Long-term clinical and radiological results in patients with chronic low back pain and degenerative disc randomised to total disc replacement or multidisciplinary rehabilitation

PhD thesis by Håvard Furunes

University of Oslo
Faculty of medicine
Oslo, Norway
2018

Series of dissertations submitted to the
Faculty of Medicine, University of Oslo

ISBN 978-82-8377-351-4

Cover: Hanne Baadsgaard Utigard.
Print production: Reprosentralen, University of Oslo.

Acknowledgements

The present work was carried out in the period 2012-2018 at FORMI (Research- and comunication unit for musculoskeletal health), Oslo University Hospital Ullevål and at Innlandet Hospital Gjøvik.

The Norwegian Total Disc Replacement study was initiated by **Oliver Grundnes** and his colleagues in 2003-2004 as a multicentre study including Oslo University Hospital Ullevål, Haukeland University Hospital, St Olav's University Hospital, Stavanger University Hospital and University Hospital of Northern Norway. In 2012, we invited the study participants to an eight-year follow-up, on which the present thesis is based.

I would like to express my sincere gratitude to everyone who has contributed to this thesis.

First, I would like to thank **Kjersti Storheim** for her enthusiasm and support as my main supervisor. She has introduced me to the scientific network at FORMI, and guided me through the entire process of planning, collection and analyses of data and publication of the results. The project could not have been completed without her insight and overview in the field of musculoskeletal health.

I will also thank **Christian Hellum** for sharing his detailed knowledge of the Norwegian TDR Study, his contagious enthusiasm and curiosity, his willingness and ability to share his experience in surgical treatment of spinal disorders, and for our interesting discussions and friendship.

Jens Ivar Brox for sharing his profound understanding of randomised studies and the pitfalls in interpreting their results. He is an expert in challanging established thoughts and our discussions have definitely brought me to a deeper insight.

Ansgar Espeland for his inexhaustible work to optimise the radiological protocol, classify the radiological findings and interprete the results. He has an impressing ability to catch important details, he is supportive and solution oriented and he is a role model in precise communication.

Tore Solberg, Jan Sture Skouen, Lars Gunnar Johnsen, Ivar Rossvoll, Linda Berg, Eric Franssen and **Helga Maria Brøgger** for their help in the collection, analyses, interpretation and publication of the findings in the studies.

Leiv Sandvik and **Milada Cvancarova Småstuen** for their invaluable help as they shared their knowledge in medical biostatistics.

Marianne Bakke Johnsen and **Maren Hjelle Guddal** for collecting and punching data, and **Eira Kathleen Ebbs** for language help.

My colleagues at **Gjøvik Hospital**; **Espen Mørk, Øyvind Sletten, Peder Bogsti, Steinar Sulheim, Karim Al-Dekany** and **Ivar Dammen**, and in particular my leaders **Inger Opheim, Bjørn Robstad** and **Dirk Clemens** for offering me the time and space needed to complete this PhD project. A special thank to **Arne Skoglund** for introducing me to the field of spinal surgery.

The **study participants** for their positive attitude, their time and their willingness to share their unique experience with our attempts to improve the treatment for low back pain.

The study was supported by indispensable grants from the Norway's **South Eastern Health Authority**, and from the **EXTRA funds** from the **Norwegian Foundation for Health and Rehabilitation** through the **Norwegian Low Back Pain Association.**

Finally, I will thank my beloved wife **Marte Mørk Furunes** for her constant support, patience and ability to organise the lives of three daughters and a husband. I will also thank **Mari**, **Oda** and **Eline** for bringing endless amusement, distraction and joy into my life.

Gjøvik

June, 2018

Table of contents

List of included papers

Paper I

Furunes H, Storheim K, Brox JI, Johnsen LG, Skouen JS, Franssen E, Solberg TK, Sandvik L, Hellum C. Total disc replacement versus multidisciplinary rehabilitation in patients with chronic low back pain and degenerative discs: 8-year follow-up of a randomized controlled multicenter trial. *Spine J 2017; 17: 1480-1488.*

Paper II

Furunes H, Hellum C, Brox JI, Rossvoll I, Espeland A, Berg L, Brøgger HM, Småstuen MC, Storheim K. Lumbar total disc replacement: predictors for long-term outcome. *Eur Spine J 2018; 27: 709-718.*

Paper III

Furunes H, Hellum C, Espeland A, Brox JI, Småstuen MC, Berg L, Storheim K. Adjacent disc degeneration after lumbar total disc replacement or non-operative treatment: A randomized study with eight-year follow-up. *Spine 2018;* May 22. doi: 10.1097/BRS.0000000000002712

Abbreviations

ADD = Adjacent disc degeneration

BMI = Body mass index

CI = Confidence interval

DDD = Degenerative disc disease

EQ-5D = EuroQol 5D

HIZ = High intensity zone

HUNT = Helseundersøkelsen i Nord-Trøndelag

IDD = Intervertebral disc degeneration

IVD = Intervertebral disc

LBP = Low back pain

MC = Modic changes

MCID = minimal clinically important difference

MDR = Multidisciplinary rehabilitation

MRI = Magnetic resonance imaging

NNT = Number needed to treat

ODI = Oswestry Disability Index

OR = Odds ratio

PROMs = Patient reported outcome measures

ROM = Range of motion

TDR = Total disc replacement

VAS = Visual Analogue Scale

YLD = Years lived with disability

Summary

Background

Lumbar total disc replacement (TDR) is a treatment option for selected patients with chronic low back pain that is non-responsive to non-operative treatment. TDR was introduced as a motion-preserving alternative to spinal fusion, which has been reported to increase the risk of adjacent disc degeneration (ADD). However, ADD may develop regardless of surgery, and previous studies have called the clinical importance of ADD into question. The long-term results of disc replacement compared to multidisciplinary rehabilitation have not been reported previously. We aimed to assess the long-term relative efficacy of lumbar TDR compared to multidisciplinary rehabilitation, to identify patient characteristics associated with a favourable long-term result and to assess the long-term ADD development following TDR compared to non-operative treatment.

Material and methods

This is an eight-year follow-up of a multicentre randomised controlled trial performed at five university hospitals in Norway. The sample consists of 173 patients aged 25-55 years with chronic low back pain and localized degenerative changes in the lumbar intervertebral discs. Self-reported outcome measures were collected eight years after treatment. The primary outcome was self-reported physical function (Oswestry Disability Index, ODI) at eight-year follow-up in the intention-to-treat (ITT) population. Secondary outcomes included self-reported low back pain (visual analogue scale, VAS), quality of life (EuroQol, EQ-5D), emotional distress (Hopkins Symptom Check List, HSCL-25), occupational status, patient satisfaction with outcome and care, drug use, complications and additional back surgery. We used $\chi 2$ test or Fisher's exact test to analyse categorical variables and an independent two-sided t test or analysis of variance to analyse continuous variables (Paper I). In a cohort of 82 patients treated with TDR, we analysed the predictive value of pre-treatment socio-demographic, clinical, psychological and radiological patient characteristics for (1) achieving a clinically important improvement ($\geq$ 15 ODI points) from baseline to eight-year follow-up and for (2) being employed at eight-year follow-up. The associations between potential predictors and outcomes were modelled using logistic regression. We also organised a prediction matrix for presenting the probabilities of being employed at eight-year follow-up (Paper II). The development of ADD was evaluated in 126 patients with magnetic resonance imaging (MRI) of the lumbar spine before treatment and at eight-year follow-up. ADD was categorized as increased or not increased based on an evaluation of Modic changes, disc height reduction, disc contour, herniation size, nucleus pulposus signal and posterior high intensity zones. We used a $\chi 2$ test or a Fisher's exact test to compare crude proportions, and multiple linear regressions to analyse the association between increased ADD (yes/no) and change in ODI from pre-treatment to eight-year follow-up (Paper III).

Results

605 patients were screened for eligibility, of whom 173 were randomly assigned treatment. 77 patients (90%) randomised to surgery and 74 patients (85%) randomised to rehabilitation responded at eight-year follow-up. Mean improvement on the ODI was 20.0 points (95% CI

16.4-23.6, p≤0.0001) in the surgery group and 14.4 points (95% CI 10.7-18.1, p≤0.0001) in the rehabilitation group. Mean difference between the groups at eight-year follow-up was 6.1 points (95% CI 1.2-11.0, p=0.02). Mean difference in favour of surgery on secondary outcomes were 9.9 points on VAS (95 % CI 0.6-19.2, p= 0.04) and 0.16 points on HSCL-25 (95 % CI 0.01-0.32, p=0.04). 18 patients (24 %) in the surgery group and four patients (6 %) in the rehabilitation group reported full recovery (p=0.002). There were no significant differences between the groups in EQ-5D, occupational status, satisfaction with care or drug use. In the per-protocol analysis, the mean difference between groups was 8.1 ODI points (95 % CI 2.3-13.9, p=0.01) in favour of surgery. 43 of 61 patients (70 %) in the surgery group and 26 of 52 patients (50 %) in the rehabilitation group had a clinically important improvement (15 ODI points or more) from baseline (p=0.03). The proportion of patients with a clinically important deterioration (six ODI-points or more) were not significantly different between the groups. 21 patients (24 %) randomised to rehabilitation had crossed over and had undergone back surgery since inclusion. 12 patients (14 %) randomised to surgery had undergone additional back surgery. One serious adverse event after disc replacement is registered (<1%) (Paper I). Of all pre-treatment patient characteristics analysed for predictive value, only presence of Modic changes (type 1 and/or 2) was statistically significantly associated with an improvement of ≥ 15 ODI points. The probability of employment at eight-year follow-up was 1 % for patients with ≥ 1 year of sick leave, comorbidity, ODI ≥ 50 and ≤ nine years of education prior to treatment, and 87 % for patients with < 1 year of sick leave, no comorbidity, ODI < 50 and higher education (Paper II). ADD increased (for at least one ADD variable) in 23 of 57 patients (40%) treated non-operatively, and 29 of 69 patients (42%) treated with TDR (p=0.86). We found no significant associations between ADD increase and the change in ODI (Paper III).

Conclusions

Substantial long-term improvement can be expected both after disc replacement and multidisciplinary rehabilitation. The difference between groups is statistically significant in favour of surgery, but smaller than the pre-specified clinical important difference of ten ODI points that the study was designed to detect. Patients with Modic changes prior to the TDR surgery were more likely to report a clinically important functional improvement at long-term follow-up. Comorbidity, low level of education, long-term sick leave and high ODI score at baseline were associated with unemployment at eight years. Increased ADD occurred with similar frequency after TDR and after non-operative treatment, and was not related to the clinical outcome at eight-year follow-up.

1 Introduction

Low back pain (LBP) is common and causes more disability than any other condition [1]. The aetiology of LBP is usually multifactorial, but intervertebral disc degeneration (IDD) is often considered as an important pain source [2]. When non-operative treatment fails, some patients suffering from LBP are treated surgically. In the presence of IDD, LBP is sometimes considered 'discogenic', although the diagnosis has always been controversial [3]. The expression 'degenerative disc disease' (DDD) is used to describe the condition of LBP when IDD is suspected as the main pain source [4]. In such cases, spinal fusion has traditionally been the preferred surgical treatment. In randomised studies however, the results of spinal fusion have been similar to those of modern multidisciplinary rehabilitation (MDR) [5]. Total disc replacement (TDR) was introduced as a motion preserving surgical alternative to spinal fusion, and disc prostheses have been commercially available since the late 1980s [6]. In addition, early reports of adjacent level disc degeneration (i.e. degenerative disc changes at the level above the fusion) occurring after spinal fusion procedures have further encouraged the development and use of TDR, even though several reports have raised doubts about the role of fusion in adjacent level disc degeneration [7 8].

This thesis is based on the long-term follow-up of the Norwegian TDR Study; the only randomised study in which TDR is compared with non-operative treatment (i.e. MDR). The thesis explores the differences between the long-term clinical outcomes of MDR and TDR, investigates predictors for long-term outcome after TDR and describes degenerative disc changes at the adjacent level after MDR and TDR.

1.1 The intervertebral disc

The lumbar intervertebral disc is a fibrocartilaginous structure that acts as a shock absorber and allows limited segmental mobility [9 10].

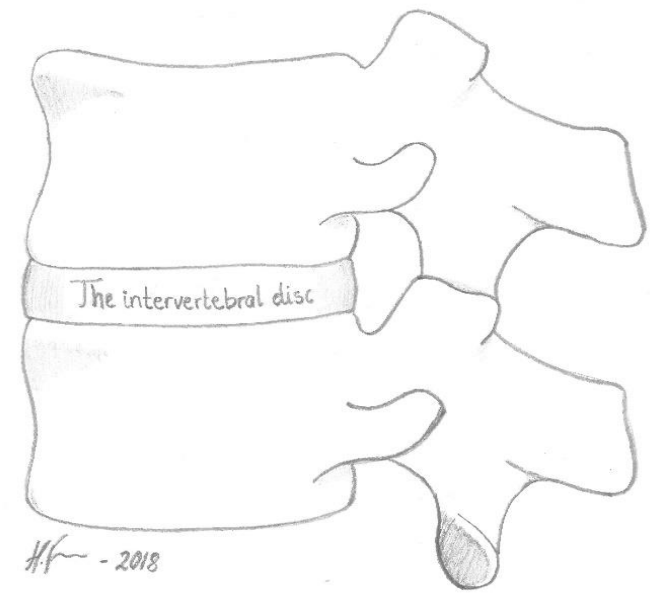


Figure 1. The intervertebral disc.

1.1.1 The normal disc

The intervertebral disc consists of an outer annulus fibrosus surrounding an inner nucleus pulposus. Collagen fibres tie the annulus to the anterior and posterior longitudinal ligaments and to the hyaline cartilage of the end plates of the superior and inferior vertebral bodies [9 11 12].

At birth, the cartilage end plates make up approximately 50 % of the intervertebral disc space, and have a rich blood supply. During the following decade, the blood circulation gradually ceases, and in adulthood the intervertebral disc is the largest avascular tissue in the body [9].

A normal adult intervertebral disc mainly consists of extracellular matrix and a small number of cells that make up approximately 1 % of the total disc volume [9]. Nucleus pulposus cells synthesize only type-II collagen and annulus fibrosus cells produce both type-I and type-II collagen. The nucleus is composed of collagen II and elastin fibres which are embedded in an aggrecan-containing gel. The aggrecan molecules are proteoglycans that interact with hyaluronan to form large aggregates that generate a high osmotic pressure, and contribute to the highly hydrated nature of the nucleus, thus maintaining disc height and distributing load across the end plates [11 13]. The annulus normally consists of 15-25 lamellae, and small amounts of elastin and type-III and type-IV collagen have been shown to have specific microanatomic locations [9 14].

1.1.2 The degenerated disc

Clefts and tears appear in the disc as part of the aging process, as well as increasing crack formation and thinning of the end plates, altered cell density, microfracture of the adjacent subchondral bone and bone sclerosis [9 15]. More advanced stages of degeneration include gross matrix changes, dehydration, increased lamellar disorganisation and fissures.

Histopathological changes include increased disc-cell proliferation, cell-cluster formation and increased cell death. At the molecular level, increased production of cytokines and matrix degrading enzymes such as metalloproteinases (MMPs) contribute to the degeneration of the disc matrix [9 13]. The distribution of structural matrix molecules like elastin and different collagen types is altered [9 11]. In earlier reports, there has been no obvious difference between the normal aging process of the disc and degenerative changes seen in younger individuals with LBP. Degenerative disc changes have therefore traditionally been considered as the early appearance of aging processes [9]. However, recent studies propose that the pathological process of disc degeneration should not be misinterpreted as a normal aging process, as both genetic features and certain environmental exposures are associated with early disc degeneration and LBP [11 16]. A Danish research group has also proposed that certain signs of disc degeneration (i.e. Modic changes) can occur due to bacterial infections [17 18]. Adams [12] suggested that IDD should be distinguished from ‘degenerative disc disease’ (DDD), as IDD simply describes a pathological process, while DDD describes a painful disc degeneration.

1.1.3 Radiological signs of IDD

Although some signs of disc degeneration can be identified on plain radiographs (e.g. reduced disc height, end plate sclerosis and osteophytes), MRI remains the gold standard for the

identification and evaluation of disc degeneration. The following characteristics are commonly used to describe disc degeneration:

1.1.3.1 Modic changes (MC)

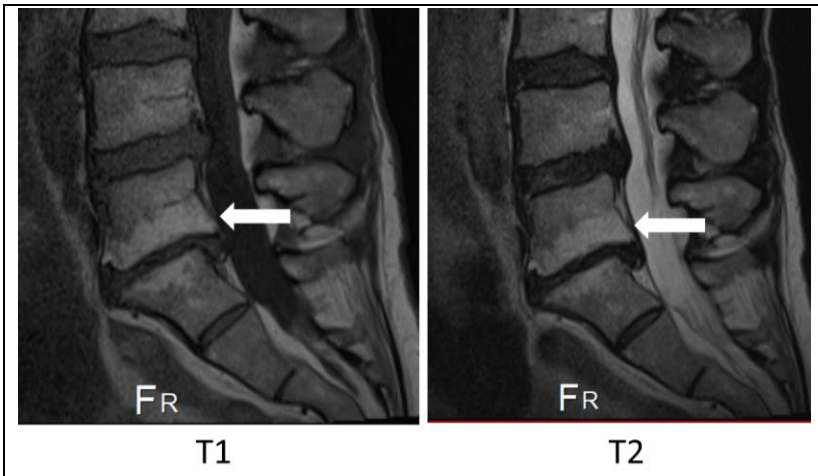


Figure 2. Vertebral end plate changes of Modic type 2 at disc level L5/S1 on T1- and T2-weighted sagittal MRI, respectively.

MC are changes in the vertebral bone marrow adjacent to the end plate, visible on MRI. MC were first described in 1988 by Michael T. Modic [19 20], who classified MC into three different types. Type 1 is characterised by a hypointense T1-signal and a hyperintense T2-signal on MRI, meaning that the changes represent bone marrow edema and inflammation [21]. Modic also did histopathological analyses of MC type 1 and found disruption and fissuring of the end plates and vascularised fibrous tissue. Type 2 is characterised by hyperintense T1-signal and iso- or hyperintense T2-signal on MRI, which are the features of fatty tissue. In the histopathological analyses of vertebral bodies with MC type 2, Modic found yellow marrow replacement. MC type 3 is characterised by hypointense T1-signal and hypointense T2-signal on MRI, and the changes are interpreted as sclerotic changes. Modic described the histopathological changes in MC type 3 as dense woven bone within the vertebral body, and demonstrated that such changes correlate with extensive bone sclerosis on plain radiographs [19]. Mixed types of MC may occur in the same vertebral end plate, and MC can transform from one type to another, suggesting that different MC types represent different stages of the same disease [22]. However, the pathway is not necessarily a sequential progression through the different types of MC, as reverse transformation from type 2 to type 1 is also observed [22 23].

Kjær et al. [24] reported a prevalence all types of MC of 22 % (15 % type 1 and 7 % type 2) in a cross-sectional study of a 40-year old Danish normal population. MC may be observed in asymptomatic individuals, but the presence of MC is reported to be associated with LBP. In a review, Jensen et al. [25] analysed 82 study samples from 77 original articles and found a median prevalence rate for any type of MC of 43 % in patients with non-specific LBP and/or sciatica and 6 % in asymptomatic populations. A positive association between MC and non-specific LBP was found in seven of ten studies from the general, working and clinical populations with ORs from 2.0 to 19.9. In a more recent review, Brinjikji et al. [26] reported a mean prevalence of MC of 12 % in asymptomatic individuals and 23 % in patients with LBP.

They found that MC type 1 was associated with LBP (OR 4.01, 95% CI 1.10 –14.55; P = 0.04), while no such association was found for all types of MC (OR 1.62, 95% CI 0.48–5.41, P = 0.43).

Modic originally considered MC to be a result of mechanical stress [20], but the aetiology of MC is still not completely understood. However, three leading hypotheses explain MC as a response to infectious, mechanical or inflammatory processes, respectively [22], the former drawing increasing attention over the last years. In 2001, Stirling et al. [27] reported that anaerobic microorganisms (Propionibacterium acnes and Corynebacterium propinquum) were isolated in samples from the nucleus pulposus of 53 % of patients operated for lumbar disc herniation. In 2013, Albert et al. [17] found Propionibacterium acnes in disc samples of 40 % of patients treated operatively for lumbar disc herniation. They also found that 80 % of those who had anaerobic bacteria isolated developed new MC adjacent to the previous disc herniation. In contrast, 0 % of those who had aerobic bacteria isolated developed new MC, and 44 % developed MC among those with negative cultures. They also included 162 patients with chronic LBP and MC type 1 in a randomised double-blind trial [18] in which patients treated with antibiotics (Bioclavid®, amoxicillin-clavulanate 500 mg / 125 mg three times a day for 100 days) had significantly better functional improvement and pain relief compared to the patients treated with placebo. A significant reduction of the size of the MC was also found in the antibiotic group, but not in the placebo group. Still, a more modern understanding of MC is that mechanical, inflammatory and infectious processes, or combinations of those processes, may all cause MC type 1 [28].

1.1.3.2 Disc height reduction

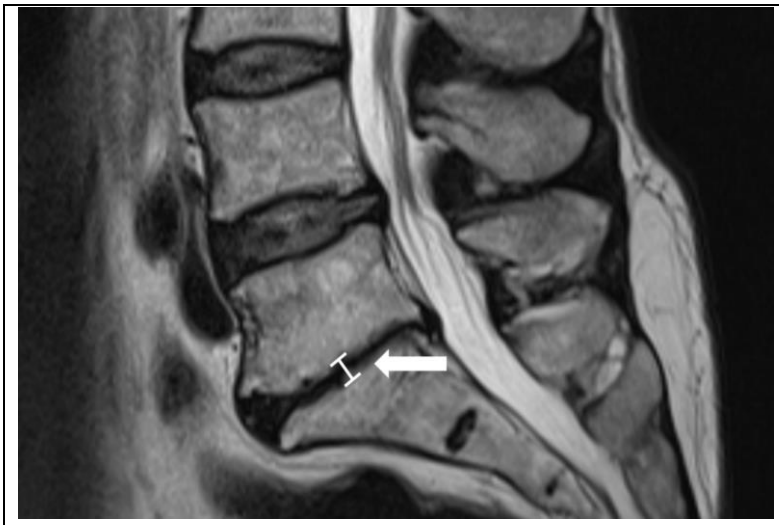


Figure 3. Disc level L5/S1 with disc height reduction compared to the disc levels above.

Disc height reduction is considered to be a sign of degeneration of the intervertebral disc [9 11]. In longitudinal studies, disc height reduction may be defined as a height reduction compared with earlier images, or, for lack of earlier images, as a proportion of the disc height in the superior level. In an earlier report from the Norwegian TDR Study, disc height reduction was defined as at least a 40 % height reduction compared to the next superior disc [29]. Masharawi et al. [30] defined the disc height as the distance between the mid-inferior

and mid-superior disc borders on a mid-sagittal MRI view, and validated this method for the evaluation of disc height. They also suggested that the measured disc height may depend on whether the patient is in standing position or laying down. Teichtahl et al. [31] reported that there is a dose-response relationship between the severity of disc degeneration measured by Pfirrmann classification [32] and intervertebral disc height. Twomey and Taylor [33] did not consider disc height reduction as a normal aging process, but rather as a pure pathologic feature. In contrast, Mannion et al. [34] found no correlation between a reduction in disc height and the clinical outcome in the long-term follow-up of 355 patients treated non-operatively or with spinal fusion. Videman et al. [35] reported a lumbar disc height decrease of 0.4 mm over five years and 1.0-1.3 mm over 15 years in a longitudinal study of Finnish monozygotic twins, and calculated the measurement error as approximately 0.6 mm. In an analysis of the degeneration of the adjacent disc two years after TDR or rehabilitation, Hellum et al. [36] reported a minimal detectable change in disc height of 2 mm.

1.1.3.3 Changed disc contour

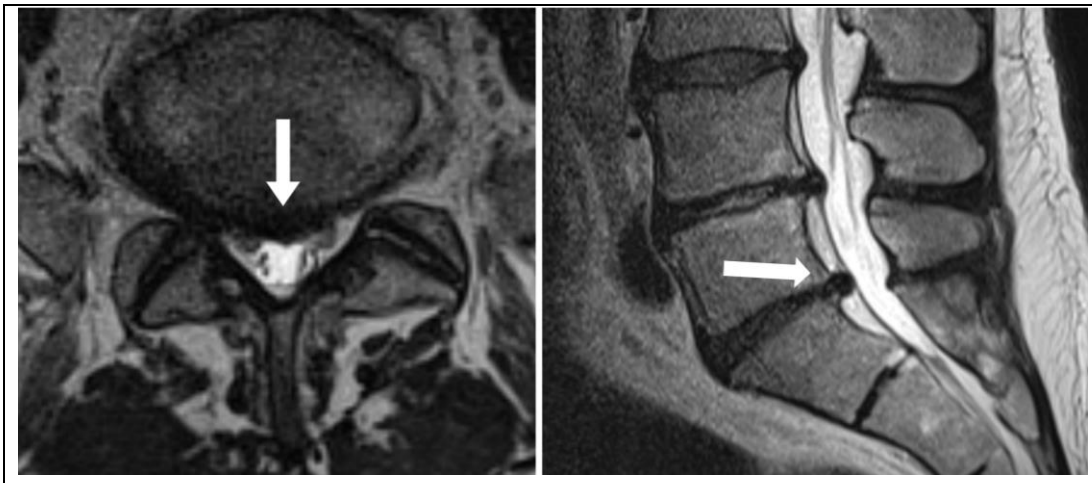


Figure 4. Changed contour of the intervertebral disc L5/S1.

The disc contour can be categorised as normal, bulging or herniated [4]. Disc bulging describes the situation where the outer annulus extends beyond the edges of the disc space in the axial plane, usually more than 25 % of the circumference of the disc and usually less than 3 mm beyond the edges of the vertebral body apophysis. Disc bulging may represent degeneration, and can be explained by loss of disc space height, ligamentous laxity, or as a response to loading or angular motion or remodelling in response to adjacent pathology [4]. Disc bulging is commonly found in asymptomatic individuals. Kjær et al. [24] reported a prevalence of 28 % in a cross-sectional study of a 40-year-old Danish normal population, and a positive association between disc bulging and LBP (OR 2.6, 95 % CI 1.4-4.4). Zou et al. [37] reported that disc bulging increased with the severity of disc degeneration in a cross-sectional study of 513 patients examined with kinematic MRI. Disc bulging should be distinguished from disc herniation, which is defined as a localised (e.g. < 25 % of the disc circumference) displacement of nucleus, cartilage, fragmented apophyseal bone or fragmented annular tissue beyond the intervertebral disc space [4]. Different types of disc herniation include disc protrusion, disc extrusion and disc sequestration [4].

1.1.3.4 Changed nucleus pulposus signal

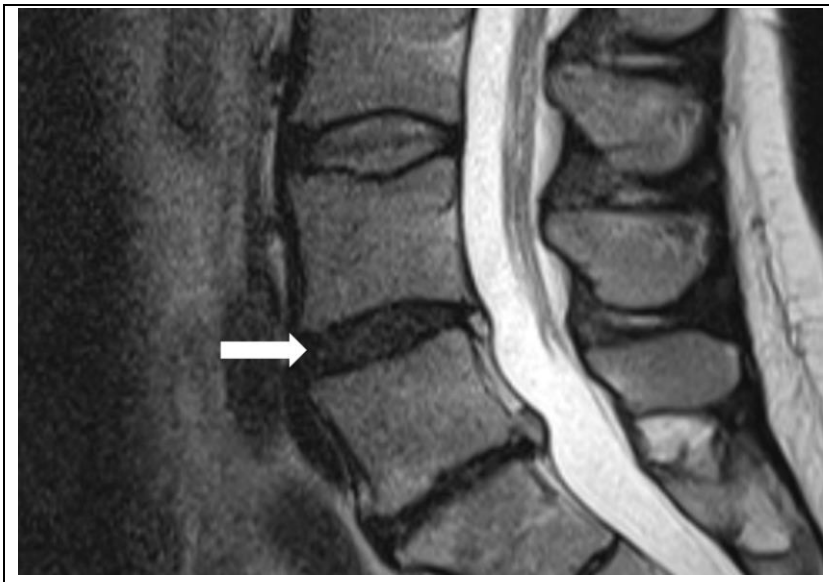


Figure 5. Reduced signal intensity of the disc L4/L5 due to dehydration.

Breakdown of hydrophilic proteoglycan content and concomitant increase in collagen inside the degenerated disc leads to dehydration of the disc [10 38], which is recognised by reduced signal intensity on T2-weighted MRI [39]. Luoma et al. [40] categorised the signal intensity in the disc, using the cerebrospinal fluid (CSF) in the adjacent dural sac as an intensity reference. They reported a significant association between a dark nucleus pulposus and a one-year incidence of LBP (OR 2.0, 95 % CI 1.2-3.1), but other researchers have not found such associations [41-43]. Decreased signal intensity in the disc is also commonly observed in patients without LBP. Kjær et al. [24] detected hypointense disc signals in 45 % of a 40-year-old Danish normal population. Moreover, the prevalence of decreased signal intensity in the disc increases with age [44].

1.1.3.5 Posterior High Intensity Zone (HIZ)

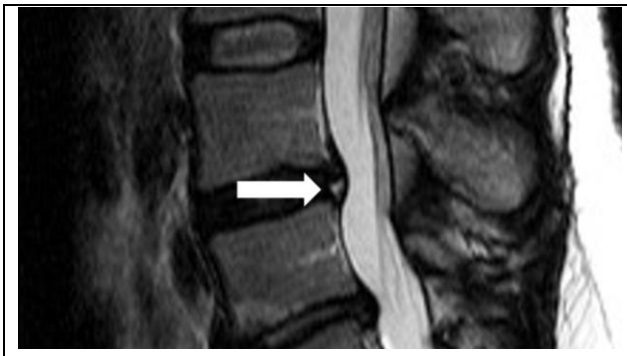


Figure 6. Posterior High Intensity Zone (HIZ).

A posterior High Intensity Zone (HIZ) was defined by Aprill and Bogduk [45] as an area of high signal intensity in the posterior annulus fibrosis that is brighter than the nucleus pulposus on T2-weighted images and surrounded superiorly, inferiorly and anteriorly by the low-intensity (black) signal of the annulus fibrosus. They claimed that HIZ was pathognomonic of an internally disrupted and symptomatic intervertebral disc, and reported a prevalence of 29

% for HIZ in a prospective study of 500 patients with LBP. Later studies have shown that HIZ may also be present in asymptomatic populations [46]. Carragee et al. [47] reported a prevalence of HIZ in 59 % of patients with LBP and 24 % in asymptomatic controls, while Liu et al. [48] observed HIZ in 46 % of patients with LBP and 20 % in asymptomatic controls. Kjær et al. [24] detected HIZ in 41 % of a 40-year-old Danish normal population, and a positive association between HIZ and care-seeking for LBP (OR 2.0, 95 % CI 1.2-3.0). Hence, HIZ can be observed in asymptomatic individuals, but seems to be more common in populations with LBP. As with other signs of IDD, the prevalence of HIZ increases with age [49].

1.1.4 Intervertebral disc degeneration and low back pain

Degenerative changes in the intervertebral disc are often found in MRI images, both of patients with LBP and of individuals without LBP. Endean et al. [50] have published a review with meta-analysis of 21 studies of the prevalence of IDD in individuals without LBP. The combined estimate of prevalence from all studies was 54 %, varying from 7 % to 85 %. However, there are also several reports of a significant association between LBP and degeneration of the lumbar disc [2 12 26 51-53]. In a recent meta-analysis of 3097 individuals by Brinjikji et al. [26], disc degeneration was more prevalent in patients below 50 years of age with LBP than in asymptomatic controls of a similar age (OR 2.24, 95 % CI 1.21-4.15). Nevertheless, no MRI lesions alone can be established as the cause of LBP, since MRI abnormalities are also common in asymptomatic individuals [2 50]. Furthermore, in a previous report from the Norwegian TDR Study, Berg et al. [54] found that more advanced IDD was not related to the degree of disability or the intensity of LBP.

1.2 Low back pain

1.2.1 Definition

In the European guidelines for the management of chronic non-specific LBP [55], LBP is defined as pain and discomfort localised between the costal margin and the inferior gluteal folds, with or without referred leg pain. LBP is considered as non-specific when it is not explained by nerve root affection or linked to specific spinal pathology such as infection, tumour, fracture, deformity or an inflammatory disorder. LBP can be classified, according to the duration of pain, as acute (< 12 weeks) or chronic (> 12 weeks) [56 57].

1.2.2 Epidemiology

LBP is reported to be the main cause of disability worldwide, with a global estimate of 72 318 000 years lived with disability (YLD) in 2013 [1] and a global point prevalence of 9.4 % [58]. The lifetime prevalence of LBP is reported to be as high as 84 %. For chronic non-specific LBP the lifetime prevalence estimate is 23 % [55]. Most episodes of LBP are self-limiting and not related to serious disease [2 59]. The epidemiological data are heterogeneous, and mean estimates need to be interpreted with caution [60]. The great variation in the reported estimates from different countries may be due to different definitions of the condition, different methods for reporting epidemiological data and different distribution of chronic LBP [60].

In Norway, a recent survey estimated the prevalence of chronic low back and neck pain as 7.9 % in women and 7.6 % in men, accounting for 22 % and 18 % % of contacts in primary care, and 1.2 % and 1.1 % of contacts in specialist health services, respectively [61]. According to the National Institute of Occupational Health, 32 % of the working population experience LBP during any one month [62].

1.2.3 Potential causes of chronic LBP

In a minority (about 10-15 %) of patients with LBP, there is a specific cause (i.e. nerve root affection, fracture, spondylolisthesis, cancer, ankylosing spondylitis infection or other) [55 63]. For a majority of patients with LBP (about 85 – 90 %) the pain has no obvious cause, and the diagnosis is based on the exclusion of specific pathology [64].

1.2.3.1 The biopsychosocial model

According to Wadell [64], low back pain is best understood from a biopsychosocial point of view. In this model, the origin of the pain is mainly pathoanatomical, but psychological factors such as the patient's personality, attitude, beliefs and psychological distress may modulate the perception of pain and influence the level of pain and disability. Environmental and social factors such as work status, socioeconomic status and social environment may also modulate pain perception and influence the experience of LBP. The biopsychosocial model also contributes to the understanding of the transition from acute to chronic LBP. Costa et al. [65] observed that the chronic stage of LBP in particular was characterised by a combination of physical, psychological and social dysfunction. Also, psychological and social factors may contribute both to the development and maintenance of pain and disability [66-68].

1.2.3.2 Potential anatomical pain sources

Chronic non-specific LBP is believed to have a multifactorial aetiology. Several somatic pain sources are reported, including the paraspinal muscles [69-71], the facet joints [72-75], the sacroiliac joints [73 76] and the degenerative disc, which is described in more detail above. Possible pathophysiological roles for tumour necrosis factor α (TNFα) and nerve growth factor have also been suggested, but the clinical implication of these findings needs further clarification [2].

1.2.4 Risk factors for LBP

There are a number of individual and environmental risk factors for LBP, of which some may be modified, and some may not [64 77].

An important risk factor for IDD is genetic inheritance. In twin studies, the heritability estimates for IDD were 29-61 % [78 79], indicating that heredity factors play a substantial role in IDD and LBP. Recent studies have identified several genes that have been associated with both the development and the progression of disc degeneration, including genes coding for different collagen types, aggrecan, Matrix-metalloproteinase-3, transmitter substances involved in pain perception such as Interleukin-1 and Interleukin-6, and vitamin D receptors [13 80-82]. Battie et al. [83] estimated that up to 25 % of the genetic effects on pain are attributed to the same genetic factors that affect disc height reduction. In the UK Twin Spine Study [53], there was a significant genetic correlation between LBP and IDD, suggesting that 11-13 % of genetic effects are shared by LBP and IDD. In a recent review of twin studies,

Ferreira et al. [84] reported heritability estimates for LBP of 21-67 %. Genes can also influence LBP through other mechanisms, such as pain perception, signalling, psychological processing and immunity [2 11 85 86]. Omair et al. [87] found that genetic factors are also partly responsible for the variation in disability levels in patients with chronic LBP.

Age is commonly considered as a risk factor for LBP [64 77]. Incidence of LBP is reported to be highest in the third decade of life [88-91], and overall prevalence increases with age until the age of 60-65 before it gradually declines [92 93]. However, for some more severe forms of LBP, such as osteoporotic vertebral fractures, tumours and spinal infections, the prevalence continues to increase with age [94 95].

Gender may be a risk factor for developing LBP, although some studies have found similar a prevalence in women and men [88 96]. Two systematic reviews have found that the prevalence of LBP was higher in women [77 97]. Women are also more likely to develop chronic LBP [98-100].

Patients' genetic constitution, age and gender are examples of risk factors that cannot be modified. In contrast, there are also a number of known risk factors that may be modified by patients and by society.

Obesity is one such risk factor. In a systematic review of twin studies, Dario et al. [101] detected a dose-response relationship between obesity and LBP. The relationship was weakened, but still significant, after adjusting for genetics and shared early environment. Later, a prospective study of Spanish twins identified no such relationship after two to four years when adjusting for genetics [102]. In a large Norwegian cross-sectional study (Helseundersøkelsen i Nord-Trøndelag, HUNT) [103], a significant positive association was found between BMI and risk of LBP among persons without LBP at baseline. The odds ratio for a BMI of 30 or more versus a BMI under 25 was 1.34 (95% CI 1.08-1.67) for men and 1.22 (95% CI, 1.03-1.46) for women, in analyses adjusted for age, education, work status, physical activity at work and in leisure time, smoking, blood pressure and serum lipid levels. A significant positive association was also established between BMI and recurrence of LBP among women. The effect of body height has also been evaluated in data from HUNT [104]. Women with no LBP at baseline and body height ≥ 170 cm) had a higher risk of LBP compared with women with body height < 160 cm after adjustment for other risk factors (relative risk 1.19, 95 % CI 1.03-1.37). No such relationship was established among men.

Physical comorbidities have been reported to affect the occurrence of LBP in several epidemiological studies [105 106]. A systematic review [107] detected a number of individual risk factors for developing disabling LBP, and an inferior general health status was among the identified risk factors. Especially in the older population, comorbidity is associated with increased prevalence of LBP [95 108 109]. Stewart Williams et al. [108] also demonstrated that individuals with more than one comorbid condition had higher odds for LBP compared to those with only one comorbid condition.

Level of physical activity may influence LBP. Kwon et al. [110] summarised eight systematic review reports and found no consistent causal relationship between physical activity at work and the risk of developing LBP. However, two recent studies of data from HUNT have

evaluated the relationship between physical activity and chronic LBP and detected a positive association between strenuous physical work and LBP [111], while physical activity in leisure time was negatively related to LBP [112]. Zadro et al. [113] found that twins with recent LBP were less likely to meet the physical activity guidelines from the World Health Organization (WHO) compared with those with no history of chronic LBP, but the relationship was not significant after adjusting for genetics and shared early environment. There are also several reports of the association between LBP and sports. In a recent systematic review of 43 studies of LBP in athletes, Trompeter et al. [114] found a large variation in the reported prevalence of LBP, and highest prevalence in rowing and cross-country skiing. Due to the methodological heterogeneity of the included studies, a detailed comparison of different sports or versus the general population was not possible. However, in two Swedish studies [115 116], elite alpine skiers had more degenerative disc changes and a similar lifetime prevalence of LBP (50 %) compared to non-athletic controls (44 %).

Smoking is associated with LBP in several cross-sectional studies [117-119]. A meta-analysis detected a higher prevalence of LBP in both former and current smokers, and a stronger association between current smoking and LBP in adolescents than in adults [118]. In a Finnish cohort study of adolescents, Mikkonen et al. [120] also demonstrated a dose-response relationship between pack-years and LBP in girls. However, the association may be confounded by differences in physical and psychological health and socio-economic status between smokers and non-smokers [64]. Further, socio-economic status can be assessed in several ways, and the methods used to indicate socio-economic status have been reported to influence the association between socio-economic status and LBP [121].

Education level may be considered as an indicator of socio-economic status, and is also reported as a risk factor for LBP. In a review of the literature on the relationship between education level and LBP, Dionne et al. [122] found that well-educated people were less likely to have disabling back pain. Later, Zadro et al. [123] performed a population-based study of Spanish twins, and found that women with higher education were less likely to develop LBP, but the association was not significant after adjusting for genetics and shared early environment.

Psychological comorbidities and hostile environment are commonly reported as risk factors for LBP. In a large prospective cohort study based on data from HUNT, Nordstoga et al. [124] demonstrated that presence of anxiety and depression reduced the probability of recovery from LBP (adjusted relative risk 0.77, 95 % CI 0.66-0.91). George et al. [125] demonstrated that patients with symptoms of depression and increased fear avoidance beliefs had a lower probability of recovery six months after an episode of LBP. Grotle et al. [126] performed a prospective cohort study and detected increased fear avoidance beliefs in patients with chronic LBP compared to patients with acute LBP, and fear avoidance beliefs predicted increased future pain and disability. Moreover, job related factors such as job dissatisfaction, monotonous tasks, poor work relations, demands, stress and low level of social support in the workplace are reported to be associated with increased occurrence of LBP [127-129]

1.2.5 Prognosis of LBP

In general, LBP is a benign and self-limiting condition. However, about 23 % of the population are expected to develop chronic LBP, and 11-12 % of the population are disabled by LBP [55]. In a Norwegian cohort study of 123 patients with acute LBP, Grotle et al. [130] found that 17 % had not fully recovered at 12-months follow-up, while Henschke et al. [131] described a slow recovery in most patients with acute LBP attending an Australian cohort study, and 28 % did not recover within a year. A review of 11 studies of the prognosis of LBP revealed that 33 % had recovered after three months, but 65 % still reported pain after a year [132]. A recent Danish cohort study with four- and eight-year follow-up [133] found that the prevalence rates of LBP were constant over time at a group level, but did not necessarily involve the same individuals. Those with more severe LBP were more likely to report future LBP. Work related factors such as low workplace support and long duration of sick leave are among the important predictors for chronicity of LBP [55]. Psychological factors such as patient expectations and psychological comorbidity may also influence the prognosis [55 134 135]. A systematic review of the role of fear avoidance beliefs suggests that they predict delayed recovery in subacute LBP [136]. In a Norwegian cohort study, Wilkens et al. [137] reported that both physical and psychological patient characteristics were associated with prolonged pain-related disability: Impaired fasting glucose tolerance, greater pain related disability, higher BMI, and lower quality of life. The predictive value of Modic changes has also been tested in a Norwegian cohort study [138], but was not prognostic for recovery from LBP.

1.3 Treatment for chronic LBP

There is great variation in treatment methods for chronic LBP. They can be divided into non-operative and operative treatments.

1.3.1 Non-operative treatment

Non-operative treatment is a heterogeneous group of treatments. Over the last years, the Cochrane Library has published several systematic reviews on different non-operative treatment methods. Paracetamol [139] was compared with placebo, and did not provide better pain relief for acute LBP, while for chronic LBP it was uncertain if paracetamol had any effect. Therapeutic ultrasound [140] was not effective in improving quality of life or relieving pain, and the small improvement of short-term physical function was considered clinically unimportant (mean difference 0.5 standard deviations). Several forms of exercise have been evaluated. Motor control exercise [141], a form of exercise that aims to restore the muscles that support the spine, provided better pain relief (mean difference 13 points on VAS) and functional improvement (mean difference 6 points on a 100-point scale) compared to minimal intervention after 12 months, but did not prove superior to other forms of exercise. Muscle energy technique [142] is a method used by some osteopaths, chiropractors and physiotherapists, and combines stretching and resisted muscle contractions. In the Cochrane review, there was no evidence for the effect of this treatment for patients with LBP. For Pilates [143], there was some evidence for the effectiveness on LBP compared to minimal intervention at 3-12 months follow-up (mean difference 10.5 points on a 100-point scale for pain and 11.2 points on a 100-point scale for disability), but there was no evidence for superior effectiveness compared to other treatments. Yoga [144] was compared to non-

exercise controls, and there was some evidence for better short-term pain relief (mean difference 4.6 points on a 100-point scale at three to four months follow-up) and physical improvement (mean difference 2.2 points on a 100-point scale at six months follow-up), but the effect sizes were small and there were also more adverse events (i.e. increased back pain) in the yoga group. Evidence for the effect of yoga compared to exercise was lacking.

Some non-operative treatment forms are based on the biopsychosocial model. Behavioural treatment aims to modify inappropriate cognitive processes and pain behaviour. Three different treatment approaches are often described: The operant approach involves the reduction of external factors that are believed to reinforce pain behaviour, such as rest, analgesic medication and disease-related attention from other people, alongside the promotion of exercise and work [145]. The cognitive approach aims to identify and modify inappropriate thoughts, feelings and beliefs that patients with chronic LBP may have [146]. Cognitive patterns may be restructured through imagery and attention diversion or education [147]. The respondent approach aims to reduce pain through reduction of muscular tension [148 149]. Behavioural treatment often consists of a combination of these approaches. A Cochrane review [150] concluded that moderate evidence exists for better short-term pain relief of behavioural treatment compared to usual care (mean difference 5.2 points on a 100-point scale). Long-term effects were equivalent to those of group exercises. Further, no specific type of behavioural treatment was more effective than another. Back schools combine patient education and exercise. However, many variations have evolved. A recent Cochrane review [151] reported generally very low quality of evidence for back schools, and found, at best, a trivial effect in the treatment of chronic LBP.

Multidisciplinary rehabilitation (MDR) is also based on the biopsychosocial model, as different treatment approaches are combined in order to improve the patient's ability to cope with the disease, to modify inappropriate disease-related behaviour, and thereby reduce pain and disability. Through identification and modification of inappropriate thoughts, beliefs, fears and behaviours, the treatment aims to reassure the patients that it is not harmful for them to move even though it is painful. They are exposed to activities that they may fear and automatically avoid, such as physical exercise and work, and will thereby experience improved confidence as they manage these tasks. A Cochrane review [152] that evaluated MDR in the treatment of chronic LBP reported better effect of MDR on pain (standardised mean difference 0.2) and disability (standardised mean difference 0.2) compared to 'usual care', and better effect of MDR on pain (standardised mean difference 0.5), disability (standardised mean difference 0.7) and work status (OR 1.9) compared to different physical treatment programs. Patients with indicators of significant psychosocial impact were supposed to be more likely to benefit from MDR. This is in agreement with current guidelines [153] that recommend MDR for patients with psychosocial obstacles to recovery and in cases where previous treatment has not been effective.

1.3.2 Surgical treatment

Spinal fusion surgery was first described in 1891 by Berthold Hadra who repaired a cervical fracture dislocation by wiring together the spinal processes of the sixth and seventh cervical

vertebrae [154]. Spinal fusion has since evolved from the treatment of fractures and tuberculosis, to the treatment of the degenerative spine [155]. Its aim is to relieve pain derived from degenerated anatomical structures such as intervertebral discs or facet joints by restricting segmental movement. Spinal fusion is still considered the 'gold standard' for the surgical treatment of the degenerated spine [156 157]. However, randomised trials find similar effects of spinal fusion and MDR both in the short- [158 159] and long-term [160], and spinal fusion has several potential adverse effects. Beside surgical complications like infections and damage to neural and vascular structures [158 159 161-164], the loss of segmental movement may reduce the patient's mobility [165 166], and is also reported to accelerate the degenerative process of the adjacent level (i.e. the level above the fusion) [167-169]. On a group level, adjacent level degeneration (ALD) does not seem to influence the clinical outcome [34]. Nevertheless, numerous motion-preserving alternatives to spinal fusion have evolved [169 170] in order to preserve the segmental motion and to reduce the risk of ALD; among them is total disc replacement (TDR).

1.3.2.1 Total disc replacement

In other orthopaedic subspecialties, joint fusion (arthrodesis) has failed and arthroplasty has been successful [171]. This has inspired the development of alternatives to spinal fusion, such as TDR, a surgical procedure in which the degenerated IVD is removed and replaced with an artificial disc. The first attempts to replace the IVD with an artificial implant were carried out by the Swedish surgeon Ulf Fernström [172]. He replaced the IVD with a spherical steel implant and achieved acceptable clinical results, even though subsidence was common and most levels fused [173]. Further development of potential motion-preserving disc implants included fluid-filled elastic chambers [174], silicon containing devices [175 176], titanium springs [177] and rubber implants [178]. However, the first disc prostheses that became commercially available were the metal-polyethylene devices Charité [6] and ProDisc I [179] in the late 1980s [157 180 181]. Today, several disc prostheses with different mechanical and geometrical properties are in use [182 183]. Different prostheses may be classified as unconstrained, semi-constrained and constrained, depending on the range of motion (ROM) permitted by the implant [181].

The Food and Drug Administration (FDA) regulated two randomised multicentre studies comparing TDR with fusion which showed that the clinical results of TDR were at least as good as those of fusion [184 185]. The results of these studies and four other randomised studies comparing TDR with fusion [186-189] were analysed in a Cochrane review [156]. Overall, the clinical results of TDR were better than those of fusion, but the differences were considered to be clinically unimportant. The estimated differences between the treatments in favour of TDR were 4.3 ODI points, 5.2 points on Visual Analogue Scale (VAS) for back pain and 8.6 and 8.8 points on VAS for patient satisfaction, for one- and two-level procedures, respectively [156].

In the Norwegian TDR study, Hellum et al. [29] found that the clinical two-year results of TDR were significantly better than those of multidisciplinary rehabilitation, but the differences between the groups (8.4 ODI points and 12.3 points on VAS for back pain) were smaller than the predefined clinically important difference (10 ODI points and 15 points on

VAS for back pain). The results of the Norwegian TDR study were also analysed in the Cochrane review [156], but since it is the only randomised study in which TDR is compared to non-operative treatment, the evidence for the comparison was considered of low or very low quality, even though the study was considered a low risk of bias trial.

One of the limitations of the Cochrane review is that it is based on two-year results only [156]. According to the authors' conclusion, long-term follow-up studies are needed since clinical long-term results are lacking and certain questions related to TDR can only be answered after longer implantation periods. Examples of such unanswered questions are related to wear and loosening of the implants, and degeneration of the facet joints and the adjacent disc.

12 prospective studies with a total of 1764 patients have more than five years follow-up [168 190-200] (Appendix, Table 1). Mean follow-up time was five to 12 years, and follow-up rate was 43-99 %. ODI change was used as an outcome measure in seven of the reports [190-193 196 198 200], and the mean improvement was 16-31 points. Reoperation rate was 4-39 %.

In addition, six retrospective studies with a total of 434 patients have more than five years follow-up [201-205] (Appendix, Table 2). Mean follow-up time was five to 17 years, and follow-up rate was 70-98 %. The results of the retrospective studies were not consistent. Lemaire et al. [201] and David et al. [203] reported 90 % and 82 % good or excellent clinical outcome, respectively, while Putzier et al. [202] reported a mean ODI of 42 points at last follow-up, and spontaneous ankylosis at the TDR level in 60 % of the patients. Patients with spontaneous ankylosis had better clinical results (ODI 38 versus 52 points and VAS 4.5 versus 6.1), and more frequently degenerative changes in the superior adjacent level (17 % versus 0 %) compared to those with functional disc prostheses. Reoperation rates were 5-11 %.

A recent systematic review [206] based on 59 retrospective and prospective studies with short- and long-term results of TDR found similar clinical outcomes and complication rates for TDR compared to fusion in the majority of the studies, and the authors suggest that TDR could be a reliable option for the treatment of LBP and IDD in years to come. In contrast, NICE guidelines recommend that disc replacement is not offered to patients with LBP [153].

Potential adverse effects of TDR should also be taken into account. Van den Eerenbeemt et al. [207] classified surgical complications as approach related (2-19 %, e.g. vascular injury, nerve root damage or retrograde ejaculation), implant related (2-39 %, e.g. subsidence, migration, dislocation, implant failure or end plate fracture) or related to the treatment (2-62 %, e.g. wound, pain or neuromusculoskeletal complications) [207]. Siepe et al. [208] reported more complications after two-level TDR procedures (L4/L5 and L5/S1) compared to single level procedures. In a retrospective study of 2415 patients treated with TDR, Eliasberg et al. [209] reported 0.3 % incidence of wound infections. The incidence of subsidence and migration of the prostheses is reported to have decreased, probably due to an increased surface area of the end plates covered by modern implants [183]. Possible late complications include loss of mobility, implant wear or loosening and degeneration of the facet joints and adjacent level [156].

The mobility in TDR levels is reported to decrease gradually over time, but the reduced ROM does not seem to be correlated to the clinical outcome [199 210]. Loss of mobility may be due to heterotopic ossification, which is commonly observed after disc replacement [183 198]. McAfee et al. [211] classified heterotopic ossification as 0 (no evidence of heterotopic ossification) to 4 (apparent bridging bone between the end plates). Putzier et al. [202] reported spontaneous ankylosis in 60 % of the patients treated with TDR after mean 17-year follow-up in patients operated between 1984 and 1989, but modern disc prostheses cover more of the vertebral end plates [183], which may reduce the incidence of spontaneous ankylosis. At the two-year follow-up in the Norwegian TDR Study, Johnsen et al. [212] used distortion compensated roentgen analysis (DCRA) to evaluate the mobility at the TDR level, and found that mobility was similar in a typical TDR level and in a typical degenerated disc.

The observation of complications due to implant wear in hip arthroplasty [213] has also sparked concern among spine surgeons. Some studies have found elevated metal ion levels after TDR with metal-on-metal-bearings [194 214 215] but the metal ion levels were mostly moderately elevated, indicating a low risk of complications due to toxicity. However, there are also some case reports of adverse local tissue reaction due to metal debris [216-218]. Polyethylene wear debris is also reported to induce inflammation, vascularisation and innervation in periprosthetic tissue after TDR [219], and this may be clinically relevant [220]. Baxter et al. [221] suggested that the occurence of biologically relevant polyethylene particles may be due to severe rim impingement. Although implant wear debris seems to be a smaller problem after TDR than after hip replacement [194], the clinical consequences of implant wear after TDR are not fully understood [222].

Some studies report less degeneration of the adjacent level after TDR compared to fusion [167 223 224], but adjacent level degeneration is also part of the natural course of degeneration of the spine [8 169 225], and it has probably limited clinical relevance at a group level [34 226].

Biomechanical changes after TDR have been tested in validated finite element model studies [227-229] that show increased loading on facet joints after TDR, particularly if there is malalignment of the vertebra adjacent to a disc prosthesis. Siepe et al. [230] found progressive degeneration in 20 % of the facet joints at mean four-year follow-up after TDR and inferior clinical results in patients with such progressive changes. Progressive facet joint degeneration occurred more frequently in L5/S1 than in L4/L5, and more frequently at index-levels than other levels. Park et al. [231] reported increased facet joint degeneration in 29 % of TDR levels at two-year follow-up, and found that facet joint degeneration was more common in females, patients with implant malposition and patients with two-level TDR, while Shin et al. [232] found that preoperative facet tropism (i.e. asymmetry in both facet joint angles) was associated with progressive facet joint degeneration after TDR.

Since degeneration of the adjacent level and the facet joints may occur regardless of any surgery, development of such degenerative changes following surgery should be compared to the natural course or non-operative treatment. The Norwegian TDR Study is the only study that has compared TDR with non-operative treatment. At two-year follow-up, Hellum et al. [36] found a similar development of degenerative changes in the superior adjacent level in

patients treated with TDR and patients treated non-operatively, but the incidence of progressive facet joint degeneration was significantly higher in the TDR group (34 %) than in the group treated non-operatively (4 %).

2 Aims of the thesis

The main aim of this thesis was to provide evidence-based knowledge of the long-term clinical and radiological results of lumbar total disc replacement compared to multidisciplinary rehabilitation, and to search for better selection criteria for disc replacement.

The specific aims were:

I: To evaluate the long-term efficacy of total disc replacement compared with multidisciplinary rehabilitation in patients with chronic low back pain and intervertebral disc degeneration (Paper I).

II: To identify baseline characteristics associated with (1) a clinically important improvement (≥ 15 ODI points) and with (2) employment at eight-year follow-up after total disc replacement (Paper II).

III: To assess the long-term development of adjacent disc degeneration after total disc replacement or non-operative treatment, and to analyse the association between development of adjacent disc degeneration and the clinical outcome (Paper III).

3 Material

This PhD project is an eight-year follow-up of patients included in the Norwegian TDR Study. Inclusion in the Norwegian TDR study took place from April 30, 2004 to September 27, 2007. After eight years, the patients were invited to a long-term follow-up, including collection of patient reported outcome measures (PROMs) and radiological examination. All patients were also offered a follow-up visit with a spine surgeon. The eight-year follow-up was carried out from July 3, 2012 to January 1, 2016. The patients completed all outcome questionnaires at home and returned them by mail to an independent observer before the follow-up visit.

3.1 Patients

The patients in the Norwegian TDR Study were referred as usual from local hospitals or primary care in all health regions in Norway to their nearest university hospital with no additional attempts at recruitment. They were screened according to inclusion and exclusion criteria (Table 3) at one of five university hospitals by a spine surgeon or a specialist in physical medicine and rehabilitation. If the surgeon and the specialist in physical medicine and rehabilitation agreed on inclusion after a second examination with both doctors present, the patient was included. The patients were thoroughly informed about the advantages and disadvantages of both treatment options and the fact that neither of the treatment methods was documented as clinically superior to the other. Written informed consent was obtained from all patients before inclusion. For the eight-year follow-up, new written informed consent was obtained.

Table 3. Inclusion and exclusion criteria for the Norwegian TDR study.

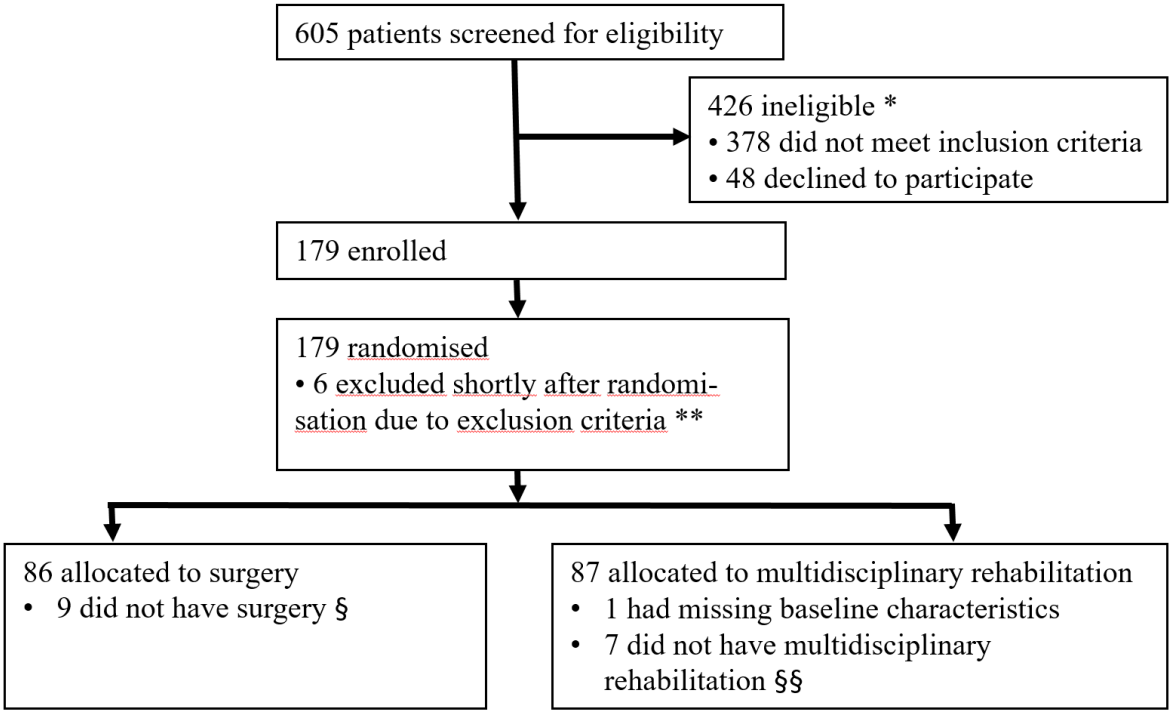
Inclusion criteria	**Exclusion criteria**
Age 25-55 years	Generalised chronic pain (e.g. fibromyalgia or widespread myofascial pain)
Low back pain as the main symptom for ≥ 12 months	Degenerative intervertebral disc changes in ≥ 3 levels
Structured physiotherapy or chiropractic treatment for ≥ 6 months without sufficient effect	Symptoms of spinal stenosis or MRI showing disc protrusion or recess stenosis with nerve root affection
Oswestry Disability Index score ≥ 30 points	Spondylolysis or isthmic spondylolisthesis
Degenerative intervertebral disc changes in L4/L5 and/or L5/S1, according to the following criteria based on MRI evaluation of the intervertebral disc:	Arthritis, osteoporosis or former fracture of L1-S1
	≥ two former microsurgical interventions
- ≥ 40 % disc height reduction	Congenital or acquired spinal deformity
- Modic changes type 1 and/or type 2 [20]	Drug abuse
- High Intensity Zone [45]	Ongoing psychiatric or somatic disease that excludes either one or both treatment alternatives
- Decreased signal intensity [40]	Unable to understand spoken or written Norwegian

The classification of the intervertebral disc as degenerated or normal was performed by two independent and experienced radiologists. When disagreement occurred, the images were evaluated by a third experienced radiologist, and the disc was classified by simple majority. Discs were classified as degenerated if the disc height was reduced by ≥ 40 % compared to the superior adjacent disc, or if at least two of the other three criteria were present (Table 3). In contrast, discs were classified as normal if disc height reduction was < 40 % and if no other criteria of IDD (Modic changes, HIZ or decreased signal intensity) were present.

Figure 7 shows how the patients were screened, enrolled, randomised and treated according to randomisation. In the inclusion period (from April 30, 2004 to September 27, 2007), no patients were treated with TDR outside the study setting at any of the participating hospitals. Patients who were eligible for the study but refused to participate were treated according to established methods (i.e. rehabilitation or spinal fusion).

Most baseline characteristics were similar in the two intervention groups, but low back pain score was significantly worse in the rehabilitation group than in the surgery group.

Figure 7: Trial profile showing the screening, enrollment, randomisation and treatment according to randomisation of the participants of the study.



* 426 patients were screened and found to be ineligible for the study due to inclusion and/or exclusion criteria:

- Insufficient degenerative intervertebral disc changes (n=29)
- Degenerative intervertebral disc changes in ≥ 3 levels (n=80)
- ODI < 30 (n=88)
- Refused the surgical treatment option (n=28)
- Refused the non-operative treatment option (n=20)
- Generalised pain (n=20)
- Previously treated with similar rehabilitation programme (n=26)
- Other reasons such as age, coccygodynia, deformity, fracture, hip arthrosis, language problems, previous surgery, psoriasis arthritis, spondylodiscitis, tumour (n=135).

** Six patients were enrolled, but excluded shortly after randomisation to rehabilitation (n=3) or TDR (n=3) due to coronary heart disease and heart attack a few days after randomisation (n=1) or obvious exclusion criteria discovered a few days after randomisation (n=5) (previous major abdominal surgery (n=1), insufficient degenerative intervertebral disc changes to satisfy inclusion criteria (n=2) or degenerative intervertebral disc changes in ≥ 3 levels (n=2)).

§ Of the 86 remaining patients randomised to TDR, nine changed their minds after randomisation and refused surgical treatment (three had social reasons for refusing the treatment, one had work related economic reasons and five wanted a guarantee of success).

§§ Of the 87 remaining patients randomised to MDR, seven changed their minds after randomisation and did not complete the rehabilitation program (two due to work related

economic reasons, two had a long travel distance and could not stay at a hotel, one had surgical treatment for lumbar disc herniation, one had social reasons and one had reasons unknown). Six patients did not complete the rehabilitation (i.e. > 50 % of the rehabilitation program) (one found the rehabilitation program disappointing, one did not manage to get through the training program, one developed diabetes just before or under treatment, one had surgical treatment for lumbar disc herniation, one had psychosocial reasons and one had hypertension and their family doctor did not recommend exercise). One had an untraceable baseline questionnaire.

4 Methods

4.1 Design

We used a multicentre randomised controlled design to compare the results of lumbar TDR with MDR in patients with chronic LBP and IDD.

4.2 Randomisation

A statistician not involved in the trial created a computer generated random list (1:1 allocation and random block sizes of two to eight). Allocation was performed using a website hosted by the Medical Faculty at the Norwegian University of Science and Technology (NTNU). Once a patient was included, a coordinating secretary, blinded to the patients' characteristics and not involved in the inclusion process, logged the patients' information on the randomisation website and performed randomisation. The treating unit and the patient were informed about the allocation shortly after randomisation and patients were treated within the next three months. Randomisation was stratified by centre and by previous back surgery (microsurgical decompression) or not. Independent observers collected and entered data.

4.3 Follow-up

PROMs were obtained before randomisation and at follow-up consultations at six weeks, three months, six months, one year, two years and eight years after the intervention was completed. At eight-year follow-up, the patients completed all outcome questionnaires at home and returned them by mail to an independent observer. After that, they were offered a follow-up visit, at which they could ask questions and receive information about the results of the radiological examinations.

Papers I and III: Prospective controlled studies. In Paper I, the randomised design was maintained in the main analyses since they were performed according to intention-to-treat principles. In Paper III, the patients were classified as being treated non-surgically or with TDR, regardless of randomisation, according to as-treated principles.

Paper II: A prospective cohort study of patients treated with TDR (including those who were originally randomised to MDR).

4.4 Interventions

The two treatment options are described below. Regardless of allocated treatment and compliance to the study protocol, the patients were invited to follow-up visits at six weeks, three months, six months, one year, two years and eight years after treatment.

4.4.1 Multidisciplinary rehabilitation

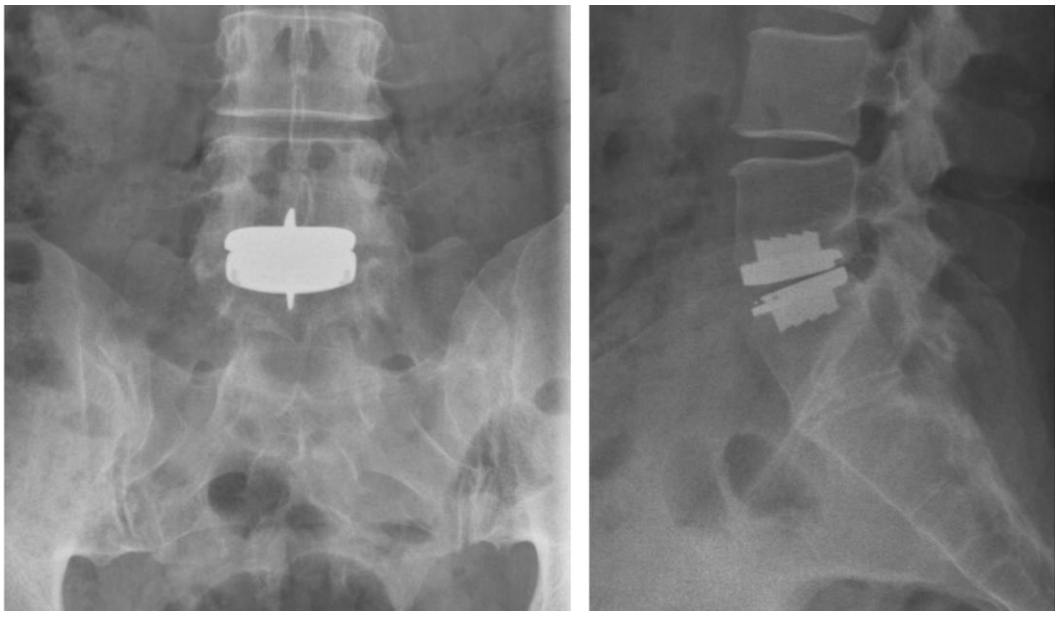
The multidisciplinary rehabilitation program consisted of a cognitive approach and supervised physical exercise as described by Brox et al. [158]. This modern multidisciplinary rehabilitation program has previously been reported to have long-term clinical results similar to spinal fusion [160]. The intervention was standardised through three seminars for the personnel responsible for the treatment, as well as videos and lecture sessions for the treatment providers before the study. A team of physiotherapists and specialists in physical medicine and rehabilitation at each study centre led the rehabilitation, and could be assisted by nurses, social workers and psychologists. The rehabilitation was organised as an outpatient

treatment in groups and lasted for about 60 hours over 12 to 15 days within a period of three to five weeks. The treatment consisted of lectures, individual discussions focusing on relevant topics (e.g. anatomical and physiological aspects of the spine, coping strategies, diagnostics, family and social life, imaging, normal reactions, pain medicine and working conditions), daily workouts to increase physical capacity, endurance, strength, coordination, and specific training of the abdominal muscles and the lumbar multifidus muscles, as well as challenging patients' thoughts about, and participation in, physical activities which were previously not recommended (such as lifting, jumping, vacuuming, dancing and ball games).

4.4.2 Total disc replacement

The surgical intervention consisted of the replacement of the degenerative intervertebral lumbar disc with an artificial disc (ProDisc II, Synthes Spine) (Figure 8). All hospitals participating in the study used the same artificial lumbar disc device. The ProDisc consists of two metal end plates (cobalt chromium molybdenum alloy) and a polyethylene (ultra-high molecular weight polyethylene, UHMWPE) core that is fixed to the inferior end plate after insertion. The implants were available in two sizes (large, medium), two shapes (6º, 11º) and three heights (10 mm, 12 mm, 14 mm). Anterior approach was used with a Pfannenstiel, median or para-median incision, depending on the surgeon's preference, and retroperitoneal dissection. When TDR was performed at the disc level L4/L5, the ascending lumbar vein was ligated. In order to avoid violation of the visceral organs and the sympathetic nerves, the disc was carefully exposed. A thorough discectomy with removal of the cartilaginous end plates and release of the posterior longitudinal ligaments was performed in order to ensure disc space mobilisation. A fluoroscope was used to ensure proper midline and posterior positioning of the prosthesis. Surgeons were required to have inserted at least six disc prostheses before performing surgery in the study, and one surgeon at each centre had the main responsibility for the operation. There were no major postoperative restrictions. Patients were not referred for postoperative physiotherapy, but if requested they could be referred for general mobilisation and non-specific exercises at six-week follow-up.

Figure 8. Study participant with a typical ProDisc II artificial disc implanted at L4/L5.



4.5 Outcome measures

The collected data included background variables, generic and specific back related questionnaires and radiological data (Table 4), according to international recommendations at the time of patient inclusion [233]. The chosen outcome measures are still in line with current recommendations [234 235].

Table 4. Outcome measures used in Papers I-III

Outcome measure	**Used in paper**
Patient-reported questionnaires	
Physical function and pain (Oswestry Disability Index, ODI)	I, II, III
Low back pain (Visual Analogue Scale, VAS)	I
Quality of life (EuroQol, EQ-5D)	I
Psychological distress (Hopkins Symptom Check-List, HSCL-25)	I
Work participation	I, II
Satisfaction with the result of the treatment (7-point Likert scale)	I
Satisfaction with care (5-point Likert scale)	I
Reoperations	I, III
Complications	I
Daily use of analgesic	I
Radiological outcome measures	
Extent of Modic changes	III
Disc height	III
Disc contour	III
Disc herniation size	III
Nucleus pulposus signal	III
Posterior high intensity zone (HIZ)	III

4.5.1 Primary outcome measure

4.5.1.1 Oswestry Disability Index

Oswestry Disability Index (ODI) version 2.0 was used as the primary outcome measure in Papers I and II. Fairbank and colleagues published version 1.0 in 1980 [236] and version 2.0 in 2000 [237]. ODI has since been translated into Norwegian and found to be a reliable and valid tool for assessment of physical function in Norwegian speaking patients with LBP [238]. ODI consists of ten questions about back specific pain and physical function, with six alternative answers for each question (0-5). Six questions represent different physical functioning activities, while the remaining four represent other health constructs such as pain intensity, sleep and social functioning. The total score ranges from 0 to 100, with higher scores indicating lower physical function. Mean ODI in asymptomatic, normal populations is approximately 10 points [237].

4.5.1.2 Adjacent disc degeneration

In Paper III, increased adjacent disc degeneration (ADD) evaluated with magnetic resonance imaging (MRI) was used as the primary outcome. Six ADD variables were analysed at the nearest level above the implanted or degenerated index level, i.e. at L3/L4 or L4/L5: Modic changes, disc height reduction, disc contour, herniation size, nucleus pulposus signal and posterior HIZ. ADD was defined as increased if at least one ADD variable had an increased rating value from pre-treatment to eight-year follow-up, as described by Hellum et al. [36]. Decreased ADD (yes/no) was defined as a decreased rating value for at least one of the six ADD variables. Unchanged ADD implied that all rating values were unchanged. Two patients who had an increased rating value for one ADD variable and a decreased rating value for another were both classified as having increased ADD.

4.5.2 Secondary clinical outcome measures

Secondary clinical outcome measures were patient reported and included pain, quality of life, psychological distress, work participation, patient satisfaction, additional treatment including reoperations, complications and daily use of analgesics.

4.5.2.1 Pain

LBP and lower limb pain were assessed with two separate horizontal visual analogue scales (VAS), which are unidimensional measures of pain intensity [239]. The scales are 100 mm in length and anchored by ‘no pain’ (score 0) and ‘worst pain imaginable’ (score 100). The VAS has been found to be reliable and responsive in populations with chronic LBP [234].

4.5.2.2 Health related quality of life

Health related quality of life was measured with EuroQol (EQ-5D) [240], a standardized tool for assessment of generic health status. Health status is measured in terms of five dimensions (5D): Mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Scores range from -0.59 to 1, and a value of 1 represents perfect health. A preference weight for health status can be estimated based on EQ-5D, and when that weight is combined with time, Quality Adjusted Life Years (QALY) can be calculated and used in cost-utility analysis. Due to its general health properties, EQ-5D has been used to compare the utility of spine surgery with other orthopaedic surgical procedures [241].

4.5.2.3 Psychological distress

Psychological distress was assessed with the Hopkins Symptom Check-List (HSCL-25) [242]. It consists of 25 items, ten items for anxiety symptoms and 15 items for depression symptoms. The scale for each question is rated 1 (‘not at all’) to 4 (‘extremely’), and the total score is the average of all 25 items.

4.5.2.4 Work participation

Self-reported work status was an outcome measure in Papers I and II. The patients could categorise themselves as 1 (‘employed’), 2 (‘on sick leave’), 3 (‘on active sick leave’), 4 (‘on part-time sick leave’), 5 (‘homemaker’), 6 (‘student’), 7 (‘retired’), 8 (‘unemployed’), 9 (“on vocational rehabilitation’) or 10 (‘receiving disability pension’). The data were dichotomised in both studies: Patients who reported full- or part-time employment, or were students, were

categorised as ‘employed’, while patients who gave any other answer were categorised as ‘not employed’.

4.5.2.5 Patient satisfaction

Satisfaction with the result of the treatment was assessed with a 7-point Likert scale [243], where the patients were asked to categorise the result of the treatment, compared to their situation before the treatment, as 1 (‘full recovery’), 2 (‘much better’), 3 (‘slightly better’), 4 (‘no change’), 5 (‘slightly worse’), 6 (‘much worse’) or 7 (‘worse than ever’). Satisfaction with care was assessed with a 5-point Likert scale [243], where the patients could rate their satisfaction as 1 (‘satisfied’), 2 (‘somewhat satisfied’), 3 (‘mixed’), 4 (‘somewhat dissatisfied’) or 5 (‘dissatisfied’).

4.5.2.6 Additional treatment including reoperations

The patients were asked if they had had any back surgery after the allocated treatment was completed. If so, they were asked to specify the year, the type of reoperation and the hospital at which the treatment was performed. We also asked for permission to contact that hospital and obtain information about the reoperation. Further, the patients were asked if they had received any other medical or physical treatment after the allocated treatment was completed, and if they could specify the types of treatment and quantify the number of treatments.

4.5.2.7 Complications

Information about early complications related to the treatments was collected at two-year follow-up. At eight-year follow-up, the patients were not specifically asked about complications, but they were asked to describe any symptoms that they had not mentioned elsewhere in the questionnaire. Information about late complications that required reoperations or other additional treatment was obtained.

4.5.2.8 Daily use of analgesics

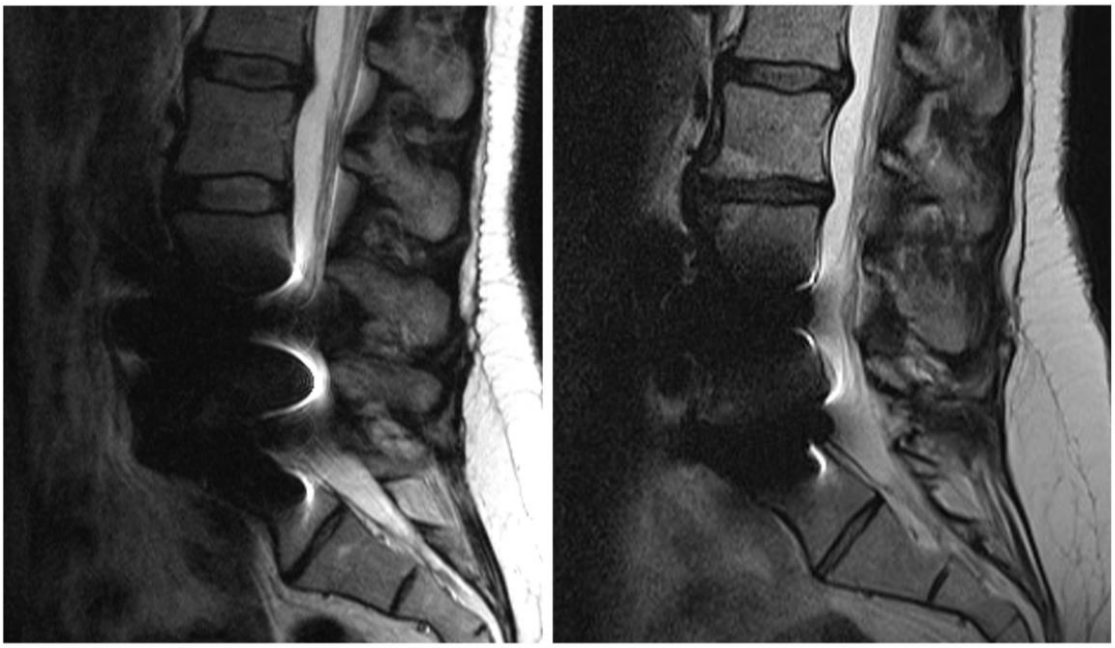
The patients were asked to quantify their analgesic consumption (type and amount of medication) during any day of a typical week.

4.5.3 Secondary radiological outcome measures

Secondary radiological outcome measures included Modic changes, the height and contour of the disc, the size of the disc herniation, the signal intensity of the nucleus pulposus and the presence of posterior HIZ. MRI variables for evaluating ADD were analysed at the nearest level above the implanted or degenerated index level, i.e. at L3/L4 or L4/L5. MRI examinations (> 95 % on 1.5-T magnets) before treatment and at eight-year follow-up included sagittal T2-weighted fast spin echo and/or DRIVE images (fast spin echo with 90° flip-back pulse), sagittal T1-weighted spin echo or fast fluid-attenuated inversion-recovery images and axial images of the lower lumbar levels (T2-, T1-, or proton density-weighted). Metal artefact reducing techniques [244] were used on 97 % of follow-up MRIs (Figure 9). Pre-treatment and follow-up images were anonymized, presented together in random order, and evaluated independently by two radiologists from different institutions who had more than 15 years of experience in spine imaging. The observers could not be blinded to the type of treatment, but were blinded to all clinical data. They rated changes in MRI findings by comparing eight-year follow-up and pre-treatment images on a clinical picture archiving and

communication system (PACS) unit. Conclusive ADD ratings were based on both observers' independent ratings and consensus between them in all cases of disagreement (review and discussion of images and findings). Interobserver agreements for each outcome measure are presented in Table 5.

Figure 9. The extent of metal artefacts on MRI at two-year follow-up (conventional MRI technique) and at eight-year follow-up (metal artefact reducing MRI technique).



Two-year follow-up Eight-year follow-up

The metal artefact reducing MRI technique used at eight-year follow-up yields a reduction in the extent of metal artefacts and permits better evaluation of the anatomy dorsally (the spinal canal), cranially (L4) and caudally (S1) compared to the conventional MRI technique used at two-year follow-up.

Table 5. Agreement between two observers evaluating the development of adjacent disc degeneration from pre-treatment to eight-year follow-up

Variable	Rating categories analysed for agreement	% 'yes'*	Kappa#	95 % CI
Modic changes	Present at eight-year follow up (yes/no)	14	0.70	0.52-0.88
	Developed or larger since pre-treatment (yes/no)	12	0.70	0.50-0.89
Disc height	≥ 2 mm lower since pre-treatment (yes/no)	10	0.60	0.38-0.81
Disc contour	Contour at follow-up (normal / bulge / herniation)		0.45	0.31-0.59
	Increased rating since pre-treatment (yes/no)	21	0.37	0.18-0.55
Disc herniation	Present at follow up (yes/no)	26	0.58	0.43-0.73
	Developed or larger since pre-treatment (yes/no)	10	0.62	0.39-0.85
Nucleus pulposus signal	Signal at follow-up (bright / grey / dark or black)		0.72	0.62-0.81
	Increased rating since pre-treatment (yes/no)	27	0.74	0.60-0.87
Posterior HIZ	Present at follow-up (yes/no)	10	0.78	0.59-0.96
	Developed since pre-treatment (yes/no)	7	0.76	0.54-0.99

CI=confidence interval, mm=millimetres, HIZ=high intensity zone

* Mean % of 'yes' reported by the two observers

Unweighted kappa for yes/no responses, otherwise linearly weighted kappa, calculated using WinPepi 11.60 (http://www.brixtonhealth.com/pepi4windows.html) and interpreted as poor (kappa ≤ 0.20), fair (0.21-0.40), moderate (0.41-0.60), good (0.61-0.80) and very good (0.81-1.00) agreement beyond chance [245].

4.5.3.1 Modic changes

Changes in the vertebral bone marrow adjacent to the end plate were classified as 0 (no change), 1 (type 1, hypointense T1-signal and hyperintense T2-signal), 2 (type 2, hyperintense T1-signal and iso- or hyperintense T2-signal) and 3 (type 3, hypointense T1-signal and hypointense T2-signal) [20]. The maximal craniocaudal (CC) extent of Modic changes was rated according to The Nordic Modic Consensus Group Classification [246] as 0 (no signal changes), 1 (located at the end plate only (minimal or small dots)), 2 (less than 25% of the vertebral body height), 3 (25 % to 50 % of the vertebral body height) or 4 (more than 50 % of the vertebral body height). We also assessed the maximal anteroposterior (AP) extent of Modic changes and rated the extent as 0 (no signal changes), 1 (less than 25 % of the AP diameter), 2 (25 % to 50 % of the AP diameter) or 3 (more than 50 % of the AP diameter). Intra- and interobserver agreement is reported as very good (kappa 0.8-1.0) [246].

4.5.3.2 Disc height

The height of the intervertebral disc was measured between the mid-inferior and the mid-superior disc borders on the mid-sagittal T2-weighted image [30]. At two-year follow-up, the smallest detectable change in disc height was calculated as 2 mm [36]. Thus, all changes < 2 mm from pre-treatment to follow-up were rated as 'no change'.

4.5.3.3 Disc contour

The shape of the intervertebral disc was rated as 0 (normal), 1 (bulging, > 1/4 of disc circumference) or 2 (herniated, including protrusion, extrusion, and sequestration) [4].

4.5.3.4 Disc herniation size

The size of the disc herniation was rated as 1 (< 1/3 of the spinal canal diameter), 2 (1/3-2/3 of the spinal canal diameter) or 3 (> 2/3 of the spinal canal diameter) [4].

4.5.3.5 Nucleus pulposus signal

The signal intensity of the nucleus pulposus was visually graded on sagittal T2-weighted images, using cerebrospinal fluid as intensity reference, as: 0 (bright), 1 (grey), 2 (dark) or 3 (black) [40].

4.5.3.6 Posterior high intensity zone (HIZ)

A posterior HIZ is an area of high-signal intensity in the posterior annulus fibrosus, brighter than or as bright as cerebrospinal fluid on sagittal T2-weighted images, and surrounded superiorly, inferiorly and anteriorly by the low-intensity signal of the annulus fibrosus [45]. The posterior HIZ was rated as 0 (not present) or 1 (present).

4.6 Exposure variables

In Paper II, we performed an analysis on the association between selected baseline characteristics and the treatment outcome. The selected baseline characteristics were tested for predictive value and included socio-demographic variables, clinical variables, psychological variables, pain and radiological variables.

4.6.1 Socio-demographic variables

All socio-demographic variables were patient reported. Patients were categorised as manual or non-manual workers [247]. In addition, we collected information on educational level [248], work status, duration of sick leave, smoking, gender and age.

4.6.2 Clinical variables

Clinical variables included prior discectomy, level(s) operated on with TDR, presence of comorbidity, ODI and body mass index (BMI). The variables were patient reported, except level(s) operated on, which was reported by the surgeon.

4.6.3 Psychological variables and pain

Psychological variables were Hopkins Symptom Check List (HSCL-25) [242], Fear-Avoidance Belief Questionnaire (FABQ) [249] and the Mental Component Scale (MCS) part of SF-36 [250]. Pain variables were LBP intensity (VAS), pain drawing categorised as pain below waist or pain above waist (with or without pain below waist) [251], duration of LBP and daily consumption of analgesics (yes/no).

4.6.4 Radiological variables

Pelvic incidence [252] was measured on radiographs obtained at the last follow-up by an experienced radiologist blinded to the clinical data, and was analysed as a baseline variable, since it describes the fixed relationship between the femoral heads and the end plate of the sacrum, which should remain unaltered after TDR. Pelvic incidence was dichotomised as $< / \geq 55$, as recommended by Prof. Le Huec (personal communication). All other radiological variables (Modic changes [20], disc height reduction [30], nucleus pulposus grade [40], facet arthropathy [253] and posterior HIZ [45]) were evaluated independently on pre-treatment images by three experienced radiologists blinded to the clinical data. The outcome was decided by simple majority, by mean value or by a fourth radiologist when majority or mean was unsuitable (Modic type) [254].

4.7 Ethical considerations

The first part of the study, in which the patients were included and treated, was approved by the Regional Committee for Medical Research Ethics in Eastern Norway (REK 1). It was conducted in accordance with the Helsinki Declaration and the ICH-GCP guidelines, and registered at www.clinicaltrial.gov under the identifier NCT00394732 before it commenced.

The eight-year follow-up was approved by the Norwegian Regional Ethical Committee South East C (2011/2177). The project was conducted in accordance with the Helsinki Declaration and the ICH-GCP guidelines and registered at www.clinicaltrial.gov under the identifier NCT01704677 before it commenced. Since this study does not evaluate drugs, a data monitoring committee is not mandatory according to Norwegian regulations, but it was overseen by a scientific board. A written and spoken informed consent was obtained from all patients before they were included in the long-term follow-up.

In Papers I and III, the results were reported according to the CONSORT standard for reporting randomised trials, while in Paper II, the results were reported according to the STROBE standard for reporting cohort studies.

4.8 Statistical analyses

The statistical analyses in Paper I were performed using SPSS version 22.0 (IBM SPSS Statistics for Windows, IBM corp., Armonk, NY), while the statistical analyses in Papers II and III were performed using SPSS version 24.0.

4.8.1 Power

The first phase of the trial was designed to have 80 % power in order to detect the significant difference ($p < 0.05$) of a change of at least 10 points in the mean ODI score between the intervention groups at two-year follow-up. Baseline standard deviation was estimated at 18 [237]. Adding 25 % for a multicentre study design and 30 % for possible dropouts, the plan was to include 180 patients.

4.8.2 Paper I

The main statistical analysis was in the intention-to-treat population at the eight-year follow-up. Missing values were replaced with multiple imputation. Patients who received treatment similar to the opposite treatment arm (crossovers) were kept in the group they were randomised to. We used $\chi 2$ test or Fisher's exact test to analyse categorical variables and an independent two-sided t test or analysis of variance to analyse continuous variables. A significance level of 5 % was used throughout. We did not adjust for significantly different baseline scores.

In a per-protocol analysis of the primary outcome variable (ODI) we excluded patients who did not receive the intervention they were randomised to, and those who had undergone back surgery or MDR after the study intervention was complete. Missing data were not replaced. In addition to calculating the mean change in each group, we also calculated the proportion of patients whose condition was classified as either improved or deteriorated. According to the FDA criteria, an individual improvement in ODI of at least 15 points can be considered a clinically important improvement [190 194], and this threshold value was used in Papers I and II. A decrease of six ODI points represented a 'change for the worse' [255] in Paper I. We calculated number needed to treat (NNT), which represents the number of patients treated with TDR instead of MDR needed to provide a clinically important improvement in one patient.

In an additional subanalysis of ODI we excluded patients who did not receive the intervention they were randomised to, and used the last value before crossover or reoperation in those who had undergone back surgery or MDR after the study intervention was complete. Other missing data were not replaced. We reported the number and types of reoperations and presented the survival until the first spinal operation after the end of the allocated treatment with a Kaplan-Meier plot.

4.8.3 Paper II

Continuous variables were described as medians and ranges, categorical variables as proportions and percentages. Outcome variables (clinical improvement (yes / no) and employment (yes / no)) were modelled as the dependent variables, and selected baseline covariates as the independent variables.

Some exposure variables were re-categorised. For example, patients were categorised as manual or non-manual workers according to the Norwegian Standard Classification of Socioeconomic Status [247]. The classification originally consisted of six groups, but since there were few patients in each group, they were dichotomised as manual or non-manual workers. The predictive value of a threshold level in baseline ODI of 55 points has been tested previously [256]. Since there were too few patients with ODI ≥ 55 points at baseline in the present sample, we chose to test a threshold level of 50 points. Work status was categorised as employed (part-time or full-time) or not employed.

Possible associations between selected variables and outcomes were modelled using binary logistic regression. Potential predictors that were highly associated with each other were excluded to avoid multicollinearity. Due to a limited sample size and the low number of patients who improved / were employed, we fit models with a maximum of four covariates to avoid overfitting. Therefore, only baseline characteristics that were statistically significantly ($p < 0.05$) associated with the outcome in univariate analyses were entered into the final multiple model. Further, the results from the multiple model were used to compute probabilities for the outcome given any selected value of the covariates, and the probabilities were expressed in a prediction matrix. The results were expressed as odds ratios (OR) with 95 % confidence intervals (CI). Since the sample size was limited, we were not able to set aside a test set for validation, and instead performed a leave-one-out cross-validation [257]. A sensitivity analysis was performed, excluding patients who were originally randomised to rehabilitation and patients who had received additional spinal surgery after the TDR. All tests were two-sided and p-values < 0.05 were considered statistically significant. Since our study was exploratory, no correction for multiple testing was performed.

4.8.4 Paper III

For each ADD variable, we compared changes in ratings between the treatment groups. We performed crude comparisons using a $\chi 2$ test or a Fisher's exact test for categorical variables and an independent two-sided t test for continuous variables.

In a sensitivity analysis, we compared the proportions of patients with overall increased ADD in each group after excluding patients randomised to rehabilitation who later received TDR, according to per-protocol principles.

In order to analyse possible associations between increased ADD and the clinical outcome adjusted for possible confounders, we fitted a multiple linear regression model. In this model, we excluded those who were not treated according to randomisation, since the treatments influence the clinical outcome. The model included ODI change from baseline to follow-up as the dependent variable and the following independent variables: Developed / increased extent of Modic changes (yes/no), disc height reduction (yes/no), disc contour worsening (yes/no), decreased nucleus pulposus signal (yes/no), developed HIZ (yes/no), age, gender and type of treatment (non-operative / TDR). Increased herniation size occurred in only one patient and was therefore not part of the regression model. The model fit was good, with normally distributed residuals. The results were presented as an estimate of beta with 95% confidence intervals (CI). We also performed a multiple logistic regression with the same independent variables as above and 'a satisfactory symptom state' (ODI ≤ 22 points at follow-up) (yes/no)

[258] as the dependent variable. A significance level of 5% was used for all analyses. All analyses were considered exploratory so no correction for multiple testing was done.

Post hoc, we analysed the severity of ADD development based on the number of worsened ADD variables. We compared the proportions of patients from each treatment group with increased rating for none, one, two, three, four, five and six ADD variables.

Table 6. Statistical methods used in Papers I-III.

Method	Purpose	Paper
$\chi 2$ test or Fisher's exact test	To analyse categorical variables.	I, III
Independent two-sided t test or analysis of variance	To analyse continuous variables.	I, III
Number needed to treat (NNT)	To calculate the number of patients treated with disc replacement instead of rehabilitation needed to provide a clinically important improvement in one patient.	I
Survival analysis (Kaplan-Meier plot)	To assess the survival without crossover or reoperation after rehabilitation or disc replacement, respectively.	I
Logistic regression	Determines the contribution of one (univariate) or several (multivariate) factors to a single binary outcome. Calculates odds ratio (OR).	II, III
Odds ratio (OR)	A measure of an effect size.	II
Linear regression	Determines the contribution of one (univariate) or several (multivariate) factors to a single continuous outcome.	III
Prediction matrix	To express the probabilities of job participation or of achieving a minimal clinically important long-term improvement.	II

4.9 Funding

The study was funded by Oslo University Hospital, South Eastern Norway Regional Health Authority, and EXTRA funds from the Norwegian Foundation for Health and Rehabilitation through the Norwegian Back Pain Association. The funders had no role in the study design, data collection, data analysis, data interpretation or writing of the reports.

5 Results

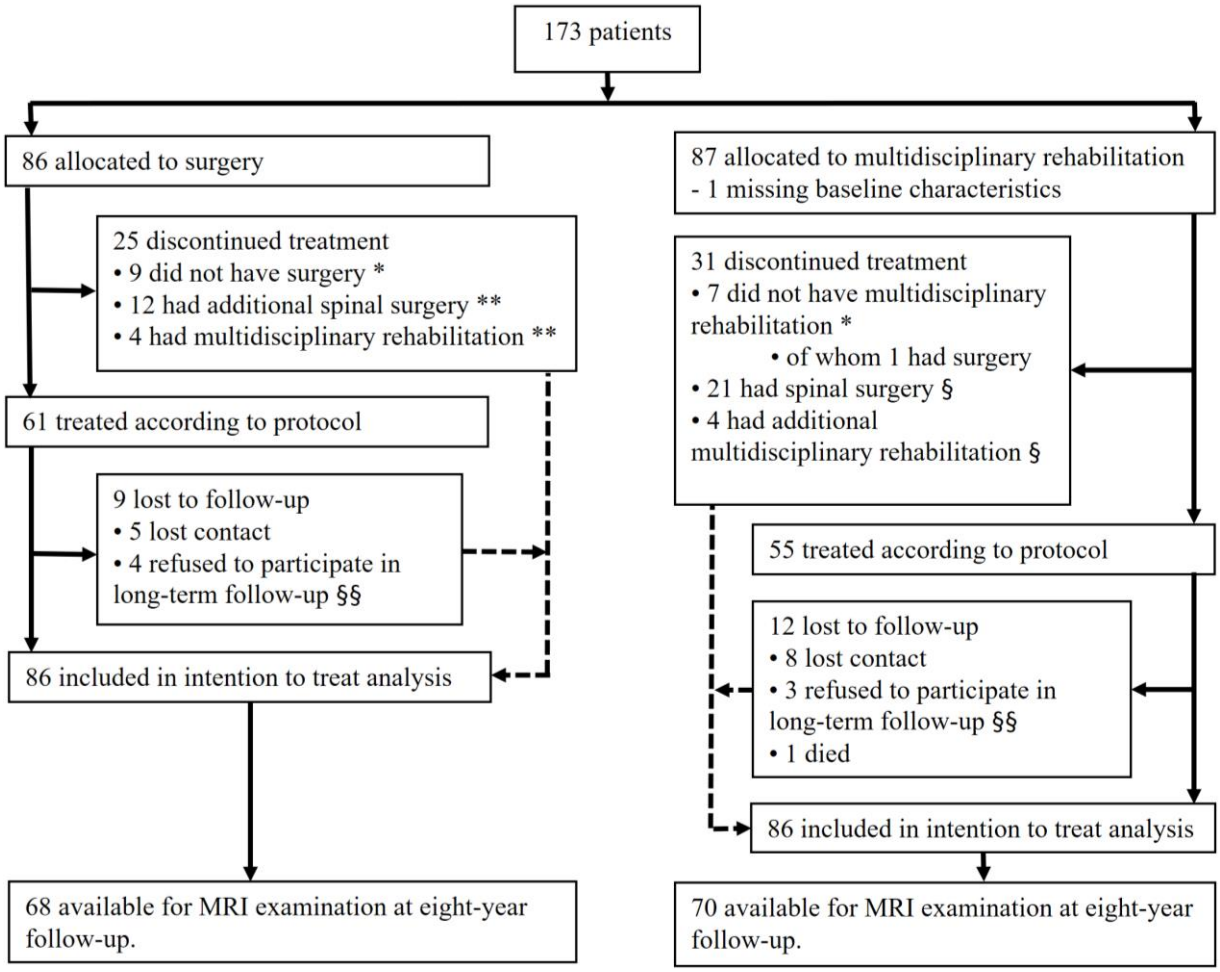
Of the 86 patients randomised to TDR, 12 patients (14 %) had additional spinal surgery (one reoperation because of implant dislocation (leading to a serious complication with a vascular injury and a leg amputation), six with decompression of spinal stenosis (one of whom had not received TDR), three with spinal fusion of the TDR level, one discectomy (had not received TDR) and one neurostimulator implantation). 77 patients (90 %) were available for eight-year follow-up, five did not respond to the follow-up invitation and four refused to participate in the long-term follow-up. 68 patients (79 %) were willing to undergo MRI examination at eight-year follow-up.

Of the 87 patients randomised to MDR, 21 (24 %) had crossed over and received spinal surgery since inclusion (14 with TDR, five with spinal fusion, and two with discectomy). 74 patients (85 %) were available for eight-year follow-up, eight did not respond to the follow-up invitation, three refused to participate in the long-term follow-up, and one had died due to cancer. 70 patients (80 %) were willing to undergo MRI examination at eight-year follow-up. For more details, see Figure 10.

After completing the allocated treatment, 25 patients (29 %) in the rehabilitation group and 35 patients (41 %) in the TDR group had received some non-operative treatment (e.g. physical, chiropractic, osteopathic or naprapathic treatment or acupuncture) and four patients (5 %) in each group had received structured multidisciplinary rehabilitation (p=0.24).

One serious complication was registered in the first two years [29]. During revision surgery for a dislocated polyethylene inlay three months postoperatively, an injury to the left common iliac artery led to a compartment syndrome and a subsequent lower leg amputation. Minor complications within two years are described by Hellum and colleagues [29].

Figure 10: Trial profile showing the treatment during eight years and follow-up of the participants of the study.



* For details, see Figure 7.

** 12 of 86 patients (14 %) had additional spinal surgery (one reoperation because of implant dislocation (leading to a serious complication with a vascular injury and a leg amputation), six with decompression of spinal stenosis (one of whom had not received TDR), three with spinal fusion of the TDR level, one discectomy (had not received TDR) and one neurostimulator implantation). Four patients received MDR.

§ 21 of 87 patients (24 %) crossed over and received spinal surgery after inclusion (14 with TDR, five with spinal fusion, and two with discectomy). Four patients received repeated MDR.

§§ Four patients randomised to surgery and three patients randomised to rehabilitation refused to participate in the long-term follow-up. Their reasons for not participating in the long-term follow-up were not explored.

5.1 Paper I

151 participants (87 %) were available for eight-years follow-up, 74 (85 %) in the rehabilitation group and 77 (90 %) in the surgery group. Mean improvement on the ODI from baseline to eight-year follow-up was 20.0 points (95 % CI 16.4-23.6, $p \leq 0.0001$) in the surgery group and 14.4 points (95 % CI 10.7-18.1, $p \leq 0.0001$) in the rehabilitation group. Mean difference in change from baseline to eight-year follow-up between the groups at eight-year follow-up was 6.1 points (95 % CI 1.2-11.0, p=0.02) in favour of surgery. Mean differences in favour of surgery on secondary outcomes were 9.9 points on VAS (95 % CI 0.6-19.2, p= 0.04) and 0.16 points on HSCL-25 (95 % CI 0.01-0.32, p=0.04). 18 patients (24 %) in the surgery group and four patients (6 %) in the rehabilitation group reported full recovery (p=0.002). There were no significant differences between the groups in terms of change from baseline to eight-year follow-up on the EQ-5D, or in occupational status, satisfaction with care or drug use (Table 7).

Table 7. Categorical secondary outcomes eight years after rehabilitation or disc replacement, with proportions and p-values.

	Rehabilitation		*Disc replacement*		*P-value*
	Number of patients	*Data*	*Number of patients*	*Data*	
Working or studying	73	29 (40 %)	77	37 (48 %)	0.33
Satisfaction with result of treatment*	73		76		
Full recovery		4 (6 %)		18 (24 %)	0.002
Much better		26 (36 %)		29 (38 %)	0.87
No / minimal change **		36 (49 %)		22 (29 %)	0.01
Much worse		5 (7 %)		1 (1 %)	0.11
Worse than ever		2 (3 %)		6 (8 %)	0.28
Satisfied with care ***	73	50 (69 %)	76	5 (66 %)	0.73
Daily analgesic medication	72	28 (39 %)	76	31 (41 %)	0.87
Daily opioid medication	72	15 (21 %)	76	15 (20 %)	1.00

* 7 point Likert scale.
** Including ‘slightly better’, ‘no change’ and ‘slightly worse’.
*** 5 point Likert scale, not including ‘slightly satisfied’ as satisfied with care.

In the per-protocol analysis, the mean difference between groups was 8.1 ODI points (95 % CI 2.3-13.9, p=0.01) in favour of surgery. 43 of 61 patients (70 %) in the surgery group and 26 of 52 patients (50 %) in the rehabilitation group had a clinically important improvement (15 ODI points or more) from baseline (p=0.03). The proportions of patients with a clinically important deterioration (six ODI-points or more) were not significantly different between the groups. Mean change in ODI was 13.9 (95 % CI 6.5-21.3) points for patients who crossed

over from rehabilitation to surgery and 5.4 (-4.8-15.6) points for patients randomised to surgery who underwent spinal reoperation.

5.2 Paper II

Of the 87 patients randomised to rehabilitation, one had missing baseline characteristics, seven did not receive rehabilitation, 21 had received spinal surgery before the eight-year follow-up, and 12 were lost to follow-up. Two participants randomised to surgery did not receive surgery, but completed the rehabilitation program, and were included in the rehabilitation group according to as-treated principles. Consequently, 55 patients were tentatively analysed in the rehabilitation group, but we did not detect any statistically significant associations between baseline characteristics and long-term outcomes after rehabilitation, possibly due to insufficient statistical power. Therefore, in Paper II we only evaluated potential predictors in patients treated with TDR.

Of the 86 patients randomised to surgery, nine did not receive the surgical treatment and nine were lost to follow-up (five lost contact, four withdrew consent). Hence, 71 patients were analysed eight years postoperatively. In addition, we included 14 patients randomised to rehabilitation who crossed over and were treated with TDR. Of these, 11 were available for follow-up (median time since surgery was 72 (range 41-88) months). Consequently, 82 patients (82 %) were included in the final cohort analyses. Nine of these 82 patients (11 %) had been reoperated, and median time since reoperation was 37 (range 1-103) months. Overall, 52 patients (63 %) achieved a clinically important improvement of ≥ 15 ODI points, and 42 patients (51 %) were employed eight years after they were included in the study.

Baseline variables significantly associated with the clinically important improvement were the presence of Modic changes (type 1 and/or 2) (OR 5.0, 95 % CI 1.4-18.2, p=0.01) and the extent of Modic changes (> 50 % of vertebral body height) (OR 3.8, 95 % CI 1.3-11.5, p=0.02). However, the presence and the extent of Modic changes were significantly associated with each other (p=0.01) and could not be included in the same model. Therefore, we did not proceed with the fitting of a prediction model.

Baseline variables significantly associated with the status of being employed at eight-year follow-up were < 12 months of sick leave before treatment (OR 4.1, 95 % CI 1.6-10.6, p=0.003), absence of comorbidity (OR 4.4, 95 % CI 1.4-13.8, p=0.01), ODI < 50 points (OR 3.6, 95 % CI 1.0-12.5) and high level of education (> nine years) (OR 3.6, 95 % CI 1.1-11.2, p=0.03). In addition, FABQ-work was statistically significantly associated with employment at eight-year follow-up (OR 0.9, 95 % CI 0.9-1.0, p=0.01). However, in the multivariate analysis with comorbidity, education level, ODI ≥ 50 and ≥ 12 months sick leave, including FABQ-work weakened the predictive power of the model, and we therefore did not include FABQ-work in the final multiple model. We found significant differences in the probabilities of being employed corresponding to different combinations of baseline variables. The probability of employment at the last follow-up was 1 % (95 % CI 0-4 %) for patients with ≥ 12 months sick leave, comorbidity, ODI ≥ 50 and ≤ nine years of education prior to treatment, and 87 % (95 % CI 80-94 %) for patients with < 12 months sick leave, no comorbidity, ODI < 50 and higher education (Figure 11) (Paper II, Figure 2). According to the classification table (confusion matrix), the model correctly classified whether patients were employed or not in

72.2 % of cases. The results were confirmed in a sensitivity analysis, in which patients who were reoperated or who had crossed over from the rehabilitation group were excluded, according to per-protocol analyses.

Figure 11. Prediction matrix.

		Low education		High education	
		Comorbidity	No comorbidity	Comorbidity	No comorbidity
≥ 12 months sick leave	ODI ≥ 50	**1 %** (0-4)	**9 %** (3-15)	**4 %** (0-8)	**24 %** (15-33)
	ODI < 50	**4 %** (0-8)	**25 %** (16-35)	**12 %** (5-19)	**52 %** (41-63)
< 12 months sick leave	ODI ≥ 50	**7 %** (2-13)	**38 %** (28-49)	**20 %** (12-29)	**67 %** (56-77)
	ODI < 50	**22 %** (13-31)	**68 %** (58-78)	**47 %** (36-58)	**87 %** (80-94)

Probability of working (95 % CI) at long-term follow-up after total disc replacement using a probability matrix model. Educational level (≤ 9 years or > 9 years, presence of comorbidity, duration of sick leave before treatment (< 12 months or ≥ 12 months) and Oswestry Disability Index (ODI, < 50 points or ≥ 50 points).

5.3 Paper III

Of 86 patients randomised to TDR, nine did not receive TDR, 14 were lost to follow-up (five lost contact and nine were not willing to undergo MRI examination), four could not be included in the TDR group as they had been operated with lumbar spinal fusion and one had no pre-treatment MRI. Therefore, 58 patients were analysed eight years postoperatively. In addition, 11 patients randomised to rehabilitation who crossed over and were treated with TDR (median time since surgery was 74 (range 61-85) months) were included in the analyses. Consequently, 69 patients treated with TDR were analysed.

Of 87 patients randomised to rehabilitation, 15 were lost to follow-up (eight lost contact and seven were not willing to undergo MRI examination). 19 could not be included in the non-operatively treated group as they had been treated with lumbar TDR (n=14) or spinal fusion (n=5) and one had no pre-treatment MRI. This resulted in 52 patients able to be included in the analysis after eight years. In addition, we included five patients randomised to surgery who were not operated. Thus, 57 patients treated non-operatively were analysed.

The two treatment groups had similar pre-treatment clinical, demographical and radiological characteristics. Furthermore, the 126 patients included in the analyses had similar pre-treatment clinical, demographical and radiological characteristics, as well as similar outcome measures at eight-year follow-up, compared to the 47 patients who could not be included.

At eight-year follow-up, 23 patients (40 %) in the non-operative group and 29 patients (42 %) in the TDR group had increased ADD (p=0.86). Three patients (5 %) treated non-operatively versus two patients (3 %) treated with TDR had decreased ADD (p=0.66). Regression of ADD was due to disappearance of HIZ in three patients, disappearance of Modic changes in one patient and regression of disc herniation in one patient. The change in rating from pre-treatment to eight-year follow-up did not differ significantly between the treatment groups for any of the ADD variables. The sensitivity analysis, in which we excluded the 11 patients randomised to rehabilitation and treated with TDR, showed similar proportions of patients with increased ADD (23 patients (40 %) in the non-operative group and 24 patients (41 %) in the TDR group (p=0.89)). In the multiple linear regression analysis of the association between increased ADD and the clinical outcome, the only variable that was significantly associated with change in ODI at follow-up was the type of treatment (non-operative or TDR) (B=7.2, 95 % CI 0.5-13.8, p=0.04). Therefore, we analysed the treatment groups separately. However, we did not find any significant association between increased rating in any ADD variable and change in ODI (R^2=0.06 (p=0.85) in patients treated non-operatively and R^2=0.10 (p=0.50) in patients treated with TDR). Multiple logistic regression analysis did not reveal any association between the increase in any ADD variable and ODI ≤ 22 points. The two treatment groups did not differ significantly in the proportions of patients with increased rating values for one, two, three or four ADD variables (p=0.38).

6 Discussion

In this first randomised trial comparing TDR with non-operative treatment, we found a significant long-term improvement in physical function and pain relief after both TDR and MDR. TDR was significantly more effective than MDR, but the clinical significance of the difference will be discussed below, since it was smaller than the prespecified clinically important difference of 10 points on ODI. We also found a positive association between Modic changes and a clinically important individual improvement (≥ 15 points on ODI). Shorter duration of sick leave, absence of comorbidity, lower ODI score and higher education at baseline increased the probability of employment at eight-year follow-up in patients treated with TDR. Furthermore, we found no difference in ADD development between the treatment groups, and the ADD development was not related to the clinical outcome (ODI change).

6.1 Methodological considerations

The internal validity of our study is defined by its ability to establish a causal conclusion about the relationship between the treatments and their clinical results. The main threats to the internal validity of the study are systematic errors (bias). Examples of systematic errors are selection bias, performance bias, attrition bias, detection bias and reporting bias [259] (Table 8). Further, improper use of outcome measures and statistical methods may cause systematic errors.

Table 8. Sources of systematic errors and their potential impact on the study.

Sources of systematic errors *	**Potential impact on the Norwegian TDR Study**
Selection bias	The randomised design with concealed treatment allocation should protect the study from major selection bias.
Performance bias	Neither patients nor care providers could be blinded to the treatment, which allows for different placebo effects in the treatment groups. However, the patients were told that neither of the treatments were documented as superior to the other. Co-interventions were not similar in the two groups. 24 % of those randomised to rehabilitation received surgical treatment, while additional non-operative treatment was received by 29 % of those randomised to rehabilitation and 41 % of those randomised to surgery. Effective co-intervention may lead to an overestimated effect of the allocated treatment.
Attrition bias	Dissimilarities between the groups may occur when subgroups of participants with divergent baseline characteristics or outcomes are lost to follow-up. The follow-up rates were good in both groups (85 % in the rehabilitation group and 90 % in the surgery group), which reduces the risk of attrition bias. As well, we found no significant differences in baseline characteristics among those who participated and those who did not participate in the long-term follow-up. Attrition bias may also occur due to crossovers, but the intention-to-treat analyses in Paper I protect against such bias. The as-treated analyses in Papers II and III allow for such bias, but we found similar results in the sensitivity analyses (per-protocol) and in the main analyses in Papers II and III.
Detection bias	Detection bias may occur when the outcome assessor is not blinded to the treatment. This should not be a problem in Paper I, as we used patient reported outcomes. In Papers II and III, we also used radiological outcomes, and the radiologists were not blinded to the treatment. However, they were blinded to the clinical outcomes, and had no conflicts of interest. Detection bias may also occur when the timing of outcome assessment is not similar in both groups. The outcomes were assessed eight years after treatment start in both groups, but those who had crossed over from rehabilitation to surgery had a shorter follow-up time (median six years follow-up time) compared to those treated surgically according to randomisation.
Reporting bias	Reporting bias is the result of selective reporting of outcomes. All outcomes in our study were pre-specified, which reduces the risk of reporting bias. Further, none of the authors had any conflicts of interest.

* Furlan et al. [259].

6.2 Study design

Randomised controlled trials (RCTs) are considered the 'gold standard' for testing the effects of one treatment against another. The randomised design protects the internal validity of the study by minimizing bias in treatment assignment (selection bias) [260]. An RCT may also provide information of causality, while observational studies provide associations. Furthermore, they allow active treatment to be compared to placebo, and, if the patients and investigators are properly blinded, the true treatment effect may be measured. Several treatments may be chosen as a control group for a study on the effect of TDR. Since no studies have compared TDR with sham surgery, the contribution of the placebo effect in the effect of TDR has not been investigated. The placebo effect in surgery should not be underestimated, and previous studies have found a significant placebo effect of spine surgery [261 262]. Neither have any studies compared TDR with 'usual care'. In a prospective cohort study, Mirza et al. [263] reported better results of spinal fusion compared to unstructured care, while no obvious benefit of fusion was observed when compared to multidisciplinary rehabilitation [263]. The relative efficacy of TDR compared to spinal fusion has been tested in several studies [156 206], and the Norwegian TDR Study is the first to compare TDR with non-surgical treatment. The design of the Norwegian TDR Study has been controversial. After the review process of our Paper II, the European Spine Journal published a comment by professor Mayer [264], who described the objectives of the present study as 'simply wrong', since TDR, in his opinion, should be considered as an alternative to spinal fusion and not to rehabilitation. Our attempts to officially defend the use of RCTs to compare surgical and non-surgical treatments have been stopped by the Editor-in-Chief. Although the results of MDR and TDR are compared 'head-to-head' in the present study, they should not be considered as competing treatments. Alternative study designs could include a randomised comparison of MDR versus MDR combined with TDR, or a pragmatic design in which all patients were offered MDR, and TDR could be offered to those with an insufficient treatment effect of MDR. 'Failed non-operative treatment' is often referred to as an indication for TDR [264], but several types of non-operative treatment exist, and patients receiving MDR are likely to experience less pain and disability than those receiving usual care or only physical treatment, according to a Cochrane review [152]. Therefore, the term 'failed non-operative treatment' must be considered as imprecise, as it does not always mean that modern non-operative treatment with a documented effect on LBP has been provided.

In our Paper I, the patients were analysed according to randomisation (i.e. intention-to-treat), thus preserving the advantages of an RCT in terms of protection against selection bias [260]. Our Paper II was conducted as a prospective cohort study, excluding patients who had not received TDR or who had been reoperated, and including patients who had crossed over from rehabilitation and received TDR (i.e. as-treated). The randomised design was broken, since the aim of the study was to identify patient characteristics that could predict favourable long-term treatment outcomes of TDR. Due to loss of statistical power in the rehabilitation group, the analyses were only performed on patients treated with TDR. In Paper III, we also chose to use an as-treated analysis as the main analysis, breaking the randomisation and allowing for some selection bias. However, the analysed treatment groups had similar pre-treatment clinical, demographical and radiological characteristics. Since 24 % of the patients

randomised to rehabilitation were treated surgically, and 10 % of those randomised to TDR were not operated [265], we considered an intention-to-treat analysis to be unsuitable for analysing the influence of TDR on ADD.

6.3 Patients

Systematic errors may arise from dissimilarities between the treatment groups [260]. Our concealed randomisation process could not be manipulated by investigators or patients, and the goal of the randomisation was to eliminate the possibility of selection bias. However, systematic withdrawal from one treatment group could allow for some systematic error. Attrition bias is the term used to describe bias caused by attrition (i.e. loss of participants) [259 266 267], and may occur due to dropout, crossover, withdrawal or nonresponse.

6.3.1 Protocol deviation

Although all patients were informed that neither of the treatment methods was documented as being superior to the other, they were recruited as candidates for disc replacement, and some might have participated in the trial with the hope of receiving surgery. The number of patients who did not complete the treatment they were randomised to was similar in the two groups, but we did not assess patients' treatment expectations before randomisation. After eight years, there was a relatively high crossover rate, especially from rehabilitation to surgery (24 %), and this allows for some systematic error. However, the crossover rate is not higher than in other studies with long-term follow-up comparing spine surgery with non-operative treatment [160 268 269]. The crossovers probably have only a small impact on the result in the intention-to-treat analysis, since a similar change in ODI among patients randomised to rehabilitation was found in the per-protocol analysis (14.1 points) and in the intention-to-treat analysis (14.4 points).

A substantial number of patients treated with TDR were reoperated after eight years (14 %), but the reoperation rate was in line with other studies of TDR with long-term follow-up (Appendix, Tables 1 and 2). A high reoperation rate makes it difficult to untangle the results of TDR from the results of the reoperation in the intention-to-treat analysis. In the per-protocol analysis, reoperated patients were excluded, thus probably removing the most inferior results of surgery from this analysis. Therefore, we added a sensitivity analysis in which values before crossover or reoperation were carried forward. This analysis may reflect the true results of the treatment the patients were randomised to, but carrying short-term results forward also represents a limitation.

We have not detected any systematic differences in baseline characteristics between those randomised to rehabilitation who crossed over to surgery and those who did not. Furthermore, the long-term results among those who crossed over were similar to those of the participants who did not cross over. Neither did we detect any systematic differences in the baseline characteristics between those randomised to surgery who were reoperated and those who were not. However, the long-term results among those who were reoperated were inferior to those who were not. Inferior results in patients reoperated after TDR are also reported in several other studies [192 195 196 270].

In Papers II and III, we included patients who crossed over from rehabilitation to TDR in the main analyses. These patients had an observation time shorter than eight years after TDR, which may allow for detection bias [259], and thereby affect the associations between baseline characteristics and outcomes (Paper II) and the proportions with increased ADD (Paper III). However, the results of the sensitivity analyses, in which these patients were excluded, were similar to those of the main analyses in both papers.

Systematic errors may also occur if the baseline characteristics or the outcomes are different in those who respond to follow-up and those who do not respond [271]. However, in a Norwegian Spine Registry study, there were no differences in the outcomes between those who responded and those who did not respond to follow-up invitation [272].

Nevertheless, a substantial follow-up rate is of major importance for the internal validity of the study. A general threshold for an acceptable follow-up rate can hardly be established [273], but, according to Furlan et al. [259], the proportion of dropouts and withdrawals should not exceed 20 % for short-term follow-up or 30 % for long-term follow-up. Studies with long-term follow-up are generally at higher risk of loss to follow-up [274]. The follow-up rate in Paper I (questionnaire) was 87 %, which we consider to be satisfactory. In Paper II, the follow-up rate was 82 %. The analyses were performed in the per-protocol population, which made it necessary to exclude some of the patients who were included in Paper I. In Paper III (radiological outcomes), the follow-up rate was 73 %. The main reason for loss to follow-up in Paper III was unwillingness to undergo MRI at follow-up. In addition, we could not include patients who had been operated with fusion, which further reduced the proportion of patients available for follow-up. For all three papers, we found no differences in baseline characteristics among patients who participated in the long-term follow-up and those who did not. In Paper III, we could not find any difference in clinical long-term outcomes among those who participated and those who did not.

6.4 Primary outcome measures

Even if an RCT is protected against selection bias, several other threats to the internal validity of the study exist. Detection bias may occur when the investigator influences the outcome measure [259 275 276]. To avoid this, the outcome measures in Papers I and II were patient reported and collected by a person not involved in the study. In addition, the financing of the study was public, and none of the authors of Papers I, II or III had any conflicts of interest. Reporting bias occurs when the reported outcomes are selected based on the results [259 277 278]. To avoid reporting bias, all outcome measures should be defined in the protocol before the study begins. The outcome measures in our study were registered at www.clinicaltrial.gov under the identifier NCT01704677 before the study commenced.

6.4.1 Oswestry Disability Index (ODI)

ODI is an internationally recommended outcome measure for physical function [233-235], and was used as the primary outcome in Papers I and II, and as a secondary outcome in Paper III. The choice of ten ODI points as a cutoff value for a clinically important difference between treatment groups was based on recommendations at the time the Norwegian TDR Study was designed [255]. Different threshold values were used in other studies at that time, for example, Fairbank and colleagues chose a cutoff value of four points [159]. There is no

consensus on the size of a minimal clinically important difference (MCID) in ODI when comparing one treatment group to another [279], but estimates range from 4 to 17 points [280]. According to Glassman and colleagues [281], the threshold value of ten ODI points indicates a clinically important improvement in an individual, and should not be misinterpreted as a measure of a difference between groups. The authors advocate reporting the proportion of patients achieving a minimal or substantial clinical difference in each group, rather than reporting mean group differences. Therefore, we did an additional analysis and calculated the proportion of patients with a clinically important improvement, which in FDA studies is defined as a minimum of 15 points improvement on ODI [190 194]. An individual minimal clinically important improvement is also commonly defined as a 30 % improvement on ODI [156 243], but we chose to use the FDA criteria as they had been used in an earlier report from the present trial [29]. At two-year follow-up in the present study, the individual minimal clinical improvement was calculated as 12.88 ODI points based on Receiver Operator Curve (ROC) analysis [282]. Van Hooff et al. [258] have advocated reporting the proportion of patients achieving 'a satisfactory symptom state' (ODI ≤ 22 points at follow-up) rather than reporting proportions of patients with a MCID.

In Paper II, we used a MCID of 15 points to define patients with a favourable clinical outcome (main outcome), and, in Paper III, we used linear regression with ODI as the dependent variable to test the association between increased ADD and the clinical outcome.

6.4.2 Adjacent disc degeneration (ADD)

The concern about increased ADD has been a major motivation for the development and implementation of TDR. Increased ADD was used as the primary outcome in Paper III, and was based on the increased rating in at least one of six ADD variables measured on MRI. The choice of the primary outcome variable may be debated, and we studied a broad range of separate ADD variables. There are several ways to describe ADD, and no gold standard exists. The commonly used Pfirrmann system [32] provides a single rating of disc degeneration based on the height, structure and signal of the disc, and the distinction of nucleus and annulus. This system does not separate disc signal from disc height, and it does not include disc contour/herniation or HIZ – nor Modic changes, which were related to clinical outcome after TDR in our cohort in both the short- [256] and long-term [283]. Increased ADD (yes/no) was a mainly qualitative variable. Only one of the six underlying ADD variables (disc height) was based on an actual measurement. Our study still provides more information than studies restricted to disc height measurement alone. By defining increase in ADD as a dichotomized variable, we did not use information on the degree of the increase. It is unclear how a variable reflecting the overall degree of increase in ADD can be constructed and weighted based on increases in different underlying variables (disc signal, disc contour, HIZ, etc.). However, we compared the number of increased ADD variables between groups in a post hoc analysis. As at 2-year follow-up [36], we accepted an increase in only one ADD variable as indicating increased ADD, and were thus able to detect even small increases and differences in ADD.

6.5 Secondary outcome measures

6.5.1 Low back pain (LBP)

The intensity of LBP was measured with VAS. Pain intensity can also be measured with a numeric rating scale (NRS), which is reported to have superior responsiveness compared to VAS [284]. However, we chose to use VAS since it was used to measure pain intensity in the early phase of the study [29]. Ostelo et al. [285] suggested 15 points as a MCID in VAS for LBP.

6.5.2 Health related quality of life

Health related quality of life was measured with EuroQol (EQ-5D). EQ-5D is documented as a valid, reliable and responsive instrument in a Norwegian population treated surgically for degenerative disorders in the lumbar spine [286]. A five-level EQ-5D (EQ-5D-5L) was introduced in 2009 in order to improve the sensitivity of the instrument and to reduce ceiling effects, as compared to the three-level EQ-5D (EQ-5D-3L). Our questionnaire included both versions. EQ-5D-5L data were collected for use in a future methodological study comparing the three-level and five-level version. In our Papers I and II, we have only reported the results from EQ-5D-3L since it was the version used in the early phase of the study [29 256].

6.5.3 Psychological distress

Psychological distress was assessed with a Hopkins Symptom Check-List (HSCL-25). The total score is highly correlated with severe emotional distress of unspecified diagnoses, and its reliability and validity is documented [242]. HSCL-25 was considered a valuable tool at the time when the study started, but was not included in a list of instruments more recently recommended for the assessment of psychological distress [234 235].

6.5.4 Work participation

Self-reported work status was a secondary outcome measure in Papers I and II. In a Norwegian study of patients with chronic LBP [287], self-reported work status was a reliable tool for assessing whether patients were working or not. However, in an other Norwegian study [288], poor agreement was found between self-reported duration of sick-leave and data from a public registry. Since the participants could have several reasons for having a different kind of work after eight years, we chose to report work participation rate rather than the proportion who had returned to their previous work. As reported by Hellum et al., the work participation rate before treatment was similar in the rehabilitation group (26 %) and the TDR group (28 %).

6.5.5 Patient satisfaction

Satisfaction with the result of the treatment was assessed with a 7-point Likert scale. Since there were no differences between the groups in terms of the proportion of patients reporting ‘slightly better’, ‘no change’ or ‘slightly worse’, these categories were merged into one and reported as ‘no/minimal change’.

6.5.6 Additional treatment including reoperations

The patients were asked if they had received any additional non-operative treatment or had any back surgery after the allocated treatment was completed. In the main analysis, we registered all spinal surgical procedures in those randomised to TDR as ‘reoperations’. It is

worth noting that two of the 12 patients registered as ‘reoperated’ had initially refused the TDR, and were in fact not reoperated, but had received primary spinal surgery. On the other hand, two of the 14 patients who were randomised to MDR and crossed over to TDR, were reoperated later (one with spinal fusion and one with decompression at the TDR level). The reoperation rate is an important outcome measure, since several studies report that patients who have been reoperated have worse results than those who have not been reoperated [192 195 196 270]. Moreover, anterior revision procedures are technically difficult and involve a high risk of vascular injury [289].

6.5.7 Complications

Information about complications related to the treatments was collected at two-year follow-up, and has been reported in detail [29]. Later complications may be hard to detect, but clinically relevant late complications should be reflected in the long-term PROMs and the reoperation rate. We have obtained images for the detection of radiological changes such as ADD (Paper III) and facet joint degeneration (not published). Heterotopic ossification and reduced ROM may also represent late complications. We have obtained radiographs with flexion and extension for the evaluation of long-term ROM, but these data have not yet been analysed.

6.5.8 Radiological outcome measures

In Paper III, we analysed a broad range of radiological outcome measures with MRI. The strengths of this evaluation include the use of a metal artefact reducing MRI protocol, the experienced radiologists blinded to the clinical outcome who independently performed the MRI evaluation and the direct comparison of post- and pre-treatment MRIs to assess changes in ADD. Such comparison can reduce overrating of changes due to ambiguous findings or small differences in MRI techniques, and can also improve agreement on changes in ratings [254]. In this study, the agreement was mostly good, despite instances where a low prevalence of change tended to reduce many of the kappa values.

On the other hand, observer bias may have occurred since the radiologists could not be blinded to the treatment group. They were not blinded to post-treatment images when assessing pre-treatment images, and this may have influenced their pre-treatment ratings. Disc prostheses leave metal artefacts on the images close to the implant, but, according to previous reports [290-292], such metal artefacts barely affect the evaluation of the adjacent disc level. Moreover, the metal artefact reducing MRI protocol further reduced the extent of the artefacts (Figure 9).

Although MRI remains the gold standard for imaging of the degenerated disc, some radiological findings only occur in standing position, and MRI is usually performed in supine position [293], including in the papers that the present thesis in based upon. Tarantino et al. [294] demonstrated that one third of patients with normal supine MRI had detectable IDD in upright MRI.

6.5.9 Daily use of analgesics

The patients were asked to quantify their analgesic consumption during any day of a typical week. The reported outcome in Paper I included any kind of analgesic medication, but the

proportion who uses opioids may also be used as an outcome measure. Despite obvious side effects, opioids have become the most commonly prescribed drug class for back pain in the US [295]. Smith et al. [296] have recently reported that depressed patients with LBP are more likely to be prescribed opioids and receive higher dosages. Different explanations for this trend have been discussed, but pain expression and behaviour may affect opioid-prescribing patterns [297].

6.6 Exposure variables

In Paper II, we tested several baseline socio-demographical, clinical, psychological and radiological variables for their association with the clinical outcome and the patients' work participation. In the literature, several other patient characteristics are reported to be associated with these outcomes, as detailed in the discussion of the results. However, when testing a very large number of patient characteristics for predictive value, there is an increased risk of detecting significant associations by chance (type 1 error). Therefore, we analysed a selection of baseline variables for which an association with the outcome measures was plausible. In addition to the variables collected before treatment, we wished to test the predictive value of pelvic incidence [252], which has drawn special interest as a potential predictor for outcome after back surgery. Pelvic incidence was measured on radiographs obtained at eight-year follow-up, but was analysed as a baseline variable since it describes the fixed relationship between the femoral heads and the end plate of the sacrum, and should remain unaltered after TDR.

6.7 Statistics

6.7.1 Sample size

The trial was powered to detect a difference of ten ODI points between the treatment groups, and 25 % was added for a multicentre study design and 30 % for possible dropouts. Since the dropout rate was less than 30 % after eight years, the power of the study of the clinical effect (Paper I) was considered satisfactory. In Paper II, we used regression analyses to detect baseline characteristics associated with a favourable long-term outcome after TDR, and for this purpose the sample size was limited. A larger simple size would have allowed us to fit a larger prediction model, perform a validation and possibly identify further variables associated with the outcome. We had planned to analyse the association between baseline characteristics and a favourable outcome in the rehabilitation group as well, but the large number of crossovers left only 55 patients available for the per-protocol analysis. The statistical power was therefore considered insufficient for this purpose. For the analysis of development of ADD in Paper III, the sample size was also limited. Thus, we could not analyse subgroups of patients, and statistical significance was hard to achieve for each specific ADD-variable.

6.7.2 Statistical methods

In Paper I, we had planned to perform mixed model analyses of the long-term clinical results, in addition to ANOVA. This was also registered at www.clinicaltrials.gov, but, after a thorough discussion within the research group, we chose to focus on the long-term results rather than on the entire eight-year clinical course. Therefore, we replaced missing values

with multiple imputation, performed the ANOVA analysis, and did not perform mixed model analyses.

In Paper II, the use of cut-off values for the independent variables may be questioned. In order to create a prediction matrix that could help clinicians and patients choose the right treatment for chronic LBP, the independent variables had to be dichotomised. Due to the limited sample size, the cut-off values were not only based on clinical recommendations, but also on statistical properties that gave the best separation among subgroups of patients. The associations might have been weakened if we had used other cut-off values for the independent variables. Further, according to the Hosmer and Lemeshow [298], any variable whose univariate test has a $p < 0.25$ is a candidate for the multiple model. However, due to the limited sample size and the low number of patients who improved / were employed, we fit models with a maximum of four covariates to avoid overfitting. Therefore, only baseline characteristics that were statistically significantly ($p < 0.05$) associated with the outcome in univariate analyses were entered into our final multiple model. In the univariate analysis, we found a statistically significant association between baseline FABQ-work and employment at eight-year follow-up, but we did not include FABQ-work in the final multiple model since this weakened the predictive power of the model.

6.8 Interventions

The study was designed to compare TDR with the best-known rehabilitation program. In a Cochrane review [152], patients treated with MDR experienced less pain and disability than patients who received 'usual care' or physical treatment. A larger treatment effect of TDR would therefore be expected if TDR was compared to a less comprehensive rehabilitation program or 'usual care'.

Different prosthesis designs have different properties and allow for different ROM. To our knowledge, no disc prosthesis has been documented as superior to any other, but different clinical results may have been achieved if another prosthesis had been used. The Pro-Disc II is classified as semi-constrained [182]. Since the mobility in the treated level can affect the development of ADD [202 226], the development of ADD may also have been different if another prosthesis design had been used.

6.9 Main results

6.9.1 Primary outcome measures

6.9.1.1 Oswestry Disability Index (ODI)

The mean improvement in ODI from baseline to eight-year follow-up was 14.4 points in the rehabilitation group and 20.0 points in the TDR group, and both must be considered statistically and clinically significant in both groups, as they are larger than the MCID that was calculated at two-year follow-up [282]. The mean difference between the groups of 6.1 points in favour of surgery was statistically significant, but smaller than the prespecified 10-point difference that the study was designed to detect.

We found a clinically important improvement in a significantly larger proportion of patients in the TDR group (70 %) than in the MDR group (50 %) at eight-year follow-up. A minimum

improvement of 15 ODI points has previously been reported by 68-87 % of patients 5-8 years after TDR [190 192 194].

The mean 14.4 points improvement in the rehabilitation group was in line with the mean improvement of 12.6 points at 11-years follow-up after similar rehabilitation in patients with LBP and degenerative discs in three randomised trials reported by Mannion and colleagues [160]. The mean 20.0 ODI points improvement in the TDR group was comparable to the findings in previous studies [190 192-194 201]. The short-term (two years) mean improvement on ODI from baseline was 12.8 points in the rehabilitation group and 20.6 points in the surgery group [29], similar to the findings at eight-year follow-up, thus indicating a persistent long-term treatment effect for both rehabilitation and disc replacement.

6.9.1.2 Adjacent disc degeneration (ADD)

In Paper III, we found increased ADD in 40 % of the patients treated non-operatively and in 42 % of the patients treated with TDR, and the difference was not significant (p=0.86). The post hoc analysis in which we compared the number of increased ADD variables did not reveal any significant difference between the groups (Paper III, Figure 3).

In each treatment group, a larger proportion of patients had increased ADD at the eight-year follow-up than at the two-year follow-up (40% versus 19% after non-operative treatment and 42% versus 13% after TDR) [36].) A further increase is expected with longer follow-up.

Reduced ROM in the prosthesis has previously been reported to be associated with an increased prevalence of ADD [202 226]. We did not measure ROM at eight-year follow-up, but at two-year follow-up [212] segmental ROM was similar for an average disc prosthesis as for a degenerated index level disc. This may have contributed to similar ADD development in both treatment groups.

In a previous review of ADD following back surgery, Harrop et al. [167] found a large variation in reported ADD (0-24 % of the patients 3 to 17 years after TDR). However, the included studies were heterogeneous and had major limitations. The large variation in ADD most likely reflects differences in patient characteristics (e.g. age), follow-up time and ADD assessment methods. Such methods were either not reported or included disc height, osteophyte formation or instability on flexion-extension images. The variation in ADD assessment makes it difficult to compare the proportions of patients with ADD increase, including in comparisons between recent studies. Zigler et al. [224] based ADD on disc height reduction, end plate sclerosis, osteophytes and spondylolisthesis on radiographs. They found increased ADD in 9% of patients five years after TDR. We found a much higher proportion (42%) with increased ADD at eight-year follow-up. The difference may partly be due to our use of MRI to detect changes not visible on radiographs.

Regardless of allocation, 91 patients in our cohort have been treated with TDR [265], of whom one (1%) has been re-operated with fusion including the adjacent segment. This is in agreement with other studies reporting re-operation due to ADD in 0-1% of patients two to five years following TDR [188 190 299].

We found no association between increased ADD and the clinical outcome (ODI). This is in line with a previous report from Huang et al. [226] and with the results of a recent cross-sectional analysis of long-term follow-up data from four randomised trials comparing non-operative treatment with fusion for chronic LBP [34].

6.9.2 Secondary outcome measures

6.9.2.1 Low back pain (LBP)

We found a mean difference in pain intensity (VAS) between the treatment groups of 9.9 points in favour of TDR, which was statistically significant (p=0.04), but not clinically important according to Ostelo et al. [285].

6.9.2.2 Work participation

We found no significant difference between the groups in work participation, as 40 % of the patients in the MDR group and 48 % in the TDR group were still working or studying at eight-year follow-up (p=0.33). Sköld et al. [192] found that 78 % were working full- or part-time at five-year follow-up after TDR, as compared to 37 % before treatment. Siepe et al [193] reported that 67 % of their patients had some kind of work, while Guyer et al [190] found that 66 % were working full-time after five years, as compared to 51 % before treatment. David et al. [203] reported that 90 % of the patients who were working before the treatment returned to work at follow-up after mean 13 years, but their calculation did not include those who were not working before treatment. Lemaire et al. [201] found that 92 % of the patients who were 'eligible to return to work', were working at follow-up after mean 11 years, but did not include the patients who were retired in the calculations. The smaller long-term work participation rate following TDR in our study may be due to a smaller preoperative work participation rate, different calculation methods or the generous disability pensions in Norway [300].

6.9.2.3 Patient satisfaction

A notable difference between the groups was the proportions who reported a 'full recovery' (6 % in the rehabilitation group and 24 % in the TDR group). In comparison, Sköld et al. [192] reported that 38 % of the participants in their study were totally pain-free at five-year follow-up, measured on a five-point scale. We found no significant difference between the groups in the proportions of patients who classified themselves as 'worse than ever', but there was a trend towards more deterioration in the TDR group (8 %) than in the rehabilitation group (3 %) (p=0.28). Similar proportions of patients in the rehabilitation group (69 %) and the TDR group (66 %) reported that they were satisfied or slightly satisfied with care.

6.9.2.4 Additional treatment including reoperations

Large proportions of the patients in the rehabilitation group (47 %) and TDR group (57 %) had received non-operative treatment after completing the allocated treatment, and 4 patients in each group had received structured multidisciplinary rehabilitation. The reoperation rates reported in other studies with long-term follow-up after TDR range from 4 to 39 % (Appendix, Table 1). The great variation in reoperation rates may partly be due to different ways of reporting reoperations. Tropiano et al. [168] reported 5 % posterior fusions at mean nine-years follow-up after TDR, while other reoperations were not reported. Guyer et al. [190] reported 8 % 'device related' reoperations after five years. The highest reoperation rate

(39 %) was reported at mean 10-years follow-up after implantation of the Acroflex elastomeric core prosthesis [196], which was withdrawn after implantation in only 28 patients due to early mechanical failure. However, a reoperation rate of 33 % has been reported at mean 11-years follow-up after implantation of Prodisc II prostheses [195].

6.9.2.5 Complications

One serious complication occurred in an early reoperation due to implant failure, and a vascular injury led to a lower limb amputation. The complication rates reported in other studies with long-term follow-up after TDR range from 9 to 23 % (Appendix, Table 1), but late complications such as heterotopic ossification and facet joint degeneration may not have been among those reported. Heterotopic ossification has been commonly observed after implantation of earlier generations of disc prostheses [202], but Lu et al. [301] have recently reported varying grades of heterotopic ossification in 74 % of patients at mean 15-years follow-up after implantation of the Charité III prosthesis. Increased facet joint degeneration was observed in 34 % of patients treated with TDR and 4 % of patients treated non-operatively at two-year follow-up in the Norwegian TDR Study [36]. A more severe facet joint degeneration is expected to occur after TDR in the long-term.

6.9.2.6 Radiological outcome measures

In Paper III, there were no significant differences between the groups in any ADD variable, but there was a trend towards more disc height reduction in the superior adjacent disc in the TDR group than in the group treated non-operatively. However, the observed difference of 0.3 mm is much smaller than the minimal detectable change of 2 mm [36], and is thus not likely to be clinically relevant. In patients treated non-operatively or with TDR, similar proportions had increased ADD (40 % versus 42 %) and decreased ADD (5 % versus 3 %). Spontaneous disappearance of MC and regression of disc herniations have been described previously. Albert et al. [302] reported that 16 % of the patients with any type of MC had no detectable MC at 14-month follow-up. Yang et al. [303] have suggested that regression of disc herniations may be caused by disc dehydration, retraction of herniated disc material into the disc space or enzymatic catabolism and phagocytosis.

6.9.2.7 Daily use of analgesics

We found that approximately 40 % of the patients in each group used analgesic drugs daily, and there was no difference between the groups in analgesic consumption. We did not report the proportion of patients who used daily opioids in Paper I, but, post hoc, we found that approximately 20 % of the patients in each group reported daily opioid use (Table 7).

6.9.3 Exposure variables

In Paper II, preoperative MC were positively associated with a clinically significant improvement of ≥ 15 ODI points. The extent of MC (> 50 % of the vertebral body height) was significantly associated with both the presence of MC and the outcome (≥ 15 points improvement in ODI score). Therefore, the extent of MC may be just as important for the clinical outcome as the presence of MC. Our findings should be interpreted in light of the findings in a recent systematic review on the impact of MC on outcome after lumbar spine surgery [304]. This review identified four TDR studies (including the two-year results from the Norwegian TDR study [29]). One study found no association between MC and ODI or

LBP after TDR, and the remaining three had conflicting findings about which types of MC (type 1, type 2, or both types combined) were related to ODI or pain after TDR. Although MC may be associated with improved outcome after TDR, the association is not consistent between different studies or outcomes.

In the literature, several other patient characteristics may also influence the clinical outcome after TDR. Gornet et al. [305] found significantly less improvement in ODI score at two- and five-year follow-up after TDR in patients with workers' compensation. They also found a statistically significant association between a favourable outcome measured with ODI at five-year follow-up and higher grades of disc degeneration preoperatively, presence of Modic type 2 changes and a smaller proportion of the overall lumbar lordosis (L1-S1) at the treatment level. Park et al. [205] showed inferior long-term results of TDR in patients with spondylolisthesis, facet joint arthritis, lateral recess stenosis and patients treated with TDR at the adjacent level of a fused segment.

Shorter duration of sick leave, absence of comorbidity, lower ODI score and higher education at baseline increased the probability of employment at eight-year follow-up in our prediction matrix. We also found that a preoperative ODI score ≥ 50 points was associated with lower probability of work participation at eight-year follow-up.

Our findings are plausible, but in the literature there is no consensus on baseline characteristics that predict return to work after surgery in patients with chronic LBP. In agreement with our findings, Hellum et al. [256] reported that high preoperative ODI scores predicted similar results of MDR and TDR, strongly indicating that MDR should be the first choice of treatment for LBP – particularly in patients with high levels of disability. In populations including mostly non-operated patients with LBP or sciatica, Cougot et al. [306] found that the patient's profession was the only predictor for return to work in health care workers with LBP. In patients with sciatica, Grøvle et al. [307] found that lower age, better general health, less baseline sciatica bothersomeness, lower score on the FABQ-work and a negative straight leg raising test result were significantly associated with a higher probability of returning to work. McGirth et al. [308] found that preoperative depression, arthritis and prolonged preoperative opioid use reduced the likelihood of returning to work in patients labelled as having degenerative chronic LBP without workers' compensation. In a longitudinal study of women, Nordeman et al. [309] found that the six-minute walk test, depression and earlier ability to work predicted the ability to work at two-year follow-up. Hence, the biopsychosocial factors at baseline associated with employment at follow-up in our study find broad support in the literature.

Pelvic incidence had no predictive value in our study. Salzmann et al. [157] suggested that patients with pelvic incidence of more than 65º are prone to developing facet joint degeneration after TDR, and considered such patients as bad candidates for TDR. Arunakul et al. [310] found no association between radiological features such as pelvic incidence and sacral slope, and clinical outcomes measured by back pain (VAS) and function (ODI) after TDR. In contrast, Laouissat et al. [311] reported inferior results after TDR in patients with sagittal curvature of Roussouly type 4 [312].

6.10 External validity

The multicentre design of this study improves the external validity of its results. However, the inclusion of only 179 of the 605 patients (30 %) screened for eligibility may call the external validity of the study into question. Our findings may not apply to the general population with chronic LBP. The most important exclusion criteria were nerve root involvement or presence of generalised disc degeneration. On the other hand, TDR is only indicated in selected patients, and we believe that the participants of this study are representative as candidates for TDR. Spine registry studies or other observational studies may be used to test whether the results of RCTs are valid for a larger group of patients receiving the same treatment, since observational studies generally include less strictly selected patients [156]. In a recent international spine registry study [313], the results of cervical disc replacement were in accordance with those in the published RCTs. Furthermore, in a Swiss registry study with five-year follow-up after lumbar TDR [197], the main results are similar to those in RCTs with a follow-up time of at least five years (Appendix, Table 1). Our results are in line with two small single centre Norwegian observational studies [314 315], while, in the Norwegian Spine Registry (2007-2013), the one-year mean improvement following TDR was 26.0 ODI points as compared to 22.5 ODI points after one year in the Norwegian TDR Study [316].

6.11 The role of placebo

The placebo effect results from the patients' expectations of a treatment effect [317], and the first attempts to quantify the placebo effect were made in the mid-19th century [318]. In the field of spine surgery, a considerable placebo effect has been reported for vertebroplasty compared to sham surgery [261]. On the other hand, Freeman and colleagues did not find any significant clinical response from intradiscal electrothermal therapy or sham surgery among patients with chronic LBP [319]. Since no studies comparing TDR with sham surgery have been published, the placebo effect of TDR cannot be quantified.

Obviously, the patients in our study could not be blinded to treatment, which is described as a potential source of performance bias [259], and may have led to a difference in placebo effect between the groups. However, all eligible patients were given balanced information on the treatments before allocation, and neither treatment was presented as superior to the other. Similar proportions of patients from both treatment groups refused the allocated treatment, possibly indicating that the patients' expectations of treatment effect were similar in both groups.

6.12 The role of the natural course of LBP

When evaluating the long-term clinical effect of treatments for LBP, the natural course of the condition should be considered. Since many patients with LBP seek different treatment methods [320], the 'no treatment' option may not represent a natural course. Frizell et al. [161] studied the effect of spinal fusion on LBP compared with non-operative treatment methods that were not specifically designed to treat LBP. The control group may therefore represent the natural course of LBP. After two years in that study, back pain (VAS) was reduced by 7 % and physical function (ODI) was improved by 6 % in patients treated non-operatively. In a systematic review of prospective cohort studies set in primary care, Itz et al. [132] found that 65 % of the patients still reported LBP one year after the onset of symptoms.

Still, we have limited knowledge on the natural course of chronic LBP over eight years. However, Peng et al. [321] observed a small and clinically unimportant improvement from 46.4 to 44.0 points on ODI over four years in an observational study of patients with chronic LBP. Therefore, we may assume that the change in physical function in our cohort is mainly caused by the intervention, and only minimally influenced by the natural course of LBP.

7 Conclusions and clinical implications

7.1 Paper I

- Significant long-term clinical improvement was observed after both TDR and MDR.
- The long-term functional improvement and pain relief were statistically significantly better after TDR compared to MDR, but the clinical significance of the difference is questionable.
- The long-term results of both TDR and MDR seem to be acceptable, and in line with short-term results.
- Considering the risk of surgical complications, and the significant number of patients who achieve a clinically important improvement after rehabilitation, the first choice of treatment should be MDR.

7.2 Paper II

- Patients with Modic changes were more likely to achieve a clinically important long-term improvement (≥ 15 ODI points).
- Patients with a shorter duration of sick leave, absence of comorbidity, lower ODI score and higher education were more likely to be employed at eight-year follow-up.

7.3 Paper III

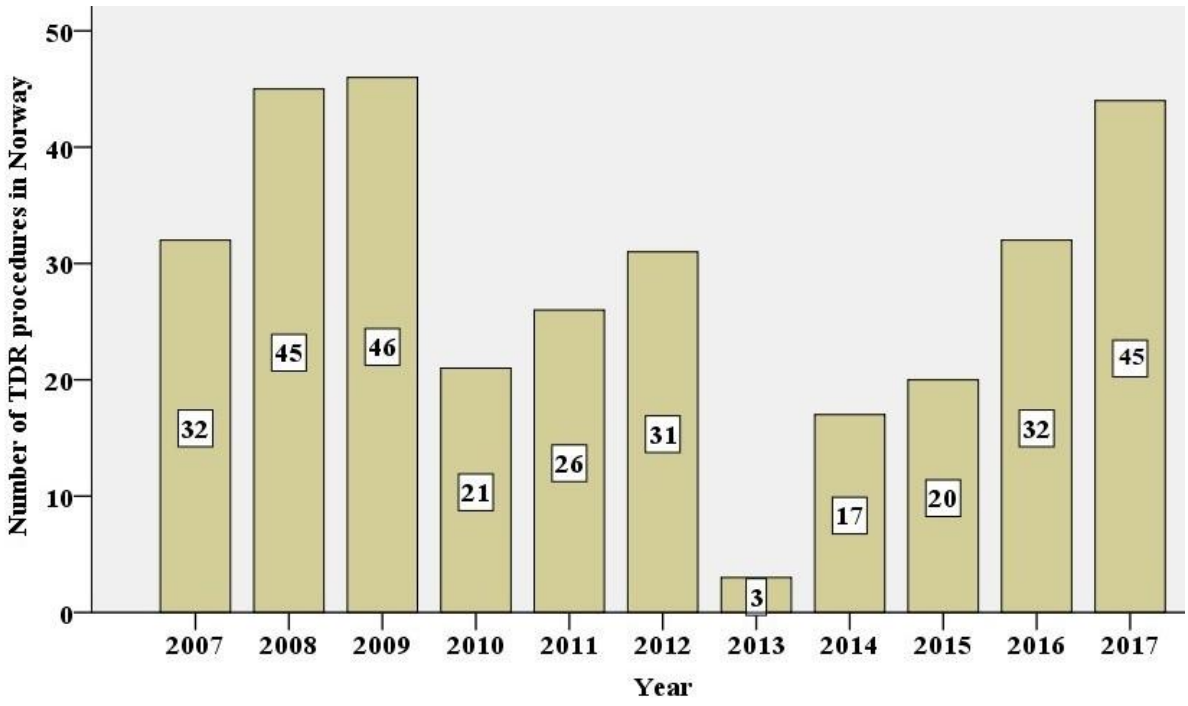
- The development of ADD at eight-year follow-up was similar in patients treated non-operatively and in patients treated with TDR.
- The ADD development was not related to the clinical outcome, and the risk of ADD should have little impact on the choice of TDR versus non-operative treatment for chronic LBP.

8 Future perspectives

Despite initial enthusiasm after the approval of the first TDR devices in 2004, the number of primary TDR procedures performed in the USA has declined by 86 % from 3059 patients in 2005 to 420 patients in 2013 [157 322 323]. According to Salzmann et al. [157], possible explanations for this trend include strict indications for use, challenging surgical techniques, lack of device selection, fear of late complications or revision surgeries and reimbursement issues. In the USA, health insurance companies have not provided coverage for single level lumbar TDR because it is considered investigational. This has provoked patients and spine surgeons, and there has been debate on the reason for this practice [324]. In a report from the First Annual Lumbar TDR Summit in 2016 [325], consensus statements included that reliable tools exist for identifying patients with discogenic back pain [326], sufficient data are available for the long-term efficacy and safety of TDR [223] and the cost to US health insurers were expected to remain unchanged if coverage for lumbar TDR were provided [324]. The future role of TDR is difficult to predict, but in a recent systematic review of 59 clinical trials, Formica et al. [206] concluded that it could be a reliable option for the treatment of LBP in patients with IDD in years to come. In contrast, 2016 NICE guidelines recommend that disc replacement is not offered to patients with LBP [153].

In Norway, lumbar TDR has never come into widespread use. Data from the Norwegian Spine Registry shows that the annual number of TDR procedures has remained below 50, but no obvious trend can be observed (Figure 12).

Figure 12. The number of TDR procedures in Norway from 2007 to 2017. Data from the Norwegian Spine Registry online (http://helseregister.no (01.05.18))



8.1 Suggestions for future research

Although growing evidence exists for the long-term effects of TDR, there are still several unanswered questions that need to be addressed. The following are suggestions for future research.

8.1.1 Prospective studies with even longer follow-up time

Most patients receiving TDR are young or middle-aged adults. Mean age at the time of surgery in the present study was 41 years. Our follow-up time of eight years may therefore be too short to reveal very late complications like implant loosening and implant debris reactions, both well known complications in the field of hip and knee arthroplasty [327-329]. Hence, prospective studies with even longer follow-up times are still needed in order to assess the long-term safety and efficacy of TDR.

8.1.2 Predictor analyses

The limited sample size was an important limitation of our study of baseline patient characteristics associated with a favourable long-term outcome (Paper II). There is still a need for improved patient selection for TDR. Detailed information on baseline characteristics of more than 300 patients treated with TDR is stored in the Norwegian Spine Registry (Figure 12), which could be a source of information about baseline characteristics predicting the outcome. The most important baseline characteristics in our study – the presence and extent of Modic changes – are not among the baseline characteristics collected by the Norwegian Spine Registry. Moreover, the registry collects PROMs at three months and one year postoperatively, meaning that long-term outcomes cannot be extracted from the Norwegian Spine Registry at the moment. However, the registry has recently been allowed to obtain the patients' written consent for long-term follow-up. Further research is needed in order to identify patients with increased probability of both favourable and unfavourable outcomes. International spine registries may be an important source of such information.

8.1.3 The long-term mobility of the artificial disc

The preservation of segmental mobility has been among the main reasons for the introduction of TDR as an alternative to spinal fusion. However, the methods used for the measurement of the mobility of the artificial disc has not been standardised. Previous reports of the correlation between ROM and clinical outcome have been non-consistent [212 330 331]. Heterotopic ossification (HO) is observed in several studies with long-term follow-up. Increasing grades of HO may lead to increasing loss of segmental ROM, but only HO grade 4 indicates segmental ankylosis [211], and the exact influence of other HO grades on segmental ROM has not been established. In a recent prospective study of 51 patients treated with TDR (Prodisc II), Wuertinger et al. [199] reported that the mobility of the artificial disc was maintained, but gradually decreased over an average follow-up period of 7.8 years. However, no correlation was observed between the decrease in ROM and the clinical outcome. Johnsen et al. [212] used distortion compensated roentgen analysis (DCRA) to measure the segmental ROM at two-year follow-up. The DCRA method is documented as a precise tool for the assessment of segmental motion [332], and long-term segmental ROM should be analysed using the same method.

8.1.4 Facet joint arthropathy

Several researchers have shared their concern about facet joint degeneration following TDR [227-232]. Degeneration of the facet joints may be part of the natural course, but at two-year follow-up in the Norwegian TDR Study, Hellum et al. [36] found increased facet joint degeneration in 34 % of patients treated with TDR and 4 % of patients treated non-operatively, indicating a significantly increased risk of facet joint degeneration following TDR compared to the natural course. The assessment of facet joint degeneration was based on MRI, and metal artefacts from the prostheses may have affected the evaluation of the facet joints. We are planning an evaluation of the facet joint degeneration following TDR or non-operative treatment based on computer tomography (CT) at eight-year follow-up in the present study. Further, the CT-based evaluation of the facet joints will be compared to an evaluation based on MRI performed with a metal artefact reducing protocol [244].

8.1.5 Long-term health economic analysis

Since LBP is the leading cause of global disability [1], the health economics aspects of different treatments are of great importance for public health providers. Analyses from Sweden [333] and the USA [334] have found a potential economic benefit of TDR compared to fusion. In the Norwegian TDR Study, Johnsen et al. [335] found that TDR was cost-effective compared to MDR when using EQ-5D, but not SF-6D, for assessing QALY. The indirect costs related to loss of productivity were much higher than the direct treatment-related costs. The observed difference in the work participation rate between the TDR group (48 %) and the MDR group (40 %) was not significant at eight-year follow-up, but could have profound impact on health economy analyses. Therefore, we plan to analyse the cost-effectiveness of TDR compared to MDR in our eight-year follow-up in the Norwegian TDR Study.

8.1.6 The role of different prosthesis designs, constructs and surgical approaches

Several different artificial lumbar disc designs are available for TDR. In a review of different TDR design concepts, Galbusera et al. [182] found evidence for increased load through the facet joints in semi-constrained and unconstrained prostheses. Semi-constrained prostheses were able to share a greater part of the load, but were also more susceptible to wear. Moreover, segmental lordosis alterations were observed both for unconstrained and constrained prostheses. A few prospective studies comparing different prosthesis designs have not yet been able to identify one prosthesis design with superior clinical results [194 336 337]. Different designs may have favourable properties for different patients, but so far there is no evidence for the superiority of a particular prosthesis design in patients with specific anatomic properties [338].

Siepe et al. [208] reported more complications after bisegmental TDR (L4/L5 and L5/S1) compared to monosegmental TDR (L4/L5 or L5/S1), with better results achieved after TDR at L4/L5 than at L5/S1. In a prospective randomised trial, Hoff et al. [339] found better pain relief immediately after the treatment of two-level disc degeneration with a hybrid technique (anterior fusion in L5/S1 and TDR in L4/L5) compared to two-level transforaminal lumbar interbody fusion (TLIF). In a systematic review and meta-analysis of outcomes in hybrid constructs for multilevel disc degeneration, Lackey et al. [340] found a tendency towards

better back pain relief following hybrid surgery compared to multilevel TDR or multilevel fusion. However, the future role of the hybrid technique should be based on larger prospective studies with longer follow-up times.

Tohmeh and Smith [341] have published acceptable clinical outcomes after TDR performed through a mini-invasive lateral trans-psoas approach, and Pimenta et al. [342] have presented promising preliminary data showing long-term pain relief and improved physical function. The approach has theoretical advantages when revision surgery is indicated [343]. At the moment, however, the evidence for increased use of the lateral approach in TDR is weak.

Hence, the role of different prosthesis designs, constructs and surgical approaches should also be explored in future research.

References

1. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2015;**386**(9995):743-800 doi: 10.1016/s0140-6736(15)60692-4[published Online First: Epub Date]|.
2. Balague F, Mannion AF, Pellise F, Cedraschi C. Non-specific low back pain. Lancet 2012;**379**(9814):482-91 doi: 10.1016/s0140-6736(11)60610-7[published Online First: Epub Date]|.
3. Nachemson A, Zdeblick TA, O'Brien JP. Lumbar disc disease with discogenic pain. What surgical treatment is most effective? Spine 1996;**21**(15):1835-8
4. Fardon DF, Williams AL, Dohring EJ, Murtagh FR, Gabriel Rothman SL, Sze GK. Lumbar disc nomenclature: version 2.0: recommendations of the combined task forces of the North American Spine Society, the American Society of Spine Radiology, and the American Society of Neuroradiology. Spine 2014;**39**(24):E1448-65 doi: 10.1097/BRS.0b013e3182a8866d[published Online First: Epub Date]|.
5. Froholdt A, Reikeraas O, Holm I, Keller A, Brox JI. No difference in 9-year outcome in CLBP patients randomized to lumbar fusion versus cognitive intervention and exercises. Eur. Spine J. 2012;**21**(12):2531-8 doi: 10.1007/s00586-012-2382-0[published Online First: Epub Date]|.
6. Buttner-Janz K, Schellnack K, Zippel H. [An alternative treatment strategy in lumbar intervertebral disk damage using an SB Charite modular type intervertebral disk endoprosthesis]. Z. Orthop. Ihre Grenzgeb. 1987;**125**(1):1-6 doi: 10.1055/s-2008-1039666[published Online First: Epub Date]|.
7. Hilibrand AS, Robbins M. Adjacent segment degeneration and adjacent segment disease: the consequences of spinal fusion? The spine journal : official journal of the North American Spine Society 2004;**4**(6 Suppl):190S-94S doi: 10.1016/j.spinee.2004.07.007[published Online First: Epub Date]|.
8. Lund T, Oxland TR. Adjacent level disk disease--is it really a fusion disease? Orthop. Clin. North Am. 2011;**42**(4):529-41, viii doi: 10.1016/j.ocl.2011.07.006[published Online First: Epub Date]|.
9. Roberts S, Evans H, Trivedi J, Menage J. Histology and pathology of the human intervertebral disc. J. Bone Joint Surg. Am. 2006;**88 Suppl 2**:10-4 doi: 10.2106/jbjs.f.00019[published Online First: Epub Date]|.
10. Adams A, Roche O, Mazumder A, Davagnanam I, Mankad K. Imaging of degenerative lumbar intervertebral discs; linking anatomy, pathology and imaging. Postgrad. Med. J. 2014;**90**(1067):511-9 doi: 10.1136/postgradmedj-2013-132193[published Online First: Epub Date]|.
11. Dowdell J, Erwin M, Choma T, Vaccaro A, Iatridis J, Cho SK. Intervertebral Disk Degeneration and Repair. Neurosurgery 2017;**80**(3S):S46-S54 doi: 10.1093/neuros/nyw078[published Online First: Epub Date]|.
12. Adams MA, Roughley PJ. What is intervertebral disc degeneration, and what causes it? Spine 2006;**31**(18):2151-61 doi: 10.1097/01.brs.0000231761.73859.2c[published Online First: Epub Date]|.
13. Feng Y, Egan B, Wang J. Genetic Factors in Intervertebral Disc Degeneration. Genes & diseases 2016;**3**(3):178-85 doi: 10.1016/j.gendis.2016.04.005[published Online First: Epub Date]|.
14. Weber KT, Jacobsen TD, Maidhof R, et al. Developments in intervertebral disc disease research: pathophysiology, mechanobiology, and therapeutics. Curr. Rev. Musculoskelet. Med. 2015;**8**(1):18-31 doi: 10.1007/s12178-014-9253-8[published Online First: Epub Date]|.
15. Boos N, Weissbach S, Rohrbach H, Weiler C, Spratt KF, Nerlich AG. Classification of age-related changes in lumbar intervertebral discs: 2002 Volvo Award in basic science. Spine

2002;**27**(23):2631-44 doi: 10.1097/01.brs.0000035304.27153.5b[published Online First: Epub Date]|.
16. Adams MA, Dolan P. Intervertebral disc degeneration: evidence for two distinct phenotypes. J. Anat. 2012;**221**(6):497-506 doi: 10.1111/j.1469-7580.2012.01551.x[published Online First: Epub Date]|.
17. Albert HB, Lambert P, Rollason J, et al. Does nuclear tissue infected with bacteria following disc herniations lead to Modic changes in the adjacent vertebrae? Eur. Spine J. 2013;**22**(4):690-6 doi: 10.1007/s00586-013-2674-z[published Online First: Epub Date]|.
18. Albert HB, Sorensen JS, Christensen BS, Manniche C. Antibiotic treatment in patients with chronic low back pain and vertebral bone edema (Modic type 1 changes): a double-blind randomized clinical controlled trial of efficacy. Eur. Spine J. 2013;**22**(4):697-707 doi: 10.1007/s00586-013-2675-y[published Online First: Epub Date]|.
19. Modic MT, Masaryk TJ, Ross JS, Carter JR. Imaging of degenerative disk disease. Radiology 1988;**168**(1):177-86
20. Modic MT, Steinberg PM, Ross JS, Masaryk TJ, Carter JR. Degenerative disk disease: assessment of changes in vertebral body marrow with MR imaging. Radiology 1988;**166**(1 Pt 1):193-9
21. Rahme R, Moussa R. The modic vertebral endplate and marrow changes: pathologic significance and relation to low back pain and segmental instability of the lumbar spine. AJNR Am. J. Neuroradiol. 2008;**29**(5):838-42 doi: 10.3174/ajnr.A0925[published Online First: Epub Date]|.
22. Dudli S, Fields AJ, Samartzis D, Karppinen J, Lotz JC. Pathobiology of Modic changes. Eur. Spine J. 2016;**25**(11):3723-34 doi: 10.1007/s00586-016-4459-7[published Online First: Epub Date]|.
23. Marshman LA, Trewhella M, Friesem T, Bhatia CK, Krishna M. Reverse transformation of Modic type 2 changes to Modic type 1 changes during sustained chronic low-back pain severity. Report of two cases and review of the literature. J. Neurosurg. Spine 2007;**6**(2):152-5 doi: 10.3171/spi.2007.6.2.152[published Online First: Epub Date]|.
24. Kjaer P, Leboeuf-Yde C, Korsholm L, Sorensen JS, Bendix T. Magnetic resonance imaging and low back pain in adults: a diagnostic imaging study of 40-year-old men and women. Spine 2005;**30**(10):1173-80
25. Jensen TS, Karppinen J, Sorensen JS, Niinimaki J, Leboeuf-Yde C. Vertebral endplate signal changes (Modic change): a systematic literature review of prevalence and association with non-specific low back pain. Eur. Spine J. 2008;**17**(11):1407-22 doi: 10.1007/s00586-008-0770-2[published Online First: Epub Date]|.
26. Brinjikji W, Diehn FE, Jarvik JG, et al. MRI Findings of Disc Degeneration are More Prevalent in Adults with Low Back Pain than in Asymptomatic Controls: A Systematic Review and Meta-Analysis. AJNR Am. J. Neuroradiol. 2015;**36**(12):2394-9 doi: 10.3174/ajnr.A4498[published Online First: Epub Date]|.
27. Stirling A, Worthington T, Rafiq M, Lambert PA, Elliott TS. Association between sciatica and Propionibacterium acnes. Lancet 2001;**357**(9273):2024-5 doi: 10.1016/s0140-6736(00)05109-6[published Online First: Epub Date]|.
28. Crockett MT, Kelly BS, van Baarsel S, Kavanagh EC. Modic Type 1 Vertebral Endplate Changes: Injury, Inflammation, or Infection? AJR Am. J. Roentgenol. 2017;**209**(1):167-70 doi: 10.2214/ajr.16.17403[published Online First: Epub Date]|.
29. Hellum C, Johnsen LG, Storheim K, et al. Surgery with disc prosthesis versus rehabilitation in patients with low back pain and degenerative disc: two year follow-up of randomised study. BMJ 2011;**342**:d2786 doi: 10.1136/bmj.d2786[published Online First: Epub Date]|.
30. Masharawi Y, Kjaer P, Bendix T, et al. The reproducibility of quantitative measurements in lumbar magnetic resonance imaging of children from the general population. Spine 2008;**33**(19):2094-100 doi: 10.1097/BRS.0b013e31817f19f7[published Online First: Epub Date]|.
31. Teichtahl AJ, Urquhart DM, Wang Y, Wluka AE, Heritier S, Cicuttini FM. A Dose-response relationship between severity of disc degeneration and intervertebral disc height in the

lumbosacral spine. Arthritis Res. Ther. 2015;**17**:297 doi: 10.1186/s13075-015-0820-1[published Online First: Epub Date]|.
32. Pfirrmann CW, Metzdorf A, Zanetti M, Hodler J, Boos N. Magnetic resonance classification of lumbar intervertebral disc degeneration. Spine 2001;**26**(17):1873-8
33. Twomey L, Taylor J. Age changes in lumbar intervertebral discs. Acta Orthop. Scand. 1985;**56**(6):496-9
34. Mannion AF, Leivseth G, Brox JI, Fritzell P, Hagg O, Fairbank JC. ISSLS Prize winner: Long-term follow-up suggests spinal fusion is associated with increased adjacent segment disc degeneration but without influence on clinical outcome: results of a combined follow-up from 4 randomized controlled trials. Spine 2014;**39**(17):1373-83 doi: 10.1097/brs.0000000000000437[published Online First: Epub Date]|.
35. Videman T, Battie MC, Gibbons LE, Gill K. Aging changes in lumbar discs and vertebrae and their interaction: a 15-year follow-up study. The spine journal : official journal of the North American Spine Society 2014;**14**(3):469-78 doi: 10.1016/j.spinee.2013.11.018[published Online First: Epub Date]|.
36. Hellum C, Berg L, Gjertsen O, et al. Adjacent level degeneration and facet arthropathy after disc prosthesis surgery or rehabilitation in patients with chronic low back pain and degenerative disc: second report of a randomized study. Spine 2012;**37**(25):2063-73 doi: 10.1097/BRS.0b013e318263cc46[published Online First: Epub Date]|.
37. Zou J, Yang H, Miyazaki M, et al. Dynamic bulging of intervertebral discs in the degenerative lumbar spine. Spine 2009;**34**(23):2545-50 doi: 10.1097/BRS.0b013e3181b32998[published Online First: Epub Date]|.
38. Roughley PJ. Biology of intervertebral disc aging and degeneration: involvement of the extracellular matrix. Spine 2004;**29**(23):2691-9
39. Ellingson AM, Nagel TM, Polly DW, Ellermann J, Nuckley DJ. Quantitative T2* (T2 star) relaxation times predict site specific proteoglycan content and residual mechanics of the intervertebral disc throughout degeneration. J. Orthop. Res. 2014;**32**(8):1083-9 doi: 10.1002/jor.22633[published Online First: Epub Date]|.
40. Luoma K, Riihimaki H, Luukkonen R, Raininko R, Viikari-Juntura E, Lamminen A. Low back pain in relation to lumbar disc degeneration. Spine 2000;**25**(4):487-92
41. Videman T, Battie MC, Gibbons LE, Maravilla K, Manninen H, Kaprio J. Associations between back pain history and lumbar MRI findings. Spine 2003;**28**(6):582-8 doi: 10.1097/01.brs.0000049905.44466.73[published Online First: Epub Date]|.
42. Horton WC, Daftari TK. Which disc as visualized by magnetic resonance imaging is actually a source of pain? A correlation between magnetic resonance imaging and discography. Spine 1992;**17**(6 Suppl):S164-71
43. Buirski G, Silberstein M. The symptomatic lumbar disc in patients with low-back pain. Magnetic resonance imaging appearances in both a symptomatic and control population. Spine 1993;**18**(13):1808-11
44. Hu X, Chen M, Pan J, Liang L, Wang Y. Is it appropriate to measure age-related lumbar disc degeneration on the mid-sagittal MR image? A quantitative image study. Eur. Spine J. 2017 doi: 10.1007/s00586-017-5357-3[published Online First: Epub Date]|.
45. Aprill C, Bogduk N. High-intensity zone: a diagnostic sign of painful lumbar disc on magnetic resonance imaging. Br. J. Radiol. 1992;**65**(773):361-9 doi: 10.1259/0007-1285-65-773-361[published Online First: Epub Date]|.
46. Khan I, Hargunani R, Saifuddin A. The lumbar high-intensity zone: 20 years on. Clin. Radiol. 2014;**69**(6):551-8 doi: 10.1016/j.crad.2013.12.012[published Online First: Epub Date]|.
47. Carragee EJ, Paragioudakis SJ, Khurana S. 2000 Volvo Award winner in clinical studies: Lumbar high-intensity zone and discography in subjects without low back problems. Spine 2000;**25**(23):2987-92
48. Liu C, Cai HX, Zhang JF, Ma JJ, Lu YJ, Fan SW. Quantitative estimation of the high-intensity zone in the lumbar spine: comparison between the symptomatic and asymptomatic population. The

spine journal : official journal of the North American Spine Society 2014;**14**(3):391-6 doi: 10.1016/j.spinee.2013.06.078[published Online First: Epub Date]|.
49. Stadnik TW, Lee RR, Coen HL, Neirynck EC, Buisseret TS, Osteaux MJ. Annular tears and disk herniation: prevalence and contrast enhancement on MR images in the absence of low back pain or sciatica. Radiology 1998;**206**(1):49-55 doi: 10.1148/radiology.206.1.9423651[published Online First: Epub Date]|.
50. Endean A, Palmer KT, Coggon D. Potential of magnetic resonance imaging findings to refine case definition for mechanical low back pain in epidemiological studies: a systematic review. Spine 2011;**36**(2):160-9 doi: 10.1097/BRS.0b013e3181cd9adb[published Online First: Epub Date]|.
51. Mok FP, Samartzis D, Karppinen J, Fong DY, Luk KD, Cheung KM. Modic changes of the lumbar spine: prevalence, risk factors, and association with disc degeneration and low back pain in a large-scale population-based cohort. The spine journal : official journal of the North American Spine Society 2016;**16**(1):32-41 doi: 10.1016/j.spinee.2015.09.060[published Online First: Epub Date]|.
52. de Schepper EI, Damen J, van Meurs JB, et al. The association between lumbar disc degeneration and low back pain: the influence of age, gender, and individual radiographic features. Spine 2010;**35**(5):531-6 doi: 10.1097/BRS.0b013e3181aa5b33[published Online First: Epub Date]|.
53. Livshits G, Popham M, Malkin I, et al. Lumbar disc degeneration and genetic factors are the main risk factors for low back pain in women: the UK Twin Spine Study. Ann. Rheum. Dis. 2011;**70**(10):1740-5 doi: 10.1136/ard.2010.137836[published Online First: Epub Date]|.
54. Berg L, Hellum C, Gjertsen O, et al. Do more MRI findings imply worse disability or more intense low back pain? A cross-sectional study of candidates for lumbar disc prosthesis. Skeletal Radiol. 2013;**42**(11):1593-602 doi: 10.1007/s00256-013-1700-x[published Online First: Epub Date]|.
55. Airaksinen O, Brox JI, 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 doi: 10.1007/s00586-006-1072-1[published Online First: Epub Date]|.
56. McIntosh G, Hall H. Low back pain (acute). BMJ clinical evidence 2011;**2011**
57. Chou R. Low back pain (chronic). BMJ clinical evidence 2010;**2010**
58. Hoy D, March L, Brooks P, et al. The global burden of low back pain: estimates from the Global Burden of Disease 2010 study. Ann. Rheum. Dis. 2014;**73**(6):968-74 doi: 10.1136/annrheumdis-2013-204428[published Online First: Epub Date]|.
59. Koes BW, van Tulder MW, Thomas S. Diagnosis and treatment of low back pain. BMJ 2006;**332**(7555):1430-4 doi: 10.1136/bmj.332.7555.1430[published Online First: Epub Date]|.
60. Hoy D, Bain C, Williams G, et al. A systematic review of the global prevalence of low back pain. Arthritis Rheum. 2012;**64**(6):2028-37 doi: 10.1002/art.34347[published Online First: Epub Date]|.
61. Kinge JM, Knudsen AK, Skirbekk V, Vollset SE. Musculoskeletal disorders in Norway: prevalence of chronicity and use of primary and specialist health care services. BMC Musculoskelet. Disord. 2015;**16**:75 doi: 10.1186/s12891-015-0536-z[published Online First: Epub Date]|.
62. Fakta om ryggsmerter. Fakta om arbeid og helse. In: Health NIoO, ed., 2010.
63. van Tulder MW, Assendelft WJ, Koes BW, Bouter LM. Spinal radiographic findings and nonspecific low back pain. A systematic review of observational studies. Spine 1997;**22**(4):427-34
64. Waddell G. The Back Pain Revolution. First ed. London: Churchill Livingstone 1998
65. Costa Lda C, Maher CG, McAuley JH, et al. Prognosis for patients with chronic low back pain: inception cohort study. BMJ 2009;**339**:b3829 doi: 10.1136/bmj.b3829[published Online First: Epub Date]|.
66. den Hollander M, de Jong JR, Volders S, Goossens ME, Smeets RJ, Vlaeyen JW. Fear reduction in patients with chronic pain: a learning theory perspective. Expert Rev. Neurother. 2010;**10**(11):1733-45 doi: 10.1586/ern.10.115[published Online First: Epub Date]|.

67. Linton SJ, Shaw WS. Impact of psychological factors in the experience of pain. Phys. Ther. 2011;**91**(5):700-11 doi: 10.2522/ptj.20100330[published Online First: Epub Date]|.
68. Nicholas MK, Linton SJ, Watson PJ, Main CJ. Early identification and management of psychological risk factors ("yellow flags") in patients with low back pain: a reappraisal. Phys. Ther. 2011;**91**(5):737-53 doi: 10.2522/ptj.20100224[published Online First: Epub Date]|.
69. Ranger TA, Cicuttini FM, Jensen TS, et al. Are the size and composition of the paraspinal muscles associated with low back pain? A systematic review. The spine journal : official journal of the North American Spine Society 2017;**17**(11):1729-48 doi: 10.1016/j.spinee.2017.07.002[published Online First: Epub Date]|.
70. Goubert D, Oosterwijck JV, Meeus M, Danneels L. Structural Changes of Lumbar Muscles in Non-specific Low Back Pain: A Systematic Review. Pain physician 2016;**19**(7):E985-E1000
71. Kalichman L, Carmeli E, Been E. The Association between Imaging Parameters of the Paraspinal Muscles, Spinal Degeneration, and Low Back Pain. BioMed research international 2017;**2017**:2562957 doi: 10.1155/2017/2562957[published Online First: Epub Date]|.
72. Maas ET, Juch JN, Ostelo RW, et al. Systematic review of patient history and physical examination to diagnose chronic low back pain originating from the facet joints. Eur. J. Pain 2017;**21**(3):403-14 doi: 10.1002/ejp.963[published Online First: Epub Date]|.
73. Allegri M, Montella S, Salici F, et al. Mechanisms of low back pain: a guide for diagnosis and therapy. F1000Research 2016;**5** doi: 10.12688/f1000research.8105.2[published Online First: Epub Date]|.
74. Vekaria R, Bhatt R, Ellard DR, Henschke N, Underwood M, Sandhu H. Intra-articular facet joint injections for low back pain: a systematic review. Eur. Spine J. 2016;**25**(4):1266-81 doi: 10.1007/s00586-016-4455-y[published Online First: Epub Date]|.
75. Cohen SP, Raja SN. Pathogenesis, diagnosis, and treatment of lumbar zygapophysial (facet) joint pain. Anesthesiology 2007;**106**(3):591-614
76. Schwarzer AC, Aprill CN, Bogduk N. The sacroiliac joint in chronic low back pain. Spine 1995;**20**(1):31-7
77. Hoy D, Brooks P, Blyth F, Buchbinder R. The Epidemiology of low back pain. Best Pract. Res. Clin. Rheumatol. 2010;**24**(6):769-81 doi: 10.1016/j.berh.2010.10.002[published Online First: Epub Date]|.
78. Battie MC, Videman T, Levalahti 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):2801-8 doi: 10.1097/BRS.0b013e31818043b7[published Online First: Epub Date]|.
79. Kalichman L, Hunter DJ. The genetics of intervertebral disc degeneration. Familial predisposition and heritability estimation. Joint Bone Spine 2008;**75**(4):383-7 doi: 10.1016/j.jbspin.2007.11.003[published Online First: Epub Date]|.
80. Kalichman L, Hunter DJ. The genetics of intervertebral disc degeneration. Associated genes. Joint Bone Spine 2008;**75**(4):388-96 doi: 10.1016/j.jbspin.2007.11.002[published Online First: Epub Date]|.
81. Vo NV, Hartman RA, Yurube T, Jacobs LJ, Sowa GA, Kang JD. Expression and regulation of metalloproteinases and their inhibitors in intervertebral disc aging and degeneration. The spine journal : official journal of the North American Spine Society 2013;**13**(3):331-41 doi: 10.1016/j.spinee.2012.02.027[published Online First: Epub Date]|.
82. Perera RS, Dissanayake PH, Senarath U, Wijayaratne LS, Karunanayake AL, Dissanayake VH. Single Nucleotide Variants of Candidate Genes in Aggrecan Metabolic Pathway Are Associated with Lumbar Disc Degeneration and Modic Changes. PLoS One 2017;**12**(1):e0169835 doi: 10.1371/journal.pone.0169835[published Online First: Epub Date]|.
83. Battie MC, Videman T, Levalahti E, Gill K, Kaprio J. Heritability of low back pain and the role of disc degeneration. Pain 2007;**131**(3):272-80 doi: 10.1016/j.pain.2007.01.010[published Online First: Epub Date]|.

84. Ferreira PH, Beckenkamp P, Maher CG, Hopper JL, Ferreira ML. Nature or nurture in low back pain? Results of a systematic review of studies based on twin samples. Eur. J. Pain 2013;**17**(7):957-71 doi: 10.1002/j.1532-2149.2012.00277.x[published Online First: Epub Date]|.
85. Reimann F, Cox JJ, Belfer I, et al. Pain perception is altered by a nucleotide polymorphism in SCN9A. Proc. Natl. Acad. Sci. U. S. A. 2010;**107**(11):5148-53 doi: 10.1073/pnas.0913181107[published Online First: Epub Date]|.
86. Tegeder I, Lotsch J. Current evidence for a modulation of low back pain by human genetic variants. J. Cell. Mol. Med. 2009;**13**(8B):1605-19 doi: 10.1111/j.1582-4934.2009.00703.x[published Online First: Epub Date]|.
87. Omair A, Mannion AF, Holden M, et al. Catechol-O-methyltransferase (COMT) gene polymorphisms are associated with baseline disability but not long-term treatment outcome in patients with chronic low back pain. Eur. Spine J. 2015;**24**(11):2425-31 doi: 10.1007/s00586-015-3866-5[published Online First: Epub Date]|.
88. Kopec JA, Sayre EC, Esdaile JM. Predictors of back pain in a general population cohort. Spine 2004;**29**(1):70-7; discussion 77-8 doi: 10.1097/01.brs.0000103942.81227.7f[published Online First: Epub Date]|.
89. Waxman R, Tennant A, Helliwell P. A prospective follow-up study of low back pain in the community. Spine 2000;**25**(16):2085-90
90. Reigo T, Timpka T, Tropp H. The epidemiology of back pain in vocational age groups. Scand. J. Prim. Health Care 1999;**17**(1):17-21
91. Hurwitz EL, Morgenstern H. Correlates of back problems and back-related disability in the United States. J. Clin. Epidemiol. 1997;**50**(6):669-81
92. Loney PL, Stratford PW. The prevalence of low back pain in adults: a methodological review of the literature. Phys. Ther. 1999;**79**(4):384-96
93. Lawrence RC, Helmick CG, Arnett FC, et al. Estimates of the prevalence of arthritis and selected musculoskeletal disorders in the United States. Arthritis Rheum. 1998;**41**(5):778-99 doi: 10.1002/1529-0131(199805)41:5<778::aid-art4>3.0.co;2-v[published Online First: Epub Date]|.
94. Dionne CE, Dunn KM, Croft PR. Does back pain prevalence really decrease with increasing age? A systematic review. Age Ageing 2006;**35**(3):229-34 doi: 10.1093/ageing/afj055[published Online First: Epub Date]|.
95. Wong AY, Karppinen J, Samartzis D. Low back pain in older adults: risk factors, management options and future directions. Scoliosis and spinal disorders 2017;**12**:14 doi: 10.1186/s13013-017-0121-3[published Online First: Epub Date]|.
96. Toroptsova NV, Benevolenskaya LI, Karyakin AN, Sergeev IL, Erdesz S. "Cross-sectional" study of low back pain among workers at an industrial enterprise in Russia. Spine 1995;**20**(3):328-32
97. Meucci RD, Fassa AG, Faria NM. Prevalence of chronic low back pain: systematic review. Rev. Saude Publica 2015;**49** doi: 10.1590/s0034-8910.2015049005874[published Online First: Epub Date]|.
98. Linton SJ, Hellsing AL, Hallden K. A population-based study of spinal pain among 35-45-year-old individuals. Prevalence, sick leave, and health care use. Spine 1998;**23**(13):1457-63
99. Smith BH, Elliott AM, Hannaford PC, Chambers WA, Smith WC. Factors related to the onset and persistence of chronic back pain in the community: results from a general population follow-up study. Spine 2004;**29**(9):1032-40
100. Thomas E, Silman AJ, Croft PR, Papageorgiou AC, Jayson MI, Macfarlane GJ. Predicting who develops chronic low back pain in primary care: a prospective study. BMJ 1999;**318**(7199):1662-7
101. Dario AB, Ferreira ML, Refshauge KM, Lima TS, Ordonana JR, Ferreira PH. The relationship between obesity, low back pain, and lumbar disc degeneration when genetics and the environment are considered: a systematic review of twin studies. The spine journal : official

journal of the North American Spine Society 2015;**15**(5):1106-17 doi: 10.1016/j.spinee.2015.02.001[published Online First: Epub Date]|.
102. Dario AB, Loureiro Ferreira M, Refshauge K, Luque-Suarez A, Ordonana JR, Ferreira PH. Obesity does not increase the risk of chronic low back pain when genetics are considered. A prospective study of Spanish adult twins. The spine journal : official journal of the North American Spine Society 2017;**17**(2):282-90 doi: 10.1016/j.spinee.2016.10.006[published Online First: Epub Date]|.
103. Heuch I, Heuch I, Hagen K, Zwart JA. Body mass index as a risk factor for developing chronic low back pain: a follow-up in the Nord-Trondelag Health Study. Spine 2013;**38**(2):133-9 doi: 10.1097/BRS.0b013e3182647af2[published Online First: Epub Date]|.
104. Heuch I, Heuch I, Hagen K, Zwart JA. Association between body height and chronic low back pain: a follow-up in the Nord-Trondelag Health Study. BMJ open 2015;**5**(6):e006983 doi: 10.1136/bmjopen-2014-006983[published Online First: Epub Date]|.
105. Gore M, Sadosky A, Stacey BR, Tai KS, Leslie D. The burden of chronic low back pain: clinical comorbidities, treatment patterns, and health care costs in usual care settings. Spine 2012;**37**(11):E668-77 doi: 10.1097/BRS.0b013e318241e5de[published Online First: Epub Date]|.
106. Manchikanti L, Singh V, Falco FJ, Benyamin RM, Hirsch JA. Epidemiology of low back pain in adults. Neuromodulation 2014;**17 Suppl 2**:3-10 doi: 10.1111/ner.12018[published Online First: Epub Date]|.
107. Chou R, Shekelle P. Will this patient develop persistent disabling low back pain? JAMA 2010;**303**(13):1295-302 doi: 10.1001/jama.2010.344[published Online First: Epub Date]|.
108. Stewart Williams J, Ng N, Peltzer K, et al. Risk Factors and Disability Associated with Low Back Pain in Older Adults in Low- and Middle-Income Countries. Results from the WHO Study on Global AGEing and Adult Health (SAGE). PLoS One 2015;**10**(6):e0127880 doi: 10.1371/journal.pone.0127880[published Online First: Epub Date]|.
109. Jacobs JM, Hammerman-Rozenberg R, Cohen A, Stessman J. Chronic back pain among the elderly: prevalence, associations, and predictors. Spine 2006;**31**(7):E203-7 doi: 10.1097/01.brs.0000206367.57918.3c[published Online First: Epub Date]|.
110. Kwon BK, Roffey DM, Bishop PB, Dagenais S, Wai EK. Systematic review: occupational physical activity and low back pain. Occup. Med. (Lond.) 2011;**61**(8):541-8 doi: 10.1093/occmed/kqr092[published Online First: Epub Date]|.
111. Heuch I, Heuch I, Hagen K, Zwart JA. Physical activity level at work and risk of chronic low back pain: A follow-up in the Nord-Trondelag Health Study. PLoS One 2017;**12**(4):e0175086 doi: 10.1371/journal.pone.0175086[published Online First: Epub Date]|.
112. Heuch I, Heuch I, Hagen K, Zwart JA. Is there a U-shaped relationship between physical activity in leisure time and risk of chronic low back pain? A follow-up in the HUNT Study. BMC Public Health 2016;**16**(1):306 doi: 10.1186/s12889-016-2970-8[published Online First: Epub Date]|.
113. Zadro JR, Shirley D, Amorim A, Perez-Riquelme F, Ordonana JR, Ferreira PH. Are people with chronic low back pain meeting the physical activity guidelines? A co-twin control study. The spine journal : official journal of the North American Spine Society 2017;**17**(6):845-54 doi: 10.1016/j.spinee.2017.01.015[published Online First: Epub Date]|.
114. Trompeter K, Fett D, Platen P. Prevalence of Back Pain in Sports: A Systematic Review of the Literature. Sports Med. 2017;**47**(6):1183-207 doi: 10.1007/s40279-016-0645-3[published Online First: Epub Date]|.
115. Witwit WA, Kovac P, Sward A, et al. Disc degeneration on MRI is more prevalent in young elite skiers compared to controls. Knee Surg. Sports Traumatol. Arthrosc. 2018;**26**(1):325-32 doi: 10.1007/s00167-017-4545-3[published Online First: Epub Date]|.
116. Todd C, Aminoff AS, Agnvall C, et al. No difference in prevalence of spine and hip pain in young Elite skiers. Knee Surg. Sports Traumatol. Arthrosc. 2017 doi: 10.1007/s00167-017-4733-1[published Online First: Epub Date]|.

117. Garcia JB, Hernandez-Castro JJ, Nunez RG, et al. Prevalence of low back pain in Latin America: a systematic literature review. Pain physician 2014;**17**(5):379-91
118. Shiri R, Karppinen J, Leino-Arjas P, Solovieva S, Viikari-Juntura E. The association between smoking and low back pain: a meta-analysis. Am. J. Med. 2010;**123**(1):87 e7-35 doi: 10.1016/j.amjmed.2009.05.028[published Online First: Epub Date]|.
119. Yang H, Haldeman S. Behavior-Related Factors Associated With Low Back Pain in the US Adult Population. Spine 2018;**43**(1):28-34 doi: 10.1097/brs.0000000000001665[published Online First: Epub Date]|.
120. Mikkonen P, Leino-Arjas P, Remes J, Zitting P, Taimela S, Karppinen J. Is smoking a risk factor for low back pain in adolescents? A prospective cohort study. Spine 2008;**33**(5):527-32 doi: 10.1097/BRS.0b013e3181657d3c[published Online First: Epub Date]|.
121. Fliesser M, De Witt Huberts J, Wippert PM. The choice that matters: the relative influence of socioeconomic status indicators on chronic back pain- a longitudinal study. BMC Health Serv. Res. 2017;**17**(1):800 doi: 10.1186/s12913-017-2735-9[published Online First: Epub Date]|.
122. Dionne CE, Von Korff M, Koepsell TD, Deyo RA, Barlow WE, Checkoway H. Formal education and back pain: a review. J. Epidemiol. Community Health 2001;**55**(7):455-68
123. Zadro JR, Shirley D, Pinheiro MB, et al. Does educational attainment increase the risk of low back pain when genetics are considered? A population-based study of Spanish twins. The spine journal : official journal of the North American Spine Society 2017;**17**(4):518-30 doi: 10.1016/j.spinee.2016.10.021[published Online First: Epub Date]|.
124. Nordstoga AL, Nilsen TIL, Vasseljen O, Unsgaard-Tondel M, Mork PJ. The influence of multisite pain and psychological comorbidity on prognosis of chronic low back pain: longitudinal data from the Norwegian HUNT Study. BMJ open 2017;**7**(5):e015312 doi: 10.1136/bmjopen-2016-015312[published Online First: Epub Date]|.
125. George SZ, Beneciuk JM. Psychological predictors of recovery from low back pain: a prospective study. BMC Musculoskelet. Disord. 2015;**16**:49 doi: 10.1186/s12891-015-0509-2[published Online First: Epub Date]|.
126. Grotle M, Vollestad NK, Brox JI. Clinical course and impact of fear-avoidance beliefs in low back pain: prospective cohort study of acute and chronic low back pain: II. Spine 2006;**31**(9):1038-46 doi: 10.1097/01.brs.0000214878.01709.0e[published Online First: Epub Date]|.
127. Hoogendoorn WE, van Poppel MN, Bongers PM, Koes BW, Bouter LM. Systematic review of psychosocial factors at work and private life as risk factors for back pain. Spine 2000;**25**(16):2114-25
128. Linton SJ. Occupational psychological factors increase the risk for back pain: a systematic review. Journal of occupational rehabilitation 2001;**11**(1):53-66
129. Ganesan S, Acharya AS, Chauhan R, Acharya S. Prevalence and Risk Factors for Low Back Pain in 1,355 Young Adults: A Cross-Sectional Study. Asian spine journal 2017;**11**(4):610-17 doi: 10.4184/asj.2017.11.4.610[published Online First: Epub Date]|.
130. Grotle M, Brox JI, Glomsrod B, Lonn JH, Vollestad NK. Prognostic factors in first-time care seekers due to acute low back pain. Eur. J. Pain 2007;**11**(3):290-8 doi: 10.1016/j.ejpain.2006.03.004[published Online First: Epub Date]|.
131. Henschke N, Maher CG, Refshauge KM, et al. Prognosis in patients with recent onset low back pain in Australian primary care: inception cohort study. BMJ 2008;**337**:a171 doi: 10.1136/bmj.a171[published Online First: Epub Date]|.
132. Itz CJ, Geurts JW, van Kleef M, Nelemans P. Clinical course of non-specific low back pain: a systematic review of prospective cohort studies set in primary care. Eur. J. Pain 2013;**17**(1):5-15 doi: 10.1002/j.1532-2149.2012.00170.x[published Online First: Epub Date]|.
133. Kjaer P, Korsholm L, Leboeuf-Yde C, Hestbaek L, Bendix T. Individual courses of low back pain in adult Danes: a cohort study with 4-year and 8-year follow-up. BMC Musculoskelet. Disord. 2017;**18**(1):28 doi: 10.1186/s12891-016-1377-0[published Online First: Epub Date]|.

134. Campbell P, Foster NE, Thomas E, Dunn KM. Prognostic indicators of low back pain in primary care: five-year prospective study. J. Pain 2013;**14**(8):873-83 doi: 10.1016/j.jpain.2013.03.013[published Online First: Epub Date]|.
135. Pinheiro MB, Ferreira ML, Refshauge K, et al. Symptoms of depression as a prognostic factor for low back pain: a systematic review. The spine journal : official journal of the North American Spine Society 2016;**16**(1):105-16 doi: 10.1016/j.spinee.2015.10.037[published Online First: Epub Date]|.
136. Wertli MM, Rasmussen-Barr E, Weiser S, Bachmann LM, Brunner F. The role of fear avoidance beliefs as a prognostic factor for outcome in patients with nonspecific low back pain: a systematic review. The spine journal : official journal of the North American Spine Society 2014;**14**(5):816-36 e4 doi: 10.1016/j.spinee.2013.09.036[published Online First: Epub Date]|.
137. Wilkens P, Scheel IB, Grundnes O, Hellum C, Storheim K. Prognostic factors of prolonged disability in patients with chronic low back pain and lumbar degeneration in primary care: a cohort study. Spine 2013;**38**(1):65-74 doi: 10.1097/BRS.0b013e318263bb7b[published Online First: Epub Date]|.
138. Keller A, Boyle E, Skog TA, Cassidy JD, Bautz-Holter E. Are Modic changes prognostic for recovery in a cohort of patients with non-specific low back pain? Eur. Spine J. 2012;**21**(3):418-24 doi: 10.1007/s00586-011-1964-6[published Online First: Epub Date]|.
139. Saragiotto BT, Machado GC, Ferreira ML, Pinheiro MB, Abdel Shaheed C, Maher CG. Paracetamol for low back pain. Cochrane database of systematic reviews (Online) 2016(6):CD012230 doi: 10.1002/14651858.cd012230[published Online First: Epub Date]|.
140. Ebadi S, Henschke N, Nakhostin Ansari N, Fallah E, van Tulder MW. Therapeutic ultrasound for chronic low-back pain. Cochrane database of systematic reviews (Online) 2014(3):CD009169 doi: 10.1002/14651858.CD009169.pub2[published Online First: Epub Date]|.
141. Saragiotto BT, Maher CG, Yamato TP, et al. Motor control exercise for chronic non-specific low-back pain. Cochrane database of systematic reviews (Online) 2016(1):CD012004 doi: 10.1002/14651858.cd012004[published Online First: Epub Date]|.
142. Franke H, Fryer G, Ostelo RW, Kamper SJ. Muscle energy technique for non-specific low-back pain. Cochrane database of systematic reviews (Online) 2015(2):CD009852 doi: 10.1002/14651858.CD009852.pub2[published Online First: Epub Date]|.
143. Yamato TP, Maher CG, Saragiotto BT, et al. Pilates for low back pain. Cochrane database of systematic reviews (Online) 2015(7):CD010265 doi: 10.1002/14651858.CD010265.pub2[published Online First: Epub Date]|.
144. Wieland LS, Skoetz N, Pilkington K, Vempati R, D'Adamo CR, Berman BM. Yoga treatment for chronic non-specific low back pain. Cochrane database of systematic reviews (Online) 2017;**1**:CD010671 doi: 10.1002/14651858.CD010671.pub2[published Online First: Epub Date]|.
145. Fordyce WE. *Behavioral methods for chronic pain and illness*. St. Louis: Mosby, 1976.
146. Pincus T, Vogel S, Burton AK, Santos R, Field AP. Fear avoidance and prognosis in back pain: a systematic review and synthesis of current evidence. Arthritis Rheum. 2006;**54**(12):3999-4010 doi: 10.1002/art.22273[published Online First: Epub Date]|.
147. Turner JA, Jensen MP. Efficacy of cognitive therapy for chronic low back pain. Pain 1993;**52**(2):169-77
148. Turk DC, Flor H. Etiological theories and treatments for chronic back pain. II. Psychological models and interventions. Pain 1984;**19**(3):209-33
149. Vlaeyen JW, Haazen IW, Schuerman JA, Kole-Snijders AM, van Eek H. Behavioural rehabilitation of chronic low back pain: comparison of an operant treatment, an operant-cognitive treatment and an operant-respondent treatment. Br. J. Clin. Psychol. 1995;**34 (Pt 1)**:95-118
150. Henschke N, Ostelo RW, van Tulder MW, et al. Behavioural treatment for chronic low-back pain. Cochrane database of systematic reviews (Online) 2010(7):CD002014 doi: 10.1002/14651858.CD002014.pub3[published Online First: Epub Date]|.

151. Parreira P, Heymans MW, van Tulder MW, et al. Back Schools for chronic non-specific low back pain. Cochrane database of systematic reviews (Online) 2017;**8**:CD011674 doi: 10.1002/14651858.CD011674.pub2[published Online First: Epub Date]|.
152. Kamper SJ, Apeldoorn AT, Chiarotto A, et al. Multidisciplinary biopsychosocial rehabilitation for chronic low back pain. Cochrane database of systematic reviews (Online) 2014(9):CD000963 doi: 10.1002/14651858.CD000963.pub3[published Online First: Epub Date]|.
153. NICE Guidelines. Low back pain and sciatica in over 16s: assessment and management. UK National Institute for Health and Care Excellence., 2016.
154. Hadra BE. Wiring of the vertebrae as a means of immobilization in fracture and Potts' disease. Med Times and Register 1891(22):1-8
155. Tarpada SP, Morris MT, Burton DA. Spinal fusion surgery: A historical perspective. Journal of orthopaedics 2017;**14**(1):134-36 doi: 10.1016/j.jor.2016.10.029[published Online First: Epub Date]|.
156. Jacobs W, Van der Gaag NA, Tuschel A, et al. Total disc replacement for chronic back pain in the presence of disc degeneration. Cochrane database of systematic reviews (Online) 2012;**9**:CD008326 doi: 10.1002/14651858.CD008326.pub2[published Online First: Epub Date]|.
157. Salzmann SN, Plais N, Shue J, Girardi FP. Lumbar disc replacement surgery-successes and obstacles to widespread adoption. Curr. Rev. Musculoskelet. Med. 2017;**10**(2):153-59 doi: 10.1007/s12178-017-9397-4[published Online First: Epub Date]|.
158. Brox JI, Sorensen R, Friis A, et al. 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):1913-21 doi: 10.1097/01.brs.0000083234.62751.7a[published Online First: Epub Date]|.
159. Fairbank J, Frost H, Wilson-MacDonald J, Yu LM, 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 doi: 10.1136/bmj.38441.620417.8F[published Online First: Epub Date]|.
160. Mannion AF, Brox JI, Fairbank JC. Comparison of spinal fusion and nonoperative treatment in patients with chronic low back pain: long-term follow-up of three randomized controlled trials. The spine journal : official journal of the North American Spine Society 2013;**13**(11):1438-48 doi: 10.1016/j.spinee.2013.06.101[published Online First: Epub Date]|.
161. Fritzell P, Hagg 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):2521-32; discussion 32-4
162. Katz JN. Lumbar spinal fusion. Surgical rates, costs, and complications. Spine 1995;**20**(24 Suppl):78S-83S
163. Wong AP, Smith ZA, Nixon AT, et al. Intraoperative and perioperative complications in minimally invasive transforaminal lumbar interbody fusion: a review of 513 patients. J. Neurosurg. Spine 2015;**22**(5):487-95 doi: 10.3171/2014.10.spine14129[published Online First: Epub Date]|.
164. Mobbs RJ, Phan K, Malham G, Seex K, Rao PJ. Lumbar interbody fusion: techniques, indications and comparison of interbody fusion options including PLIF, TLIF, MI-TLIF, OLIF/ATP, LLIF and ALIF. Journal of spine surgery (Hong Kong) 2015;**1**(1):2-18 doi: 10.3978/j.issn.2414-469X.2015.10.05[published Online First: Epub Date]|.
165. Sanchez-Raya J, Bago J, Pellise F, Cuxart A, Villanueva C. Does the lower instrumented vertebra have an effect on lumbar mobility, subjective perception of trunk flexibility, and quality of life in patients with idiopathic scoliosis treated by spinal fusion? Journal of spinal disorders & techniques 2012;**25**(8):437-42 doi: 10.1097/BSD.0b013e3182318622[published Online First: Epub Date]|.

166. Luk KD, Chow DH, Evans JH, Leong JC. Lumbar spinal mobility after short anterior interbody fusion. Spine 1995;**20**(7):813-8
167. Harrop JS, Youssef JA, Maltenfort M, et al. Lumbar adjacent segment degeneration and disease after arthrodesis and total disc arthroplasty. Spine 2008;**33**(15):1701-7 doi: 10.1097/BRS.0b013e31817bb956[published Online First: Epub Date]|.
168. Tropiano P, Huang RC, Girardi FP, Cammisa FP, Jr., Marnay T. Lumbar total disc replacement. Seven to eleven-year follow-up. J. Bone Joint Surg. Am. 2005;**87**(3):490-6 doi: 10.2106/jbjs.c.01345[published Online First: Epub Date]|.
169. Helgeson MD, Bevevino AJ, Hilibrand AS. Update on the evidence for adjacent segment degeneration and disease. The spine journal : official journal of the North American Spine Society 2013;**13**(3):342-51 doi: 10.1016/j.spinee.2012.12.009[published Online First: Epub Date]|.
170. Wang JC, Arnold PM, Hermsmeyer JT, Norvell DC. Do lumbar motion preserving devices reduce the risk of adjacent segment pathology compared with fusion surgery? A systematic review. Spine 2012;**37**(22 Suppl):S133-43 doi: 10.1097/BRS.0b013e31826cadf2[published Online First: Epub Date]|.
171. Learmonth ID, Young C, Rorabeck C. The operation of the century: total hip replacement. Lancet 2007;**370**(9597):1508-19 doi: 10.1016/s0140-6736(07)60457-7[published Online First: Epub Date]|.
172. Fernstrom U. Arthroplasty with intercorporal endoprothesis in herniated disc and in painful disc. Acta Chir. Scand. Suppl. 1966;**357**:154-9
173. Fisahn C, Burgess B, Iwanaga J, Chapman JR, Oskouian RJ, Tubbs RS. Ulf Fernstrom (1915-1985) and his Contributions to the Development of Artificial Disc Replacements. World Neurosurg. 2017;**98**:278-80 doi: 10.1016/j.wneu.2016.10.135[published Online First: Epub Date]|.
174. Froming N. 1974.
175. Stubstad JA, Urbaniak JR, Kahn P. 1975.
176. Fassio B, Ginestie JF. [Discal prosthesis made of silicone: experimental study and 1st clinical cases]. Nouv. Presse Med. 1978;**7**(3):207
177. Hellier WG, Hedman TP, Kostuik JP. Wear studies for development of an intervertebral disc prosthesis. Spine 1992;**17**(6 Suppl):S86-96
178. Steffee A. *The Steffee artificial disc. In Weinstein (ed). Clinical Efficacy and Outcome in the Diagnosis and Treatment of Low Back Pain*. New York: Raven Press, 1992.
179. Marnay T. *Lumbar disc replacement: 7–11 year clinical experience*. New York: Spine Solutions, Inc., 2000.
180. Bono CM, Garfin SR. History and evolution of disc replacement. The spine journal : official journal of the North American Spine Society 2004;**4**(6 Suppl):145S-50S doi: 10.1016/j.spinee.2004.07.005[published Online First: Epub Date]|.
181. Errico TJ. Lumbar disc arthroplasty. Clin. Orthop. Relat. Res. 2005(435):106-17
182. Galbusera F, Bellini CM, Zweig T, et al. Design concepts in lumbar total disc arthroplasty. Eur. Spine J. 2008;**17**(12):1635-50 doi: 10.1007/s00586-008-0811-x[published Online First: Epub Date]|.
183. Abi-Hanna D, Kerferd J, Phan K, Rao P, Mobbs R. Lumbar Disk Arthroplasty for Degenerative Disk Disease: Literature Review. World Neurosurg. 2018;**109**:188-96 doi: 10.1016/j.wneu.2017.09.153[published Online First: Epub Date]|.
184. Blumenthal S, McAfee PC, Guyer RD, et al. A prospective, randomized, multicenter Food and Drug Administration investigational device exemptions study of lumbar total disc replacement with the CHARITE artificial disc versus lumbar fusion: part I: evaluation of clinical outcomes. Spine 2005;**30**(14):1565-75; discussion E387-91
185. Zigler J, Delamarter R, Spivak JM, et al. Results of the prospective, randomized, multicenter Food and Drug Administration investigational device exemption study of the ProDisc-L total disc replacement versus circumferential fusion for the treatment of 1-level degenerative disc

disease. Spine 2007;**32**(11):1155-62; discussion 63 doi: 10.1097/BRS.0b013e318054e377[published Online First: Epub Date]|.
186. Moreno P, Boulot J. [Comparative study of short-term results between total artificial disc prosthesis and anterior lumbar interbody fusion]. Rev. Chir. Orthop. Reparatrice Appar. Mot. 2008;**94**(3):282-8 doi: 10.1016/j.rco.2007.12.011[published Online First: Epub Date]|.
187. Sasso RC, Foulk DM, Hahn M. Prospective, randomized trial of metal-on-metal artificial lumbar disc replacement: initial results for treatment of discogenic pain. Spine 2008;**33**(2):123-31 doi: 10.1097/BRS.0b013e31816043af[published Online First: Epub Date]|.
188. Berg S, Tullberg T, Branth B, Olerud C, Tropp H. Total disc replacement compared to lumbar fusion: a randomised controlled trial with 2-year follow-up. Eur. Spine J. 2009;**18**(10):1512-9 doi: 10.1007/s00586-009-1047-0[published Online First: Epub Date]|.
189. Gornet MF, Burkus JK, Dryer RF, Peloza JH. Lumbar disc arthroplasty with Maverick disc versus stand-alone interbody fusion: a prospective, randomized, controlled, multicenter investigational device exemption trial. Spine 2011;**36**(25):E1600-11 doi: 10.1097/BRS.0b013e318217668f[published Online First: Epub Date]|.
190. Guyer RD, McAfee PC, Banco RJ, et al. Prospective, randomized, multicenter Food and Drug Administration investigational device exemption study of lumbar total disc replacement with the CHARITE artificial disc versus lumbar fusion: five-year follow-up. The spine journal : official journal of the North American Spine Society 2009;**9**(5):374-86 doi: 10.1016/j.spinee.2008.08.007[published Online First: Epub Date]|.
191. Zigler JE, Delamarter RB. Five-year results of the prospective, randomized, multicenter, Food and Drug Administration investigational device exemption study of the ProDisc-L total disc replacement versus circumferential arthrodesis for the treatment of single-level degenerative disc disease. J. Neurosurg. Spine 2012;**17**(6):493-501 doi: 10.3171/2012.9.spine11498[published Online First: Epub Date]|.
192. Skold C, Tropp H, Berg S. Five-year follow-up of total disc replacement compared to fusion: a randomized controlled trial. Eur. Spine J. 2013 doi: 10.1007/s00586-013-2926-y[published Online First: Epub Date]|.
193. Siepe CJ, Heider F, Wiechert K, Hitzl W, Ishak B, Mayer MH. Mid- to long-term results of total lumbar disc replacement: a prospective analysis with 5- to 10-year follow-up. The spine journal : official journal of the North American Spine Society 2014;**14**(8):1417-31 doi: 10.1016/j.spinee.2013.08.028[published Online First: Epub Date]|.
194. Guyer RD, Pettine K, Roh JS, et al. Five-Year Follow-Up of a Prospective, Randomized Trial Comparing Two Lumbar Total Disc Replacements. Spine 2016;**41**(1):3-8 doi: 10.1097/brs.0000000000001168[published Online First: Epub Date]|.
195. Laugesen LA, Paulsen RT, Carreon L, Ernst C, Andersen MO. Patient-reported Outcomes and Revision Rates at a Mean Follow-up of 10 Years After Lumbar Total Disc Replacement. Spine 2017;**42**(21):1657-63 doi: 10.1097/brs.0000000000002174[published Online First: Epub Date]|.
196. Meir AR, Freeman BJ, Fraser RD, Fowler SM. Ten-year survival and clinical outcome of the AcroFlex lumbar disc replacement for the treatment of symptomatic disc degeneration. The spine journal : official journal of the North American Spine Society 2013;**13**(1):13-21 doi: 10.1016/j.spinee.2012.12.008[published Online First: Epub Date]|.
197. Aghayev E, Etter C, Barlocher C, et al. Five-year results of lumbar disc prostheses in the SWISSspine registry. Eur. Spine J. 2014;**23**(10):2114-26 doi: 10.1007/s00586-014-3418-4[published Online First: Epub Date]|.
198. Lu SB, Hai Y, Kong C, et al. An 11-year minimum follow-up of the Charite III lumbar disc replacement for the treatment of symptomatic degenerative disc disease. Eur. Spine J. 2015;**24**(9):2056-64 doi: 10.1007/s00586-015-3939-5[published Online First: Epub Date]|.
199. Wuertinger C, Annes RDA, Hitzl W, Siepe CJ. Motion preservation following total lumbar disc replacement at the lumbosacral junction: a prospective long-term clinical and radiographic

investigation. The spine journal : official journal of the North American Spine Society 2018;**18**(1):72-80 doi: 10.1016/j.spinee.2017.06.035[published Online First: Epub Date]|.
200. Lazennec JY, Rakover JP, Rousseau MA. Five-year follow-up of clinical and radiological outcomes of LP-ESP elastomeric lumbar total disc replacement in active patients. The spine journal : official journal of the North American Spine Society 2018 doi: 10.1016/j.spinee.2018.05.023[published Online First: Epub Date]|.
201. Lemaire JP, Carrier H, Sariali el H, Skalli W, Lavaste F. Clinical and radiological outcomes with the Charite artificial disc: a 10-year minimum follow-up. Journal of spinal disorders & techniques 2005;**18**(4):353-9
202. Putzier M, Funk JF, Schneider SV, et al. Charite total disc replacement--clinical and radiographical results after an average follow-up of 17 years. Eur. Spine J. 2006;**15**(2):183-95 doi: 10.1007/s00586-005-1022-3[published Online First: Epub Date]|.
203. David T. Long-term results of one-level lumbar arthroplasty: minimum 10-year follow-up of the CHARITE artificial disc in 106 patients. Spine 2007;**32**(6):661-6 doi: 10.1097/01.brs.0000257554.67505.45[published Online First: Epub Date]|.
204. Lee WT, Liu G, Thambiah J, Wong HK. Clinical outcomes of single-level lumbar artificial disc replacement compared with transforaminal lumbar interbody fusion in an Asian population. Singapore Med. J. 2015;**56**(4):208-11 doi: 10.11622/smedj.2015032[published Online First: Epub Date]|.
205. Park SJ, Lee CS, Chung SS, Lee KH, Kim WS, Lee JY. Long-Term Outcomes Following Lumbar Total Disc Replacement Using ProDisc-II: Average 10-Year Follow-Up at a Single Institute. Spine 2016;**41**(11):971-7 doi: 10.1097/brs.0000000000001527[published Online First: Epub Date]|.
206. Formica M, Divano S, Cavagnaro L, et al. Lumbar total disc arthroplasty: outdated surgery or here to stay procedure? A systematic review of current literature. J. Orthop. Traumatol. 2017;**18**(3):197-215 doi: 10.1007/s10195-017-0462-y[published Online First: Epub Date]|.
207. van den Eerenbeemt KD, Ostelo RW, van Royen BJ, Peul WC, van Tulder MW. Total disc replacement surgery for symptomatic degenerative lumbar disc disease: a systematic review of the literature. Eur. Spine J. 2010;**19**(8):1262-80 doi: 10.1007/s00586-010-1445-3[published Online First: Epub Date]|.
208. Siepe CJ, Mayer HM, Heinz-Leisenheimer M, Korge A. Total lumbar disc replacement: different results for different levels. Spine 2007;**32**(7):782-90 doi: 10.1097/01.brs.0000259071.64027.04[published Online First: Epub Date]|.
209. Eliasberg CD, Kelly MP, Ajiboye RM, SooHoo NF. Complications and Rates of Subsequent Lumbar Surgery Following Lumbar Total Disc Arthroplasty and Lumbar Fusion. Spine 2016;**41**(2):173-81 doi: 10.1097/brs.0000000000001180[published Online First: Epub Date]|.
210. Park HJ, Lee CS, Chung SS, et al. Radiological and clinical long-term results of heterotopic ossification following lumbar total disc replacement. The spine journal : official journal of the North American Spine Society 2017 doi: 10.1016/j.spinee.2017.09.003[published Online First: Epub Date]|.
211. McAfee PC, Cunningham BW, Devine J, Williams E, Yu-Yahiro J. Classification of heterotopic ossification (HO) in artificial disk replacement. Journal of spinal disorders & techniques 2003;**16**(4):384-9
212. Johnsen LG, Brinckmann P, Hellum C, Rossvoll I, Leivseth G. Segmental mobility, disc height and patient-reported outcomes after surgery for degenerative disc disease: a prospective randomised trial comparing disc replacement and multidisciplinary rehabilitation. The bone & joint journal 2013;**95-B**(1):81-9 doi: 10.1302/0301-620x.95b1.29829[published Online First: Epub Date]|.
213. Liow MHL, Kwon YM. Metal-on-metal total hip arthroplasty: risk factors for pseudotumours and clinical systematic evaluation. Int. Orthop. 2017;**41**(5):885-92 doi: 10.1007/s00264-016-3305-1[published Online First: Epub Date]|.

214. Bisseling P, Zeilstra DJ, Hol AM, van Susante JL. Metal ion levels in patients with a lumbar metal-on-metal total disc replacement: should we be concerned? J. Bone Joint Surg. Br. 2011;**93**(7):949-54 doi: 10.1302/0301-620x.93b7.26392[published Online First: Epub Date]|.
215. Gornet MF, Burkus JK, Harper ML, Chan FW, Skipor AK, Jacobs JJ. Prospective study on serum metal levels in patients with metal-on-metal lumbar disc arthroplasty. Eur. Spine J. 2013;**22**(4):741-6 doi: 10.1007/s00586-012-2581-8[published Online First: Epub Date]|.
216. Guyer RD, Shellock J, MacLennan B, et al. Early failure of metal-on-metal artificial disc prostheses associated with lymphocytic reaction: diagnosis and treatment experience in four cases. Spine 2011;**36**(7):E492-7 doi: 10.1097/BRS.0b013e31820ea9a2[published Online First: Epub Date]|.
217. Berry MR, Peterson BG, Alander DH. A granulomatous mass surrounding a Maverick total disc replacement causing iliac vein occlusion and spinal stenosis: a case report. J. Bone Joint Surg. Am. 2010;**92**(5):1242-5 doi: 10.2106/jbjs.h.01625[published Online First: Epub Date]|.
218. Golish SR, Anderson PA. Bearing surfaces for total disc arthroplasty: metal-on-metal versus metal-on-polyethylene and other biomaterials. The spine journal : official journal of the North American Spine Society 2012;**12**(8):693-701 doi: 10.1016/j.spinee.2011.05.008[published Online First: Epub Date]|.
219. Veruva SY, Lanman TH, Isaza JE, Freeman TA, Kurtz SM, Steinbeck MJ. Periprosthetic UHMWPE Wear Debris Induces Inflammation, Vascularization, and Innervation After Total Disc Replacement in the Lumbar Spine. Clin. Orthop. Relat. Res. 2017;**475**(5):1369-81 doi: 10.1007/s11999-016-4996-8[published Online First: Epub Date]|.
220. Punt IM, Cleutjens JP, de Bruin T, et al. Periprosthetic tissue reactions observed at revision of total intervertebral disc arthroplasty. Biomaterials 2009;**30**(11):2079-84 doi: 10.1016/j.biomaterials.2008.12.071[published Online First: Epub Date]|.
221. Baxter RM, Macdonald DW, Kurtz SM, Steinbeck MJ. Severe impingement of lumbar disc replacements increases the functional biological activity of polyethylene wear debris. J. Bone Joint Surg. Am. 2013;**95**(11):e751-9 doi: 10.2106/jbjs.k.00522[published Online First: Epub Date]|.
222. Veruva SY, Steinbeck MJ, Toth J, Alexander DD, Kurtz SM. Which design and biomaterial factors affect clinical wear performance of total disc replacements? A systematic review. Clin. Orthop. Relat. Res. 2014;**472**(12):3759-69 doi: 10.1007/s11999-014-3751-2[published Online First: Epub Date]|.
223. Janssen M, Garcia R, Miller L, et al. Challenges and Solutions for Lumbar Total Disc Replacement Implantation. Spine 2017;**42 Suppl 24**:S108-S11 doi: 10.1097/brs.0000000000002454[published Online First: Epub Date]|.
224. Zigler JE, Glenn J, Delamarter RB. Five-year adjacent-level degenerative changes in patients with single-level disease treated using lumbar total disc replacement with ProDisc-L versus circumferential fusion. J. Neurosurg. Spine 2012;**17**(6):504-11 doi: 10.3171/2012.9.spine11717[published Online First: Epub Date]|.
225. Song KJ, Choi BW, Jeon TS, Lee KB, Chang H. Adjacent segment degenerative disease: is it due to disease progression or a fusion-associated phenomenon? Comparison between segments adjacent to the fused and non-fused segments. Eur. Spine J. 2011;**20**(11):1940-5 doi: 10.1007/s00586-011-1864-9[published Online First: Epub Date]|.
226. Huang RC, Tropiano P, Marnay T, Girardi FP, Lim MR, Cammisa FP, Jr. Range of motion and adjacent level degeneration after lumbar total disc replacement. The spine journal : official journal of the North American Spine Society 2006;**6**(3):242-7 doi: 10.1016/j.spinee.2005.04.013[published Online First: Epub Date]|.
227. Rohlmann A, Lauterborn S, Dreischarf M, et al. Parameters influencing the outcome after total disc replacement at the lumbosacral junction. Part 1: misalignment of the vertebrae adjacent to a total disc replacement affects the facet joint and facet capsule forces in a probabilistic finite element analysis. Eur. Spine J. 2013;**22**(10):2271-8 doi: 10.1007/s00586-013-2909-z[published Online First: Epub Date]|.

228. Rundell SA, Auerbach JD, Balderston RA, Kurtz SM. Total disc replacement positioning affects facet contact forces and vertebral body strains. Spine 2008;**33**(23):2510-7 doi: 10.1097/BRS.0b013e318186b258[published Online First: Epub Date]|.
229. Zander T, Rohlmann A, Bergmann G. Influence of different artificial disc kinematics on spine biomechanics. Clin. Biomech. (Bristol, Avon) 2009;**24**(2):135-42 doi: 10.1016/j.clinbiomech.2008.11.008[published Online First: Epub Date]|.
230. Siepe CJ, Zelenkov P, Sauri-Barraza JC, et al. The fate of facet joint and adjacent level disc degeneration following total lumbar disc replacement: a prospective clinical, X-ray, and magnetic resonance imaging investigation. Spine 2010;**35**(22):1991-2003 doi: 10.1097/BRS.0b013e3181d6f878[published Online First: Epub Date]|.
231. Park CK, Ryu KS, Jee WH. Degenerative changes of discs and facet joints in lumbar total disc replacement using ProDisc II: minimum two-year follow-up. Spine 2008;**33**(16):1755-61 doi: 10.1097/BRS.0b013e31817b8fed[published Online First: Epub Date]|.
232. Shin MH, Ryu KS, Hur JW, Kim JS, Park CK. Association of facet tropism and progressive facet arthrosis after lumbar total disc replacement using ProDisc-L. Eur. Spine J. 2013;**22**(8):1717-22 doi: 10.1007/s00586-012-2606-3[published Online First: Epub Date]|.
233. Bombardier C. Outcome assessments in the evaluation of treatment of spinal disorders: summary and general recommendations. Spine 2000;**25**(24):3100-3
234. Chapman JR, Norvell DC, Hermsmeyer JT, et al. Evaluating common outcomes for measuring treatment success for chronic low back pain. Spine 2011;**36**(21 Suppl):S54-68 doi: 10.1097/BRS.0b013e31822ef74d[published Online First: Epub Date]|.
235. Chiarotto A, Terwee CB, Ostelo RW. Choosing the right outcome measurement instruments for patients with low back pain. Best Pract. Res. Clin. Rheumatol. 2016;**30**(6):1003-20 doi: 10.1016/j.berh.2017.07.001[published Online First: Epub Date]|.
236. Fairbank JC, Couper J, Davies JB, O'Brien JP. The Oswestry low back pain disability questionnaire. Physiotherapy 1980;**66**(8):271-3
237. Fairbank JC, Pynsent PB. The Oswestry Disability Index. Spine 2000;**25**(22):2940-52; discussion 52
238. Grotle M, Brox JI, Vollestad NK. Cross-cultural adaptation of the Norwegian versions of the Roland-Morris Disability Questionnaire and the Oswestry Disability Index. J. Rehabil. Med. 2003;**35**(5):241-7
239. McCormack HM, Horne DJ, Sheather S. Clinical applications of visual analogue scales: a critical review. Psychol. Med. 1988;**18**(4):1007-19
240. EuroQol--a new facility for the measurement of health-related quality of life. Health Policy 1990;**16**(3):199-208
241. Hansson T, Hansson E, Malchau H. Utility of spine surgery: a comparison of common elective orthopaedic surgical procedures. Spine 2008;**33**(25):2819-30 doi: 10.1097/BRS.0b013e31818e2914[published Online First: Epub Date]|.
242. Derogatis LR, Lipman RS, Rickels K, Uhlenhuth EH, Covi L. The Hopkins Symptom Checklist (HSCL): a self-report symptom inventory. Behav. Sci. 1974;**19**(1):1-15
243. Ostelo RW, de Vet HC. Clinically important outcomes in low back pain. Best Pract. Res. Clin. Rheumatol. 2005;**19**(4):593-607 doi: 10.1016/j.berh.2005.03.003[published Online First: Epub Date]|.
244. Marshman LA, Strong G, Trewhella M, Kasis A, Friesem T. Minimizing ferromagnetic artefact with metallic lumbar total disc arthroplasty devices at adjacent segments: technical note. Spine 2010;**35**(2):252-6 doi: 10.1097/BRS.0b013e3181c838cc[published Online First: Epub Date]|.
245. Altman D. *Practical statistics for medical research*. New York: Chapman & Hall, 1991.
246. Jensen TS, Sorensen JS, Kjaer P. Intra- and interobserver reproducibility of vertebral endplate signal (modic) changes in the lumbar spine: the Nordic Modic Consensus Group classification. Acta Radiol. 2007;**48**(7):748-54 doi: 10.1080/02841850701422112[published Online First: Epub Date]|.

247. Statistics Norway (1984) Standard Classifications of Socioeconomic Status.
248. Statistics Norway (1998) Standard Classification of Occupations.
249. Waddell G, Newton M, Henderson I, Somerville D, Main CJ. A Fear-Avoidance Beliefs Questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. Pain 1993;**52**(2):157-68
250. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med. Care 1992;**30**(6):473-83
251. Uden A, Astrom M, Bergenudd H. Pain drawings in chronic back pain. Spine 1988;**13**(4):389-92
252. Legaye J, Duval-Beaupere G, Hecquet J, Marty C. Pelvic incidence: a fundamental pelvic parameter for three-dimensional regulation of spinal sagittal curves. Eur. Spine J. 1998;**7**(2):99-103
253. Fujiwara A, Tamai K, Yamato M, et al. The relationship between facet joint osteoarthritis and disc degeneration of the lumbar spine: an MRI study. Eur. Spine J. 1999;**8**(5):396-401
254. Berg L, Neckelmann G, Gjertsen O, et al. Reliability of MRI findings in candidates for lumbar disc prosthesis. Neuroradiology 2012;**54**(7):699-707 doi: 10.1007/s00234-011-0963-y[published Online First: Epub Date]|.
255. Hagg 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 doi: 10.1007/s00586-002-0464-0[published Online First: Epub Date]|.
256. Hellum C, Johnsen LG, Gjertsen O, et al. Predictors of outcome after surgery with disc prosthesis and rehabilitation in patients with chronic low back pain and degenerative disc: 2-year follow-up. Eur. Spine J. 2012;**21**(4):681-90 doi: 10.1007/s00586-011-2145-3[published Online First: Epub Date]|.
257. Geisser S. *Predictive Inference.* New York, NY: Chapman and Hall, 1993.
258. van Hooff ML, Mannion AF, Staub LP, Ostelo RW, Fairbank JC. Determination of the Oswestry Disability Index score equivalent to a "satisfactory symptom state" in patients undergoing surgery for degenerative disorders of the lumbar spine-a Spine Tango registry-based study. The spine journal : official journal of the North American Spine Society 2016;**16**(10):1221-30 doi: 10.1016/j.spinee.2016.06.010[published Online First: Epub Date]|.
259. Furlan AD, Malmivaara A, Chou R, et al. 2015 Updated Method Guideline for Systematic Reviews in the Cochrane Back and Neck Group. Spine 2015;**40**(21):1660-73 doi: 10.1097/brs.0000000000001061[published Online First: Epub Date]|.
260. Pocock SJ. *Clinical Trials: A Practical Approach*: John Wiley & Sons Ltd. , 2013.
261. Buchbinder R, Golmohammadi K, Johnston RV, et al. Percutaneous vertebroplasty for osteoporotic vertebral compression fracture. Cochrane database of systematic reviews (Online) 2015(4):CD006349 doi: 10.1002/14651858.CD006349.pub2[published Online First: Epub Date]|.
262. Fischgrund JS, Rhyne A, Franke J, et al. Intraosseous basivertebral nerve ablation for the treatment of chronic low back pain: a prospective randomized double-blind sham-controlled multi-center study. Eur. Spine J. 2018 doi: 10.1007/s00586-018-5496-1[published Online First: Epub Date]|.
263. Mirza SK, Deyo RA, Heagerty PJ, Turner JA, Martin BI, Comstock BA. One-year outcomes of surgical versus nonsurgical treatments for discogenic back pain: a community-based prospective cohort study. The spine journal : official journal of the North American Spine Society 2013;**13**(11):1421-33 doi: 10.1016/j.spinee.2013.05.047[published Online First: Epub Date]|.
264. Mayer M. Reviewer's comment concerning "Lumbar total disc replacement: predictors for 8-year outcome" (by H. Furunes et al. [2017] Eur Spine J). Eur. Spine J. 2017 doi: 10.1007/s00586-017-5380-4[published Online First: Epub Date]|.
265. Furunes H, Storheim K, Brox JI, et al. Total disc replacement versus multidisciplinary rehabilitation in patients with chronic low back pain and degenerative discs: 8-year follow-up of a randomized controlled multicenter trial. The spine journal : official journal of the North

American Spine Society 2017;**17**(10):1480-88 doi: 10.1016/j.spinee.2017.05.011[published Online First: Epub Date]|.
266. Nunan D, Aronson J, Bankhead C. Catalogue of bias: attrition bias. BMJ evidence-based medicine 2018;**23**(1):21-22 doi: 10.1136/ebmed-2017-110883[published Online First: Epub Date]|.
267. Juni P, Egger M. Commentary: Empirical evidence of attrition bias in clinical trials. Int. J. Epidemiol. 2005;**34**(1):87-8 doi: 10.1093/ije/dyh406[published Online First: Epub Date]|.
268. Lurie JD, Tosteson TD, Tosteson AN, et al. Surgical versus nonoperative treatment for lumbar disc herniation: eight-year results for the spine patient outcomes research trial. Spine 2014;**39**(1):3-16 doi: 10.1097/brs.0000000000000088[published Online First: Epub Date]|.
269. Lurie JD, Tosteson TD, Tosteson A, et al. Long-term outcomes of lumbar spinal stenosis: eight-year results of the Spine Patient Outcomes Research Trial (SPORT). Spine 2015;**40**(2):63-76 doi: 10.1097/brs.0000000000000731[published Online First: Epub Date]|.
270. Geisler FH, Guyer RD, Blumenthal SL, et al. Patient selection for lumbar arthroplasty and arthrodesis: the effect of revision surgery in a controlled, multicenter, randomized study. J. Neurosurg. Spine 2008;**8**(1):13-6 doi: 10.3171/spi-08/01/013[published Online First: Epub Date]|.
271. Gluud LL. Bias in clinical intervention research. Am. J. Epidemiol. 2006;**163**(6):493-501 doi: 10.1093/aje/kwj069[published Online First: Epub Date]|.
272. Solberg TK, Sorlie A, Sjaavik K, Nygaard OP, Ingebrigtsen T. Would loss to follow-up bias the outcome evaluation of patients operated for degenerative disorders of the lumbar spine? Acta Orthop. 2011;**82**(1):56-63 doi: 10.3109/17453674.2010.548024[published Online First: Epub Date]|.
273. Fewtrell MS, Kennedy K, Singhal A, et al. How much loss to follow-up is acceptable in long-term randomised trials and prospective studies? Arch. Dis. Child. 2008;**93**(6):458-61 doi: 10.1136/adc.2007.127316[published Online First: Epub Date]|.
274. Somerson JS, Bartush KC, Shroff JB, Bhandari M, Zelle BA. Loss to follow-up in orthopaedic clinical trials: a systematic review. Int. Orthop. 2016;**40**(11):2213-19 doi: 10.1007/s00264-016-3212-5[published Online First: Epub Date]|.
275. Hrobjartsson A, Thomsen AS, Emanuelsson F, et al. Observer bias in randomised clinical trials with binary outcomes: systematic review of trials with both blinded and non-blinded outcome assessors. BMJ 2012;**344**:e1119 doi: 10.1136/bmj.e1119[published Online First: Epub Date]|.
276. Poolman RW, Struijs PA, Krips R, et al. Reporting of outcomes in orthopaedic randomized trials: does blinding of outcome assessors matter? J. Bone Joint Surg. Am. 2007;**89**(3):550-8 doi: 10.2106/jbjs.f.00683[published Online First: Epub Date]|.
277. Every-Palmer S, Howick J. How evidence-based medicine is failing due to biased trials and selective publication. J. Eval. Clin. Pract. 2014;**20**(6):908-14 doi: 10.1111/jep.12147[published Online First: Epub Date]|.
278. Rongen JJ, Hannink G. Comparison of Registered and Published Primary Outcomes in Randomized Controlled Trials of Orthopaedic Surgical Interventions. J. Bone Joint Surg. Am. 2016;**98**(5):403-9 doi: 10.2106/jbjs.15.00400[published Online First: Epub Date]|.
279. Schwind J, Learman K, O'Halloran B, Showalter C, Cook C. Different minimally important clinical difference (MCID) scores lead to different clinical prediction rules for the Oswestry disability index for the same sample of patients. J. Man. Manip. Ther. 2013;**21**(2):71-8 doi: 10.1179/2042618613y.0000000028[published Online First: Epub Date]|.
280. Taylor SJ, Taylor AE, Foy MA, Fogg AJ. Responsiveness of common outcome measures for patients with low back pain. Spine 1999;**24**(17):1805-12
281. Glassman SD, Copay AG, Berven SH, Polly DW, Subach BR, Carreon LY. Defining substantial clinical benefit following lumbar spine arthrodesis. J. Bone Joint Surg. Am. 2008;**90**(9):1839-47 doi: 10.2106/jbjs.g.01095[published Online First: Epub Date]|.
282. Johnsen LG, Hellum C, Nygaard OP, et al. Comparison of the SF6D, the EQ5D, and the oswestry disability index in patients with chronic low back pain and degenerative disc disease. BMC

Musculoskelet. Disord. 2013;**14**:148 doi: 10.1186/1471-2474-14-148[published Online First: Epub Date]|.
283. Furunes H, Hellum C, Brox JI, et al. Lumbar total disc replacement: predictors for long-term outcome. Eur. Spine J. 2017 doi: 10.1007/s00586-017-5375-1[published Online First: Epub Date]|.
284. Grotle M, Brox JI, Vollestad NK. Concurrent comparison of responsiveness in pain and functional status measurements used for patients with low back pain. Spine 2004;**29**(21):E492-501
285. Ostelo RW, Deyo RA, 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):90-4 doi: 10.1097/BRS.0b013e31815e3a10[published Online First: Epub Date]|.
286. Solberg TK, Olsen JA, Ingebrigtsen T, Hofoss D, Nygaard OP. Health-related quality of life assessment by the EuroQol-5D can provide cost-utility data in the field of low-back surgery. Eur. Spine J. 2005;**14**(10):1000-7 doi: 10.1007/s00586-005-0898-2[published Online First: Epub Date]|.
287. Holm I, Friis A, Storheim K, Brox JI. Measuring self-reported functional status and pain in patients with chronic low back pain by postal questionnaires: a reliability study. Spine 2003;**28**(8):828-33
288. Grovle L, Haugen AJ, Keller A, Natvig B, Brox JI, Grotle M. Poor agreement found between self-report and a public registry on duration of sickness absence. J. Clin. Epidemiol. 2012;**65**(2):212-8 doi: 10.1016/j.jclinepi.2011.05.009[published Online First: Epub Date]|.
289. Gerometta A, Rodriguez Olaverri JC, Bittan F. Infection and revision strategies in total disc arthroplasty. Int. Orthop. 2012;**36**(2):471-4 doi: 10.1007/s00264-011-1436-y[published Online First: Epub Date]|.
290. Neal CJ, Rosner MK, Kuklo TR. Magnetic resonance imaging evaluation of adjacent segments after disc arthroplasty. J. Neurosurg. Spine 2005;**3**(5):342-7 doi: 10.3171/spi.2005.3.5.0342[published Online First: Epub Date]|.
291. Yang CW, Liu L, Wang J, et al. Magnetic resonance imaging of artificial lumbar disks: safety and metal artifacts. Chin. Med. J. (Engl.) 2009;**122**(8):911-6
292. Berg L, Gjertsen O, Hellum C, et al. Reliability of change in lumbar MRI findings over time in patients with and without disc prosthesis--comparing two different image evaluation methods. Skeletal Radiol. 2012;**41**(12):1547-57 doi: 10.1007/s00256-012-1394-5[published Online First: Epub Date]|.
293. Farshad-Amacker NA, Farshad M, Winklehner A, Andreisek G. MR imaging of degenerative disc disease. Eur. J. Radiol. 2015;**84**(9):1768-76 doi: 10.1016/j.ejrad.2015.04.002[published Online First: Epub Date]|.
294. Tarantino U, Fanucci E, Iundusi R, et al. Lumbar spine MRI in upright position for diagnosing acute and chronic low back pain: statistical analysis of morphological changes. J. Orthop. Traumatol. 2013;**14**(1):15-22 doi: 10.1007/s10195-012-0213-z[published Online First: Epub Date]|.
295. Deyo RA, Von Korff M, Duhrkoop D. Opioids for low back pain. BMJ 2015;**350**:g6380 doi: 10.1136/bmj.g6380[published Online First: Epub Date]|.
296. Smith JA, Fuino RL, Pesis-Katz I, et al. Differences in opioid prescribing in low back pain patients with and without depression: a cross-sectional study of a national sample from the United States. Pain reports 2017;**2**(4):e606 doi: 10.1097/pr9.0000000000000606[published Online First: Epub Date]|.
297. Turk DC, Okifuji A. What factors affect physicians' decisions to prescribe opioids for chronic noncancer pain patients? Clin. J. Pain 1997;**13**(4):330-6
298. Hosmer DW, Lemeshow S. *Applied logistic regression*. 2nd Edition ed. New York: John Wiley & Sons, Inc., 2005.
299. McAfee PC, Geisler FH, Saiedy SS, et al. Revisability of the CHARITE artificial disc replacement: analysis of 688 patients enrolled in the U.S. IDE study of the CHARITE Artificial Disc. Spine

2006;**31**(11):1217-26 doi: 10.1097/01.brs.0000217689.08487.a8[published Online First: Epub Date]|.
300. OECD. OECD Economic surveys Norway. Paris, 2007.
301. Lu S, Sun S, Kong C, et al. Long-term clinical results following Charite III lumbar total disc replacement. The spine journal : official journal of the North American Spine Society 2017 doi: 10.1016/j.spinee.2017.08.252[published Online First: Epub Date]|.
302. Albert HB, Manniche C. Modic changes following lumbar disc herniation. Eur. Spine J. 2007;**16**(7):977-82 doi: 10.1007/s00586-007-0336-8[published Online First: Epub Date]|.
303. Yang X, Zhang Q, Hao X, Guo X, Wang L. Spontaneous regression of herniated lumbar discs: Report of one illustrative case and review of the literature. Clin. Neurol. Neurosurg. 2016;**143**:86-9 doi: 10.1016/j.clineuro.2016.02.020[published Online First: Epub Date]|.
304. Laustsen AF, Bech-Azeddine R. Do Modic changes have an impact on clinical outcome in lumbar spine surgery? A systematic literature review. Eur. Spine J. 2016 doi: 10.1007/s00586-016-4609-y[published Online First: Epub Date]|.
305. Gornet MF, Schranck F, Wharton ND, et al. Optimizing success with lumbar disc arthroplasty. Eur. Spine J. 2014;**23**(10):2127-35 doi: 10.1007/s00586-014-3309-8[published Online First: Epub Date]|.
306. Cougot B, Petit A, Paget C, et al. Chronic low back pain among French healthcare workers and prognostic factors of return to work (RTW): a non-randomized controlled trial. J. Occup. Med. Toxicol. 2015;**10**:40 doi: 10.1186/s12995-015-0082-5[published Online First: Epub Date]|.
307. Grovle L, Haugen AJ, Keller A, Natvig B, Brox JI, Grotle M. The bothersomeness of sciatica: patients' self-report of paresthesia, weakness and leg pain. Eur. Spine J. 2010;**19**(2):263-9 doi: 10.1007/s00586-009-1042-5[published Online First: Epub Date]|.
308. McGirt MJ, Sivaganesan A, Asher AL, Devin CJ. Prediction model for outcome after low-back surgery: individualized likelihood of complication, hospital readmission, return to work, and 12-month improvement in functional disability. Neurosurg. Focus 2015;**39**(6):E13 doi: 10.3171/2015.8.focus15338[published Online First: Epub Date]|.
309. Nordeman L, Gunnarsson R, Mannerkorpi K. Prognostic factors for work ability in women with chronic low back pain consulting primary health care: a 2-year prospective longitudinal cohort study. Clin. J. Pain 2014;**30**(5):391-8 doi: 10.1097/AJP.0b013e3182a0dd06[published Online First: Epub Date]|.
310. Arunakul R, Metzger M, Kanim L, Bae H, Kropf M, Delamarter R. Radiographic Analysis of the Lumbosacral Juncture: Is There a Critical Sacral Angle for Total Disc Replacement? Asian spine journal 2017;**11**(2):249-55 doi: 10.4184/asj.2017.11.2.249[published Online First: Epub Date]|.
311. Laouissat F, Scemama C, Delecrin J. Does the type of sagittal spinal shape influence the clinical results of lumbar disc arthroplasty? Orthop. Traumatol. Surg. Res. 2016;**102**(6):765-8 doi: 10.1016/j.otsr.2016.05.012[published Online First: Epub Date]|.
312. Roussouly P, Gollogly S, Berthonnaud E, Dimnet J. Classification of the normal variation in the sagittal alignment of the human lumbar spine and pelvis in the standing position. Spine 2005;**30**(3):346-53
313. Staub LP, Ryser C, Roder C, et al. Total disc arthroplasty versus anterior cervical interbody fusion: use of the Spine Tango registry to supplement the evidence from randomized control trials. The spine journal : official journal of the North American Spine Society 2016;**16**(2):136-45 doi: 10.1016/j.spinee.2015.11.056[published Online First: Epub Date]|.
314. Rognsvag T, Strand LI, Skouen JS. [Pain and functional ability in patients with total disc replacement in the lumbar spine]. Tidsskr. Nor. Laegeforen. 2007;**127**(9):1188-91
315. Furunes H, Pedersen A, Skoglund A. Kliniske resultater etter innsettelse av lumbale skiveproteser ved SI Elverum. Ortopedisk Høstmøte, 2011.
316. Furunes H, Hellum C, Brox JI, Solberg TK, storheim K. Lumbal skiveprotesekirurgi i Norge 2007-2013. Ett års resultater fra Ryggregisteret. Ortopedisk Høstmøte, 2014.

317. Colagiuri B, Schenk LA, Kessler MD, Dorsey SG, Colloca L. The placebo effect: From concepts to genes. Neuroscience 2015;**307**:171-90 doi: 10.1016/j.neuroscience.2015.08.017[published Online First: Epub Date]|.
318. Beecher HK. The powerful placebo. J. Am. Med. Assoc. 1955;**159**(17):1602-6
319. Freeman BJ, Fraser RD, Cain CM, Hall DJ, Chapple DC. A randomized, double-blind, controlled trial: intradiscal electrothermal therapy versus placebo for the treatment of chronic discogenic low back pain. Spine 2005;**30**(21):2369-77; discussion 78
320. Ihlebaek C, Laerum E. [Hits most, costs most and gets least]. Tidsskr. Nor. Laegeforen. 2010;**130**(21):2106 doi: 10.4045/tidsskr.10.1035[published Online First: Epub Date]|.
321. Peng B, Fu X, Pang X, et al. Prospective clinical study on natural history of discogenic low back pain at 4 years of follow-up. Pain physician 2012;**15**(6):525-32
322. Yoshihara H, Yoneoka D. National trends in the surgical treatment for lumbar degenerative disc disease: United States, 2000 to 2009. The spine journal : official journal of the North American Spine Society 2015;**15**(2):265-71 doi: 10.1016/j.spinee.2014.09.026[published Online First: Epub Date]|.
323. Saifi C, Cazzulino A, Park C, et al. National Trends for Primary and Revision Lumbar Disc Arthroplasty Throughout the United States. Global spine journal 2018;**8**(2):172-77 doi: 10.1177/2192568217726980[published Online First: Epub Date]|.
324. Sandhu F, Blumenthal S, Grunch B, Kimball B, Ferko N, Hollmann S. Barriers to and Budget Impact of Lumbar Total Disc Replacement Utilization. Spine 2017;**42 Suppl 24**:S112-S14 doi: 10.1097/brs.0000000000002455[published Online First: Epub Date]|.
325. Blumenthal S, Buttermann G, Garcia R, et al. Foreword: Proceedings From the First Annual Lumbar Total Disc Replacement Summit. Spine 2017;**42 Suppl 24**:S99-S100 doi: 10.1097/brs.0000000000002451[published Online First: Epub Date]|.
326. Gornet M, Buttermann G, Guyer R, Yue J, Ferko N, Hollmann S. Defining the Ideal Lumbar Total Disc Replacement Patient and Standard of Care. Spine 2017;**42 Suppl 24**:S103-S07 doi: 10.1097/brs.0000000000002453[published Online First: Epub Date]|.
327. Gallo J, Vaculova J, Goodman SB, Konttinen YT, Thyssen JP. Contributions of human tissue analysis to understanding the mechanisms of loosening and osteolysis in total hip replacement. Acta biomaterialia 2014;**10**(6):2354-66 doi: 10.1016/j.actbio.2014.02.003[published Online First: Epub Date]|.
328. Nine MJ, Choudhury D, Hee AC, Mootanah R, Osman NAA. Wear Debris Characterization and Corresponding Biological Response: Artificial Hip and Knee Joints. Materials (Basel, Switzerland) 2014;**7**(2):980-1016 doi: 10.3390/ma7020980[published Online First: Epub Date]|.
329. Catelas I, Wimmer MA, Utzschneider S. Polyethylene and metal wear particles: characteristics and biological effects. Semin. Immunopathol. 2011;**33**(3):257-71 doi: 10.1007/s00281-011-0242-3[published Online First: Epub Date]|.
330. Cakir B, Schmidt R, Mattes T, Fraitzl CR, Reichel H, Kafer W. Index level mobility after total lumbar disc replacement: is it beneficial or detrimental? Spine 2009;**34**(9):917-23 doi: 10.1097/BRS.0b013e31819b213c[published Online First: Epub Date]|.
331. Berg S, Tropp HT, Leivseth G. Disc height and motion patterns in the lumbar spine in patients operated with total disc replacement or fusion for discogenic back pain. Results from a randomized controlled trial. The spine journal : official journal of the North American Spine Society 2011;**11**(11):991-8 doi: 10.1016/j.spinee.2011.08.434[published Online First: Epub Date]|.
332. Leivseth G, Brinckmann P, Frobin W, Johnsson R, Stromqvist B. Assessment of sagittal plane segmental motion in the lumbar spine. A comparison between distortion-compensated and stereophotogrammetric roentgen analysis. Spine 1998;**23**(23):2648-55
333. Fritzell P, Berg S, Borgstrom F, Tullberg T, Tropp H. Cost effectiveness of disc prosthesis versus lumbar fusion in patients with chronic low back pain: randomized controlled trial with 2-year

follow-up. Eur. Spine J. 2011;**20**(7):1001-11 doi: 10.1007/s00586-010-1607-3[published Online First: Epub Date]|.
334. Kurtz SM, Lau E, Ianuzzi A, et al. National revision burden for lumbar total disc replacement in the United States: epidemiologic and economic perspectives. Spine 2010;**35**(6):690-6 doi: 10.1097/BRS.0b013e3181d0fabb[published Online First: Epub Date]|.
335. Johnsen LG, Hellum C, Storheim K, et al. Cost-effectiveness of total disc replacement versus multidisciplinary rehabilitation in patients with chronic low back pain: a Norwegian multicenter RCT. Spine 2014;**39**(1):23-32 doi: 10.1097/brs.0000000000000065[published Online First: Epub Date]|.
336. Garcia R, Jr., Yue JJ, Blumenthal S, et al. Lumbar Total Disc Replacement for Discogenic Low Back Pain: Two-year Outcomes of the activL Multicenter Randomized Controlled IDE Clinical Trial. Spine 2015;**40**(24):1873-81 doi: 10.1097/brs.0000000000001245[published Online First: Epub Date]|.
337. Pettine K, Hersh A. Kineflex lumbar artificial disc versus Charite lumbar total disc replacement for the treatment of degenerative disc disease: A randomized non-inferiority trial with minimum of 2 years' follow-up. SAS journal 2011;**5**(4):108-13 doi: 10.1016/j.esas.2011.07.003[published Online First: Epub Date]|.
338. Buttner-Janz K, Guyer RD, Ohnmeiss DD. Indications for lumbar total disc replacement: selecting the right patient with the right indication for the right total disc. International journal of spine surgery 2014;**8** doi: 10.14444/1012[published Online First: Epub Date]|.
339. Hoff EK, Strube P, Pumberger M, Zahn RK, Putzier M. ALIF and total disc replacement versus 2-level circumferential fusion with TLIF: a prospective, randomized, clinical and radiological trial. Eur. Spine J. 2015 doi: 10.1007/s00586-015-3852-y[published Online First: Epub Date]|.
340. Lackey A, Phan K, Mobbs R. A systematic review and meta-analysis of outcomes in hybrid constructs for multi-level lumbar degenerative disc disease. J. Clin. Neurosci. 2016;**34**:23-29 doi: 10.1016/j.jocn.2016.06.011[published Online First: Epub Date]|.
341. Tohmeh AG, Smith WD. Lumbar total disc replacement by less invasive lateral approach: a report of results from two centers in the US IDE clinical trial of the XL TDR(R) device. Eur. Spine J. 2015;**24 Suppl 3**:331-8 doi: 10.1007/s00586-015-3843-z[published Online First: Epub Date]|.
342. Pimenta L, Marchi L, Amaral RA, et al. 342 Lumbar Total Disc Replacement by the Lateral Approach Up to 10-Year Follow-up. Neurosurgery 2016;**63 Suppl 1**:200-1 doi: 10.1227/01.neu.0000489831.09131.09[published Online First: Epub Date]|.
343. Malham GM, Parker RM. Early experience with lateral lumbar total disc replacement: Utility, complications and revision strategies. J. Clin. Neurosci. 2017;**39**:176-83 doi: 10.1016/j.jocn.2017.01.033[published Online First: Epub Date]|.

APPENDIX

Previous studies with long-term follow-up after total disc replacement

Table 1. **Prospective studies with long-term results after TDR**

Publication	**Study design**	**Study cohort (n)**	**Mean follow-up**	**Follow-up rate**	**Implant**	**Reoperation rate**	**Mean ODI change**	**Result**
Tropiano et al. (2005)	Prospective cohort (single centre)	64	9 years	86 %	Prodisc I	5 % (posterior fusion)	NA*	60 % excellent result and 15 % good result (modified Stauffer Coventry scale). Significant improvement in 3 point scales measuring pain (mean 2.7 – 1.4) and disability (mean 2.0 – 0.8) 9 % complications. Age < 45 years or prior lumbar surgery affected the outcome negatively.
Guyer et al. (2009)	Multicentre randomised (FDA regulated study comparing TDR to anterior fusion)	277	5 years	44 %	Charité	8 % (device related)	24 (vs 28 points in fusion group)	58 % overall success (vs 51 % after fusion) (FDA criteria**). Mean VAS improvement 39 points (vs 40 points after fusion). No difference between groups in overall success, functional improvement or pain relief, but a larger proportion of patients were employed after TDR (66 %) compared to fusion (47 %). 19 % had ROM < 5º and ossification of the disc space.
Zigler and Delamarter (2012)	Multicentre randomised (FDA regulated study comparing TDR to circumferen-tial fusion)	236	5 years	82 %	Prodisc-L	8 %	29 (vs 27 points in fusion group)	64 % overall success (vs 45 % after fusion) (separate criteria for the two groups). Mean VAS improvement 39 points (vs 35 points after fusion). 38 % used narcotics at follow-up (vs 40 % after fusion). Similar patient satisfaction in the two groups (77 %). 83 % “would have the same surgery again” (vs 68 % after fusion). No spontaneous ankyloses. Less severe adverse events after TDR versus fusion (0.38 versus 0.58 per patient).

Meir et al. (2013)	Prospective cohort (single surgeon)	28	10 years	82 %	Acroflex	39 %	16	61 % cumulative implant survival (to first revision surgery). Better mean ODI improvement in patients who were not reoperated (18 points) versus patients who were reoperated (12 points). 25 % revision because of implant failure. CT findings in patients with implant survival included heterotopic bone formation (85 %), osteolysis (50 %), subsidence (14 %) and adjacent disc degeneration (68 %).
Sköld et al. (2013)	Single centre randomised (comparing TDR to posterior lumbar fusion)	152	5 years	99 %	Charité, Prodisc II, Maverick	20 % (total)	25 (vs 17 points in fusion group)	38 % were pain free at follow-up (versus 15 % after fusion). Mean VAS improvement 40 points (vs 28 points after fusion). 8 % described worse pain compared to pre-treatment (vs 4 % after fusion). 78 % had ≥ 25 % ODI improvement (vs 65 % after fusion). EQ-5D 0.76 at follow-up (vs 0.68 after fusion). 79 % satisfied with the result of the treatment (vs 69 % after fusion). 78 % were employed at follow-up (vs 90 % after fusion). 16 % complications (vs 13 % in the fusion group).
Siepe et al. (2014)	Prospective cohort (single centre)	201	7 years	90 %	Prodisc II	16 % (total)	22	Mean VAS improvement 38 points at last follow-up. 86 % highly satisfied or satisfied with the outcome. Complication rate 14 %. Less satisfaction and more complications in one-level compared to two-level procedures. 79 % “would undergo surgery again”. 67 % were employed at follow-up. 2 % reoperated at adjacent level.

Aghayev et al. (2014)	Prospective cohort (Swiss spine registry)	248	5 years	51 %	Activ L, Charité, Dynardi, Maverick, Prodisc-L	4 %	NA*	Mean VAS improvement 44 points. EQ-5D improvement 0.46 points. 23 % complications. 11 % adjacent level degeneration. 13 % range of motion 0-2 º (functional x-ray).
Lu et al. (2015)	Prospective cohort study	35	12 years	91 %	Charité III	6 %	28	Mean VAS improvement 70 points. 76 % returned to work. 9 % subsidence. 71 % heterotopic ossification. No major complications, but 6 % had vascular injury, 6 % had abdominal hernia and 6 % had postoperative anhidrosis of the feet.
Guyer et al. (2016)	Multicentre randomised (FDA regulated study comparing two different TDR devices.	394	5 years	68 %	Kineflex-L, Charité	12 % (total)	40 (approxi-mately)	Serum ion levels after metal-on-metal-implants were < 20 % of recommended threshold level to merit monitoring hip replacement patients (Medicines and Healthcare Products regulatory Agency). Overall success (FDA criteria**): 77 % vs 70 %. ≥ 15 ODI points improvement: 87 % vs 85 %. Patient satisfaction: 97 % vs 96 %. Severe heterotopic ossification: 16 % vs 19 %.
Laugesen et al. (2017)	Prospective cohort study.	68	11 years	84 %	Prodisc II	33 %	NA*	Mean VAS improvement 36 points. Dallas Pain Questionnaire improvement 18 points. SF-36 MCS *** significantly worse at follow-up compared to pre-treatment score. Worse outcome among patients who were reoperated compared to patients who were not reoperated. 65 % “would have the same surgery again” or “would probably have the same surgery again”.

Wuertinger et al. (2018)	Prospective cohort study of patients treated with TDR at L5/S1. Subgroup of patients in *Siepe et al (2014).*	51	8 years	NA*	Prodisc II	NA*	21	Mean VAS improvement 51 points. 71 % satisfactory or highly satisfactory result. ROM gradually declined during follow-up.
Lazennec et al. (2018)	Prospective cohort study of patients treated with monosegmental TDR.	61	5 years	43 %	LP-ESP	8 %	31	≥ 15 ODI points improvement: 82 % Mean VAS improvement: 3.3 points *(scale 0-10, Prof. Jean-Yves Lazennec, personal communication).* No local ossification or osteolysis (x-ray). ROM at index level: 6.8º.

*Not available

** Food and Drug Administration (FDA) criteria for overall success: ODI improvement ≥ 15 points, no device failure, no major complications, no neurologic deterioration.

*** SF-36 MCS = 36-item Short Form Health Survey Mental component Summary.

Table 2. **Retrospective studies with long-term results after TDR**

Publication	**Study design**	**Study cohort (n)**	**Mean follow-up**	**Follow-up rate**	**Implant**	**Reoperation rate**	**Result**
Lemaire et al. (2005)	Retro-spective case series	107	11 years	93 %	Charité	5 % fusion	90 % good or excellent clinical outcome (modified Stauffer Coventry scale). 92 % were employed at follow-up. No spontaneous ankylosis. 21 % complications (including need for fusion (5 %), facet arthropathy (4 %), subsidence (2 %), periprosthetic ossification affecting mobility (2 %), adjacent level degeneration (2 %), vascular injury (2 %), leg ischemia (1 %), sexual dysfunction (1 %) and neurologic complications.
Putzier et al. (2006)	Retro-spective case series	71	17 years	75 %	Charité	11 %	60 % of the patients had spontaneous ankylosis, and those were more satisfied than those with functional implants. Mean ODI 42 points and mean VAS 48 points at follow-up. No difference in outcomes between Charité types I, II or III.
David et al. (2007)	Retro-spective case series	108	13 years	98 %	Charité	8 % fusion 3 % reoperated due to adjacent level degeneration	82 % good or excellent clinical outcome (modified Stauffer Coventry scale). 90 % returned to work. 9 % spontaneous ankylosis. 3 % subsidence 2 % core subluxation
Lee et al. (2015)	Retro-spective case series	54	5 years	70 % (> 2-year follow-up)	Prodisc-L	11 %	17 % approach related complications. Less blood loss, shorter operating time, shorter hospital stay and shorter time to mobilisation – but more frequent approach related surgical complications - compared to a similar group operated with fusion.

Park et al. (2016)	Retro-spective case series	64	10 years	84 %	Prodisc II	9 %	ODI improvement 22 points and VAS improvement 45 points from before treatment to last follow-up. Larger proportions with clinical success and patient satisfaction among patients considered as good versus bad candidates for TDR.
Lu et al. (2017)	Retro-spective case series	30	15 years	86 %	Charité III	6 %	ODI improvement 38 points. VAS improvement 7.0 points (scale 0-10). 90 % returned to work. 87 % "would choose the same treatment again". 11 % surgical complications. 13 % subsidence. Heterotopic ossification was observed in 74 % of TDR levels, and spontaneous ankylosis in 3 %. 23 % of TDR levels had ROM < 3º.

*Not available

I

ELSEVIER

CrossMark

The Spine Journal 17 (2017) 1480–1488

THE SPINE JOURNAL

Total disc replacement versus multidisciplinary rehabilitation in patients with chronic low back pain and degenerative discs: 8-year follow-up of a randomized controlled multicenter trial

Håvard Furunes, MD[a,b,c,*], Kjersti Storheim, PhD[b,c], Jens Ivar Brox, PhD[b,d], Lars Gunnar Johnsen, PhD[e,f], Jan Sture Skouen, PhD[g,h], Eric Franssen, MD[i], Tore K. Solberg, PhD[j,k,l], Leiv Sandvik, PhD[b], Christian Hellum, PhD[b,m]

[a]*Innlandet Hospital Gjøvik, Department of Surgery, Innlandet Hospital Gjøvik, Kyrre Grepps gate 11, 2819 Gjøvik, Norway*
[b]*University of Oslo, Postbox 1072 Blindern, 0316 Oslo, Norway*
[c]*Oslo University Hospital Ullevål, FORMI, Building 37B, Postbox 4950 Nydalen, 0424 Oslo, Norway*
[d]*Department for Physical Medicine and Rehabilitation, Oslo University Hospital, Postbox 4950 Nydalen, 0424 Oslo, Norway*
[e]*Department of Orthopaedic Surgery, St. Olav's University Hospital, Prinsesse Kristinas gate 3, 7030 Trondheim, Norway*
[f]*Department of Neuromedicine, Faculty of Medicine, Norwegian University of Science and Technology, Høgskoleringen 1, 7491 Trondheim, Norway*
[g]*Department of Physical Medicine and Rehabilitation, The Outpatient Spine Clinic, Haukeland University Hospital, Postbox 1400, 5021 Bergen, Norway*
[h]*Department of Global Public Health and Primary Care, University of Bergen, PO Box 7804, 5007 Bergen, Norway*
[i]*Orthopaedic Department, Stavanger University Hospital, Gerd Ragna Bloch Thorsens gate, 4011 Stavanger, Norway*
[j]*Department of Clinical Medicine, University of Tromsø—The Arctic University of Norway, Hansine Hansens veg 18, 9019 Tromsø, Norway*
[k]*The Norwegian Registry for Spine Surgery (NORspine), Northern Norway Regional Health Authority, University Hospital of Northern Norway, Sykehusvegen 38, 9019 Tromsø, Norway*
[l]*Department of Neurosurgery, University Hospital of Northern Norway, Sykehusvegen 38, 9019 Tromsø, Norway*
[m]*Division of Orthopaedic Surgery, Oslo University Hospital, Postbox 4950 Nydalen, 0424 Oslo, Norway*

Received 14 February 2017; revised 1 April 2017; accepted 8 May 2017

Abstract

BACKGROUND CONTEXT: Lumbar total disc replacement (TDR) is a treatment option for selected patients with chronic low back pain (LBP) that is non-responsive to conservative treatment. The long-term results of disc replacement compared with multidisciplinary rehabilitation (MDR) have not been reported previously.

PURPOSE: We aimed to assess the long-term relative efficacy of lumbar TDR compared with MDR.

DESIGN: We undertook a multicenter randomized controlled trial at five university hospitals in Norway.

PATIENT SAMPLE: The sample consisted of 173 patients aged 25–55 years with chronic LBP and localized degenerative changes in the lumbar intervertebral discs.

OUTCOME MEASURES: The primary outcome was self-reported physical function (Oswestry Disability Index [ODI]) at 8-year follow-up in the intention-to-treat population. Secondary outcomes included self-reported LBP (visual analogue scale [VAS]), quality of life (EuroQol [EQ-5D]), emotional distress (Hopkins Symptom Checklist [HSCL-25]), occupational status, patient satisfaction, drug use, complications, and additional back surgery.

METHODS: Patients were randomly assigned to lumbar TDR or MDR. Self-reported outcome measures were collected 8 years after treatment. The study was powered to detect a difference of 10 ODI points between the groups. The study has not been funded by the industry.

RESULTS: A total of 605 patients were screened for eligibility, of whom 173 were randomly assigned treatment. Seventy-seven patients (90%) randomized to surgery and 74 patients (85%) randomized to rehabilitation responded at 8-year follow-up. Mean improvement in the ODI was 20.0 points (95% confidence interval [CI] 16.4–23.6, p≤.0001) in the surgery group and 14.4 points (95% CI 10.7–18.1, p≤.0001) in the rehabilitation group. Mean difference between the groups at 8-year

FDA device/drug status: Approved (ProDisc L, Synthes Spine).

Author disclosures: ***HF:*** Nothing to disclose. ***KS:*** Nothing to disclose. ***JIB:*** Nothing to disclose. ***LGJ:*** Nothing to disclose. ***JSS:*** Nothing to disclose. ***EF:*** Nothing to disclose. ***TKS:*** Nothing to disclose. ***LS:*** Nothing to disclose. ***CH:*** Nothing to disclose.

* Corresponding author. Innlandet Hospital Gjøvik, Department of Surgery, Innlandet Hospital Gjøvik, Kyrre Grepps gate 11, 2819 Gjøvik, Norway. Tel.: +4799041763.

E-mail address: havardfurunes@gmail.com (H. Furunes)

http://dx.doi.org/10.1016/j.spinee.2017.05.011

follow-up was 6.1 points (95% CI 1.2–11.0, p=.02). Mean difference in favor of surgery on secondary outcomes were 9.9 points on VAS (95% CI 0.6–19.2, p=.04) and 0.16 points on HSCL-25 (95% CI 0.01–0.32, p=.04). There were 18 patients (24%) in the surgery group and 4 patients (6%) in the rehabilitation group who reported full recovery (p=.002). There were no significant differences between the groups in EQ-5D, occupational status, satisfaction with care, or drug use. In the per protocol analysis, the mean difference between groups was 8.1 ODI points (95% CI 2.3–13.9, p=.01) in favor of surgery. Forty-three of 61 patients (70%) in the surgery group and 26 of 52 patients (50%) in the rehabilitation group had a clinically important improvement (15 ODI points or more) from baseline (p=.03). The proportion of patients with a clinically important deterioration (six ODI points or more) was not significantly different between the groups. Twenty-one patients (24%) randomized to rehabilitation had crossed over and had undergone back surgery since inclusion, whereas 12 patients (14%) randomized to surgery had undergone additional back surgery. One serious adverse event after disc replacement is registered (<1%).

CONCLUSIONS: Substantial long-term improvement can be expected after both disc replacement and MDR. The difference between groups is statistically significant in favor of surgery, but smaller than the prespecified clinically important difference of 10 ODI points that the study was designed to detect. Future research should aim to improve selection criteria for disc replacement and MDR.

Keywords: Chronic low back pain; Degenerative disc disease; Long-term follow-up; Lumbar total disc replacement; Multidisciplinary cognitive behavioral and exercise rehabilitation; Randomized trial; Self-rated disability

Introduction

Low back pain (LBP) is common and causes more disability than any other condition [1]. The theory of a multifactorial etiology of chronic LBP is supported by good results from multidisciplinary rehabilitation (MDR) [2]. Degeneration of the intervertebral disc has been linked to back pain [3]. The traditional surgical treatment is spinal fusion, which aims to relieve pain by restricting the segmental motion.

Lumbar total disc replacement (TDR) was introduced as a motion preserving surgical alternative to spinal fusion. A Cochrane review found statistically significant results in favor of TDR compared with fusion [4], but the differences were not considered to be clinically important. A few randomized studies with mid- to long-term follow-up have now been published [5–7], so far with promising results for disc replacement.

The role of disc replacement is still controversial. Recently, most major health insurance carriers in the United States decided not to provide coverage for single-level lumbar TDR, considering TDR to be experimental, and stating that its long-term clinical outcome is unclear [8].

In patients with LBP and degenerative disc, fusion and non-surgical treatment with MDR have had similar short- [9,10] and long-term [11] results. However, TDR was significantly more effective than MDR after 2 years in the only randomized study comparing these two treatments [12]. The long-term effect of disc replacement compared with non-surgical treatment has not been reported previously.

Hence, the aim of this study was to evaluate the long-term efficacy of TDR compared with MDR in patients with chronic LBP.

Materials and methods

Study design

This study is an 8-year follow-up of a randomized multicenter study conducted at five university hospitals in Norway [12]. The 8-year follow-up was approved by the Norwegian Regional Ethical Committee South East C (2011/2177). The project was conducted in accordance with the Helsinki Declaration and the ICH-GCP guidelines and registered at www.clinicaltrials.gov under the identifier NCT01704677 before it commenced. Because this study does not evaluate drugs, a data monitoring committee is not mandatory according to Norwegian regulations, but it was overseen by a scientific board.

Results are reported according to the CONSORT standard for reporting randomized trials.

Participants

Patients were recruited from all health regions in Norway, as detailed by Hellum and colleagues [12]. Written informed consent was obtained. Eligible patients were aged 25–55 and had LBP as their main symptom for at least 1 year, were resistant to non-operative treatment, had a score of at least 30 on the Oswestry Disability Index (ODI), and degenerative intervertebral disc changes in L4–L5 or L5–S1, or both. We excluded patients with nerve root involvement, disc degeneration in more than two levels, symptoms of spinal stenosis, generalized chronic pain, former lumbar fracture, osteoporosis, spondylosis, arthritis, spinal deformity, or drug abuse.

Randomization and blinding

A statistician not involved in the trial created a computer-generated random list (1:1 allocation and random block sizes

EVIDENCE & METHODS

Context
Via a randomized controlled trial (RCT), the authors compared eight-year outcomes between total disc replacement (TDR) and multidisciplinary rehab.

Contribution
While there were subtle differences favoring surgery, these were not found to be clinically significant.

Implications
Problems with evaluating long-term functional outcomes in surgery versus nonoperative RCTs using intent-to-treat are well-recognized, as cross-over in the nonoperative group and re-operations in the surgical groups make drawing conclusions difficult. There is value (as we learned from the thoughtful design of SPORT) to also following those who choose their treatment, as opposed to only those who have been randomized. That said, this study is in line with well-known prior fusion versus CBT studies, lending some support for the findings given similar reported outcomes between TDR and fusion.

—The Editors

of two to eight). Allocation was performed using a website hosted by the medical faculty at the Norwegian University of Science and Technology. Once a patient was included, a coordinating secretary, blinded to the patients' characteristics and not involved in the inclusion process, logged their information on the randomization website and performed randomization. The treating unit and the patient were informed about the allocation shortly after randomization and patients were treated within the next 3 months. Randomization was stratified by center and by previous back surgery (microsurgical decompression) or not. Independent observers collected and entered data.

Procedures

Detailed information on the study interventions is given by Hellum and colleagues [12].

Rehabilitation

Rehabilitation was conducted according to the principles described by Brox and colleagues and consisted of a cognitive approach and supervised physical exercise [13]. The intervention was standardized through three seminars as well as videos and lecture sessions for the treatment providers before the study. Rehabilitation was organized as an outpatient treatment in groups and lasted for about 60 hours over 3–5 weeks. The treatment consisted of lectures, individual discussions, daily workouts to increase physical capacity, endurance, strength, coordination, and specific training of the abdominal muscles and the lumbar multifidus muscles, as well as challenging patients' thoughts about, and participation in, physical activities that were previously not recommended (such as lifting, jumping, vacuuming, dancing, and ball games).

Surgery

The surgical intervention consisted of replacement of the degenerative intervertebral lumbar disc with an artificial disc (ProDisc II, Synthes Spine). Anterior approach was used with a Pfannenstiel, median or paramedian incision, and retroperitoneal dissection. A fluoroscope was used to ensure the correct positioning of the prosthesis. Surgeons were required to have inserted at least six disc prostheses before performing surgery in the study. There were no major postoperative restrictions. Patients were not referred for postoperative physiotherapy, but if requested they could be referred for general mobilization and non-specific exercises at 6-weeks follow-up.

Outcomes

The primary outcome measure was pain and disability measured with Norwegian version 2.0 of the ODI [14,15]. (Scores range from 0 to 100, with a lower score indicating less pain and disability.) Secondary outcomes included LBP (measured with a visual analogue scale, ranging from 0 (no pain) to 100 (worst pain imaginable)) and EuroQol (scores ranging from –0.59 to 1 (1 represents perfect health)) [16]. We included emotional distress as a psychological variable (Hopkins Symptom Checklist [HSCL-25], scores range from 1 to 4, with lower scores indicating less severe symptoms) [17]. Work participation rate was calculated. Satisfaction with the result of the treatment was reported on a seven-point Likert scale, and satisfaction with care on a five-point Likert scale [18]. Further, reoperations, complications, and daily use of drugs were registered. Patient reported outcome measures were obtained before randomization and at follow-up consultations at 6 weeks, 3 months, 6 months, 1 year, 2 years, and 8 years after the intervention was completed. At the main end point 8 years after treatment, the patients completed all outcome questionnaires at home and returned them by mail to an independent observer before the follow-up visit.

Statistical analysis

Sample size

The first phase of the trial was designed to have 80% power to detect the significant difference of a change of at least 10 points in the mean ODI score between the intervention groups at 2-year follow-up. Baseline standard deviation was estimated at 18 [14]. Adding 25% for a multicenter study design and 30% for possible dropouts, we decided to include 180 patients.

Intention to treat

The main statistical analysis was in the intention-to-treat population at the 8-year follow-up. Missing values were replaced with multiple imputation. Patients who received

treatment similar to the opposite treatment arm (crossovers) were kept in the group they were randomized to. We used χ^2 test or Fisher exact test to analyze categorical variables and an independent two-sided *t* test or analysis of variance to analyze continuous variables. A significance level of 5% was used throughout. All statistical analyses were performed using SPSS version 22.0 (IBM SPSS Statistics for Windows, IBM Corp., Armonk, NY). We did not adjust for significantly different baseline scores.

Per protocol

In a per protocol analysis of the primary outcome variable (ODI) we excluded patients who did not receive the intervention they were randomized to, and those who had undergone back surgery or MDR after the study intervention was complete. Missing data were not replaced. In addition to calculating the mean change in each group, we also calculated the proportion of patients whose condition was classified as either improved or deteriorated. According to the Food and Drug Administration criteria, an individual improvement in ODI of at least 15 points can be considered a clinically important improvement [19,20]. A decrease of six ODI points represented a "change for the worse" [21]. We calculated number needed to treat, which represents the number of patients treated with TDR instead of MDR needed to provide a clinically important improvement in one patient.

Sensitivity analysis

In an additional subanalysis of ODI we excluded patients who did not receive the intervention they were randomized to, and used the last value before crossover or reoperation in those who had undergone back surgery or MDR after the study intervention was complete. Other missing data were not replaced.

Role of the funding sources

The funders of the study (Oslo University Hospital, South Eastern Norway Regional Health Authority, and EXTRA funds from the Norwegian Foundation for Health and Rehabilitation through the Norwegian Back Pain Association) had no role in the study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

Between April 30, 2004 and September 27, 2007, a total of 605 patients were screened for eligibility, of which 179 patients were included. Six patients who obviously fulfilled the exclusion criteria (n=3/3 in surgery/rehabilitation) were included by mistake, and were therefore excluded shortly after randomization. Hence, a total of 173 patients were included

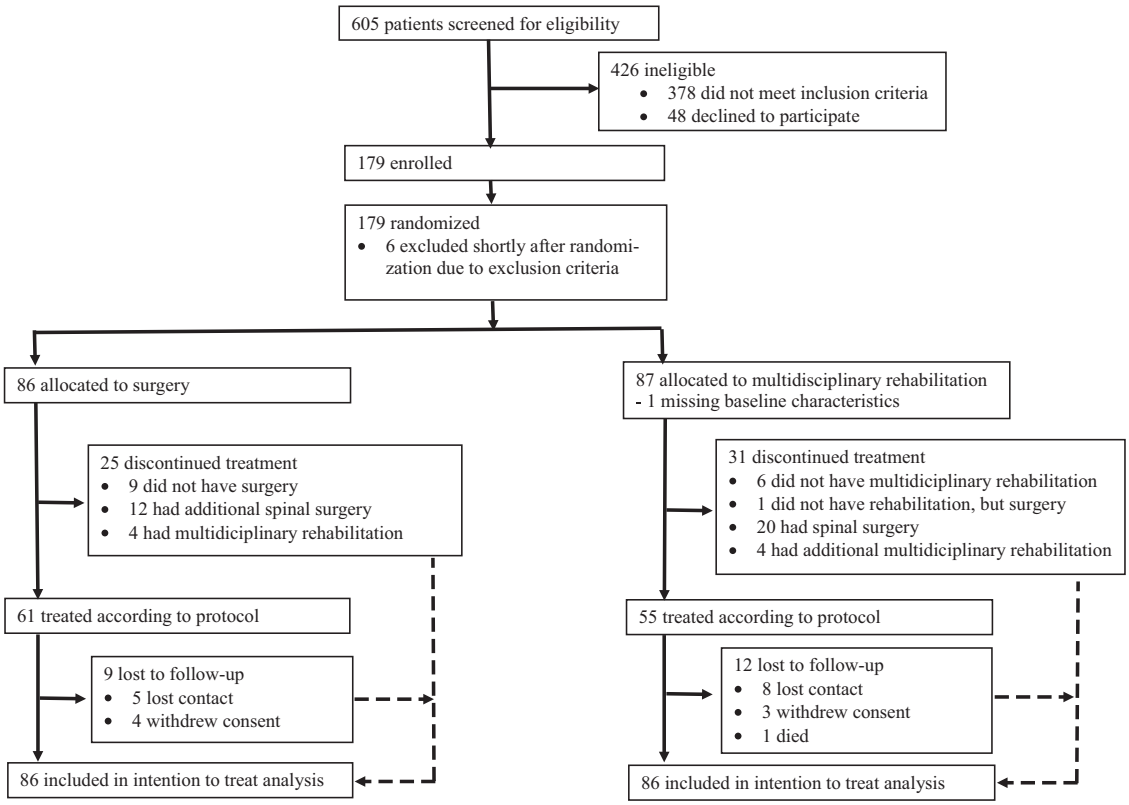


Fig. 1. Trial profile.

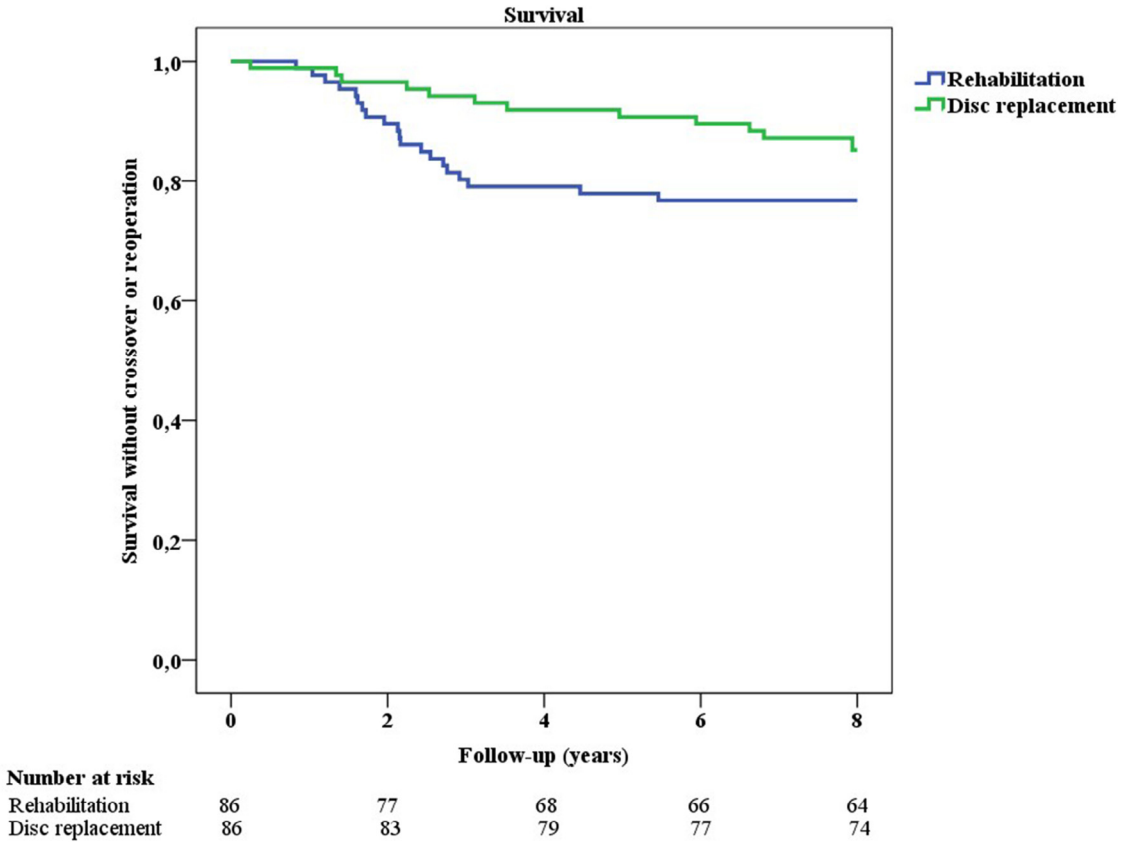


Fig. 2. Kaplan-Meier plot comparing event-free survival of concepts of rehabilitation and disc replacement. Events are crossover to surgery (rehabilitation group) and reoperation (surgery group). Number at risk represents the number of patients who were not reoperated and did not cross over.

in the study (n=86 allocated to surgery and n=87 allocated to rehabilitation) (Fig. 1).

There were 151 participants (87%) available for 8-years' follow-up, 74 (85%) in the rehabilitation group and 77 (90%) in the surgery group. Because the dropout rate was less than 30%, the power of the study was considered to be satisfactory.

In the MDR group, 21 of 87 patients (24%) had crossed over and had undergone back surgery since inclusion (14 with TDR, 5 with spinal fusion, and 2 with discectomy). In the surgery group, 12 of 86 patients (14%) had undergone additional back surgery (one reoperation because of implant dislocation, three with spinal fusion at the level of the prosthesis, six with decompression of spinal stenosis (one of whom had not received TDR), one discectomy (had not received TDR), and one neurostimulator implantation). Four patients from the rehabilitation group and four patients from the surgery group received MDR after the intervention was finished. Survival without crossover or reoperation is presented in Fig. 2. The figure is descriptive and does not include a measure of effect.

Most baseline characteristics were similar in the two intervention groups, but LBP score was significantly worse in the rehabilitation group than in the surgery group (Table 1).

In the intention-to-treat analysis of the primary outcome (ODI), the mean difference between the groups at 8-year follow-up was 6.1 points (95% CI 1.2–11.0, p=.02) in favor of surgery. Mean improvement in ODI from baseline to 8-year follow-up was 14.4 points (95% CI 10.7–18.1) in the rehabilitation group and 20.0 points (95% CI 16.4–23.6) in the surgery group.

In the intention-to-treat analysis of secondary outcomes, the mean differences between groups in favor of surgery were: 9.9 on visual analogue scale (95% CI 0.6–19.2, p=.04), 0.05 points on EuroQol 5D (95% CI −0.06 to 0.16, p=.35), and 0.16 points on Hopkins Symptom Checklist-25 (95% CI 0.01–0.32, p=.04). Categorical secondary outcomes are presented in Table 2.

One serious complication was registered in the first 2 years [12]. During revision surgery for a dislocated polyethylene inlay 3 months postoperatively, an injury to the left common iliac artery led to a compartment syndrome and subsequently a lower leg amputation. Minor complications within 2 years are also described by Hellum and colleagues [12].

Twelve patients randomized to rehabilitation underwent spinal operations between 2 and 8 years, of which one required a reoperation; the patient was first operated on with an anterior fusion, and later reoperated on with additional

Table 1
Demographics and baseline clinical characteristics

	Surgery (n=86)	Rehabilitation (n=86)	p
Mean (SD) age (y)	41.1 (7.1)	40.8 (7.1)	.73
Women (n, %)	40 (47)	51 (59)	.11
Mean (SD) duration of back pain (mo) (n, %)	76 (72)	85 (74)	.49
Education:			.08
Primary school (9 y) (n, %)	19 (22)	17 (20)	
High school (12 y) (n, %)	44 (51)	58 (67)	
College (n, %)	14 (16)	8 (9)	
University (n, %)	9 (11)	3 (4)	
Mean (SD) body mass index (BMI)	25.6 (3.1)	25.5 (3.5)	.71
Current smokers (n, %)	42 (49)	37 (43)	.40
Work status (working vs. not working) (n, %):			.24
Working (includes part time sick leave)	24 (28)	22 (26)	
On sick leave	25 (29)	34 (41)	
Rehabilitation	29 (34)	25 (29)	
Disability pension	3 (4)	0	
Homemaker	0	2 (2)	
Unemployed	1 (1)	0	
Student	3 (4)	0	
Unknown	1 (1)	3 (4)	
Comorbidity (n, %)	20 (23)	21 (24)	.86
Daily consumption of narcotics (n, %)	23 (27)	17 (20)	.26
Previous surgery (n, %)	23 (27)	25 (29)	.77
Mean (SD) ODI score	41.8 (9.1)	42.8 (9.3)	.50
Low back pain score*	64.9 (15.3)	73.6 (13.9)	<.001
Mean (SD) SF-36 score:			
Physical function	52.7 (17.6)	50.6 (17.7)	.47
Role physical	25.3 (24.2	23.9 (18.7)	.69
Bodily pain	24.9 (16.5)	24.4 (12.1)	.84
General health	57.9 (19.7)	55.9 (19.9)	.54
Vitality	37.8 (20.2)	33.1 (19.9)	.15
Social function	53.0 (30.6)	57.6 (26.7)	.32
Role emotion	72.5 (33.3)	67.6 (32.7)	.35
Mental health	71.7 (18.0)	65.8 (18.9)	.05
Physical component summary score	30.5 (7.1)	30.8 (6.5)	.79
Mental component summary score	47.7 (13.0)	45.2 (13.2)	.23
Mean (SD) HSCL-25	1.8 (0.5)	1.9 (0.5)	.37
Mean (SD) FABQ work	25.9 (11.3)	27.4 (9.9)	.38
Mean (SD) FABQ physical	14.1 (5.8)	12 (5.5)	.08

FABQ, Fear Avoidance Beliefs Questionnaire (scale ranges from 0 to 24 [physical] and from 0 to 42 [work], lower scores indicate less severe symptoms); HSCL-25, Hopkins Symptom Checklist (for emotional distress, scores range from 1 to 4, lower scores indicate less severe symptoms); ODI, Oswestry Disability Index (from 0 to 100, lower scores indicate less severe symptoms); SD, standard deviation; SF-36, Short Form-36 (from 0 to 100, higher scores indicate better health status).

* Calculated with horizontal scale ranging from 0 (no pain) to 100 (worst pain imaginable), with word anchors at the beginning and end.

posterolateral fusion caused by non-fusion of the segment. Nine patients randomized to surgery were operated on between 2 and 8 years, of which two required further reoperation; one had a decompression at the level of the prosthesis and was later reoperated on because of a dural tear, and one (who did not receive TDR) had a decompression with an interspinous device, followed by removal of the device, and subsequently an anterior fusion.

In the per protocol analysis, we found a mean difference between groups of 8.1 ODI points (95% CI 2.3–13.9, p=.01) in favor of surgery. Twenty-six of 52 patients (50%) in the rehabilitation group and 43 of 61 patients (70%) in the surgery group improved by 15 ODI points or more from baseline to 8-year follow-up (p=.03). Number needed to treat was 4.9 (95% CI 2.6–36.7). Six of 52 patients (12%) in the rehabilitation group and 3 of 61 patients (5%) in the surgery group deteriorated by six ODI points or more (p=.30).

In the sensitivity analysis, the mean difference was 10.8 ODI points (95% CI 5.5–16.2, p<.0001) in favor of surgery.

Mean change in ODI among patients who crossed over to surgery or underwent spinal reoperation, compared with patients who did not, is presented in Table 3.

Discussion

In this randomized multicenter trial, we found significant long-term improvement after both rehabilitation and disc replacement, and statistically significant long-term results in favor of disc replacement compared with rehabilitation in terms of functional improvement and pain relief. The clinical significance of the difference observed in this study will be discussed, as it is smaller than the prespecified clinically important difference of 10 points on ODI. It is also worth noting that there is still no agreement on the size of a clinically important difference between two treatment groups [22].

The choice of 10 ODI points as a threshold value for a clinically important difference between treatment groups in our study was based on recommendations at the time the study was designed [21]. Different threshold values were used in other studies at that time, for example, Fairbank and colleagues chose a threshold value of four points [10]. According to Glassman and colleagues [23], the threshold value of 10 ODI points indicates a clinically important improvement in an individual, and should not be misinterpreted as a measure of a difference between groups. The authors advocate reporting the proportion of patients achieving a minimal or substantial clinical difference in each group, rather than reporting mean group differences. Therefore, we did an additional analysis and calculated the proportion of patients with a clinically important improvement, which in FDA studies is defined as a minimum of 15 points improvement in ODI [19,20]. We found a clinically important improvement in a significantly larger proportion of patients in the TDR group (70%) than in the MDR group (50%) at 8-year follow-up. An individual minimal clinically important improvement is also commonly defined as a 30% improvement in ODI [4,18], but we chose to use the FDA criteria, because they had been used in an earlier report from the present trial [12]. At 2-year follow-up in the present study, the individual minimal clinical improvement was calculated as 12.88 ODI points based on receiver operating characteristic (ROC) curve analysis [24].

Table 2
Categorical secondary outcomes 8 years after rehabilitation or disc replacement, with proportions and p-values

	Rehabilitation		Disc replacement		
	Number of patients	Data	Number of patients	Data	p-Value
Working or studying	73	29 (40%)	77	37 (48%)	.33
Satisfaction with result of treatment*	73		76		
Full recovery		4 (6%)		18 (24%)	.002
Much better		26 (36%)		29 (38%)	.87
No/minimal change†		36 (49%)		22 (29%)	.01
Much worse		5 (7%)		1 (1%)	.11
Worse than ever		2 (3%)		6 (8%)	.28
Satisfied with care‡	73	50 (69%)	76	5 (66%)	.73
Daily analgetic medication	72	28 (39%)	76	31 (41%)	.87

* Seven-point Likert scale.
† Including "Slightly better," "No change," and "Slightly worse."
‡ Five-point Likert scale, not including "Slightly satisfied" as satisfied with care.

Table 3
Comparison of mean change on Oswestry Disability Index (ODI) from baseline to 8-year follow-up in patients who stayed in the treatment group they were randomized to, and those who crossed over or underwent reoperation

	Rehabilitation only (n=65)	Crossover from rehabilitation to surgery (n=21)	Disc replacement only (n=74)	Disc replacement, reoperated (n=12)
Mean change on ODI (95% CI)	14.6 (10.2–19.0)	13.9 (6.5–21.3)	22.4 (18.7–26.4)	5.4 (–4.8 to 15.6)

Missing values are replaced by multiple imputation.

The mean improvement from baseline in the rehabilitation group was 14.4 ODI points, which is in line with the mean improvement of 12.6 points at 11-years' follow-up after similar rehabilitation in patients with LBP and degenerative discs in three randomized trials reported by Mannion and colleagues [11]. The mean improvement from baseline in the surgery group was 20.0 ODI points, which is comparable with the findings in previous studies [5,6,19,20,25] (Table 4).

The short-term (2-year) mean improvement in ODI from baseline was 12.8 points in the rehabilitation group and 20.6 points in the surgery group [12], similar to the findings at 8-year follow-up, thus indicating a persistent long-term treatment effect for both rehabilitation and disc replacement.

The most important strengths of this study are the multicenter randomized controlled design, public financing, and data collection done by an independent research assistant. The follow-up rate of 87% after 8 years ensures robust internal validity of the results.

This study has several limitations that should be acknowledged. First, the patients could not be blinded. This may have led to a difference in placebo effect between the groups. A considerable placebo effect has been reported for vertebroplasty compared with sham surgery [26]. On the other hand, Freeman and colleagues did not find a significant change in outcome among patients with chronic LBP treated with intradiscal electrothermal therapy or sham surgery [27]. No studies comparing TDR with sham surgery have been published. Also, we do not know the natural course of LBP over 8 years, although Peng and colleagues observed a small, and not clinically important, improvement from 46.4 to 44.0 points in ODI over 4 years in an observational study [28].

Table 4
Previously published prospective trials with long-term results after TDR

	Current study	Sköld et al. (2013)	Siepe et al. (2014)	Guyer et al. (2009)	Guyer et al. (2016)	Lemaire (2005)
Study design	Multicenter randomized	Single center randomized	Single center non-randomized	Multicenter randomized	Multicenter randomized	Single center non-randomized
Study cohort	86	152	201	277	394	107
Mean follow-up	8 y	5 y	7 y	5 y	5 y	11 y
Follow-up rate	90%	99%	90%	44%	68%	93%
Reoperation rate	16% (total)	20% (total)	16% (total)	8% (device related)	12% (total)	NA
ODI change	20	25	22	24	NA	NA
Implant	Prodisc II	Charité, Prodisc II, Maverick	Prodisc II	Charité	Kineflex-L, Charité	Charité

ODI, Oswestry Disability Index; NA, not available; TDR, total disc replacement.

A second limitation of the study is that only 179 (30%) of the 605 patients screened for eligibility were included, which means that this study is only valid for a strictly defined group among patients with chronic LBP. The most important exclusion criteria were nerve root involvement or presence of generalized disc degeneration.

A third limitation is the relatively high crossover rate, especially from rehabilitation to surgery. Although all patients were informed that neither of the treatment methods were documented as being superior to the other, they were recruited as candidates for disc replacement, and some might have participated in the trial in hope of surgery. The number of patients who did not complete the treatment they were randomized to was similar in the two groups, but we did not assess patients' treatment expectations before randomization. The crossover rate is not higher than in other studies with long-term follow-up comparing spine surgery with non-operative treatment [11,29,30]. The crossovers probably have only a small impact on the result in the intention-to-treat analysis, because a similar change in ODI among patients randomized to rehabilitation was found in the per protocol analysis (14.1 points) and in the intention-to-treat analysis (14.4 points).

A forth limitation is the relatively high reoperation rate, similar to previous studies with long-term follow-up after TDR (Table 4). A high reoperation rate makes it difficult to untangle the results of TDR from the results of the reoperation in the intention-to-treat analysis. In the per protocol analysis, patients reoperated on were excluded, thus probably removing the most inferior results of surgery from this analysis.

Therefore, we added a sensitivity analysis in which values before crossover or reoperation were carried forward. This analysis may reflect the true results of the treatment the patients were randomized to, but carrying short-term results forward certainly represents a limitation in this analysis.

We have not detected any systematic differences in baseline characteristics between those randomized to rehabilitation who crossed over to surgery and those who did not, and long-term results for those who crossed over are similar to those who did not cross over. Neither did we detect any systematic differences in baseline characteristics between those randomized to surgery who were reoperated on and those who were not, but long-term results among those who were reoperated on were inferior to those who were not (Table 3).

To our knowledge, this is the only randomized trial comparing TDR with non-operative treatment, and we encourage future researchers to conduct a similar study. Because 24% of patients reported an excellent result with no symptoms of back pain at 8-year follow-up, and yet 8% described themselves as "worse than ever" (Table 2), there is a need for further research on predictors for good and bad long-term outcomes following TDR. Predictors for outcome after MDR are also needed. According to a Cochrane review [4], long-term mobility of a disc prosthesis, and the fate of the adjacent level and the facet joints following disc replacement, should also be assessed in future research.

Based on available evidence, including this study, long-term results of disc replacement seem to be acceptable, and in line with short-term results. On the other hand, long-term results of MDR are also acceptable [11]. Considering the risk of surgical complications, and the significant number of patients who achieve a clinically important improvement after rehabilitation, the first choice of treatment should be MDR.

Acknowledgments

The authors thank Marianne Bakke Johnsen and Maren Hjelle Guddal for collecting and punching data, and Eira Kathleen Ebbs for language help.

References

[1] Global Burden of Disease Study 2013 Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2015;386:743–800.

[2] Kamper SJ, Apeldoorn AT, Chiarotto A, et al. Multidisciplinary biopsychosocial rehabilitation for chronic low back pain. Cochrane Database Syst Rev 2014;(9):CD000963.

[3] Ohtori S, Inoue G, Miyagi M, Takahashi K. Pathomechanisms of discogenic low back pain in humans and animal models. Spine J 2015;15:1347–55.

[4] Jacobs W, Van der Gaag NA, Tuschel A, et al. Total disc replacement for chronic back pain in the presence of disc degeneration. Cochrane Database Syst Rev 2012;(9):CD008326.

[5] Siepe CJ, Heider F, Wiechert K, Hitzl W, Ishak B, Mayer MH. Mid- to long-term results of total lumbar disc replacement: a prospective analysis with 5- to 10-year follow-up. Spine J 2014;14:1417–31.

[6] Sköld C, Tropp H, Berg S. Five-year follow-up of total disc replacement compared to fusion: a randomized controlled trial. Eur Spine J 2013;22:2288–95.

[7] Zigler JE, Delamarter RB. Five-year results of the prospective, randomized, multicenter, Food and Drug Administration investigational device exemption study of the ProDisc-L total disc replacement versus circumferential arthrodesis for the treatment of single-level degenerative disc disease. J Neurosurg Spine 2012;17:493–501.

[8] Zigler J, Garcia R. ISASS policy statement—lumbar artificial disc. Int J Spine Surg 2015;9:7.

[9] Brox JI, Reikeras O, Nygaard O, et al. Lumbar instrumented fusion compared with cognitive intervention and exercises in patients with chronic back pain after previous surgery for disc herniation: a prospective randomized controlled study. Pain 2006;122:145–55.

[10] Fairbank J, Frost H, Wilson-MacDonald J, Yu LM, 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;330:1233.

[11] Mannion AF, Brox JI, Fairbank JC. Comparison of spinal fusion and nonoperative treatment in patients with chronic low back pain: long-term follow-up of three randomized controlled trials. Spine J 2013;13:1438–48.

[12] Hellum C, Johnsen LG, Storheim K, et al. Surgery with disc prosthesis versus rehabilitation in patients with low back pain and degenerative disc: two year follow-up of randomised study. BMJ 2011;342:d2786.

[13] Brox JI, Sorensen R, Friis A, et al. Randomized clinical trial of lumbar instrumented fusion and cognitive intervention and exercises in patients

with chronic low back pain and disc degeneration. Spine 2003;28:1913–21.
[14] Fairbank JC, Pynsent PB. The Oswestry Disability Index. Spine 2000;25:2940–52. discussion 52.
[15] Grotle M, Brox JI, Vollestad NK. Cross-cultural adaptation of the Norwegian versions of the Roland-Morris Disability Questionnaire and the Oswestry Disability Index. J Rehabil Med 2003;35:241–7.
[16] EuroQol Group. EuroQol—a new facility for the measurement of health-related quality of life. Health Policy (New York) 1990;16:199–208.
[17] Derogatis LR, Lipman RS, Rickels K, Uhlenhuth EH, Covi L. The Hopkins Symptom Checklist (HSCL): a self-report symptom inventory. Behav Sci 1974;19:1–15.
[18] Ostelo RW, de Vet HC. Clinically important outcomes in low back pain. Best Pract Res Clin Rheumatol 2005;19:593–607.
[19] Guyer RD, McAfee PC, Banco RJ, et al. Prospective, randomized, multicenter Food and Drug Administration investigational device exemption study of lumbar total disc replacement with the CHARITE artificial disc versus lumbar fusion: five-year follow-up. Spine J 2009;9:374–86.
[20] Guyer RD, Pettine K, Roh JS, et al. Five-year follow-up of a prospective, randomized trial comparing two lumbar total disc replacements. Spine 2016;41:3–8.
[21] Hagg O, Fritzell P, Nordwall A. The clinical importance of changes in outcome scores after treatment for chronic low back pain. Eur Spine J 2003;12:12–20.
[22] Schwind J, Learman K, O'Halloran B, Showalter C, Cook C. Different minimally important clinical difference (MCID) scores lead to different clinical prediction rules for the Oswestry Disability Index for the same sample of patients. J Man Manip Ther 2013;21:71–8.
[23] Glassman SD, Copay AG, Berven SH, Polly DW, Subach BR, Carreon LY. Defining substantial clinical benefit following lumbar spine arthrodesis. J Bone Joint Surg Am 2008;90:1839–47.
[24] Johnsen LG, Hellum C, Nygaard OP, et al. Comparison of the SF6D, the EQ5D, and the Oswestry Disability Index in patients with chronic low back pain and degenerative disc disease. BMC Musculoskelet Disord 2013;14:148.
[25] Lemaire JP, Carrier H, el Sariali H, Skalli W, Lavaste F. Clinical and radiological outcomes with the Charite artificial disc: a 10-year minimum follow-up. J Spinal Disord Tech 2005;18:353–9.
[26] Buchbinder R, Golmohammadi K, Johnston RV, et al. Percutaneous vertebroplasty for osteoporotic vertebral compression fracture. Cochrane Database Syst Rev 2015;(4):CD006349.
[27] Freeman BJ, Fraser RD, Cain CM, Hall DJ, Chapple DC. A randomized, double-blind, controlled trial: intradiscal electrothermal therapy versus placebo for the treatment of chronic discogenic low back pain. Spine 2005;30:2369–77. discussion 78.
[28] Peng B, Fu X, Pang X, et al. Prospective clinical study on natural history of discogenic low back pain at 4 years of follow-up. Pain Physician 2012;15:525–32.
[29] Lurie JD, Tosteson TD, Tosteson AN, et al. Surgical versus nonoperative treatment for lumbar disc herniation: eight-year results for the spine patient outcomes research trial. Spine 2014;39:3–16.
[30] Lurie JD, Tosteson TD, Tosteson A, et al. Long-term outcomes of lumbar spinal stenosis: eight-year results of the Spine Patient Outcomes Research Trial (SPORT). Spine 2015;40:63–76.