Surgical treatment and clinical outcomes in Lumbar Degenerative Spondylolisthesis

Ivar Magne Austevoll

Thesis for the degree of Philosophiae Doctor (PhD)
University of Bergen, Norway
2020

UNIVERSITY OF BERGEN

Surgical treatment and clinical outcomes in Lumbar Degenerative Spondylolisthesis

Ivar Magne Austevoll

Thesis for the degree of Philosophiae Doctor (PhD)
at the University of Bergen

Date of defense: 12.06.2020

Year: 2020

Title: Surgical treatment and clinical outcomes in Lumbar Degenerative Spondylolisthesis

Name: Ivar Magne Austevoll

Print: Skipnes Kommunikasjon / University of Bergen

Samandrag på norsk

Ei bukande mellomvirvelskive og auka storleik på ledd og ligament kan føre til tronghet for nervar i spinalkanalen. Dette blir kalla *spinal stenose* (bilete 1), og førekjem i all hovudsak etter 60 års alder. Hjå nokre har i tillegg den øvre virvelen glidd framover i forhold til virvelen under. Då heiter det *degenerativ spondylolistese* (bilete 2). Ved begge høve kan ein få smerter i korsryggen og nedover i beina. Dette innskrenkar ofte pasientane sitt funksjonsnivå. Hjå nokre kan plagene verta så store at ein vel å operera for å gjera betre plass til nervane som har det trongt. Ein fjernar da bein og seneband i bakre del av spinalkanalen. Dette inngrepet heiter *dekompresjon* (bilete 3). Hjå dei med spondylolistese kan ein i tillegg velja å gjera avstiving, det vil seia å festa den framoverglidde virvelen til virvelen under. Ved ein avstiving legg ein bein mellom virvlane, men oftast set ein inn skruar og stag i tillegg. Dette heiter *instrumentell fusjon* (bilete 4).

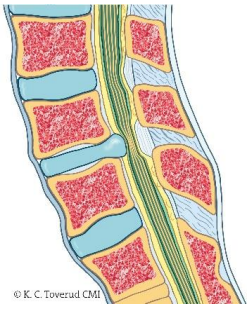
Kor vidt det er naudsynt med avstiving er omstridd. Nokre studiar tyder på at avstiving gjev betre resultat, medan andre studiar viser om lag like godt resultat om ein berre gjer dekompresjon. Praksis varierer mykje mellom ulike land. I 2013 vart om lag halvparten av pasientane i Noreg og Sverige opererte med berre dekompresjon, medan i USA fekk over 95% skruar i tillegg til dekompresjon. Å gjera instrumentell avstiving er dyrare og gir oftare komplikasjonar enn når ein berre gjer dekompresjon.

Hovudmålet med doktoravhandlinga var å undersøkja om det er naudsynt å gjera fusjon i tillegg til dekompresjon ved operasjon for degenerativ spondylolistese. I første studien fann vi at det var litt betre resultat hjå dei som hadde fått avstiving, men forskjellane mellom metodane var små og vi konkluderte med at begge metodane kunne brukast. I ein annan studie samanlikna vi berre dei to metodane som har vore mest brukt dei seinare åra; mikro-dekompresjon aleine og dekompresjon pluss instrumentell fusjon. Vi fann at det var lik del som hadde oppnådd eit vellukka resultat ved dei to metodane. Vi konkluderte med at mikro-dekompresjon aleine er

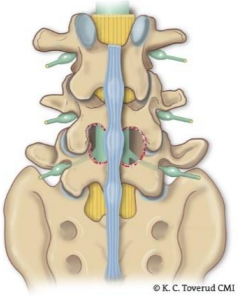
ein god nok metode, og bør vera fyrstevalet ved kirurgisk behandling av degenerativ spondylolistese. Eit vellukka resultat var definert som 30% betring av evne til å utføra vanlege aktivitetar. Grensa på 30% hadde vi berekna i ein eigen studie på pasientar med degenerativ spondylolistese.

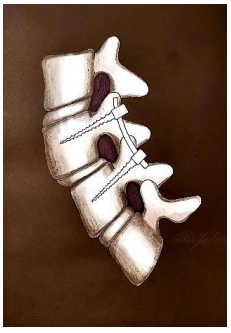
Studiane har nytta data frå Nasjonalt Kvalitetsregister for Ryggkirurgi. Registeret inneheld opplysningar om diagnosar, operasjonsmetodar og komplikasjonar, samt spørjeskjema som pasientane har svart på før operasjonen, etter tre månadar og 12 månadar etter operasjonen.

Ein vesentleg del av doktorgradsarbeidet har i tillegg vore å gjennomføra ein nasjonal studie der val av metode skulle avgjerast ved loddtrekning. I denne studien er det med 267 frivillige pasientar, kor halvparten er opererte med mikro-dekompresjon, mens andre halvparten er operert med dekompresjon pluss skruar. Resultata frå denne studien er ikkje klare enda, men den publiserte studieprotokollen, og gjennomføringa av studien, er presentert og diskutert i avhandlinga.

1. 

2. 

3. 

4. 

Scientific environment

This PhD project was initiated in 2012 with the planning of a multicentre randomised controlled trial (RCT). The trial, named NORDSTEN-DS, was a part of the NORwegian Degenerative spondylolisthesis and spinal STENosis (NORDSTEN) study. At the same time, plans for investigation of degenerative spondylolisthesis with data from the Norwegian Registry for Spine Surgery were initiated.

Members of the steering committee of the NORDSTEN-study, represented by experienced researchers from the Universities of Tromsø, Bergen and Oslo, have contributed to all four papers. Additional scientific support has been received from Rolf Gjestad – a key contributor and the responsible biostatistician in papers I, II and III, Morten Fagerland – the responsible biostatistician in paper IV, and Margreth Grotle – a key contributor to paper II.

The research has received funding from Helse Vest RHF (the Western Regional Health Authority) and Møre and Romsdal Hospital Trust.

My main supervisor was Christian Hellum, MD, PhD, Division of Orthopaedic Surgery, Oslo University Hospital.

My co-supervisors were Kari Indrekvam, MD, PhD, Kysthospitalet in Hagevik, Orthopedic Clinic, Haukeland University Hospital, and Rolf Gjestad, Psych., PhD, Research Department, Division of Psychiatry, Haukeland University Hospital.

Acknowledgements

Foremost, I would like to thank Dr. Lunde, Dr. Ylvisaker, Dr. Braaten, Dr. Algaard, and Dr. Nilsen. Witnessing your skills in handling the surgical knife, in addition to your patient handling skills, has played a crucial role in influencing my career decisions and development. You have all been role models and outstanding mentors in my surgical education.

I am aware and appreciative of the fact that my routine work at the hospitals in Hagavik and Haukeland has been tremendously stimulating. I count myself extremely fortunate to have been a part of such progressive teams. Essential contributors to my satisfaction are the highly skilled spinal surgeons at Hagavik and Haukeland. Thank you for taking on a larger share of the daily clinical activities during my study period. Thanks to Andreas for all the exciting discussions and for all your comments regarding important and unimportant aspects of life. Last but not least, for those urgent sprints from Kolstien to the operation room when called upon to assist me, struggling with life-threatening bleeding metastases. Thank you, Rune, for sharing all your experience regarding spinal surgery, second-hand shopping, and how to make fried potatoes. You are a specialist by nature at whatever you decide to turn your hand to. I am left inspired. To my somewhat younger colleagues, thank you for joining the spine team. Whilst professionally, you have much in common with each other, I cannot overstate how much I have also enjoyed benefitting from your leisure interests. Each of you able to enlighten me with in-depth knowledge of your deep passions: Thomas, a 'real' fisherman, Frode, with a zest and enthusiasm for football equal to that of the current Liverpool trainer Klopp himself, Truls with a vast interest in all forms of skiing, and Eric, a low-handicap golf player from Botswana. To all of you, thank you for all experiences at work, but just as essential, for numerous running and ski trips, for football matches both played and watched together, and for birdies and bogeys at Fana Golf.

Erland Hermansen is the founder of the NORDSTEN. The NORDSTEN collaboration and being able to conduct the studies are mainly a result of your enthusiasm and extreme executive capability. Thank you, Erland, for all the pleasure

and enjoyment when planning and organizing the studies and at numerous meetings and hospital visits.

Kari Indrekvam, my co-supervisor, is the head of the Department of Orthopedic Surgery, Kysthospitalet in Hagevik and the leader of the Administrative Executive Board of the NORDSTEN collaboration. Thank you, Kari, for your cheerful as well as professional contribution to the working environment at Hagavik and in the NORDSTEN steering group. I am grateful for all your help with numerous applications and your advice and support during the entire thesis.

Rolf Gjestad, my co-supervisor, statistician, and brother in law, is first a close friend. To me, your skills in handling SPSS and Mplus are pure art. Statistics are never boring when one sits close to you. I do not know in how many family gatherings we have geeked out discussing missing data, propensity score matching, and the true definition of a confidence interval, but it has been many. I feel a huge acknowledge is due to you for spending numerous hours analyzing registry data during this period. Nevertheless, most of all, I appreciate your personality and your crazy humor and time being together with you, Bente, and your daughters.

My main supervisor Christian Hellum has been the main contributor in planning this thesis and in the interpretation of the results. Your experience from the randomized trial of disc prosthesis has been of the utmost importance. It is impossible to describe your patience, repeatedly and skillfully editing my woeful English language. You have my sympathy. Thank you, Christian, for making this PhD period amusing and enlightening. I have really appreciated spending time with Jon, Kjersti, and yourself at your house in Kjelsås.

Thanks to co-authors of the articles, Tore Solberg, Jens Ivar Brox, Margreth Grotle, and Kjersti Storheim, for significant improvement of the manuscripts. Your world-renowned expertise is the guarantor of the quality of everything you contribute.

Thanks to all other participants of the NORDSTEN collaboration. Special thanks to study-coordinators at the hospitals and central coordinators at FORMI. Without you, we would not have been able to include this high number of patients. Also, thanks to

the Norwegian spinal surgeons and patients who have faithfully completed several forms and thus made this study possible.

Thanks to the heads of the Orthopedic Department, Kari Indrekvam, Jonas Fevang, and Kjell Matre for facilitating my PhD time. Your leadership is both friendly and professional.

Thanks to the Western Regional Health Authority (Helse Vest) and the research department at Møre and Romsdal Hospital Trust for funding my PhD and the NORDSTEN study.

I want to thank Eira Ebbs, for great linguistic assistance in writing the manuscripts and the thesis. I do not doubt that you made the text more precise and readable.

I will also thank friends and family members for showing interest in my work and for frequently asking about the progression of my studies. Thanks to my mother, Marie, and my father, Hans, for all your support throughout my life. Thanks to my siblings and their spouses, Odd Inge and Herdis, Alice and Asle, and Hanne Marie and Lars, for our close relationship. Thanks to Tom, Jostein, Arve, Rune, Bjørn, and Ingmar for regular gatherings and trips. Thanks to Erna, Lars, Heidi, Harald, Tove and Graham, and Tone and Glenn, for long-lasting friendships.

I want to give a special big thank you to my wife and closest friend, Lena. When all is said and done, you have been the main reason that has enabled my time spent working on this PhD to have been a pleasurable one.

Finally, thank you, Ingrid, Mari, and Lea for contributing to my well-being with all your exciting chats and humor, for sharing meals, and for joining me in all kinds of outdoor activities. I remind myself constantly that you won't have the same immense pleasure being with me as I have to be with you. However, I am happy to indulge myself, I enjoy every second having you around me.

Content

Abstract

Lumbar degenerative spondylolisthesis (DS) is defined as an anterior displacement of one lumbar vertebra relative to the vertebra below due to age-related changes. In clinical practice, DS are most commonly concomitant to a symptomatic lumbar spinal stenosis, i.e., a narrowing of the spinal canal, especially among the older (patient) population. Typical symptoms are low back pain and pain in the buttocks and/or lower limbs. For patients with severe pain and disability, surgery with decompression of nerve structures is the accepted treatment option.

In recent decades, the rate of surgery for DS has dramatically increased, and the original method of decompression alone has been more frequently performed with an additional fusion procedure. A fusion can be performed with the use of a bone graft only (non-instrumented fusion), but screws connected to rods are usually utilised (instrumented fusion). Fusion is more invasive, more expensive and seems to have higher complication rates than decompression alone. The evidence for adding fusion to decompression is limited, as well as equivocal, and there is remarkable variation in surgical methods. In some countries, about 50% of procedures involve decompression alone, while in others an instrumented fusion procedure is included in about 90% of cases. This discrepancy in practice indicates a clear need for further high-quality studies.

The major objective of this PhD has been to investigate whether decompression alone can be an appropriate choice or if it needs to be supported by a fusion procedure. In paper I, we used registry data to compare all methods of decompression alone to all methods of decompression with fusion. The fusion group experienced somewhat greater pain reduction, but no greater reduction in pain-related disability. In this paper, we were unable to conclude that decompression alone was as good as decompression with fusion. In papers III and IV, we intended to compare the most common operation methods in Norway: micro-decompression alone and micro-decompression with instrumented fusion. We wanted to study the effectiveness of the treatments, i.e., how they work in usual conditions, as well as their efficacy, i.e., how they work in a randomised setting, to gain complementary evidence for the best

treatment for this patient group. In paper III, we found that the effectiveness of micro-decompression alone was as good as decompression with instrumentation. In the NORwegian Degenerative spondylolisthesis and spinal STENosis study (NORDSTEN-DS), a multicentre randomised controlled trial (RCT), we have included 267 patients in the period from February 2014 to December 2017. The study protocol of the RCT is included in the thesis, but the data is not yet accessible for analysis and results will not be presented here.

An important prerequisite for the comparative clinical studies was to estimate criteria for clinical 'success' following surgery. Therefore, in paper II, we searched for criteria for a clinically important outcome assessed by Patient Reported Outcome Measurements (PROMs). Cut-offs for 'success' were estimated both for spinal stenosis with spondylolisthesis and for spinal stenosis without spondylolisthesis. We found that the percentage change in a score was able to reflect the perception of being 'cured' post-surgery more accurately than the numerical change. The results from paper II were used to determine whether a patient could be classified as a 'responder' or not, which is the primary outcome in papers III and IV.

Abbreviations

DA	Decompression alone
DF	Decompression with instrumented fusion
DS	Lumbar spinal stenosis with a concomitant degenerative spondylolisthesis
EQ-5D	EuroQol 5-dimensional questionnaire utility index
FAS	Full analysis set
GPE	Global perceived effect
ITT	Intention to treat
JOA	Japanese Orthopedic Association
LSS	Lumbar spinal stenosis without a concomitant spondylolisthesis
MAR	Missing at random
MCAR	Missing completely at random
MNAR	Missing not at random
NNT	Number needed to treat
NORDSTEN	NORwegian Degenerative Spondylolisthesis and spinal STENosis
NORSpine	The Norwegian registry for Spine Surgery
NRS	Numeric rating scale
MI	Multiple imputation
NPR	Norsk pasientregister
ODI	Oswestry disability index
PPS	Per Protocol Set
PSM	Propensity score matching

PDQ	Pain Disability Questionnaire
PROM	Patient reported outcome measurement
SAP	Statistical Analysis Plan
SF-36	36-Item Short-Form Health Survey
VAS	Visual Analogue Scale
ZCQ	Zurich claudication questionnaire –score

Follow-up score = time-point value for the actual score

Change score = Follow-up score minus baseline score

Percentage change score = [(Follow-up score minus baseline score) / Baseline score] x100

List of papers

I Austevoll IM, Gjestad R, Brox JI, Solberg TK, Storheim K, Rekeland F, Hermansen E, Indrekvam K, Hellum C (2016). **The effectiveness of decompression alone compared with additional fusion for lumbar spinal stenosis with degenerative spondylolisthesis: a pragmatic comparative non-inferiority observational study from the Norwegian Registry for Spine Surgery.** *Eur. Spine Journal. doi: 10.1007/s00586-016-4683-1*

II Austevoll IM, Gjestad R, Grotle M, Solberg T, Brox JI, Hermansen E, Rekeland F, Indrekvam K, Storheim K, Hellum C (2019). **Follow-up score, change score or percentage change score for determining clinical important outcome following surgery? An observational study from the Norwegian registry for Spine surgery evaluating patient reported outcome measures in lumbar spinal stenosis and lumbar degenerative spondylolisthesis.** *BMC Musculoskel. Disord. 20:31. doi: 10.1186/s12891-018-2386-y*

III Austevoll IM, Gjestad IM, Brox JI, Storheim K, Rekeland F, Hermansen E, Indrekvam I, Solberg T, Hellum C. **Real-World Effectiveness of Micro-decompression alone versus Decompression plus instrumented fusion in Lumbar Degenerative Spondylolisthesis.** *Under review in Jama Network Open*

IV Austevoll IM, Hermansen E, Fagerland M, Rekeland F, Solberg T, Storheim K, Brox JI, Lonne G, Indrekvam K, Aaen J, Grundnes O, Hellum C (2019) **Decompression alone versus decompression with instrumental fusion the NORDSTEN degenerative spondylolisthesis trial (NORDSTEN-DS); study protocol for a randomized controlled trial.** *BMC Musculoskel. Disord. 20:7. doi: 10.1186/s12891-018-2384-0*

1 Introduction and background

1.1 Historical overview

'Spondylolisthesis' is derived from the Greek words *σπόνδυλος* (*spondylos)*, meaning vertebra, and *ολίσθησης* (*olisthesis*), a slip or a sliding (Figure 1). The term was first used by Kilian [1], an obstetrician, who described the slowly developing forward translation of a lumbar vertebra.

In clinical practice, one distinguishes between spondylolisthesis due to pathology in the neural arch, i.e., a lysis or a dysplasia, and spondylolisthesis due to degenerative changes in the facet joints and the intervertebral discs.

Spondylolisthesis without a defect in the pars interarticularis was first described by Herbert Junghanns in 1930 [2]. "Bei der echten Spondylolisthesen ist der breit klaffende Spalt im Zwischengelenkstuck des 4. Lendenwirbels die Ursache fur das Vorgleiten, wahrend bei der Pseudospondylolisthesis das Zwischengelenkstuck unversehrt ist." This knowledge was extended by Macnab (1950), who described the clinical symptoms in patients suffering from 'pseudospondylolisthesis' [3]. The term 'degenerative spondylolisthesis' was first used by P. H. Newman in 1955 [4].

1.2 Epidemiology, pathophysiology, and aetiology

In a longitudinal survey (The Copenhagen Osteorthrosis study) of 4001 Caucasian individuals selected by a random algorithm in an older population (mean age 62), the prevalence of degenerative spondylolisthesis was found to be 6%, with a female-male ratio of 5:1. Increasing age was an independent predictor of degenerative spondylolisthesis in both sexes. An olisthesis was observed at the L4-L5 level in 67% of the cases, whereas L3/L4 and L5/S1 were each represented in 15% [5]. A study on 4000 healthy, elderly Chinese (age ≥ 65 years, 2000 women and 2000 men) recruited for a population-based screening survey of osteoporotic fractures, found 25% spondylolisthesis among women and 19% among men [6]. A study of 788 white women ≥ 65 years, representing a community-living population in Pittsburgh, US,

found an overall prevalence of a radiological forward slip ≥ 3 mm of 29% [7]. A similar study of 300 US men, most of them Caucasian with a mean age of 74 years, revealed a radiological spondylolisthesis in 29% of the participants [8]. Unfortunately, the studies from China and the United States were not designed to differentiate between lytic and degenerative spondylolisthesis. These studies were included in a recent systematic review focusing on the gender- and age-specific prevalence of DS, which concluded that DS is rare before 50 years of age, is more common among women, and usually occurs in level L4/L5. The prevalence in elderly Caucasian Americans was higher than in elderly Chinese. However, the review revealed that results from the different studies vary significantly [9].

Among patients operated for DS, the mean age has been found to be 65-69 years [10-12]. More than 80% of surgeries included the L4/L5 level [10-12].

Simple dynamic considerations explain how a forward tilt of the pelvis induces shear forces between connecting vertebras in the lower lumbar region. A forward slip of the upper vertebra over the adjacent vertebra is counteracted by the muscles, ligaments, facet joints, and the interconnecting ability of the intervertebral disc. Newman and Fitzgerald have described how the pedicle, pars interarticularis, and the inferior joint facet hook over the superior facet below and counteract the downward and forward directed forces. Due to degenerative changes in the intervertebral disc and the facet joints and weakness of supporting soft tissues, the ‘hook’ loses its efficiency and a forward slippage of the upper vertebra relative to the vertebra below may occur[4, 13]. This is in accordance with the modern definition of Degenerative Lumbar Spondylolisthesis, as stated by the North American Spine Society’s (NASS) in 2016 [14]: “An acquired anterior displacement of one vertebra over the subadjacent vertebra, associated with degenerative changes, without an associated disruption or defect in the vertebral ring.”

It is accepted that degenerative changes often begin at quite a young age with changes of the intervertebral discs [15], but an understanding of the aetiology of disc degeneration is still unclear [16]. The current belief is that the degenerative process is ‘multi-factorial’, and often results in a cascade of degenerative changes with

increasing age [15]. Several initiating factors have been suggested, for example genetic predisposition [17], the response of the intervertebral disc to mechanical stress [18], and disruption of the molecular environments of the discs [19]. Unfortunately, there is limited evidence regarding the presence of causal factors responsible for disc degeneration, and the association between individual factors and progression of disc degeneration is not well documented [17].

1.3 Symptoms

Typical symptoms of spinal stenosis are pain and/or fatigue in the gluteal region and the lower extremities with or without low back pain, exacerbated when walking and in upright and extended positions, and with relief of symptoms in a recumbent position and when bending forward [20, 21]. Less common is the presence of neurological deficits such as sensory decreases and motor weaknesses. Among patients operated for spinal stenosis with a concomitant spinal stenosis, the reported rate of neurological deficits varies from 10% [22] to 25% [12, 23]. The symptoms of spinal stenosis usually have a substantial impact on patients' daily lives. Disability is typically experienced in relation to activities such as walking, exercising, travelling, social events, cooking and housekeeping, and reduced walking capacity frequently requires several kinds of walking aids [24].

It is well accepted that spinal stenosis can induce back pain, and that patients experience an improvement in back and leg pain following decompression [25, 26]. In patients with lumbar spinal stenosis, the preoperative reported level of back pain seems to be similar in patients with and without a radiological verified slip [10, 26-29]. In the Copenhagen Ostearthrosis study (4001 patients), no statistically significant relationship was revealed between an observed radiological degenerative spondylolisthesis and the presence of low back pain [5]. Contrary to the Copenhagen study, the abovementioned Chinese epidemiological study found spondylolisthesis to be associated with low back pain, statistically significant in males but not among women [6]. Since information regarding concomitant spinal stenosis was lacking in that study, the association between a forward slipclinical low back pain might be

biased due to confounding – individuals may have back pain due to spinal stenosis. In sum, one cannot eliminate the possibility that a concomitant spondylolisthesis is responsible for an individual's back pain, but it is more likely that the slip is not the main cause [5, 11, 27, 28].

1.4 Terminology

In the literature, the label 'degenerative spondylolisthesis' is commonly used alone to define the condition of patients with clinical symptoms of spinal stenosis with a radiological verified degenerative spondylolisthesis [12, 23, 30, 31]. From here on, the use of the term lumbar degenerative spondylolisthesis (DS), without the addition of 'with a concomitant lumbar spinal stenosis', refers to the simultaneous presence of spinal stenosis.

1.5 Radiology

1.5.1 Spinal stenosis

Until Magnetic Resonance Imaging (MRI) was introduced in the early 1980s, myelography, computed tomography (CT) scans, or combination of the two, were used to visualise spinal stenosis. MRI offers more detailed information regarding the nerve morphology, the cerebrospinal fluid, the epidural fat, the intervertebral discs and the soft tissues adjacent to the spinal canal (i.e., ligaments, cysts, vessels, fat and the joint capsules), and is currently the recommended imaging modality for confirming the diagnosis of spinal stenosis [24]. The common approach to visualising a stenotic spinal canal and compromised nerve structures is to use images in the axial as well as the sagittal plane (see figure 1).

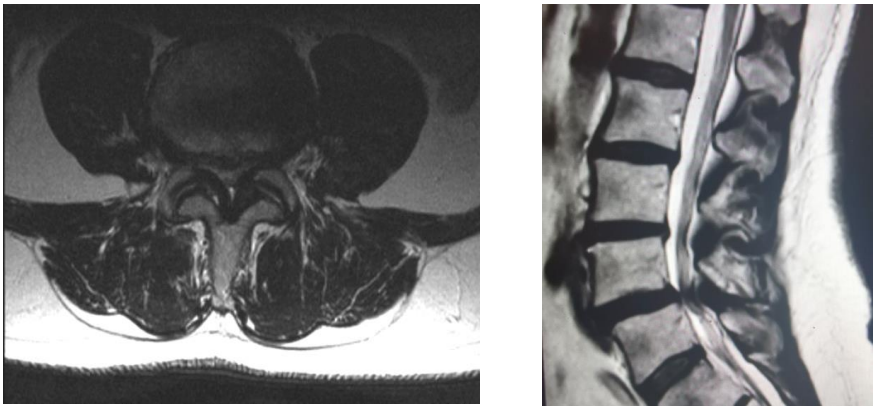


Figure 1. *MRI-scan showing spinal stenois in the axial plan (to the left) and in the sagittal plane*

Measurements used to confirm a spinal stenosis include quantitative assessments of the dural sac cross-sectional area (DSCA) [32], measurement of the length from the anterior to the posterior margin of the dural sac (A-P diameter), and qualitative grading of the dural sac morphology (i.e., Schizas classification [33]). There is lack of evidence for an association between pain intensity, functional disability, walking distance and the degree of spinal canal narrowing on MRI [20, 34]. However, a recent study among patients operated for DS showed that more severe stenosis, as measured by DSCA or by grading the morphology, was associated with better clinical 12-month outcomes [35]. A survey among Norwegian spine surgeons showed that more than 80% used the morphological cross-sectional image of the dural sac when evaluating preoperative MRI, but most of them did not measure the AP diameter or the DSCA or grade the morphology according to the Schizas classification [36].

1.5.2 Degenerative spondylolisthesis

The distinction between DS and spinal stenosis without spondylolisthesis (LSS) is based on radiological examinations (figure 2). In a research context, a forward slip of ≥ 3 mm measured on the lateral radiograph in standing position has been a commonly used criterion [23, 27, 37] for DS.

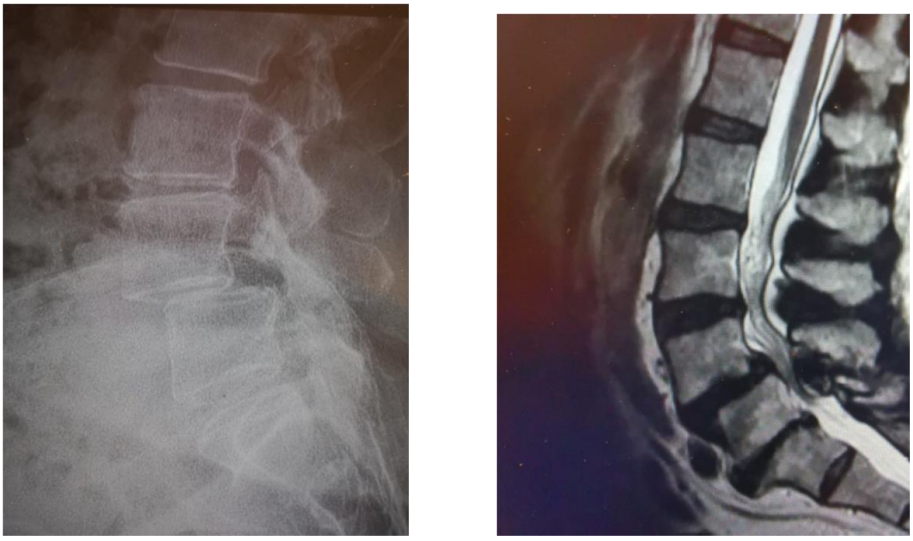


***Figure 2**. DS verified on standing lateral X-ray (to the left) and on MRI in the sagittal plane*

The terms ‘unstable’ and ‘mobile’ spondylolisthesis have also commonly been used to describe a lumbar segment’s ability to constrain dynamic translation and rotation. Several radiological definitions of instability have been proposed [38]. In clinical trials, a translation > 3mm [23] or angulation ≥ 10° [12, 23] measured on flexion-extension (flex-ex) radiographs have been used as criteria for ‘unstable’ spondylolisthesis. A recent study which aimed to evaluate factors influencing their definition of instability in DS, received responses from 226 of the 2329 American surgeons contacted [39]. Dynamic translation > 2-4 mm and a change in angulation of at least 10-15 degrees were the most frequently chosen cut-off values for instability. Less important radiological factors for defining a DA instable were disc height, facet orientation, pelvic incidence and severity of stenosis. It appears that a common radiological definition of instability does not exist, and that surgeons’ perceptions of instability diverge. Furthermore, knowledge of how often physicians routinely use quantitative assessment of a slip seems to be limited.

1.6 Treatments

Medical history and clinical examination, as well as radiological evaluation, is mandatory in diagnoses of DS. It is generally accepted that patients with tolerable

pain and without neurological palsy should wait at least 3 months before a decision to operate is made [12]. The final choice of treatment should be based on a decision-making process involving the patient and the health care provider. In such a process, it is crucial that the physician considers the patient's total burden of symptoms and complaints, as well as their expectations regarding outcome. Some patients may expect to engage in strenuous exercise and demanding sports activities whereas others are happy to comfortably perform the more limited activities of daily life. The physician is obliged to inform the patient about the probability of a satisfying result and the risk of complications and an unfavourable outcome. The most important in the decision-making process is to consider whether the expected benefits of the operation do exceed the potential risk of complications and impairment of pain and function [40].

1.6.1 Non-surgical treatment

There is limited evidence regarding the natural course of non-surgically treated DS patients. In a study by Matsunaga et al., 198 conservatively treated DS patients were followed for a period of 10 to 18 years. During this follow-up period, 53 (27%) patients were operated due to clinical deterioration, whereas 85 (43%) reported improvement of symptoms. Patients with mild symptoms at baseline had the lowest risk of deterioration [41]. In the observational cohort of the Spine Patient Outcomes Research Trial (SPORT), 33% (43 out of 130) of the conservatively treated DS patients were operated during a period of four years. The non-operated patients showed a modest mean clinical improvement at four-year follow-up [42].

Recommendations for non-surgical treatment of DS are considered to be the same as for LSS [14]. Commonly used alternatives are pharmacological treatment, physiotherapy, manipulation, lifestyle modification and multidisciplinary rehabilitation [24]. According to the NASS Guidelines for LSS [14, 20], there is insufficient evidence for or against the use of pharmacological treatment, physiotherapy and manipulation. However, the NASS working group recommend physiotherapy as an option in treatment of DS. A Cochrane review from 2016

concluded that no specific conservative treatment could be recommended over another [43].

1.6.2 Surgical treatment

Decompression

The main goal of an operation for DS is to decompress the stenotic neural elements. The original method of decompression was open laminectomy (removal of the whole lamina) and partial removal of the facet joints and ligamentum flavum [44]. An alternative to laminectomy is decompression without removal of the midline structures (i.e., the spinous process, the interspinous ligament and the lamina). Examples of such decompressions are unilateral laminotomy (in unilateral stenosis), unilateral laminotomy for bilateral decompression, bilateral laminotomy, and decompression following processus spinosus osteotomy [26, 37, 45-47] (Figures 3a to 3f). These methods are known as less- or minimal invasive decompression or micro-decompression [26, 48].

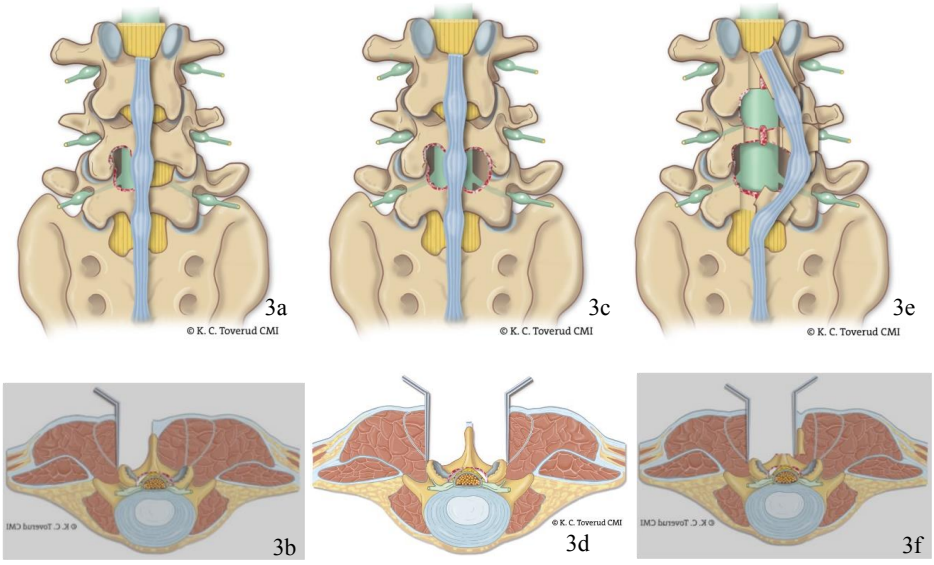


Figure 3. *Decompression without removal of the midline structures: 3a and 3b) Unilateral laminotomy; 3c and 3 d) Bilateral laminotomy; 3e and 3f) Decompression following processus spinosus osteotomy*

In a randomized trial, Thome et al. found superior results for lumbar spinal stenosis patients operated with less-invasive decompression compared with those operated

with laminectomy [49]. In a systematic review of decompression techniques for lumbar stenosis, they found some advantages for midline-preserving decompression in preventing iatrogen instability, in improvement of postoperative back pain and in perceived recovery following surgery [50]. However, the authors warned against definitive conclusions due to the limited quality of the available studies.

Despite the lack of definite scientific support for precerving the midline structures, less invasive techniques account presently for more than 90% of the decompressions in treatment of spinal stenosis in Norway [51]. Less-invasive decompression was also preferred over laminectomy for management of spinal stenosis in a survey among Dutch spine surgeons [52].

Decompression plus fusion

Several influential authors hold the common belief that a decompression should be followed by a fusion procedure in order to reduce symptoms and prevent further slippage due to 'instability' when treating patients with DS [12, 23, 53]. A fusion procedure can be performed with the use of bone without instrumentation and with a bone graft accompanied by screws and rods (i.e., instrumented fusion). Further, an instrumented fusion can be performed with a cage being inserted into the intervertebral space (figure 4). The aim of all fusion methods is to achieve a solid bony bridge between the slipped vertebra and the vertebra below it. Regarding clinical outcome, there is no evidence for or against additional instrumentation [54-56], but using screws and rods has been recommended to provide a radiological fusion [56], and has been preferred over non-instrumented fusion for the past two decades [10, 57, 58].

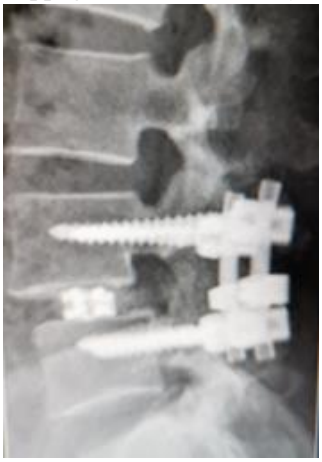


Figure 4

Decompression alone versus decompression plus fusion

The most important impetus for the growing use of additional fusion in decompression probably comes from three studies performed between 1991 and 2004. In 1991, Herkowitz et al. found more pain reduction when a laminectomy was supported by a non-instrumented fusion [59]. This study was followed by that of Fischground et al. (1997), who reported higher rates of radiological union, but no differences in clinical outcomes, when a bone graft was supported by instrumentation [60]. Finally, in a long-term follow-up study of patients from Herkowitz's and Fischground's cohorts, Kornblum and co-workers (2004) used phone interviews to evaluate patients operated with decompression plus non-instrumented fusion 5 to 14 years after surgery [54]. A comparison of improvement in leg and back pain in those with and without radiological union (i.e., a successful fusion) showed that successfully fused patients had more pain reduction than patients with radiological non-union. Based on their findings they recommended treating DS with decompression plus instrumented fusion [54]. The conclusion was debated against by Katz when commenting on the paper in the same number of *Spine* [54]. His main criticism was that only the one arm (decompression with non-instrumented fusion) of the original cohort was studied. By not including the original control groups, the study design introduced a severe risk of biases. A better design would have compared the long-term outcomes in the original arms (decompression alone versus decompression with non-instrumented fusion and decompression with non-instrumented fusion versus decompression with instrumentation). Katz argued that although higher fusion rates achieved by instrumentations might reduce back and leg pain, the instrumentation might cause pain that would violate the benefits of a successful radiological union.

Other small cohort studies published before 2000 also recommended additional fusion when operating on patients with DS [61-63]. Further, two systematic reviews both upheld instrumented fusion as the best approach in treatment of DS, especially among those with a radiologically "unstable" spondylolisthesis [55, 64]. The

proportion of patients operated with decompression with instrumented fusion increased from about 50% to nearly 90% from 1999 to 2011 in the United States. Correspondingly, decompression alone, as well as decompression with non-instrumented fusion, was performed less frequently in this period [57]. This practice was only partially in accordance with the 2008 evidence-based clinical guideline from the North American Spine Society. They recommended supporting decompression with fusion to improve clinical outcome, but did not find evidence that instrumented fusion would improve clinical outcomes compared to non-instrumented fusion [65].

Reports from the Scandinavian spine registries have shown a significant practice variation between the nations. In Sweden and Norway (2011-2013) about 50% of surgeries included a fusion procedure, whereas in Denmark fusion was utilized in nearly 90% of surgeries [11, 22, 66].

When planning this PhD project (2010 to 2013), the literature suggested that decompression with fusion was the best treatment option [55, 56, 64, 67], and the reports of national trends indicate that instrumentation was the preferred fusion method [57, 68, 69]. However, although systematic reviews and clinical guidelines [55, 56, 64, 67] recommended fusion as the preferred method, this was only supported by low-quality evidence, according to the Grading of Recommendations Assessment, Development and Evaluation (GRADE) [70]. Contrary to the abovementioned studies, a Swedish registry study did not find superior effectiveness of additional fusion compared to decompression alone [11]. Based on the limited and conflicting evidence for the best treatment, a need for more high-quality research was warranted [55, 57].

1.7 Assessment of clinical results

Treatment effects following spinal surgery are most commonly assessed by patient reported outcome measurements (PROMs) [10, 23, 27, 71, 72]. A PROM is a self-administrated questionnaire that reflects patients' experiences of the treatment effect, such as pain, pain-related disability, quality of life, and their global assessments of

the perceived effects of the treatment. An important consideration in the interpretation of outcomes is the clinical importance of results assessed by PROMs. A statistically significant mean change from baseline to follow-up reflects the amount and variability of the treatment effect as well as the size of the studied population, but does not necessarily provide meaningful clinical information [73]. To determine whether an improvement is of clinical relevance, thresholds for a clinically important benefit following treatment are introduced [74]. The minimal clinically important difference (MCID) [74], a substantial clinical benefit [75] and a satisfactory symptom state [76] are examples of metrics developed to determine whether a patient has benefited from treatment or not. Several authors have pointed out great variability and diversity in such thresholds [77-79], which may be caused by heterogeneity in studied populations [80]. In order to ensure reliable and valid estimates for a clinically important outcome when comparing treatment effects, thresholds should be derived from a context as similar to the trial as possible, i.e., in a similar patient population [73].

1.8 Different study designs

Randomised controlled trials provide high-level evidence for detecting differences in treatment effects (i.e., efficacy) solely related to the delivered treatments. However, a weakness is the limited external validity of the results. Due to strict criteria for inclusion and exclusion, many RCTs only investigate subgroups within the population of interest. Hence, the results can only be applied to a small part of the whole patient group [81-83]. To provide knowledge about ‘effectiveness’, i.e., the performance of the treatments used in daily clinical conditions, studies with data from comprehensive registries are recommended [81, 84]. Such ‘real-world knowledge’ is considered to be important for bridging the evidentiary gap between clinical research and practice [84-86].

2 Aims of the thesis

The primary objective of this thesis was to investigate whether decompression alone could be considered 'good enough' for the large majority of patients with DS, or if the more invasive and expensive method of decompression with fusion should be advocated as the preferred method. Firstly (paper I), we aimed to compare the effectiveness of the two principal methods: decompression alone versus decompression with fusion. Since the less invasive methods of decompression alone (micro-decompression) and the method of decompression plus instrumented fusion account for the majority of DS surgeries in Norway, we intended to compare these treatments in a multicentre RCT (paper IV) as well as in a registry study (paper III). The results from the RCT are not yet available, and consequently have not been included in this thesis. The published study protocol, however, is included in order to describe the importance of gaining knowledge both from situations in which treatments are chosen in a daily clinical setting and from a setting where treatments are selected through randomisation.

An important prerequisite for the assessment of the clinical outcome was to estimate the proportion of responders in each group. In order to calculate the responder rate, it was important to establish condition-specific criteria (cut-offs) for a substantial clinical improvement in PROMs.

2.1 Paper I

To compare the clinical effectiveness of the two principal surgical methods used in DS, decompression alone (includes all surgical methods for decompression) and decompression with fusion (all methods for fusion) based on data from the Norwegian Registry for Spine Surgery (NORSpine).

2.2 Paper II

To evaluate which criteria for 'success', assessed by Oswestry Disability Index (ODI), EuroQqol 5 dimensional descriptive system (EQ-5D) and Numeric Rating Scale (NRS) for leg and back pain that best match the patients' experience of being cured following an operation. The estimated criteria should be a used in study III and IV to determine whether a patient obtains a clinical important benefit following surgery or not.

2.3 Paper III

To compare the clinical effectiveness of micro-decompression alone versus decompression supplied with an instrumented fusion. The rationale for this study was to evaluate how the most commonly used methods in Norway perform when used in normal clinical conditions. Furthermore, an important rationale for this study was to compare the real-world effectiveness of similar treatment groups as investigated in the RCT (paper IV).

2.4 Paper IV

To conduct a national randomised controlled trial (RCT) for comparison of micro-decompression alone and decompression with instrumented fusion. The study protocol describes the objectives of the trial. The primary aim was to gain level 1 evidence regarding treatment efficacy of the investigated methods during the first two years after surgery. Secondary aims were to investigate whether a predictor model for the best treatment option for an individual can be developed, and to gain evidence for treatment efficacy in a longer time perspective (five and 10 years). In addition, we have planned to compare the cost-efficacy of the two treatments.

3 Materials and methods

3.1 Study design

Papers I, II and III: Multicentre observational cohort studies based on prospectively collected data.

Paper IV: A multicentre randomized controlled study. The study protocol is presented.

3.2 Patients

3.2.1 The NORSpine database

In studies I, II and III, patients were recruited from the Norwegian Registry for Spine Surgery (NORSpine) database, a national comprehensive clinical registry for quality control and research. The registry's coverage rate has increased since it was first put in place in 2007. In 2010, 30 out of 40 hospitals performing spine surgery reported to the registry whereas all 41 hospitals reported in 2016. Further, the coverage rate at an individual level was 42% in 2010 and 70% in 2017 [51]. Due to a well-known lower reporting rate of emergency surgeries and weekend surgeries, the coverage rate for elective surgeries, such as spinal stenosis and degenerative spondylolisthesis most probably exceeds 70% [51].

In papers I and II, patients operated in the period from 2007 to 2013 were included. In paper I, only DS patients were included and all types of decompression alone and decompression plus posterior fusion were included. In paper II, both DS patients and LSS (i.e., spinal stenosis without spondylolisthesis) patients were included. For both studies, patients formerly operated at index level(s) were excluded.

In paper III, patients operated for DS with micro-decompression alone or decompression with instrumented fusion in the period from 2007 to 2015 were included. Exclusion criteria were previous surgery at the index level(s), surgery in

more than two levels, and operation with an inter-spinous device or with an anterior approach.

3.2.2 The NORDSTEN database

The NORwegian Degenerative Spondylolisthesis and spinal STENosis (NORDSTEN) study is a multicentre study involving patients recruited from 18 Norwegian hospital departments. In addition to the present trial (NORDSTEN-DS), it consists of the NORDSTEN-SST (Spinal Stenosis Trial); a randomized controlled trial comparing the clinical and radiological results in three different decompression techniques in patients with LSS, and the NORDSTEN-OS; a prospective observational study aiming to gain knowledge of the natural course in non-operated patients with LSS and DS. A total of 993 patients are included in the NORDSTEN database.

Prior the start of the NORDSTEN-DS trial, surgeons and study coordinators at the participating hospitals were educated regarding the background and aims of the study and the scientific principles for performing an RCT, and received instruction regarding eligible criteria and the interventions. This education was provided at joint meetings, through visits to the participating hospitals by members of the NORDSTEN steering group, and in information sent by mail. An independent monitor methodically instructed the study coordinators in the principles of Good Clinical Practice (GCP). Prior inclusion to the study, patients received information about the background, aims, alternative treatments, and the voluntary nature of participation. If not willing to participate, the patients were offered treatment according to the surgeon's preference and the practice at the department. Patients evaluated for eligibility were registered in a "Screening Form". If eligible, a 'Consent form' was signed by the patient and by the surgeon who confirmed that spoken and written information had been given. Prior to randomization, the surgeons completed an "Inclusion to Study" form, a checklist for inclusion and exclusion. The local study coordinator received the results of the concealed computer-generated randomization from the central coordinator at FORMI by phone and by email. The treatment

allocation was documented in the patient's record. Patients, investigators and surgeons were unable to influence the randomisation.

Inclusion and exclusion criteria are listed in table 1. One amendment was made to the original eligibility criteria. Originally, ODI score < 25 was an exclusion criterion, but, due to input from participating surgeons that many patients suffering from leg and back pain scored less than 25 ODI points, the NORDSTEN steering group decided that from the 29th of August 2015 no ODI limit should be claimed.

Table 1. Eligible criteria for the NORDSTEN-DS trial

Inclusion criteria:	**Exclusion criteria:**
• Over 18 years of age. • Understand Norwegian language, spoken and written. • Spondylolisthesis, with a slip >=3 mm, verified on standing plain x-rays in lateral view. • Spinal stenosis in the level of spondylolisthesis, shown on MRI, CT scan or myelogram. • Clinical symptoms of spinal stenosis as neurogenic claudication or radiating pain into the lower limbs, not responding to at least 3 months of qualified conservative treatment. • Be able to give informed consent and to respond to questionnaires.	• Not willing to give written consent. • Participating in another clinical trial that may interfere with this trial. • ASA- grade > 3. • Older than 80 years. • Not able to fully comply with the protocol, including treatment, follow-up or study procedures (psychosocially, mentally and physically). • Cauda equina syndrome (bowel or bladder dysfunction) or fixed complete motor deficit. • A slip >=3 mm in more than one level. • An isthmic defect in pars interarticularis. • Fracture or former fusion of the thoracolumbal region. • Previous surgery in the level of spondylolisthesis. • Lumbosacral scoliosis of more than 20 degrees verified on AP-view. • Distinct symptoms in one or both legs due to other diseases, e.g. polynevropathy, vascular claudication or osteoarthtritis. • Radicular pain due to a MRI-verified foraminal stenosis in the slipped level, with deformation of the nerve root because of a bony narrowing in the vertical direction.

3.3 Data

3.3.1 Papers I, II and III

At admission to surgery, the patients completed a preoperative baseline questionnaire. The questionnaire included PROMs, patient characteristic such as age, gender, smoking habits, BMI, length of education and use of pain medication. During the hospital stay, surgical parameters such as diagnosis, comorbidity, surgical methods, complications and the length of surgery and hospital stay were recorded in the surgeon form. Before data was entered into the NORSpine database, all questionnaires were checked for completeness by dedicated study coordinators. The follow-up questionnaires for assessment of PROMs were sent by mail from NORSpine and completed by the patients three and 12 months after the operation.

3.3.2 Paper IV

The preoperative NORDSTEN questionnaire is based on the NORSpine baseline form. Some additional information is collected, such as assessment of psychological variables and more extensive measurement of disorder-specific disability. Follow-up questionnaires are completed at three months, 12 months, two years, five years and ten years. Research coordinators at each department manage the practical details regarding the registration of complications and reoperations and further collection and submission of completed follow-up forms to the central coordinator at the Section for Musculoskeletal Research (FORMI), at Oslo University Hospital. The coordinators record data from the hospital stay, and from postoperative follow-ups in Case Report Forms (CRFs). The CRFs record reoperations and complications in the periods between each follow-up. The data are stored at the Faculty of Research support, University of Oslo. Study coordinators and research personnel at FORMI and the Faculty of Research are not involved in the scientific aspects of the study. Radiological exams (X-rays including dynamic examination, MRI, and CT) are collected and stored for the assessment of predefined variables. This assessment is being performed by two surgeons and two radiologists. An overview of the collected data is shown in table 2.

Table 2. Time-frame regarding data collection

	Before operation	Hospital stay	3 months (±2 weeks)	12 months (±1 month)	2 years (±2 months)	5 years (±3 months)	10 years (±3 months)
Demographics[1]	I-IV						
PROMs[2]							
ODI	I-IV		I-IV	I-IV	IV	IV	IV
EQ-5D	I-IV		I-IV	I-IV	IV	IV	IV
NRS leg/back pain	I-IV		I-IV	I-IV	IV	IV	IV
GPE score			I-IV	I-IV	IV	IV	IV
Zurich CQ	IV		IV	IV	IV	IV	IV
HSCL-25[3]	IV						
X-ray[4]	IV		IV				
MRI scan[5]	IV						
CT scan[6]					IV		
Operation data[7]		I-IV					
Data from hospital stay[8]		I-IV					
Complications[9]		I-IV	I-IV	IV	IV	IV	IV
Reoperations[9]		IV	IV	IV	IV	IV	IV

[1] Demographics: Age; Gender; BMI; ASA; Education; First language; Smoking; Former surgery; Analgesics.
[2] Patient reported outcome measurements (PROMs) are described in the 'Outcome Measures' section.
[3] Hopkins symptom checklist (HSCL-25) [87] is a self-reported questionnaire for assessment of psychological variables, and will be collected preoperatively. It includes 25 items (i.e., questions) regarding emotional distress which are ranged from 1 to 4, with lower scores indicating less severe symptoms. The questionnaire has been completed at baseline in study IV for use in the predictor study.
[4] X-rays: Degree of spondylolisthesis [88], Segmental instability [88]; Lumbal lordosis [89]; Pelvic incidence [89].
[5] MRI: Grading of spinal stenosis (Schizas A-D) [33]; Presence of foraminal stenosis [90]; Amount of facet joint fluid [91]; Disc degeneration [92]; Modic changes [93].
[6] CT scan: Evaluation of fusion grade.
[7] From surgeon forms: Level(s) operated on; Number of level(s) operated on; Method used for decompression; Method used for instrumented fusion; Operation time; Blood loss; Complications.
[8] From Case Report Forms completed by coordinators: Level(s) operated on; Number of level(s) operated on; Method used for decompression; Method used for instrumented fusion; Operation time; Blood loss; Complications; Length of hospital stay; Control for operated level and side; Control for whether used method was in accordance with protocol.
[9] From Case Report Forms fulfilled by coordinators.
Roman numerals indicate studies I to IV

Monitoring of data collection

The trial is monitored following the Helsinki Declaration, The International Conference on Harmonisation Guideline for Good Clinical Practice (ICH GCP). An independent monitor, without influence on the scientific work, and not otherwise involved in the study, is responsible for the monitoring. Due to the non-regulated ICH GCP guideline for this trial (not including drug intervention), the risk and safety are safeguarded at the same level as data quality. All informed consent forms are being checked, and all registrations of serious events are monitored. According to the monitoring plan, selected variables are being checked. All hospitals are being visited regularly. Adapted versions of the 'Investigator's Site File (ISF)' and the 'Trial Master File (TMF)' are being checked for essential documents during the trial. Queries and deviations are being recorded and reported, and the coordinators at responsible hospitals have two months to send a written report with the required corrections to the monitor. All deviations from the protocol are subsequently being recorded in the 'Note to file form'.

3.4 Outcome measures

3.4.1 Patient reported outcome measurements

Oswestry Disability Index (ODI) V.2.0

ODI is a self-reported instrument comprising 10 items connected to pain and pain-related disability in activities of daily life [94]. For each item, six alternatives are presented in increasing order of disability. The patient is supposed to check off the most appropriate alternative. For each item, zero represents no disability, and ten represents the greatest impairment. The ODI score is calculated as the sum of the responses divided by the number of items responded to. Hence, total ODI scores range from 0 to 100, where 100 represents the greatest impairment.

The original questionnaire (version 1.0) was developed to measure disability related to back pain, and was first presented by Fairbank in 1980 [94]. This questionnaire, or modified versions, were later translated and validated for use in several languages [95]. The different versions of ODI were re-evaluated in 2000 by Fairbank and

Pynsent [96]; their publication is ranked as the third most cited paper in the field of lumbar spine surgery [95]. The authors concluded that "ODI remains a valid and vigorous measure of condition-specific disability". They recommend the use of version 2.0 [96]. This version has been translated into Norwegian, and its reliability, as well as its validity, have been found to be acceptable for the assessment of functional disability in the Norwegian population [97].

Zürich Claudication Questionnaire (ZCQ)

ZCQ is also known as the Swiss Spinal Stenosis Questionnaire, and is a self-completed disorder-specific score consisting of three domains: symptom severity (7 items ranging from 1 to 5), physical function (5 items ranging from 1 to 4), and patient satisfaction (6 items ranging from 1 to 4). A score of one point represents the lowest symptom severity, least physical impairment and best patient satisfaction [98]. The ZCQ has been translated into Norwegian by Thornes and Grotle for use among spinal stenosis patients [99]. In a methodology study, they found the reliability to be 'very good', the validity 'acceptable' and the responsiveness (i.e., the sensitivity to detect a change in symptoms or disability) to be 'good' [99].

Numeric Rating Scale (NRS) for leg pain and for back pain

NRS is a PROM that assesses self-reported pain experienced by an individual in the last week. Patients are asked to rate their pain during the last week on an 11-point Likert scale, ranging from 0 (no pain) to 10 (the maximum imaginable pain) [100]. The measurement has acceptable reliability as well as validity [101], is easy to understand, and has shown high test-retest reliability [102].

EuroQol 5 dimensional descriptive system (EQ-5D)

EQ-5D is a generic PROM that is self-completed and comprises five questions relating to mobility, self-care, usual activity, pain/discomfort, and anxiety/depression [103]. Each question has a three-point descriptive scale where 3 represents the worst possible health. Each possible combination of responses ($3^5 = 243$) represents a score between -0.59 and 1.0, with higher scores indicating better quality of life. Its validity, reliability, and responsiveness have been evaluated to be acceptable for Norwegian

patients operated for lumbar degenerative disorders [104]. EQ-5D is commonly used to compare the cost-effectiveness of different treatments [105].

Global Perceived Effect (GPE) scale

Patient-rated satisfaction with treatment outcome will be assessed using a single question with a seven-point descriptive scale including the answers 'completely recovered', 'much improved', 'slightly improved', 'unchanged', 'slightly worse', 'much worse' and 'worse than ever'. This scale is easy for patients to understand and to answer, and clinically relevant for physicians and for patients [106, 107]. Further, its test-retest reliability was found to be high in a study by Kamper et al. [107]. The Norwegian version utilized in the present studies has been evaluated by Grovle et al. [106]. They found that the GPE scale correlated well with other simultaneously assessed PROMs (VAS leg pain, VAS back pain and disability assessed by the Maine Seattle Back Questionnaire and Short Form-36 Health Survey (SF-36).

3.4.2 Other outcome measures

Duration of surgery

The time from opening to closing the skin, assessed in minutes.

Length of hospital stay

The time from surgery to hospital discharge, assessed by number of days.

Complications

Perioperative complications such as dural tears and nerve lesions, and complications during hospital stay, such as hematoma, misplaced implants, and cardiac and pulmonary complications were recorded during the hospital stay. The occurrence of complications such as wound infection, pneumonia, and thrombosis are in studies I and III reported by patients on the three-month questionnaire and in the RCT on each follow-up.

Reoperations

A new operation at the same level as the primary operation is noted as a reoperation.

3.4.3 Primary and secondary outcomes

Paper I

Primary outcomes were ODI, NRS leg pain and NRS back pain, evaluated both by mean 12-month follow-up scores and by the proportion of individuals with improvement from baseline to follow-up equal to or greater than MCID (12.8 points for ODI, 1.2 for NRS back pain and 1.6 for NRS leg pain [108]). Secondary outcomes were: 1) GPE scores trichotomised into 'substantially improved' ('completely recovered' and 'much improved'), 'unchanged' ('slightly improved', 'unchanged' and 'slightly worse') and 'substantially deteriorated' ('much worse' and 'worse than ever'); 2) duration of surgery; 3) length of hospital stay; 4) the rate of surgeon- and patient reported complications.

Paper II

This was a methodical study where the PROM variables ODI, EQ-5D, NRS leg pain and NRS back pain were evaluated against the responses from the GPE scale. We did not evaluate differences in outcomes between treatment groups. The area under the Receiver Operating Characteristics (ROC) curve was utilized to compare the ability of the different PROM instruments to determine whether a patient was a 'success' or 'non-success'.

Paper III

Primary outcome was the ODI, evaluated by the proportion of individuals with a reduction from baseline of 30% or greater in ODI score at 12-month follow-up [109]. The change scores from baseline to 3 months, from 3 to 12 months, and the 12-month follow-up scores in ODI, NRS leg pain and NRS back pain were secondary outcomes. Additional secondary outcomes were GPE scores, the duration of surgery, the length of hospital stay and the surgeon- and patient reported complications.

Paper IV

Primary endpoint is a reduction from baseline of 30% or greater in the ODI score at two-year follow-up. A reduction of ≥ 40% in the NRS leg pain, ≥ 33% in the NRS back pain [109], and an improvement in the ZCQ ≥MCID [98] are secondary

outcomes. Additional secondary outcomes are defined as: mean change scores and follow-up scores in the ODI, the NRS leg pain, the NRS back pain, the ZCQ, the EQ-5D, the GPE scale scores, the duration of surgery; the length of hospital stay; the surgeon- and patient reported complications; the volume of blood loss and blood substitution; and the rate of reoperations at the index level.

For the predictor analysis and the long-term follow-up, the primary outcome is similar to the primary endpoint in the 2-year follow-up. EQ-5D will be the primary outcome in the cost-efficacy analysis.

3.5 Interventions

3.5.1 Paper I

a. Decompression alone: All methods of decompression.

b. Decompression with an additional fusion: All methods used for posterior fusion (non-instrumented and instrumented).

3.5.2 Paper II

Not a comparative study of treatment effects.

3.5.3 Papers III and IV

a) Micro-decompression alone: A decompression preserving the midline structures was mandatory. The surgeon used a microscope or magnifying glasses.

b) Decompression and instrumented fusion: An optional technique for decompression was followed by posterolateral pedicle screw fixation with or without an additional cage. In study IV the surgeon used a microscope or magnifying glasses, this was not mandatory for study III.

3.6 Statistical methods

3.6.1 Paper I

To compare the responder rates (i.e., proportion of patients with a change score > MCID [108]), tests for non-inferiority were performed for each of the three

measurements. The null hypothesis (H0) was that the proportion of responders in the decompression alone group (n_{DA}) was at least 15 percentage points lower (the non-inferiority margin) than for the instrumented fusion group (n_{DF}); **H0: $n_{DF} - n_{DA} \geq 15$** [110, 111]. This margin corresponds to a Number Needed to Treat of seven patients (NNT = 1/0.15 = 6.67), i.e., at least seven patients need to be fused to achieve one additional responder [112]. H0 was tested by forming a 95% confidence interval (CI) for the between-group difference in responder rate, and was to be rejected if the upper bound of the CI was below the non-inferiority margin of 15 percentage points. The alternative hypothesis was that the decompression alone group was non-inferior, i.e., as good as, decompression with fusion; **H1: $n_{DF} - n_{DA} < 15$**.

In order to reduce the risk of allocation bias, propensity score matching (PSM) was used to make the distribution of observed baseline variables as similar as possible between the groups[113]. The propensity score reflects the patient's probability for being fused. We considered that the following baseline covariates might influence treatment allocation: age, gender, body mass index, smoking, ODI, NRS leg pain, NRS back pain, EQ5D, the presence of foraminal stenosis, degenerative disc disease, scoliosis, predominating back pain, number of level operated on, and neurological palsy. These variables were used in a logistic regression model to estimate the propensity score. Pairs (one from each treatment group) with similar propensity scores were formed, and patients without a 'match' were excluded.

The mean change and follow-up scores in ODI, NRS leg pain and NRS back pain scores were estimated by Latent Growth Curve (LGC) models with Full Information Maximum Likelihood [114]. The models were specified as latent difference score models, including the changes from baseline to three months, from three to 12 months, and the level at 12-month follow-up.

Statistical Package for Social Sciences (SPSS) version 22.0 was used for testing distribution of data, cross-tabulations with χ^2 test, Student t-tests and Mann-Whitney U tests, and for the PSM. Mplus 7.3 was used for analysing LGC models [115].

3.6.2 Paper II

Standard descriptive statistics were used to estimate centrality and variability of baseline characteristics. Baseline and 12-month follow-up scores in ODI, EQ-5D and NRS for leg and back pain were analysed according to the external anchor, the Global Perceived Effect (GPE) scale. For each PROM, three alternative instruments were evaluated: 1) The (raw) follow-up score; 2) The numerical change score (the absolute change from baseline to follow-up); 3) The percentage change score (the change score as a percentage of the baseline score). The ability of the instruments to discriminate between those reporting 'completely recovered' and 'much improved' from those reporting 'slightly improved', unchanged', 'slightly worse', 'much worse', and 'worse than ever' was evaluated by calculating the area under the ROC curves (AUC) [116]. The accuracy is classified as 'excellent' for AUC from 0.9 to 1.0, 'good' from 0.9 to 0.8, 'fair' from 0.8 to 0.7, 'poor' from 0.70 to 0.60, and 'failed' from 0.60 to 0.50 [117]. Further, ROC analyses were performed to estimate optimal cut-offs for 'success' for the different PROM instruments. Optimal cut-offs were values that maximised the percentage of correctly classified patients according to the anchor. Finally, we evaluated how accurately the estimated cut-offs correctly classified patients with low, medium and high PROM values at baseline. Since responder rates were to be assessed in study III and IV, and in other NORDSTEN-studies [37], it was important to define thresholds for both DS and LSS. The data were analysed using the Statistical Package for Social Sciences (SPSS) version 23.0 and Stata version 14.0

3.6.3 Paper III

An individual with a 30% or greater reduction in the ODI score from baseline to one-year follow-up was defined as a responder. Propensity score matching was performed for adjustment of differences in observed baseline variables. Descriptive statistics, including measures of centrality and variability were used to describe baseline characteristics for the unmatched and the matched cohort. The between-group differences in responder rates were estimated with the Newcombe hybrid score CI [118] both for the unmatched and matched cohort. To test non-inferiority the difference in responder rates was estimated with the Newcombe hybrid score 95% CI

[119] with a non-inferiority margin of 15 percentage point. The treatment groups were compared in both the unmatched and the matched cohort.

For the matched cohort, means and standard errors for change scores from baseline to three-months, from three- to 12-months, and scores at 12-month follow-up, were estimated by Multi-sample LGC models [114].

Due to an expected loss to follow-up of approximately 80%, an additional LGC analysis was performed following Multiple Imputation [120] of missing data. In previous studies from NORSpine, loss to follow-up was estimated to be about 20% [22, 26, 45]. Consequently, the LGC models were also analysed following Multiple Imputation (MI) [120] for missing data. Seventydata sets were generated to create complete follow-up scores for ODI, NRS leg pain and NRS. Baseline patient characteristics, operation time, length of hospital stay, baseline and follow-up scores for ODI, NRS leg pain, NRS back pain, Eq-5D, GPE, length of hospital stay, duration of surgery, and complications was used as predictors in the imputation model.

SPSS version 24 was used for descriptive statistics, analyses of continuous variables with Student-t tests or Mann-Whitney tests, depending on the distribution of data, analyses of binary variables with Fisher mid-P tests and Newcombe hybrid score confidence intervals [118], and for propensity score matching. The LGC analyses were performed with Mplus 8 [121].

3.6.4 Paper IV

Statistics for the determining efficacy

The primary outcome will be tested according to non-inferiority for micro-decompression alone at two-year follow-up. As in study III, the cut-off for being a responder is a 30% or greater improvement in ODI from baseline to follow-up. The difference in responder rates will be estimated with the Newcombe hybrid score CI and a non-inferiority margin of 15 percentage points is confirmed.

The comparison of treatment efficacy will be based on a Full Analysis Set (FAS) and a Per Protocol Set (PPS). For micro-decompression to be considered to be as good as

decompression with instrumented fusion, both the FAS and the PPS analysis are required to show non-inferiority.

In the FAS all randomised patients with primary operation according to the randomly assigned study treatment and with data on the primary outcome variable (ODI) at one or more time point(s) will be included. Missing scores necessary for dichotomising patients into responders/non-responders will be imputed by use of Multiple imputation (MI). The imputation model, using linear regression, will include the following explanatory variables: Baseline patient characteristics (age; gender; education; first language; smoking; body mass index; former spinal surgery; duration of pain; use of analgesics), radiological parameters at baseline (degree of the slip; segmental instability; Schizas grade; orientation of facet joint; disc height), operation time, length of hospital stay, baseline and follow-up scores for ODI, NRS leg pain, NRS back pain, Eq-5D, ZCQ, GPE, duration of surgery, length of hospital stay, complications, and reoperation. The imputation will be stratified by treatment group[122]. The multiply imputing will be performed before dichotomising, as recommended [123], and will generate 50 data sets with complete two-year follow-up scores for ODI, ZCQ, NRS leg pain and NRS back pain. Before the responder analyses the imputed scores will be estimated based on the 50 aggregated data sets.

The PPS will exclude patients if they: 1) Have not received operative treatment in accordance with randomized allocation; 2) Have received operative treatment in accordance with randomized allocation, but were re-operated at the same level during the follow-up period. 3) Withdrew their informed consent and asked for their data to be withdrawn from the analyses.

In addition, we will perform two sensitivity analyses: One with responder analysis of FAS without imputation (a complete case analysis), and one with responder analysis of FAS, where missing values will be replaced with values at one year follow-up, if available. Categorical secondary outcomes will be analysed with Fisher mid-P tests and Newcombe hybrid score.

Linear mixed models (LMM) will be used to estimate the between-group difference in level and change in continuous secondary outcome variables. Outcome measurement at baseline, three month-, and one-year- and two-year follow-up will be included in the models. Because most change from baseline is expected to occur in the first three months, the time development in the linear mixed models will be modelled as piecewise linear, with a knot at three months.

Statistics for secondary objectives

Predictor analysis

The predictor analysis will be performed using the pragmatic model-building approach of Hosmer et.al [124]. Patients treated with micro-decompression alone and decompression with fusion will be analysed. The following baseline variables will be tested for their association with the primary outcome variable 'responder': treatment; age; gender; comorbidity (ASA group); body mass index; smoking; ODI score; NRS back pain score; NRS leg pain score; HSCL-25 score; magnitude of olisthesis; segmental instability; presence of foraminal stenosis; orientation of the facet joint; amount of facet joint fluid; disc degeneration; disc height in the level of olisthesis; lumbal lordosis; pelvic incidence. From the final models, the predicted probability of being a responder will be estimated for each combination of the covariates. The risk estimates will be used for building matrixes for an individual's overall risk of being a responder, depending on which treatment is choosen (i.e., decompression alone or decompression with instrumented fusion). Previously, risk matrix models for predicting probability, given a set of established predictors, has been constructed for other medical conditions [125, 126].

Long-time follow-up analyses

For comparing the efficacy of the treatments at five- and 10-year follow-up we will use the same statistical methods as in the analyses at two-year follow-up.

Further details

Detailed information regarding statistical methods for study IV are recorded in ClinicalTrials.gov (Identifier: NCT02051374; 'Statistical Analysis Plan for NORDSTEN-DS').

3.6.5 Sample sizes

For the comparative studies, the sample size was computed by using the Blackwelder methodology [127]. A type 1 error (α) = 0.05 and a non-inferiority limit (δ) of 15 percentage points were set for the studies. Based on these assumptions, the required group sizes were computed to 155 for study I (power = 0.80, dropouts = 25%), 196 for study III (power = 0.90, dropouts = 25%), and 128 for study IV (power = 0.80, dropouts = 10%). Power calculations were performed by using https://www.sealedenvelope.com/power/binary-noninferior/

4 Results

4.1 Paper I

Following the eligibility criteria, 616 patients were included in the study. After propensity score matching, 260 patients from each treatment group remained for analysis. Of those, 73% returned the forms at three months, 85% at 12 months and 94% had at least one follow-up registration.

Fifty-nine percent of the decompression alone group and 67% of the fusion group achieved a clinically important improvement in the ODI (12.8 points), a difference in the proportion of responders of 8 percentage points. This difference corresponds to a Number Needed to Treat of 12 patients. For NRS leg pain and NRS back pain, the responder rate was 7% (67% vs 74%) and 11% (63% vs 74%) lower in the decompression group than in the fusion group, respectively. The upper bounds of the 95% CI for differences in responder rate were 18% for ODI, 16% for NRS leg pain, and 20% for NRS back pain, all of which exceed the proposed limit for non-inferiority of 15 percentage points. Hence, the null-hypothesis could not be rejected, and we could not claim the effectiveness of decompression alone to be statistically significantly non-inferior to decompression plus fusion.

The fusion group rated their pain slightly but statistically significantly lower than the decompression alone group (leg pain 3.0 and 3.6 respectively, mean difference -0.6, 95% CI -1.2 to -0.05, p=0.03 and back pain 3.3 and 3.9 respectively, mean difference -0.6, 95% CI -1.1 to -0.1, p=0.02). ODI was not statistically significantly different between the groups (21.0 vs 23.3, mean difference -2.3, 95% CI -5.8 to 1.1, p=0.18). The scores at baseline and at three and 12 month follow-up are illustrated in figure 5.

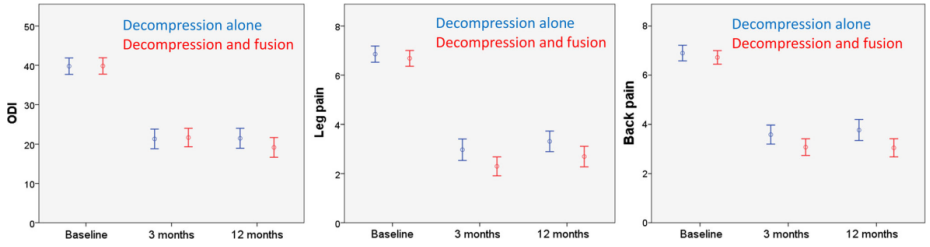


Figure 5. *Error bars for the propensity score matched cohort representing means and 95 % confidence interval for ODI, NRS back pain and NRS leg pain at baseline, 3 and 12 months*

We found no statistically significant differences in perioperative or postoperative complications between the groups. The duration of surgery (68 min. vs 103 min., mean difference 68 min., 95% CI 58 to 78, $p<0.01$) and the length of hospital stay (2.9 days vs 7.1 days, mean difference 4.2 days, 95% CI 3.5 to 4.8, $p<0.01$) were statistically significantly lower in the decompression alone group compared to the fusion group. According to the GPE scale, 4% of both groups reported their condition to be deteriorated ('much worse' or 'worse than ever') at one-year follow-up.

4.2 Paper II

Following the eligibility criteria, 3859 patients with spinal stenosis (LSS) and 617 patients with degenerative spondylolisthesis (DS) were included in the study.

For all PROMs, the accuracy of identifying 'completely recovered' and 'much better' patients was generally high, but lower for the numerical change score than for the follow-up score and the percentage change score, especially among patients with low and high PROM scores at baseline.

Estimated cut-offs for the follow-up score for a clinically important outcome were ≤ 24 for ODI, ≥ 0.69 for EQ-5D, ≤ 3 for NRS leg pain, and ≤ 4 for NRS back pain, and, for the percentage change score, ≥ 30% for ODI, ≥ 40% for NRS leg pain, and ≥ 33% for NRS back pain. These cut-offs were similar for LSS and for DS. For the numerical change score the cut-offs were ≥ 13 for ODI, ≥ 3 for NRS leg pain, ≥ 2 for NRS back pain for LSS, and > 3 for NRS back pain for DS.

4.3 Paper III

According to the eligibility criteria, 794 out of 1376 patients were included in the analyses. Of these, 476 (60%) were operated with micro-decompression alone (mean age 67.5 years, 65% female) and 318 with decompression plus instrumented fusion (mean age 63.5 years, 76% female). In this unmatched cohort, the responder rate (proportion with 30% or greater reduction in ODI at 12-month follow-up) was 71% in

the micro-decompression group and 70% in the instrumentation group (difference 1%, 95 CI -7% to 8%).

After propensity score matching, 285 patients from the micro-decompression group (mean age 64.6 years, 72% female) and 285 patients from the instrumented fusion group (mean age 64.8 years, 73% female) remained for further analyses. The follow-up rate was 423/570 (74%) at three months and 438/570 (77%) at 12 months, and 479/570 (84%) participants had at least one follow-up registration. The responder rate was 68% in the micro-decompression group and 72% in the instrumented fusion group. The lower bound of the 95% CI (-12% to 5%) for the between-group difference of -4% did not cross the -15% limit of non-inferiority. An absolute difference of 4% corresponds to a Number Needed to Treat of 25 patients (95% CI 8 to ∞).

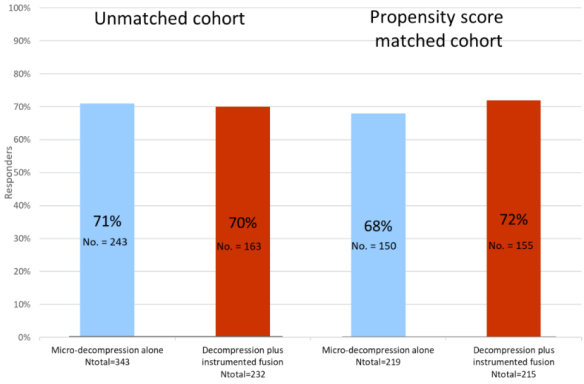


Figure 6. *The proportion with 30% or greater reduction in ODI at 12-month follow-up*

There was no statistically significant difference in mean ODI scores between micro-decompression alone and instrumented fusion at 12 months (mean [SD] 22.2 [18.2] and 20.5 [17.7], respectively, mean difference 1.7, 95% CI -2.4 to 5.8; p=0.42). The micro-decompression group had slightly but statistically significantly higher scores for NRS leg pain (mean 3.5 vs 2.7) and NRS back pain (mean 3.8 vs 3.3) than the instrumented fusion group. The micro-decompression group had fewer perioperative complications but higher patient-reported rate of superficial wound infection than the instrumentation group. Similar to paper I, the operation time and the length of hospital stay were shorter for the micro-decompression group.

4.4 Paper IV

The trial has been conducted in accordance to the the present study protocol. From February 12, 2014 to December 18, 2017, 267 patients at 16 Norwegian departments were included and randomised (table 3). Of these, 261 (98%) completed the questionnaires at baseline, 258 (97%) at 3 months, and 251 (94%) at 12 months.

The interim analysis for safety did not show between-group differences that exceeded the predefined criteria for terminating further inclusion of patients.

Table 3. Participating hospitals with numbers of patients included

Hospital	**Number of patients**
Kysthospitalet i Hagevik, Orth. dept.	78
Bærum Hospital, Ort. dept.	11
Oslo University Hospital, Ort. dept.	25
Stavanger University Hospital, Ort. dept.	42
Lillehammer Hospital	2
Arendal Hospital	7
Gjøvik Hospital	12
Skien Hospital	12
Ålesund Hospital	24
Haukeland University Hospital, Ort. dept.	9
Haukeland University Hospital, Neurosurg. dept.	12
St. Olavs Hospital	4
Akershus University Hospital, Ort. dept.	12
Kristiansand Hospital	1
University Hospital of Northern Norway, Neurosurg. dept.	11
Elverum Hospital	2
Total	**267**

5 Discussion

5.1 Discussion of main findings

5.1.1 Paper I

We aimed to investigate whether the effectiveness of decompression alone was 'as good as' decompression with additional fusion. For all primary outcomes (ODI, NRS leg pain, and NRS back pain), the upper bound of 95% CI for the difference in responder rates was above the predefined margin of 15%. Hence, statistically significant non-inferiority for decompression alone could not be claimed. However, based on the longer duration of surgery and hospital stay, the non-significant difference in the mean ODI score, and the small, perhaps clinically irrelevant, differences in pain scores at follow-up, we suggested that decompression alone is an appropriate treatment for a considerable number of patients. Commonly cited papers comparing the general concept of decompression alone with decompression and fusion are summarised in table 4.

In Herkowitz's study (1991), 50 consecutive patients with LDS were alternately assigned to either decompressive laminectomy alone or decompressive laminectomy followed by fusion without instrumentation [59]. After a mean of three years (2.4 to 4 years) they found statistically significantly better results for back pain and pain in the lower limbs in the fusion group. In a study by Bridwell et al., 44 patients were allocated to three treatment groups: Decompression alone (nine patients), decompression and fusion without instrumentation (11 patients), and decompression followed by instrumented fusion (24 patients) [128]. They found less progression of slip and more patient satisfaction in the instrumented group. These two studies are often classified as RCTs, however, the allocation process has been questioned for both studies [55]. In Herkowitz's study, the participants were alternately assigned into treatments, and in Bridwell's study the randomization procedure was not described. In both studies, the questionnaires were not self-administered and outcome measurements did not assess functional status. Further weaknesses of Bridwell's study are the low number of patients in the decompression alone group and the

limited assessment of clinical outcome. The only clinical measurement used was the patient's ability to walk. These two studies, the previously described studies from Fischground et al., [60] and Kornblum et al. [54], a systematic review [55], a meta-analysis [129], and guidelines [55, 64, 65, 67, 130] all suggest that an additional fusion should be the first choice of treatment for DS. Instrumented fusion was recommended to improve the radiological fusion rates. Although none of the papers showed a superior ability of instrumented fusion to improve clinical outcomes, the number of surgeries with instrumented fusion increased during the 1990s and the 2000s [68]. Some researchers warned against this practice, and ties between equipment suppliers and surgeons have been questioned [131, 132]. Both Deyo and Carragee suggested that surgeons' increasing use of complex fusion procedures might be motivated by generous financial reimbursement from the surgical implant manufacturers [131, 132].

The next two RCTs in the field were simultaneously published in the *New England Journal of Medicine* in April 2016. Försth et al. found no difference in clinical outcome measures between decompression alone and decompression with additional fusion, whereas Ghogawala et al. found better functional status for the fusion group [23, 27]. The contradictory conclusions of the trials [23, 27] sparked debate regarding the evidence for the best choice of treatment for DS [133-137]. In the study by Försth et al., the surgical methods were decided by the surgeons. Of the fusion group, 90% were operated with pedicle screw instrumentation, and in 18% of the decompressions the midline structures were preserved. The patient characteristics and the baseline scores for ODI, leg pain, and back pain were comparable to the baseline scores revealed in a registry study from Swespine (2013; n= 1306) [11]. In the registry study, the two-year follow-up scores did not differ between decompression alone and decompression with fusion. The ability to use knowledge from both a randomised study and a registry study was a major strength of the research from Försth's study group. Thus, the similar results provided evidence with acceptable internal as well as external validity. The external validity has been further confirmed in our study. The study cohort in paper I is in accordance with the cohorts defined in the Swedish studies. In contrast to our study, the Swedish studies did not demonstrate different

levels of pain relief between the groups. In the study by Ghogawala et al., 35 patients underwent a decompression alone without preserving the midline structures (i.e., laminectomy), and 31 patients had a laminectomy plus an instrumented fusion. The decompression alone group had statistically significantly less improvement in the physical-component summary score of the 36-Item Short-Form Health Survey (SF-36) than the instrumented group. Further, a borderline statistically significant difference in ODI improvement was revealed (26.3 reduction in the instrumented fusion group vs 17.9 points in the decompression alone group; p=0.06). The results were in accordance with a previous cohort study (n=34) of similar treatment groups published by Ghogawala et al. [138]. Compared to the RCTs by Ghogawala and Försth, our results from paper I corresponded best with the latter. One reason may be that Ghogawala et al. used narrow inclusion criteria and several clinical as well as radiological exclusion criteria, unlike the more pragmatic design of the Swedish study [139].

Another Swedish registry study found more improvement in ODI, VAS for leg pain, VAS for back pain, and EQ-5D among fused patients (n=594) compared to only decompressed patients, (n=245) when back pain dominated over leg pain preoperatively. In patients with predominate leg pain, the fusion group had greater back pain relief compared to the decompression alone group. These results were based on data from one-year follow-up. At two-year follow-up, no statistically significant differences between the treatment groups were found [71].

Two recent meta-analyses and one review included the two RCTs (Försth and Ghogawala) as well as paper I in the analyses. Liang et al. [140] reported more satisfaction and leg pain relief for decompression with fusion. They detected no differences in ODI, back pain, complication rate, or reoperation rate. Chen et al. [141] did not find any differences in outcome scores between the groups. Both studies found less blood loss and shorter operation time and length of hospital stay in the decompression alone group. In a review of current concepts, Samuel et al. [142] found non-different outcomes in ODI, but greater pain improvement in the fusion group was upheld. The authors summarised that “posterolateral spinal fusion remains

the treatment of choice", and that instrumentation with intervertebral cages as well as decompression alone may be an appropriate option in subgroups.

Table 4. Presentation of RCTs and observational studies, with description of samples, intervention types, and conclusions

Years	Study	Nationality	Inter-vensions (N)	Conclusion
	RCT			
1993	Herkowitz	US (one site)	DA/DFni (25/25)	DA < DFni (NRS leg/back pain, 0-5)
2016	Ghogawala	US (five sites)	DA/DFi (35/31)	DA < DFi (SF-36, physical health)
2016	Försth	Sweden (multicentre)	DA/DF (67/66)	DA = DF (ODI, ZCQ, VAS leg, VAS back)
	Observational cohort studies			
2004	Ghogawala	United States (one site)	DA/DFi (20/24)	DA < DFi (SF-36, physical health)
2011	Kleinstueck	Switzerland (one site)	DA/DF (56/157)	DA < DF (COMI, VAS leg, VAS back)
2013	Försth	Sweden (registry)	DA/DF (651/480)	DA = DF (ODI, VAS leg, VAS back)
2015	Sigmundsson	Sweden (registry)	DA/DF 245/594	DF > DA in patients with predominant back pain (ODI.VAS leg, VAS back, EQ-5D)
2017	Austevoll	Norway (registry)	DA/DF (260/260)	DA not non-inferior DF (ODI, NRS leg, NRS back)
2016	Alvin	US (one site)	DA/DFi (25/75)	DA > DFi (cost-effectiveness)
	Reviews/ Meta-analysis/ Guidelines			
1994	Mardjetko	United States	DA/DF	Fusion significantly improves patient satisfaction and instrumentation increases fusion rates.
2005	Resnick	United States	DA/DF/DFi	DFi is recommended.
2007	Martin	United States	DA/DF	DF may lead to better clinical outcome than DA. Moderate evidence that instrumentation fusion rates.
2009	Watters	United States	DA/DF/DFi	DF > DA. Instrumentation improves fusion rates but is not superior to non-instrumented fusion in regard to clinical outcome.
2014	Resnic	United States	DA/DF/DFi	DFi is recommended.

Years	Study	Nationality	Inter-vensions (N)	Conclusion
	Reviews/ Meta-analysis/ Guidelines			
2014	Steiger	Switzer-land	DA vs DF	"Insufficient evidence to draw conclusions concerning clear indications for specific types of surgical treatment. There remains a need to establish a decision-making tool to assure appropriate treatment for patients with LDS."
2015	Joaquim	United States	DA vs DF	"Satisfactory clinical outcome can be achieved with DA in selected patients."
2017	Chang	China	DA vs DF	DF did not show better clinical results for DF than DA. Longer duration of operation, more blood loss, and a higher risk of complications for DF.
2017	Liang	China	DA vs DF	DF had more improvement of clinical satisfaction and leg pain, but more blood loss, operation time and hospital stay. No differences in ODI, back pain scores, complication rate, and reoperation rate.
2018	Chen	China	DA vs DF	DF did not yield better clinical outcomes than DA. Longer duration of operation, more blood loss, and a higher risk of complications for DF

Abbreviations: DA, Decompression alone; DF, Decompression and fusion; DFi, Decompression and instrumented fusion; DFni, Decompression and non-instrumented fusion.

5.1.2 Paper II

PROMs are key instruments for assessment of clinical outcomes following spinal surgery. In this paper, we evaluated how accurate three alternative PROM instruments could distinguish patients with a clinically important outcome (i.e. responders) from those without (i.e. non-responders). For this purpose, the study showed that the follow-up score and the percentage change score were more accurate than the numerical change score, especially among patients with low and high baseline scores. We recommend not using the numerical change score for determining clinical importance following surgeries in DS and LSS.

The use of numerical change scores has been criticized for not considering the relative relationship between baseline and follow-up [76, 77, 143]. Mathematically,

any given amount out of a large amount is a smaller proportion than the given amount out of a small amount Thus, a numerical change from high baseline values would constitute a smaller improvement than a corresponding change from a lower baseline. For example, a numerical change from 8 to 6 in NRS leg pain will most likely represent less improvement than a reduction from 4 to 2.

The follow-up score and the percentage change score had similar accuracy for discriminating between those perceived as 'cured' and 'not cured'. The estimated cut-offs for the follow-up score (ODI ≤ 24 points, leg pain ≤ 3 points, and back pain ≤ 4 points) are clearly in accordance with estimated cut-offs for what constitutes a 'satisfactory symptom state' [144] and an 'acceptable pain level'[76], as derived from the EUROSPINE Spine Tango Registry. An advantage of using the follow-up score to define 'success', is that only post-operative measurements are needed. Hence, being considered a responder only depends on whether a patient reaches a score lower than a defined threshold or not. Such an approach is particularly suitable in retrospective studies without recorded baseline data, for example when evaluating outcomes following emergency trauma surgeries. For studies that collect baseline scores as well as follow-up scores, the use of the percentage change score may be more in line with the aim of the treatment – to reduce the level of the patient's pain and disability from before the operation (baseline) to follow-up. Our suggested cut-off for the percentage change score was in accordance with thresholds suggested by Ostelo et al. in a study based on a literature review and expert panel discussions [145].

While the construct MCID is interpreted as a minimal clinically important difference, the cut-offs from the present study conceptually represent thresholds for a substantial improvement ('completely recovered' or 'much better') following surgery. Since surgery for spinal stenosis has a considerable risk for complications and undesirable outcomes, a "minimal clinically important difference" does not reflect the target of an operation. The construct "substantial clinical improvement" better represents the goal of the surgery [75].

It is crucial to recognize that the present criteria cannot be applied to evaluate whether mean differences in treatment effects are clinically significant or not. Such use is described by Katz et al. [73] as a pitfall in the interpretation of clinical trial data: “In clinical trials, the MCID yardstick should be applied to changes in individual subjects, not to group changes; applying individual clinically important changes on a group level is misleading”. Several other authors dealing with this subject have also strongly warned against using benchmarks for clinically important treatment benefits to directly compare treatment effects between groups [73]. Instead, the criteria should be utilized to compare the proportion of patients (responder rate) achieving an outcome of clinical relevance and importance [73, 75, 77, 145, 146].

5.1.3 Paper III

We aimed to investigate whether the effectiveness of micro-decompression alone was ‘as good as’ decompression with fusion. Statistically significant non-inferiority for micro-decompression alone was revealed both in the unmatched cohort and in the propensity score matched cohort. For the matched cohort, the micro-decompression group had a 4 percentage points (95% CI -12 to 5) lower responder rate than the instrumented group, corresponding to a NNT of 25 (95% CI 8 to ∞). If the “extreme low value” (i.e., the lower bound of the 95% CI) of NNT represents the true value of the population, eight patients needed decompression with instrumented fusion instead of micro-decompression alone to achieve one extra responder. Advantages of micro-decompression were the shorter duration of surgery and hospital stay, and the lower rate of preoperative complications. The responses on the GPE scale indicate that the patients’ overall outcome ratings were comparable between treatment groups. As for study I, the fusion group had somewhat less leg and back pain relief at one-year follow-up.

There is no definite answer as to why non-inferiority could be claimed for micro-decompression alone in this paper but not for decompression alone in paper I. The different criterion for being a responder (12.8 points improvement in study I and 30% improvement in study III) might be an explanation. Another possibility is that the technique for decompression alone impacted the outcome. In a recent meta-analysis,

a minimally invasive (midline-preserving) decompression (n=485) showed greater patient satisfaction and lower reoperation rates than laminectomy (n=671) in treatment of DS [147]. Another recent review reported that mini-invasive techniques are increasingly utilised with promising results regarding slip progression and subsequent fusion surgery [142]. Nevertheless, it is important to recognize that the results from our studies do not provide evidence that micro-decompression alone is a better method than laminectomy alone. Those two techniques have not been compared in the present studies.

In recent years, an increasing number of papers involving less-invasive decompression alone techniques have been published. A list of frequently cited papers is given in table 5. In a randomised trial, Inose et al. compared midline-preserving decompression (n=29), decompression with instrumented fusion (n=31), and decompression with dynamic fusion (n=25) [148]. They reported no statistical difference in patient-reported outcome between the groups at one- and five-year follow-up. The results were graphically presented, so no numerical information regarding PROM scores was given in the paper. Regardless of potentially non-significant differences in outcomes, the small sample size might introduce a serious risk of type II error, meaning an erroneous conclusion of no between-group differences.

Several observational studies have been published. Madsudaira et al. compared clinical and radiological outcomes at two-year follow-up between 19 patients operated with pedicle screw instrumentation, 18 with only a midline-preserving decompression (named ‘laminoplasty’), and 16 conservatively treated patients. Only patients with DS grade I (a slip of less than 25% of the length of the adjacent lying vertebra) [149] were included. The surgically treated groups showed statistically significantly more alleviation of symptoms than the conservatively treated group. Although instrumented fusion prevented postoperative progression of spondylolisthesis, the clinical outcome of this group was not improved compared to those with midline-preserving decompression. The authors suggested that

decompression with preserving posterior midline structures can be useful for treating patients with DS grade I.

In 2002, Park et al., retrospectively evaluated patients operated for DS. Patients without clinically relevant back pain operated with micro-decompression (n=20; median follow-up, 63 months) were compared to 25 patients suffering from both leg and back pain treated by decompression and instrumented fusion. No differences were found in reduction of leg pain or in functional improvement [150] between the treatment groups.

Rampersaud et al. demonstrated that those with 'stable' grade 1 spondylolisthesis (n=46; median follow-up, 63 months) operated with midline-preserving decompression had a reduction in pain and functional improvement which was similar to that of 133 patients with more complex structural radiological pathology operated with instrumented fusion [151].

In a study of 140 patients the effectiveness of decompression plus instrumented fusion was evaluated in relation to altered indications for performing micro-decompression alone (n=60; mean follow up, 78 months). The authors suggested that midline-preserving decompression alone is a suitable choice of treatment, even for patients with severe back pain and preoperative spondylolisthesis of more than 5 mm [152].

Chang et al. (n=59 at baseline) found no difference in clinical outcomes or progression of olisthesis between patients operated with less-invasive decompression alone and decompression with instrumented fusion, neither at 12-month (n=56) nor at 60-month (n=23) follow-up [153].

Finally, one study has demonstrated favourable cost-effectiveness for a selected patient group (i.e., leg dominant symptoms, grade 1 'stable' spondylolisthesis, and 'favourable facet joints') allocated to midline-preserving decompression compared to a more 'complex' group of instrumented fused patients [154].

The conclusions from the above mentioned studies were in accordance with the conclusion in paper III; clinical performance seems to be comparable between micro-

decompression alone and decompression with fusion. Nevertheless, we consider our study to have strengths that these studies lack. In the other studies, the decompression group was generally older, had a larger proportion of women, had fewer one-level operations, and had lower mean back pain at baseline compared to the instrumented group. In our study, due to the propensity score matching, the mean baseline characteristics were similar between the groups. In addition, due to the multicentre design and the considerably larger sample size, our study provides a higher external validity.

Table 5. Presentation of RCT and observational studies, with description of samples, intervention types, and conclusions

Years	**Study**		**Nationality (Sample)**	**Inter-vensions (no)**	**Conclusion**
	RCT				
2018	Inose		Japan (one site)	DAm/DFi (23/28)	DA = DF (JOA, VAS leg, VAS back)
2020?	Austevoll		Norway multicentre)	DAm/DFi (133/133)	?
	Observational cohort studies				
2005	Matsudaira		Japan	DAm/DFi (19/18)	DAm = DFi (JOA)
2012	Kim		Canada	DAm/DFi (57/58)	DAm > DFi (cost-effectiveness)
2014	Rampersaud		Canada	DAm/DFi (46/133)	DAm = DFi (SF-36 Physical component summary)
2012	Park		Korea	DAm/DFi (20/25)	DAm = DFi (ODI, NRS leg/back pain, SF-36)
2015	Inui		Japan	DAm/DFi (60/80)	DAm = DFi (JOA, back pain)
2020	Austevoll		Norway (Norspine registry)	DAm/DFi (285/285)	DAm non-inferior DFi (ODI, NRS leg, NRS back)

Abbreviations: DAm, Micro-decompression alone; DFi, Decompression and instrumented fusion

5.2 Methodical considerations

Several critical considerations were of importance in planning the studies: 1) The *current evidence* on surgical treatment of DS should advocate new investigations; 2) The *objective* should be relevant, i.e., the findings should be of importance for future treatment of patients or for future investigation of the population; 3) The investigated

patients (the study cohort/sample) should represent the population (i.e., all patients the study intends to gain knowledge on) as accurately as possible. The extent to which a study can be extrapolated to a population as a whole is denoted by the grade of external validity. Population validity describes how well the analysed sample represents the target population, and is central when evaluating the external validity of results from a study; 4) The *data* should be of high quality. This requires high completeness of collected data, and that data are accurately transferred into the database used for analyses; 5) The utilized *outcome measurements* should be reliable, valid, and responsive; 6) The *sample size* should be large enough to answer the relevant questions with statistically robust and clinically relevant conclusions; 7) The planned *design* should be in accordance with the objective of the study; 8) The *statistical methods* should be able to account for systematic errors that could otherwise lead to fallacious (biased) conclusions.

5.2.1 Patients

Papers I-III

Patients selected in papers I-III were found in the Norwegian Registry for Spine Surgery (NORspine) based on a form filled out by the surgeon. Patients were included if the surgeon ticked off both the 'spinal stenosis' box and the 'degenerative spondylolisthesis' box. The registry does not provide further information regarding diagnoses. The registration is based on the surgeons' evaluation of MRI and/or standard X-rays or a radiological report. We have not retrospectively controlled the radiological diagnoses against the radiological reports, nor have the radiological examinations been evaluated retrospectively by the investigators. Regarding misclassified diagnosis, it is more likely that a 'minor' DS would be classified as LSS than the other way around. If the surgeon has ticked off the DS box, it is likely that a radiological spondylolisthesis is present. Regarding the reported surgical treatment methods, a study from 2010 revealed a 97% agreement between the registry and hospital records [51].

Since the introduction of the registry in 2007, its coverage rate has steadily increased. In order to evaluate the coverage rate, the annual number of registered NORspine

patients has been compared to the number of annual spine surgeries registered by the National Registry and Statistics Norway (Norsk pasientregister (NPR), organised by the Norwegian Directorate of Health). The NPR contains information on admissions to Norwegian hospitals, but, unfortunately, this registry neither does cover 100% of all operated patients. For calculating the coverage rate the following formula has been used: $(N_{NOR\ spine\ alone} + N_{NORspine\ and\ NKR}) / (N_{NOR\ spine\ alone} + N_{NKR\ alone} + N_{NORspine\ and\ NKR})$. For studies I-II the coverage rate was below 50% at the beginning of inclusion, but did subsequently increase, and was calculated to be at least 60% in 2014 [155]. Some patients do not wish to participate, and the motivation for completing registration varies between hospitals and surgeons. However, in studies from NORspine, a wide spectrum of hospitals, surgeons, and patients participated, providing results and conclusions with a high population validity.

Paper IV

Patients included in paper IV were recruited through the NORDSTEN study collaboration. The DS trial is one of two randomized NORDSTEN trials. The Spinal Stenosis Trial compares clinical outcomes of three different surgical methods [37]. In addition, a prospective observational study of non-operated patients with LSS or DS is being conducted (ClinicalTrials.gov Identifier: NCT03562936).

The short period of enrolment compared to previous RCTs [23, 27, 148] indicates that a relatively large proportion of eligible patients were enrolled into the NORDSTEN-DS trial. This contributes to high external validity of the study.

The eligibility criteria were thoroughly discussed by the NORDSTEN steering group before patients were included. The inclusion criteria are generally in concordance with similar RCTs [12, 23, 27]. Although we intended to make a pragmatic study design, some exclusion criteria were needed to ensure reliable assessments on a disease-specific condition and provide patients that would adhere to the follow-ups. We decided to exclude those with radiological findings of scoliosis > 20 degrees, a slip ≥ 3 mm in more than one level, or a foraminal deformation in vertical direction of a nerve root (i.e., grade 3 foraminal stenosis according to Lee) [90]. For these patients, additional arguments may exist for performing decompression accompanied

by instrumented fusion. Loosening of screws is a well-known complication for patients with osteoporosis, and since former thoracolumbar fractures are associated with osteoporosis, they were excluded. A non-specific eligibility criterion states that patients "not able to fully comply with the protocol", including treatment and follow-up, should be excluded. Hence, the participating surgeons were recommended not to include patients that not were able to adhere to the protocol.

The steering committee decided that from August 2015, patients should not be excluded due to ODS scores below 25. Compared to the original design, this enhances the external validity but can lead to lower mean ODI score at inclusion and, consequently, smaller effect sizes and potentially smaller differences between groups.

5.2.2 Data

Utilizing data of high quality is crucial to provide conclusions with high validity. The quality of data depends on the completeness of the variables one intends to analyse, and how accurately the data represent the 'true' value of a participant. Regarding collected data in the NORspine database, the completeness of baseline characteristics as well as PROMs has been evaluated [51]. The patient age at baseline was reported in 99.2% of the patients, the gender in 100%, the BMI in 97.0%, and the PROM scores (ODI, NRS leg pain, NRS back pain and EQ-5D) in 94.3 to 99.5%. Further, the error rate of punching patients' baseline forms was calculated to be 0.3% and errors of scanning follow-up questionnaires to be 0.04% [51].

The quality of data is not yet evaluated for the NORDSTEN-DS trial (paper IV). Since this study is closely monitored we anticipate a high completeness of the collected data.

5.2.3 Outcome variables

The main objective in surgical treatment of spinal degenerative disorders is to reduce pain and improve function. In the present studies, we have utilized PROM questionnaires recommended for assessment of spinal stenosis and degenerative spondylolisthesis [56]. A major strength of using PROMS is that the completion of the questionnaires does not influence people involved in the delivery of the treatment.

The measurements of ODI and the NRS for leg and back pain are included in all papers. Although the original version of the ODI was intended to assess disability related to back pain, the questionnaire has also been a key instrument in modern landmark studies of patients suffering from pain in the lower limbs (i.e., lumbar spinal stenosis and lumbar disc herniation) [12, 21, 23, 25, 27, 156]. In the translated version 2.0, the patients are asked how their pain influences their daily lives, without specifying whether the pain originates in the back or from the lower limbs.

For assessment of pain the Visual Analogue Scale (VAS, which range from 0 to 100) is a frequently used alternative [27, 157] with acceptable reliability as well as validity [101]. Nevertheless, compared to the VAS, the NRS is suggested to be easier to understand and has shown higher test-retest stability (higher reliability) [102].

The ZCQ is specifically constructed to assess the degree of symptoms and functional disability in patients suffering from spinal stenosis [98]. In addition, it assesses how satisfied patients are with their treatment. The questionnaire is not included in the NORSpine registry forms. Defining a single primary outcome in the RCT, we did a literature review and thoroughly discussed whether to choose the ZCQ or the ODI. We could not conclude better psychometric properties of ZCQ compared with ODI [158]. Since ODI is more commonly used in related studies [11, 23, 27], and as it would be of importance to directly compare the results between the RCT and study III, we defined the ODI to be the primary outcome and the ZCQ to be a secondary outcome in study IV.

For measurement of health related quality of life (HRQoL), the EQ-5D questionnaire [103] is a commonly used tool to compare cost-utility between treatments [105], and will be utilized in a planned cost-efficacy study with data from the RCT. Although acceptable validity, reliability, and responsiveness in the assessment of lumbar degenerative disorders [104], it does not assess the symptoms typical of spinal stenosis, and was therefore not utilised in studies I and III.

The mean change in PROMs, from baseline to follow-up, has been routinely used in similar studies [12, 25, 138, 156]. However, in order to provide meaningful clinical information, we have focused more on the proportion of patients with a clinically

important improvement (responder rate) [73]. This approach is recommended for comparing treatment effects between groups [73, 75, 77]. In paper I we used an MCID derived from a mixed population of spinal disorders by Copay et al. [108]. Strength of studies III and IV is the use of the condition-specific criteria for a clinically important improvement, derived from a representative study cohort (paper II).

5.2.4 Study design

Studies I, II, and III are longitudinal observational cohort studies, with prospectively collected data. Although retrospectively analysed, the aim of the studies, the eligibility criteria, the primary and secondary outcomes, and the statistical methods were defined and recorded in study protocols before data were available to the investigators.

Paper IV describes a longitudinal randomised controlled multicentre trial.

Papers I, III, and IV are comparative studies, whereas paper II is a methodological study which evaluates the ability of different PROM instruments to detect a clinically relevant outcome.

RCT versus observational study – efficacy versus effectiveness

To provide the highest level of evidence regarding treatment effects, a randomised clinical trial is required. The study design, its structure, and the way the study is conducted should be clearly associated with the purpose of the study. If conducted under ideal conditions, tightly monitored, and with strict criteria for inclusion and exclusion, the trial is explanatory. The principal objective is to investigate whether a treatment works compared to placebo or an established or well-documented treatment, i.e., the efficacy of a treatment. Eliminating all other elements means that a difference in efficacy is solely contributed to by the treatments. The internal validity is high, meaning that a causal conclusion linked to treatment is warranted. However, limitations exist for explanatory RCTs. Carefully selected participants with well-defined characteristics create a homogenous study cohort, which is likely to be different from the population of interest. Furthermore, firmly monitored follow-up

routines and extensive use of outcome measurements usually diverge from usual clinical practice. Hence, the implications of such evidence could be limited by a low generalizability from the study cohort to the target population [159]. Studies of effectiveness can be assessed in so-called pragmatic studies. In such studies, the eligibility criteria and study setting should mimic ordinary clinical practice as closely as possible. Patients should be recruited from practices of multiple surgeons and institutions, and outcomes of interest should be relevant to patients and clinicians; this leads to a high generalizability of results to the general population [82, 139]. In sum, an explanatory trial measures efficacy and answers the question: Does the treatment works under ideal conditions? A pragmatic study intends to assess the effectiveness of treatments, i.e., how they work under usual conditions.

Clinical studies are hardly ever purely explanatory or pragmatic. In an explanatory study there may be aspects of the participants or the interventions that are beyond the investigator's control. Similarly, a comprehensive collection of data from standard questionnaire forms exemplifies how a pragmatic study setting can diverge from the normal clinical setting. Thorpe et al. introduced "A pragmatic-explanatory continuum indicator summary (PRECIS)" as a tool to help investigators define a trial in a multidimensional continuum [83]. When planning a study, the PRECIS tool can help investigators to define the study population, the eligibly criteria, the requirements for expertise among the practitioners, the follow-up routines and outcomes, and the study design. It can also help the readers (e.g., physicians and health care providers) to understand the strengths and limitations in scientific papers, and to deduce the grade of internal and external validity of the results [83].

Although the NORDSTEN trial is a closely monitored efficacy study using selected inclusion and exclusion criteria, according to the PRECIS tool this trial has several pragmatic aspects. It is a multicentre study including small and large departments. Some flexibility in performing surgery is accepted, and no specific qualifications are required of the surgeons except that they perform spine surgery regularly. Furthermore, the outcome is assessed by measurements considered to be clinically important for the patients and for physicians in real life. However, several exclusion

criteria, restrictions in the surgical interventions, more extensive collection of follow-up data, and the follow-up time diverge to some extent from usual care. For example, the results cannot be directly applied to patients over 80 years of age, patients with ASA grade > 3, those with hip and knee arthrosis, those with olisthesis in more than one level, patients formerly operated at index level, and those with a high grade foraminal stenosis. Neither can the findings be directly applied to decompression with standard laminectomy and decompression plus non-instrumented fusion. Hence, the external validity of the results from the NORDSTEN-DS trial might be somewhat limited. Papers I and III contribute real-world knowledge about the treatments with higher external validity. Assessing both the efficacy and the effectiveness of a treatment is recommended to ensure knowledge with high internal as well as external validity [84, 85, 159]. The different study designs of an RCT and an observational study will bring complementary information to the interpretation of how the treatments work [159]. To highlight questions that neither an RCT nor an observational study would have the ability to solve separately, both kinds of studies are recommended for meta-analyses searching for the best treatment option [159].

5.2.5 Statistical methods

Paper II

Since paper II is conceptually unlike papers I, III, and IV, it will be discussed separately.

For the estimate of cut-off values for ‘success’ assessed by different PROMs, we have used an anchor-based method. The anchor-based method is advocated by Dworkin et al. (the ‘IMMPACT Recommendation’) [77]. In this method, an external criterion (‘anchor’) for a clinically important outcome (‘success’) was compared with the PROM scores. The responses ‘completely recovered’ or ‘much improved’ on the GPE scale (1-year follow-up) were considered to indicate an outcome that clearly reflects a substantial change for the patients [77] and were utilized as the gold standard for ‘success’. This is analogue to how the threshold for ‘sickness’ is estimated for a diagnostic test on a continuous variable. An example is the measurement of the serum prostate-specific antigen (s-PSA) for detecting prostate

adenocarcinoma; to dichotomize patients into ‘potentially sick’ and ‘healthy’ cases, the threshold of PSA can be estimated with a histological exam as anchor [160]. The chosen threshold for PSA is one that optimises the relationship between sensitivity (i.e., the probability of the test being positive) and specificity (i.e., the probability of the test being negative). Similarly, we have estimated cut-offs for PROMs that most accurately distinguish ‘cured’ patients (the responses ‘much better’ or ‘completely recovered’ at the GPE scale) from those who are not ‘cured’ following surgery.

Another method for determining a clinically important change is the distribution-based method. This is a statistical method using the variability in a measurement within the observed cohort to determine whether a given change in the variable is of clinical importance [73, 77]. The most common distribution-based method uses the standard error of measurements (SEM = SD$*\sqrt{1-r}$) [SD, standard deviation of the baseline scores; r, test-retest reliability]) to define the lowest value in a change score that is larger than the measurement error. Because the SEM value is an estimate of the precision of a change in a measurement, the threshold for a clinically important improvement should be above the SEM value [73]. We have not used a distribution-based method; however, our estimated cut-offs for the change scores are above (i.e., stricter than) estimated cut-offs for previously distribution-derived MCIDs [161, 162]. This indicates that the criteria for the change scores in the present study most probably surpass thresholds for measurement errors.

To evaluate PROMs against GPE scores is recommended [77], and this is the most frequently used approach for determining thresholds for clinical importance [75, 76, 108, 143, 163, 164]. However, evaluating PROMs against another self-evaluating instrument as ‘the gold standard’ has been criticized [79]. Alternative anchors for ‘success’ would be walking distance [165], return to work [79], or simply to preoperatively assess patients’ expectations of improvement. The latter recommended approach is described as the ‘benefit-harm trade-off method’ [166, 167].

Papers I, III, and IV

The choice of non-inferiority design

Test for superiority

To test whether treatment A is better or worse than treatment B (placebo or a control treatment), a superiority trial is designed. The null hypothesis asserts that the two interventions are similar (A = B). If A ≠ B, the null hypothesis will be rejected, and the alternative hypothesis is supported (A superior to B, or B superior to A). In a power calculation one needs to define the magnitude of between-group difference that is of clinical importance (γ), the Type I error rate (α; risk of falsely rejecting the null hypothesis), and the Type II error rate (β; risk of falsely accepting the null hypothesis). In hypothesis testing the null hypothesis will be rejected if the (100- α) confidence interval does not include zero (i.e., the whole CI should be positive or negative, dependent on the direction of the calculation) and/or the p-value should be less than 0.05. This is called two-sided testing for superiority. In a superiority study, automatically claiming evidence for equivalence of treatment if the null hypothesis is not rejected would be a misinterpretation of results. Erroneously accepting the null hypothesis (type II error) might be explained by a lack of power to detect a statistically, and perhaps clinically, relevant difference [168].

Test for non-inferiority

For the comparative studies we have used a non-inferiority design. Non-inferiority trials intend to show that one treatment (normally a new treatment) is not inferior to a control treatment (normally the standard treatment). In other words, it tests whether one intervention is ‘as good as’ or ‘not unacceptably worse than’ another. The null-hypothesis is turned around in that H0 tests whether treatment A is worse than treatment B by more than –δ, where δ is the predefined margin of non-inferiority. H1 states that the between-treatment difference is less than –δ. By testing H0 a 95% (or a 90%) CI for the between-treatment group is estimated. If the lower bound of the CI is above –δ, H0 is rejected and non-inferiority for H1 is shown. Hence, the new treatment could be said to be ‘as good as’ treatment B. Figure 7 shows the possible

conclusions from comparing the treatment effect of decompression alone (DA) and decompression with fusion (DF).

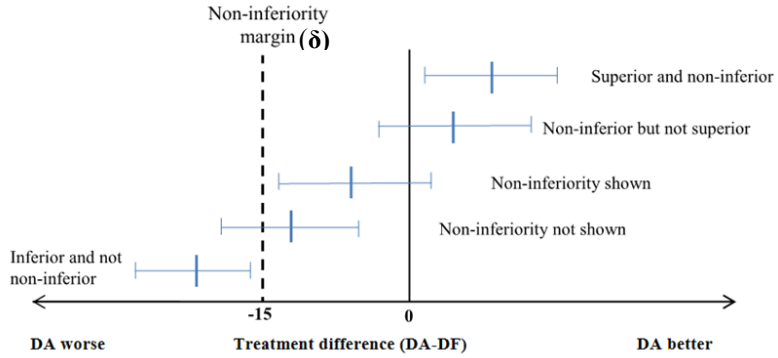


Figure 7. The X-axis shows the treatment difference between decompression alone (DA) and decompression with fusion (DF). The vertical lines indicate zero (no between-group difference) and –δ (the predefined non-inferiority margin). For five possible outcomes, the mean difference with corresponding confidence interval is indicated with horizontal lines. For the two lowermost lines the lower bound of the CI are below –δ and non-inferiority for DA could not be claimed. For the next two lines the lower bound of the CI are above –δ, and non-inferiority is shown. The upper line demonstrates a scenario where both on-inferiority as well as superiority for DA could be claimed.

In the present studies, there were several reasons for choosing a non-inferiority design. Based on previous studies our expectation was that decompression alone would not be better than decompression with fusion. We wanted to focus on decompression alone and intended to test whether the clinical outcomes for DA were not unacceptably worse than for DF. To prove a hypothesis, an alternative hypothesis has to be rejected. To prove that DA is not unacceptably worse than DF, a non-inferiority design had to be chosen. If we had chosen a standard superiority trial and failed to prove superiority of one treatment over the other, we could not conclude that the treatments were equal or that one treatment was non-inferior; “no evidence of difference is not evidence of no difference”.

An important prerequisite for a non-inferiority trial is that the efficacy of the standard treatment, i.e., decompression plus instrumented fusion [142], which the new

treatment is tested against, should be scientifically proven. For DS, a modest grade of evidence exists for surgical treatment being better than non-surgical treatment. (SPORT) [12, 42]. In these studies, the majority of patients (94%) were operated with decompression with an additional fusion. Another criterion for conducting a non-inferiority study is that the alternative treatment has some obvious advantages compared to the standard treatment. Decompression alone is associated with lower perioperative complications and lower hospital costs than decompression plus fusion [58, 68]. Decompression alone could therefore be advocated if the clinical outcomes following surgery were found to be not unacceptably worse than for decompression with fusion. Unfortunately, a clinically reliable margin for 'unacceptably worse' does not exist. Neither is it possible to estimate the margin directly based on statistical assumptions. After thorough discussions in the study group based on former literature, we defined the margin of non-inferiority to be a 15 percentage points difference in responder rate, corresponding to a number needed to treat of seven patients (NNT = 100/15 = 6.67) [112]. In other words, if seven patients or more needed fusion to achieve one extra responder, we considered the advantages of decompression (e.g., less invasive and cheaper) to surpass the disadvantages of instrumented fusion (e.g., higher complication rate and costs). This margin is in accordance with a prospective, randomized, multicentre Food and Drug Administration investigational trial comparing lumbar total disc replacement and lumbar fusion in patients with degenerative disc disease [111].

5.2.6 Sample sizes

Study I is among the three largest observational studies comparing decompression alone and decompression with fusion **[11, 71]**. Study III is the largest study comparing micro-decompression alone with decompression and fusion. The NORDSTEN-DS trial has a considerably larger sample then present published RCTs **[23, 27, 148]**. For the three comparative studies we consider the sample sizes large enough to answer the questions of the investigation.

5.2.7 Risks of bias

Bias can be defined as any systematic error that results in incorrect estimates of the association between treatments and outcomes, or incorrect interpretation of revealed estimates. Several kinds of biases might threaten the validity of the present studies.

Information bias

Information bias can occur due to inaccuracy in data management and due to low reliability, low validity, and poor responsiveness in the measurements used in the assessment of outcomes [169]. Misclassification of diagnosis, inaccurate reporting of surgical method, and errors in punching data from patient forms introduce some risk of bias in the present studies, especially in the registry studies. In the thoroughly conducted and monitored RCT these risks are lower.

Although the utilised PROMs are well accepted for DS [56], and the properties are validated for use in the Norwegian language, some bias due to inaccurate measurement of treatment effect may exist. Since a common goal of surgery is to improve walking distance and physical activity in general, an objective assessment of ambulation would be beneficial [165, 170].

Two more common biases we have to consider in our studies are *selection bias* and *allocation bias*.

Selection bias

Ideally, all participants eligible for a study should be included. Unfortunately, some patients do not wish to participate, and physicians often treat patients without keeping ongoing studies in mind or do not wish to fill in required forms for a study or a registry. If patients recruited in a study systematically differ from those not recruited, selection bias can occur [171].

In the registry studies, we have no information about patients not registered in the NORSpine database. Although the registries have a relative high coverage rate, the extent to which the sample is representative of the population is not guaranteed. If selection bias is present, the estimates of the treatment effectiveness for the studied cohorts might be over- or underestimated. We did not have the opportunity to explore

whether a selection bias was present in studies I and III; inaccurate estimates of relative effectiveness might therefore exist due to selection bias.

Selection bias might also exist in the multicentre RCT. Due to a low number of included patients at some hospitals, it is likely that a considerable number of patients were not screened for eligibility. Hence the representation of included patients might not be in accordance with the defined study population [171]. Further, the treatment effect for the cohort as a whole might be under or overestimated.

Attrition bias - missing data

Missing data has been defined as 'values that are not available and that would be meaningful for analysis if they were observed' [172]. Attrition refers to the loss of participants during the period from inclusion to follow-up and is a common cause of missing data [173]. Another cause would be an incomplete questionnaire. If relevant data is missing, there is no analytic approach that can provide estimates without a risk of bias [172].

Three principle mechanisms cause missing data [174]: 1) Missing completely at random (MCAR): The cause of missing data is unrelated to any other observed variable as well as to the missing variable itself. The missing values have randomly disappeared from the data set, for example by an accidental random deletion of part of a data matrix. In a situation of MCAR, distribution of observed parameters will be similar between the 'missing group' and the 'complete case' group. MCAR will reduce power, but will not be a potential source of bias; 2) Missing at random (MAR): Missing data should be related solely to other observed data and not to the value of the missing data itself. For example, only the oldest patients have missing data for EQ-5D. If the values of the missing EQ-5D data were known, MAR would be confirmed if missingness was related to age but unrelated to the missing values of EQ-5D. Unfortunately, since the missing values of EQ-5D are unknown, MAR cannot be confirmed; 3) Missing not at random (MNAR): The mechanism of missing data is related to the missing variable itself. It can be, but is not necessarily, related to other variables. For example, the assumption of MNAR would be satisfied if data were missing for EQ-5D, and those with missing data had lower EQ-5D than those

without. In the registry studies, approximately 75% filled out the three-month form, 80% the 12-month form, and about 90% had at least one completed follow-up form. Less than 5% loss is suggested to not introduce bias, more than 20% is cause for concern, and loss between 5% and 20% creates a potential risk of bias [173].

In studies I and III, the proportions with a clinically important outcome (the responder rates) were estimated by complete case analysis, i.e., only patients with complete data for the outcome parameter(s) were analysed. A prerequisite for complete case analyses is that missingness should occur completely at random (MCAR). Since this assumption could not be verified, a risk of biased estimates cannot be excluded. However, a previous study from NORSpine found that 142 participants (22%) evaluated with extraordinary follow-up routines did not have statistically significantly different clinical outcomes compared to 491 participants (78%) who completed the two-year follow-ups according to the standard routines for the registry [175]. The study indicated that missing data was not related to outcome scores. Hence, a MAR assumption seems to be reasonable. For estimating the mean changes from baseline to three months, the mean change from three months to 12 months, and the mean scores at 12 months, the data were analysed by Latent Growth Curve (LGC) models estimated with Full Information Maximum Likelihood (FIML) [114]. Including all three follow-ups, and hence utilizing all available data under the MAR assumption, the statistical power was enhanced [176]. In study III we performed an additional analysis to ensure that the between-group differences were not altered by the imputation of missing data. Compared to analysis prior to imputation, analysis of the imputed data set did not alter the between-group differences for the outcome scores in the PROMs. This indicates that the risk for biased estimates due to missing data was low in paper III.

In the RCT, we have taken some preventive steps to reduce the risk of attrition bias. The trial is strictly conducted according to follow-up routines, and the importance of adherence to the study has been conveyed to study coordinators and participants. However, some degree of missingness is expected. Under the assumption of missingness at random (MAR), missing values necessary for estimating responder

rates at two-year follow-up will be imputed by Multiple imputation. For continuous secondary outcomes variables, missing data will be managed by using Linear mixed-effects models with Full Information Maximum Likelihood to estimate differences between the treatment groups. Although missing at one or more time point, all available measurements from inclusion to 2-year follow-up will be included in the analysis.

Allocation bias – Propensity score matching

Whereas selection bias deals with recruitment of the studied sample, allocation bias deals with how included participants are assigned to treatment arms [171]. If systematic differences in treatment allocation exist, the relative treatment effect might be confounded by other observed or unobserved variables. For example, if only males receive treatment A and females receive treatment B, a difference in outcome might be associated with gender and not with treatment.

For the RCT, the participants were assigned in a 1:1 ratio to one of two arms. The randomisation was block-permuted (randomly selected block size of 4 and 6 cases) and centre-stratified. Details of block size, allocation sequence generation, and randomisation were unavailable to those who enrolled patients or assigned treatment. Due to the comprehensive allocation concealment, the risk for allocation bias is nearly eliminated for the NORDSTEN-DS trial. In a less concealed randomisation process, where the surgeon responsible for inclusion can sort out the next assigned treatment, some risk of allocation bias can be introduced.

For studies I and III, the allocation to treatment was decided by clinicians in usual clinical conditions. Decision for one treatment over the other was based on the surgeons' experience, their perception of current evidence, the policy/routine at the institution, and sometimes on the patients' wishes. A survey among members of the Lumbar Spine Research Society and the AOSpine reported that patient-related factors such as higher age, absence of low back pain, and absence of instability at extension-flexion radiographs had the highest impact on the decision for performing decompression alone or not [177, 178]. A similar survey from Germany found that the academic status of the hospital and speciality of the surgeon (orthopaedic surgery

vs neurosurgery) in addition to patient age had the most impact on the decision to fuse or not [178]. In papers I and III, allocation to treatment was most likely influenced by patient characteristics, clinical symptoms, and radiological findings. Hence, subjects in the treatment group may systematically differ from those in the control group. For adjustment of such differences in studies I and III, propensity score matching (PSM) [179] was performed before comparing the treatment groups. More specifically, PSM was used to make the distribution of observed baseline patient characteristics between the decompression and fusion group as similar as possible.

Some advantages for case-mix adjustment with PSM, compared to the more commonly used regression adjustment, are upheld [113, 179]. First, it allows for assessment of whether variables included in the model for adjustment successfully balance the observed baseline data of interest. As shown in tables for baseline characteristics (table 1, papers I and II), the matching did successfully create groups with similar distribution of observed baseline parameters. Second, the PSM allows the separation of the case-mix adjustment from the analyses. Following matching, all standard well-known statistical methods (Student T-test, Cross Tabulation with a Chi-Square-Test, etc.) can be used to directly compare treatment groups. This simplifies the interpretation of the results of the analyses.

Although PSM is described as a tool for analysing a non-randomized study so that it "mimics some of the particular characteristics of a randomised study" [113], important limitations should be considered and conclusions should be interpreted with caution. For the present studies, although matching did equalize the baseline scores regarding the observed parameters, the distribution of unobserved parameters may differ between the groups and be a source of bias. Unfortunately, radiological parameters such as the degree of the slip, the presence of 'instability', the disc height, and the amount of fluid and orientation of the facet joint might be unevenly distributed between the decompression alone and the fusion group. Therefore, unlike the RCT, the studies cannot provide unbiased evidence for the causal effect (the

efficacy) of the treatment. Again, the studies generate knowledge of how the treatment works in the real world.

Summarised, we consider the risk of bias to be low for the conducted RCT. For the observational studies, the risk is limited by the high quality of data collected and the use of statistical methods recommended for assessment of effectiveness in observational studies [159]. The methodological quality of paper I was considered high according to the risk of bias when evaluated in two meta-analyses with use of the Newcastle-Ottawa Scale [180]. The study received eight out of nine stars by Liang et al. (the loss of one star was due to lack of long-term follow-up) [140], and nine out of nine by Chen et al. [141].

6 Ethical considerations

6.1 Papers I-III

All patients registered in NORSpine signed a consent form upon admission to surgery. The Norwegian Committee for Medical and Health Research Ethics Midt (2014/344) has approved the studies.

6.2 Paper IV

The study protocol has been approved by the Norwegian Committee for Medical and Health Research Ethics Midt (2013/366).

All patients received information about the study before inclusion. Background, aims, the alternative treatments, and the voluntary nature of participation were presented. If unwilling to participate, the patients were offered treatment according to the surgeon's preference and the practice of the department. Before inclusion, the consent form (appendix x) was signed by the patient and by the surgeon who also confirmed that spoken and written information had been given.

Due to ethical considerations in agreement with the Norwegian Committee for Medical and Health Research Ethics Midt, an interim analysis for safety was performed when 75 patients in each group had completed the 12-month follow-up. An independent statistician blinded for treatment adherence performed the analysis. Only data on reoperations and on the primary outcome measure (ODI) was available to the statistician. Following the analysis, the statistician informed the steering committee, via the central coordinator, that the study could continue. Further information about the analysis has not been disclosed to the study group.

7 Conclusion, implications and future perspectives

This thesis has several important conclusions and implications. Regarding the evaluation of patient reported outcome measures and the criteria for being a responder with a clinically important benefit of the operation, the percentage change score or the absolute follow-up, but not the numerical change score, should be used. Compared to mean outcomes in a measurement, responder rates are considered more clinically relevant, easier to interpret, and are more understandable for patients and health care providers [73]. Comparing two groups, possible advantages and disadvantages should also be taken into account. If one treatment has lower costs or fewer complications, a larger difference in the success rate should be required to conclude that one treatment is superior to another [75, 77, 112, 146]. The Number Needed to Treat can be calculated as the inverse of the difference in responder rate, and, hence, could be included in the discussion about clinically relevant group differences [77, 146]. For example, for DS, how many patients need fusion in addition to decompression to achieve one extra responder? Further, as part of a shared decision-making process, knowledge of the 'success rate' for a surgical treatment will be of great importance for patients and physicians. When calculating responder rates, it is important to utilize condition specific thresholds [73]. For comparing 'success'-rates across studies, it would be beneficial to use identical criteria for 'success'.

Among patients without a clinically relevant treatment effect, it would be informative to distinguish between those who experience themselves to be unchanged or minorly changed and those reporting a substantial deterioration. Unfortunately, such criteria have not been established for spinal stenosis or degenerative spondylolisthesis; further research is therefore needed.

The results from the two observational effectiveness studies (studies I and III) indicate that the majority of DS patients can be operated without fusion. Although non-inferiority was not revealed in study I, we carefully suggested that "a considerable number of patients can be treated with decompression alone". This

statement was extended into a suggestion that micro-decompression alone should be considered prior to decompression with instrumented fusion in study III. Both studies revealed small between-group differences in responder rates, but statistically significant non-inferiority was only claimed for micro-decompression (paper III). These studies showed somewhat more reduction in leg and back pain among fused patients. However, the similar improvement in disability, the higher costs of implants, and the longer operation time and hospital stay suggest that fusion is not necessary for the majority of patients.

The results of the studies have been presented at the Annual Meeting of the Norwegian Orthopedic Federation in 2014, 2016, and 2018, respectively for paper I, II, and III, and at the Annual EUROSPINE Meeting and Congress in 2015, 2017, and 2018, respectively. The research has also been presented regularly in other national meetings and in web-based shared teaching between Norwegian hospitals. There is no indication that the papers or the presentations have impacted the treatment of DS from a global perspective, but they might have contributed to changes in practice in Norway. Data from NORSpine (Figure 8a) shows that from 2013 to 2018 the nationwide rate of fusion procedures decreased with a concomitant increase in the rate of micro-decompression. Interestingly, the change of practice has not altered the clinical outcomes as illustrated in figure 8b.

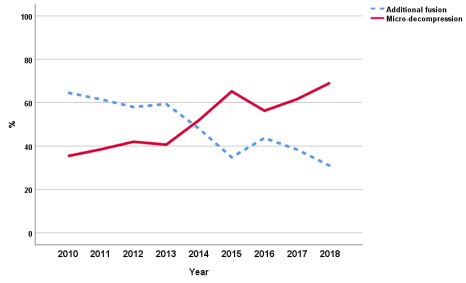


Figure 8a. *Rate of fusion procedures*

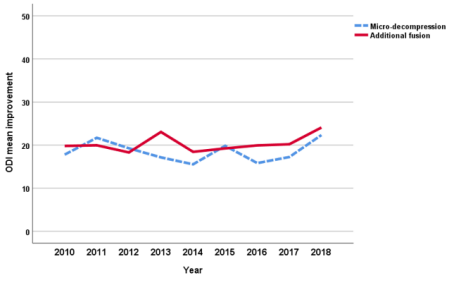


Figure 8b.*Improvement in ODI at 12-month follow-up*

Important future research is also represented by the randomised trial (study IV), as it will provide Level 1evidence of whether decompression alone should be advocated as the preferred method of treatment. The published study protocol was commented on in *The Back Letter* of April 2019. With the title “Major Surgical Question Has No

Answer", it was upheld as a promising high-quality RCT with low risk of bias. Further, the study will investigate whether patient characteristics, preoperative symptoms, and radiological parameters can be used to predict outcome for an individual when operating with micro-decompression and decompression with instrumented fusion, respectively. Hopefully, the planned risk matrix will enable physicians to choose the most appropriate treatment for an individual. The five- and 10-year follow-ups will contribute important long-term knowledge of treatment-related efficacy. Although some loss to follow-up is expected in this older study population, the trial sample will still almost certainly be much greater then currently published RCTs. Finally, health economic analyses will provide important knowledge about the cost-utility related to the investigated treatments.

8 References

1. Kilian, H.F., *De Spondylolisthesi gravissimae Pelvvagustiae caussa nuper detecta commentatio anatomico-obstetriticia.* . Bonnae, 1854.
2. Junghanns, H., *Spondylolisthesen ohne Spalt im Zwischengelenkstück.* Archiv für orthopädische und Unfall-Chirurgie, mit besonderer Berücksichtigung der Frakturenlehre und der orthopädisch-chirurgischen Technik, 1931. **29**(1): p. 118-127.
3. Macnab, I., *Spondylolisthesis with an intact neural arch; the so-called pseudo-spondylolisthesis.* J Bone Joint Surg Br, 1950. **32-b**(3): p. 325-33.
4. Newman, P.H., *Spondylolisthesis, Its Cause and Effect.* Ann. Roy. Coll. Surg, 1955. **19**: p. 305-323.
5. Jacobsen, S., et al., *Degenerative lumbar spondylolisthesis: an epidemiological perspective: the Copenhagen Osteoarthritis Study.* Spine (Phila Pa 1976), 2007. **32**(1): p. 120-5.
6. He, L.C., et al., *Prevalence and risk factors of lumbar spondylolisthesis in elderly Chinese men and women.* Eur Radiol, 2014. **24**(2): p. 441-8.
7. Vogt, M.T., et al., *Lumbar olisthesis and lower back symptoms in elderly white women. The Study of Osteoporotic Fractures.* Spine (Phila Pa 1976), 1998. **23**(23): p. 2640-7.
8. Denard, P.J., et al., *Lumbar spondylolisthesis among elderly men: prevalence, correlates, and progression.* Spine (Phila Pa 1976), 2010. **35**(10): p. 1072-8.
9. Wang, Y.X.J., et al., *Lumbar degenerative spondylolisthesis epidemiology: A systematic review with a focus on gender-specific and age-specific prevalence.* J Orthop Translat, 2017. **11**: p. 39-52.
10. Austevoll, I.M., et al., *The effectiveness of decompression alone compared with additional fusion for lumbar spinal stenosis with degenerative spondylolisthesis: a pragmatic comparative non-inferiority observational study from the Norwegian Registry for Spine Surgery.* Eur Spine J, 2016.
11. Forsth, P., K. Michaelsson, and B. Sanden, *Does fusion improve the outcome after decompressive surgery for lumbar spinal stenosis?: A two-year follow-up study involving 5390 patients.* Bone Joint J., 2013. **95-B**(7): p. 960-965.
12. Weinstein, J.N., et al., *Surgical versus nonsurgical treatment for lumbar degenerative spondylolisthesis.* N Engl J Med, 2007. **356**(22): p. 2257-70.
13. Fitzgerald, J.A. and P.H. Newman, *Degenerative spondylolisthesis.* J Bone Joint Surg Br, 1976. **58**(2): p. 184-92.
14. Matz, P.G., et al., *Guideline summary review: an evidence-based clinical guideline for the diagnosis and treatment of degenerative lumbar spondylolisthesis.* Spine J, 2015.
15. Vergroesen, P.P., et al., *Mechanics and biology in intervertebral disc degeneration: a vicious circle.* Osteoarthritis Cartilage, 2015. **23**(7): p. 1057-70.
16. Rustenburg, C.M.E., et al., *Prognostic factors in the progression of intervertebral disc degeneration: Which patient should be targeted with regenerative therapies?* JOR Spine, 2019. **2**(3): p. e1063.

17. Sambrook, P.N., A.J. MacGregor, and T.D. Spector, *Genetic influences on cervical and lumbar disc degeneration: a magnetic resonance imaging study in twins.* Arthritis Rheum, 1999. **42**(2): p. 366-72.
18. Wang, D.L., S.D. Jiang, and L.Y. Dai, *Biologic response of the intervertebral disc to static and dynamic compression in vitro.* Spine (Phila Pa 1976), 2007. **32**(23): p. 2521-8.
19. Kadow, T., et al., *Molecular basis of intervertebral disc degeneration and herniations: what are the important translational questions?* Clin Orthop Relat Res, 2015. **473**(6): p. 1903-12.
20. Kreiner, D.S., et al., *An evidence-based clinical guideline for the diagnosis and treatment of degenerative lumbar spinal stenosis (update).* Spine J, 2013. **13**(7): p. 734-43.
21. Amundsen, T., et al., *Lumbar spinal stenosis: conservative or surgical management?: A prospective 10-year study.* Spine (Phila Pa 1976), 2000. **25**(11): p. 1424-35; discussion 1435-6.
22. Austevoll, I.M., et al., *The effectiveness of decompression alone compared with additional fusion for lumbar spinal stenosis with degenerative spondylolisthesis: a pragmatic comparative non-inferiority observational study from the Norwegian Registry for Spine Surgery.* Eur Spine J, 2017. **26**(2): p. 404-413.
23. Ghogawala, Z., et al., *Laminectomy plus Fusion versus Laminectomy Alone for Lumbar Spondylolisthesis.* N Engl J Med, 2016. **374**(15): p. 1424-34.
24. Lurie, J. and C. Tomkins-Lane, *Management of lumbar spinal stenosis.* Bmj, 2016. **352**: p. h6234.
25. Weinstein, J.N., et al., *Surgical versus nonsurgical therapy for lumbar spinal stenosis.* N Engl J Med, 2008. **358**(8): p. 794-810.
26. Nerland, U.S., et al., *Minimally invasive decompression versus open laminectomy for central stenosis of the lumbar spine: pragmatic comparative effectiveness study.* Bmj, 2015. **350**: p. h1603.
27. Forsth, P., et al., *A Randomized, Controlled Trial of Fusion Surgery for Lumbar Spinal Stenosis.* N Engl J Med, 2016. **374**(15): p. 1413-23.
28. Pearson, A., et al., *Degenerative spondylolisthesis versus spinal stenosis: does a slip matter? Comparison of baseline characteristics and outcomes (SPORT).* Spine (Phila Pa 1976), 2010. **35**(3): p. 298-305.
29. Hasegawa, K., et al., *Lumbar degenerative spondylolisthesis is not always unstable: clinicobiomechanical evidence.* Spine (Phila Pa 1976), 2014. **39**(26): p. 2127-35.
30. Herkowitz, H.N., - *Degenerative lumbar spondylolisthesis: evolution of surgical management.* - Spine J.2009 Jul;9(7):605-6., 2007: p. 6.
31. Kleinstueck, F.S., et al., *To fuse or not to fuse in lumbar degenerative spondylolisthesis: do baseline symptoms help provide the answer?* Eur Spine J, 2012. **21**(2): p. 268-75.
32. Hamanishi, C., et al., *Cross-sectional area of the stenotic lumbar dural tube measured from the transverse views of magnetic resonance imaging.* J Spinal Disord, 1994. **7**(5): p. 388-93.

33. Schizas, C., et al., *Qualitative grading of severity of lumbar spinal stenosis based on the morphology of the dural sac on magnetic resonance images.* Spine (Phila Pa 1976), 2010. **35**(21): p. 1919-24.
34. Andrasinova, T., et al., *Is there a Correlation Between Degree of Radiologic Lumbar Spinal Stenosis and its Clinical Manifestation?* Clinical Spine Surgery, 2018. **31**(8): p. E403-E408.
35. Mannion, A.F., et al., *Dural sac cross-sectional area and morphological grade show significant associations with patient-rated outcome of surgery for lumbar central spinal stenosis.* Eur Spine J, 2017. **26**(10): p. 2552-2564.
36. Weber, C., et al., *Surgical management of lumbar spinal stenosis: a survey among Norwegian spine surgeons.* Acta Neurochir (Wien), 2017. **159**(1): p. 191-197.
37. Hermansen, E., et al., *Study-protocol for a randomized controlled trial comparing clinical and radiological results after three different posterior decompression techniques for lumbar spinal stenosis: the Spinal Stenosis Trial (SST) (part of the NORDSTEN Study).* BMC Musculoskelet Disord, 2017. **18**(1): p. 121.
38. Iguchi, T., et al., *Lumbar instability and clinical symptoms: which is the more critical factor for symptoms: sagittal translation or segment angulation?* J Spinal Disord Tech, 2004. **17**(4): p. 284-90.
39. Spina, N., et al., *Defining Instability in Degenerative Spondylolisthesis: Surgeon Views.* Clin Spine Surg, 2019.
40. Mannion, A.F., et al., *Development of appropriateness criteria for the surgical treatment of symptomatic lumbar degenerative spondylolisthesis (LDS).* Eur Spine J, 2014. **23**(9): p. 1903-17.
41. Matsunaga, S., K. Ijiri, and K. Hayashi, *Nonsurgically managed patients with degenerative spondylolisthesis: a 10- to 18-year follow-up study.* J Neurosurg, 2000. **93**(2 Suppl): p. 194-8.
42. Weinstein, J.N., et al., *Surgical compared with nonoperative treatment for lumbar degenerative spondylolisthesis. four-year results in the Spine Patient Outcomes Research Trial (SPORT) randomized and observational cohorts.* J Bone Joint Surg Am, 2009. **91**(6): p. 1295-304.
43. Ammendolia, C., et al., *Nonoperative treatment for lumbar spinal stenosis with neurogenic claudication.* Cochrane Database Syst Rev, 2013(8): p. Cd010712.
44. Cauchoix, J., M. Benoist, and V. Chassaing, *Degenerative spondylolisthesis.* Clin Orthop Relat Res, 1976(115): p. 122-9.
45. Hermansen, E., et al., *Does surgical technique influence clinical outcome after lumbar spinal stenosis decompression? A comparative effectiveness study from the Norwegian Registry for Spine Surgery.* Eur Spine J, 2017. **26**(2): p. 420-427.
46. Kuo, C.C., et al., *In Degenerative Spondylolisthesis, Unilateral Laminotomy for Bilateral Decompression Leads to Less Reoperations at 5 Years When Compared to Posterior Decompression With Instrumented Fusion: A Propensity-matched Retrospective Analysis.* Spine (Phila Pa 1976), 2019. **44**(21): p. 1530-1537.

47. 120 Thome, C.F., et al., - *Outcome after less-invasive decompression of lumbar spinal stenosis: a randomized comparison of unilateral laminotomy, bilateral laminotomy, and laminectomy.* - J Neurosurg Spine.2005 Aug;3(2):129-41., 2002: p. 41.
48. Park, H.K. and J.C. Chang, *Microdecompression in spinal stenosis: a review.* J Neurosurg Sci, 2014. **58**(2): p. 57-64.
49. Thome, C., et al., *Outcome after less-invasive decompression of lumbar spinal stenosis: a randomized comparison of unilateral laminotomy, bilateral laminotomy, and laminectomy.* J Neurosurg Spine, 2005. **3**(2): p. 129-41.
50. Overdevest, G., et al., *Effectiveness of posterior decompression techniques compared with conventional laminectomy for lumbar stenosis.* Eur Spine J, 2015. **24**(10): p. 2244-63.
51. Solberg, T., L.R. Olsen, and M.B. Berglund, *The Norwegian registry for spine surgery, Annual Report.* 2018.
52. Overdevest, G.M., et al., *Management of lumbar spinal stenosis: a survey among Dutch spine surgeons.* Acta Neurochir (Wien), 2014. **156**(11): p. 2139-45.
53. Pisano, A.J., et al., *Does Surgically Managed Grade I Degenerative Lumbar Spondylolisthesis Require Fusion?* Clin Spine Surg, 2018.
54. Kornblum, M.B., et al., *Degenerative lumbar spondylolisthesis with spinal stenosis: a prospective long-term study comparing fusion and pseudarthrosis.* Spine (Phila Pa 1976), 2004. **29**(7): p. 726-33; discussion 733-4.
55. Martin, C.R., et al., *The surgical management of degenerative lumbar spondylolisthesis: a systematic review.* Spine (Phila Pa 1976), 2007. **32**(16): p. 1791-8.
56. Watters, W.C., 3rd, et al., *An evidence-based clinical guideline for the diagnosis and treatment of degenerative lumbar spondylolisthesis.* Spine J, 2009. **9**(7): p. 609-14.
57. Kepler, C.K., et al., *National trends in the use of fusion techniques to treat degenerative spondylolisthesis.* Spine (Phila Pa 1976), 2014. **39**(19): p. 1584-9.
58. Kim, C.H., et al., *Increased Proportion of Fusion Surgery for Degenerative Lumbar Spondylolisthesis and Changes in Reoperation Rate: A Nationwide Cohort Study with a Minimum 5-Year Follow-Up.* Spine (Phila Pa 1976), 2018.
59. Herkowitz, H.N. and L.T. Kurz, *Degenerative lumbar spondylolisthesis with spinal stenosis. A prospective study comparing decompression with decompression and intertransverse process arthrodesis.* J Bone Joint Surg Am, 1991. **73**(6): p. 802-8.
60. Fischgrund, J.S., et al., *1997 Volvo Award winner in clinical studies. Degenerative lumbar spondylolisthesis with spinal stenosis: a prospective, randomized study comparing decompressive laminectomy and arthrodesis with and without spinal instrumentation.* Spine (Phila Pa 1976), 1997. **22**(24): p. 2807-12.

61. Chang, P., K.H. Seow, and S.K. Tan, *Comparison of the results of spinal fusion for spondylolisthesis in patients who are instrumented with patients who are not.* Singapore Med J, 1993. **34**(6): p. 511-4.
62. Lee, T.C., *Complications of transpedicular reduction and stabilization of the thoracolumbar spine.* J Formos Med Assoc, 1995. **94**(12): p. 738-41.
63. Lombardi JS, F.A.U., et al., - *Treatment of degenerative spondylolisthesis.* - Spine (Phila Pa 1976).1985 Nov;10(9):821-7., 2012: p. 7.
64. Resnick, D.K., et al., *Guidelines for the performance of fusion procedures for degenerative disease of the lumbar spine. Part 9: fusion in patients with stenosis and spondylolisthesis.* J Neurosurg Spine, 2005. **2**(6): p. 679-85.
65. Watters, W.C., 3rd, et al., *Degenerative lumbar spinal stenosis: an evidence-based clinical guideline for the diagnosis and treatment of degenerative lumbar spinal stenosis.* Spine J, 2008. **8**(2): p. 305-10.
66. Lonne, G., et al., *Lumbar spinal stenosis: comparison of surgical practice variation and clinical outcome in three national spine registries.* Spine J, 2019. **19**(1): p. 41-49.
67. Resnick, D.K., et al., *Guideline update for the performance of fusion procedures for degenerative disease of the lumbar spine. Part 9: lumbar fusion for stenosis with spondylolisthesis.* J Neurosurg Spine, 2014. **21**(1): p. 54-61.
68. Deyo, R.A., et al., *Trends, major medical complications, and charges associated with surgery for lumbar spinal stenosis in older adults.* Jama, 2010. **303**(13): p. 1259-65.
69. Bae, H.W., S.S. Rajaee, and L.E. Kanim, *Nationwide trends in the surgical management of lumbar spinal stenosis.* Spine (Phila Pa 1976), 2013. **38**(11): p. 916-26.
70. Andrews, J.C., et al., *GRADE guidelines: 15. Going from evidence to recommendation-determinants of a recommendation's direction and strength.* J Clin Epidemiol, 2013. **66**(7): p. 726-35.
71. Sigmundsson, F.G., B. Jonsson, and B. Stromqvist, *Outcome of decompression with and without fusion in spinal stenosis with degenerative spondylolisthesis in relation to preoperative pain pattern: a register study of 1,624 patients.* Spine J, 2015. **15**(4): p. 638-46.
72. Birkmeyer, N.J., et al., *Design of the Spine Patient outcomes Research Trial (SPORT).* Spine (Phila Pa 1976), 2002. **27**(12): p. 1361-72.
73. Katz, N.P., F.C. Paillard, and E. Ekman, *Determining the clinical importance of treatment benefits for interventions for painful orthopedic conditions.* J Orthop Surg Res, 2015. **10**: p. 24.
74. Jaeschke, R., J. Singer, and G.H. Guyatt, *Measurement of health status. Ascertaining the minimal clinically important difference.* Control Clin Trials, 1989. **10**(4): p. 407-15.
75. Glassman, S.D., et al., *Defining substantial clinical benefit following lumbar spine arthrodesis.* J Bone Joint Surg Am, 2008. **90**(9): p. 1839-47.
76. Fekete, T.F., et al., *What level of pain are patients happy to live with after surgery for lumbar degenerative disorders?* Spine J, 2016.

77. Dworkin, R.H., et al., *Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations.* J Pain, 2008. **9**(2): p. 105-21.
78. Gatchel, R.J. and T.G. Mayer, *Testing minimal clinically important difference: consensus or conundrum?* Spine J, 2010. **10**(4): p. 321-7.
79. Terwee, C.B., et al., *Mind the MIC: large variation among populations and methods.* Journal of Clinical Epidemiology, 2010. **63**(5): p. 524-534.
80. Wright, A., et al., *Clinimetrics corner: a closer look at the minimal clinically important difference (MCID).* J Man Manip Ther, 2012. **20**(3): p. 160-6.
81. Merali, Z. and J.R. Wilson, *Explanatory Versus Pragmatic Trials: An Essential Concept in Study Design and Interpretation.* Clin Spine Surg, 2017. **30**(9): p. 404-406.
82. Sedgwick, P., *Explanatory trials versus pragmatic trials.* Bmj, 2014. **349**: p. g6694.
83. Thorpe, K.E., et al., *A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers.* Cmaj, 2009. **180**(10): p. E47-57.
84. Corrigan-Curay, J., L. Sacks, and J. Woodcock, *Real-World Evidence and Real-World Data for Evaluating Drug Safety and Effectiveness.* Jama, 2018. **320**(9): p. 867-868.
85. Bui, T.B.V., et al., *Real-World Effectiveness of Palbociclib Versus Clinical Trial Results in Patients With Advanced/Metastatic Breast Cancer That Progressed on Previous Endocrine Therapy.* Breast Cancer (Auckl), 2019. **13**: p. 1178223418823238.
86. Kaye, I.D., et al., *Spine Registries: Where Do We Stand?* Clin Spine Surg, 2017.
87. Derogatis, L.R., et al., *The Hopkins Symptom Checklist (HSCL): a self-report symptom inventory.* Behav Sci, 1974. **19**(1): p. 1-15.
88. Dupuis, P.R., et al., *Radiologic diagnosis of degenerative lumbar spinal instability.* Spine (Phila Pa 1976), 1985. **10**(3): p. 262-76.
89. Schwab, F., et al., *Sagittal plane considerations and the pelvis in the adult patient.* Spine (Phila Pa 1976), 2009. **34**(17): p. 1828-33.
90. Lee, S., et al., *A practical MRI grading system for lumbar foraminal stenosis.* AJR Am.J.Roentgenol., 2010. **194**(4): p. 1095-1098.
91. Cho, I.Y., et al., *MRI findings of lumbar spine instability in degenerative spondylolisthesis.* J Orthop Surg (Hong Kong), 2017. **25**(2): p. 2309499017718907.
92. Pfirrmann, C.W., et al., *Magnetic resonance classification of lumbar intervertebral disc degeneration.* Spine (Phila Pa 1976.), 2001. **26**(17): p. 1873-1878.
93. Modic, M.T., et al., *Degenerative disk disease: assessment of changes in vertebral body marrow with MR imaging.* Radiology, 1988. **166**(1 Pt 1): p. 193-9.
94. Fairbank, J.C., et al., *The Oswestry low back pain disability questionnaire.* Physiotherapy, 1980. **66**(8): p. 271-3.
95. Steinberger, J., et al., *The top 100 classic papers in lumbar spine surgery.* Spine (Phila Pa 1976), 2015. **40**(10): p. 740-7.

96. Fairbank, J.C. and P.B. Pynsent, *The Oswestry Disability Index.* Spine (Phila Pa 1976), 2000. **25**(22): p. 2940-52; discussion 2952.
97. Grotle, M., J.I. Brox, and N.K. Vollestad, *Cross-cultural adaptation of the Norwegian versions of the Roland-Morris Disability Questionnaire and the Oswestry Disability Index.* J.Rehabil.Med., 2003. **35**(5): p. 241-247.
98. Tuli, S.K., S.A. Yerby, and J.N. Katz, *Methodological approaches to developing criteria for improvement in lumbar spinal stenosis surgery.* Spine (Phila Pa 1976), 2006. **31**(11): p. 1276-80.
99. Thornes, E. and M. Grotle, *Cross-cultural adaptation of the Norwegian version of the spinal stenosis measure.* Eur.Spine J., 2008. **17**(3): p. 456-462.
100. Hjermstad, M.J., et al., *Studies comparing Numerical Rating Scales, Verbal Rating Scales, and Visual Analogue Scales for assessment of pain intensity in adults: a systematic literature review.* J Pain Symptom Manage, 2011. **41**(6): p. 1073-93.
101. Froud, R., et al., *Responsiveness, Reliability, and Minimally Important and Minimal Detectable Changes of 3 Electronic Patient-Reported Outcome Measures for Low Back Pain: Validation Study.* J Med Internet Res, 2018. **20**(10): p. e272.
102. Gallasch, C.H. and N.M. Alexandre, *The measurement of musculoskeletal pain intensity: a comparison of four methods.* Rev Gaucha Enferm, 2007. **28**(2): p. 260-5.
103. *EuroQol--a new facility for the measurement of health-related quality of life.* Health Policy, 1990. **16**(3): p. 199-208.
104. Solberg, T.K., et al., *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): p. 1000-7.
105. Lonne, G., et al., *Comparing cost-effectiveness of X-Stop with minimally invasive decompression in lumbar spinal stenosis: a randomized controlled trial.* Spine (Phila Pa 1976), 2015. **40**(8): p. 514-20.
106. Grovle, L., et al., *Patients' ratings of global perceived change during 2 years were strongly influenced by the current health status.* J Clin Epidemiol, 2014. **67**(5): p. 508-15.
107. Kamper, S.J., et al., *Global Perceived Effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status.* J.Clin.Epidemiol., 2010. **63**(7): p. 760-766.
108. Copay, A.G., et al., *Minimum clinically important difference in lumbar spine surgery patients: a choice of methods using the Oswestry Disability Index, Medical Outcomes Study questionnaire Short Form 36, and pain scales.* Spine J, 2008. **8**(6): p. 968-74.
109. Austevoll, I.M., et al., *Follow-up score, change score or percentage change score for determining clinical important outcome following surgery? An observational study from the Norwegian registry for Spine surgery evaluating patient reported outcome measures in lumbar spinal stenosis and lumbar degenerative spondylolisthesis.* BMC Musculoskelet Disord, 2019. **20**(1): p. 31.

110. Austevoll, I.M., et al., *Decompression alone versus decompression with instrumental fusion the NORDSTEN degenerative spondylolisthesis trial (NORDSTEN-DS); study protocol for a randomized controlled trial.* BMC Musculoskelet Disord, 2019. **20**(1): p. 7.
111. Blumenthal, S., 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 (Phila Pa 1976), 2005. **30**(14): p. 1565-75; discussion E387-91.
112. Katz, N., F.C. Paillard, and R. Van Inwegen, *A review of the use of the number needed to treat to evaluate the efficacy of analgesics.* J Pain, 2015. **16**(2): p. 116-23.
113. Austin, P.C., *An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies.* Multivariate.Behav.Res., 2011. **46**(3): p. 399-424.
114. Bollen, K.A. and P.J. Curran, *Latent curve models: A structural equation perspective*. 2006, Hoboken, N.J.: Wiley-Interscience. XII, 285 s.
115. Muthén, L.K. and B.O. Muthén, *Mplus 7.3*. 2014, Muthén & Muthén, 3463 Stoner Avenue, CA 90066: Los Angeles.
116. Altman, D.G. and J.M. Bland, *Diagnostic tests 3: receiver operating characteristics plots.* BMJ, 1994. **309**: p. 188.
117. Tape, T.G., *Interpreting diagnostic tests.* http://gim.unmc.edu/dxtests/Default.htm, 2006 Dec 18.
118. Fagerland, M.W., S. Lydersen, and P. Laake, *Recommended confidence intervals for two independent binomial proportions.* Stat.Methods Med.Res., 2011.
119. Fagerland, M.W., S. Lydersen, and P. Laake, *Recommended confidence intervals for two independent binomial proportions.* Stat Methods Med Res, 2011.
120. Schafer, J.L. and J.W. Graham, *Missing data: our view of the state of the art.* Psychol Methods, 2002. **7**(2): p. 147-77.
121. Muthen, L.K. and M.O. Muthen, *Mplus 8. Los Angeles: Muthén & Muthén, 3463 Stoner Avenue, CA 90066.* 2017.
122. Yamaguchi, Y., et al., *Multiple imputation for longitudinal data in the presence of heteroscedasticity between treatment groups.* Journal of Biopharmaceutical Statistics, 2020. **30**(1): p. 178-196.
123. Floden, L. and M.L. Bell, *Imputation strategies when a continuous outcome is to be dichotomized for responder analysis: a simulation study.* BMC Med Res Methodol, 2019. **19**(1): p. 161.
124. Hosmer, D.W. and S. Lemeshow, *Applied Logistic Regression, 2nd Edition* New York:Wiley, 2000.
125. Conroy, R.M., et al., *Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project.* Eur Heart J, 2003. **24**(11): p. 987-1003.

126. Solberg, I.C., et al., *Risk matrix model for prediction of colectomy in a population-based study of ulcerative colitis patients (the IBSEN study).* Scand J Gastroenterol, 2015. **50**(12): p. 1456-62.
127. Blackwelder, W.C. and M.A. Chang, *Sample size graphs for "proving the null hypothesis".* Control Clin Trials, 1984. **5**(2): p. 97-105.
128. Bridwell, K.H., et al., *The role of fusion and instrumentation in the treatment of degenerative spondylolisthesis with spinal stenosis.* J Spinal Disord, 1993. **6**(6): p. 461-72.
129. Mardjetko, S.M., P.J. Connolly, and S. Shott, *Degenerative lumbar spondylolisthesis. A meta-analysis of literature 1970-1993.* Spine (Phila Pa 1976), 1994. **19**(20 Suppl): p. 2256s-2265s.
130. Matz, P.G., et al., *Guideline summary review: An evidence-based clinical guideline for the diagnosis and treatment of degenerative lumbar spondylolisthesis.* Spine J, 2016. **16**(3): p. 439-48.
131. Carragee, E.J., *The increasing morbidity of elective spinal stenosis surgery: is it necessary?* Jama, 2010. **303**(13): p. 1309-10.
132. Deyo, R.A., *Fusion surgery for lumbar degenerative disc disease: still more questions than answers.* Spine J, 2015. **15**(2): p. 272-4.
133. Pearson, A.M., *Fusion in degenerative spondylolisthesis: how to reconcile conflicting evidence.* Journal of Spine Surgery, 2016. **2**(2): p. 143-145.
134. Peul, W.C. and W.A. Moojen, *Fusion for Lumbar Spinal Stenosis--Safeguard or Superfluous Surgical Implant?* N Engl J Med, 2016. **374**(15): p. 1478-9.
135. Weinstein, J. and A. Pearson, *Fusion in degenerative spondylolisthesis becomes controversial...again.* Evid Based Med, 2016. **21**(4): p. 148-9.
136. Forsth, P., K. Michaelsson, and B. Sanden, *Fusion Surgery for Lumbar Spinal Stenosis.* N Engl J Med, 2016. **375**(6): p. 599-600.
137. Forsth, P., K. Michaelsson, and B. Sanden, *More on Fusion Surgery for Lumbar Spinal Stenosis.* N Engl J Med, 2016. **375**(18): p. 1806-1807.
138. Ghogawala, Z., et al., *Prospective outcomes evaluation after decompression with or without instrumented fusion for lumbar stenosis and degenerative Grade I spondylolisthesis.* J Neurosurg Spine, 2004. **1**(3): p. 267-72.
139. Merali, Z. and J.R. Wilson, *Explanatory Versus Pragmatic Trials: An Essential Concept in Study Design and Interpretation.* Clin Spine Surg, 2017.
140. Liang, H.F., et al., *Decompression plus fusion versus decompression alone for degenerative lumbar spondylolisthesis: a systematic review and meta-analysis.* Eur Spine J, 2017. **26**(12): p. 3084-3095.
141. Chen, Z., et al., *Decompression Alone Versus Decompression and Fusion for Lumbar Degenerative Spondylolisthesis: A Meta-Analysis.* World Neurosurg, 2018. **111**: p. e165-e177.
142. Samuel, A.M., H.G. Moore, and M.E. Cunningham, *Treatment for Degenerative Lumbar Spondylolisthesis: Current Concepts and New Evidence.* Curr Rev Musculoskelet Med, 2017. **10**(4): p. 521-529.
143. de Vet, H.C., et al., *Minimally important change values of a measurement instrument depend more on baseline values than on the type of intervention.* J Clin Epidemiol, 2015. **68**(5): p. 518-24.

144. van Hooff, M.L., et al., *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.* Spine J, 2016. **16**(10): p. 1221-1230.
145. Ostelo, R.W., et al., *Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change.* Spine (Phila Pa 1976), 2008. **33**(1): p. 90-4.
146. Guyatt, G.H., et al., *Methods to explain the clinical significance of health status measures.* Mayo Clin Proc, 2002. **77**(4): p. 371-83.
147. Scholler, K., et al., *Lumbar Spinal Stenosis Associated With Degenerative Lumbar Spondylolisthesis: A Systematic Review and Meta-analysis of Secondary Fusion Rates Following Open vs Minimally Invasive Decompression.* Neurosurgery, 2017. **80**(3): p. 355-367.
148. Inose, H., et al., *Comparison of Decompression, Decompression Plus Fusion, and Decompression Plus Stabilization for Degenerative Spondylolisthesis: A Prospective, Randomized Study.* Clin Spine Surg, 2018. **31**(7): p. E347-e352.
149. Matsudaira, K., et al., *Spinal stenosis in grade I degenerative lumbar spondylolisthesis: a comparative study of outcomes following laminoplasty and laminectomy with instrumented spinal fusion.* J Orthop Sci, 2005. **10**(3): p. 270-6.
150. Park, J.H., et al., *A comparison of unilateral laminectomy with bilateral decompression and fusion surgery in the treatment of grade I lumbar degenerative spondylolisthesis.* Acta Neurochir (Wien), 2012. **154**(7): p. 1205-12.
151. Rampersaud, Y.R., et al., *Health-related quality of life following decompression compared to decompression and fusion for degenerative lumbar spondylolisthesis: a Canadian multicentre study.* Can J Surg, 2014. **57**(4): p. E126-33.
152. Inui, T., et al., *Lumbar Degenerative Spondylolisthesis: Changes in Surgical Indications and Comparison of Instrumented Fusion With Two Surgical Decompression Procedures.* Spine (Phila Pa 1976), 2017. **42**(1): p. E15-e24.
153. Chang, H.S., et al., *Degenerative spondylolisthesis does not affect the outcome of unilateral laminotomy with bilateral decompression in patients with lumbar stenosis.* Spine (Phila Pa 1976), 2014. **39**(5): p. 400-8.
154. Kim, S., et al., *Cost-utility of lumbar decompression with or without fusion for patients with symptomatic degenerative lumbar spondylolisthesis.* Spine J, 2012. **12**(1): p. 44-54.
155. Solberg, T., L.R. Olsen, and M.B. Berglund, *The Norwegian Registry for Spinal Surgery. Annual Report.* 2015.
156. Weinstein, J.N., et al., *Surgical versus nonoperative treatment for lumbar disc herniation: four-year results for the Spine Patient Outcomes Research Trial (SPORT).* Spine (Phila Pa 1976), 2008. **33**(25): p. 2789-800.
157. Stromqvist, B., et al., *The Swedish Spine Register: development, design and utility.* Eur Spine J, 2009. **18 Suppl 3**: p. 294-304.
158. Cleland, J.A., et al., *Psychometric properties of selected tests in patients with lumbar spinal stenosis.* Spine J, 2012. **12**(10): p. 921-31.

159. Faraoni, D. and S.T. Schaefer, *Randomized controlled trials vs. observational studies: why not just live together?* BMC Anesthesiol, 2016. **16**(1): p. 102.
160. Shakir, N.A., et al., *Identification of threshold prostate specific antigen levels to optimize the detection of clinically significant prostate cancer by magnetic resonance imaging/ultrasound fusion guided biopsy.* J Urol, 2014. **192**(6): p. 1642-8.
161. Carreon, L.Y., et al., *Differentiating minimum clinically important difference for primary and revision lumbar fusion surgeries.* J Neurosurg Spine, 2013. **18**(1): p. 102-6.
162. Hagg, O., P. Fritzell, and A. Nordwall, *The clinical importance of changes in outcome scores after treatment for chronic low back pain.* Eur Spine J, 2003. **12**(1): p. 12-20.
163. Solberg, T., et al., *Can we define success criteria for lumbar disc surgery? : estimates for a substantial amount of improvement in core outcome measures.* Acta Orthop, 2013. **84**(2): p. 196-201.
164. Mannion, A.F., et al., *Could less be more when assessing patient-rated outcome in spinal stenosis?* Spine (Phila Pa 1976), 2015. **40**(10): p. 710-8.
165. Malmivaara, A., et al., *Surgical or nonoperative treatment for lumbar spinal stenosis? A randomized controlled trial.* Spine (Phila Pa 1976), 2007. **32**(1): p. 1-8.
166. Barrett, B., et al., *Sufficiently important difference: expanding the framework of clinical significance.* Med Decis Making, 2005. **25**(3): p. 250-61.
167. Ferreira, M.L., et al., *People with low back pain typically need to feel 'much better' to consider intervention worthwhile: an observational study.* Aust J Physiother, 2009. **55**(2): p. 123-7.
168. Altman, D.G. and J.M. Bland, *Absence of evidence is not evidence of absence.* Bmj, 1995. **311**(7003): p. 485.
169. Kesmodel, U.S., *Information bias in epidemiological studies with a special focus on obstetrics and gynecology.* Acta Obstet Gynecol Scand, 2018. **97**(4): p. 417-423.
170. Smuck, M., et al., *Objective measurement of function following lumbar spinal stenosis decompression reveals improved functional capacity with stagnant real-life physical activity.* Spine J, 2018. **18**(1): p. 15-21.
171. Sedgwick, P., *Selection bias versus allocation bias.* BMJ Evid Based Med, 2013.
172. Ware, J.H., et al., *Missing Data.* 2012. **367**(14): p. 1353-1354.
173. Nunan, D., J. Aronson, and C. Bankhead, *Catalogue of bias: attrition bias.* BMJ Evid Based Med, 2018. **23**(1): p. 21-22.
174. Jakobsen, J.C., et al., *When and how should multiple imputation be used for handling missing data in randomised clinical trials – a practical guide with flowcharts.* 2017. **17**(1): p. 162.
175. Solberg, T.K., et al., *Would loss to follow-up bias the outcome evaluation of patients operated for degenerative disorders of the lumbar spine?* Acta Orthop., 2011. **82**(1): p. 56-63.

176. Wang, J. and X. Wang, *Structural Equation Modeling: Applications Using Mplus.* West Sussex, UK: Wiley, A John Wiley & Sons, Ltd., Publication., 2012.
177. Schroeder, G.D., et al., *Rationale for the Surgical Treatment of Lumbar Degenerative Spondylolisthesis.* Spine (Phila Pa 1976), 2015. **40**(21): p. E1161-6.
178. Strube, P., et al., *To fuse or not to fuse: a survey among members of the German Spine Society (DWG) regarding lumbar degenerative spondylolisthesis and spinal stenosis.* Arch Orthop Trauma Surg, 2019. **139**(5): p. 613-621.
179. Rosenbaum PR and R. DB., *The central role of the propensity score in observational studies for causal effects.* Biometrika1983; 70:41-55, 1985.
180. Stang, A., *Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses.* Eur J Epidemiol, 2010. **25**(9): p. 603-5.

9 Appendices

Appendix I

Surgeon form from the Norwegian Registry for Spine Surgery

SKJEMA 2A:
SYKEPLEIER/LEGEOPPLYSNINGER PREOPERATIVT
(Fylles ut av lege samtidig med operasjonsbeskrivelsen og suppleres evt. ved utstrivelse eller ved innrapportering)

Registreringsskjema for pasienter som opereres i ryggen

Nasjonalt Kvalitetsregister for Ryggkirurgi

Nasjonalt kvalitetsregister for ryggkirurgi
Degenerativ LS

Senter for Klinisk Dokumentasjon og Evaluering - Helse Nord RHF
E-post: ryggregisteret@unn.no
Hjemmeside: www.ryggregisteret.no

1108 - Versjon 2

Operasjonsdato *(Må fylles ut)* Dag ☐☐ Måned ☐☐ År ☐☐

Dato for utfylling Dag ☐☐ Måned ☐☐ År ☐☐

Pasientdata (Barkode)

Navn

Fødselsnr. (11 siffer) ☐☐☐☐☐☐ ☐☐☐☐☐

Sykehistorie

Tidligere ryggoperert?

☐ Ja, samme nivå ☐ Ja, annet nivå ☐ Nei

- Pasienten har vært operert ☐☐ ganger tidligere i LS-kolumna

Andre relevante sykdommer, skader eller plager

☐ Nei

Ja, spesifiser:

- ☐ Reumatoid artritt
- ☐ Mb. Bechterew
- ☐ Annen reumatisk sykdom
- ☐ Hofte- eller kneartrose
- ☐ Depresjon / Angst
- ☐ Kroniske smerter i muskel-skjelettsystemet
- ☐ Kronisk nevrologisk sykdom
- ☐ Cerebrovaskulær sykdom
- ☐ Hjerte eller karsykdom
- ☐ Vaskulær Claudicatio
- ☐ Kronisk lungesykdom
- ☐ Kreftsykdom
- ☐ Osteoporose
- ☐ Hypertensjon
- ☐ Diabetes Mellitus
- ☐ Annen endokrin sykdom

Annet, spesifiser _____

Radiologisk vurdering (Sett evntuelt flere kryss)

1. Undersøkelse

- ☐ CT
- ☐ MR
- ☐ Radikulografi
- ☐ Diskografi
- ☐ Diagnostisk blokade
- ☐ Røntgen LS-columna
- ☐ Med fleksjon/ekstensjon

2. Funn

- ☐ Normal
- ☐ Skiveprolaps
- ☐ Sentral spinalstenose
- ☐ Lateral spinalstenose
- ☐ Foraminal stenose
- ☐ Degenerativ rygg/skivedegenerasjon
- ☐ Istmisk spondylolistese
- ☐ Degenerativ spondylolistese
- ☐ Degenerativ skoliose
- ☐ Synovial syste
- ☐ Pseudomeningocele

☐ Annet, spesifiser _____

Operasjonsindikasjon (Sett evntuelt flere kryss)

- ☐ Smerter
  - ☐ Rygg-/hoftesmerter
  - ☐ Bensmerter
  - ☐ Begge deler
- ☐ Parese, Grad (0-5): ..... Se eventuelt rettledning
- ☐ Cauda equina syndrom
- ☐ Annet, spesifiser _____

Ved tidlig reoperasjon (innen 90 dager), årsak: *(Kun ett kryss)*

- ☐ Recidiv prolaps
- ☐ Durarift
- ☐ Hematom
- ☐ Dyp infeksjon
- ☐ Overfladisk infeksjon
- ☐ Postoperativ spondylolisthese
- ☐ Løsning/feilplassering av osteosyntesemateriale

☐ Annet, spesifiser _____

Operasjonskategori

☐ Elektiv ☐ Øyeblikkelig hjelp ☐ ½ øyeblikkelig hjelp

Dagkirurgi (ingen døgnopphold på avdelingen)

☐ Ja ☐ Nei

ASA-klassifisering

- ☐ I — Ingen organisk, fysiologisk, biokjemisk eller psykisk forstyrrelse. Den aktuelle lidelsen er lokalisert og gir ikke generelle systemforstyrrelser
- ☐ II — Moderat sykdom eller forstyrrelse som ikke forårsaker funksjonelle begrensninger
- ☐ III — Alvorlig sykdom eller forstyrrelse som gir definerte funksjonelle begrensninger
- ☐ IV — Livstruende organisk sykdom som ikke behøver å være knyttet til den aktuelle kirurgiske lidelse eller som ikke bedres ved det planlagte kirurgiske inngrepet
- ☐ V — Døende pasient som ikke forventes å overleve 24 timer uten kirurgi

LUNDBLAD MEDIA AS, TROMSØ – O-92101

SNU

Operasjonsmetode (Sett evt. flere kryss)

Har operatøren brukt mikroskop eller lupebriller?

☐ Ja ☐ Nei

Prolapsekstirpasjon?

☐ Nei

☐ Ja, med tømming av skive (diskektomi)

☐ Ja, uten tømming av skive

Kirurgisk dekompresjon

☐ Dekompresjon med bevaring av midtlinjestrukturer

- ☐ Unilateral
- ☐ Bilateral med unilateral tilgang
- ☐ Bilateral med bilateral tilgang

☐ Laminektomi

☐ Fasettektomi i ett eller flere nivåer

- ☐ Unilateral
- ☐ Bilateral

Andre operasjonsmetoder

- ☐ Endoskopi
- ☐ Minimal invasiv prosedyre (tube kirurgi)
- ☐ Ekspanderende interspinøst implantat
- ☐ Fjerning av ekspanderende interspinøst implantat
- ☐ Skiveprotese
- ☐ Nukleus implantat
- ☐ Nukleutomi
- ☐ Kjemonukleolyse
- ☐ Revisjon av osteosyntesematerialet
- ☐ Fjerning av osteosyntesemateriale

Annet, spesifiser _____

Tilgang (sett eventuelt flere kryss)

☐ Midtlinje

☐ Lateral tilgang (Wiltze)

☐ Fremre

Ved fusjonskirurgi (sett eventuelt flere kryss)

☐ Posterolateral fusjon

- ☐ Instrumentell
- ☐ Bengraft

☐ ALIF

- ☐ Bur (cage)
- ☐ Benblokk i skiverom

☐ PLIF

- ☐ Bur (cage)
- ☐ Kun benblokk

☐ TLIF

- ☐ Bur (cage)
- ☐ Kun benblokk

Annet, spesifiser _____

Type bengraft (sett eventuelt flere kryss)

☐ Autograft

☐ Bensubstitutt

☐ Bank-ben

Operert nivå og side (Sett eventuelt flere kryss)

☐ L2/3	☐ Hø.	☐ Ve.
☐ L3/4	☐ Hø.	☐ Ve.
☐ L4/5	☐ Hø.	☐ Ve.
☐ L5/S1	☐ Hø.	☐ Ve.

Annet, spesifiser _____

Antibiotikaprofylakse

☐ Ja ☐ Nei

Sårdren

☐ Ja ☐ Nei

Knivtid (hud til hud)

Opr. start (klokkeslett) ☐☐ ☐☐ (timer/min)

Opr. slutt (klokkeslett) ☐☐ ☐☐ (timer/min)

Evt. samlet knivtid (kalkuleres atuomatisk). ☐☐ ☐☐ (timer/min)

Peroperative komplikasjoner:

☐ Durarift/liquorlekasje

☐ Nerverotskade

☐ Operert på feil nivå/side

☐ Feil plassering av implantat

☐ Transfusjonskrevende peroperativ blødning

☐ Respiratoriske komplikasjoner

☐ Kardiovaskulære komplikasjone

☐ Anafylaktisk reaksjon

☐ Annet, spesifiser _____

Oppgi inntil to operasjonskoder som best beskriver inngrepet (NCSP):

☐☐☐ ☐☐

☐☐☐ ☐☐

Fylles ut ved endt opphold/utskrivelse

Antall liggedøgn i forbindelse med inngrepet

☐☐☐ (dager)

Ved dødsfall under oppholdet, oppgi årsak *(Kun ett kryss)*

☐ Cardiogen årsak

☐ Lungeemboli

☐ Pneumoni

☐ Annen infeksjon

☐ Anafylaksi

☐ Cerebrovaskulær årsak

☐ Blødning

☐ Annet, spesifiser _____

Appendix II

Preoperative patient form from the Norwegian Registry for Spine Surgery

SKJEMA 1A: PASIENTOPPLYSNINGER PREOPERATIVT
(Fylles ut av pasienten før operasjonen)

Spørreskjema for pasienter som skal opereres i ryggen

Nasjonalt Kvalitetsregister for Ryggkirurgi
Degenerativ rygg

Nasjonalt Kvalitetsregister for Ryggkirurgi
E-post: ryggregisteret@unn.no
Hjemmeside: www.ryggregisteret.no

1108 - Versjon 2

Pasientdata (Barkode)

Navn

Fødselsnr. (11 siffer) ☐☐☐☐☐☐ ☐☐☐☐☐

Adresse

E-post
(For bruk ved etterkontroll)

Mobil ☐☐☐ ☐☐ ☐☐☐☐
(For bruk ved etterkontroll)

Formålet med dette spørreskjemaet er å gi leger, sykepleiere og fysioterapeuter bedre forståelse av ryggpasienters plager og gi dem muligheter til å vurdere effekter av behandling. Din utfylling av skjemaet vil og være til stor nytte for å kunne gi et best mulig behandlingstilbud til ryggpasienter i fremtiden.

Spørreskjemaet har fire deler. Første del omhandler ulike sider ved din utdanning og familie samt dine smerter og plager. De neste delene består av tre ulike sett spørsmål for måling av din nåværende helse. Det første av disse (kalt Oswestry-skåre) måler hvordan ryggplagene påvirker dine dagligdagse gjøremål. Det andre (kalt EQ-5D) måler din helserelaterte livskvalitet. Den siste delen er en skala der du skal merke av hvor god eller dårlig din helsetilstand er.

Dato for utfylling ☐☐ ☐☐ ☐☐
Dag Måned År

Røyker du? ☐ Ja ☐ Nei

Høyde og vekt

Høyde ☐,☐☐ (m) Vekt ☐☐☐ (kg)

Familie og barn

1. Sivilstatus (sett kun ett kryss)
- ☐ Gift
- ☐ Samboende
- ☐ Enslig

2. Hvor mange barn har du? ☐☐

Utdanning og yrke

1. Hva er din høyeste fullførte utdanning? (Sett kun ett kryss)

- ☐ Grunnskole 7-10 år, framhaldsskole eller folkehøyskole
- ☐ Yrkesfaglig videregående skole, yrkesskole eller realskole
- ☐ Allmennfaglig videregående skole eller gymnas
- ☐ Høyskole eller universitet (mindre enn 4 år)
- ☐ Høyskole eller universitet (4 år eller mer)

Morsmål

- ☐ Norsk
- ☐ Samisk
- ☐ Annet, angi hvilket

LUNDBLAD MEDIA AS, TROMSØ – O-102117

Hvor sterke smerter har du hatt siste uke?

Hvordan vil du gradere smertene du har hatt i rygg/hofte i løpet av den siste uken? Sett ring rundt ett tall.

0 1 2 3 4 5 6 7 8 9 10

Ingen smerter — Så vondt som det går an å ha

Hvordan vil du gradere de smertene du har hatt i benet (ett eller begge) i løpet av den siste uken? Sett ring rundt ett tall.

0 1 2 3 4 5 6 7 8 9 10

Ingen smerter — Så vondt som det går an å ha

Funksjonsscore (Oswestry)

Disse spørsmålene er utarbeidet for å gi oss informasjon om hvordan dine smerter har påvirket dine muligheter til å klare dagliglivet ditt. Vær snill å besvare spørsmålene ved å sette kryss (kun ett kryss for hvert avsnitt) i de rutene som passer best for deg.

1. Smerte

- ☐ Jeg har ingen smerter for øyeblikket
- ☐ Smertene er veldig svake for øyeblikket
- ☐ Smertene er moderate for øyeblikket
- ☐ Smertene er temmelig sterke for øyeblikket
- ☐ Smertene er veldig sterke for øyeblikket
- ☐ Smertene er de verste jeg kan tenke meg for øyeblikket

2. Personlig stell

- ☐ Jeg kan stelle meg selv på vanlig måte uten at det forårsaker ekstra smerter
- ☐ Jeg kan stelle meg selv på vanlig måte, men det er veldig smertefullt
- ☐ Det er smertefullt å stelle seg selv, og jeg gjør det langsomt og forsiktig
- ☐ Jeg trenger noe hjelp, men klarer det meste av mitt personlige stell
- ☐ Jeg trenger hjelp hver dag til det meste av eget stell
- ☐ Jeg kler ikke på meg, har vanskeligheter med å vaske meg og holder sengen

3. Å løfte

- ☐ Jeg kan løfte tunge ting uten å få mer smerter
- ☐ Jeg kan løfte tunge ting, men får mer smerter
- ☐ Smertene hindrer meg i å løfte tunge ting opp fra gulvet, men jeg greier det hvis det som skal løftes er gunstig plassert, for eksempel på et bord
- ☐ Smertene hindrer meg i å løfte tunge ting, men jeg klarer lette og middels tunge ting, hvis det er gunstig plassert
- ☐ Jeg kan bare løfte noe som er veldig lett
- ☐ Jeg kan ikke løfte eller bære noe i det hele tatt

4. Å gå

- ☐ Smerter hindrer meg ikke i å gå i det hele tatt
- ☐ Smerter hindrer meg i å gå mer enn 1 ½ km
- ☐ Smerter hindrer meg i å gå mer enn ¾ km
- ☐ Smerter hindrer meg i å gå mer enn 100 m
- ☐ Jeg kan bare gå med stokk eller krykker
- ☐ Jeg ligger for det meste i sengen, og jeg må krabbe til toalettet

5. Å sitte

- ☐ Jeg kan sitte så lenge jeg vil i en hvilken som helst stol
- ☐ Jeg kan sitte så lenge jeg vil i min favorittstol
- ☐ Smerter hindrer meg i å sitte i mer enn en time
- ☐ Smerter hindrer meg i å sitte i mer enn en halv time
- ☐ Smerter hindrer meg i å sitte i mer enn ti minutter
- ☐ Smerter hindrer meg i å sitte i det hele tatt

6. Å stå

- ☐ Jeg kan stå så lenge jeg vil uten å få mer smerter
- ☐ Jeg kan stå så lenge jeg vil, men får mer smerter
- ☐ Smerter hindrer meg i å stå i mer enn en time
- ☐ Smerter hindrer meg i å stå i mer enn en halv time
- ☐ Smerter hindrer meg i å stå i mer enn ti minutter
- ☐ Smerter hindrer meg i å stå i det hele tatt

7. Å sove

- ☐ Søvnen min forstyrres aldri av smerter
- ☐ Søvnen min forstyrres av og til av smerter
- ☐ På grunn av smerter får jeg mindre enn seks timers søvn
- ☐ På grunn av smerter får jeg mindre enn fire timers søvn
- ☐ På grunn av smerter får jeg mindre enn to timers søvn
- ☐ Smerter hindrer all søvn

8. Seksualliv

- ☐ Seksuallivet mitt er normalt og forårsaker ikke mer smerter
- ☐ Seksuallivet mitt er normalt, men forårsaker noe mer smerter
- ☐ Seksuallivet mitt er normalt, men svært smertefullt
- ☐ Seksuallivet mitt er svært begrenset av smerter
- ☐ Seksuallivet mitt er nesten borte på grunn av smerter
- ☐ Smerter forhindrer alt seksualliv

9. Sosialt liv (omgang med venner og kjente)

- ☐ Det sosiale livet mitt er normalt og forårsaker ikke mer smerter
- ☐ Det sosiale livet mitt er normalt, men øker graden av smerter
- ☐ Smerter har ingen betydelig innvirkning på mitt sosiale liv, bortsett fra at de begrenser mine mer fysisk aktive sider, som sport osv.
- ☐ Smerter har begrenset mitt sosiale liv, og jeg går ikke så ofte ut
- ☐ Smerter har begrenset mitt sosiale liv til hjemmet
- ☐ På grunn av smerter har jeg ikke noe sosialt liv

10. Å reise

- ☐ Jeg kan reise hvor som helst uten smerter
- ☐ Jeg kan reise hvor som helst, men det gir mer smerter
- ☐ Smertene er ille, men jeg klarer reiser på to timer
- ☐ Smerter begrenser meg til korte reiser på under en time
- ☐ Smerter begrenser meg til korte, nødvendige reiser på under 30 minutter
- ☐ Smerter forhindrer meg fra å reise, unntatt for å få behandling

Beskrivelse av helsetilstand (EQ-5D)

Vis hvilke utsagn som passer best på din helsetilstand i dag ved å sette kun ett kryss i en av rutene for hvert punkt nedenfor.

1. Gange

- ☐ Jeg har ingen problemer med å gå omkring
- ☐ Jeg har litt problemer med å gå omkring
- ☐ Jeg er sengeliggende

2. Personlig stell

- ☐ Jeg har ingen problemer med personlig stell
- ☐ Jeg har litt problemer med å vaske meg eller kle meg
- ☐ Jeg er ute av stand til å vaske meg eller kle meg

3. Vanlige gjøremål (f.eks. arbeid, studier, husarbeid, famile- eller fritidsaktiviteter)

- ☐ Jeg har ingen problemer med å utføre mine vanlige gjøremål
- ☐ Jeg har litt problemer med å utføre mine vanlige gjøremål
- ☐ Jeg er ute av stand til å utføre mine vanlige gjøremål

4. Smerte og ubehag

- ☐ Jeg har hverken smerte eller ubehag
- ☐ Jeg har moderat smerte eller ubehag
- ☐ Jeg har sterk smerte eller ubehag

5. Angst og depresjon

- ☐ Jeg er hverken engstelig eller deprimert
- ☐ Jeg er noe engstelig eller deprimert
- ☐ Jeg er svært engstelig eller deprimert

Smertestillende medisiner

Bruker du smertestillende medisiner på grunn av dine rygg- og/eller beinsmerter?

☐ Ja ☐ Nei

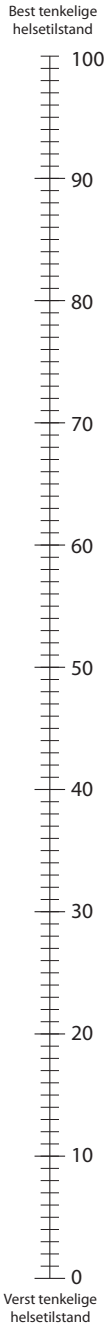
Hvis du har svart ja: Hvor ofte bruker du smertestillende medisiner? (Sett kun ett kryss)

- ☐ Sjeldnere enn hver måned
- ☐ Hver måned
- ☐ Hver uke
- ☐ Daglig
- ☐ Flere ganger daglig

Helsetilstand

For at du skal kunne vise oss hvor god eller dårlig din helsetilstand er, har vi laget en skala (nesten som et termometer), hvor den beste helsetilstanden du kan tenke deg er markert med 100 og den dårligste med 0.

Vi ber om at du viser din helsetilstand ved å trekke ei linje fra boksen nedenfor til det punkt på skalaen som passer best med din helsetilstand.



Nåværende helsetilstand

Symptomvarighet

Varighet av nåværende rygg-/hoftesmerter(sett kun ett kryss):

- ☐ Jeg har ingen rygg-/hoftesmerter
- ☐ Mindre enn 3 måneder
- ☐ 3 til 12 måneder
- ☐ 1 til 2 år
- ☐ Mer enn 2 år

Varighet av nåværende utstrålende smerter:

- ☐ Jeg har ingen utstrålende smerter
- ☐ Mindre enn 3 måneder
- ☐ 3 til 12 måneder
- ☐ 1 til 2 år
- ☐ Mer enn 2 år

Varighet sykemelding/attføring/ rehabilitering pga aktuelle plager ☐☐☐ (uker)

Arbeidsstatus

- ☐ I arbeid
- ☐ Hjemmeværende, ulønnet
- ☐ Student/skoleelev
- ☐ Alderspensjonist
- ☐ Arbeidsledig
- ☐ Sykemeldt
- ☐ Aktivt sykemeldt
- ☐ Delvis sykemeldt
- ---------- % sykemeldt
- ☐ Attføring/rehabilitering
- ☐ Uføretrygdet
- evt --------- % uføretrygdet

Har du søkt om uføretrygd?

(Sett kun ett kryss)

- ☐ Ja
- ☐ Nei
- ☐ Planlegger å søke
- ☐ Er allerede innvilget

Har du søkt om erstatning fra forsikringsselskap eller folketrygden (eventuelt yrkesskadeerstatning)?

(Sett kun ett kryss)

- ☐ Ja
- ☐ Nei
- ☐ Planlegger å søke
- ☐ Er allerede innvilget

Appendix III

Postoperative patient form from the Norwegian Registry for Spine Surgery

Nasjonalt Kvalitetsregister for Ryggkirurgi
Degenerativ rygg

Deres ref: Vår ref: Dato:

Kontroll 3 og 12 måneder etter ryggoperasjon

Hei
Da du ble operert i ryggen, ble det registrert en del opplysninger knyttet til rygglidelsen din og disse er lagret avidentifisert ved Nasjonalt Kvalitetsregister for Ryggkirurgi (NKR).
For å vurdere kvaliteten på den kirurgiske behandlingen som tilbys ryggpasienter i Norge i dag, er det viktig å innhente opplysninger etter operasjonen. NKR har ansvaret for 3 og 12 måneders kontroll av samtlige pasienter som opereres i ryggen i Norge. De opplysninger som samles inn vil gjøres tilgjengelig for den avdeling som behandlet deg, slik at de kan få kunnskap om hvordan det har gått med deg. Dette vil bidra til at den enkelte avdeling kan kvalitetssikre egen virksomhet.

Du har nå fått tilsendt et skjema som vi håper du har anledning til å fylle ut. Ved å returnere skjemaet samtykker du samtidig med at data gjøres tilgjengelig for den avdeling som har behandlet deg.
Skjemaet skal skannes (leses inn av en datamaskin) direkte inn i registeret og vi ber derfor om at du passer på å fylle det ut på korrekt måte. Du *skal ikke* skrive navn eller personnummer på skjemaet da dette er kodet og skal returneres alene i den vedlagte frankerte konvolutten.

Hvis du trenger et nytt skjema eller har spørsmål kan du sende en e-post eller et vanlig brev til NKR. For generell informasjon, kan du også besøke vår internettside. Se for øvrig på baksiden av arket for informasjon om hvordan skjemaet bør fylles ut.

Med vennlig hilsen
Nasjonalt kvalitetsregister for ryggkirurgi

Tore Solberg

Registerleder

Nasjonalt kvalitetsregister for ryggkirurgi
Postboks 20
Universitetssykehuset Nord-Norge
9038 Tromsø

Besøksadresse
Plan 5, fløyC0
Sykehusveien 38
9038 Tromsø

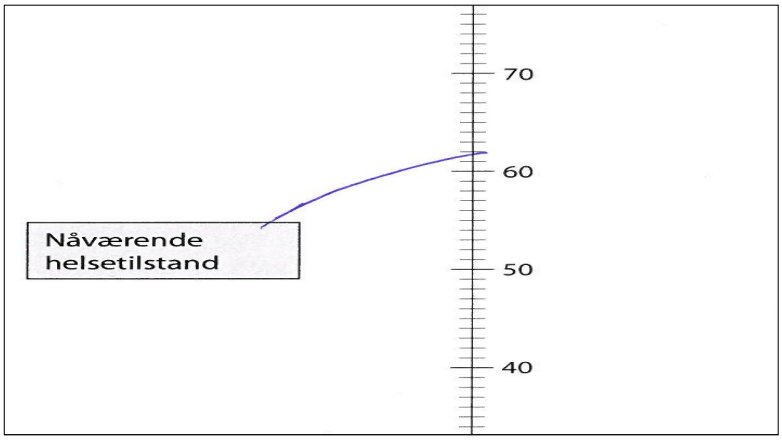
Kontakt
Telefon: 776 69 015
e-post: ryggregisteret@unn.no
webside: www.ryggregisteret.no

Utfylling av skjemaet

Kryss av slil ☒

og ikke slik ☒

For markering av helsetilstand er det viktig at det markeres på tvers av skalaen slik at verdien skannes korrekt.



Figur: Utsnitt av side fire av svarskjemaet som viser eksempel på korrekt markering.

Med «friskmeldt dato» er vi interessert i datoen da du kom tilbake i arbeid, selv om du på det tidspunktet var delvis sykemeldt. På samme måte gjelder «varighet av sykemelding» den perioden du var 100 prosent sykemeldt etter operasjonen.

For «Arbeidsstatus» angis den arbeidsstatus du har når du fyller ut skjemaet.

Pas. id

SKJEMA B1

Nasjonalt Kvalitetsregister for Ryggkirurgi
Degenerativ rygg

Nasjonalt Kvalitetsregister for Ryggkirurgi

Senter for Klinisk Dokumentasjon
og Evaluering - Helse Nord RHF

E-post: ryggregisteret@unn.no
Hjemmeside: www.ryggregisteret.no

Spørreskjema for pasienter 3 måneder etter ryggoperasjon

Formålet med dette spørreskjemaet er å gi leger, sykepleiere og fysioterapeuter bedre forståelse av ryggpasienters plager og å vurdere effekter av behandling. Din utfylling av skjemaet vil være til stor nytte for å kunne gi et best mulig behandlingstilbud til ryggpasienter i fremtiden.

Spørreskjemaet har fem deler. Første del omhandler dine smerter og plager. De neste delene består av tre ulike sett spørsmål for måling av din nåværende helse. Det første av disse (kalt Oswestry-skåre) måler hvordan ryggplagene påvirker dine dagligdagse gjøremål. Det andre (kalt EQ-5D) måler din helserelaterte livskvalitet, mens den neste er en skala der du skal merke av hvor god eller dårlig din helsetilstand er.

Vi ønsker også informasjon om eventuelle komplikasjoner som kan knyttes til inngrepet, samt trygd- og arbeidsstatus.

Dato for utfylling ☐☐.☐☐.☐☐
Dag Måned År

Hvilken nytte mener du at du har hatt av operasjon?

(Sett *kun ett* kryss)

- ☐ Jeg er helt bra
- ☐ Jeg er mye bedre
- ☐ Jeg er litt bedre
- ☐ Ingen forandring
- ☐ Jeg er litt verre
- ☐ Jeg er mye verre
- ☐ Jeg er verre enn noen gang før

Hvor fornøyd er du med behandlingen du har fått på sykehuset?

(Sett *kun ett* kryss)

- ☐ Fornøyd
- ☐ Litt fornøyd
- ☐ Hverken fornøyd eller misfornøyd
- ☐ Litt misfornøyd
- ☐ Misfornøyd

Hvor sterke smerter har du hatt siste uke?

Hvordan vil du gradere smertene du har hatt i rygg/hofte i løpet av den siste uken? Sett kryss ved ett tall.

0	1	2	3	4	5	6	7	8	9	10
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Ingen smerter — Så vondt som det går an å ha

Hvordan vil du gradere smertene du har hatt i benet (ett eller begge) i løpet av den siste uken? Sett kryss ved ett tall.

0	1	2	3	4	5	6	7	8	9	10
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Ingen smerter — Så vondt som det går an å ha

Funksjonsscore (Oswestry)

Pas. id

Disse spørsmålene er utarbeidet for å gi oss informasjon om hvordan dine smerter har påvirket dine muligheter til å klare dagliglivet ditt. Vær så snill å besvare spørsmålene ved å sette kryss (*kun ett* kryss for hvert avsnitt) i de rutene som passer best for deg.

1. Smerte

- ☐ Jeg har ingen smerter for øyeblikket
- ☐ Smertene er veldig svake for øyeblikket
- ☐ Smertene er moderate for øyeblikket
- ☐ Smertene er temmelig sterke for øyeblikket
- ☐ Smertene er veldig sterke for øyeblikket
- ☐ Smertene er det verste jeg kan tenke meg for øyeblikket

2. Personlig stell

- ☐ Jeg kan stelle meg selv på valig måte uten at det forårsaker ekstra smerter
- ☐ Jeg kan stelle meg selv på vanlig måte, men det er veldig smertefullt
- ☐ Det er smertefullt å stelle seg selv, og jeg gjør det langsomt og forsiktig
- ☐ Jeg trenger noe hjelp, men klarer det meste av mitt personlige stell
- ☐ Jeg trenger hjelp hver dag til det meste av eget stell
- ☐ Jeg kler ikke på meg, har vanskeligheter med å vaske meg og holder sengen

3. Å løfte

- ☐ Jeg kan løfte tunge ting uten å få mer smerter
- ☐ Jeg kan løfte tunge ting, men får smerter
- ☐ Smertene hindrer meg i å løfte tunge ting opp fra gulvet, men jeg greier det hvis det som skal løftes er gunstig plassert, for eksempel på et bord
- ☐ Smertene hindrer meg i å løfte tunge ting, men jeg klarer lette og middels tunge ting, hvis det er gunstig plassert
- ☐ Jeg kan bare løfte noe som er veldig lett
- ☐ Jeg kan ikke løfte eller bære noe i det hele tatt

4. Å gå

- ☐ Smerter hindrer meg ikke i å gå i det hele tatt
- ☐ Smerter hindrer meg i å gå mer enn 1 ½ km
- ☐ Smerter hindrer meg i å gå mer enn ¾ km
- ☐ Smeter hindrer meg i å gå mer enn 100 m
- ☐ Jeg kan bare gå med stokk eller krykker
- ☐ Jeg ligger for det meste i sengen, og jeg må krabbe til toalettet

5. Å sitte

- ☐ Jeg kan sitte så lenge jeg vil i en hvilken som helst stol
- ☐ Jeg kan sitte så lenge jeg vil i min favorittstol
- ☐ Smerter hindrer meg i å sitte mer enn en time
- ☐ Smerter hindrer meg i å sitte mer enn en halv time
- ☐ Smerter hindrer meg i å sitte mer enn ti minutter
- ☐ Smerter hindrer meg i å sitte i det hele tatt

6. Å stå

- ☐ Jeg kan stå så lenge jeg vil uten å få mer smerter
- ☐ Jeg kan stå så lenge jeg vil, men får mer smerter
- ☐ Smerter hindrer meg i å stå mer enn en time
- ☐ Smerter hindrer meg i å stå mer enn en halv time
- ☐ Smerter hindrer meg i å stå mer enn ti minutter
- ☐ Smerter hindrer meg i å stå i det hele tatt

7. Å sove

- ☐ Søvnen min forstyrres aldri av smerter
- ☐ Søvnen min forstyrres av og til av smerter
- ☐ På grunn av smerter får jeg mindre enn seks timers søvn
- ☐ På grunn av smerter får jeg mindre en fire timers søvn
- ☐ På grunn av smerter får jeg mindre enn to timers søvn
- ☐ Smerter hindre all søvn

8. Seksualliv

- ☐ Seksuallivet mitt er normalt og forårsaker ikke mer smerter
- ☐ Seksuallivet mitt er normalt, men forårsaker noe mer smerter
- ☐ Seksuallivet mitt er normalt, men svært smertefult
- ☐ Seksuallivet mitt er svært begrenset av smerter
- ☐ Seksuallivet mitt er nesten borte på grunn av smerter
- ☐ Smerter forhindrer alt seksualliv

Pas. id

9. Sosialt liv (omgang med venner og kjente)

- ☐ Det sosiale livet mitt er normalt og forårsaker ikke mer smerter
- ☐ Det sosiale livet mitt er normalt, men øker graden av smerter
- ☐ Smerter har ingen betydelig innvirkning på mitt sosiale liv, bortsett fra at de begrenser mine mer fysiske aktive sider, som sport osv.
- ☐ Smerter har begrenset mitt sosiale liv, og jeg går ikke så ofte ut
- ☐ Smerter har begrenset mitt sosiale liv til hjemmet
- ☐ På grunn av smerter har jeg ikke noe sosialt liv

10. Å reise

- ☐ Jeg kan reise hvor som helst uten smerter
- ☐ Jeg kan reise hvor som helst, men det gir mer smerter
- ☐ Smertene er ille, men jeg klarer reiser på to timer
- ☐ Smerter begrenser meg til korte reiser på under en time
- ☐ Smerter begrenser meg til korte, nødvendige reiser på under 30 minutter
- ☐ Smerter forhindrer meg fra å reise, unntatt for å få behandling

Beskrivelse av helsetilstand (EQ-5D)

Vis hvilke utsagn som passer best på din helsetilstand i dag ved å sette *kun ett* kryss i en av rutene for hvert punkt nedenfor.

1. Gange

- ☐ Jeg har ingen problemer med å gå omkring
- ☐ Jeg har litt problemer med å gå omkring
- ☐ Jeg er sengeliggende

2. Personlig stell

- ☐ Jeg har ingen problemer med personlig stell
- ☐ Jeg har litt problemer med å vaske meg eller kle meg
- ☐ Jeg er ute av stand til å vaske meg eller kle meg

3. Vanlige gjøremål

- ☐ Jeg har ingen problemer med å utføre mine vanlige gjøremål
- ☐ Jeg har litt problemer med å utføre mine vanlige gjøremål
- ☐ Jeg er ute av stand til å utføre mine vanlige gjøremål

4. Smerte og ubehag

- ☐ Jeg har hverken smerte eller ubehag
- ☐ Jeg har moderat smerte eller ubehag
- ☐ Jeg har sterk smerte eller ubehag

5. Angst og depresjon

- ☐ Jeg er hverken engstelig eller deprimert
- ☐ Jeg er noe engstelig eller deprimert
- ☐ Jeg er svært engstelig eller deprimert

Smertestillende medisiner

Bruker du smertestillende medisiner på grunn av dine rygg- og/eller beinsmerter?

☐ Ja ☐ Nei

Hvis du har svart ja: Hvor ofte bruker du smertestillende medisiner? (Sett *kun ett* kryss)

- ☐ Sjeldnere enn hver måned
- ☐ Hver måned
- ☐ Hver uke
- ☐ Daglig
- ☐ Flere ganger daglig

Arbeidsstatus

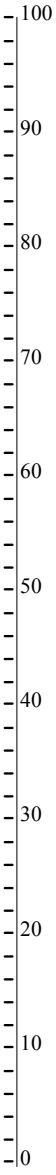
- ☐ I arbeid
- ☐ Hjemmeværende (ulønnet)
- ☐ Student/skoleelev
- ☐ Alderspensjonist
- ☐ Arbedisledig
- ☐ Sykemeldt
- ☐ Aktiv sykemeldt
- ☐ Delvis sykemeldt
- ☐ ☐ % sykemeldt
- ☐ Attføring/rehabilitering
- ☐ Uføretrygdet
- evt. ☐ ☐ % uføretrygdet

Helsetilstand

For at du skal kunne vise oss hvor god eller dårlig din helsetilstand er, har vi laget en skala (nesten som et termometer), hvor den beste helsetilstanden du kan tenke deg er markert med 100 og den dårligste med 0.

Vi ber om at du viser din helsetilstand ved å trekke ei linje fra boksen nedenfor til det punkt på skalaen som passer best med din helsetilstand.

Best tenkelige helsetilstand



Verst tenkelige helsetilstand

Nåværende helsetilstand

Pas. id

Friskmeldt? (tilbake i arbeid, helt eller delvis)

Hvis ja, angi dato ☐☐ . ☐☐ . ☐☐

Dag Måned År

Varighet av sykemelding etter operasjon ☐☐☐ (uker)

Komplikasjoner til inngrepet? (Sett evt. flere kryss)

- ☐ Oppsto det uventet blødning som medførte blod-overføring eller ny operasjon?
- ☐ Ble du behandlet med antibiotika for en urinveisinfeksjon i løpet av de nærmeste 4 ukene etter operasjonen?
- ☐ Ble du behandlet med antibiotika for en lungebetennelse i løpet av de nærmeste 4 ukene etter operasjonen?
- ☐ Har du i løpet av 3 måneder etter operasjonen, fått diagnosen "dyp vene trombose" (blodpropp i benet) og vært behandlet for dette?
- ☐ Har du i løpet av 3 måneder etter operasjonen, fått diagnosen lungeemboli (blodpropp i lungen) og blitt behandlet for dette?
- ☐ Ble du behandlet med antibiotika for en overfladisk infeksjon i operasjonssåret i løpet av de første 4 ukene etter operasjonen?
- ☐ Har du blitt eller blir du behandlet i over 6 uker med antibiotika for dyp infeksjon i operasjonssåret?
- ☐ Har du opplevd nytilkommet svakhet/lammelse i fot eller ben som kan tilskrives operasjonen?
- ☐ Har du som følge av operasjonen utviklet problemer med ufrivillig vannlating eller avføring?

Har du søkt om uføretrygd?

(Sett *kun ett* kryss)

- ☐ Ja
- ☐ Nei
- ☐ Planlegger å søke
- ☐ Er allerede innvilget

Har du søkt om erstatning fra forsikringsselskap eller folketrygden (eventuelt yrkesskadeerstatning)?

(Sett *kun ett* kryss)

- ☐ Ja
- ☐ Nei
- ☐ Planlegger å søke
- ☐ Er allerede innvilget

Paper II

II

Austevoll et al. BMC Musculoskeletal Disorders (2019) 20:31
https://doi.org/10.1186/s12891-018-2386-y

BMC Musculoskeletal Disorders

RESEARCH ARTICLE Open Access

Follow-up score, change score or percentage change score for determining clinical important outcome following surgery? An observational study from the Norwegian registry for Spine surgery evaluating patient reported outcome measures in lumbar spinal stenosis and lumbar degenerative spondylolisthesis

Ivar Magne Austevoll[1,2,3*], Rolf Gjestad[4], Margreth Grotle[7,10], Tore Solberg[3,6], Jens Ivar Brox[3,5], Erland Hermansen[1,2,8], Frode Rekeland[1], Kari Indrekvam[1,2], Kjersti Storheim[7] and Christian Hellum[3,9]

Abstract

Background: Assessment of outcomes for spinal surgeries is challenging, and an ideal measurement that reflects all aspects of importance for the patients does not exist. Oswestry Disability Index (ODI), EuroQol (EQ-5D) and Numeric Rating Scales (NRS) for leg pain and for back pain are commonly used patients reported outcome measurements (PROMs). Reporting the proportion of individuals with an outcome of clinical importance is recommended. Knowledge of the ability of PROMs to identify clearly improved patients is essential. The purpose of this study was to search cut-off criteria for PROMs that best reflect an improvement considered by the patients to be of clinical importance.

Methods: The Global Perceived Effect scale was utilized to evaluate a clinically important outcome 12 months after surgery. The cut-offs for the PROMs that most accurately distinguish those who reported 'completely recovered' or 'much improved' from those who reported 'slightly improved', unchanged', 'slightly worse', 'much worse', or 'worse than ever' were estimated. For each PROM, we evaluated three candidate response parameters: the (raw) follow-up score, the (numerical) change score, and the percentage change score.

(Continued on next page)

* Correspondence: imau@helse-bergen.no
[1]Kysthospitalet in Hagevik, Orthopedic Clinic, Haukeland, University Hospital, Hagaviksbakken 25, 5217 Hagevik, Bergen, Norway
[2]Department of Clinical Medicine, University of Bergen, Christies gate 6, 5007 Bergen, Bergen, Norway
Full list of author information is available at the end of the article

(Continued from previous page)
Results: We analysed 3859 patients with Lumbar Spinal Stenosis [(LSS); mean age 66; female gender 50%] and 617 patients with Lumbar Degenerative Spondylolisthesis [(LDS); mean age 67; 72% female gender]. The accuracy of identifying 'completely recovered' and 'much better' patients was generally high, but lower for EQ-5D than for the other PROMs. For all PROMs the accuracy was lower for the change score than for the follow-up score and the percentage change score, especially among patients with low and high PROM scores at baseline.
The optimal threshold for a clinically important outcome was ≤24 for ODI, ≥0.69 for EQ-5D, ≤3 for NRS leg pain, and ≤ 4 for NRS back pain, and, for the percentage change score, ≥30% for ODI, ≥40% for NRS leg pain, and ≥ 33% for NRS back pain. The estimated cut-offs were similar for LSS and for LDS.

Conclusion: For estimating a 'success' rate assessed by a PROM, we recommend using the follow-up score or the percentage change score. These scores reflected a clinically important outcome better than the change score.

Keywords: Lumbar spinal stenosis (LSS), Lumbar degenerative spondylolisthesis (LDS), Patient reported outcome measures (PROMs), Oswestry disability index (ODI), Leg pain, Back pain, Success criteria, Minimal clinically important difference (MCID)

Background

The success of surgical treatment of spinal degenerative disorders is basically determined by reduction of pain and improvement of function. In clinical studies, treatment effects are most commonly assessed by patient reported outcome measures (PROMs) [1–5]. Widely used PROMs for evaluating outcomes after surgery for lumbar spinal stenosis (LSS) with and without degenerative spondylolisthesis (LDS) are the Oswestry Disability Index (ODI) [1, 2, 4, 5], the numeric rating scales (NRS) for leg- and back pain [1, 6–9], and a generic measure of health-related quality of life such as the EQ-5D [8–10]. However, these outcome measures do not necessarily cover all areas of interest to the patient. Even though items like personal care and walking distance are addressed by the ODI, more specific disabilities such as problems with personal hygiene, posture imbalance and slow walking speed may not be detected.

Due to the frequent use of PROMS, the statistical application and the interpretation of the clinical importance of the outcomes should be evaluated [11]. The clinical effect of a treatment is usually only presented as the mean change from baseline to follow-up [1, 4, 5]. However, a statistically significant mean group difference does not necessarily provide meaningful clinical information when comparing two methods. A large improvement in a few individuals in one of the treatment groups can dramatically enhance the mean change of the group, even if the majority had no improvement or even a slight worsening of their complaints [11, 12]. Rather than discussing the relevance of mean changes alone, the proportion of individuals with a clinically relevant reduction in pain and disability (i.e., a 'success' rate) can be employed as a comprehendible metric for patients and physicians to use in clinical decision-making [11–13].

To calculate 'success' rates assessed by PROMs, we need criteria that reflect the patients' perceptions of important benefits following operations [11–13]. The Minimal clinical important difference (MCID) was the first metric developed for this purpose [14, 15]. Minimal important changes (MIC) [16], a substantial clinical benefit [17] and a satisfactory symptom state [18, 19] are other metrics developed to distinguish whether patients have achieved a clinically important effect of treatment or not. Several authors have pointed out the great variability and diversity of such thresholds [12, 20, 21], which may be caused by the heterogeneity in the populations studied [22]. The objective of the present study was to estimate the thresholds for ODI, EQ-5D and NRS leg- and back pain that best identify the patients who perceived a clinically important outcome following surgery for LSS and LDS. Receiver Operating Characteristic (ROC) analyses were evaluated to explore how accurately 'success' assessed by a single question on the Global Perceived Effect (GPE) scale [23] would be reflected in the PROMs. Despite limited evidence for the validity of the GPE scale [12, 24], it is widely used [17, 18, 25–28] and recommended [12, 29] in such analyses. For each PROM three alternative response parameters were evaluated: the follow-up score, the change score and the percentage change score. LSS and LDS were analysed separately.

Methods

Study population

Data were obtained from the Norwegian Registry for Spine Surgery (NORSpine). NORSpine is a government-funded, comprehensive, clinical registry for quality control and research. The registry receives no funding from the industry. Informed consent is obtained from all patients. The patient form consists of PROMs completed before surgery (baseline) and at 3- and 12-month follow-up. During the hospital stay, data concerning diagnosis, treatment and comorbidity were recorded by the surgeons on a standard form.

Inclusion criteria: (1) Patients registered in NORSpine in the period 2007–2013; (2) Patients assessed by the surgeon to have spinal stenosis with or without degenerative spondylolisthesis; (3) Patients operated with a decompression procedure or with decompression in combination with posterior fusion. Patients with a former operation at index level were excluded.

Patient reported outcome measures (PROMs)

1. The Oswestry Disability Index (ODI) V.2.0 [30, 31] has been translated and validated for application among Norwegian patients [32]. It is found to be an appropriate instrument for assessing treatment outcome in patients with spinal stenosis with and without a degenerative spondylolisthesis [33]. It is a self-reported instrument comprising 10 questions about pain related disability in activities of daily life. The sum score ranges from 0 (no disability) to 100 points (bedridden).
2. The EuroQol (EQ-5D) [34] is a generic measurement for assessing health-related quality of life. It evaluates mobility, self-care, usual activity, pain/discomfort and anxiety/discomfort. For each component the patients can choose between three answers; none, mild to moderate, and severe. This gives $3^5 = 243$ possible sets of answers, and each unique combination corresponds to a value between − 0.59 and 1.0, where 1.0 represents perfect health.
3. Numeric Rating Scale (NRS) for back- and leg pain assesses self-reported pain level in the last week ranging from 0 (no pain) to 10 (worst conceivable pain) [30].
4. Global Perceived Effect (GPE) is a single question measuring patient-rated assessment of treatment outcome [23]. The patient may choose between seven response alternatives: 'completely recovered', 'much improved', 'slightly improved', unchanged', 'slightly worse', 'much worse', and 'worse than ever'.

Definition of 'success' according to GPE scale

Patients who rated themselves as 'completely recovered' or 'much improved' on the GPE scale (the anchor) at 12-month follow-up were considered to have gained a clinically important outcome following the surgery ('success'), whereas patients that replied 'slightly improved', unchanged', 'slightly worse', 'much worse', and 'worse than ever' were considered to have not benefited from their operation ('non-success') [12, 17, 18, 35].

Statistics

For each PROM three alternative response parameters were evaluated: 1) the (raw) follow-up score; 2) the (numerical) change score (i.e., the absolute change from baseline to follow-up); 3) the percentage change score (i.e., the change score as a percentage of the baseline score). In order to evaluate whether 'success' on the GPE scale (the anchor) would be reflected in a PROM, Receiver Operating Characteristics (ROC) [36] curve analyses were performed. Analogue to a diagnostic test, the sensitivity refers to the probability of detecting a condition. In the present setting it refers to the probability of correctly classifying an individual replying 'completely recovered' or 'much improved' (GPE) as a 'success' when assessed by a PROM. Correspondingly, the specificity refers to the probability of correctly classifying a patient reporting less than 'completely recovered' or 'much improved' as a 'non-success'. Depending on the level of a cut-off, the sensitivity and specificity will vary. A ROC curve was made by plotting the sensitivity against 1 minus the specificity, for all possible cut-off values for 'success'. The cut-off that maximized the proportion of correctly classified patients according to the anchor was chosen as the threshold for 'success'. If more than one cut-off value maximized the percentage of correct classification we prioritized the relation between sensitivity and specificity that balanced the ratio between false negatives and false positives [13, 36]. If possible, still with the assumption of maximum correct classification and a balanced false negatives/false positives ratio, we intended to choose common cut-off values for LSS and LDS.

For all PROMs, the area under the ROC curves (AUC) with 95% confidence interval (CI) was estimated for the alternative response parameters. The AUC describes the test's accuracy in correctly classifying a case according to the anchor – the larger the AUC, the greater the accuracy of the test. The AUC is classified as 'excellent' from 1.0 to 0.90, 'good' from 0.90 to 0.80, 'fair' from 0.80 to 0.70, 'poor' from 0.70 to 0.60, and 'failed' from 0.60 to 0.50 [37].

Since cut-off values for clinical improvement tend to be dependent on the baseline level of a measurement [26], a sensitivity analysis was performed. For each of the estimated cut-off values the percentage of correct classification was calculated for patient groups with low, medium, and high baseline scores respectively. The split values were chosen to ensure equal proportions of patients in each group (tertiary split). For ODI the split values between groups were 32 and 46 points, for EQ-5D they were 0.1 and 0.6. For NRS leg- and back pain the low baseline group had scores of 1–5, the medium baseline group, 6–7 and the high baseline group, 8–10.

Baseline characteristics and PROMs were reported as means and standard deviations of continuous variables and as percentages of categorical variables. The mean 12-month follow-up scores and the mean changes from baseline to follow-up were assessed against the

categories of the GPE scale. To evaluate the predictive validity of PROMs, correlations between the response on the GPE scale and the PROMs were analysed using the Spearman rank coefficient.

In a previous study from NORSpine, no differences in outcome were found when comparing compliers and non-compliers at follow-up [38]. We therefore assumed that missing data were comparable to data from those who answered, and performed the analysis based on the listwise deletion method [39].

The statistical analyses were performed using the Statistical Package for Social Sciences (SPSS) version 23.0 and by Stata version 14.0.

Results

Of 5238 eligible patients from 32 clinics, 4476 met the inclusion criteria. Of these, 617 had a degenerative spondylolisthesis. At 12-month follow-up, 3093 with LSS and 517 with LDS had answered the questionnaire, a follow-up rate of 81% (Fig. 1).

The mean age (±SD) was 66 (±11) years for LSS and 67 (±10) years for LDS, and the percentage of females was 50 and 72%, respectively. Further patient demographics and surgical data are presented in Table 1.

The mean (±SD) ODI changed from 40 (±15) at baseline (Table 1) to 23 (±18) at 12-month follow-up (Table 2) for LSS, and from 41 (±15) to 22 (±18) for LDS. Respectively for LSS and LDS, EQ-5D changed from 0.37 (±0.32) to 0.64 and from 0.34 (±32) to 0.67, NRS leg pain from 6.6 (±2.2) to 3.5 (±3.0) and 6.7 (±2.2) to 3.2 (±2.9) and NRS back pain from 6.4 (± 2.2) to 3.8 (±2.8) and 6.9 (±2.2) to 3.6 (±2.8). On the GPE-scale 58 and 65% replied that they were 'completely recovered' or 'much improved' (LSS and LDS, respectively). The Spearman rank coefficients between the GPE ratings and the 12-month follow-up measures were 0.77 and 0.78 for ODI, 0.73 and 0.78 for EQ-5D, 0.72 and 0.68 for NRS leg pain and 0.76 and 0.78 for NRS back pain, respectively for LSS and LDS; $p < 0.001$ for all correlations (Table 2).

Figures 2, 3, 4 and 5 show the ROC curves for each of the response parameters for ODI, EQ-5D and NRS leg- and back pain. For all PROMs the graphs for the follow-up scores and the percentage change scores illustrate larger areas under the curves (AUC) than for the (numerical) change scores, indicating that the change scores were less accurate in matching 'successes' from the GPE scale.

In general, the computed AUC showed good or excellent test accuracy (AUC from 0.82 to 0.92) for the three alternative scores for all measurements except for the EQ-5D's change score [AUC = 0.76 (fair accuracy)]. However, for all PROMs, the AUC was generally lower for the change scores than for the follow-up scores and the percentage change scores, and in most cases this

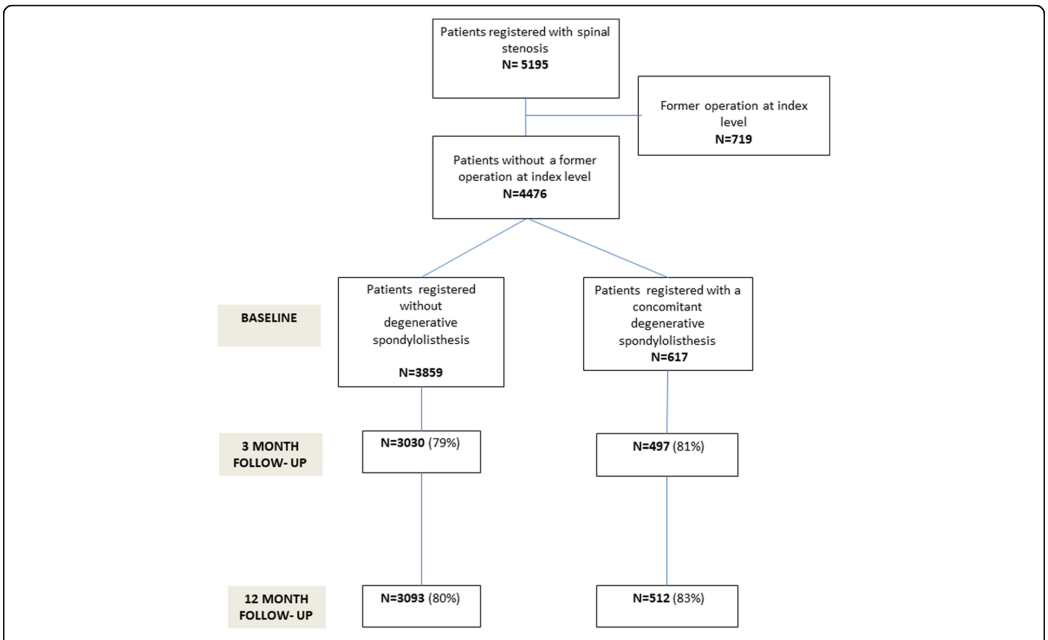


Fig. 1 Flow chart for patients registered with spinal stenosis in NORSpine in the period 2007–2013

Table 1 Patient demographics and surgical data for patients operated for spinal stenosis and for degenerative spondylolisthesis

	Spinal stenosis		Degenerative spondylolisthesis	
	N		N	
Age; Yr ± SD	3858	66 ± 11	617	67 ± 10
Female, no (%)	3859	1919 (50%)	617	444 (72%)
ASA level (1–4); Mean ± SD	3759	2.0 ± 0.6	608	2.0 ± 0.5
ASA level 1, no (%)		681 (18%)	82 (13%)	
ASA level 2, no (%)		2349 (61%)	429 (71%)	
ASA level 3, no (%)		753 (19%)	97 (16%)	
ASA level 4, no (%)		12(0.3%)	0	
Body Mass Index; Mean (SD)	3547	27 ± 4	560	27.0 ± 5
Smokers, no (%)	3808	877 (23%)	609	115 (19%)
Laminectomy, no (%)	3859	1024 (27%)	617	239 (39%)
Midline preserving decompression, no (%)	3859	2835 (73%)	617	378 (61%)
Fusion, no (%)	3859	214 (6%)	617	297 (48%)
ODI; Mean (SD)	3837	40 ± 15	617	41 ± 15
EQ-5D; Mean (SD)	3535	0.37 ± 0.32	564	0.34 ± 0.32
NRS leg pain; Mean (SD)	3559	6.6 ± 2.2	569	6.7 ± 2.2
NRS back pain; Mean (SD)	3597	6.4 ± 2.2	573	6.9 ± 2.1

N number of patient with data for the evaluated parameter

difference was statistically significant (i.e., without overlap of the 95% CI (Table 3). For LSS, the AUC for ODI was 0.90 (95% CI 0.89–0.91) for the follow-up score, 0.86 (95% CI 0.84–-0.87) for the numerical change score and 0.91(95% CI 0.90–0.92) for the percentage change score, and, respectively, 0.92 (95% CI 0.89–0.94), 0.86 (95% CI 0.82–0.89) and 0.92 (95% CI 0.90–0.94) for LDS. The AUCs for all PROMs are listed in Table 3.

Except for the NRS back pain change score, the cut-off values for a clinically important outcome were identical for LSS and LDS (Table 3). The following cut-offs were estimated, with the correct classification rates (for LSS and LDS respectively) listed in parentheses:

ODI

follow-up score ≤ 24 points (82%, 85%), change score ≥ 13 points (78%, 80%), percentage change ≥30% (83%, 85%).

EQ-5D

follow-up score ≥ 0.692 points (78%, 84%), change score ≥ 0.105 points (73%, 76%). Because the EQ-5D questionnaire values ranged from − 0.6 to 1.0 on a categorical scale, it was not possible to find a mathematically adequate method to evaluate the percentage change score.

NRS leg pain

follow-up score ≤ 3points (81%, 79%), change score ≥ 3 points (77%, 76%), percentage change ≥40% (81%, 78%).

NRS back pain

follow-up score ≤ 4 points (82%, 83%), change score ≥ 2 points for LSS (75%) and ≥ 3 points for LDS (79%), percentage change ≥33% (80%, 82%).

The sensitivity and specificity for each cut-off value are listed in Table 4.

In the sensitivity analysis, a ≤ 24 point cut-off for the ODI follow-up score gave 80% correctly classified patients in low, 85% in medium and 80% in high baseline levels for LSS, respectively 87, 85 and 84% for LDS. The corresponding rates for the ODI change score were 72, 84 and 78% for LSS, and 77, 86 and 75% for LDS, and, for the percentage change score, 83, 85 and 80% for LSS, and 88, 85 and 82% for LDS. Table 4 shows that also for the other PROMs, the change scores for patients with low and high baseline values were the least accurate in matching 'successes' from the GPE scale.

Discussion

We evaluated how accurately four frequently used PROMs would reflect the patients' global assessment of being completely recovered or much better at 12-month follow-up. All outcome scores for the PROMs were highly correlated to the GPE score, indicating good predictive validity. The accuracy for correct classification of a GPE 'success' as a 'success' assessed by the PROMs was generally high, however, lower for the (numerical) change score than for the follow-up score and the percentage change score, especially among patients with low and high preoperative PROM values. All estimated

Table 2 Follow-up scores and the change scores for PROMs according to the GPE-scale

		Spinal stenosis								Degenerative spondylolisthesis							
		N	(%)	1 year Follow-up Mean	(SD)	Spearman's rho	Change score Mean	(SD)	Spearman's rho	N	(%)	1 year Follow- up Mean	(SD)	Spearman's rho	Change score Mean	(SD)	Spear- man's rho
	All	3060		23	(18)	0.77 *	16	(18)	0.66*	509		22	(18)	0.78*	19	(17)	0.64*
O	Compl.recovered	599	(20%)	4	(9)		32	(16)		117	(23%)	4	(7)		33	(15)	
D	Much improved	1176	(38%)	17	(12)		21	(15)		213	(42%)	17	(13)		23	(14)	
I	Slightly improved	658	(21%)	32	(12)		9	(13)		105	(21%)	36	(13)		9	(12)	
	Unchanged	283	(9%)	38	(13)		0	(10)		33	(6%)	38	(14)		5	(13)	
	Slightly worse	181	(6%)	42	(13)		0	(12)		21	(4%)	41	(13)		3	(13)	
	Much worse	117	(4%)	49	(12)		−3	(12)		11	(2%)	51	(11)		−8	(13)	
	Worse than ever	46	(2%)	59	(15)		−11	(12)		9	(2%)	57	(17)		−7	(15)	
	Missing	799								108							
	All	2464		0.64	(0.31)	0.73*	0.25	0.36	0.50*	419		0.67	(0.30)	0.78*	0.32	(0.34)	0.48*
E	Compl.recovered	463	(19%)	0.92	(0.15)		0.47	(0.32)		97	(23%)	0.93	(0.16)		0.51	(0.30)	
Q	Much improved	945	(38%)	0.74	(0.17)		0.34	(0.32)		175	(42%)	0.75	(0.16)		0.37	(0.32)	
-	Slightly improved	543	(22%)	0.55	(0.26)		0.19	(0.33)		89	(21%)	0.46	(0.29)		0.18	(0.31)	
5	Unchanged	230	(9%)	0.41	(0.31)		0.03	(0.29)		26	(6%)	0.40	(0.30)		0.08	(0.33)	
D	Slightly worse	148	(6%)	0.33	(0.32)		0.00	(0.32)		17	(4%)	0.36	(0.30)		0.13	(0.29)	
	Much worse	100	(4%)	0.15	(0.23)		0.15	(0.32)		8	(2%)	0.30	(0.34)		0.02	(0.08)	
	Worse than ever	35	(1%)	0.04	(0.22)		0.24	(0.37)		7	(2%)	0.08	(0.24)		0.03	(0.11)	
	Missing	1395								198							
L	All	2988		3.5	(3.0)	0.72*	3.1	(3.3)	0.63*	493		3.2	(2.9)	0.68*	3.5	(3.2)	0.58*
E	Compl.Recovered	580	19%	0.6	(1.5)		5.9	(2.5)		112	(23%)	0.6	(2.2)		6.0	(2.5)	
G	Much improved	1159	39%	2.5	(2.2)		4.0	(2.7)		208	(42%)	2.6	(2.2)		4.0	(2.7)	
	Slightly improved	640	21%	4.9	(2.2)		1.8	(2.6)		102	(20%)	4.8	(2.4)		1.8	(2.6)	
P	Unchanged	275	9%	6.3	(2.1)		0.1	(2.3)		33	(7%)	6.1	(4.7)		0.4	(2.2)	
A	Slightly worse	176	6%	6.4	(2.1)		0.7	(2.6)		18	(4%)	5.2	(3.0)		1.0	(2.9)	
I	Much worse	114	4%	7.5	(2.1)		−0.5	(2.6)		11	(2%)	6.6	(2.3)		0.4	(3.8)	
N	Worse than ever	44	1%	7.7	(2.1)		−0.4	(2.9)		9	(2%)	7.8	(1.9)		0.0	(2.1)	
	Missing	871								124							
B	All	3033		3.8	(2.8)	0.76*	3.3	(2.9)	0.62*	507		3.6	(2.8)	0.78*	3.3	(2.9)	0.64*
A	Compl. recovered	592	20%	0.6	(1.4)		5.4	(2.5)		117	(23%)	0.7	(2.0)		5.8	(2.5)	
C	Much improved	1171	38%	3.0	(2.0)		3.2	(2.5)		214	(42%)	3.0	(2.0)		3.7	(2.5)	
K	Slightly improved	648	21%	5.2	(1.9)		1.4	(2.3)		105	(21%)	5.7	(1.7)		1.6	(1.8)	
	Unchanged	278	9%	6.5	(2.0)		0.5	(2.0)		32	(6%)	6.0	(2.0)		1.4	(2.1)	
P	Slightly worse	182	6%	6.7	(1.8)		0.1	(2.0)		20	(4%)	6.7	(1.6)		0.3	(1.6)	
A	Much worse	116	4%	7.4	(2.1)		−0.1	(2.2)		11	(2%)	7.3	(2.19		−0.2	(1.3)	
I	Worse than ever	46	2%	8.3	(1.9)		−0.8	(2.3)		8	(2%)	8.5	(1.3)		−0.4	(1.4)	
N	Missing	826								110							

Mean 1 year follow-up scores and mean change scores from baseline to follow-up for ODI, EQ-5D, NRS leg pain, and NRS back pain [positive values indicate decreased disability (ODI), improved health-related quality of life (EQ-5D), and reduced pain (NRS)]. Results are given for all patients, and for patients stratified according to the Global Perceived Effect (GPE) scale. The association between the outcome measurements and the GPE responses are given by Spearman's rank correlation coefficients (Spearman's rho)

*$p<0.005$

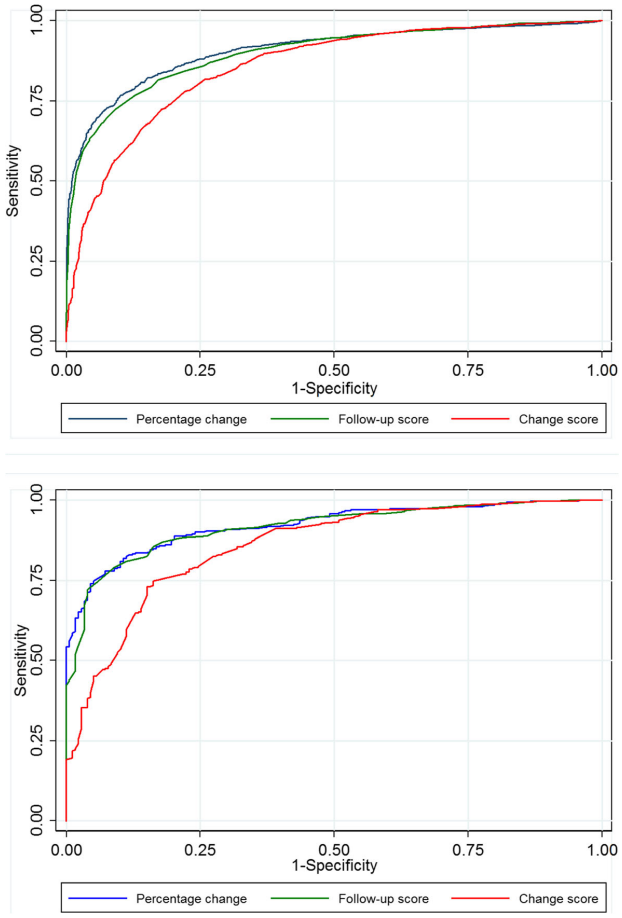


Fig. 2 Receiver Operating Characteristic curves for ODI. Legend: The closer the curve is in the upper left corner, the higher accuracy for determining whether a patients is cured ('completely recovered' or 'much improved') or not. 2**a**. Spinal stenosis; 2**b**. Degenerative spondylolisthesis

cut-off values were the same for LSS and LDS, except for the change score for NRS back pain.

Other studies

Follow-up score

In a study with a similar methodology to the present study, Fekete et al. [18] suggested that a follow-up score of ≤3 points is the best cut-off value for an acceptable level of leg pain and back pain following surgery for spinal stenosis with ($n = 910$) and without degenerative spondylolisthesis ($n = 1625$). This is in accordance with our estimate for leg pain and one point lower than our estimate for back pain. In a study [19] on patients with degenerative lumbar spine disorders operated with decompression ($n = 1288$), the estimated cut-off for ODI for a satisfactory symptom state was ≤22, nearly equivalent to our own criterion (≤24). Furthermore, they found the same cut-off estimates at 1-year and 2-year follow-up [19].

Change score

Carreon et al. [40] analysed patients operated with primary fusion surgery – 332 for spinal stenosis with spondylolisthesis (including both isthmic and degenerative cases) and 153 for spinal stenosis without spondylolisthesis. They evaluated the change score and found the minimum detectable change (smallest change above the upper limit of a 95% CI for the measurement error) to be 12.5 for ODI,

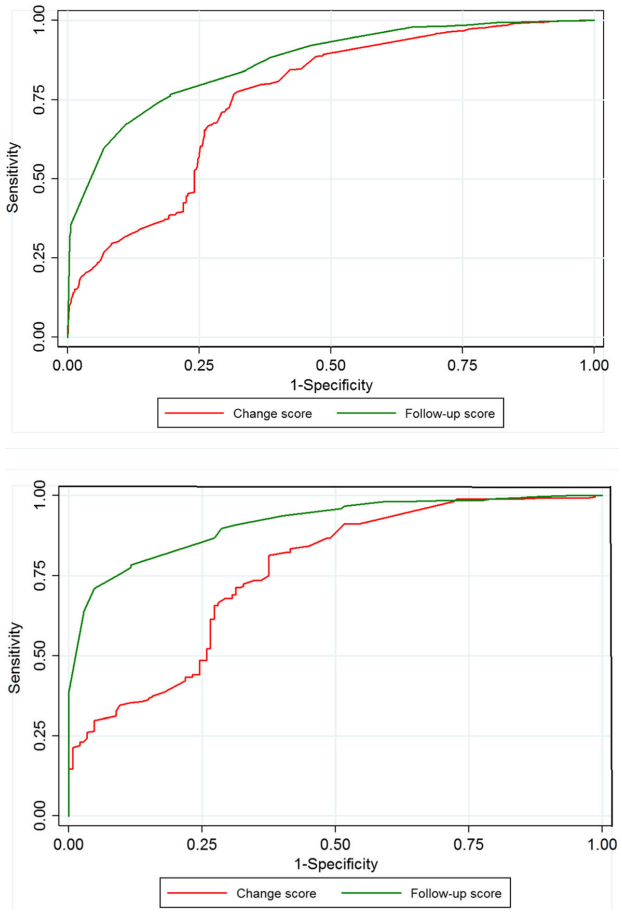


Fig. 3 Receiver Operating Characteristic curves for EQ-5D. Legend: The closer the curve is in the upper left corner, the higher accuracy for determining whether a patients is cured ('completely recovered' or 'much improved') or not. 3**a**. Spinal stenosis; 3**b**. Degenerative spondylolisthesis

1.2 for NRS leg pain and 1.1 for NRS back pain. All these thresholds were below our estimated thresholds. Glassman et al. [17] found 18.8 for ODI, 2.5 for NRS leg pain and 2.5 for NRS back pain to be cut-offs for a substantial clinical improvement for patients ($n = 357$) treated with fusion surgery for several spinal disorders. Their ODI limit was higher than in our study, whereas their thresholds for pain were in accordance with our results.

The use of change scores for benchmarking has been criticized for not taking into account the patients' baseline scores [12, 18, 41]. A numerical change from high baseline scores is probably of less importance than a change from low baseline scores.

In the present study, the change scores' weak ability to correctly classify patients in the upper and lower baseline groups lends support to this criticism.

Percentage change score

In order to account for the influence of the baseline score on the outcome score, using the percentage change score has been recommended [12, 42]. Based on a literature review and an expert panel decision, Ostelo et al. [42] concluded that a > 30% change from baseline to follow-up was the best threshold for identifying clinically meaningful improvement in ODI and NRS back pain. Their cut-off for ODI is identical to our estimate, and their threshold for pain is in accordance with our estimate

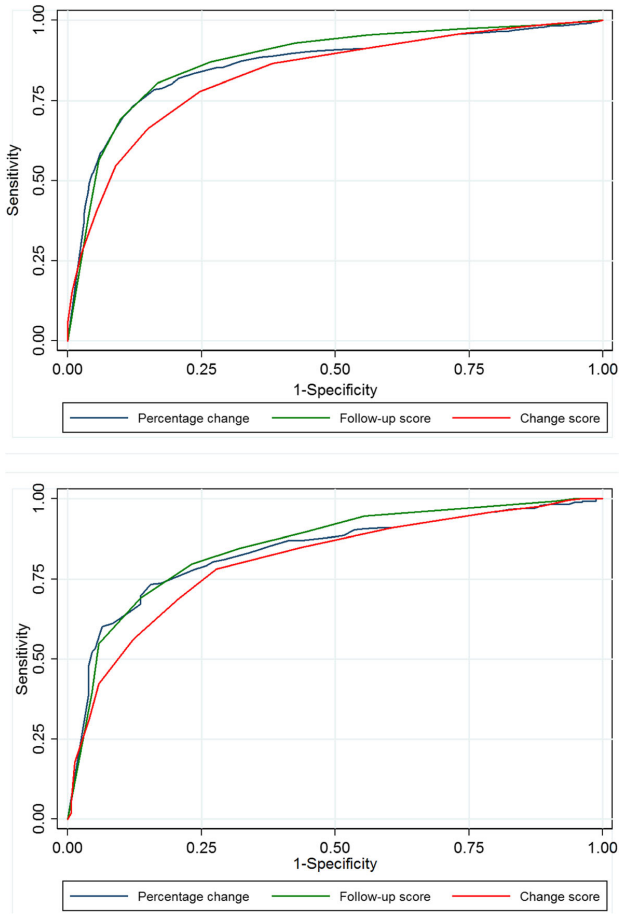


Fig. 4 Receiver Operating Characteristic curves for NRS leg pain. Legend: The closer the curve is in the upper left corner, the higher accuracy for determining whether a patients is cured ('completely recovered' or 'much improved') or not. 4**a**. Spinal stenosis; 4**b**. Degenerative spondylolisthesis

(> 33%). Dworkin et al. [12] suggested a 30% reduction in pain to be moderately important and a 50% reduction to be substantially important for patients treated for chronic pain. Our cut-off estimates for NRS leg- and back pain for LSS and LDS were between these suggestions.

Methodical challenges

Because the EQ-5D questionnaire values ranged from − 0.59 to 1.0, it was not possible to adequately calculate the percentage change score. Hence, only the 12-month follow-up score and the change score could be provided for the EQ-5D.

Application of the thresholds

As for other metrics developed for determining a clinically relevant outcome following treatment (i.e., MCID [8], (MIC) [27], a substantial clinical benefit [11] and a satisfactory symptom state [28]), it is essential to recognize that the thresholds from the present study cannot be directly applied to comparisons of mean outcome scores between groups [12, 13, 17, 43]. The thresholds are developed to determine whether an individual has achieved an important preoperative to postoperative benefit/improvement and should be used in the same context when comparing treatment effects [13]. Assuming a mean between-group

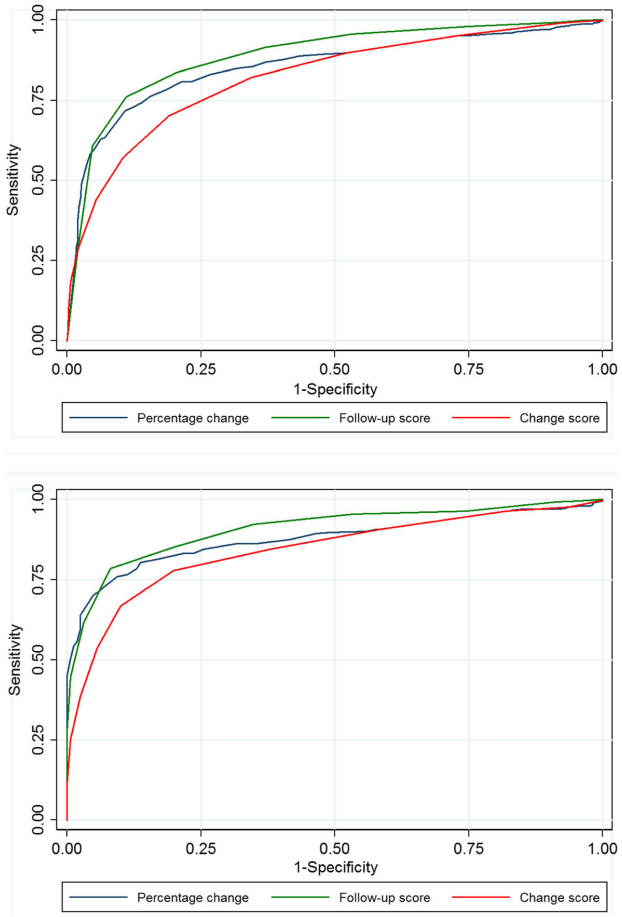


Fig. 5 Receiver Operating Characteristic curves for NRS back pain. Legend: The closer the curve is in the upper left corner, the higher accuracy for determining whether a patients is cured ('completely recovered' or 'much improved') or not. 5**a**. Spinal stenosis; 5**b**. Degenerative spondylolisthesis

difference in a PROM less than MCID to be clinically unimportant and a difference above MCID to be clinically important is warned against [12, 13, 43]. Instead the proportion of patients reaching the threshold for clinical improvement (the 'success' rate) should be calculated for each treatment group. Then the 'success' rates should be compared between the groups. This approach is advocated as a statistically and clinically useful tool for evaluating treatment effects [12, 16, 17, 24, 43, 44]. In discussion with patients, knowledge of the 'success' rate for a treatment can be employed as clinically relevant information in a shared decision-making process [17]. Furthermore, knowing the difference in the 'success' rates of two treatment groups makes it is possible to calculate the number needed to treat to obtain one extra patient with 'success' in an investigated group compared to a control group (NNT = 100 divided by the absolute difference in 'success' rate) [6, 12, 44]. For example, in patients with degenerative spondylolisthesis treated with either decompression alone or decompression with fusion, assessed by ODI, how many patients must be fused to get one more patient with a clinically relevant outcome? [6]. Finally, assumptions regarding the difference in the 'success' rate between groups provide the opportunity to estimate a statistically and clinically relevant sample size when planning a clinical trial [6, 12].

Table 3 ROC analyses for determining AUC (95% CI) and for estimating cut-off values for 'success

	ODI				EQ-5D				NRS	Leg pain		NRS	Back pain		
	AUC	(95% CI)	Max corr. Class	Cut-off	AUC	(95% CI)	Max corr. Class	Cut-off	AUC		Max corr. Class.	Cut-off	AUC		Max corr. class.	Cut-off
Spinal stenosis																
Follow-up score (points)	0.90	(0.89–0.91)	82%	≤24	0.87	(0.85–0.88)	78%	0.692	0.87	(0.86–0.89)	81%	≤4	0.89	(0.87–0.90)	82%	≤4
Change score (points)	0.86	(0.84–0.87)	78%	≥13	0.76	(0.74–0.78)	73%	0.105	0.83	(0.82–0.85)	77%	≥3	0.82	(0.81–0.84)	75%	≥2
Percentage change (%)	0.91	(0.90–0.92)	83%	≥30					0.86	(0.85–0.88)	81%	≥40	0.86	(0.84–0.87)	79%	≥33
Degenerative Spondylolisthesis																
Follow-up score (points)	0.92	(0.89–0.94)	85%	≤24	0.92	(0.89–0.94)	84%	≥0.692	0.86	(0.82–0.89)	79%	≤3	0.90	(0.88–0.93)	83%	≤4
Change score (points)	0.86	(0.82–0.89)	80%	≥13	0.76	(0.76–0.81)	76%	≥0.105	0.81	(0.77–0.91)	76%	≥3	0.84	(0.81–0.88)	79%	≥3
Percentage change (%)	0.92	(0.90–0.94)	85%	≥30					0.84	(0.80–0.87)	78%	≥40	0.88	(0.85–0.91)	80%	≥33

The area under the curve (AUC) with 95% confidence interval (CI) describes a candidate score's ability to classify patients who replied 'completely recovered' or 'much improved' on the GPE scale into 'success' and those replied 'slightly improved', 'unchanged', 'slightly worse', 'much worse', and 'worse than ever' into 'non-success' at 12 month follow-up. The larger the AUC, the better the accuracy of the score [range from 0.5 (no ability) to 1.0 (perfect ability)]. A cut-off corresponds to the threshold that gave rise to the maximum percentage of patients correctly classified (max corr. Class.) into 'success' and 'non-success'. Results are given for ODI, EQ-5D, NRS leg pain, and NRS back pain for spinal stenosis and for degenerative spondylolisthesis. Because the EQ-5D questionnaire values ranged from −0.6 to 1.0 on a categorical scale, it was not mathematically possible to evaluate the percent change score

Table 4 Sensitivity and specificity for estimated cut-off values. Correct classification rate in different PROM baseline groups

	Spinal stenosis				Degenerative spondylolisthesis			
	Estimated cut-off	Correct classification	Sensitivity	Specificity	Estimated cut-off	Correct classification	Sensitivity	Specificity
ODI follow-up score	≤24		**0.83**	**0.80**	≤24		**0.85**	**0.84**
Low baseline		80%				87%		
Medium		85%				85%		
High baseline		80%				84%		
ODI change score	≥13		**0.78**	**0.77**	≥13		**0.83**	**0.71**
Low baseline		72%				77%		
Medium		84%				86%		
High baseline		78%				75%		
ODI percentage change	≥30		**0.87**	**0.77**	≥30		**0.89**	**0.77**
Low baseline		83%				88%		
Medium		85%				85%		
High baseline		80%				82%		
EQ-5D follow-up score	≥0.692		**0.76**	**0.81**	≥0.692		**0.80**	**0.88**
Low baseline		75%				81%		
Medium		79%				80%		
High baseline		80%				82%		
EQ-5D change score	≥0.105		**0.77**	**0.68**	≥0.105		**0.81**	**0.63**
Low baseline		73%				74%		
Medium		75%				80%		
High baseline		72%				71%		
Leg pain follow-up score	≤3		**0.80**	**0.83**	≤3		**0.79**	**0.78**
Low baseline		82%				81%		
Medium		82%				76%		
High baseline		81%				79%		
Leg pain change score	≥3		**0.78**	**0.75**	≥3		**0.78**	**0.72**
Low baseline		69%				70%		
Medium		82%				76%		
High baseline		78%				80%		
Leg pain percentage change	≥40		**0.82**	**0.80**	≥40		**0.80**	**0.73**
Low baseline		79%				75%		
Medium		81%				76%		
High baseline		81%				81%		
Back pain follow-up score	≤4		**0.84**	**0.79**	≤4		**0.85**	**0.79**
Low baseline		81%				82%		
Medium		83%				80%		
High baseline		82%				87%		
Back pain change score	≥2		**0.82**	**0.66**	≥3		**0.78**	**0.80**
Low baseline		72%				67%		
Medium		83%				81%		
High baseline		71%				83%		
Back pain percentage change	≥33%		**0.81**	**0.79**	≥33%		**0.83**	**0.78**
Low baseline		76%				78%		
Medium		83%				80%		
High baseline		80%				85%		

The sensitivity describes the probability of correctly classifying an individual replying 'completely recovered' or 'much improved' (GPE) as a 'success' when assessed by the estimated cut-offs for the PROMs. The specificity describes the probability for detecting a 'non-success' patient (one with a lower response at the GPE scale)
For each estimated cut-off values the percentage of correctly classified patients (correct classification) into 'success' and 'non-success' according to the anchor are given separately for patients with low (ODI; 0–32, EQ-5D; −0.59-0.1, NRS leg and back pain; 0–5), medium (ODI; 32–46, EQ-5D; 0.1–0.6, NRS leg and back pain; 6–7), and high (ODI; 46 to 100, EQ-5D; 0.6–1.0, NRS leg and back pain; 8–10) baseline scores

The proposed threshold from the present study is derived from populations with LSS and LDS. The threshold is condition-specific [13] and should be applied solely to these conditions.

Strengths and limitations of the study

Strengths of this study are the large sample size and the collection of data through a comprehensive and well-structured registry. More than 90% of the national centres performing spinal stenosis surgery report to the registry, and currently more than 65% of operations for spinal stenosis are registered. The follow-up rate was good and in accordance with recommendations for spine registries [45].

For research on effectiveness and efficacy following treatment in a specific patient group it is recommended to use criteria for clinical improvement derived from populations similar to the one being studied [13]. The estimated thresholds derived from patients operated for LSS and LDS ensure reliable estimates for these conditions. Finally, we consider the evaluation of all scores in the same study and the consecutive sub-group analysis of the three baseline groups to be strengths.

There are several major limitations in the method used for determining the thresholds. As long as we know, the validity of a single-item rating (GPE scale) of how the patients are doing one year after spine surgery is not proven specifically for LSS and LDS. However there are some arguments in its favour. Using global assessment to evaluate patients' satisfaction with treatment outcome in spinal disorders is recommended by international panels of experts in the field [12, 46, 47]. The global assessment of 'pain free or much better' and 'much or very much improved' has been considered to be an appropriate reference criterion for a successful outcome following spinal surgery [35]. In a Norwegian study of the validity of the GPE scale, the GPE replies were strongly associated with the follow-up scores for PROMs [48].

Another limitation is the evaluation of self-report measurements (ODI, EQ-5D, and NRS leg- and back pain) against another related self-report instrument (GPE) as a criterion [20]. Alternatively, an objective functional 'non-self-report' outcome, such as return to work, has been recommended [20]. However, this criterion is also criticized as return to work is not necessarily the primary goal for all patients, and it is not a relevant measurement for an elderly patient group [49]. Walking capacity is another criterion used to assess functional outcome in patients with spinal stenosis. In addition to asking about the walking distance before and after surgery, an objective assessment of walking distance could be recorded [50]. The differences in activity levels preoperatively and the patients expectations or anticipated activity level after surgery should also be taken into account. Patients' who are happy to perform their limited activities of daily living, most probably accept more disability than patients involved in more demanding activities such as running and playing tennis. A suggested method, the 'benefit-harm trade-off method' [51, 52], in which the patients are asked to estimate how much benefit they would consider sufficient to justify the risk of getting worse after surgery, would take into account the patients' accepted physical performance level. For the future this may be a suitable alternative approach for determining 'success'- criteria.

The method used in the present study is described in detail and advocated by the 'IMMPACT Recommendation' [12], and is the most frequently used method for determining thresholds for clinical importance [17, 18, 25–28]. Furthermore, according to US FDA-recommended methodology for defining thresholds for PROMs, the GPE scale is considered a suitable anchor [29].

It is essential that the estimated PROM thresholds should be utilised and interpreted with caution. The evaluated PROMs do not assess all aspects that may be considered important for an individual. A patient who obtains an outcome in a PROM which exceeds the threshold for clinical importance may have non-observed complaints that are not detected; for example, loss of agility, slow walking speed and general stiffness of the back. Furthermore, objective data such as measured walking distance and muscle strength are not recorded in the registry questionnaire. Therefore, when reporting a 'success' rate it should be made clear that it is only an estimate of the proportion of patients reaching a threshold for improvement in a PROM considered to be of importance for a patient. An ideal PROM that covers all relevant domains of importance for all kind of patients will give a more accurate estimate of the 'success' rate.

Conclusion

For estimating 'success' rates assessed by PROMs for patients operated for LSS and LDS we recommend using the follow-up score or the percentage change score. These scores reflect a clinically important outcome more accurately than the change score.

Abbreviations

AUC: Area under curve; CI: Confidence interval; EQ-5D: EuroQol 5-dimensional questionnaire utility index; GPE: Global perceived effect; LDS: Lumbar degenerative spondylolisthesis; LSS: Lumbar spinal stenosis; NNT: Number needed to treat; NRS: Numeric rating scale; ODI: Oswestry disability index; PROMs: Patient reported outcome measurements; ROC: Receiver operating characteristics

Acknowledgements

Thanks to Eira Kathleen Ebbs for linguistic assistance in writing the manuscript.

Availability of data and material

The data that support the findings of this study are available from the Norwegian Registry for Spine Surgery but restrictions apply to the availability

of these data, which were used under license for the current study and are therefore not publicly available. Data are, however, available to researchers with the permission of the Norwegian Committee for Medical and Health Research Ethics and the Norwegian Registry for Spine Surgery.

Funding
The project has received funding from Helse Vest RHF (the Western Regional Health Authority). The funder has no influence on study design, management and interpretation of data or the decision to submit data.

Authors' contributions
IMA, RG, MG, TS, KS, JIB, EH, FR, KI and CH have been involved in planning the study and in drafting the manuscript. All authors read and approved the final manuscript. All authors meet the ICMJE guidelines for authorship.

Ethics approval and consent to participate
All patients have signed an informed consent form.
The protocol has been approved by the Norwegian Committee for Medical and Health Research Ethics Midt (2014/344).

Consent for publication
Not applicable.

Competing interests
None of the authors have any conflicts of interest.

Publisher's Note

Author details
[1]Kysthospitalet in Hagevik, Orthopedic Clinic, Haukeland, University Hospital, Hagaviksbakken 25, 5217 Hagevik, Bergen, Norway. [2]Department of Clinical Medicine, University of Bergen, Christies gate 6, 5007 Bergen, Bergen, Norway. [3]The Norwegian Registry for Spine Surgery (NORspine), Northern Norway Regional Health Authority, Postboks 20, 9038 Tromsø, Bodø, Norway. [4]Research Department, Division of Psychiatry, Haukeland University Hospital, Sanviksleitet 1, 5036 Bergen, Bergen, Norway. [5]Department of Physical Medicine and Rehabilitation, Oslo University Hospital, PB 4950 Nydalen, 0424 Oslo, Oslo, Norway. [6]Department of Neurosurgery, University Hospital of Northern Norway, Sykehusvegen 38, 90919 Tromsø, Tromsø, Norway. [7]Research and Communication Unit for Musculoskeletal Health (FORMI), Oslo University Hospital, PB 4950 Nydalen, 0424 Oslo, Oslo, Norway. [8]Department of Orthopedic Surgery, Ålesund Hospital, Møre and Romsdal Hospital Trust, Ålesund, Norway. [9]Division of Orthopaedic Surgery, Oslo University Hospital, 4950 Nydalen, 0424 Oslo, PB, Norway. [10]Faculty of Health Science, OsloMet – Oslo Metropolitan University, PO box 4 St. Olavs plass, 0130 Oslo, Oslo, Norway.

Received: 24 November 2017 Accepted: 19 December 2018
Published online: 18 January 2019

References

1. Forsth P, Olafsson G, Carlsson T, Frost A, Borgstrom F, Fritzell P, Ohagen P, Michaelsson K, Sanden B. A randomized, controlled trial of fusion surgery for lumbar spinal stenosis. NEJM. 2016;374:1413–23. https://doi.org/10.1056/NEJMoa1513721.
2. Ghogawala Z, Dziura J, Butler WE, Dai F, Terrin N, Magge SN, Coumans JV, Harrington JF, Amin-Hanjani S, Schwartz JS, Sonntag VK, Barker FG 2nd, Benzel EC. Laminectomy plus fusion versus laminectomy alone for lumbar spondylolisthesis. NEJM. 2016;374:1424–34. https://doi.org/10.1056/NEJMoa1508788.
3. Watters WC 3rd, Baisden J, Gilbert TJ, Kreiner S, Resnick DK, Bono CM, Ghiselli G, Heggeness MH, Mazanec DJ, O'Neill C, Reitman CA, Shaffer WO, Summers JT, Toton JF. Degenerative lumbar spinal stenosis: an evidence-based clinical guideline for the diagnosis and treatment of degenerative lumbar spinal stenosis. Spine J. 2008;8:305–10. https://doi.org/10.1016/j.spinee.2007.10.033.
4. Weinstein JN, Lurie JD, Tosteson TD, Hanscom B, Tosteson AN, Blood EA, Birkmeyer NJ, Hilibrand AS, Herkowitz H, Cammisa FP, Albert TJ, Emery SE, Lenke LG, Abdu WA, Longley M, Errico TJ, Hu SS. Surgical versus nonsurgical treatment for lumbar degenerative spondylolisthesis. NEJM. 2007;356:2257–70. https://doi.org/10.1056/NEJMoa070302.
5. Weinstein JN, Tosteson TD, Lurie JD, Tosteson AN, Blood E, Hanscom B, Herkowitz H, Cammisa F, Albert T, Boden SD, Hilibrand A, Goldberg H, Berven S, An H. Surgical versus nonsurgical therapy for lumbar spinal stenosis. NEJM. 2008;358:794–810. https://doi.org/10.1056/NEJMoa0707136.
6. Austevoll IM, Gjestad R, Brox JI, Solberg TK, Storheim K, Rekeland F, Hermansen E, Indrekvam K, Hellum C. The effectiveness of decompression alone compared with additional fusion for lumbar spinal stenosis with degenerative spondylolisthesis: a pragmatic comparative non-inferiority observational study from the Norwegian registry for Spine surgery. ESJ. 2017;26:404–13. https://doi.org/10.1007/s00586-016-4683-1.
7. Forsth P, Michaelsson K, Sanden B (2013) Does fusion improve the outcome after decompressive surgery for lumbar spinal stenosis?: A two-year follow-up study involving 5390 patients. Bone Joint J 95-B:960–965. doi: 95-B/7/960 [pii];https://doi.org/10.1302/0301-620X.95B7.30776 [doi].
8. Hellum CF, – Johnsen LG FAU - Storheim K, Storheim KF, – Nygaard OP FAU - Brox JI, – Brox JI FAU - Rossvoll I, Rossvoll IF, Ro MF, Sandvik LF, Grundnes O (2011)Surgery with disc prosthesis versus rehabilitation in patients with low back pain and degenerative disc: two year follow-up of randomised study. BMJ 19;342:d2786. doi: 101136/bmjd2786:BMJ.
9. Hermansen E, Austevoll IM, Romild UK, Rekeland F, Solberg T, Storheim K, Grundnes O, Aaen J, Brox JI, Hellum C, Indrekvam K. Study-protocol for a randomized controlled trial comparing clinical and radiological results after three different posterior decompression techniques for lumbar spinal stenosis: the spinal stenosis trial (SST) (part of the NORDSTEN study). BMC musculoskel Disord. 2017;18:121. https://doi.org/10.1186/s12891-017-1491-7.
10. Lonne G, Johnsen LG, Aas E, Lydersen S, Andresen H, Ronning R, Nygaard OP. Comparing cost-effectiveness of X-stop with minimally invasive decompression in lumbar spinal stenosis: a randomized controlled trial. Spine. 2015;40:514–20. https://doi.org/10.1097/brs.0000000000000798.
11. Guyatt GH, Juniper EF, Walter SD, Griffith LE, Goldstein RS. Interpreting treatment effects in randomised trials. BMJ (Clinical research ed). 1998;316:690–3.
12. Dworkin RH, Turk DC, Wyrwich KW, Beaton D, Cleeland CS, Farrar JT, Haythornthwaite JA, Jensen MP, Kerns RD, Ader DN, Brandenburg N, Burke LB, Cella D, Chandler J, Cowan P, Dimitrova R, Dionne R, Hertz S, Jadad AR, Katz NP, Kehlet H, Kramer LD, Manning DC, McCormick C, McDermott MP, McQuay HJ, Patel S, Porter L, Quessy S, Rappaport BA, Rauschkolb C, Revicki DA, Rothman M, Schmader KE, Stacey BR, Stauffer JW, von Stein T, White RE, Witter J, Zavisic S. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. J Pain. 2008;9:105–21. https://doi.org/10.1016/j.jpain.2007.09.005.
13. Katz NP, Paillard FC, Ekman E. Determining the clinical importance of treatment benefits for interventions for painful orthopedic conditions. J Orthop Surg. 2015;10:24. https://doi.org/10.1186/s13018-014-0144-x.
14. Stratford PW, Binkley JM, Riddle DL, Guyatt GH. Sensitivity to change of the Roland-Morris Back pain questionnaire: part 1. Phys Ther. 1998;78:1186–96.
15. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials. 1989;10:407–15.
16. van Kampen DA, Willems WJ, van Beers LW, Castelein RM, Scholtes VA, Terwee CB. Determination and comparison of the smallest detectable change (SDC) and the minimal important change (MIC) of four-shoulder patient-reported outcome measures (PROMs). JOrthop Surg Res. 2013;8:40. https://doi.org/10.1186/1749-799x-8-40.
17. 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 volume. 2008;90:1839–47. https://doi.org/10.2106/jbjs.g.01095.
18. Fekete TF, Haschtmann D, Kleinstuck FS, Porchet F, Jeszenszky D, Mannion A. What level of pain are patients happy to live with after surgery for lumbar degenerative disorders? Spine J. 2016. https://doi.org/10.1016/j.spinee.2016.01.180.
19. 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. Spine J. 2016;16:1221–30. https://doi.org/10.1016/j.spinee.2016.06.010.
20. Gatchel RJ, Mayer TG. Testing minimal clinically important difference: consensus or conundrum? Spine J. 2010;10:321–7. https://doi.org/10.1016/j.spinee.2009.10.015.
21. Terwee CB, Roorda LD, Dekker J, Bierma-Zeinstra SM, Peat G, Jordan KP, Croft P, de Vet HCW. Mind the MIC: large variation among populations and methods. J Clin Epidemiol. 2010;63:524–34. https://doi.org/10.1016/j.jclinepi.2009.08.010.

22. Wright A, Hannon J, Hegedus EJ, Kavchak AE. Clinimetrics corner: a closer look at the minimal clinically important difference (MCID). J Man Manip Ther20. 2012:160–6. https://doi.org/10.1179/2042618612y.0000000001.
23. Kamper SJ, Ostelo RW, Knol DL, Maher CG, de Vet HC, Hancock MJ (2010) Global Perceived Effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. J Clin Epidemiol 63:760–766. doi: S0895-4356(09)00304–7 [pii];https://doi.org/10.1016/j.jclinepi.2009.09.009 [doi].
24. Ostelo RW, de Vet HC. Clinically important outcomes in low back pain. Best Pract Res Clin Rheumatol. 2005;19:593–607. https://doi.org/10.1016/j.berh.2005.03.003.
25. Copay AG, Glassman SD, Subach BR, Berven S, Schuler TC, Carreon LY. Minimum clinically important difference in lumbar spine surgery patients: a choice of methods using the Oswestry disability index, medical outcomes study questionnaire short form 36, and pain scales. Spine J. 2008;8:968–74. https://doi.org/10.1016/j.spinee.2007.11.006.
26. de Vet HC, Foumani M, Scholten MA, Jacobs WC, Stiggelbout AM, Knol DL, Peul WC. Minimally important change values of a measurement instrument depend more on baseline values than on the type of intervention. J Clinical Epidemiol. 2015;68:518–24. https://doi.org/10.1016/j.jclinepi.2014.07.008.
27. Solberg T, Johnsen LG, Nygaard OP, Grotle M. Can we define success criteria for lumbar disc surgery? estimates for a substantial amount of improvement in core outcome measures Acta Orthop. 2013;84:196–201. https://doi.org/10.3109/17453674.2013.786634.
28. Mannion AF, Fekete TF, Wertli MM, Mattle M, Nauer S, Kleinstuck FS, Jeszenszky D, Haschtmann D, Becker HJ, Porchet F. Could less be more when assessing patient-rated outcome in spinal stenosis? Spine. 2015;40:710–8. https://doi.org/10.1097/brs.0000000000000751.
29. McLeod LD, Coon CD, Martin SA, Fehnel SE, Hays RD. Interpreting patient-reported outcome results: US FDA guidance and emerging methods. Expert Rev Pharmacoecon Outcomes Res. 2011;11:163–9. https://doi.org/10.1586/erp.11.12.
30. Baker DJ, Pynsent PB, J F (1989) The Oswestry disability index revisited: its reliability, repeatability and validity, and a comparison with the St Thomas's disability index. Roland MO, Jenner JR, eds New approaches to rehabilitation and education Manchester: Manchester University Press:174–186.
31. Fairbank JC, Pynsent PB. The Oswestry disability index. Spine. 2000;25:2940–52 discussion 2952.
32. 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.
33. Watters WC, 3rd, Bono CM, Gilbert TJ, Kreiner DS, Mazanec DJ, Shaffer WO, Baisden J, Easa JE, Fernand R, Ghiselli G, Heggeness MH, Mendel RC, O'Neill C, Reitman CA, Resnick DK, Summers JT, Timmons RB, Toton JF, North American Spine S (2009) An evidence-based clinical guideline for the diagnosis and treatment of degenerative lumbar spondylolisthesis. Spine J 9:609–614. doi: https://doi.org/10.1016/j.spinee.2009.03.016.
34. Dolan P, Gudex C, Kind P, Williams A. The time trade-off method: results from a general population study. Health Econ. 1996;5:141–54. https://doi.org/10.1002/(sici)1099-1050(199603)5:2<141::aid-hec189>3.0.co;2-n.
35. Parai C, Hagg O, Lind B, Brisby H. The value of patient global assessment in lumbar spine surgery: an evaluation based on more than 90,000 patients. ESJ. 2018;27:554–63. https://doi.org/10.1007/s00586-017-5331-0.
36. Altman DG, Bland JM. Diagnostic tests 3: receiver operating characteristics plots. BMJ (Clinical research ed). 1994;309:188.
37. Tape TG (2006 Dec 18) Interpreting diagnostic tests. http://gim.unmc.edu/dxtests/ROC3.htm.
38. 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;82:56–63. https://doi.org/10.3109/17453674.2010.548024 [doi].
39. Schafer JL, Graham JW. Missing data: our view of the state of the art. Psychol Methods. 2002;7:147–77.
40. Carreon LY, Bratcher KR, Canan CE, Burke LO, Djurasovic M, Glassman SD. Differentiating minimum clinically important difference for primary and revision lumbar fusion surgeries. J Neurosurg Spine. 2013;18:102–6. https://doi.org/10.3171/2012.10.spine12727.
41. de Vet HC, Ostelo RW, Terwee CB, van der Roer N, Knol DL, Beckerman H, Boers M, Bouter LM. Minimally important change determined by a visual method integrating an anchor-based and a distribution-based approach. Qual Life Res. 2007;16:131–42. https://doi.org/10.1007/s11136-006-9109-9.
42. Ostelo RW, Deyo RA, Stratford P, Waddell G, Croft P, Von Korff M, Bouter LM, de Vet HC. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. Spine. 2008;33:90–4. https://doi.org/10.1097/BRS.0b013e31815e3a10.
43. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR. Methods to explain the clinical significance of health status measures. Mayo Clinic Proc. 2002;77:371–83. https://doi.org/10.1016/s0025-6196(11)61793-x.
44. Katz N, Paillard FC, Van Inwegen R. A review of the use of the number needed to treat to evaluate the efficacy of analgesics. J Pain. 2015;16:116–23. https://doi.org/10.1016/j.jpain.2014.08.005.
45. van Hooff ML, Jacobs WC, Willems PC, Wouters MW, de Kleuver M, Peul WC, Ostelo RW, Fritzell P. Evidence and practice in spine registries. Acta Orthop. 2015;86:534–44. https://doi.org/10.3109/17453674.2015.1043174.
46. Bombardier C. Outcome assessments in the evaluation of treatment of spinal disorders: summary and general recommendations. Spine. 2000;25:3100–3.
47. Hudak PL, Wright JG. The characteristics of patient satisfaction measures. Spine. 2000;25:3167–77.
48. Grovle L, Haugen AJ, Hasvik E, Natvig B, Brox JI, Grotle M. Patients' ratings of global perceived change during 2 years were strongly influenced by the current health status. J Clin Epidemiol. 2014;67:508–15. https://doi.org/10.1016/j.jclinepi.2013.12.001.
49. Glassman SD, Carreon LY. Thresholds for health-related quality of life measures: reality testing. Spine J. 2010;10:328–9. https://doi.org/10.1016/j.spinee.2009.12.026.
50. Malmivaara A, Slatis P, Heliovaara M, Sainio P, Kinnunen H, Kankare J, Dalin-Hirvonen N, Seitsalo S, Herno A, Kortekangas P, Niinimaki T, Ronty H, Tallroth K, Turunen V, Knekt P, Harkanen T, Hurri H. Surgical or nonoperative treatment for lumbar spinal stenosis? A randomized controlled trial Spine. 2007;32:1–8. https://doi.org/10.1097/01.brs.0000251014.81875.6d.
51. Barrett B, Brown D, Mundt M, Brown R. Sufficiently important difference: expanding the framework of clinical significance. Med Decis Mak. 2005;25:250–61. https://doi.org/10.1177/0272989x05276863.
52. Ferreira ML, Herbert RD, Ferreira PH, Latimer J, Ostelo RW, Nascimento DP, Smeets RJ. A critical review of methods used to determine the smallest worthwhile effect of interventions for low back pain. J Clin Epidemiol. 2012;65:253–61. https://doi.org/10.1016/j.jclinepi.2011.06.018.

Paper IV

IV

Austevoll et al. BMC Musculoskeletal Disorders (2019) 20:7
https://doi.org/10.1186/s12891-018-2384-0

BMC Musculoskeletal Disorders

STUDY PROTOCOL Open Access

Decompression alone versus decompression with instrumental fusion the NORDSTEN degenerative spondylolisthesis trial (NORDSTEN-DS); study protocol for a randomized controlled trial

Ivar Magne Austevoll[1,2*], Erland Hermansen[1,2,3], Morten Fagerland[4], Frode Rekeland[1], Tore Solberg[5,6,7], Kjersti Storheim[8], Jens Ivar Brox[10], Greger Lønne[12], Kari Indrekvam[2,3], Jørn Aaen[2,11], Oliver Grundnes[9] and Christian Hellum[13]

Abstract

Background: Fusion in addition to decompression has become the standard treatment for lumbar spinal stenosis with degenerative spondylolisthesis (DS). The evidence for performing fusion among these patients is conflicting and there is a need for further investigation through studies of high quality. The present protocol describes an ongoing study with the primary aim of comparing the outcome between decompression alone and decompression with instrumented fusion. The secondary aim is to investigate whether predictors can be used to choose the best treatment for an individual. The trial, named the NORDSTEN-DS trial, is one of three studies in the Norwegian Degenerative Spinal Stenosis (NORDSTEN) study.

Methods: The NORDSTEN-DS trial is a block-randomized, controlled, multicenter, non-inferiority study with two parallel groups. The surgeons at the 15 participating hospitals decide whether a patient is eligible or not according to the inclusion and exclusion criteria. Participating patients are randomized to either a midline preserving decompression or a decompression followed by an instrumental fusion.
Primary endpoint is the percentage of patients with an improvement in Oswestry Disability Index version 2.0 of more than 30% from baseline to 2-year follow-up. Secondary outcome measurements are the Zürich Claudication Questionnaire, Numeric Rating Scale for back and leg pain, Euroqol 5 dimensions questionnaire, Global perceived effect scale, complications and several radiological parameters. Analysis and interpretation of results will also be conducted after 5 and 10 years.

(Continued on next page)

* Correspondence: imau@helse-bergen.no
[1]Kysthospitalet in Hagevik, Orthopedic Clinic, Haukeland University Hospital, Hagavik, N- 5217 Bergen, Norway
[2]Department of Clinical Medicine, University of Bergen, N- 5007 Bergen, Norway
Full list of author information is available at the end of the article

(Continued from previous page)
Conclusion: The NORDSTEN/DS trial has the potential to provide Level 1 evidence of whether decompression alone should be advocated as the preferred method or not. Further on the study will investigate whether predictors exist and if they can be used to make the appropriate choice for surgical treatment for this patient group.

Trial registration: ClinicalTrials.gov Identifier: NCT02051374. First Posted: January 31, 2014. Last Update Posted: February 14, 2018.

Keywords: Spinal stenosis, Degenerative spondylolisthesis, Randomized controlled trial, Decompression, Fusion, Clinical outcomes, NORDSTEN

Background

Lumbar degenerative spondylolisthesis (LDS) is the forward slip of one vertebra over another caused by degeneration and instability of facet joints, and degeneration of ligaments and intervertebral discs [1]. Most patients suffer from symptoms related to a concomitant spinal stenosis, such as back pain, radiating pain to the lower extremities, and, typically, increased pain when walking upright and decreased pain when bending forward [2, 3].

Several meta-analyses and systematic reviews have been published with the purpose of providing guidelines on how to surgically treat patients with degenerative spondylolisthesis. Based largely on a pseudorandomized study from 1991 [4], they conclude that there is moderate evidence for a tendency towards better outcome when decompression is combined with fusion [3, 5–7]. A recently published randomized controlled trial (RCT) has lent support to this evidence [8]. However, several cohort studies [9–11] and another recently published RCT [12], have introduced evidence against additional fusion when operating for LDS.

The current evidence cannot support any definite advice on operation method [13–16]. Although challenging, it is important to investigate how to treat this patient group.

Objectives

Primary objective

The primary objective is to detect whether the intervention-related difference in outcome between decompression alone (DA) and decompression with an additional instrumented fusion (DF) 2 years after surgery, is large enough to justify the use of instrumentation. Our hypothesis is that DA is "as good as" DF for the treatment of spinal stenosis with degenerative spondylolisthesis.

Secondary objectives

a. Health economic analysis: To compare the cost-utility of the investigated treatments DA and DF [17].
b. Predictor analysis: To evaluate whether radiological parameters and patient characteristics in the future can be used by clinicians to choose between DA and DF.
c. Long-time follow-up studies: The analyses performed at 2-year follow-up will be repeated at 5- and 10-year follow-up.

Trial design

The proposed trial is a 1:1 block-randomized, controlled, multicenter, non-inferiority trial, with two parallel groups.

The study is one of three trials in the NORDSTEN study, a Norwegian multicenter study on patients with lumbar spinal stenosis [18].

Methods

The SPIRIT checklist [19] has been used as a template for the present protocol. One exclusion criterion has been detached from the original study protocol (Version 1.0) received January 10, 2014 in Clinicaltrials.gov (Identifier: NCT02051374), see 'Amendment'.

The report of the trial will be based on an adapted Consolidated Standards of Reporting Trials (CONSORT) checklist for reporting non-inferiority trials [20].

Participants

The surgeons at the 15 participating hospitals (Table 1) are following the inclusion and exclusion criteria to decide whether a patient is eligible or not.

The patients are given verbal and written information about the study and the alternative treatment options. If willing to participate, the patients sign an informed consent form. If a patient does not want to participate in the study, he/she will not be included in the study and will receive normal care and be treated following the hospital's established procedures. Criteria for inclusion and exclusion are given in Table 2.

All eligible patients are being registered, and the reasons that some are not included are being documented and interpreted. A CONSORT flow chart is illustrated in Fig. 1.

Table 1 Recruiting hospitals

Oslo University Hospital, Orthopedic dept.
Akershus University Hospital, Orthopedic dept.
Bærum Hospital, Orthopedic dept.
Skien Hospital, Orthopedic dept.
Arendal Hospital, Orthopedic dept.
Gjøvik Hospital, Orthopedic dept.
Lillehammer Hospital, Orthopedic dept.
Stavanger University Hospital, Orthopedic dept. and dept. for Neurosurgery
Haukeland University Hospital, Orthopedic dept. and dept. for Neurosurgery
Kysthospitalet i Hagevik, Haukeland University Hospital, Orthopedic dept.
Ålesund Hospital, Orthopedic dept.
St. Olav University Hospital, dept. for Neurosurgery
University Hospital of Northern Norway, dept. for Neurosurgery
Kristiansand Hospital, Orthopedic dept.
Elverum Hospital, Orthopedic dept.

Interventions

Decompression alone

Posterior approach with decompression after microsurgical principles will be performed, and the midline structures will be preserved. The surgeons will either use a microscope or magnifying glasses.

Decompression and instrumental fusion

Posterior approach with decompression will be performed, followed by posterolateral pedicle screw fixation with or without an additional cage. The surgeons will either use a microscope or magnifying glasses.

Both groups will receive perioperative intravenous antibiotic prophylaxis. Postoperative care and mobilization will follow each hospital's normal practices and routines.

Outcomes

Patient reported outcome measures (PROMs) will be collected preoperatively and at 3 months, 12 months, 2 years, 5 years and 10 years postoperatively. Primary endpoint is at 2-year follow-up. To evaluate the long-term results (5- and 10-year follow-up) we will use the same primary and secondary outcome measurements as at 2-year follow-up. The time schedule for collection of data is shown in Table 3.

The primary outcome is the proportion of responders assessed by the Oswestry Disability Index (ODI) V.2.0 [21, 22]. ODI scores range from 0 to 100, where 100 represent the greatest impairment. Based on former studies [23, 24] and a presently not submitted study from The Norwegian Registry for Spine Surgery (NORSpine), an individual ODI improvement of 30% or more from baseline to follow-up has been chosen as the cut-off for being a responder. Mean scores at follow-up and mean change scores from baseline to follow-up for the ODI scores will be secondary outcomes.

Other secondary outcome measurements are the mean scores at follow-up, the mean changes from baseline to follow-up and the responder rates assessed by the Zürich Claudication Questionnaire [25] [ZCQ; which ranges from 1 to 4 (worst disability)], and by the Numeric Rating Scale for back and leg pain (NRS; which ranges from 0 to 10 (worst pain imaginable)]. Cut-off values for being a responder for ZCQ are defined by Tully et al. [25]. Based on data from NORSpine, the individual thresholds

Table 2 Criteria for inclusion and exclusion for the NORDSTEN/DS trial

Inclusion criteria:	Exclusion criteria:
To be eligible for the study the participants must:	The participants will be excluded from the study if they:
Be over 18 years of age. Understand Norwegian language, spoken and written. Have a spondylolisthesis, with a slip > = 3 mm, verified on standing plain x-rays in lateral view. Have a spinal stenosis in the level of spondylolisthesis, shown on MRI, CT scan or myelogram. Have clinical symptoms of spinal stenosis as neurogenic claudication or radiating pain into the lower limbs, not responding to at least 3 months of qualified conservative treatment. Be able to give informed consent and to respond to the questionnaires.	Are not willing to give written consent. Are participating in another clinical trial that may interfere with this trial. Are ASA- grade > 3. Are older than 80 years. Are not able to fully comply with the protocol, including treatment, follow-up or study procedures (psychosocially, mentally and physically). Have cauda equina syndrome (bowel or bladder dysfunction) or fixed complete motor deficit. Have a slip > = 3 mm in more than one level. Have an isthmic defect in pars interarticularis. Have a fracture or former fusion of the thoracolumbal region. Have had previous surgery in the level of spondylolisthesis. Have a lumbosacral scoliosis of more than 20 degrees verified on AP-view. Have distinct symptoms in one or both legs due to other diseases, e.g. polynevropathy, vascular claudication or osteoarthtritis. Have radicular pain due to a MRI-verified foraminal stenosis in the slipped level, with deformation of the nerve root because of a bony narrowing in the vertical direction.

MRI Magnetic resonance imaging, *CT* Computed tomography, *AP* anterior- posterior, *ASA* American Society of Anesthesiologists

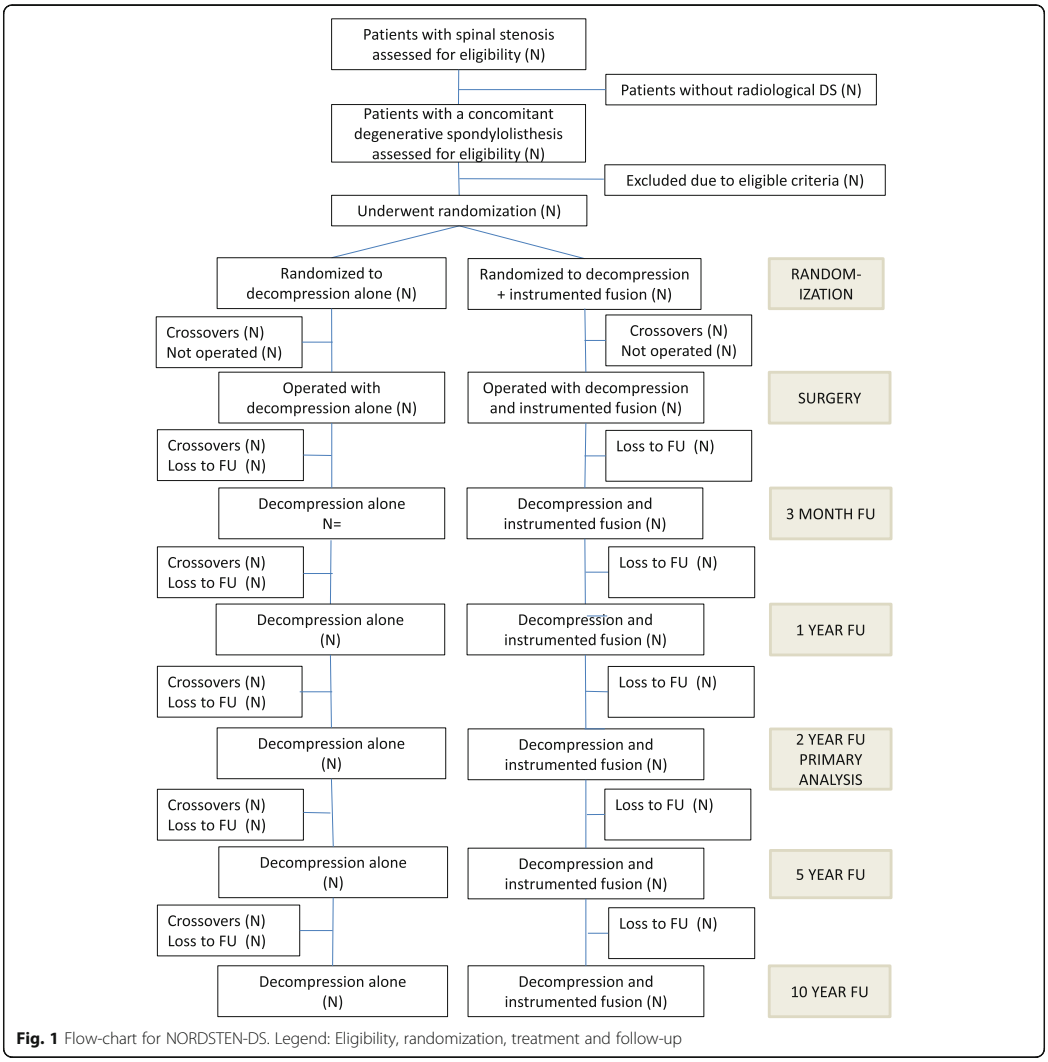


Fig. 1 Flow-chart for NORDSTEN-DS. Legend: Eligibility, randomization, treatment and follow-up

for being a responder are defined as a 40% reduction in the NRS leg pain and a 33% reduction in NRS back pain.

Additional secondary PROMs are the mean scores on the Euroqol 5-D [26] (EQ-5D; which ranges from − 0.6 to 1, with higher scores indicating better quality of life) and the scores on the Global Perceived Effect scale [27] (GPE; a global assessment of patient-rated satisfaction with treatment outcome, with the answers 'completely recovered', 'much improved', 'slightly improved', 'unchanged', 'slightly worse', 'much worse' and 'worse than ever'). For comparing the failure rate between the groups, the proportion of patients replying 'much worse' or 'worse than ever' on the GPE scale will be calculated and compared between the groups.

In addition, we will compare the rates of complications and adverse effects (Table 4), the volume of blood loss, the use of blood transfusion perioperatively and postoperatively, the duration of the surgeries from the skin being opened to when it is closed, and the length of hospital stays. Any new surgery in the lumbosacral column from the time of the index operation to follow-up will be recorded and the reoperation rates will be compared. We will distinguish between an operation at the same level as the primary operation and an operation in a new segment.

For descriptive interpretation, and for the predictor analyses, The Hopkins symptom check list (HSCL-25; a self-reported questionnaire for assessment of psychological variables) [28], data concerning age, gender, education, work, smoking habits, comorbidity, osteoporosis, the American Society of Anesthesiologists (ASA) grade and prior history of spinal surgery will be recorded preoperatively. For radiological evaluations we will assess the grade of spinal stenosis [29], the foraminal stenosis [30], the magnitude of the olisthesis [31], the segmental instability [31], the orientation of the facet joint [32], the amount of facet joint fluid [33], the degree of disc degeneration [34, 35], the disc height in the level of listhesis [36], the lumbar lordosis [37] and the pelvic parameters (the sacral slope, the pelvic tilt and the pelvic incidence) [37]. A CT scan will be performed at the 2 year follow up for assessment of fusion for the DF group [38]. The time schedule for radiological examinations is given in Table 3. The radiological evaluations will be performed by at least one spine surgeon and one radiologist.

Sample size

The sample size calculation for efficacy is based on the hypothesis that the 2-year results for the decompression alone group will be at least as good as those from the fusion group when comparing the proportions of responders in each group. The sample size is computed by using the Blackwelder methodology [39]. Based on data from the Norwegian Spine Register, the proportion of responders for the whole treatment group is expected to be 0.70. Choosing a type 1 error = 0.05, power = 0.80 and non-inferiority limit (δ) = 0.15 gives a sample size of 116. Considering these assumptions and adding 10% for possible dropouts, a total of 128 patients are required in each group.

Recruitment

To ensure a standardized system of enrollment, one or two research coordinators at each hospital manage the practical details regarding registration, collection and further submission of patient data to the central coordinator at the Section for musculoskeletal research (FORMI), Division of neuroscience, at Oslo University Hospital.

Allocation

The computer generated 1:1 randomization is block-permuted and center-stratified. After the patient has signed the informed consent form, the randomization is performed within the 6 weeks before treatment. The computer generated randomization procedure is concealed and administered by the central coordinator at FORMI, and communicated by phone and by email to the local research coordinator. The coordinator documents the result of the randomization in the patient's records and assigns the allocated surgical procedure to the surgeon in charge. The randomization process cannot be influenced by the patients, the investigators, the surgeons or any other persons involved in the study.

Blinding

The treatment given is not blinded for the patients. For analysis and testing of the efficacy variables, the statistician will be blinded for treatment adherence.

Data collection

The study coordinators are responsible for the collection and administration of data at baseline and at 3-month follow-up. Data from 12-month 2-year, 5-year and 10-year follow-up is collected by the central coordinator at FORMI. All data will be stored at the Faculty of Research support, University of Oslo. The data will be inaccessible to the research group until the first analysis at 2-year follow-up.

Statistical methods

The first analyses will be performed 2 years after surgery. Long-term follow-up analyses will be performed at 5 and 10 years after surgery.

Table 3 Time schedule for collection of data for the NORDSTEN/DS trial

	Before operation	Hospital stay	3 months	12 months	2 years	5 years	10 years
X-rays	x		x		x	x	x
MRI scan	x						
CT scan					x		
Demographics	x						
Lifestyles	x						
PROMs	x		x	x	x	x	x
Operation data		x					
Data from hospital stay		x					
Complications, and reoperations		x	x	x	x	x	x

MRI Magnetic resonance imaging, *CT* Computed tomography, *PROMs* Patient reported outcome measures

Table 4 Complications and side effects registered during the hospital stay

Perioperative	Postoperative
Dural tear	Liquor leakage
Nerve root lesion	Superficial infection
Operated on the wrong side	Neurological deterioration
Operated on the wrong level	Hematoma requiring reoperation
Amount of bleeding	Use of blood transfusion
Cardiopulmonary complications	Deep infection
Anaphylactic reaction	Thromboembolic episode
Death	Cardiopulmonary complication
Other	Urological complication
	Wrong level/side revealed postoperatively
	Death
	Other

For the primary objective, the proportion of patients with a reduction in ODI of 30% or more from baseline to 2-year follow-up (responder-rate) is defined as the primary outcome [23, 40]. The null hypothesis (H0) is that the responder rate in the decompression alone group is inferior the responder rate in the decompression and fusion group with an amount of 0.15. H0 will be tested by forming a 95% confidence interval (CI) for the difference of proportions, and H0 will be rejected if the upper limit of the confidence interval (CI) is less than 0.15.

The alternative hypothesis is that the responder rate in the DA group is non- inferior the responder rate in the DF group (Fig. 2).

We have predefined the non- inferior margin to be 0.15 of 1.0, i.e., a 15 percentage difference in the responder rate [41]. With this margin it will be necessary to treat 7 patients or more with fusion in addition to decompression in order to prevent one responder. (Number needed to treat = 1.0:0.15 = 6.67) [42].

The statistical analysis will be done according to intention-to-treat principles (ITT). A sensitivity analysis will be conducted where patient's crossing over from one treatment to another will receive the last score before crossover. To recommend DA, both the ITT and the sensitivity analysis are required to show non-inferiority.

Descriptive statistics, including measures of centrality and variability, will be used to describe the baseline characteristics of the two treatment groups.

The difference in the proportions of responders (the primary outcome) will be estimated with the Newcombe hybrid score CI [43]. Categorical secondary outcomes will be analyzed with Fisher mid-P tests and Newcombe hybrid score intervals. The GPE responses will be analyzed with a proportional odds logistic regression model. We will use linear mixed models to estimate the difference between the treatment groups for the continuous secondary outcomes (all follow-up measurements from inclusion to 2-year follow-up will be included). Because most change from baseline is expected to occur the first three months, the time development in the linear mixed models will be modelled as piecewise linear, with a knot at 3 months. The models will include fixed effects for treatment group, time, and treatment group x time interaction. A random intercept will be used, and – if possible – a random effect for treatment group.

Missing data

For the primary outcome, the primary analysis will be a complete case analysis. If there are patients with missing data in the primary outcome, sensitivity analyses with different imputation scenarios will be performed. The scenarios include all DA patients (with missing data) are responders and none of the DF patients (with missing data) are responders, and vice versa; all DA and all DF patients are responders; all DA and all DF patients are non-responders. Missing data for the continuous secondary outcomes will be handled by the linear mixed

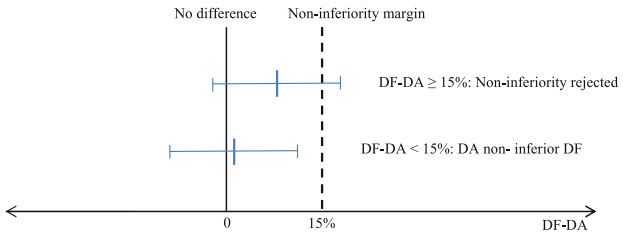


Fig. 2 Test for non-inferiority. Legend: The figure shows two alternative results for the primary outcome. DA and DF indicate the proportion of responders in the decompression alone group and decompression plus instrumented fusion group, respectively. The bars indicate the absolute difference in proportion of responders (DF-DA) with 95% confidence interval (CI) limits. Non-inferiority for DA is shown if the upper limit of the 95% CI for the difference is less than 15%

models, which include all patients with a measurement at at least one time point.

Complete case analyses will be performed on the categorical secondary outcomes. A significance level of 5% will be used throughout.

Analyses of secondary objectives

Predictor analysis The predictor analysis will be performed by use of a pragmatic model-building approach of Hosmer et.al [44]. This method is advocated when risk factor modelling is of interests and not just prediction [45]. Patients treated with decompression alone and decompression with fusion will be analyzed in separate cohorts. For each cohort the following purposefully selected baseline variables will be tested for their association to the primary outcome variable 'responder': 1) Patient age; 2) Gender; 3) Comorbidity (ASA group); 4) Body Mass Index; 5) Smoking; 6) ODI score; 7) NRS back pain score; 8) NRS leg pain score; 9) Hopkins symptom check list (HSCL-25); 10) The magnitude of olisthesis; 11) Segmental instability; 12) Presence of foraminal stenosis; 13) Orientation of the facet joint; 14) Amount of facet joint fluid; 15) Disc degeneration; 16) Disc height in the level of olisthesis; 17) Lumbal lordosis; 18) Pelvic incidence.

From a univariate screening, variables with $P < 0.25$ will be included in the multivariate analyses. Since age and gender will be of interests for clinicians when searching for the best choice of treatment, these variables will be included throughout the multivariate analysis. In the second step, the iterative process, covariates are removed if they are non- significant predictors at the 0.1 alpha level and not a confounder. Confounding is defined as a change in any remaining covariate more than 15% when removing a covariate from the model. The covariates will be deleted in descending rang according to the *p*-value. After deleting and refitting, the model will contain only significant covariates and confounders. In the next step, the covariates not selected from the univariate analysis one by one will be tested for their contributions in the presence of variables from the retained model. If significant at alpha level 0.15 they are included for further fitting of the multivariate model. Finally the model is iteratively reduced as before, but only variables additionally added will be excluded. From the final best fitted model for each treatment group, predicted probabilities of being a responder will be estimated for each combination of the covariates. The risk estimates will be used for building matrixes for an individual's overall risk for being a responder following surgery. Previously, risk matrix models for predicting probability given a set of established predictors has been constructed for other conditions [46, 47].

Cost-utility analysis Cost-utility will be analyzed as the difference in costs between the two treatment groups divided by their difference in Quality adjusted life years (QALYs) gained [48, 49]. QALYs will be estimated by combining EQ-5D index and time, calculating the area under the curve using the trapezoidal method. The results will be presented as an incremental cost-effectiveness ratio (ICER), meaning the cost for each unit of effect (QALY) gained from decompression alone instead of decompression with instrumented fusion. The presentation will be done from a health provider perspective based on data from two-year follow-up.

Clinical monitoring of the trial

The trial is monitored following the Helsinki Declaration, The International Conference on Harmonisation Guideline for Good Clinical Practice (ICH GCP) [50]. An independent monitor affiliated with Møre and Romsdal Health Trust, without influence on the scientific work, will be responsible for the monitoring. Due to the non-regulated ICH GCP guideline for this trial (not including drug intervention) the risk and safety will be safeguarded at the same level as data quality. All informed consent forms will be checked and all registrations of serious events will be monitored. According to the monitoring plan selected variables will be checked. All hospitals will be visited regularly. Adapted versions of the 'Investigator's Site File (ISF)' and the 'Trial Master File (TMF)' will be checked for essential documents during the trial. Queries and deviations will be recorded and reported, and the coordinator at the responsible hospital will have two months to send a written report with the required corrections to the monitor.

Interim analysis and stopping rules

Due to ethical considerations in agreement with the Norwegian Committee for Medical and Health Research Ethics Midt, an interim analysis for safety will be performed when 75 patients in each group have completed the 12-month follow- up. If one of the proposed stop criteria is fulfilled the study will be terminated:

1. The proportion of patients needing reoperation due to any condition in the operated level(s) is statistically significantly higher in one of the groups.
2. The proportion of responders in the DF group, assessed by the primary outcome measure, is higher than in the DA group by an amount of 0.20.

The interim analysis will be conducted by an independent statistician blinded for treatment adherence. Only data on reoperations and on the primary outcome measure (ODI) will be available to the statistician. The statistician will inform the steering committee, via the central coordinator, whether the study can be continued or not. Further

information about the analysis will not be disclosed and will not be available to anyone until the main analysis at 2-year follow-up.

Ethics and dissemination

The protocol has been approved by the Norwegian Committee for Medical and Health Research Ethics Midt (2013/366).

Storage of data is approved by the Norwegian Data Inspectorate. Written informed consent is obtained from the patients. The project is in accordance with the Helsinki Declaration.

None of the principal investigators have any financial or other competing conflicts of interest.

Trial results will be communicated at national and international conferences and published in well-recognized journals.

Discussion

The rationale, design and method for this prospective randomized clinical multi-center trial on patients with LDS are presented in the current protocol.

We have chosen a non-inferiority design in order to investigate whether clinical outcomes for decompression alone are not worse than decompression with fusion by more than an acceptable amount. Superiority for decompression alone is not considered to be necessary; it would be an additional benefit [20].

The present study will be the largest powered study comparing decompression alone and decompression with instrumented fusion in a randomized setting. It is designed and powered to provide Level 1 evidence for whether decompression alone can be advocated as the preferred method for surgical treatment of DS or not. We also aim to investigate whether patients can be assigned to the most appropriate surgical method. Finally, results at 5- and 10-year follow-up will provide high level evidence for long-time results for the two methods.

We anticipate enclosing the inclusion by the end of 2017.

Abbreviations

CT: Computed tomography; DA: Decompression alone; DF: Decompression with an additional fusion; DS: Degenerative Spondylolisthesis; EQ-5D: EuroQol 5-dimensional questionnaire utility index; FU: Follow-up; GPE: Global perceived effect; HSCL-25: Hopkins symptom check list; LSS: Lumbar spinal stenosis; MRI: Magnetic resonance imaging; NNT: Number needed to treat; NRS: Numeric rating scale; ODI: Oswestry disability index; PROMs: Patient reported outcome measures; SAP: Statistical Analysis Plan; ZCQ: Zurich claudication questionnaire –score

Acknowledgements

Thanks to Eira Kathleen Ebbs for linguistic assistance in writing the manuscript. Thanks to study coordinators at the participating hospitals and the central coordinators at FORMI, for an extensive contribution in keeping track of the patients in the study.

Amendment

From the start of inclusion (April 15, 2014) patients with ODI scores of less than 25 were excluded. Due to the experiences of participating surgeons, a considerable number of the patients suffering from leg and back pain, but with ODI of less than 25, were found eligible for operation but not for inclusion in the study. To enhance the external validity of the study, the steering committee decided that from date 29th August 2015, the patients should not be excluded due to ODS- score lower than 25.

Funding

Helse Vest RHF (the Western Regional Health Authority) has provided funds for the present study. The funder has no influence on study design, management and interpretation of data or the decision to submit data.

Availability of data and materials

All data will be stored at the Faculty of Research support, University of Oslo. The data will be inaccessible to the research group until the first analysis at 2-year follow-up. Storage of data is approved by the Norwegian Data Inspectorate.

Authors' contributions

IMA, EH, MF, FR, TS, KS, OG, JIB, JA, GL, KI and CH have been involved in planning the study and in drafting the manuscript. CH was a major contributor in writing the manuscript. All authors read and approved the final manuscript. All authors meet the ICMJE guidelines for authorships.

Ethics approval and consent to participate

The patients are given verbal and written information about the study. If willing to participate, the patients sign an informed consent form. If a patient does not want to participate, he/she will receive normal care and be treated following the hospital's established procedures.
The protocol has been approved by the Norwegian Committee for Medical and Health Research Ethics Midt (2013/366).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Author details

[1]Kysthospitalet in Hagevik, Orthopedic Clinic, Haukeland University Hospital, Hagavik, N- 5217 Bergen, Norway. [2]Department of Clinical Medicine, University of Bergen, N- 5007 Bergen, Norway. [3]Department of Orthopedic Surgery, Ålesund Hospital, Møre and Romsdal Hospital Trust, N-6026 Ålesund, Norway. [4]Oslo Centre for Biostatistics and Epidemiology, Research Support Services, Oslo University Hospital, N-0424 Oslo, Norway. [5]Department of Neurosurgery, University Hospital of Northern Norway, N-9019 Tromsø, Norway. [6]Department of Clinical Medicine, University of Tromsø - The Arctic University of Norway, N-9019 Tromsø, Norway. [7]The Norwegian Registry for Spine Surgery (NORspine), Northern Norway Regional Health Authority, N-9038 Tromsø, Bodø, Norway. [8]Research and Communication Unit for Musculoskeletal Health (FORMI), Oslo University Hospital, N-0424 Oslo, Oslo, Norway. [9]Department of Orthopedics, Akershus University Hospital, N-1474 Lørenskog, Oslo, Norway. [10]Department of Physical Medicine and Rehabilitation, Oslo University Hospital, N-0424 Oslo, Norway. [11]Department of Research, Levanger Hospital, Nord-Trøndelag Hospital Trust, N-7600 Levanger, Norway. [12]Department of Orthopedic Surgery, Innlandet Hospital Trust, N-2609 Lillehammer, Lillehammer, Norway. [13]Division of Orthopaedic Surgery, Oslo University Hospital, N-0424 Oslo, Norway.

Received: 17 November 2017 Accepted: 18 December 2018
Published online: 05 January 2019

References

1. Farfan HF (2012) - The pathological anatomy of degenerative spondylolisthesis. A cadaver study. - spine (Phila pa 1976)1980;5(5):412-8:Oct.

2. Fitzgerald JA, Newman PH. Degenerative spondylolisthesis. J Bone Joint Surg Br. 1976;58:184–92.
3. Watters WC 3rd, Bono CM, Gilbert TJ, Kreiner DS, Mazanec DJ, Shaffer WO, Baisden J, Easa JE, Fernand R, Ghiselli G, Heggeness MH, Mendel RC, O'Neill C, Reitman CA, Resnick DK, Summers JT, Timmons RB, Toton JF, North American Spine S. An evidence-based clinical guideline for the diagnosis and treatment of degenerative lumbar spondylolisthesis. Spine J. 2009;9:609–14. https://doi.org/10.1016/j.spinee.2009.03.016.
4. Herkowitz HN, Kurz LT. Degenerative lumbar spondylolisthesis with spinal stenosis. A prospective study comparing decompression with decompression and intertransverse process arthrodesis. J Bone Joint Surg Am. 1991;73:802–8.
5. Martin CR, Gruszczynski AT, Braunsfurth HA, Fallatah SM, O'Neil J, Wai EK. The surgical management of degenerative lumbar spondylolisthesis: a systematic review. Spine. 2007;32:1791–8. https://doi.org/10.1097/BRS.0b013e3180bc219e.
6. Resnick DK, Watters WC 3rd, Sharan A, Mummaneni PV, Dailey AT, Wang JC, Choudhri TF, Eck J, Ghogawala Z, Groff MW, Dhall SS, Kaiser MG. Guideline update for the performance of fusion procedures for degenerative disease of the lumbar spine. Part 9: lumbar fusion for stenosis with spondylolisthesis. J Neurosurg Spine. 2014;21:54–61. https://doi.org/10.3171/2014.4.spine14274.
7. Steiger F, Becker HJ, Standaert CJ, Balague F, Vader JP, Porchet F, Mannion AF. Surgery in lumbar degenerative spondylolisthesis: indications, outcomes and complications. A systematic review. Eur Spine J. 2014;23:945–73. https://doi.org/10.1007/s00586-013-3144-3.
8. Ghogawala Z, Dziura J, Butler WE, Dai F, Terrin N, Magge SN, Coumans JV, Harrington JF, Amin-Hanjani S, Schwartz JS, Sonntag VK, Barker FG 2nd, Benzel EC. Laminectomy plus fusion versus laminectomy alone for lumbar spondylolisthesis. N Engl J Med. 2016;374:1424–34. https://doi.org/10.1056/NEJMoa1508788.
9. Chang HS, Fujisawa N, Tsuchiya T, Oya S, Matsui T. Degenerative spondylolisthesis does not affect the outcome of unilateral laminotomy with bilateral decompression in patients with lumbar stenosis. Spine. 2014;39:400–8. https://doi.org/10.1097/BRS.0000000000000161.
10. Forsth P, Michaelsson K, Sanden B (2013) Does fusion improve the outcome after decompressive surgery for lumbar spinal stenosis?: a two-year follow-up study involving 5390 patients. Bone joint J 95-B:960-965. 95-B/7/960; https://doi.org/10.1302/0301-620X.95B7.30776.
11. Austevoll IM, Gjestad R, Brox JI, Solberg TK, Storheim K, Rekeland F, Hermansen E, Indrekvam K, Hellum C. The effectiveness of decompression alone compared with additional fusion for lumbar spinal stenosis with degenerative spondylolisthesis: a pragmatic comparative non-inferiority observational study from the Norwegian registry for spine surgery. Eur Spine J. 2016. https://doi.org/10.1007/s00586-016-4683-1.
12. Forsth P, Olafsson G, Carlsson T, Frost A, Borgstrom F, Fritzell P, Ohagen P, Michaelsson K, Sanden B. A randomized, controlled trial of fusion surgery for lumbar spinal stenosis. N Engl J Med. 2016;374:1413–23. https://doi.org/10.1056/NEJMoa1513721.
13. Weinstein J, Pearson A. Fusion in degenerative spondylolisthesis becomes controversial … again. Evid Based Med. 2016;21:148–9. https://doi.org/10.1136/ebmed-2016-110474.
14. Peul WC, Moojen WA. Fusion for lumbar spinal stenosis--safeguard or superfluous surgical implant? N Engl J Med. 2016;374:1478–9. https://doi.org/10.1056/NEJMe1600955.
15. Pearson AM. Fusion in degenerative spondylolisthesis: how to reconcile conflicting evidence. Journ Spine Surg. 2016;2:143–5.
16. Joaquim AF, Milano JB, Ghizoni E, Patel AA. Is there a role for decompression alone for treating symptomatic degenerative lumbar spondylolisthesis?: a systematic review. Clin Spine Surg. 2016;29:191–202. https://doi.org/10.1097/bsd.0000000000000357.
17. Angevine PD, Berven S. Health economic studies: an introduction to cost-benefit, cost-effectiveness, and cost-utility analyses. Spine. 2014;39:S9–15. https://doi.org/10.1097/brs.0000000000000576.
18. Hermansen E, Austevoll IM, Romild UK, Rekeland F, Solberg T, Storheim K, Grundnes O, Aaen J, Brox JI, Hellum C, Indrekvam K. Study-protocol for a randomized controlled trial comparing clinical and radiological results after three different posterior decompression techniques for lumbar spinal stenosis: the spinal stenosis trial (SST) (part of the NORDSTEN study). BMC Musculoskelet Disord. 2017;18:121. https://doi.org/10.1186/s12891-017-1491-7.
19. Agha RA, Altman DG, Rosin D. The SPIRIT 2013 statement--defining standard protocol items for trials. Int J Surg. 2015;13:288–91. https://doi.org/10.1016/j.ijsu.2014.12.007.
20. Piaggio GF, – Elbourne DR FAU - Altman D, – Altman DG FAU - Pocock S, – Pocock SJ FAU - Evans S, Evans SJ (2010) - Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. - Jama 2006;295(10):1152–60:60.
21. Fairbank JC, Couper J, Davies JB, O'Brien JP. The Oswestry low back pain disability questionnaire. Physiotherapy. 1980;66:271–3.
22. Baker DJ, Pynsent PB, J F (1989) The Oswestry disability index revisited: its reliability, repeatability and validity, and a comparison with the St Thomas's disability index. Roland MO, Jenner JR, eds New approaches to rehabilitation and education Manchester: Manchester University Press: 174–186.
23. Dworkin RH, Turk DC, Wyrwich KW, Beaton D, Cleeland CS, Farrar JT, Haythornthwaite JA, Jensen MP, Kerns RD, Ader DN, Brandenburg N, Burke LB, Cella D, Chandler J, Cowan P, Dimitrova R, Dionne R, Hertz S, Jadad AR, Katz NP, Kehlet H, Kramer LD, Manning DC, McCormick C, McDermott MP, McQuay HJ, Patel S, Porter L, Quessy S, Rappaport BA, Rauschkolb C, Revicki DA, Rothman M, Schmader KE, Stacey BR, Stauffer JW, von Stein T, White RE, Witter J, Zavisic S. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. J Pain. 2008;9:105–21. https://doi.org/10.1016/j.jpain.2007.09.005.
24. Ostelo RW, Deyo RA, Stratford P, Waddell G, Croft P, Von Korff M, Bouter LM, de Vet HC. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. Spine. 2008;33:90–4. https://doi.org/10.1097/BRS.0b013e31815e3a10.
25. Tuli SK, Yerby SA, Katz JN. Methodological approaches to developing criteria for improvement in lumbar spinal stenosis surgery. Spine. 2006;31:1276–80. https://doi.org/10.1097/01.brs.0000217615.20018.6c.
26. (1990) EuroQol--a new facility for the measurement of health-related quality of life. Health policy (Amsterdam, Netherlands) 16:199–208.
27. Kamper SJ, Ostelo RW, Knol DL, Maher CG, de Vet HC, Hancock MJ (2010) Global perceived effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. J Clin Epidemiol 63:760-766. Doi: S0895-4356(09)00304-7.
28. Derogatis LR FAU, Lipman RS FAU, Rickels KF, Uhlenhuth EH FAU, Covi L, Derogatis LR FAU, Lipman RS FAU, Rickels KF, Uhlenhuth EH FAU, Covi L (2000) - The Hopkins symptom checklist (HSCL). A measure of primary symptom dimensions - the Hopkins symptom checklist (HSCL): a self-report symptom inventory Mod Probl Pharmacopsychiatry 1974;7(0):79–110:110.
29. Schizas C, Theumann N, Burn A, Tansey R, Wardlaw D, Smith FW, Kulik G. Qualitative grading of severity of lumbar spinal stenosis based on the morphology of the dural sac on magnetic resonance images. Spine. 2010;35:1919–24. https://doi.org/10.1097/BRS.0b013e3181d359bd.
30. Lee S, Lee JW, Yeom JS, Kim KJ, Kim HJ, Chung SK, Kang HS (2010) A practical MRI grading system for lumbar foraminal stenosis. AJR Am J Roentgenol 194:1095-1098. doi: 194/4/1095;https://doi.org/10.2214/AJR.09.2772 [doi].
31. Dupuis PR, Yong-Hing K, Cassidy JD, Kirkaldy-Willis WH Radiologic diagnosis of degenerative lumbar spinal instability. Spine. 1985; 10:262–276.
32. Berlemann UF, Jeszenszky DJ FAU, Buhler DW FAU, Harms J (2005) - Facet joint remodeling in degenerative spondylolisthesis: an investigation of joint orientation and tropism. - Eur Spine J 1998;7(5):376-80:80.
33. Cho IY, Park SY, Park JH, Suh SW, Lee SH. MRI findings of lumbar spine instability in degenerative spondylolisthesis. J Orthop Surg (Hong Kong). 2017;25:2309499017718907. https://doi.org/10.1177/2309499017718907.
34. Pfirrmann CW, Metzdorf A, Zanetti M, Hodler J, Boos N. Magnetic resonance classification of lumbar intervertebral disc degeneration. Spine (Phila Pa 1976). 2001;26:1873–8.
35. 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 166:193–199. https://doi.org/10.1148/radiology.166.1.3336678.
36. Masharawi Y, Kjaer P, Bendix T, Manniche C, Wedderkopp N, Sorensen JS, Peled N, Jensen TS. The reproducibility of quantitative measurements in lumbar magnetic resonance imaging of children from the general population. Spine. 2008;33:2094–100. https://doi.org/10.1097/BRS.0b013e31817f19f7.

37. Schwab F, Lafage V, Patel A, Farcy JP. Sagittal plane considerations and the pelvis in the adult patient. Spine. 2009;34:1828–33. https://doi.org/10.1097/BRS.0b013e3181a13c08.
38. Bridwell KH, Lenke LG, McEnery KW, Baldus C, Blanke K. Anterior fresh frozen structural allografts in the thoracic and lumbar spine. Do they work if combined with posterior fusion and instrumentation in adult patients with kyphosis or anterior column defects? Spine. 1995;20:1410–8.
39. Blackwelder WC, Chang MA. Sample size graphs for "proving the null hypothesis". Control Clin Trials. 1984;5:97–105.
40. Katz NP, Paillard FC, Ekman E. Determining the clinical importance of treatment benefits for interventions for painful orthopedic conditions. J Orthop Surgery Res. 2015;10:24. https://doi.org/10.1186/s13018-014-0144-x.
41. Blumenthal S, McAfee PC, Guyer RD, Hochschuler SH, Geisler FH, Holt RT, Garcia R Jr, Regan JJ, Ohnmeiss DD. 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;30:1565–75 discussion E1387-1591.
42. Katz N, Paillard FC, Van Inwegen R. A review of the use of the number needed to treat to evaluate the efficacy of analgesics. J Pain. 2015;16:116–23. https://doi.org/10.1016/j.jpain.2014.08.005.
43. Fagerland MW, Lydersen S, Laake P (2011) Recommended confidence intervals for two independent binomial proportions. Stat Methods Med Res. Doi: 0962280211415469 ;https://doi.org/10.1177/0962280211415469 [doi].
44. Hosmer DW, Lemeshow S. Applied logistic regression. 2nd ed. New York: John Wiley & Sons, Inc, Hoboken; 2000.
45. Bursac Z, Gauss CH, Williams DK, Hosmer DW. Purposeful selection of variables in logistic regression. Source Code Biol Med. 2008;3:17. https://doi.org/10.1186/1751-0473-3-17.
46. Conroy RM, Pyorala K, Fitzgerald AP, Sans S, Menotti A, De Backer G, De Bacquer D, Ducimetiere P, Jousilahti P, Keil U, Njolstad I, Oganov RG, Thomsen T, Tunstall-Pedoe H, Tverdal A, Wedel H, Whincup P, Wilhelmsen L, Graham IM. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. Eur Heart J. 2003;24:987–1003.
47. Solberg IC, Hoivik ML, Cvancarova M, Moum B. Risk matrix model for prediction of colectomy in a population-based study of ulcerative colitis patients (the IBSEN study). Scand J Gastroenterol. 2015;50:1456–62. https://doi.org/10.3109/00365521.2015.1064991.
48. Drummond M (2005) Methods for the economic evaluation of health care Programmes. 3rd ed Oxford, NY : Oxford medical publications, Oxford University press;
49. Lonne G, Johnsen LG, Aas E, Lydersen S, Andresen H, Ronning R, Nygaard OP. Comparing cost-effectiveness of X-stop with minimally invasive decompression in lumbar spinal stenosis: a randomized controlled trial. Spine. 2015;40:514–20. https://doi.org/10.1097/brs.0000000000000798.
50. ICH, Harmonised, Tripartite, Guideline (2014) Guideline for good clinical practice E6(R1). 1996. http://www.ich.org/products/guidelines/efficacy/efficacy-single/article/good-clinical-practice.html.

Graphic design: Communication Division, UiB / Print: Skipnes Kommunikasjon AS

uib.no

ISBN: 9788230869468 (print)
9788230845196 (PDF)