation to predict clinical outcome (study VI)

Patients with chronic low back pain and an equivocal indi-

Table 3. TETF: Descriptives and mean VAS scores for pain in the patient groups with a negative and those with a positive external fixation test result

	Negative test	Positive test
No. of patients (n=301)	176	125
Age (years)	42 (9.3)	41 (8.7)
Duration of complaints (years)	6.1 (6.7)	5.5 (5.1)
Length of follow-up (months)	81 (35)	76 (34)
Previous spine surgery (%)	65	62
VAS before test	74 (15)	74 (18)
VAS at follow-up	63 (27)	54 (32)
Difference (before – follow-up)	11 (28)	20 (32)

cation for fusion after routine diagnostic tests, such as plain radiographs, MRI, plaster cast immobilisation or discography, were selected for this test. On the basis of the data from routine testing, the symptomatic levels to be immobilised were determined. The patients were hospitalised for 7 to 10 days. Under general anesthesia and fluoroscopic control, self-tapping 5 mm diameter Schanz's screws were inserted transpedicularly. The screw ends were fixed externally by bars either vertically with immobilisation of the vertebral segments involved or horizontally, in which the levels were not immobilised (dynamised state). Each position was maintained for 2 to 3 days and the patient was unaware whether the levels involved were fixed or dynamised. At the end of the test, the screws were removed under general anesthesia.

Suitable candidates for fusion were those patients who experienced adequate pain relief in fixation and definitely less pain relief in the dynamised position.

Before the test, pain was registered on a visual analogue scale (VAS) from 0 (no pain) to 100 points (maximal pain). Similar pain assessment was performed at follow-up by means of a written questionnaire. The test result, type of treatment after the test (fusion or conservative), age, gender, duration of symptoms, and preoperative diagnosis were the variables tested.

Between October 1989 and October 2000, 330 patients (123 male, 207 female, mean age of 42 (SD 9) years) with a mean duration of symptoms of 6 (5) years, underwent the test (Table 3). The median follow-up period was 79 (15–144) months.

Questionnaire data were missing for 29 patients (Two patients did not complete the test due to complications, two deceased, 8 could not be found 17 forms were not returned). Thus the response was 92% with 108 male and 193 female patients who were used for analysis.

The test result was determined by the VAS score at fixation subtracted from the VAS in the dynamised position. A reduction of 30 points or more was considered as a positive test result.

The VAS score at follow-up after treatment subtracted from the VAS score before the test was used as the outcome parameter. A reduction of more than 30% or more was considered to be a clinically relevant improvement in pain. Nominal and

ordinal-scaled variables were analysed using chi-square tests. Differences in means of continuous and interval scaled variables were analysed by paired and unpaired Student's *t* test and analyses of variance, with alpha set at 0.05.

Systematic review of the literature (study VII)

Data sources and searches

A literature search was conducted according to the guidelines by Devillé et al. (2002). We explored Pubmed (1966 – November 2010) and Embase (1974 – November 2010) databases and used search terms for relevant test procedures, study design and patient population. For the tests the following terms were used: immobilisation, thoracolumbosacral orthosis, surgical cast(s), provocative discography, discography, temporary external fixation, facet joint blocks, zygapophyseal joint blocks, imaging and MRI. For study design we used the terms prognosis, prognostic, accuracy, predictive, diagnosis, diagnostic test(s), and diagnostic technique(s), and for patient population the terms lumbar spine, lumbar vertebrae, lumbar-sacral, spinal, low back pain, degenerative disc disease, intervertebral disc(k), disc degeneration, failed back surgery syndrome, spondylosis, spinal fusion and spondylodesis. Both Mesh terms and free text words were entered.

Study selection

Two authors (PW, JBS) screened the titles and abstracts of all references identified by the search to determine whether they met the following inclusion criteria:

- Patients should suffer from low back pain for at least 3 months.
- Studies should contain both patients with a positive and patients with a negative index test result, who subsequently underwent spinal fusion.
- Clinical outcome after fusion, which was considered as the reference standard, should be presented per individual patient in such a way that a relevant clinical improvement cut-off could be defined for analysis and outcome could be dichotomized into “success” or “failure”.
- Pain, subjective improvement, back-specific disability, disability for work, or patient satisfaction should have been incorporated as a clinically-relevant outcome measure.
- Studies should include at least 20 patients.
- There were no restrictions by language.
- Study populations with objective neurologic motor deficit, fracture, infectious disease, ankylosing spondylitis, neoplasm, congenital or adolescent idiopathic scoliosis, kyphosis, or adult scoliosis were excluded.

Full publications of studies which were considered as potentially relevant by both authors were retrieved. The articles were read and checked for final inclusion independently. Any disagreement with regard to study selection was discussed in consensus meetings. In cases where disagreement persisted, a

Table 4. QUADAS for prognostic studies: criteria to assess risk of bias

1. Was the spectrum of patients representative of the patients who will receive the index test in practice?
2. Were selection criteria clearly described?
3. Were the outcomes used to assess recovery collected by means of validated measures of acceptable quality? ^a
4. Was a sufficiently long follow-up period (two years or more) used to assess the outcome of the spinal fusion operation? ^a
5. Did all patients receive spinal fusion followed by the outcome assessment regardless of the index test result?
6. Was the execution of the index test described in sufficient detail in order to permit replication of the test?
7. Was a clear cut-off point used to qualify positive versus negative results of the index test? ^a
8. Did the effect sizes that were used to consider patients as being recovered (i.e., the reference standard) meet accepted standards of clinical relevance, i.e., a minimal important change (MIC) of 30% or more? ^a
9. Were the clinical outcomes after fusion assessed without knowledge of the results of the index test?
10. Were the same clinical data available when index test results were interpreted as would be available when the test is used in practice?
11. Were uninterpretable results of the index test reported?
12. Were withdrawals from the study explained?

QUADAS = quality assessment of diagnostic accuracy studies (20)

^a Items 3, 4, 7, and 8 are modified items.

third reviewer was consulted for the final decision. The references of the articles identified by the search were checked for additional eligible studies.

Data extraction and assessment of bias

Relevant study data were retrieved by the same two reviewers using standardised forms. Extracted information included: standard reference data (first author, journal and publication year), number of patients, characteristics of study population before surgery (i.e., age, gender, severity and duration of pain and/or disability), index test, spinal fusion method, outcome measures, and clinical outcome.

The two reviewers independently assessed the risk of bias of included studies by means of a modified version of the Quality Assessment for Diagnostic Accuracy Studies (QUADAS) checklist (Whiting et al. 2003). The QUADAS is a generally acknowledged checklist to assess the quality of primary studies of diagnostic accuracy. As there are no gold standard criteria for quality assessment of studies of prognostic accuracy, we modified the QUADAS checklist, as listed in Table 4. Disagreements between both reviewers were discussed and resolved in a consensus meeting.

Data synthesis and analysis

By combining outcome (dichotomized into “success” or “failure”) with the test results (positive or negative), two by two tables with four cells (true positives, false negatives, false positives and true negatives) could be generated, and test qualifiers, such as sensitivity, specificity, predictive values and likelihood

ratios with 95% confidence intervals were calculated. Calculations were done with MetaDisc statistical software version 1.4 (Zamora et al. 2006). Statistical pooling was only performed if studies on a specific test were not hampered by statistical or clinical heterogeneity. Statistical heterogeneity was defined as non-overlapping 95% confidence intervals for estimates of sensitivity and specificity, and a difference in these estimates

among the studies of more than 20% (Jellema et al. 2010; van der Windt et al. 2010). We considered studies as clinically heterogeneous when patient groups, outcome measures, or the execution of index tests were different. In cases of statistical or clinical heterogeneity we refrained from statistical pooling and the results were presented per individual study.

Findings and discussion

Survey Dutch spine surgeons (study I)

Of 132 active spine surgeons, 93 (70%) completed and returned the questionnaire. Thirty-one of the 93 respondents (33%) declared not to perform spinal fusion for low back pain and were excluded from further analysis. The characteristics of the final group of 62 respondents are listed in Table 5. The level of experience for neurosurgeons and orthopaedic surgeons was equal: 11 of 16 versus 33 of 46 worked ten years or more in clinical practice, respectively.

Prognostic factors

The respondents' comments on prognostic factors are listed in Table 6. For 7 of the 11 items there was no consensus (significantly less than 70% uniformity of opinion).

More than 70% of the respondents would fuse patients over 60 years old for back pain. Years of clinical experience or specialty did not appear to be of influence ($p=0.5$, and $p=0.7$, respectively). Only 1 of 15 neurosurgeons fused patients below 20 for back pain, versus 14 of 46 orthopaedic surgeons ($p=0.06$).

Eighteen orthopaedic surgeons performed fusion of 3 or more levels for low back pain, whereas no neurosurgeon did ($p=0.003$).

Tests for patient selection

The surgeons' appreciation and use of predictive tests are listed in Tables 7a and 7b, respectively. Apart from MRI, there

was no uniformity regarding the value of these tests for clinical decision making.

Mainly orthopaedic surgeons (21 of 46, versus 2 of 16 neurosurgeons, $p=0.03$) considered provocative discography to be a valid predictor of fusion. Spine surgeons working in general hospitals (20 of 43), appeared to believe more in the test than academic surgeons did (3 of 18, $p=0.03$). There was no relation with clinical experience ($p=0.4$). Apart from the use of discography, differences in discipline or clinical experience did not appear to be of significant influence on treatment strategy. In the evaluation of chronic low back pain no other predictive tests than those mentioned in Tables 7a and 7b were used on a regular basis.

Table 6. Survey: Respondents' opinion to what extent patient-specific prognostic factors influence their clinical decision making in the treatment of CLBP. The numbers listed are percentages of valid responses

	Patient-specific prognostic factors			p-value ^a
Maximum number of levels for fusion	1 level 18 (30.5)	2 levels 23 (39.0)	≥ 3 levels 18 (30.5)	<0.001
Minimum age patient	< 20 yrs 15 (24.6)	20–30 yrs 25 (41.0)	≥ 30 yrs 21 (34.4)	<0.001
Maximum age patient	40–50 yrs 5 (8.1)	50–60 yrs 12 (19.4)	≥ 60 yrs 45 (72.5)	NS
Minimal length conservative therapy	< 6 mo 3 (4.8)	0.5–1 yr 36 (58.1)	≥ 1 yr or 23 (37.1)	NS
Maximum Body Mass Index (BMI)	< 31 29 (46.8)	31–37 18 (29.0)	≥ 37 15 (24.2)	<0.001
Maximum number of cigarettes / day	0 29 (47.5)	1 to 20 7 (11.4)	20 or more 25 (40.9)	<0.001
Referral overweight patients to dietician	Always 29 (46.8)	Sometimes 20 (32.3)	Never 13 (21.0)	<0.001
Psychological screening referral	Always 10 (16.2)	Sometimes 28 (45.2)	Never 24 (38.7)	<0.001
Different criteria for primary DDD versus prior spine surgery	Agree 44 (71.0)	Neutral 8 (12.9)	Disagree 10 (16.1)	NS
Work status affects outcome	Agree 29 (46.7)	Neutral 17 (27.4)	Disagree 16 (25.9)	<0.001
Litigation procedures affect outcome	Agree 43 (69.3)	Neutral 9 (14.5)	Disagree 10 (16.2)	NS

Abbreviation: DDD = degenerative disc disease, NS = not significant.

^a Chi-square test: $p<0.05$ means significantly less than 70% consensus, NS implies uniformity.

Table 5. Survey: Characteristics of the 62 respondents

	Orthopaedic surgeons (n)	Neurosurgeons (n)	All respondents (n)
No. of respondents	46	16	62
Age			
< 50 years	22	10	32
≥ 50 years	24	6	30
Clinical experience			
< 10 years	13	5	18
≥ 10 years	33	11	44
Type of hospital			
University/specialised	13	5	18
General	33	11	44
No. of fusions for CLBP/year			
1–10	24	9	33
10–25	9	6	15
25–50	7	1	8
≥ 50	6	0	6

Abbreviation: CLBP = chronic low back pain; n = number

Table 7a. Survey: Respondents' opinion on predictive tests for clinical decision making. The numbers listed are valid responses and respective percentages

Predictive test	Agree (%)	Neutral (%)	Disagree (%)	p-value ^a
MRI sufficient for decision making	10 (16)	11 (18)	41 (66)	NS
Cast immobilisation valuable test	25 (40)	15 (24)	22 (35)	<0.001
Cast immobilisation too unpleasant	11 (18)	16 (26)	35 (56)	0.03
PD proven valuable test	23 (38)	16 (26)	22 (36)	<0.001
PD too many complications	3 (5)	14 (23)	44 (72)	NS
TETF valuable test	8 (13)	33 (55)	19 (33)	0.01
TETF too many complications	20 (33)	31 (51)	10 (16)	0.001

Abbreviations: MRI = magnetic resonance imaging, PD = provocative discography, TETF = temporary external transpedicular fixation, NS = not significant.

^a Chi-square test: $p < 0.05$ means significantly less than 70% consensus, NS implies uniformity.

Table 7b. Survey: The use of predictive tests by the surgeons in clinical practice. The numbers listed are valid responses and their respective percentages

Use of test	Always (%)	Sometimes (%)	Never (%)	p-value ^a
Facet joint blocks	5 (8)	32 (52)	25 (40)	0.002
Cast immobilisation	20 (33)	23 (38)	18 (30)	<0.001
PD	25 (42)	10 (17)	24 (41)	<0.001
TETF	0 (0)	3 (5)	58 (95)	NS

Abbreviations: PD = provocative discography, TETF = temporary external transpedicular fixation, NS = not significant.

^a Chi-square test: $p < 0.05$ means significantly less than 70% consensus, NS implies uniformity.

Individual decision making in clinical practice

Figure 5 shows the importance of predictive tests and prognostic factors for clinical decision making as rated on a scale from 0 to 10. Patient history and imaging were valued significantly higher than predictive tests, psychological screening or patient preferences (all respective comparisons: $p < 0.01$, Independent t-test).

Twenty-seven (45%) surgeons responded to have a protocol for decision making to which they frequently or always adhered. Of those 35 respondents who did not have such a protocol, 23 (68%) replied that there should be guidelines. In other words, 50 respondents (83%) felt that clinical guidelines in the management of CLBP patients are prerequisite.

Comparison with related research

Older age is an acknowledged predictor of poor outcome (DeBerard et al. 2001). Nevertheless, almost three quarters

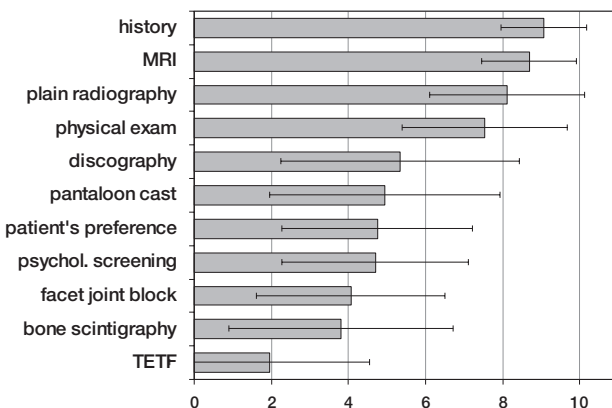


Figure 5. Survey: The importance of listed factors in clinical decision making (presented as mean ± SD) as rated by the respondents on a scale from 0 (no importance) to 10 (maximal importance). Abbreviations: MRI = magnetic resonance imaging, TETF = temporary external transpedicular fixation.

(73%) of the respondents fused patients above 60 for low back pain.

Despite the fact that 2 or 3-level fusions have been reported to have proven higher rates of pseudarthrosis with lower patient satisfaction as compared to single level fusions (Turner et al. 1992), over 30% of the respondents would consider fusion of 3 levels or more.

Although fusion surgery is not recommended unless 2 years of conservative treatment have failed (Airaksinen et al. 2006), 63% of the surgeons felt that less than 1 year of conservative therapy is enough to consider fusion.

Obesity is an independent risk factor for low back pain, and surgery in these patients is significantly associated with major complications, such as thrombo-embolism and infection (Patel et al. 2007). Nevertheless, 53% of the respondents would operate for chronic low back pain on obese patients and 24% on the morbid obese. Less than half of the surgeons (47%) consistently referred overweight patients to a dietician.

There was no consensus regarding smoking, which is known to be an independent risk factor for low back pain (Deyo and Bass 1989) and associated with worse results of spinal fusion (Wilson-MacDonald et al. 2008). About 41% would fuse heavy smokers, whereas 48% would not operate smokers for back pain.

Psychologically stressful work has been associated with low back pain and disability (Katz 2006), and it has been reported that psychological distress, depressive mood and somatisation lead to an increased risk of chronicity (Pincus et al. 2002). In addition, presurgical depression is associated with worse patient outcome after lumbar fusion (DeBerard et al. 2001). Only 16% of the respondents referred patients routinely for psychological screening and 39% never referred for this purpose at all.

There is strong evidence that clinical interventions are not effective in returning patients back to work once they have

Table 8. Motion analysis: Patient characteristics and angles of lumbosacral sagittal rotation (degrees) in the unrestricted condition and in both plaster cast conditions, while performing spinal flexion-extension, pelvic tilt, and walking

Patient no.	Sex	Age	External fixation level	Spinal flexion-extension			Pelvic tilt			Walking		
				No plaster cast	Cast – hip fixation	Cast + hip fixation	No plaster cast	Cast – hip fixation	Cast + hip fixation	No plaster cast	Cast – hip fixation	Cast + hip fixation
1	M	43	L4–S1	3.0	4.9	0.6	3.5	3.4	0.2	2.2	1.0	2.6
2	M	46	L4–S1	6.7	2.4	2.2	0.2	0.3	1.9	1.2	2.8	1.8
3	M	41	L4–S1	6.3	1.7	1.3	0.5	1.2	0.6	2.1	2.1	2.0
4	M	30	L4–L5	0.5	1.9	2.0	0.5	2.7	1.9	1.1	0.9	1.1
5	F	34	L5–S1	0.9	0.5	0.5	2.1	0.9	0.9	1.6	1.7	1.4
6	M	49	L5–S1	2.8	0.1	0.2	0.4	1.6	0.7	2.4	2.7	1.7
7	F	25	L5–S1	0.1	0.3	0.6	0.7	1.0	0.8	1.0	1.0	0.8
8	M	37	L5–S1	0.7	0.4	1.1	1.2	1.5	1.2	0.8	6.4	2.1
9	F	28	L5–S1	1.1	0.7	1.2	1.2	1.6	1.4	3.2	1.1	3.5
10	M	34	L5–S1	2.0	2.3	0.8	0.7	0.4	0.7	2.5	3.7	1.9

been off work for a longer time (Airaksinen et al. 2006). About half of the respondents agreed that the work status of patients with low back pain affects outcome considerably and 69% acknowledged that litigation or workers' compensation are of great influence on decision making, as they have been associated with persisting pain and disability (Katz 2006).

Two-thirds (66%) of the respondents considered findings on plain radiographs and MRI-scan alone to be insufficient for surgical decision making (Table 7a). This is in accordance with the literature indicating that degenerative or black discs on MRI do not appear to have a strong clinical relevance (Boden et al. 1990; Jarvik et al. 2001) and that there is no correlation between radiographic signs of degeneration and clinical outcome (van Tulder et al. 1997).

Opinion differed about trial immobilisation with a pantaloons cast: 40% of the respondents agreed that it is a valuable test and 36% disagreed. This resembles conflicting reports from the literature claiming that the test is not predictive of fusion outcome (Axelsson et al. 1995) or that only in highly selected patient groups the pantaloons cast test may be of value (Willems et al. 2006).

Provocative discography is a controversial test, which is highly variable in chronic pain patients and can also be positive in pain-free individuals (Carragee et al. 2000). Its value in predicting the outcome of fusion for low back pain is debated (Colhoun et al. 1988; Willems et al. 2007), which was reflected in the completely contradictory respondents' opinions. Trial immobilisation with a temporary external fixator is known for its high complication rate (Bednar 2001) and because of ambiguous results, its use is not recommended (Elmans et al. 2005). In the present survey, external fixation was not frequently used (94% never used it) and only 13% believed in its predictive value.

Lumbar facet injections have been reported not to be predictive of either arthrodesis or nonsurgical treatment of back pain (Esses and Moro 1993). Accordingly, only 8% used facet joint blocks on a regular basis as a predictor of spinal fusion.

The effect of an orthosis on lumbosacral motion (study II)

Translations were around 1 mm which was too small given the accuracy of the system, and were excluded from analysis. Noteworthy rotations of more than 1.0 degrees, were only found in the sagittal plane (Table 8). Although larger LS motion has been reported while measuring with the patient in lateral decubitus position, we chose to measure patients in standing and walking positions because for us, the immobilising effect of a plaster cast in common daily life activities was of interest.

In the flexion-extension exercise (Figure 6a), the patients with motion measured across two segments (L4–S1, patients 1, 2 and 3) showed considerably higher rotation and significantly more reduction of motion by both plaster casts than the other patients ($F(1,8) = 6.1$; $p = 0.04$). During walking (6c), considerably more sagittal rotation was found for both cast conditions than in the other two exercises, reaching almost the level of significance ($F(1,9) = 4.99$; $p = 0.05$). The decrease in motion by both plaster casts in the static measurements was therefore not present during walking. This interesting observation deserves further attention, although the small number of patients and the generally low values of sagittal rotation in this study may have limited statistical value.

Our findings are consistent with those of a study (Lumsden and Morris 1968) in which Steinmann pins were inserted into the spinous process of L5 and the posterior superior iliac spines and a lumbar orthosis appeared to rather increase than decrease lumbosacral rotations during normal levels of walking. Buchalter et al. (1988) reported that casts and orthoses do not provide sufficient rigidity to adequately immobilise the lumbar spine. With a rigid support, optimal restriction of movement is achieved approximately midway in length and decreases toward the ends (Fidler and Plasmans 1983). Adding hip fixation to the support should thus provide maximal stabilisation of the LS joint. However, Axelsson et al. (1993), using

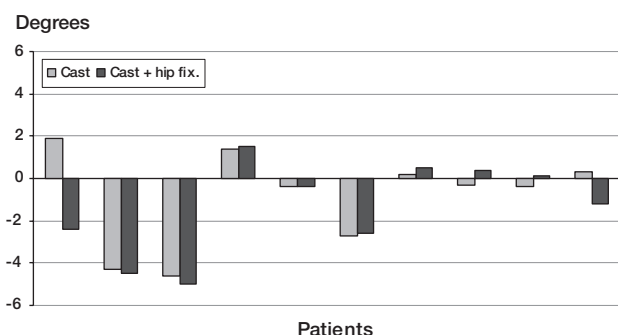


Figure 6a. The differences (Delta) in range of sagittal rotation from both plaster casts to the range obtained in each patient in the condition with no support, during spinal flexion-extension. Note that a negative delta represents an immobilising effect.

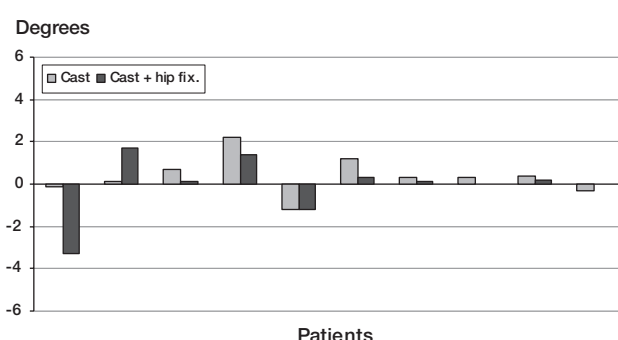


Figure 6b. The differences (Delta) in range of sagittal rotation from both plaster casts to the range obtained in each patient in the condition with no support, during pelvic tilt. Note that a negative delta represents an immobilising effect.

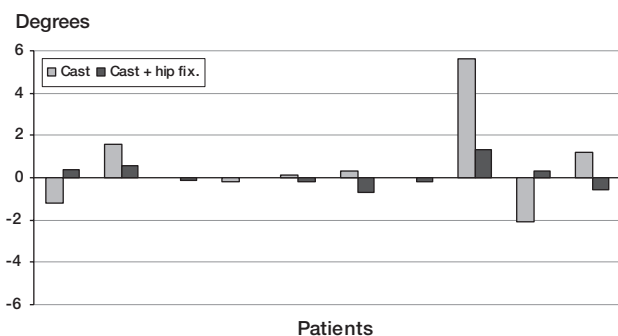


Figure 6c. The differences (Delta) in range of sagittal rotation from both plaster casts to the range obtained in each patient in the condition with no support, during walking. Note that a negative delta represents an immobilising effect.

roentgen stereophotogrammetry, found no significant stabilising effect by a molded, rigid TLSO with unilateral hip immobilisation on sagittal, vertical, or transverse translations. They suggested that plaster casts restrict gross motions of the trunk rather than exert a specific stabilising effect on intervertebral mobility of the lumbosacral joint.

The accuracy of the PRIMAS system for static and moving markers in our experimental set-up appeared to be 0.3 degrees,

which implies that rotational ranges less than 0.6 degrees should be analysed with great caution.

Additionally, the standardised position in which the patients were placed in the field of measurement, may have differed between the exercise sequences. Calculation showed that a deviation of 10 degrees in position produced an error of 1.6% in sagittal rotation, which we deemed an acceptable limitation for the experiment. In that respect, PRIMAS seems to be an accurate method for continuous registration of lumbosacral kinematics during dynamic exercises in patients with chronic low back pain.

Cohort study: Value of an orthosis to predict clinical outcome (study III)

Of the 257 analyzed patients, 98 experienced significant pain relief when wearing the pantaloone cast, whereas 159 patients had a negative test result. Apart from inconvenience, no complications from wearing the cast were noted. There was no significant correlation between the pantaloone cast test result and the advice from the TETF trial ($\rho=0.01$, $p=0.891$). Based on the TETF, 107 patients actually underwent lumbar fusion: 79 patients underwent an instrumented posterolateral intertransverse process fusion and 28 patients had an anterior interbody fusion. For both operation techniques about 40% of the patients had a clinically relevant improvement.

According to the logistic regression model, only the interactions between treatment and the pantaloone cast test result ($p=0.08$) and that between treatment and whether the patient had undergone prior spine surgery ($p=0.004$) were factors contributing to successful pain relief in the study population. These interaction relations were tested by a separate Chi-square analysis on clinical outcome x pantaloone cast test result both for patients with and without a history of prior spine surgery. For the group of 99 patients with no prior spine surgery, the patients with a positive pantaloone cast test who underwent a lumbar fusion, had a significantly better outcome than those managed conservatively ($p = 0.002$, Table 9a). Those with a negative test showed no difference in outcome between fusion and conservative treatment. In the group of 158 patients with prior spine surgery no difference in outcome was found (Table 9b). None of the patient groups showed a significant difference in outcome after lumbar fusion between the patients with a positive and those with a negative pantaloone cast test.

The use of a plaster cast test in surgical decision-making for chronic LBP patients remains controversial. In search for scientific proof, the clinical outcome of 257 chronic LBP patients who either underwent lumbar fusion or conservative management was assessed in the present cohort study. In patients with no history of prior spine surgery and a positive pantaloone cast test, the success rate for lumbar fusion was significantly higher than that for conservative management ($p=0.002$), although the number of patients in both subgroups

Table 9a. Orthosis: Summary of results in the 99 patients without prior spine surgery, classified by treatment and pantaloon cast test result

Pantaloon cast test:	Lumbar fusion		Conservative management	
	A. Positive	B. Negative	C. Positive	D. Negative
Number of patients	15	22	20	42
Median duration of pain, years (range)	3.0 (0.5–10)	3.8 (1.0–18)	2.5 (0.5–20)	3.3 (0.5–22)
Median follow-up, months (range)	66 (40–137)	77 (18–124)	52 (32–115)	69 (30–144)
Baseline VAS, mean (SD)	74 (15)	72 (18)	66 (17)	72 (14)
Post-treatment VAS, mean (SD)	32 (26)	43 (34)	67 (21)	56 (29)
Pain relief of $\geq 30\%$ on VAS ^a , n (%)	11 (73%)	13 (59%)	4 (20%)	17 (40%)

^a Significantly more pain relief of $\geq 30\%$ on VAS in Group A compared to Group C ($p = 0.002$).

No significant difference in pain relief between the Groups A and B ($p = 0.4$), B and D ($p = 0.2$) or C and D ($p = 0.1$)

Table 9b. Orthosis: Summary of results in the 158 patients with prior spine surgery, classified by treatment and pantaloon cast test result

Pantaloon cast test:	Lumbar fusion		Conservative management	
	A. Positive	B. Negative	C. Positive	D. Negative
Number of patients	28	42	35	53
Median duration of pain, years (range)	3.5 (1.0–20)	4.0 (0.5–31)	4.0 (1.0–22)	3.0 (1.0–21)
Median follow-up, months (range)	53 (18–138)	89 (23–137)	89 (30–136)	89 (23–137)
Baseline VAS, mean (SD)	76 (19)	77 (16)	74 (17)	77 (16)
Post-treatment VAS, mean (SD)	61 (28)	67 (29)	65 (29)	63 (27)
Pain relief of $\geq 30\%$ on VAS ^a , n (%)	8 (29%)	12 (29%)	9 (26%)	21 (40%)

^a No significant differences in pain relief between any of the groups

was relatively small (15 and 20 patients, respectively, Table 9a). Within the lumbar fusion treatment group, the outcome between patients with a positive pantaloon cast test was not significantly different from that for patients with a negative test (73% versus 59%, respectively, Table 9a). Thus, a negative pantaloon cast test does not mean that lumbar fusion will not be successful. In patients with prior spine surgery, the success rate for lumbar fusion was low and the pantaloon cast test was of no value (Table 9b). If in these patients equivocal results exist after routine diagnostic tests, great caution with regard to lumbar fusion is warranted. These findings are in accordance with other reports, stating that there are no reliable tests for preoperative selection of failed back surgery patients and that, consequently, the results of lumbar fusion are discouraging (Lehmann et al. 1987; Turner et al. 1992; Wetzel et al. 1994).

Although we realise that pain is merely one of the many factors related to the problem “chronic low back pain”, the criterion of pain reduction on a Visual Analogue Scale was chosen to assess successful outcome. Despite the inherent inaccuracy in how the patient puts a dot on the line, it is our opinion that, because the patient serves as his own control, this is an objective method to evaluate outcome of treatment in chronic pain (Price et al. 1983). Because of the heterogeneity in VAS scores within each group, the difference between pain on VAS before allocation and at follow-up was calculated on an individual

level and not on a group level. In this way, it was possible to assess clinically relevant improvements in pain. It should be acknowledged that the cut-off level of 30% pain reduction to define success or failure is arbitrary and debatable and one should realise that, as we know from clinical practice, many of the cases defined as success in the present study still suffer from residual low back pain after treatment.

The length of follow-up did not appear to be a contributing factor to pain reduction in the logistic regression ($p=0.86$). Nevertheless, the large range in follow-up and thus the great variety in interval between the pre- and posttreatment pain assessment, should be regarded as a limitation of the present cohort study.

TETF is an invasive test and high incidences of pin tract infections and neurological complications have been reported (Jeanneret et al. 1994; Lund et al. 2002). Therefore, the TETF trial was only used as a final diagnostic test in patients for whom the indication for lumbar fusion remained unclear after routine diagnostic tests. The patient population in the current study might thus be biased because “clear-cut” cases that were indicated for fusion straight away were not included in the trial. This may limit the generalisability of the current study’s findings in other settings and may have led to lower overall results for lumbar fusion. Our aim, however, was not to assess the success of lumbar fusion in chronic low back pain patients, but to evaluate whether a pantaloon cast test can help select

those patients who will benefit more from lumbar fusion than from nonsurgical treatment.

Naturally, a randomised study would be the best way to evaluate a preoperative selection test. Within the limits of current clinical practice we considered the present cohort study design to be a reasonable alternative as the pantaloon cast test result had no correlation with the surgical decision-making based on the TETF trial ($\rho=0.01$, $p=0.9$). In this way, whether after a positive or a negative pantaloon cast test, the groups of fused and conservatively managed patients were fairly equal in size and had similar patient characteristics making analysis between these groups more valid.

Founded on the surgeon's personal belief or experience rather than supported by scientific literature, the pain-relieving effect obtained by spinal immobilization in a lumbar orthosis or pantaloon cast is widely used for patient selection in clinical practice. The mechanical basis, however, is poor, as several studies could not determine a stabilising effect on intervertebral mobility (Axelsson et al. 1992; Axelsson et al. 1993; Willems et al. 1997). In fact, radiographic studies have even suggested that motions at lower lumbar levels actually increase when orthoses are worn (Axelsson et al. 1993; Tuong et al. 1998). The main effect produced by a corset is a reduction in gross motion of the trunk (Lantz and Schultz 1986; Buchalter et al. 1988) rather than the fixation at one individual level. This gross effect makes comparison to a single level fixation questionable, which is supported by the fact that there appeared to be no correlation between the results of the TETF trial and the pantaloon cast test. The pain relief provided by a corset is probably not caused by motion restriction but by an increase in passive trunk stiffness of up to 40% (Cholewicki et al. 1999; Cholewicki et al. 2003) with a possible general unloading effect on the trunk structures and reduced intradiscal pressures (Nachemson and Morris 1964). Such pain relief caused by this unloading effect does not appear to be specific, as we found that a negative pantaloon cast test does not at all exclude the possibility that the patient will benefit from lumbar fusion.

Adjacent level discography to predict the clinical outcome of fusion (study IV)

No serious complications related to the discography procedure were registered. If degenerative discs are considered to be painful and normal discs to be asymptomatic, morphology predicted the discographic pain response in about 75% of the patients (21 painful and degenerative discs, 126 asymptomatic discs with normal morphology, Table 10). About two-thirds of the degenerative adjacent discs were painful as compared to a quarter of the adjacent discs with normal morphology ($p = 0.000$ chi-squared test, OR 6.2, 95% CI 2.7–13.9). There was no correlation ($\rho=0.1$, $p=0.3$) between the result of provocative discography and the TETF trial result. According to

Table 10. Adjacent level discography: Cross-tabulation of the morphologic (Adams' criteria) and the symptomatic status of the adjacent discs assessed by provocative discography in all 197 patients

	Symptomatic	Not symptomatic	Total
Degenerative	21	11	32
Normal morphology	39	126	165
Total number of patients	60	137	197

the logistic regression model, only treatment (lumbar fusion versus conservative management) appeared to be a contributing factor to successful pain relief ($p=0.01$, OR 2.2, 95% CI 1.2–3.9). One can conclude that a patient in this study who underwent a fusion operation had approximately twice as much chance of having at least a 30% reduction in pain as a patient who was treated conservatively. The other variables in the logistic regression did not contribute significantly to having at least a 30% reduction in pain: age ($p=0.5$), gender ($p=0.2$), preoperative VAS score ($p=0.8$), previous spine surgery ($p=0.06$), interaction previous spine surgery and treatment ($p=0.98$), provocative discography ($p=0.99$) and length of follow-up ($p=0.5$).

For the 82 fused patients the distributions of the VAS scores before lumbar fusion and at follow-up are shown in Figure 7. The success rate of those fused with asymptomatic adjacent discs was equal to the success rate of those fused with painful adjacent discs (45% versus 45%, respectively, $p=0.6$, Table 11). When considering the effect of discographic morphology on the clinical outcome for lumbar fusion between those patients with degenerative adjacent discs versus those with normal adjacent discs, the Pearson's Chi Square test was not significant ($p=0.07$). Stratification for patients with or without prior spine surgery did not reveal any effect of discography result on outcome (Pearson's Chi Square: no prior surgery, $p=0.8$; prior surgery, $p=0.7$). The method of fusion, whether anterior interbody fusion or instrumented posterolateral fusion, had no effect on outcome (success rates of 46% versus 45%, respectively). The outcome by percentage of VAS improvement and the 4 categories of patient satisfaction (excellent, good, fair or poor) appeared to correlate well ($\rho=0.73$, $p=0.000$). Patient satisfaction, which we consider to be a sufficient quality indicator rather than another validated measure of health outcome, also did not delineate differences between the two operative groups ($p=0.3$).

There is no true consensus in the management of adjacent segments in spinal fusion. Some authors advocate to limit the number of segments fused as much as possible because two- or three-level fusions have a proven higher rate of pseudarthrosis with less function (Dawson et al. 1981; Jackson et al. 1985; Herkowitz et al. 1999) and lower patient satisfaction compared to single level fusions (Turner et al. 1992). Others suggest that surgical fusion of lumbar vertebrae increases the load on adjacent discs which may accelerate degeneration at

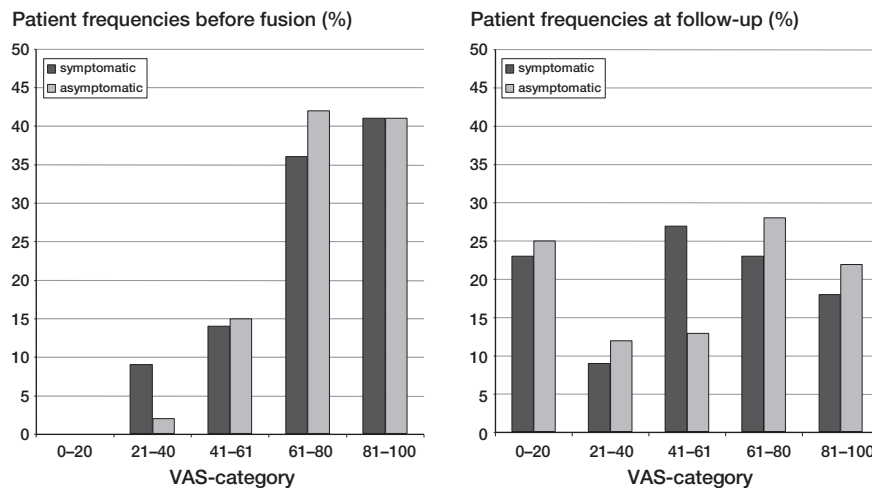


Figure 7. Adjacent level discography: The distribution of the VAS scores before lumbar fusion and at follow-up for all fused patients with symptomatic and asymptomatic adjacent discs at preoperative discography. The VAS scores have been categorised as follows: 0–20, 21–40, 41–60, 61–80, and 81–100.

Table 11. Adjacent level discography: Variables and clinical outcome of the 82 patients who underwent lumbar fusion classified by the result of provocative discography of the adjacent segments

	Asymptomatic adjacent disc(s)	Symptomatic adjacent disc(s)
Number of patients	60	22
Mean age (years)	39 (7.8)	39 (8.5)
Gender (male / female)	18 / 42	6 / 16
Mean follow-up in months (range)	80 (36) (15–137)	76 (38) (18–144)
Mean initial VAS	76 (16)	72 (18)
Mean VAS at follow-up	53 (33)	52 (29)
Patients with pain relief of $\geq 30\%$ on VAS	27 (45%)	10 (45%)

these transitional levels (Lee and Langrana 1984; Lehmann et al. 1987; Sato and Kikuchi 1993; Chow et al. 1996). However, in a long-term follow-up study following surgery for lumbar degenerative disc disease, no correlation between radiographic signs of degeneration at transitional levels and clinical outcome or functional impairment was reported (Kumar et al. 2001b). In a recent retrospective study Throckmorton et al. reviewed the outcome after lumbar fusion for degenerative instability in 25 patients (Throckmorton et al. 2003). They reported no adverse impact on clinical outcome when the lumbar fusion ended adjacent to a degenerative motion segment as determined from preoperative morphologic MRI data. As a criticism one might argue that radiographically degenerative lumbar discs or so-called black discs on MRI do not appear to have clinical relevance as they are observed in up to a third of asymptomatic individuals between 20 and 39 years of age (Modic et al. 1988; Boden et al. 1990; Hambly et al. 1998).

Because degeneration itself appears to be related to the process of aging, whereas symptomatic degeneration may point to deterioration, one could assume that physiologic data from provocative discography would reflect the clinical scenario better in chronic LBP patients. This more reliable preoperative analysis of planned future transitional levels should lead to a better prediction of the clinical outcome of lumbar fusion (Gillet 2003). In the present study, however, clinical outcome did not appear to be affected by the preoperative status of adjacent levels as assessed by provocative discography. There are, however, methodological limitations of the present study that should be addressed.

Provocative discography is a controversial test and its validity has not been established consistently in literature (Nachemson 1989). It has been shown that healthy, previously pain-free patients can also experience pain from a discogram and the discographic pain response appears to be highly variable in patients with chronic pain and somatization disorders (Carragee et al. 2000). It should be stressed that discography is not a test at the level of a gold standard to identify whether a disc is truly a clinically significant pain generator in a chronic LBP patient. It is no more than an assumption frequently given to the discography result which can be compared to clinical outcome.

In the present study, provocative discography was executed by residents who had performed a sufficient number of discography procedures under supervision to be judged capable of executing the procedure alone. Nevertheless, it has been documented that false test results increase with poor injection technique implying that the performance and evaluation of the test by residents may be less reliable than by experienced discographers who perform discography on a routine base.

The decision of which levels were to be fixed by TETF was based on the combination of conventional radiographs, MRI/CT and provocative discography that included the assessment of generally two adjacent discs. In this way, in the individual patient, an originally adjacent disc that was symptomatic on injection, may have become a suspect disc and subsequently been included in the TETF (=proposed fusion) levels. This may have resulted a bias in the selection of both fusion groups. For the present protocol, adjacent levels were defined retrospectively and the discographic responses of only these levels, truly adjacent to TETF (=fusion), were studied. Consequently, in some patients only one adjacent level could be scored for analysis, which may limit the interpretation of the present study's findings.

However, in contrast to most studies, the design of the present study had the advantage that the decision to perform

lumbar fusion was not based on the results of adjacent level discography itself (as defined retrospectively) but on the TETF results, which were not correlated ($\rho=0.1$, $p=0.3$). In this way, a group of fused patients with symptomatic or degenerative adjacent discs and a group of fused patients with normal adjacent discs could be compared.

Because TETF is an invasive diagnostic procedure with a considerable risk of pin tract infection and neurological complications (Soini and Seitsalo 1993; Jeanneret et al. 1994), only patients for whom routine diagnostic testing was equivocal were included. Such a selection manner is, of course, prone to bias by the judgment of the individual clinician. Moreover, whether the TETF trial was positive (sufficient pain relief to decide for lumbar fusion), was determined by the physician who coordinated the TETF trial. For his decisions there is no validity other than clinical experience. These methods of selection limit the generalisability of the present study to settings in which patients are selected for lumbar fusion in a different manner. The aim, however, was not to assess the success of fusion in chronic LBP patients but to evaluate whether the clinical outcome of lumbar fusion is affected by the preoperative status of adjacent discs as determined by provocative discography.

Although pain is only one aspect of the problem of “chronic low back pain”, it remains the principal indication for operative treatment. For this reason, the criterion of significant pain reduction on a VAS was chosen to define successful outcome. The graphic representation of pain over time by the VAS has been shown to be well suited for longitudinal outcomes assessment (Price et al. 1983; Zanoli et al. 2001; Hagg et al. 2003). We chose for the change in pain per individual to define outcome because of the heterogeneity in VAS scores within each patient group. It should be acknowledged that the choice for the cut-off level of 30% improvement in pain to define success or failure is arbitrary and debatable as it implies (as we also know from current clinical practice) that many of the patients defined as a success in the present study still suffer from residual LBP after treatment.

The spread in time between the pre- and posttreatment pain assessments could be a confounding factor (in both fusion groups there were patients with less than 24 months of follow-up). In the logistic regression, however, length of follow-up was not found to contribute to the clinical outcome of successful pain relief ($p=0.5$).

The total number of 82 fused patients in the present study is small and could lead to a potential type II error. A sample size calculation (power=0.8, alpha set at 0.05) showed that 67 patients are sufficient to show a 15% difference in the rate of success between the two groups (a difference less than 15% would be of little clinical relevance). Therefore, we feel confident that the results presented here justify our critical view concerning the value of evaluation of adjacent levels by provocative discography prior to spinal fusion.

Postdiscography discitis (study V)

Discography was performed in exactly 200 patients (88 men, 112 women) with a mean age of 41 (16–62) years. Forty patients had discography at one level, 85 patients at two levels, and in 75 patients three levels were involved. Thus, in total 435 discs were injected. Three patients had reported signs of discomfort or fever at follow-up visit. Laboratory controls, however, showed no significant elevation of ESR or CRP and MRI was negative for discitis. As their symptoms resolved spontaneously, discitis was excluded. As all patients could be reached by phone, 100% follow-up was achieved. No patient reported complaints or fever or discomfort regarding his or her back in the first 3 months after discography, and thus an overall incidence of 0% postdiscography discitis was registered.

Literature search

The search yielded 263 references, of which 217 papers were excluded on the basis of an abstract and title. After applying the inclusion criteria to the remaining 46 papers, and after screening additional references, 10 papers, spanning a period of 40 years, were finally selected (Table 12). Two of these studies (Osti et al. 1990; Carragee, Chen et al. 2000) actually reported a minimum follow-up of 3 months, whereas the remaining studies did not specify the exact term of follow-up. In only one study prophylactic antibiotics had been used, and a postdiscography incidence of 0% was found (Osti et al. 1990). In the other nine studies, no prophylactic antibiotics had been administered. The studies revealed 12 cases of discitis in 4891 patients or 12,770 discs with an incidence of 0.25% by patient and 0.09% by disc. Three of these studies (Patrick 1973; Jackson et al. 1989; Carragee et al. 2000) as well as the current study, reported no case of discitis. Combining the literature with the current patient study leads to an overall incidence of 0.24% by patient and 0.09% by disc.

Postdiscography discitis is considered to be due to bacterial infection (Fraser et al. 1986; Fraser et al. 1987) and should not be confused with the common exacerbation of backache after intradiscal injection most probably due to a mechanical or chemical cause. The most common causative agents are *Staphylococcus aureus* and *S. epidermidis*, although the incidence of gram-negative bacteria appears to be rising (Rimoldi and Haye 1996; Hoelscher et al. 2000; Tai et al. 2002). Fraser et al. (1987) isolated bacteria from the disc in three of four discitis patients who had open biopsy <6 weeks after discography, and they suggested bacteria introduced by the needle tip to be the initiators of the disease. To eliminate the risk of injecting a core of soft tissue including skin, the use of stilette needles and a two-needle technique was recommended, with which the incidence of discitis could be reduced from 2.7% to 0.7%. In a prospective clinical trial of 127 patients undergoing lumbar discography with cephazolin added to the contrast dye, Osti et al. (1990) reported no case of discitis and recommended the use of a single dose of prophylactic broad-

Table 12. Postdiscography discitis: Incidences in selected literature and current study by patient and by disc

Study	Year	Prophylactic antibiotics	No. of patients	No. of discs	Discitis cases	Incidence by patient (%)	Incidence by disc (%)
Collis and Gardner	1962	No	1,014	2,187	1	0.10	0.05
Patrick	1973	No	123	341	0	0.00	0.00
Colquhoun	1977	No	716	2,000	1	0.14	0.05
Milette and Melanson	1982	No	500	1009	3	0.60	0.30
Fraser et al.	1987	No	210	417	4	1.90	0.98
Guyer et al.	1988	No	2,014	6,042	2	0.10	0.05
Jackson et al.	1989	No	124	231	0	0.00	0.00
Simmons et al.	1991	No	164	465	1	0.61	0.22
Carragee et al.	2000	No	26	78	0	0.00	0.00
Current study	2003	No	200	435	0	0.00	0.00
Literature combined with current study	–	No	5,091	13,205	12	0.24	0.09
Osti et al.	1990	Yes	127	337	0	0.00	0.00

spectrum antibiotic either added to the radiographic contrast material or administered intravenously. However, Eismont et al. (1987) reported no detectable levels in discs of rabbits after intravenous administration of either cephalothin or oxacillin, and Gibson et al. (1987) was unable to retrieve any cephradine or flucloxacillin from human intervertebral discs intraoperatively in 25 scoliosis patients. Boscardin et al. (1992) found therapeutic levels in only a short period after intravenous administration of high doses of cefazolin, whereas Rhoten et al. (1995) found quantifiable levels of cefazolin and oxacillin in only 40% of human cervical discs. As the disc is an avascular structure with blood vessels found only at its margins (Urban et al. 1977), intradiscal nutrition and levels of permeation of drugs depend on passive diffusion through the adjacent endplates and the surrounding annulus fibrosis. The disc has a negatively charged structure, which implies that positively charged antibiotics, such as gentamycin may diffuse more rapidly and be more effective than negatively charged antibiotics, such as cephalosporins. In a prospective study of 20 patients undergoing spinal surgery, Tai et al. (2002) could detect no therapeutic level of cefuroxime in the disc, whereas gentamycin showed a 50% penetration from blood to disc, well within the therapeutic range. Like other researchers (Langmayr et al. 1995; Rhoten et al. 1995), they suggested that high doses of cephalosporins are required to achieve therapeutic levels in the disc. Important in this respect are the findings from Hoelscher et al. (2000) that high doses of antibiotics, including both cephalosporins and aminoglycosides, can have deleterious effects on the survival of cultured disc cells, cell proliferation, and metabolic rates. This may have serious implications if antibiotics are to be mixed with contrast dye and injected directly into the disc, creating local peak concentrations. Moreover, side effects of aminoglycosides are well known and sensitisation or anaphylactic reactions should always be anticipated.

In the present study, lumbar discography was performed in 200 consecutive patients by means of a stilette two-needle

technique under strict aseptic conditions without the use of prophylactic antibiotics. A discitis incidence of 0% was found. As method of study, the routine postclinical assessment was used, which we consider crucial in the detection of postdiscography discitis. It has been stated that underestimation of discitis incidence is possible because of the latent period between discography and the onset of symptoms, lack of clinical contact between patient and clinician, and lack of awareness by the clinician (Fraser et al. 1987; Rohde et al. 1998). Therefore, a routine clinical follow-up visit is mandatory. As the onset of symptoms of discitis after uneventful discography normally develops within a few weeks and at least within 10–12 weeks (Piotrowski et al. 1994; Rohde et al. 1998), we believe no case of discitis has gone undetected by our method.

In a review of 15 studies, The North American Spine Society (NAS) calculated an overall incidence of <0.15% by patient and <0.08% by disc. Not all of these studies reported the exact number of patients and those lost to follow-up, and no distinction was made between studies using a one- or a two-needle technique. In the present literature search, without the use of prophylactic antibiotics, overall incidences of 0.25% by patient and 0.09% by disc were found (Table 12). In one of these selected papers (Fraser et al. 1987), unusually high incidences (1.90% and 0.98%, respectively) in 210 patients were reported. This may have been caused by the fact that in 61 of these patients, discography was performed by less experienced physicians, leading to 3 cases of discitis (4.9%), as opposed to 1 case in 149 patients (0.7%) if the procedure had been carried out by an experienced discographer. In the present literature search, only one study of lumbar discography with the use of prophylactic antibiotics, in which discitis was specifically looked for, was selected (Osti et al. 1990). In this series of 127 patients (337 discs), no case of discitis was registered. However, to prove that additive antibiotics actually prevent postdiscography discitis, given the incidence of 0.25% by patient, a randomised trial of approximately 9000 patients would be needed to reach significance (calculated sample size

Table 13. TETF: Characteristics and mean VAS scores for those patients with a negative and those with a positive test result.

	Negative test	Positive test
Total no. of patients = 301	176	125
Age (years)	41.8 (9.3)	40.7 (8.7)
Duration of complaints, year	6.1 (6.7)	5.5 (5.1)
Length of follow-up, months	81 (35)	76 (34)
Previous spine surgery (%)	65	62
Pain on VAS		
– before TETF test	74 (15)	74 (18)
– at follow-up	63 (27)	54 (32)
– improvement (before – follow-up)	11 (28)	20 (32)

for dichotomous response variables). Taking into account the possible adverse effects of antibiotics, we feel there is not enough evidence to regard the routine use of prophylactic antibiotics indicated in lumbar discography. At present, as has been stressed before (Fraser 1996), it is considered imperative that discography be carried out by a physician who is very experienced with the procedure, performs it on a routine basis, and is meticulous with the two-needle technique and aseptic conditions.

Value of temporary external fixation to predict clinical outcome (study VI)

The main complications were pin track infection (20 of 330 patients, 2 needed surgical debridement, the other patients were treated with antibiotics) and nerve root irritation (26 patients (8%), all but 2 completely recovered after the test). Two patients with nerve root irritation had intractable pain and the external fixator had to be removed before completion of the test.

A positive test result was found in 125 patients, and a negative result in 176 patients. From those with a positive test, 70% (88 of 125) were treated by fusion, and 80% (141 of 176) of those with a negative test were not operated on. Both groups were comparable with regard to diagnosis, and there were no significant differences in VAS scores at follow-up between those with a positive and those with a negative test result (Table 13).

Spinal fusion was performed in 123 patients (41%) and 178 (59%) were treated conservatively. Median follow-up was 78 (18–144) months in the fusion group and 76 (15–144) months in the conservative group. Ninety operations were posterolateral fusions with pedicle screw instrumentation and 33 (27%) anterior fusions. At follow-up, no differences in results between these groups were found.

Within the fusion group, patients with a negative test showed an improvement on VAS of 11.2 (SD 33.5) points, whereas those with a positive test had a mean improvement of 22.2 (30.6) points. Because of the large within-group variations,

Table 14. Fusion group: Numbers and percentages of patients with a negative external fixation test and those with a positive test, respectively, who had a clinically relevant improvement of 30% or more on the VAS for pain

	Improvement on VAS (before test – follow-up)		
	< 30%	≥ 30%	Total
Negative test, n (%)	25 (71%)	10 (29%)	35 (100%)
Positive test, n (%)	48 (55%)	40 (45%)	88 (100%)
Total, n	73	50	123

this 11-point difference was not statistically significant ($df = 121$; $p = 0.08$).

Looking for a clinically relevant improvement, 45% of the fused patients with a positive test had a decrease in VAS of 30% or more, versus 29% of those with a negative test (Table 14). Again, this 16% difference was not significant, probably because of the same large within-group variation.

A serious drawback of the external fixation test is that it is an invasive procedure under general anesthesia. Misplacement of the screws can lead to neurological damage. The 8% nerve root irritation in this study is in accordance with the 0–9% reported in literature (Soini and Seitsalo 1993; Jeanneret et al. 1994; Faraj et al. 1997; Bednar 2001).

Continuous antibiotic prophylaxis appears mandatory: Since May 1997, flucloxacillin was administered orally in this trial for as long as the external fixator was in situ, which coincided with a drop in infection rate from 7% to 0%.

The results of this study show that in this heterogeneous group of patients with chronic low back pain, 62% of whom had earlier spine surgery, the external fixator test, with and without dynamisation, fails to predict the long-term outcome of spinal fusion, and that there is a risk of complications. Therefore, this test is not recommended as a tool for surgical decision making in clinical practice.

Systematic review of the literature (study VII)

Figure 8 shows the flow diagram of studies from initial results of database searches to final inclusion, according to the PRISMA guidelines 2009 (Liberati et al. 2009). Of the 22 selected full articles, six studies in which only patients with a positive index test had been selected for lumbar fusion, were excluded (Bednar and Raducan 1996; Lovely and Rastogi 1997; Madan et al. 2002; Axelsson et al. 2003; Ohtori et al. 2009; Peng et al. 2009). We also excluded six other studies, in which test accuracy could not be determined because only mean values of recovery were reported without proportions of patients with success or failure of treatment (Esses and Moro 1993; Christensen et al. 1996; Buttermann et al. 1997; Chataigner et al. 1998; Derby et al. 1999; Esposito et al. 2006). Finally, ten studies met the inclusion criteria.

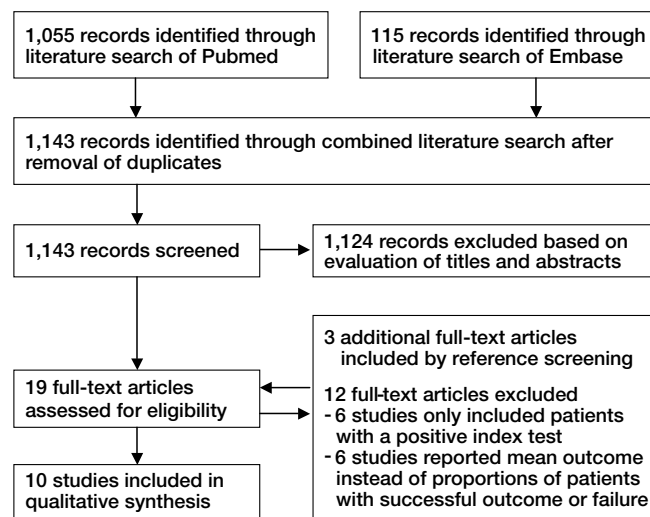


Figure 8. Systematic review: PRISMA flow diagram of combined literature search and selection.

Characteristics of included studies

Study characteristics are listed in Table 15. Three papers concerned immobilisation by an orthosis, either a fiberglass pantaloon cast, (Rask and Dall 1993) a canvas corset (Axelsson et al. 1995), or a plaster pantaloon cast (Willems et al. 2006). Four papers reported on discography, of which two studies focused on provocative discography of suspect levels (Colhoun et al. 1988), one study on provocation of the levels adjacent to the intended fusion (Willems et al. 2007), and the fourth study focused on the amount of degeneration as registered at discography in a group of patients with a positive discographic pain response (Gill and Blumenthal 1992). Three papers evaluated trial-immobilisation by external fixation, either with (Jeanneret et al. 1994; Elmans et al. 2005) or without dynamisation (Heini et al. 2004).

The sample sizes of the included studies ranged from 22 to 162. The length of follow-up ranged from 6 months or “when fusion was noted” (Esses et al. 1989) to 12 years. Either anterior or posterolateral fusion was performed. In those studies in which both procedures were performed, no difference in outcome between the two types of fusion was reported (Colhoun et al. 1988; Willems et al. 2006). Pain appeared to be the only measure of outcome that was consistently reported in all included papers. Five studies used a VAS to score pain, of which three studies defined a cut-off point of at least 30% decrease in pain as a clinically relevant improvement, one study a decrease of at least 75%, and one study “little or no pain on a VAS” (Table 16). The other studies used a subjective pain scale (“pain free” or “significant pain relief” versus “insignificant or no pain relief”). No study reporting on facet joint blocks or MRI could meet the inclusion criteria.

Risk of bias

The two reviewers agreed on 77 of the 120 items (64%)

scored. After discussion, consensus could be reached on all items. Most disagreements were because of reading errors or ambiguous reporting. The risk of bias ratings are listed in Table 16. The most prevalent shortcomings were: Uninterpretable index test results were not reported, no clear cut-off point was defined for positive and negative index test results, and in all but 3 studies not all patients who were tested underwent fusion regardless of the index test result (verification bias).

Test accuracy

Table 17 summarises test accuracies from the included studies.

Orthosis. An orthosis or cast immobilisation could neither confirm nor rule out a good outcome after spinal fusion with variable sensitivity (0.43 to 0.94; range positive likelihood ratio (LR), 0.94 to 1.13) and specificity (0.14 to 0.61; range negative LR, 0.39 to 1.12).

Provocative discography. The four studies reporting on the prognostic value of discography showed variable sensitivity (range 0.40–0.88; range positive LR 0.70–1.71) and specificity (0.25–0.48; range negative LR 0.24–1.40). In one study likelihood ratios were statistically significant (positive LR 1.71; 95% CI 1.21–2.41 and negative LR 0.24; 95% CI 0.13–0.43), but specificity was low, 0.48, meaning that only half of the patients who would not improve after fusion could be detected.

Temporary external transpedicular fixation. Sensitivity was generally high (range, 0.80–0.93; range positive LR 1.02–1.74), but specificity was low (range 0.20–0.47; range negative LR 0.15–0.94). In one study likelihood ratios were statistically significant (positive LR 1.74; 95% CI 1.07–2.83 and negative LR 0.15; 95% CI 0.03–0.65), but with a specificity of only 0.47.

Statistical pooling was not feasible because of different test protocols with no clear cut-off point for a positive versus negative result, variability in outcome assessment, and heterogeneous patient populations (varying diagnoses and different mix of patients with or without prior spine surgery between studies).

Limitations

As pain appeared to be the only measure of outcome that was consistently incorporated in all studies, our analysis was limited to pain. Although it represents only one aspect of the complex chronic low back pain problem, pain is the primary indication for operative treatment.

Substantial risk of bias in most selected studies precluded firm conclusions. A major drawback was that in all but 3 studies, a proportion of patients who had undergone the index test with a negative test result, were denied fusion and had been excluded from analysis (verification bias). Ideally, fusion should have been performed completely independent of the index test result. Because of the substantial risk of bias, heterogeneous patient populations, and differences in test protocols, a meta-analysis would very likely have produced mis-

Table 15. Systematic review: Characteristics of primary prognostic studies on the clinical outcome of lumbar spinal fusion

A	B	C	D	E	F	G	H	I	J	K	L
Thoracolumbosacral orthosis											
Axelsson et al, 1995	T UH	50	50	44 (20–68)	28 (56%)	Intractable LBP: spondylolisthesis (n=24), facet/disc degeneration (n=11) or postlaminectomy syndrome (n=15), duration not specified	R	Thoracolumbosacral orthosis or canvas corset; Positive in case of $\geq 50\%$ subjective pain relief	24	PLF with autograft	Pain free or significantly improved on a 5-point pain scale and satisfied on a 3-point satisfaction scale ^b
Rask and Dall, 1993	T UH	45	25	44 (20–61)	NS	> 6 months back pain (mean 3.9 years), no neurological motor deficit, herniation or olisthesis >2mm, 38% with prior spine surgery	R	Fiberglass pantaloons cast; Positive in case of pain relief that returned after removal of the cast	Minimum of 6	PLF with autograft	Significant subjective pain relief ^b
Willems et al, 2006	T SH	257	107	40 [8.8]	39 (36%)	Incapacitating LBP, mean 3.7 (0.5–31) years, no neurological motor deficit and routine testing indecisive, 65% with prior spine surgery (N=70)	R	Pantaloons plaster cast; Positive in case of subjective substantial pain relief in the cast	Median of 76 (15–144)	PLF (n=79) or AF (n=28)	$\geq 30\%$ decrease in pain on a VAS (0–100)
Provocative discography											
Colhoun et al., 1988	T OH	195	168	39 (17–70)	86 (51%)	Persistent LBP, no previous back surgery, duration not specified	R	Provocative discography; Positive in case of typical pain reproduction which was not present in adjacent control discs	Mean of 43 (24–120)	AF or PLF ^a	Complete pain relief or significant subjective improvement, resumption of work/normal duties and no intake of analgesics ^b
Esses et al., 1989	T UH	32	22	41 (31–57)	18 (84%)	Longstanding LBP, mean duration of 6.2 (1–20) years, 54% with prior spine surgery	P	Provocative discography, no control discs; Positive in case of typical pain reproduction	“When fusion was noted”	PLF with autograft	Complete or significant relief of pain ^b
Gill and Blumenthal, 1991	T OH	53	53	34 (21–50)	36 (68%)	LBP with a mean disability of 11 (3–120) months, all selected by concordant pain response provoked at discography L5–S1	NS	Discography image of L5–S1, no control disc(s); Positive in case of annular tear extending to the periphery	Mean of 36 (24–54)	AF with allograft (n=48) or autograft (n=5)	Relief of $\geq 75\%$ of initial back pain on VAS, return to work and no use of narcotics
Willems et al., 2007	T SH	197	82	40 [8.5]	26 (32%)	Incapacitating LBP > 1 year (mean duration not specified), no neurological motor deficit and routine testing indecisive, 65% with prior spine surgery	R	Provocative discography of adjacent to intended fusion; Positive in case of no or unfamiliar pain reproduction	Mean of 80 (15–144)	AF or PLF ^a	$\geq 30\%$ decrease in pain on a VAS (0–100)

leading results. Therefore, we refrained from statistical pooling of included studies.

Relevance and implications for clinical practice and future research

Several studies have reported that cognitive behavioural therapy or intensive exercise programs (Brox et al. 2003; Fairbank

et al. 2005; Guzman et al. 2001) have treatment results similar to those of spinal fusion, but with considerably less complications, morbidity and costs (Fritzell et al. 2004; Deyo et al. 2011). The findings of the present review show that the currently used tests do not improve the results of fusion by better patient selection, which makes it hard to propose spinal fusion as a standard treatment for chronic low back pain. Currently

Table 15. Continued

A	B	C	D	E	F	G	H	I	J	K	L
Temporary external transpedicular fixation (TETF)											
Elmans et al., 2005	T SH	330	123	42 [9]	45 (37%)	Incapacitating LBP > 1 year (mean duration of 6 ± 5 years) with inconclusive routine testing, 62% with prior spine surgery	P	TETF with dynamization; Positive if VAS in placebo was ≥ 30 points more than VAS in fixation	Median of 79 (15–44)	AF (n=33) or PLF (n=90)	≥ 30% decrease in pain on a VAS (0–100)
Heini et al., 2004	T UH	63	36	48 (26–67)	22 (62%)	LBP with a mean duration of 5 (1–20) years, 67% with prior spine surgery	P	TETF without dynamization; Positive if pain on VAS and use of analgesics decreased sufficiently (estimated by surgeon)	Mean of 32 (23–60)	PLF, except 3 dynamic fixations and 1 anterior fusion	No or little pain on a VAS and no pain medication ^b
Jeanneret et al., 1994	S RH	101	43	48 (22–74)	25 (58%)	Chronic LBP with or without leg pain, duration not specified, routine testing indeterminate	NS	TETF with dynamization; Positive if pain was reduced at stabilization and returned at destabilization	Mean of 50 (24–92)	Posterior or anterior fusion (numbers not specified)	Almost completely pain free, no pain medication and working ^b
A Source B Setting S = Secondary T = Tertiary OH = Orthopedic hospital RH = Regional hospital SH = Specialized hospital UH = University hospital C Included patients, n D Patients for current analysis, n E Mean age [SD] (range) F Gender: male, n (%) NS = Not specified G Patient characteristics LBP = low back pain						H Study design NS = Not specified P = Prospective R = Retrospective I Index test and criterion for positive result J Follow-up in months (range) K Method of fusion PLF = Postero-lateral fusion AF = Anterior fusion ^a numbers not specified L Criteria for positive reference standard (=fusion outcome) VAS = visual analog scale ^b No validated outcome measure reported					

Table 16. Systematic Review: Risk of Bias: Number of Quality Criteria of the Modified QUADAS Checklist ^a, that were met (+ = yes; – = no; ? = unclear)

Source	Quality criteria												Total
	1	2	3	4	5	6	7	8	9	10	11	12	
Thoracolumbosacral orthosis													
Axelsson et al., 1995	+	–	–	+	+	+	+	–	?	+	?	?	6
Rask and Dall, 1993	+	+	–	–	–	+	–	–	+	+	–	+	6
Willems et al., 2006	–	+	+	+	–	+	–	+	+	+	–	+	8
Provocative discography													
Colhoun et al., 1988	+	?	–	+	+	+	–	–	?	+	?	?	5
Esses et al., 1989	+	–	–	–	–	–	?	–	–	+	–	?	2
Gill and Blumenthal, 1991	–	–	?	+	+	?	+	–	–	?	?	+	4
Willems et al., 2007	–	+	+	+	–	+	?	+	+	–	–	+	7
Temporary external transpedicular fixation													
Elmans et al., 2005	–	–	+	+	–	+	+	+	+	?	?	+	7
Heini et al., 2004	–	–	+	+	–	+	–	–	–	+	+	+	6
Jeanneret et al., 1994	+	?	–	+	–	+	–	–	–	+	–	?	4

^a See Table 1 for complete modified QUADAS checklist.

Table 17. Systematic review: Summary prognostic accuracy of orthosis, provocative discography, and external transpedicular fixation for spinal fusion outcome

Source	Sample size	Sensitivity	Specificity	Positive predictive value	Negative predictive value	Positive LR (95% CI)	Negative LR (95% CI)
Thoracolumbosacral orthosis							
Axelsson et al., 1995	50	0.61	0.35	0.64	0.32	0.94 (0.60 to 1.46)	1.12 (0.52 to 2.41)
Rask and Dall, 1993	25	0.94	0.14	0.74	0.50	1.10 (0.80 to 1.52)	0.39 (0.03 to 5.40)
Willems et al., 2006	107	0.43	0.61	0.44	0.61	1.13 (0.71 to 1.80)	0.92 (0.67 to 1.28)
Provocative discography							
Colhoun et al., 1988	168	0.88	0.48	0.88	0.48	1.71 (1.21 to 2.41)	0.24 (0.13 to 0.43)
Esses et al., 1989	22	0.40	0.43	0.60	0.25	0.70 (0.29 to 1.70)	1.40 (0.54 to 3.62)
Gill and Blumenthal, 1991	53	0.81	0.41	0.74	0.50	1.37 (0.89 to 2.10)	0.47 (0.20 to 1.13)
Willems et al., 2007	82	0.73	0.27	0.45	0.55	0.99 (0.76 to 1.30)	1.01 (0.49 to 2.08)
TETF							
Elmans et al., 2005	123	0.80	0.34	0.46	0.71	1.22 (0.98 to 1.51)	0.58 (0.31 to 1.11)
Heini et al., 2004	36	0.81	0.20	0.45	0.57	1.02 (0.74 to 1.40)	0.94 (0.24 to 3.60)
Jeanneret et al., 1994	43	0.93	0.47	0.77	0.78	1.74 (1.07 to 2.83)	0.15 (0.03 to 0.65)

CI = confidence interval; LR = likelihood ratio; TETF = Temporary external transpedicular fixation.

used tests for patient selection should not be recommended for surgical decision making in standard care.

Future research should focus on identifying subgroups of patients for whom spinal fusion is a predictable and effective treatment. Studies should include both positively and negatively tested patients in a randomised design between fusion and programmed conservative care. Test protocols should be clearly described and clinical outcome should be defined by a consensus cut-off point of improvement in pain and functional status, a so-called minimal clinically important change (Ostelo et al. 2008). Additionally, the role of MRI, as well as

the relation between treatment outcome and patient risk factors for persistent disabling low back pain, should be further elucidated. If the results of fusion can be improved by better patient selection, there could be a role for spinal fusion as the treatment of choice for some patients. A reliable prediction of surgical outcome would enable the instalment of clinical guidelines for surgical decision making. Such guidelines are needed, not only for patient counselling, but also for communication with insurers, policymakers and other health care providers who are involved in the management of chronic low back pain.

General discussion

The aims of the present thesis were to investigate the value, use and appreciation of three prognostic tests (i.e., immobilization by a thoracolumbosacral orthosis (TLSO), provocative discography, and trial immobilization by temporary external transpedicular fixation (TETF)), which are commonly used in daily practice for patient selection. These tests aim to predict the outcome of lumbar spinal fusion either by mimicking a fusion, or by trying to identify the source of pain in the individual patient.

Aim 1: Is there consensus among spine surgeons regarding the use of prognostic tests for lumbar spinal fusion in CLBP patients?

A nationwide survey among spine surgeons in the Netherlands revealed a lack of uniformity in the use and appreciation of prognostic tests for lumbar fusion in CLBP patients. Prognostic factors known from the literature were not consistently incorporated in the surgeons' decision making process either.

Aim 2: Does immobilization by a pantaloone cast truly minimize lumbosacral motion?

The hypothesized working mechanism of a pantaloone cast (minimisation of lumbosacral joint mobility) was studied in patients who were admitted for a temporary external transpedicular fixation test. During dynamic test conditions (walking), wearing a pantaloone cast did not significantly decrease lumbosacral joint motion.

Aim 3: Can immobilization by a thoracolumbosacral orthosis (TLSO) predict the clinical outcome of spinal fusion for CLBP?

A prospective cohort study was performed in order to assess the value of a TLSO in surgical decision-making. It appeared that only in a small group of CLBP patients with no prior spine surgery, a pantaloone cast test with substantial pain relief suggests a favorable outcome of lumbar fusion as compared to conservative treatment. In patients with prior spine surgery the test was of no value.

Aim 4: Does provocative discography of adjacent segments predict the long-term clinical outcome of spinal fusion for CLBP?

In chapter 4a, , was studied. The results of preoperative discography of levels adjacent to a spinal fusion were analyzed retrospectively. It was concluded that in a cohort of patients for whom the decision to perform lumbar fusion was based on a TETF trial, the discographic status of adjacent discs could not predict the clinical outcome of fusion at long-term follow-up.

Aim 5: What is the incidence of postdiscography discitis and is there a need for routine antibiotic prophylaxis?

In search for clinical guidelines, the risk of postdiscography discitis was studied by means of a systematic literature review and a cohort of 200 consecutive patients. Without the use of prophylactic antibiotics, an overall incidence of discitis of 0.25% was found in the literature. To prove that antibiotics actually prevent postdiscography discitis, a randomized trial of 9,000 patients would be needed to reach significance. Given the possible adverse effects of antibiotics, it was concluded that the routine use of prophylactic antibiotics in lumbar discography is not indicated.

Aim 6: Can temporary external transpedicular fixation (TETF) help to predict the clinical outcome of spinal fusion in CLBP patients with an equivocal indication for surgery?

Using strict and objective criteria of pain reduction on a visual analogue scale, the TETF test failed to predict the clinical outcome after fusion in a group of patients for whom there was doubt about the indication for fusion after routine diagnostic testing.

Aim 7: What is the prognostic accuracy of the most commonly used tests in clinical practice to predict the outcome of spinal fusion for CLBP?

In a systematic literature review concerning the prognostic accuracy of magnetic resonance imaging (MRI), TLSO immobilization, TETF, provocative discography and facet joint infiltration, only 10 studies could be selected. It appeared that the overall methodological quality of the included studies was poor and the determined accuracy of the prognostic tests was low.

Based on the studies in this thesis, there are no reliable prognostic tests at present, to aid surgeons in the clinical decision whether or not to perform a lumbar fusion in a patient with CLBP.

Approach of CLBP patients in daily practice

If it appears from the literature that lumbar spinal fusion as a treatment for CLBP is no better or worse than structured conservative care programs, and if current prognostic tests fail to improve overall results by better patient selection, should spinal fusion still be considered as a standard treatment for CLBP?

Basically, the surgical point of view, i.e., the search for an underlying anatomic cause or pain source, should not be our

first approach in the evaluation and treatment of CLBP. The traditional biomedical treatment model does not appear to work well in these patients. This model suggests the recognition of symptoms and signs, identification of underlying injury or disease, treatment with specific therapy and the expectation that the patient will recover. Such an approach indeed works well when dealing with patients with specific LBP in whom causality for the pain can be identified, such as in spondylolisthesis, tumor or trauma. In the case of CLBP, however, causality is more complex and back pain is rather a symptom than a diagnosis in patients with different stages of impairment and disability. These patients appear to be better evaluated by the so-called biopsychosocial model (Waddell 1996), which posits that biologic, social and psychological factors all are implicated in any given state of health or disease. The model aims to identify factors associated with delayed recovery (Vlaeyen and Linton 2000) and calls for assessment of cultural beliefs, anxiety, kinesiophobia, depression, stress from work or family, job dissatisfaction, somatisation, lack of control, and a high perception of disability. All these factors may affect the patient's symptoms and resulting disability. Aiming to treat underlying psychosocial factors and empowering the patient to take responsibility for managing his or her own condition, should be key factors in the initial treatment of CLBP patients. Chronic pain has much less impact on the well-being of a patient if he or she can function normally. Treatment should thus primarily focus on functional capacity with return to work and usual activities, and secondarily on pain relief.

In recent years, strong evidence has emerged that cognitive behavioral therapy should be used in patients with CLBP (Ostelo et al. 2005), especially if psychosocial signs or symptoms are present. In occupational settings, it has been shown that individual graded behavioral-oriented exercise programs can significantly reduce days lost at work (Staal et al. 2004; van den Hout et al. 2003). Especially intensive multidisciplinary biopsychosocial rehabilitation with a functional restoration approach reduces pain significantly more than inpatient or outpatient non-multidisciplinary programs or usual care (Guzman et al. 2001).

Future care for CLBP patients: The case for a multidisciplinary Spine Centre

In current medical practice, patients seeking care can contact clinicians from different medical disciplines with expertise at treating CLBP. These doctors all have their own specific training skills, opinions and therapy proposals, which may differ substantially. There is not even assurance that clinicians within the same medical discipline will offer a similar treatment. This can be both confusing and frustrating for patients. Ideally, clinicians should be aware of all commonly used therapies and able to counsel their patients which treatment may be preferable at that moment. Consequently, they should be

willing to refer patients to colleagues, either within or outside their own discipline, who may be better trained to provide the most appropriate treatment (Haldeman and Dagenais 2008).

Current evidence points to improved triage of low back pain patients in a multidisciplinary setting or Spine Center with a standardized intake, which includes thorough history taking, physical examination, psychosocial evaluation and adequate imaging. Patients expect diagnostic explanations, adequate pain relief from treatment, instructions on how to deal with persisting pain and disability, and perhaps help with sickness certification. Good patient counselling that provides a reliable estimation of the effectiveness of a planned therapy, should be evidence based, i.e., determined by the individual physician's expertise and expectation of treatment success (Campbell and Guy 2007), combined with the best available clinical evidence from systematic research (Sackett et al. 1996). Patient education on the choice of treatment and the expected outcomes seems to be a key factor to success and clinical negotiation appears to be an essential tool when it comes to discussing pain as an active process involving patient expectations and coping strategies (Risdon et al. 2003).

Multidisciplinary therapy should provide a treatment regimen of evidence-based interventions in a stepwise active approach (Guzman et al. 2001), starting with intensive exercise programs or cognitive behavioral interventions that may be combined with pain intervention (medication or neural ablative procedures) or psychotherapy.

What could be the role of spinal fusion in this multidisciplinary approach?

In the 2006 European guidelines it was stated that "only if a prolonged period of acknowledged conservative treatment modalities fails to relieve symptoms, fusion surgery for CLBP may be considered" (Airaksinen et al. 2006). However, we found that fusion results remain unpredictable and current prognostic tests are not able to improve surgical outcome. For patient selection, proper diagnostics that include the assessment of the individual patient's potential to benefit from surgery, are prerequisite in order to conceive realistic predictions of treatment outcome. So, if we believe there is a role for spinal fusion in the treatment of CLBP, more high quality trials to determine prognostic accuracy of tests, as well as the treatment's safety and its effect on pain, function and return to work will be needed (Deyo and Nachemson 2004). RCTs provide a high level of evidence in comparing effectiveness of two treatment modalities, but may be restricted by a limited external validity, which means that the results cannot be generalised for all CLBP patients (Rothwell 2005). Therefore, observational data with indeed a lower level of evidence but a higher feasibility of being closer to daily practice, should be collected as well. The implementation of currently developing well-organized spine registries (Roder et al. 2006;

Strömqvist et al. 2001) should enable the spinal community to acquire such representative data of high validity. As clinicians, we need such data, not only to counsel our patients, but also to inform other health care providers, insurers and policy

makers whether spinal fusion may truly be an effective treatment strategy for a particular subgroup of CLBP patients, and whether it can be embedded in the current, widely supported multidisciplinary active approach of CLBP.

References

- Adams M E, McDevitt C A, Ho A, Muir H. Isolation and characterization of high-buoyant-density proteoglycans from semilunar menisci. *J Bone Joint Surg (Am)* 1986; 68(1): 55-64.
- Airaksinen O, Brox J I, Cedraschi C, et al. Chapter 4. European guidelines for the management of chronic nonspecific low back pain. *Eur Spine J* 2006; 15 (Suppl 2): S192-300.
- Aprill C, Bogduk N. High-intensity zone: a diagnostic sign of painful lumbar disc on magnetic resonance imaging. *Br J Radiol* 1992; 65(773): 361-369.
- Axelsson P, Johnsson R, Strömquist B. Effect of lumbar orthosis on intervertebral mobility. A roentgen stereophotogrammetric analysis. *Spine* 1992; 17(6): 678-681.
- Axelsson P, Johnsson R, Strömquist B. Lumbar orthosis with unilateral hip immobilization. Effect on intervertebral mobility determined by roentgen stereophotogrammetric analysis. *Spine* 1993; 18(7): 876-879.
- Axelsson P, Johnsson R, Strömquist B, Nilsson L T, Akeson M. Orthosis as prognostic instrument in lumbar fusion: no predictive value in 50 cases followed prospectively. *J Spinal Disord* 1995; 8(4): 284-288.
- Axelsson P, Johnsson R, Strömquist B, Andréasson H. Temporary external pedicular fixation versus definitive bony fusion: a prospective comparative study on pain relief and function. *Eur Spine J* 2003; 12(1): 41-47.
- Battié M C, Videman T, Levälähti E, Gill K, Kaprio J. Genetic and environmental effects on disc degeneration by phenotype and spinal level: a multivariate twin study. *Spine* 2008; 33(25): 2801-2808.
- Bednar D A. Failure of external spinal skeletal fixation to improve predictability of lumbar arthrodesis. *J Bone Joint Surg (Am)* 2001; 83-A(11): 1656-1659.
- Bednar D A, Raducan V. External spinal skeletal fixation in the management of back pain. *Clin Orthop* 1996; (322): 131-139.
- Bigos S J, Battié M C, Fisher L D. Methodology for evaluating predictive factors for the report of back injury. *Spine* 1991; 16(6): 669-670.
- Boden S D, McCowin P R, Davis D O, Dina T S, Mark A S, Wiesel S. Abnormal magnetic-resonance scans of the cervical spine in asymptomatic subjects. A prospective investigation. *J Bone Joint Surg (Am)* 1990; 72(8): 1178-1184.
- Boscardin J B, Ringus J C, Feingold D J, Ruda S C. Human intradiscal levels with cefazolin. *Spine* 1992; 17(6 Suppl): S145-148.
- Brox J I, Sørensen R, Friis A, et al. (2003). Randomized clinical trial of lumbar instrumented fusion and cognitive intervention and exercises in patients with chronic low back pain and disc degeneration. *Spine* 2003; 28(17): 1913-1921.
- Buchalter, D., N. Kahanovitz, et al. (1988). Three-dimensional spinal motion measurements. Part 2: A noninvasive assessment of lumbar brace immobilization of the spine. *J Spinal Disord* 1(4): 284-286.
- Burke J G, Watson R W, McCormack D, Dowling F E, Walsh M G, Fitzpatrick J M. Intervertebral discs which cause low back pain secrete high levels of proinflammatory mediators. *J Bone Joint Surg (Br)* 2002; 84(2): 196-201.
- Buttermann G R, Heithoff K B, Ogilvie J W, Transfeldt E E, Cohen M. Vertebral body MRI related to lumbar fusion results. *Eur Spine J* 1997; 6(2): 115-120.
- Campbell C, Guy A. Why can't they do anything for a simple back problem? A qualitative examination of expectations for low back pain treatment and outcome. *J Health Psychol* 2007; 12 : 641-652.
- Carragee E J. Clinical practice. Persistent low back pain. *N Engl J Med* 2005; 352(18): 1891-1898.
- Carragee E J, Chen Y, Tanner C M, Truong T, Lau E, Brito J L. Provocative discography in patients after limited lumbar discectomy: A controlled, randomized study of pain response in symptomatic and asymptomatic subjects. *Spine* 2000; 25(23): 3065-3071.
- Carragee E J, Tanner C M, Khurana S, et al. The rates of false-positive lumbar discography in select patients without low back symptoms. *Spine* 2000; 25(11): 1373-1380; discussion 1381.
- Chataigner H, Onimus M, Polette A. [Surgery for degenerative lumbar disc disease. Should the black disc be grafted?]. *Rev Chir Orthop Reparatrice Appar Mot* 1998; 84(7): 583-589.
- Cherkin D C, Deyo R A, Loeser J D, Bush T, Waddell G. An international comparison of back surgery rates. *Spine* 1994; 19(11): 1201-1206.
- Cheung K M, Chan D, Karppinen J, et al. Association of the Taq I allele in vitamin D receptor with degenerative disc disease and disc bulge in a Chinese population. *Spine* 2006; 31(10): 1143-1148.
- Cholewicki J, Alvi K, Silfies S P, Bartolomei J. Comparison of motion restriction and trunk stiffness provided by three thoracolumbosacral orthoses (TLSOs). *J Spinal Disord Tech* 2003; 16(5): 461-468.
- Cholewicki J, Juluru K, Radebold A, Panjabi M M, McGill S M. Lumbar spine stability can be augmented with an abdominal belt and/or increased intra-abdominal pressure. *Eur Spine J* 1999; 8(5): 388-395.
- Chow D H, Luk K D, Evans J H, Leong J C. Effects of short anterior lumbar interbody fusion on biomechanics of neighboring unfused segments. *Spine* 1996; 21(5): 549-555.
- Christensen F B, Karlsmose B, Hansen E S, Bünger C E. Radiological and functional outcome after anterior lumbar interbody spinal fusion. *Eur Spine J* 1996; 5(5): 293-298.
- Colhoun E, McCall I W, Williams L, Cassar Pullicino V N. Provocation discography as a guide to planning operations on the spine. *J Bone Joint Surg (Br)* 1988; 70(2): 267-271.
- Coppes M H, Marani E, Thomeer R T, Groen G J. Innervation of "painful" lumbar discs. *Spine* 1997; 22(20): 2342-2349; discussion 2349-2350.
- Dawson E G, Lotysch M 3rd, Urist M R. (1981). Intertransverse process lumbar arthrodesis with autogenous bone graft. *Clin Orthop* 1981; (154): 90-96.
- DeBerard M S, Masters K S, Colledge A L, Schleusener R L, Schlegel J D. Outcomes of posterolateral lumbar fusion in Utah patients receiving workers' compensation: a retrospective cohort study. *Spine* 2001; 26(7): 738-746; discussion 747.
- Derby R, Howard M W, Grant J M, Lettice J J, Van Peteghem P K, Ryan DP. The ability of pressure-controlled discography to predict surgical and nonsurgical outcomes. *Spine* 1999; 24(4): 364-371; discussion 371-362.
- Devillé W L, Buntinx F, Bouter L M, et al. Conducting systematic reviews of diagnostic studies: didactic guidelines. *BMC Med Res Methodol* 2002; 2: 9.
- Deyo R A, Bass J E. Lifestyle and low-back pain. The influence of smoking and obesity. *Spine* 1989; 14(5): 501-506.
- Deyo R A, Rainville J, Kent D L. What can the history and physical examination tell us about low back pain? *JAMA* 1992; 268(6): 760-765.
- Deyo RA, Nachemson A, Mirza S K. Spinal-fusion surgery – the case for restraint. *N Engl J Med* 2004; 350 : 722-726.
- Deyo R A, Gray D T, Kreuter W, Mirza S, Martin B I. United States trends in lumbar fusion surgery for degenerative conditions. *Spine* 2005; 30(12): 1441-1445; discussion 1446-1447.
- Eismont F J, Wiesel S W, Brighton C T, Rothman R H. Antibiotic penetration into rabbit nucleus pulposus. *Spine* 1987; 12(3): 254-256.
- Elmans L, Willems P C, Anderson P G, van Limbeek J, de Kleuver M, van der Schaaf D. Temporary external transpedicular fixation of the lumbosacral spine: a prospective, longitudinal study in 330 patients. *Spine* 2005; 30(24): 2813-2816.

- Esposito P, Pinheiro-Franco J L, Froelich S, Maitrot D. Predictive value of MRI vertebral end-plate signal changes (Modic) on outcome of surgically treated degenerative disc disease. Results of a cohort study including 60 patients. *Neurochirurgie* 2006; 52(4): 315-322.
- Esses S I, Moro J K. The value of facet joint blocks in patient selection for lumbar fusion. *Spine* 1993; 18(2): 185-190.
- Esses S I, Botsford D J, Kostuik J P. The role of external spinal skeletal fixation in the assessment of low-back disorders. *Spine* 1989; 14(6): 594-601.
- Fairbank J, Frost H, Wilson-MacDonald J, Yu L M, Barker K, Collins R. Randomised controlled trial to compare surgical stabilisation of the lumbar spine with an intensive rehabilitation programme for patients with chronic low back pain: the MRC spine stabilisation trial. *BMJ* 2005; 330(7502): 1233.
- Faraj A A, Akasha K, Mulholland R C. Temporary external fixation for low back pain: is it worth doing? *Eur Spine J* 1997; 6(3): 187-190.
- Fidler M W, Plasmans C M (1983). The effect of four types of support on the segmental mobility of the lumbosacral spine. *J Bone Joint Surg (Am)* 1983; 65(7): 943-947.
- Franson R C, Saal J S, Saal J A. Human disc phospholipase A2 is inflammatory. *Spine* 1992; 17(6 Suppl): S129-132.
- Fraser R D. The North American Spine Society (NASS) on lumbar discography. *Spine* 1996; 21(10): 1274-1276.
- Fraser R D, Osti O L, Vernon-Roberts B. Discitis following chemonucleolysis. An experimental study. *Spine* 1986; 11(7): 679-687.
- Fraser R D, Osti O L, Vernon-Roberts B. Discitis after discography. *J Bone Joint Surg (Br)* 1987; 69(1): 26-35.
- Fritzell P, Hägg O, Wessberg P, Nordwall A. 2001 Volvo Award Winner in Clinical Studies: Lumbar fusion versus nonsurgical treatment for chronic low back pain: a multicenter randomized controlled trial from the Swedish Lumbar Spine Study Group. *Spine* 2001; 26(23): 2521-2532; discussion 2532-2524.
- Fritzell P, Hägg O, Jonsson D, Nordwall A. Cost-effectiveness of lumbar fusion and nonsurgical treatment for chronic low back pain in the Swedish Lumbar Spine Study: a multicenter, randomized, controlled trial from the Swedish Lumbar Spine Study Group. *Spine* 2004; 29(4): 421-434; discussion Z423.
- Gibson M J, Karpinski M R, Slack R C, Cowlishaw W A, Webb J K. The penetration of antibiotics into the normal intervertebral disc. *J Bone Joint Surg (Br)* 1987; 69(5): 784-786.
- Gibson J N, Grant I C, Waddell G. The Cochrane review of surgery for lumbar disc prolapse and degenerative lumbar spondylosis. *Spine* 1999; 24(17): 1820-1832.
- Gill K, Blumenthal S L. Functional results after anterior lumbar fusion at L5-S1 in patients with normal and abnormal MRI scans. *Spine* 1992; 17(8): 940-942.
- Gillet P. The fate of the adjacent motion segments after lumbar fusion. *J Spinal Disord Tech* 2003; 16(4): 338-345.
- Guzmán J, Esmail R, Karjalainen K, Malmivaara A, Irvin E, Bombardier C. Multidisciplinary rehabilitation for chronic low back pain: systematic review. *BMJ* 2001; 322(7301): 1511-1516.
- Hadra B E. Wiring of the spinous process in injury and Pott's disease. *Trans Am Orthop Assoc* 1891; 4: 206-210.
- Haefeli M, Elfering A, Aebi M, et al. What comprises a good outcome in spinal surgery? A preliminary survey among spine surgeons of the SSE and European spine patients. *Eur Spine J* 2008; 17(1): 104-116.
- Hägg O, Fritzell P, Nordwall A. The clinical importance of changes in outcome scores after treatment for chronic low back pain. *Eur Spine J* 2003; 12(1): 12-20.
- Haldeman S, Dagenais S. A supermarket approach to the evidence-informed management of chronic low back pain. *Spine J* 2008; 8(1): 1-7.
- Hambly M F, Wiltse L L, Raghavan N, Schneiderman G, Koenig C. The transition zone above a lumbosacral fusion. *Spine* 1998; 23(16): 1785-1792.
- Heini P F, Gahrlich U, Orlor R. The external fixator: a tool for evaluation of complex low back pain problems. *J Spinal Disord Tech* 2004; 17(1): 8-14.
- Herkowitz H N, Abraham D J, Albert T J. Management of degenerative disc disease above an L5-S1 segment requiring arthrodesis. *Spine* 1999; 24(12): 1268-1270.
- Hermann J. PRIMAS precision motion analysis system: A short description. In: Boeninck U, Nader M, eds. *Gangbildanalyse, stand der Messtechnik und bedeutung für die Orthopädie-technik*. D. Duderstadt: Mecke, und Verlag 1990; pp: 60-65.
- Hoelscher G L, Gruber H E, Coldham G, Grigsby J H, Hanley E N Jr. Effects of very high antibiotic concentrations on human intervertebral disc cell proliferation, viability, and metabolism in vitro. *Spine* 2000; 25(15): 1871-1877.
- Jackson R K, Boston D A, Edge A J. Lateral mass fusion. A prospective study of a consecutive series with long-term follow-up. *Spine* 1985; 10(9): 828-832.
- Jackson R P, Cain J E Jr, Jacobs R R, Cooper B R, McManus G E. The neuroradiographic diagnosis of lumbar herniated nucleus pulposus: II. A comparison of computed tomography (CT), myelography, CT-myelography, and magnetic resonance imaging. *Spine* 1989; 14(12): 1362-1367.
- Jarvik J J, Hollingworth W, Heagerty P, Haynor D R, Deyo R A. The Longitudinal Assessment of Imaging and Disability of the Back (LAIDBack) Study: baseline data. *Spine* 2001; 26(10): 1158-1166.
- Jeanneret B, Jovanovic M, Magerl F. Percutaneous diagnostic stabilization for low back pain. Correlation with results after fusion operations. *Clin Orthop* 1994; (304): 130-138.
- Jellema P, van der Windt D A, Bruinvels D J, et al. Value of symptoms and additional diagnostic tests for colorectal cancer in primary care: systematic review and meta-analysis. *BMJ* 2010; 340: c1269.
- Jensen M C, Brant-Zawadzki M N, Obuchowski N, Modic M T, Malkasian D, Ross J S. Magnetic resonance imaging of the lumbar spine in people without back pain. *N Engl J Med* 1994; 331(2): 69-73.
- Kang J D, Georgescu H I, McIntyre-Larkin L, Stefanovic-Racic M, Donaldson W F 3rd, Evans C H. Herniated lumbar intervertebral discs spontaneously produce matrix metalloproteinases, nitric oxide, interleukin-6, and prostaglandin E2. *Spine* 1996; 21(3): 271-277.
- Kang J D, Stefanovic-Racic M, McIntyre LA, Georgescu H I, Evans C H. Toward a biochemical understanding of human intervertebral disc degeneration and herniation. Contributions of nitric oxide, interleukins, prostaglandin E2, and matrix metalloproteinases. *Spine* 1997; 22(10): 1065-1073.
- Katz J N. (1995). Lumbar spinal fusion. Surgical rates, costs, and complications. *Spine* 1995; 20(24 Suppl): 78S-83S.
- Katz J N. Lumbar disc disorders and low-back pain: socioeconomic factors and consequences. *J Bone Joint Surg (Am)* 2006; 88(Suppl 2): 21-24.
- Krismer M. Fusion of the lumbar spine. A consideration of the indications. *J Bone Joint Surg (Br)* 2002; 84(6): 783-794.
- Kumar M N, Baklanov A, Chopin D. Correlation between sagittal plane changes and adjacent segment degeneration following lumbar spine fusion. *Eur Spine J* 2001a; 10(4): 314-319.
- Kumar M N, Jacquot F, Hall H. Long-term follow-up of functional outcomes and radiographic changes at adjacent levels following lumbar spine fusion for degenerative disc disease. *Eur Spine J* 2001b; 10(4): 309-313.
- Langmayr J J, Obwegeser A A, Schwarz A B, Laimer I, Ulmer H, Ortler M. Intrathecal steroids to reduce pain after lumbar disc surgery: a double-blind, placebo-controlled prospective study. *Pain* 1995; 62(3): 357-361.
- Lantz S A, Schultz A B. Lumbar spine orthosis wearing. II. Effect on trunk muscle myoelectric activity. *Spine* 1986; 11(8): 838-842.
- Lee C K, Langrana N A. Lumbosacral spinal fusion. A biomechanical study. *Spine* 1984; 9(6): 574-581.
- Lehmann T R, Spratt K F, Tozzi JE, et al. Long-term follow-up of lower lumbar fusion patients. *Spine* 1987; 12(2): 97-104.
- Liberati A, Altman D G, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med* 2009; 6(7): e1000100.
- Lovely T J, Rastogi P. The value of provocative facet blocking as a predictor of success in lumbar spine fusion. *J Spinal Disord* 1997; 10(6): 512-517.

- Lumsden R M 2nd, Morris J M. An in vivo study of axial rotation and immobilization at the lumbosacral joint. *J Bone Joint Surg (Am)* 1968; 50(8): 1591-1602.
- Lund T, Oxland T R, Nydegger T, Schlenzka D, Laine T, Heini P. Is there a connection between the clinical response after an external fixation test or a subsequent lumbar fusion and the pre-test intervertebral kinematics? *Spine* 2002; 27(23): 2726-2733.
- Madan S, Gundanna M, Harley J M, Boeree N R, Sampson M. Does provocative discography screening of discogenic back pain improve surgical outcome? *J Spinal Disord Tech* 2002; 15(3): 245-251.
- Magerl F P. Stabilization of the lower thoracic and lumbar spine with external skeletal fixation. *Clin Orthop* 1984; (189): 125-141.
- McCulloch J A, Waddell G. Lateral lumbar discography. *Br J Radiol* 1978; 51(607): 498-502.
- McIntosh G, Frank J, Hogg-Johnson S, Bombardier C, Hall H. Prognostic factors for time receiving workers' compensation benefits in a cohort of patients with low back pain. *Spine* 2000; 25(2): 147-157.
- Modic M T, Masaryk T J, Ross J S, Carter J R. Imaging of degenerative disk disease. *Radiology* 1988; 168(1): 177-186.
- Nachemson A. Lumbar discography—where are we today? *Spine* 1989; 14(6): 555-557.
- Nachemson A, Morris J M. In vivo measurements of intradiscal pressure. Discometry, a method for the determination of pressure in the lower lumbar discs. *J Bone Joint Surg (Am)* 1964; 46: 1077-1092.
- Nelson M A, Allen P, Clamp S E, de Dombal F T. Reliability and reproducibility of clinical findings in low-back pain. *Spine* 1979; 4(2): 97-101.
- Ohtori S, Kinoshita T, Yamashita M, et al. Results of surgery for discogenic low back pain: a randomized study using discography versus discoblock for diagnosis. *Spine* 2009; 34(13): 1345-1348.
- Olerud S, Sjöström L, Karlström G, Hamberg M. Spontaneous effect of increased stability of the lower lumbar spine in cases of severe chronic back pain. The answer of an external transpeduncular fixation test. *Clin Orthop* 1986; (203): 67-74.
- Ostelo R W, van Tulder M W, Vlaeyen J W, Linton S J, Morley S J, Assendelft W J. Behavioural treatment for chronic low-back pain. *Cochrane Database Syst Rev* 2005; (1): CD002014.
- Ostelo R W, Deyo R A, Stratford P, et al. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. *Spine* 2008; 33(1): 90-94.
- Osti O L, Fraser R D, Vernon-Roberts B. Discitis after discography. The role of prophylactic antibiotics. *J Bone Joint Surg (Br)* 1990; 72(2): 271-274.
- Patel N, Bagan B, Vadera S, et al. Obesity and spine surgery: relation to perioperative complications. *J Neurosurg Spine* 2007; 6(4): 291-297.
- Patrick B S. Lumbar discography: a five year study. *Surg Neurol* 1973; 1(5): 267-273.
- Peng B, Wu W, Hou S, Li P, Zhang C, Yang Y. The pathogenesis of discogenic low back pain. *J Bone Joint Surg (Br)* 2005; 87(1): 62-67.
- Peng B, Chen J, Kuang Z, Li D, Pang X, Zhang X. Diagnosis and surgical treatment of back pain originating from endplate. *Eur Spine J* 2009; 18(7): 1035-1040.
- Pincus T, Burton A K, Vogel S, Field A P. A systematic review of psychological factors as predictors of chronicity/disability in prospective cohorts of low back pain. *Spine* 2002; 27(5): E109-120.
- Piotrowski W P, Krombholz M A, Mühl B. Spondylodiscitis after lumbar disk surgery. *Neurosurg Rev* 1994; 17(3): 189-193.
- Price D D, McGrath P A, Rafii A, Buckingham B. The validation of visual analogue scales as ratio scale measures for chronic and experimental pain. *Pain* 1983; 17(1): 45-56.
- Rask B, Dall B E. Use of the pantaloone cast for the selection of fusion candidates in the treatment of chronic low back pain. *Clin Orthop* 1993; (288): 148-157.
- Resnick D K, Choudhri T F, Dailey A T, et al. Guidelines for the performance of fusion procedures for degenerative disease of the lumbar spine. Part 2: assessment of functional outcome. *J Neurosurg Spine* 2005; 2(6): 639-646.
- Rhoten R L, Murphy M A, Kalfas I H, Hahn J F, Washington J A. Antibiotic penetration into cervical discs. *Neurosurgery* 1995; 37(3): 418-421.
- Rimoldi R L, Haye W. The use of antibiotics for wound prophylaxis in spinal surgery. *Orthop Clin North Am* 1996; 27(1): 47-52.
- Risdon A, Eccleston C, Crombez G, McCracken L. How can we learn to live with pain? A Q-methodological analysis of the diverse understandings of acceptance of chronic pain. *Soc Sci Med* 2003; 56: 375-386.
- Rivero-Arias O, Campbell H, Gray A, Fairbank J, Frost H, Wilson-MacDonald J. Surgical stabilisation of the spine compared with a programme of intensive rehabilitation for the management of patients with chronic low back pain: cost utility analysis based on a randomised controlled trial. *BMJ* 2005; 330(7502): 1239.
- Röder C, Müller U, Aebi M. The rationale for a spine registry. *Eur Spine J* 2006; 15 (Suppl 1): S52-S56.
- Rohde V, Meyer B, Schaller C, Hassler W E. Spondylodiscitis after lumbar discectomy. Incidence and a proposal for prophylaxis. *Spine* 1998; 23(5): 615-620.
- Rothwell P M. External validity of randomised controlled trials: To whom do the results of this trial apply? *Lancet* 2005; 365(9453): 82-93.
- Sackett D L, Rosenberg W M, Gray J A, Haynes R B, Richardson W S. Evidence based medicine: what it is and what it isn't. *BMJ* 1996; 312(7023): 71-72.
- Sato H, Kikuchi S. The natural history of radiographic instability of the lumbar spine. *Spine* 1993; 18(14): 2075-2079.
- Schulitz K P, Assheuer J. Discitis after procedures on the intervertebral disc. *Spine* 1994; 19(10): 1172-1177.
- Slobbe L C J, Smit J M, Groen J, Meerding W J, Polder J J. Kosten van ziekten in Nederland 2003. RIVM rapport 270751010. Bilthoven, RIVM 2006.
- Soini J R, Seitsalo S K. The external fixation test of the lumbar spine. 30 complications in 25 of 100 consecutive patients. *Acta Orthop Scand* 1993; 64(2): 147-149.
- Staal J B, Hlobil H, Twisk J W, Smid T, Köke A J, van Mechelen W. Graded activity for low back pain in occupational health care: a randomized, controlled trial. *Ann Intern Med* 2004; 140(2): 77-84.
- Strömquist B, Jönsson B, Fritzell P, Hägg O, Larsson B E, Lind B. The Swedish National Registry for lumbar spine surgery: Swedish Society for Spinal Surgery. *Acta Orthop Scand* 2001; 72(2): 99-106.
- Tai C C, Want S, Quraishi N A, Batten J, Kalra M, Hughes S P. Antibiotic prophylaxis in surgery of the intervertebral disc. A comparison between gentamicin and cefuroxime. *J Bone Joint Surg (Br)* 2002; 84(7): 1036-1039.
- Throckmorton T W, Hilibrand A S, Mencia G A, Hodge A, Spengler D M. The impact of adjacent level disc degeneration on health status outcomes following lumbar fusion. *Spine* 2003; 28(22): 2546-2550.
- Tuong N H, Dansereau J, Maurais G, Herrera R. Three-dimensional evaluation of lumbar orthosis effects on spinal behavior. *J Rehabil Res Dev* 1998; 35(1): 34-42.
- Turner JA, Ersek M, Herron L, et al. Patient outcomes after lumbar spinal fusions. *JAMA* 1992; 268(7): 907-911.
- Urban J P, Holm S, Maroudas A, Nachemson A. Nutrition of the intervertebral disk. An in vivo study of solute transport. *Clin Orthop* 1977; (129): 101-114.
- van den Hout J H, Vlaeyen J W, Heuts P H, Zijlema J H, Wijnen J A. Secondary prevention of work-related disability in nonspecific low back pain: does problem-solving therapy help? A randomized clinical trial. *Clin J Pain* 2003; 19(2): 87-96.
- van der Schaaf D B, van Limbeek J, Pavlov P W. Temporary external transpedicular fixation of the lumbosacral spine. *Spine* 1999; 24(5): 481-484; discussion 484-485.
- van der Windt D A, Jellema P, Mulder C J, Kneepkens C M, van der Horst H E. Diagnostic testing for celiac disease among patients with abdominal symptoms: a systematic review. *JAMA* 2010; 303(17): 1738-1746.
- van Tulder M W, Assendelft W J, Koes B W, Bouter L M. Spinal radiographic findings and nonspecific low back pain. A systematic review of observational studies. *Spine* 1997; 22(4): 427-434.

- van Tulder M W, Koes B W, Bouter L M. A cost-of-illness study of back pain in The Netherlands. *Pain* 1995; 62(2): 233-240.
- Vlaeyen J W, Linton S J. Fear-avoidance and its consequences in chronic musculoskeletal pain: a state of the art. *Pain* 2000; 85(3): 317-332.
- Waddell G. Low back pain: a twentieth century health care enigma. *Spine* 1996; 21(24): 2820-2825.
- Walker B F. The prevalence of low back pain: a systematic review of the literature from 1966 to 1998. *J Spinal Disord* 2000; 13(3): 205-217.
- Weinstein J N, Lurie J D, Olson P R, Bronner K K, Fisher E S. United States' trends and regional variations in lumbar spine surgery: 1992–2003. *Spine* 2006; 31(23): 2707-2714.
- Weishaupt D, Zanetti M, Hodler J, et al. Painful lumbar disk derangement: Relevance of endplate abnormalities at MR imaging. *Radiology* 2001; 218(2): 420-427.
- Wetzel F T, LaRocca S H, Lowery G L, Aprill C N. The treatment of lumbar spinal pain syndromes diagnosed by discography. *Lumbar arthrodesis*. *Spine* 1994; 19(7): 792-800.
- Whiting P, Rutjes A W, Reitsma J B, Bossuyt P M, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003; 3: 25.
- Willems P C, Nienhuis B, Sietsma M, van der Schaaf D B, Pavlov P W. The effect of a plaster cast on lumbosacral joint motion. An in vivo assessment with precision motion analysis system. *Spine* 1997; 22(11): 1229-1234.
- Willems P C, Jacobs W, Duinkerke E S, De Kleuver M. Lumbar discography: should we use prophylactic antibiotics? A study of 435 consecutive discograms and a systematic review of the literature. *J Spinal Disord Tech* 2004; 17(3): 243-247.
- Willems P C, Elmans L, Anderson P G, Jacobs W C, van der Schaaf D B, de Kleuver M. The value of a pantaloon cast test in surgical decision making for chronic low back pain patients: a systematic review of the literature supplemented with a prospective cohort study. *Eur Spine J* 2006; 15(10): 1487-1494.
- Willems P C, Elmans L, Anderson P G, van der Schaaf D B, de Kleuver M. Provocative discography and lumbar fusion: is preoperative assessment of adjacent discs useful? *Spine* 2007; 32(10): 1094-1099; discussion 1100.
- Wilson-MacDonald J, Fairbank J, Frost H, Yu L M, Barker K, Collins R, Campbell H. The MRC spine stabilization trial: surgical methods, outcomes, costs, and complications of surgical stabilization. *Spine* 2008; 33(21): 2334-2340.
- Zamora J, Abaira V, Muriel A, Khan K, Coomarasamy A. Meta-DiSc: a software for meta-analysis of test accuracy data. *BMC Med Res Methodol* 2006; 6: 31.
- Zanoli G, Strömqvist B, Jönsson B. Visual analog scales for interpretation of back and leg pain intensity in patients operated for degenerative lumbar spine disorders. *Spine* 2001; 26(21): 2375-2380.