Doktoravhandlinger ved NTNU, 2023:210

Ole Kristian Alhaug

Lumbar spinal stenosis, assessing failure and worsening after surgery. Identifying predictive factors with critical use of data from a national spine registry (NORspine).

Doktoravhandling

NTNU
Norges teknisk-naturvitenskapelige universitet
Avhandling for graden
philosophiae doctor
Fakultet for medisin og helsevitenskap
Institutt for nevromedisin og bevegelsesvitenskap

NTNU
Kunnskap for en bedre verden

Ole Kristian Alhaug

Lumbar spinal stenosis, assessing failure and worsening after surgery. Identifying predictive factors with critical use of data from a national spine registry (NORspine).

Avhandling for graden philosophiae doctor

Trondheim, August 2023

Norges teknisk-naturvitenskapelige universitet
Fakultet for medisin og helsevitenskap
Institutt for nevromedisin og bevegelsesvitenskap

NTNU
Kunnskap for en bedre verden

NTNU
Norges teknisk-naturvitenskapelige universitet

Avhandling for graden philosophiae doctor

Fakultet for medisin og helsevitenskap
Institutt for nevromedisin og bevegelsesvitenskap

ISBN 978-82-326-7122-9 (trykt utg.)
ISBN 978-82-326-7121-2 (elektr. utg.)
ISSN 1503-8181 (trykt utg.)
ISSN 2703-8084 (online ver.)

Doktoravhandlinger ved NTNU, 2023:210

Trykket av NTNU Grafisk senter

Ph.D. dissertation

Lumbar spinal stenosis, assessing failure and worsening after surgery. Identifying predictive factors with critical use of data from a national spine registry (NORspine).

Ole Kristian Alhaug

Table of contents

1 Acknowledgements

We started planning this work in the autumn of 2016, applying for funding and formal approvals from the authorities. My mentor Greger Lønne already had many research questions about lumbar spinal stenosis patients. As an excellent researcher and experienced clinician with a broad network within the spine community, he supported my application. Innlandet Hospital Trust was always supportive and flexible in making this PhD possible.

After starting the work in the autumn of 2018, it became clear that the entire study group was skilled and more supportive than I had ever hoped. Dr Filip Dolatowski always answers all types of questions. Dr Tore Solberg has outstanding competence in medical research. As head of the NORspine registry, he has been central in all phases of this project. Professor Milada Cvarkcova and statistician Jurate Saltyte Benth have been of unevaluable help in planning and performing the statistics. Dr Ivar Austevoll was originally not a part of the study group, but as an experienced spine researcher, he has been of great help from the beginning. Ivar had first-hand experience with some of the topics in my thesis. Dr Simran Kaur joined the group to assist in some of the studies and wrote one of the articles; her enthusiasm is truly inspiring.

As a clinician working in a team at a large hospital, I always rely on my colleagues. Working part-time at the hospital has forced my colleagues to take a more significant part of the clinical workload for many years. I hope my interest in spine surgery and research has had some effect on the daily work and enthusiasm, but this can never justify the enormous effort my clinical colleagues have made to make my studies possible.

Early in my PhD education, I was lucky to be invited to one of the research groups in Innlandet Hospital Trust; the Research Centre for Age-related Functional Decline and Disease. This group consists of a variety of new and experienced PhD students, post-docs, and experienced senior researchers. This environment has been of great importance to me; the opportunity to discuss challenges with peer PhD students or more experienced ones has been rewarding. The social aspect of joining a research centre was also essential.

Research is not 9 to 5 work, sometimes it can be quiet, and sometimes everything comes as a crash simultaneously (courses, deadlines, and revisions). I have spent many nights, many weekends and many holidays working on this project, and this would never have been possible without the support

of my family. Liv, Arn, Ask and Eir, you have my most heart-filled thanks for standing out with me during these years.

2 Abstract

2.1 Norsk sammendrag

Spinal stenose er en vanlig lidelse som skyldes trang ryggmargskanal og karakteriseres av smerter i rygg og bein og redusert gangfunksjon. Operasjon er ofte nødvendig og spinal stenose er den hyppigste årsaken til ryggkirurgi i Norge (3, 23). Resultatene etter kirurgi er noe sprikende: de fleste blir bedre, noen blir ikke bedre, og enkelte blir verre (5, 6, 7).

Det er utfordrende å måle resultat etter behandling for smertetilstander fordi det ikke finnes klare konkrete endepunkt. Pasient-rapporterte resultater er sentrale, man kan bruke smerteskalar eller spørreskjema på funksjon og livskvalitet. Fortolkning av svar i skala-form kan være krevende, det fordres en viss endring på skalaene for at endringen skal være klinisk relevant. Man kan lette fortolkningen med å lage kategorier der pasientene klassifiseres som enten bedre, uendret eller verre.

Vi har brukt data fra Norsk kvalitetsregister for ryggkirurgi og analysert pasienter operert for spinal stenose. Registeret inneholder data om pasientforhold og plager før operasjon, operasjonstekniske forhold og resultater 3-og 12 måneder etter operasjon. Gjennom dette har vi tilegnet oss ny kunnskap om spinal stenose pasienter.

Registerdata er beheftet med flere usikkerhetsområder, mange pasienter faller fra og svarer ikke på oppfølgingene, og data kan i tillegg bli feilregistrert. Vi har derfor undersøkt kvaliteten på registerdata som ble brukt i denne doktorgradsavhandlingen.

Det er kjent at ikke alle pasienter blir kvitt plagene etter kirurgi for spinal stenose, og vi fant at om lag 20% rapporterte at plagene var uendret eller verre etter kirurgi. Videre fant vi de grenseverdiene som definerte mislykket kirurgi (uendret eller verre) og forverring på de mest brukte skalaene med størst nøyaktighet.

Vi testet samsvar av registerdata ved å kontrollere opp mot journaldata og fant at datakvaliteten i registeret var vekslende. Pasientrapporterte data og operasjonstekniske faktorer hadde høyt samsvar, mens andre helseforhold og komplikasjoner hadde dårligere samsvar med pasientjournalen. Pasienter som ikke svarte på oppfølgingsskjemaer fra registeret skilte seg noe fra de som svarte; de var litt yngre og oftere røykere. Resultatene etter operasjon var like i de to gruppene.

Den vanligste komplikasjonen til kirurgi for spinal stenose er rift på nervehinnen, dette medfører lekkasje av spinalvæske, eksponering av nervetråder og noen ganger behov for reoperasjon og forlenger sengeleie. Vi fant noe dårligere resultater etter operasjon hos pasienter som fikk rift på nervehinnen.

Det kan være vanskelig å beslutte om kirurgi er riktig for den enkelte pasient. Vi identifiserte noen faktorer som øker risikoen for mislykket kirurgi og forverring (alder over 70 år, tidligere ryggkirurgi og ryggsmerter over 12 mnd., samt noen sosioøkonomiske variabler). Disse faktorene kan bidra til bedre pasient informasjon og slik gi støtte til beslutning om operasjon eller ikke operasjon.

Vi håper våre resultater er nyttige for klinikere og at de bidrar til bedre informasjon til pasienter samt gode behandlingsvalg. Vi håper også resultatene kan gi grunnlag for videre forskning på ryggkirurgi.

2.2 English abstract

Background

Results after surgery for lumbar spinal stenosis (LSS) vary; most patients improve, but some do not, and some even worsen. Some patients also suffer from complications. Previous studies have identified certain factors that may predict outcomes after surgery for LSS. Development in surgical technique may have reached a ceiling because new techniques fail to prove better; this emphasizes focus on careful patient selection to improve the overall results.

National medical registries collect a large number of data and reflect daily practice. Because of the large number of participants, registry studies are optimal for studying complications of surgery. However, registry data are vulnerable to wrong recordings and loss of follow-up. Hence, registry data should be assessed for bias before conclusions are drawn.

Methods

We reviewed patients operated on for LSS in Norway for ten years (2007-2017). Prospectively collected data from the NORspine registry was the foundation of the observational studies included in the thesis. We also supplemented registry data with data from patient records and performed a cross-sectional study.

We included patients treated over two years from four hospitals to assess data accuracy. Data was re-captured from electronic patient records, and we assessed the agreement between the two data sources using kappa statistics.

To assess potential bias due to loss to follow-up, we compared baseline variables between patients completing follow-up and those who did not. We also contacted patients lost to follow-up to see if they reported different clinical outcomes. We used simple descriptive statistics and compared baseline data and clinical outcomes between the groups with student T-tests.

We defined criteria for failure and worsening using a transition scale (Global Perceived Effect (GPE)) as an external anchor and receiver operating characteristic (ROC) curve analyses to identify the best cut-offs on PROMs commonly used to assess the effect of spine surgery.

We also studied if a dural tear affected the clinical outcome, defined as failure or worsening, using logistic regression analyses and adjusting for possible confounding factors.

Finally, we tried to identify variables that could predict failure and worsening using multiple logistic regression analyses with the cut-offs identified earlier in our project. We selected baseline variables with acceptable accuracy according to an early part of our project.

Results

The study population comprised 11873 patients, and 8919 (75%) completed 12 months of follow-up. We reviewed 474 patient records to assess NORspine accuracy and the impact of loss to follow-up.

Patient-recorded variables and surgeon-reported surgical details displayed moderate to good accuracy; however, surgeon-reported complications and comorbidity were underreported. Patients lost to follow-up were younger and, more often, were smokers. However, there were no statistically significant differences in clinical outcomes.

The following PROM cut-offs most accurately defined patient-reported failure (and worsening): ODI final score of more than 31 (39), ODI percentage improvement of less than 20% (9%) and ODI improvement of less than 8 (4) points. These cut-offs had good to excellent accuracies (AUC= 0.86-0.91).

Dural tears occurred in nearly 5%. Patients who suffered a dural tear increased the odds of failure (and worsening) with an odds ratio of 1,45 (1,50).

After LSS surgery, a proportion of 33 % was defined as failure and 22 % as worse. Age over 70 years, previous spinal surgery, and duration of back pain over 12 months were essential baseline variables associated with failure and worsening (Odds ratio 1,85 – 2,21); socioeconomic factors also affected the odds for failure and worsening (OR 1,26 – 1,67).

Conclusions

There are concerns regarding data quality in the spine registry; data should be used and interpreted with care. Patients lost to follow-up reported similar clinical outcomes as those who completed follow-up, and missing data from loss to follow-up can most likely be treated as missing at random. Cut-offs for failure and worsening are accurate and can be used in future research and clinical work. LSS patients over 70 years, with previous spine surgery and duration of back pain over 12 months, had increased odds for failure and worsening; this could aid in patient selection.

3 Lists

3.1 List of papers

Paper 1

Alhaug OK, Kaur S, Dolatowski F, Småstuen MC, Solberg TK, Lønne G. **Accuracy and agreement of national spine registry data for 474 patients compared to corresponding electronic patient records**. Eur Spine J. 2022 Mar;31(3):801-811. doi: 10.1007/s00586-021-07093-8. Epub 2022 Jan 6. PMID: 34989877.

Paper 2

Kaur S, Alhaug OK, Dolatowski FC, Solberg TK, Lønne G: **Characteristics and outcomes of patients who did not respond to a national spine surgery registry.** Accepted BMC Musculoskeletal Disorders in 23.th February 2023.

Paper 3

Alhaug OK, Dolatowski FC, Solberg TK, Lønne G. **Criteria for failure and worsening after surgery for lumbar spinal stenosis: a prospective national spine registry observational study**. Spine J. 2021 Sep;21(9):1489-1496. doi: 10.1016/j.spinee.2021.04.008. Epub 2021 Apr 17. PMID: 33848690.

Paper 4

Alhaug OK, Dolatowski F, Austevoll I, Mjønes S, Lønne G. **Incidental dural tears associated with worse clinical outcomes in patients operated for lumbar spinal stenosis**. Acta Neurochir (Wien). 2022 Nov 18. doi: 10.1007/s00701-022-05421-5. Epub ahead of print. PMID: 36399189.

Paper 5

Alhaug OK, Dolatowski FC, Solberg TK, Lønne G. **Predictors for failure after surgery for LSS, a prospective observational study**. Spine J. 2022 Nov 4:S1529-9430(22)00992-5. doi: 10.1016/j.spinee.2022.10.010. Epub ahead of print. PMID: 36343913.

3.2 List of abbreviations

ASA American society of anesthesiologists

BMI Body Mass Index

CI confidence interval

CT Computed tomography

EPR electronic patient record

GPE Global Perceived Effect

LBP Low back pain

LSS Lumbar Spinal stenosis

MIS Mini Invasive Surgery

MRI Magnetic resonance imaging

NORspine Norwegian Registry for Spine Surgery

NRS Numeric rating Scale

NSAIDs Non-steroid anti-inflammatory drugs

ODI Oswestry Disability Index

OR odds ratio

PPC Proportion Correctly Classified

PROM patient reported outcome scale

RCT randomized controlled trial

UNN University hospital in Northern Norway

ZCQ Zurich claudication questionnaire

3.3 Project group

PhD candidate

Ole Kristian Alhaug. MD, Orthopedic consultant, Akershus University Hospital.

Main supervisor

Greger Lønne. MD, PhD, Orthopedic consultant, Innlandet Hospital Trust.

Supervisor

Tore K Solberg. MD, PhD, Professor. Neurosurgeon, University hospital of North Norway.

Supervisor

Filip C Dolatowski. MD, PhD. Orthopedic consultant, Oslo University Hospital.

Statistician

Milada Cvancarova Småstuen. Professor, Oslo Metropolitan University.

4 Introduction - why the thesis?

4.1 Background lumbar spinal stenosis

Lumbar Spinal Stenosis (LSS) is a common disorder contributing to a large proportion of all spinal surgery (1, 2). The prevalence of clinical LSS is estimated to be 11 % and increases with age (3, 4). The clinical results after surgery vary; 62-75% report success (5, 6, 7, 8, 9, 10, 11). Former studies have focused on success and improvement, although a considerable proportion of patients do not improve, and some get worse or experience complications from surgery. Furthermore, spinal surgery imposes high costs on society. Avoiding unnecessary surgery is essential; hence, optimizing patient selection is central. From this point of view, there is a need to explore non-success after surgery for LSS. Which patients are at risk of failure or worsening? What is the significance of surgical complications?

Medical registries have become more important in clinical research during the last decades. However, the quality of registry data must be high to minimize the risk of bias; possible pitfalls are low coverage, poor completeness, high loss of follow-up, and compromised accuracy and reliability. Some quality domains are well explored, but others need systematic investigation.

In this thesis, we aimed to explore failure after surgery for LSS using registry data. A critical review of the data we used in the clinical observational studies was natural; was NORspine data sufficient to answer our research questions?

4.2 Anatomy

The lumbar spine consists of five vertebrae connected by soft tissue (intervertebral disc, ligaments), joints (facet joints), and musculature; hence, the spine is a flexible and long structure. The nerves pass through a canal posterior to the vertebral bodies; the canal is partly made of bone and partly of soft tissue, and the standard diameter is 15-23mm (12). The neural structures lie within the thecal sac, surrounded by a membrane, the dura.

According to an evidence-based guideline from 2013 and clinical guidelines for LSS published by the North American Spine Society in 2011, LSS is a disorder where degenerative changes in the spine narrow the spinal canal and compromise the nerves and vascular structures (13. 14). LSS is a progressive disorder involving all types of tissue in the spine (15). Degeneration leads to ticker ligaments, and ligament hypertrophy is “the main aetiology” of LSS (16); a degenerative ligament consists of a lower proportion of elastic fibres and more fibrotic tissue (17). As part of the degeneration, the height of the intervertebral discs decreases, and the intervertebral discs and ligaments can bulge into the spinal canal (Figure 1, 2, 3). As the disc height decrease, the facet joints become incongruent and can develop osteoarthritis and exostoses. The joint capsule and exostoses can also bulge into the spinal canal, compressing the nerves. Furthermore, the joint capsule can develop cysts that can compromise the nerve roots. Additionally, vertebral fractures (osteoporotic or traumatic) can lead to deformation of the spinal canal with less space for the nerves.
LSS causes neurologic symptoms in the legs, such as pain, numbness, and weakness. LSS often result in decreased walking capability, and LSS can cause back pain. The spine's position is essential as extending the spine (standing or walking) leads to a narrower spinal canal and more symptoms. On the contrary, flexing forward results in an increased space in the spinal canal and relieves symptoms (18).

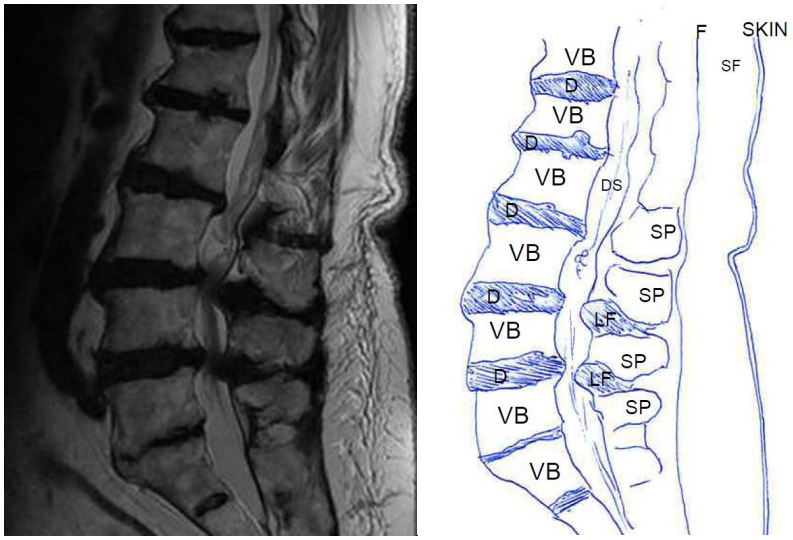


Figure 1. Saggital MRI of a lumbar spine with LSS in some segments. (All MRI images are used with the allowance from the radiologic department at Akershus University Hospital). VB = Vertebral body, D = Intervertebral disc, DS = Dural sac, SP = Spinous process, LF = Ligamentum flavum, F = Fascia, SF= Subcutaneous fat, FJ = Facet joint, PM = Psoas muscle, ESM = Erector spinae muscle.

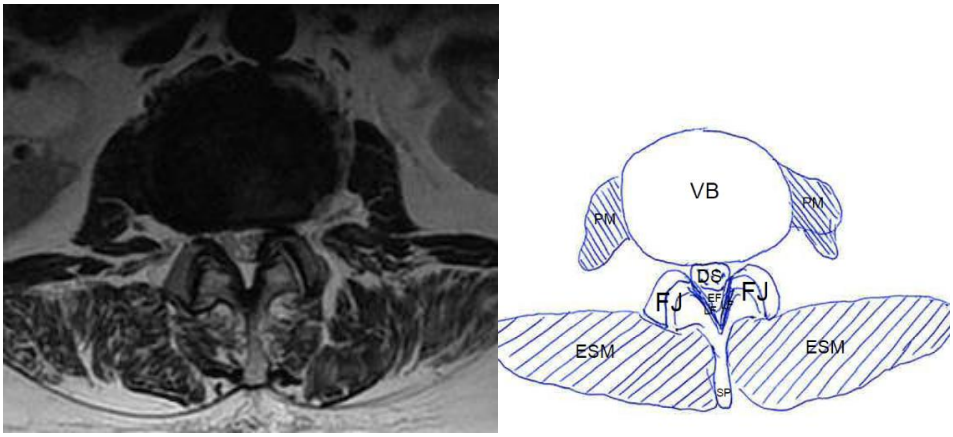


Figure 2. Transverse view (axial) of the lumbar spine in MRI in an asymptomatic segment; the cross-sectional area for the dural sac (DS) is slightly reduced but still sufficient. VB = Vertebral body, D = Intervertebral disc, DS = Dural sac, SP = Spinous process, LF = Ligamentum flavum, F = Fascia, SF= Subcutaneous fat, FJ = Facet joint, PM = Psoas muscle, ESM = Erector spinae muscle.

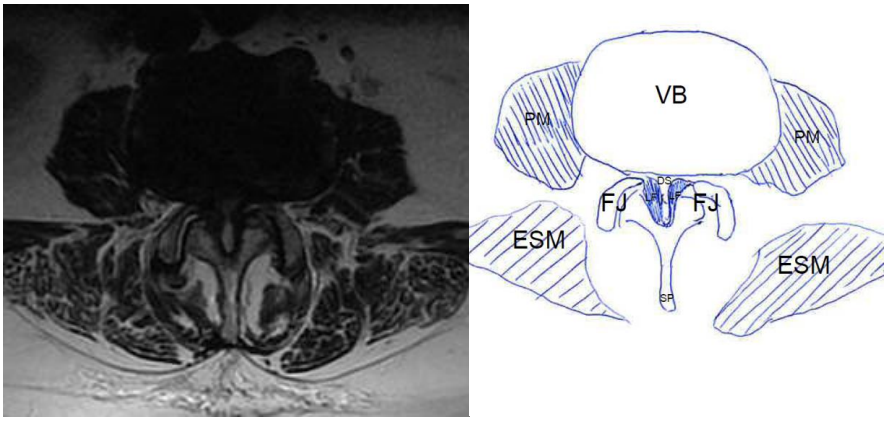


Figure 3. Transverse view (axial) of the lumbar spine in MRI in a stenotic segment; the area of the dural sac (DS) is restricted by thicker ligamentum flavum (LF) and osteophytes from the facet joints (FJ). VB = Vertebral body, D = Intervertebral disc, DS = Dural sac, SP = Spinous process, LF = Ligamentum flavum, F = Fascia, SF= Subcutaneous fat, FJ = Facet joint, PM = Psoas muscle, ESM = Erector spinae muscle.

Different types of LSS

The nerve roots run through different parts of the spinal canal. All the nerve fibres run vertically through the central spinal canal. Nerve filaments form nerve roots laterally in the spinal canal, often termed the subarticular recess. When emerging from the spinal canal, the nerve roots pass through the neural foramina in a more horizontal direction. The nerve structures can be affected in all these places. Central stenosis is the typical spinal stenosis with bilateral leg symptoms; several nerve roots are often affected. The lateral subarticular recess stenosis can result in symptoms from a specific nerve root and may be one (or two) sided. The foraminal stenosis affects single nerve roots and is often one-sided. Foraminal stenosis is often associated with structural deformity at the affected spinal segment, such as spondylolysis, spondylolisthesis, or degenerative scoliosis.
Some uncommon types of LSS that are not primarily caused by age-related degeneration. Congenital stenosis typically develops from a narrow canal due to short pedicles in patients between 30-60 years. Lipomatosis refers to a pathologically increased volume of epidural fat in the spinal canal, compromising the nerves and resulting in symptoms of spinal stenosis. Other rare disorders, such as acromegaly and Paget's disease, may also be associated with narrowing the spinal canal and LSS.

The exact **pathophysiological** mechanism of LSS remains unknown; however, two main theories have been suggested.
(1) Olmarker examined porcine nerve roots in the1980s-the 1990s and found a decrease in blood flow, impairing nutrition to the nerve tissue when introducing pressure on the cauda equine (19, 20). A review article from Katz from 2008 supports this vascular mechanism view, pointing to the reversibility of symptoms (21). The ischemia may be accompanied by venous congestion of the epidural veins located in the spinal canal.
(2) Another theory points to the development of structural changes in the nerves. Sekiguchi performed rat experiments implanting silicon spacers in the spinal canal. They reported demyelination associated with pressure on the nerves; however, no allodynia was associated with demyelination (22). The authors point to apoptosis (cell death) in the dorsal root ganglia as a possible pain mechanism. Other structural changes can be oedema, fibrosis, and axonal degeneration. These structural changes are not reversible; while symptoms of LSS usually are reversible, the structural changes in the nerves might contribute to failure after surgery for LSS (i.e. patients still suffering from LSS symptoms even after surgical decompression).

4.3 Epidemilogy

The prevalence of LSS is difficult to estimate. Different estimates for clinical LSS and radiologic LSS are proposed. A systematic review from 2020 found a pooled prevalence of clinical LSS of 11%; however, they also reported high risks for bias in two-thirds of the studies (3). Another study found a prevalence of radiologic LSS of 9 %, increasing with age to 47 % over 60 years (4). The progressive pathological degeneration described above supports the increasing prevalence with increasing age. LSS has become the most frequent indication for spinal surgery (1, 2). The NORspine records about 2000 LSS operations annually (2587 in 2021); this estimates that 0.05% of the total population is operated for LSS annually. The coverage of the NORspine registry is only about 80 %; hence, this probably underestimates the total surgical activity (23). A Swiss study reported about 1400 decompression operations yearly in a population of 1.3 million inhabitants (about 0.1% of the population operated annually) (24).

4.4 Diagnostics

4.4.1 Patient history

The spinal canal's space changes with the spine's position, and the symptoms often vary accordingly. The typical LSS patient experiences symptoms when standing or walking (spine extended). The symptoms may include buttock pain, radiating leg pain, numbness, and weakness of the legs. The walking distance is typically reduced to a few hundred meters, and "neurogenic claudication" refers to this phenomenon as the patients take breaks to relieve pain. Bending forward, sitting down, leaning over a shopping cart, and bicycling alleviate symptoms as this can increase the spinal canal's cross-sectional area and the space for the neural structures (13, 18, 21, 25).

4.4.2 Clinical examination

There is no specific clinical test to diagnose LSS, and neurologic clinical findings are uncommon. Clinical examination is most helpful to exclude other diagnoses (hip osteoarthritis, vascular claudication, neuropathy, trochanteric bursitis). The North American Spine Society clinical guidelines

found insufficient evidence to make recommendations for or against specific clinical findings to diagnose LSS (14).

4.4.3 Radiologic evaluation

MRI is considered the gold standard for radiological examination of LSS as it provides an excellent presentation of both the bony canal and the tissues limiting the spinal canal and the nerve structures (13, 14). MRI helps to confirm the diagnosis of LSS and to exclude other possible causes of back and leg pain (26). However, the correlation between the severity of symptoms and the degree of spinal stenosis on MRI is weak (13, 27, 28, 29). Furthermore, 20% of individuals over 60 years may have radiological findings of LSS despite suffering no symptoms (26, 30). MRI might be contradicted for patients with implants such as pacemakers; CT or CT myelography are good alternatives for these patients.

The width of the spinal canal can be measured by antero-posterior diameter (AP diameter), or the cross-sectional area of the spinal canal can be quantified (Dural sac cross-sectional area (DSCA)) (31). A cross-sectional area under 70 mm^2 has been defined as absolute spinal stenosis and an area between 70 and 100 mm^2 as relative spinal stenosis (32). A qualitative classification with grading according to dural sac morphology has also been suggested, the Shizas classification (33).

LSS is not easy to evaluate by plain x-ray films. Injecting a contrast medium into the spinal canal (radiculography) displays more of the spinal canal and neural foramina. Computer tomography (CT) displays the spinal canal quite well (focusing on the bony canal), and the visualization of neural structures can be further enhanced by spinal injection of contrast medium resulting in a CT myelography to visualize the space in the spinal canal. CT and x-ray (CT myelography or radiculography) are used when MRI is contraindicated or when positional or dynamic imaging is necessary.

Neurophysiological examination has not been shown to add value in diagnosing LSS, and the NASS guidelines do not recommend neurophysiological examination (14).

In conclusion, the diagnosis of LSS is challenging; it is a clinical diagnosis supported by radiologic findings (26, 34).

4.5 Treatment

LSS can be treated conservatively or surgically. There is a paucity of evidence regarding the natural course of LSS; an expert consensus stated that natural history could be favourable in about 30-50% of LSS patients (13). A Cochrane review from 2016 states that the evidence is too sparse to conclude which is best; surgical or non-surgical treatment (35).

4.5.1 Conservative treatment

Amundsen showed that half the patients (n=50) selected for conservative treatment had excellent or fair results after four years (36). Weinstein reported an improvement in a non-surgical group (ODI improved by 9.3 points) at two years follow-up (37). Guidelines from 2013 advise conservative treatment for patients with mild symptoms (13). Conservative treatment includes physiotherapy, orthosis, pain medication, and injections. However, there is a lack of evidence to make specific recommendations among these treatments (13, 38). A review from 2016 found no conservative treatment superior to another (39).

A lumbosacral corset might increase the walking distance (13). Pharmaceutical treatment includes standard pain medications such as NSAIDs and Paracetamol. In addition, Gabapentin/Pregabalin may be used to alleviate neuropathic pain. However, no study has compared medical therapy to the placebo; there is insufficient evidence to recommend pharmacological treatment for LSS as the reported effect can be the natural history of LSS (13, 14). The most recent review found low quality of evidence regarding the effect of oral medications (40). Epidural corticoid injections may relieve leg pain in the short term (13, 14, 41); however, the updated Cochrane review found that injections were ineffective (40).

When we planned this thesis, there was insufficient evidence to make recommendations for or against physical therapy according to two systematic reviews from 2013 and the clinical guidelines from North American Spine Society (13, 14, 42). However, during our project period, a recently updated systematic review reported moderate evidence for a multimodal approach, including manual therapy and exercise (40). A Norwegian multicenter RCT comparing conservative treatment (physiotherapy) to surgical treatment for LSS (Physical Therapy vs Surgical Decompression for Lumbar Spinal Stenosis) was started in 2020, but no results have been published yet (43).

In conclusion, there is a need for further research on conservative treatment for LSS. There is moderate evidence that a combination of manual therapy and exercise may be effective, no other conservative treatment has been shown to have an effect in the long term, and evidence is of low quality (41).

4.5.2 Surgical treatment – history and development

Verbiest first described LSS in 1954, and the first lumbar discectomy was described in 1909 by Krause and Oppenheim (44, 45). Early surgery included a rather extensive removal of spinal processes and laminae, i.e. laminectomy (Figure 4). Such extensive resection could destabilize the spine mechanically and damage muscle tissue (46). To minimize surgical trauma, Yasargil and Caspar independently developed a microsurgical interlaminar approach in 1977 (47, 48). The less invasive surgical approaches reduced the soft tissue damage and aimed to preserve the structural stability of the spine (5, 49, 50, 51).

The development of technical aids such as loupes and microscopes has driven spinal surgery towards mini-invasive surgery (MIS), where most stabilizing structures are preserved and only the tissue compromising the neural structures is removed. One of the most used techniques today is “Cross over” (“over the top”)- decompression, a technique where both recesses may be decompressed through a lesser unilateral approach by tilting the operating table and using visual aids (Figure 4). Parallel to the technical development, spine registries have shown promising clinical results using modern techniques (cross-over aided by microscope). A recent RCT compared three different minimally invasive surgical techniques for decompression (Cross over, bilateral laminotomy, and spinous process osteotomy), finding no differences in clinical outcome (49).

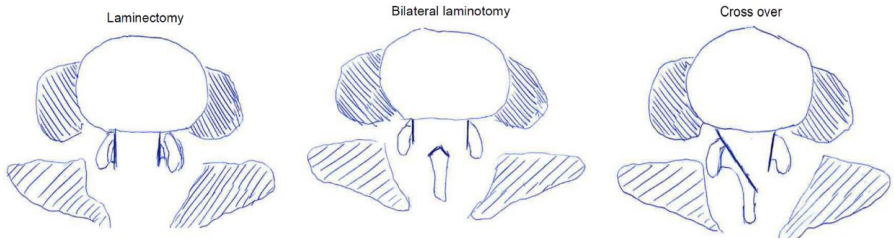


Figure 4: Three different surgical techniques for decompression of the lumbar spine, laminectomy, bilateral laminotomy and cross-over decompression.

4.5.3 Surgical treatment – current practice

Surgical treatment consists of decompression of the nerves in the spinal canal by removing some of the constricting tissue. MIS techniques are the most popular, yet open laminectomy is still performed. Visual aids are commonly used (23). An RCT from 2008 showed that surgery was more effective than conservative treatment; the surgical group reported an improvement in the Oswestry disability index (ODI) of 20.5 points compared to 9.3 points for the non-surgical group (37). Amundsen also found better results for patients randomized to surgery (vs conservative) in a small study from 2000 (36). Several studies have assessed the effect of surgical treatment, and between 62 and 75% of the patients report good results (5, 6, 7, 8, 9, 10, 11). The guidelines from 2013 stated that decompressive surgery improves outcomes in both the short and long term in patients with moderate or severe LSS symptoms (13). The NASS guidelines recommend surgical treatment based on “fair” evidence for short-term effects and “poor” evidence for long-term effects (14).

However, a systemic review from 2016 concluded that evidence that supports surgical treatment is insufficient; the authors called for trials on surgery vs sham surgery for LSS (52). Currently, one such trial is ongoing (no results have so far been published) (53). Some literature on LSS is old (from the 1990s and early 2000), and the surgical technique is constantly developing. The literature should be read in the context of the different technical eras.

Surgical treatment always involves risks. The most common perioperative complication is an unintended tear in the membrane covering the nerves in the spinal canal, a dural tear, with an incidence of 4-10% (54, 55, 56, 57, 58, 59). Other known complications are neural damage resulting in a neurologic deficit, postoperative hematoma, wound infection, and general surgical complications such as urinary tract infection, pneumonia, venous thromboembolism and ileus (60, 61).

Decompression can be supplemented with fusion; however, the benefit of additional fusion is debated. Several observational studies and one RCT have not reported any benefit of adding fusion to decompression, even in patients with a degenerative slippage (62, 63, 64). The NASS guidelines found fair evidence to recommend decompression alone (without fusion) for patients without instability (14). During our study period, an additional RCT showed decompression alone as non-inferior to decompression with fusion (65). There are differences in surgical practice concerning fusion between countries, yet no differences in clinical outcomes have been reported (i.e., fusion is more common in the USA than in the Nordic countries) (66, 67).

Patients with foraminal stenosis make up a small subgroup of LSS patients. The narrowing around the exiting nerveroot is often caused by loss of disc height or slippage that results in axial compression of the nerve root between the pedicles or one pedicle and the intervertebral disc. Decompression may not be sufficient to relieve the pressure. Patients with foraminal stenosis need consideration of different surgical approaches; fusion is one alternative.

4.5.4 Surgical treatment - future perspective

The latest technical development is endoscopic decompression, using only 5 mm skin incisions and surgically developing a space inside the spinal canal. Percutaneous procedures for disc herniations started in 1975 and developed in parallel with orthopaedic arthroscopic surgery using arthroscopic gear in the late 1980s and with saline water in the early 2000s (68, 69, 70). Dr S Ruetten widened the indications for endoscopic spine surgery during the 2000s to include LSS (51, 71). The endoscopic method is technically demanding, has a long learning curve, and has not yet been proven superior to classic surgery (72, 73, 74). However, complication rates may be inferior in endoscopic decompression (74).

Other alternative surgical techniques have been tested without lasting sound clinical effects. Interspinous devices aim to indirectly increase the space in the spinal canal by lifting the spinal segment between two spinous processes in a small operation. The procedure showed clinical effect but was associated with higher reoperation rates and total health care costs than decompression. This procedure was found improper for treating LSS; hence this technique has been abounded (75, 76). 3D-printed patient-specific cut guides to assist decompression have also been tested in a cadaveric study and showed no additional benefit when tested by experienced surgeons (77).

4.5.4 Surgical treatment – patient selection

The surgical development seems to have reached a limit where further technical development halts or does not improve the results after surgery further. The results after LSS surgery are only satisfactory for some patients. Even if 62-75% report significant improvement, some patients do not improve after surgery; hence, surgery might not benefit all LSS patients (5, 6, 7, 8, 9, 10, 11). Furthermore, some patients deteriorate after surgery, and some may suffer from complications; reducing the number of unnecessary and inefficient operations is essential. Indications for surgical treatment for disorders with pain and functional disability are often only relative. Indications must be thoroughly considered, possibly using higher thresholds for surgery than in other life-threatening diseases (i.e. cancer, infections, trauma). Careful patient selection and shared decision-making are mandatory to avoid overtreatment and have gained more focus in recent years. Sound evidence-based information about the expected result, including individualized chances for success, failure, and complications, should guide the decision.

Many studies have looked into predictors for success after surgery for LSS, evaluating the predictive value of different clinical and socioeconomic variables. Typical predictors of outcome after spinal surgery that have been reported are age, preoperative level of disability, smoking, working status, income, and level of education (6, 7, 8, 10, 11, 78). The predictors may vary with spinal diagnosis, procedures and populations across different parts of the world. *We aimed to explore non-success and the predictors for failure and worsening after surgery for LSS in Norway.*

4.6 Evaluation of outcomes

4.6.1 Clinical outcomes

Evaluating the effect of treatment for disorders dominated by pain and disability is a central question, primarily how to assess the impact of surgical treatment on pain. In contrast to other surgical treated diseases (i.e., cancer, infection, fracture), there are no obvious hard clinical outcomes in LSS (i.e. 5-year survival, infection eradicated, fracture healed). There is no objective way to measure the effect after surgery for LSS; i.e., MRI can display the area in the spinal canal after surgery, but this is not well correlated with the clinical result (27). The sensation and experience of pain depend on several factors, i.e., psychological status, mood, and social situation (79). Gender also affects pain perception and disability (80, 81, 82).
Over the last decades, the assessment of surgical outcomes has altered from doctor-centred and radiological measures to patient-centred (83). Patient-reported outcome measures (PROM) have been developed to measure these subjective symptoms and has become the gold standard for evaluating clinical result after spine surgery. PROMs are multidimensional questionnaires that provide information about pain, disability and health-related quality of living; prospectively collected PROMs can assess treatment effectiveness.

PROM instruments in spine surgery have been thoroughly explored and perform well for different spine diagnoses (84, 85). Although inferior methodical quality according to the Newcastle Ottawa scale, PROMs are recommended as outcome measures in spine surgery (86, 87). A combination of a general PROM measuring life quality, organ-specific PROMs measuring spine-related disability, and pain scales is essential and advised in spine surgery (88). Return to work is also a relevant tool, but as patients suffering from LSS often are old and retired, return to work is not that central in the LSS population.

The Norwegian Registry for Spine Surgery (NORspine) uses the most widely used PROMs in spine surgery – Oswestry Disability Index (ODI), and Numeric Rating Scale (NRS) for back - and leg pain, and EuroQoL5 Dimensions (EQ-5D) for measuring the quality of life (88, 89, 90). This choice is supported by the IMMPACT recommendation from a consensus meeting regarding the assessment of treatment effect for chronic pain; they recommend using at least two of the following; pain intensity (NRS), physical function (multidimensional scales), emotional functioning and overall improvement (transitional scale) (91).

The Oswestry Disability Index (ODI) is a disease-specific PROM; it is a validated and widely used PROM in spine surgery. ODI has been translated into Norwegian and tested for psychometric properties (84, 92). Although initially designed for low back pain (LBP), it is the most widely used in LSS (93, 94). The ODI PROM consist of ten questions answered between 1 and 5. The total score is calculated based on the questions answered and will be between 0 (no disability) and 100 (bed bound). Population-based data indicate a normative mean ODI of 8.8 for adults (95). Other PROMs for spine disorders exist and are used although not as widely (Roland Morris disability index, Zurich claudication questionnaire (Swiss spinal stenosis questionnaire, SSSQ)) (96, 97). Criticism has been raised against using ODI in LSS since it was developed for LBP (98). However, the problem with LSS-specific PROMS (i.e., SSSQ) is that they are not generalizable for all degenerative spine disorders; hence they are problematic to use in general spine registries (99). The Norwegian ODI form is shown in the appendix (appendix 1).

EQ-5D is a generic measure of health-related quality of life. This PROM is based on five questions, and the total score is calculated using a specific algorithm. The total score is between -0.59 ("worse than dead") and 1.00 ("full health") (90, 100). The Norwegian EQ-5D form is shown in the appendix, Appendix 2).

The numeric Rating Scale (NRS) for pain (back pain and leg pain) uses scales from zero (no pain) to ten (worst pain) to quantify pain in either the back or in the leg; this scale is validated and widely used (85, 91, 101). The Norwegian NRS form is shown in the appendix (Appendix 3).

In addition to PROMs measuring the status at a specific time, transitional scales add valuable information and can serve as anchors regarding the effect of treatment. Global Perceived Effect (GPE) scale is a measure of health transition. The GPE scale is a seven step-scale (1=worse than ever, 2 = much worse, 3 = somewhat worse, 4 = unchanged, 5 = somewhat better, 6 = much better, and 7 = completely recovered) (102). The Norwegian GPE scale is shown in the appendix (Appendix 4).

There are some essential considerations using PROMs as outcome variables; the instrument should be valid and reliable, meaning the instrument should measure correctly what we intend to measure, and the results should be identical when measurements are repeated. Furthermore, the scales should be calibrated to measure in the relevant clinical area, i.e. without floor - or ceiling effects. A lack of focus on these aspects could result in information bias. The PROMs are often continuous variables, and the distribution is essential; it has implications for statistical methods as some methods require a natural distribution.

Another vital issue using registry data is the size of the registry and the possibility of revealing statistically significant findings that are of no clinical importance. There are two types of error in research, type 1 and 2. Type 1 error is to detect a difference between groups that do not exist in the broader population (false positive). The p-value is set to minimize the risk of type 1 error, a p-value of 0.05 means there is a 5 % chance that a detected difference is a false positive difference. Type 2 error describes the risk of not detecting a difference in the study population that exists in the larger population (false negative) and is a question of statistical power. In registry studies, both error types are rare since the study population is ideally the same as the total population. Medical registries can reveal statistical differences that are clinically irrelevant; hence, the effect estimates and confidence intervals are more interesting than the p-values (103).

A common way to handle these challenges is to categorize the outcome measure; however, defining clinically relevant change – i.e., does an ODI improvement from 30 to 25 mean this patient has had a relevant clinical improvement? – can be problematic. Patients can perceive and weigh the different items in the PROM instrument (ODI) differently. A concept of minimum clinically important difference (MCID) has been established; this refers to the change in the PROM score necessary to describe a meaningful clinical improvement (104). The MCID could be considerably higher than the minimal detectable change (MDC), especially in large study populations. Different values for MCID for ODI have been reported between 8-20 (84, 105, 106, 107, 108.). Different cut-offs for success have also been reported at 22-24 points in the final ODI (109, 110). Dichotomous clinical outcome, defined by specific cut-offs, makes calculations of proportions and risks of a particular outcome possible. Categorical outcomes could also be easier to communicate with patients. The GPE scale had a categorical design originally.

There are several ways to calculate the change in the PROM score. One can consider only the follow-up value, irrespectively of the baseline value, i.e., Patient acceptable symptom state (PASS), or one can calculate different change scores, either numeric change (follow-up score – preoperative score) or percentage change (change / preoperative score). A percentage change of 30% improvement from baseline was considered clinically meaningful at a consensus in 2008 (107). A systematic review from 2021 explored this topic: the definition of MCID and methods to identify MCID cut-offs (111). They advised higher MCID for surgical conditions.

A central point of using MCID and cut-off values is that they only apply to individual differences and not to mean differences for groups of patients (91). When comparing groups, analyses according to MCID or cut-offs might camouflage clinically relevant differences for some patients.
Success criteria for ODI score after spinal surgery, such as PASS score, MCID, or percentage change, have been explored; however, criteria for failure and worsening are not well explored (109, 110). A Norwegian study on LSS patients used an increase in ODI of 8 points - equal to the MCID for ODI - to define failure/worsening (78). The assumption that MCID for deterioration would be equal to improvement is not founded on scientific evidence. Another Norwegian study explored failure cut-offs using an external anchor, but only for patients suffering from disc herniations (112). We focused on *failure after surgery for LSS and aimed to identify the most accurate PROMs that defined failure and to calculate the cut-offs for this patient group in Norway.*

4.6.2 Complications

Spinal surgery implies a risk for complications; elderly patients have reduced physiological reserves and are prone to infections, postoperative hematoma, micturition problems and venous thrombosis. The most common perioperative complication in LSS surgery is an incidental dural tear (ID), a laceration of the membrane that covers the nerves in the spinal canal. Former studies have shown higher postoperative ODI scores among patients that suffered an ID compared to those that did not, although the difference between the groups was minor and below the MCID (58). An ID may involve exposition of nerve filaments and leakage of cerebrospinal fluid. A repair of the dura with suture may damage the nerve filaments. Also, continuous cerebrospinal fluid leakage after closure may necessitate several days of bed rest. Hence, we expect that ID may lead to inferior results.

Some methodological issues should be considered when studying the impact of ID on PROM after LSS surgery. Utilizing MCID on a group level might be misleading as patients exposed to the factor examined (ID), although not clinically affected, will pull the mean score in a neutral direction. Therefore, categorical outcomes may be more appropriate to use than mean values when assessing the impact of a complication on a particular outcome. A categorical outcome allow us to calculate the proportions of patients with a particular outcome. To explore the difference between continuous and categorical *outcomes, we aimed to examine the effect of ID on clinical outcomes both as a categorical outcome using the GPE scale and on a continuous outcome using ODI.*

4.7 Registry data

Medical registries have grown in popularity over the years. They provide large patient populations and reflect everyday practice in several surgical units; hence they report the effectiveness of the treatment in the entire population (ideally, no selection bias) (103, 113, 114). Furthermore, registry-based studies supplement results from more narrow studies, i.e., RCTs, as registry-based studies provide high external validity. Moreover, registries also have the population size and statistical power to analyze rare events such as complications (114, 115).

Most medical registries face common challenges. High coverage (the proportion of reporting treating centres) and completeness (the proportion of eligible patients registered) are essential to optimize external validity and minimize the risk of selection bias. Loss to follow-up is also a relevant problem in medical registries; patients not responding may introduce an attrition bias when considering clinical outcomes, and follow-up rates of 60-80% are recommended (88). Several studies have explored loss to follow-up in spine registries and report some differences in baseline data between patients who respond compared to those who do not. However, similar clinical outcomes are reported by responders and non-responders (116. 117, 118). This has implications on how to handle missing data, as data missing at random can be analyzed differently than data missing not at random. The NORspine registry analyzed loss to follow up in a study from 2011 (116). However, the data used was from 2004 and from one single treatment centre. *We found it essential to repeat the analyses of non-responders since the NORspine has developed into a national registry during the last decade.*

NORspine is designed so that patients and surgeons complete different parts of the data set. As many different persons provide the registry data, data can be misclassified. Inaccurate entries in the registry might introduce information bias (103). A recent study reported low to moderate data accuracy in a spine registry (German spine society (DWG)) and advised against using data from this registry (119). NORspine conducts a data quality assessment every second year, and a former control identified the following problematic variables: ASA classifications, comorbidities and reoperations (120). Data quality is a paramount concern, and w*e aimed to explore the accuracy of registry data in NORspine; this is important when reporting and interpreting results based on registry data. Data from NORspine is the basis for this thesis*.

5 Aims of the thesis

Our primary goals were to explore failure and worsening after surgery for LSS using NORspine data and to identify predictors for failure and worsening that could assist in patient information and help in patient selection with an emphasis on shared decision-making.

5.1 Aims paper 1 (data accuracy in NORspine).

Assess data quality (accuracy and agreement) of the NORspine registry. Which variables are sufficiently accurate to use in data analyses?

5.2 Aims paper 2 (characteristics of non-responders).

Examine if loss to follow-up introduces attrition bias that may affect the assumption of clinical outcome.

5.3 Aims paper 3 (criteria for failure).

Identify the most accurate cut-offs for typical PROMs that define failure and worsening after surgery for LSS.

5.4 Aims paper 4 (dural tear).

Examine the effect of incidental dural tear on clinical outcomes, using both categorical and continuous outcomes.

5.5 Aims paper 5 (predictors for failure).

Identify and quantify predictors for failure and worsening after surgery for LSS in Norway.

6. Patients and methods

6.1 Study design

This thesis is based on retrospective analyses of prospectively collected data in a national spine registry. Paper 1 was a cross-sectional study.

6.2 Patients

6.2.1 NORspine database

NORspine was established in 2000 as a local spine registry at the University hospital in Northern Norway (UNN). It developed into a national spine registry in 2007 and is central to research and health administration in Norway. In Norway, all treating centres are obliged to participate in the registry. However, the NORspine is based on informed consent by the patients; all patients receive and sign an information sheet before inclusion. The exclusion criteria in NORspine are age under 18, insufficient understanding of the Norwegian language, and tumours, fractures and primary infections (120).

A NORspine data set consists of baseline data provided partly by the patients (socioeconomics and preoperative symptoms, including ODI score and NRS back –and leg pain; see Appendix 5) and partly by the surgeons (previous surgery, comorbidity; see Appendix 6). Spine-related diagnostics and surgical details are provided by the surgeons immediately after the surgery; see Appendix 6. The patients receive two follow-up forms at three - and twelve months after surgery regarding the clinical outcome and complications at three months. These forms record symptom severity, including ODI-score, NRS back and - leg pain, health-related quality of living (EQ-5D) and clinical improvement using the ordinal GPE scale (Appendix 7 and 8).

The NORspine registry generally has a loss to follow-up of 26% (23). NORspine registry complies with most of the recommendations for spine registries proposed by Van Hoof in 2015 (88). Repeated spine surgeries are registered as revision surgery if done within three months after the primary

surgery and as a completely new case if the reoperation is done later than three months. NORspine does not record patients that are treated conservatively.

6.2.2 Patients

We studied patients operated for LSS registered in NORspine, and included patients treated from 1. st January 2007 to 1. st April 2017. We found 11873 patients, and 8919 (75%) answered at 12 months follow-up. A flowchart illustrating the study population is displayed in figure 5.

The selection of patients for our studies was based on diagnosis and surgical procedure to make the study population meet the requirements: patients with LSS who received decompressive surgery. Additional diagnoses and surgical procedures exist for some patients.

In the data validation studies (papers 1 and 2), we assessed all LSS patients registered in NORspine operated at four hospitals (Akershus University Hospital, Elverum Hospital, Gjøvik Hospital and Lillehammer Hospital) during two years (2015 and 2016). We identified 474 patients (12.3% of the NORspine population) in the validation studies.

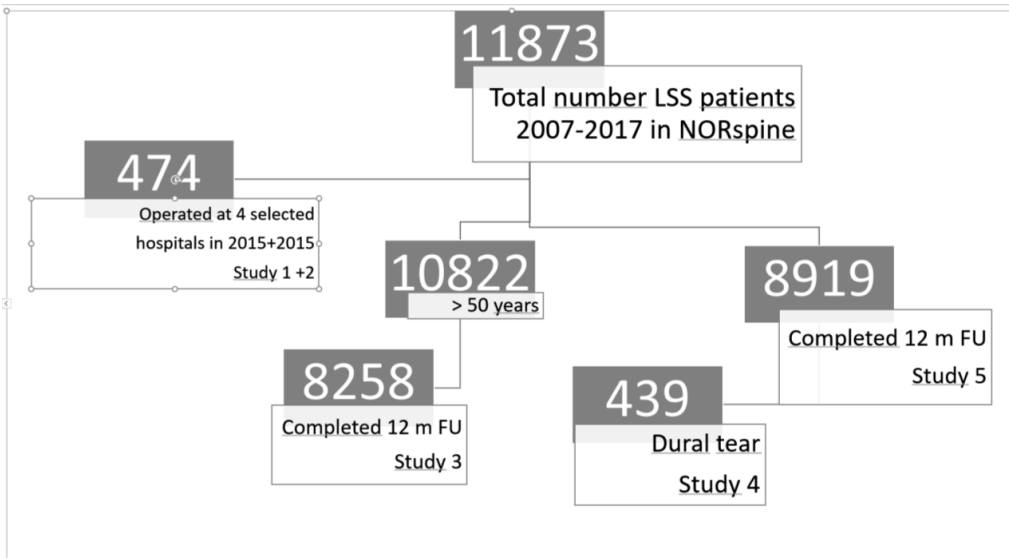


Figure 5: Flowchart of the total study population, displaying subgroups for paper 1 and 2 (data accuracy in NORspine and characteristics of non-responders).

6.2.3 Surgical methods

All patients received some type of decompressive surgery. Surgical techniques recorded by the NORspine registry were laminectomy or laminotomy (one-sided, two-sided, and “cross over” (bilateral decompression by a unilateral surgical approach)). Some patients had an additional fusion.

6.2.4 Data

In addition to data from the NORspine registry, we collected data from electronic patient records (EPR) for paper 1. Data from EPRs were collected using a standard empty NORspine form, and the investigators had no access to corresponding data previously recorded in the NORspine. The study group selected a set of NORspine variables that could be re-captured from EPRs. Two raters independently reviewed the EPRs of 22 patients to estimate inter-observer reliability in our study.

Supplementary follow-up forms were mailed to a subgroup of non-responders for paper 2. Patients who did not complete the 12-month follow-up were contacted by regular mail twice, including one additional SMS reminder.

6.3 Statistics

We report central tendency in terms of mean (95% CI) for continuous data with normal distribution and numbers and proportions (%) for categorical data. We did not impute any missing data. To assess selection bias, we did non-responder analyses of baseline characteristics in paper 3 and paper 5, in addition to paper 2, as recommended by van Hoof (88).
All statistical analyses were done with SPSS versions 25 and 26 (IBM Corp. released in 2017. IBM SPSS Statistics for Windows, Version 25. Armonk, NY) and MedCalc Software Ltd. relative risk calculator. https://www.medcalc.org/calc/relative_risk.php (Version 20.027; accessed March 14, 2022).

Some of the statistical analyses have requirements regarding the data and distribution. The outcome variables had a normal distribution. For the logistic regression, the following conditions were met: categorical outcome variable, no correlation between the explanatory variables used in the models. We also tested for multicollinearity and linear relationship between continuous explanatory variables and the logit of the outcome variable.

Paper 1 (data accuracy in NORspine)

We calculated concordance in terms of agreement when comparing the structured NORspine data with EPR data; we also calculated accuracy for dichotomous variables, using EPR as the gold standard (Figure 6). We chose to report both accuracy and agreement because certain EPR variables could be questioned as reference (e.g., smoking and comorbidity).

	EPR +	EPR -
NORspine +	A	B
NORspine -	C	D

Figure 6. Table displaying the terms used in accuracy (EPR = Electronic patient records). Sensitivity = A/ (A+C), specificity = D/ (B+D), and Proportion Correctly Classified, PCC = A+D / (A+B+C+D).

Accuracy was presented as sensitivity and proportion correctly classified. Agreement between NORspine and EPRs was assessed by Cohen's kappa (ƙ) or Fleiss weighted kappa (ƙ) for categorical variables (dichotomous and ordinal variables (ASA classification was analyzed as an ordinal variable, ranging from 1 to 5, in the agreement analysis)). For continuous variables, we calculated the intraclass correlation coefficient (ICC) using a two-way mixed model to assess absolute agreement. We classified agreement (ƙ-value) as minimal (0.21–0.39), weak (0.40–0.59), moderate (0.60–0.79), strong (0.80–0.90) and almost perfect (> 0.90) (121). The agreement, according to ICC (values), was classified as poor (< 0.50), moderate (0.50–0.75), strong (0.75–0.90) and excellent (< 0.90) (122). Finally, we calculated the prevalence of missing values for each variable. The results are presented as point estimates with 95% confidence intervals (CI).

Paper 2 (characteristics of non-responders)

We performed a separate study on patients lost to follow-up, comparing baseline data and clinical outcomes between non-responders and responders. To compare clinical outcomes, we collected continuous variables such as ODI, NRS back- and leg pain and one categorized PROM: GPE (transitional scale). We analyzed between-group differences by mean difference (95%CI) and Stud T-test for continuous variables, or relative risk (95%CI) and z-statistics for categorical variables.

Paper 3 (criteria for failure)

We defined failure after surgery for LSS as GPE 4-7 (“unchanged” or any degree of worsening) at 12 months after surgery and worsening as GPE 6-7 (“much worse” and “worse than ever”). These categorical outcomes were used as external anchors to identify cut-offs for failure and worsening. We used Receiver Operating Characteristics (ROC) analyses for each PROM derivate to identify cut-off values for failure and worsening. We determined the cut-off with the highest sensitivity and specificity using the closest point to the upper left corner of the ROC curve (Figure 9). We calculated the areas under the respective curves (AUC) to determine how accurately the PROM derivates classified the outcomes as failure vs non-failure and worsening vs non-worsening. AUC values and

corresponding grades of accuracy were interpreted as follows: < 0.7 = poor, 0.7 - 0.8 = fair, 0.8 - 0.9 = good, and ≥0.9 = excellent accuracy (123).

To evaluate the consistency of our results across subgroups, we performed ancillary analyses for age and preoperative ODI score quartiles, as well as different surgical treatments (decompression versus decompression and fusion). We performed the subgroup analysis only for the failure group, as the worsening group was considered too small.

Paper 4 (dural tear)

Primary outcome: To estimate the association between incidental dural tear (ID) and clinical outcomes, we used multiple logistic regression with failure and worsening (defined by GPE) as dependent variables, ID (yes/no), and potential confounders as independent variables. Based on previously published data, we adjusted the primary analysis by the following potential confounders: age, gender, BMI, smoking, ASA (dichotomized as grade 1 and 2 vs grade 3, 4, and 5), preoperative PROMs, duration of leg pain before surgery, previous surgery (at the same lumbar level), multilevel surgery, and fusion (in addition to decompression) (6, 78, 124). The potential confounders were decided a priori and not by statistical testing. We provided unadjusted and adjusted estimates for odds ratios (OR) with 95% CIs. This is in line with previous recommendations; multivariate methods with adjustment for covariates are recommended (88, 89, 125).

Secondary outcomes: To examine the secondary outcomes, we repeated the regression analysis using the different dichotomous outcomes (defined by ODI final score, ODI absolute change, and ODI percentage change). To quantify the association between ID and the mean ODI final score and NRS leg pain score, we used multiple linear regression with ODI final score and NRS leg pain as dependent variables, adjusting for the possible confounders. We also analyzed the association between ID and length of hospital stay and patient-reported postoperative complications, using multiple linear regression and multiple logistic regression, adjusting for possible confounders.

Paper 5 (predictors for failure)

We defined failure and worsening according to our findings in paper 3, with failure as ODI>31 at 12 months and worsening as ODI>39 at 12 months; ODI score as the primary outcome is recommended by van Hoof (88).

We assessed predictors using uni- and multivariate logistic regression, with backward conditional stepwise selection with an entry and removal threshold of 0.01. Failure and worsening were used as dependent variables (outcome). Covariates in the predictor analyses were chosen according to previous literature: age, gender, smoking, ASA classification, BMI, educational level, civil status, Norwegian speakers, disability benefit, former spinal surgery, MRI findings, preoperative ODI score, duration of symptoms, multilevel surgery (6, 78, 88, 124). Among the covariate variables, some were dichotomized to improve the data-to-model fit and facilitate the interpretation of the analyses (age, BMI, ASA classification, and educational level). There was no strong (<0.7) correlation between the covariates, and only preoperative ODI had a non-linear relationship to logit failure, as displayed in figure 1 in paper 1. There were no statistically significant effects of the interactions between the covariates.

6.4 Etichs

Participation in the registry is voluntary and includes written consent. The study was also approved by The Norwegian Regional Committee for medical and health research ethics ((2017/2157). Data protection officers at the involved hospitals for data re-capture (Innlandet Hospital Trust and Akershus University Hospital) approved the study. The study was conducted following the Helsinki declaration and is presented according to the STROBE statement (126).

7 RESULTS

7.1 Overall results

The mean (95%CI) age in the entire study group was 65.8 (65.3-66.0); 52% were female, and 21% were ASA class 3 or 4. MRI showed central stenosis in 70%, lateral stenosis in 57% and foraminal stenosis in 10%. 25% of the patients had had previous spine surgery, and 75% had had back pain for more than 12 months. The mean (95%CI) preoperative ODI was 40 (40-41).

In the subgroup used in studies 1 and 2, the mean age (95%CI) was 66 (65.3–67.2) years, and 254 (54%) were females. 26% were ASA class 3-4. 28% had had previous spine surgery, and the preoperative ODI score was 41 (40 – 42).

In the entire study group, the mean (95%CI) ODI 12 months after surgery was 23.9 (23.5 – 24.2), and the mean (95%CI) improvement in ODI was 15.9 (15.5 – 16.3) points (paper 5). The distribution of preoperative ODI score is displayed in figure 7, ODI scores 12 months after surgery is displayed in figure 8.

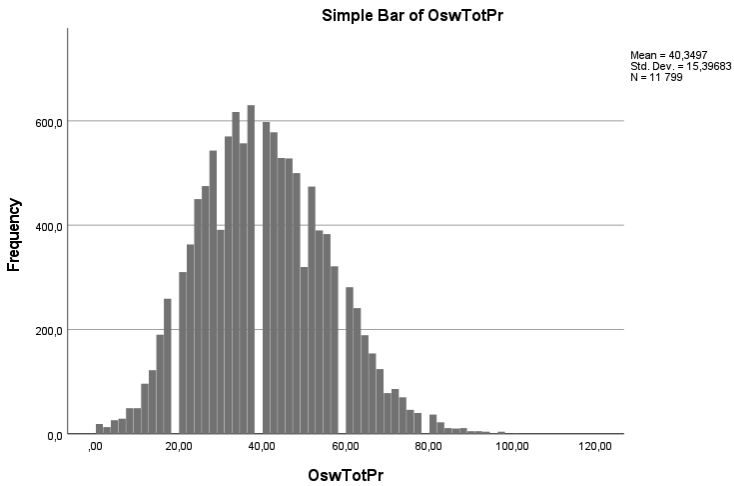


Figure 7: Distribution of ODI before surgery for LSS.

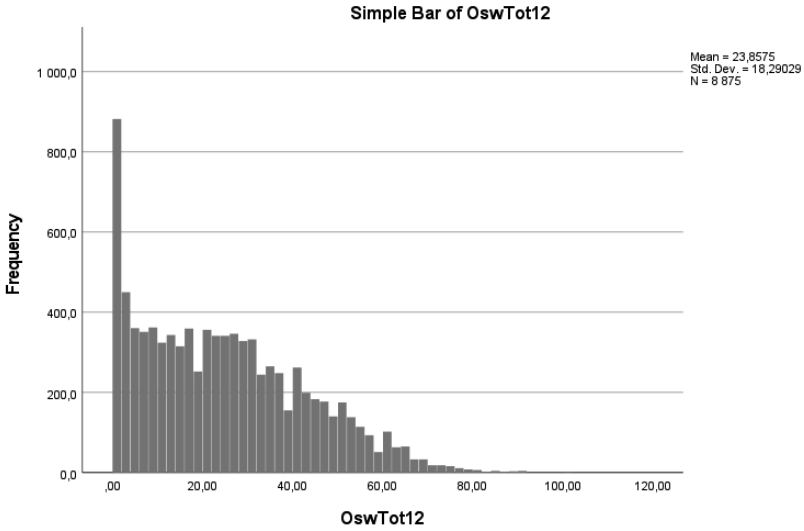


Figure 8: Distribution of ODI score 12 months after surgery for LSS.

At 12 months follow-up, the outcomes of 1,683 patients (20%) were classified as failures according to the GPE (GPE 4-7) and the outcomes of 476 patients (6%) as worse (GPE 6-7) (paper 3). According to the ODI final score cut-offs, 2950 (33.2%) patients were categorized as failures, including 1921 (21.6%) classified as worse 12 months after surgery (paper 5). When we used the ODI change score as cut–offs, 2893 (32.8%) reported failure, and 2132 (24.2%) reported worse.

According to NORspine data, 3.2 % had perioperative complications (paper 1), and 13.7% had postoperative complications (paper 4).

7.2 Results paper 1 (data accuracy in NORspine)

The total missing data was 0.9% in NORspine and 2.8% in EPRs. For a sample of 22 patients, the interrater reliability for the two authors that reviewed EPR variables was almost perfect.

Perioperative complications were recorded for 15 patients (3.2%) in NORspine, and 30 patients (6.5%) in the EPRs. The agreement between NORspine and EPR for perioperative complications was weak (ƙ (95%CI) = 0.51 (0.33–0.69)), the sensitivity for recording a complication (95%CI) was 40% (23–58%) (Table 1). Perioperative details (method of decompression, fusion, surgical access, spinal level operated) recorded by the surgeon showed moderate to excellent agreement between NORspine and EPR (ƙ= 0.76 to 0.98), and high proportions were classified correctly (93–99%), the sensitivity for the recording of perioperative details was high (92–99%) (Table 1). Smoking status had an almost perfect agreement (ƙ (95%CI) = 0.93 (0.89–0.97)), a proportion correctly classified of 97.2%, and a sensitivity of 92.0% (Table 1).

Some surgical details, i.e. previous surgery (yes or no), had an almost perfect agreement (ƙ (95%CI) = 0.93 (0.89–0.97)), a proportion classified correctly of 97.2%, and a sensitivity of 95.8% (Table 1). However, the number of previous surgeries showed only moderate agreement (ƙ (95%CI) =0.62 (0.48–0.75)) (Table 2).

ASA-classification (1-5) showed moderate agreement (ƙ (95%CI) = 0.73 (0.66–0.80)) (Table 2). The prevalences of comorbidities differed in the two data sources; NORspine underreported comorbidities compared to EPRs. Furthermore, the patients' height, weight, and BMI showed excellent agreement between NORspine and EPRs (ICC= 0.99 to 0.99) (Table 2).

Table 1. Accuracy and agreement of NORspine data for 474 spinal stenosis patients compared to their electronic patient records.

Variable (missing, n)	Prevalence * n (%)	PCC**	Sensitivity (95%CI)	Kappa (95%CI) ***
Perioperative complications (11)	30 (6.4%)	96%	40% (23-58)	0.51 (0.33-0.69)
Previous spinal surgery (14)	120 (26.1%)	97%	96% (92-99)	0.93 (0.89-0.97)
Additional fusion	51 (10.8%)	99%	94% (88-100)	0.93 (0.88-0.99)
Access, posterior midline (26)	414 (92.4%)	93%	93% (91-96)	0.19 (0.03-0.35)
Level L2-3 (14)	74 (16.1%)	99%	99% (96-100)	0.98 (0.95-1.00)
Level L3-4 (13)	193 (41.9%)	99%	98% (96-100)	0.97 (0.95-0.99)
Level L4-5 (13)	312 (67.7%)	98%	99% (98-100)	0.95 (0.92-0.98)
Level L5-S1 (13)	48 (10.4%)	99%	92% (84-100)	0.92 (0.86-0.98)
Smoking (43)****	112 (26.0%)	97%	92% (87-97)	0.93 (0.89-0.97)

*prevalence according to EPR
** Proportion correctly classified
***Cohens Kappa
****Smoking was registered by patients on the preoperative form; the remaining variables were registered by the surgeon on the postoperative form.

Table 2. Agreement for NORspine data for 474 spinal stenosis patients compared to their electronic patient records, ordinal or continuous variables.		
Data source	Variable	Agreement* (95%CI)
Surgeon, postoperative form	ASA classification**	0.73(0.66 - 0.80)
	Number of previous surgeries	0.62 (0.48 – 0.75)***
	Number of levels operated	0.91 (0.84 – 0.99)
	Type of surgery****	0.90 (0.82 – 0.98)
	Method of decompression*****	0.76 (0.68 – 0.84)
Patient, preoperative form	Height (centimeters)	0.99 (0.98 - 0.99)
	Weight (kilograms)	0.99 (0.99 - 0.99)
	BMI (calculated)	0.99 (0.98 – 0.99)

*Fleiss weighted kappa for ordinal data, Intraclass correlation coefficient (ICC) for continuous data. ICC was calculated using a two-way mixed model and absolute agreement (average measures).
** ASA: American Society of Anesthesiologists - classification of physical status (1 - 5). Mean ASA score was 2.17 in NORspine and 2.14 in EPR.
***Mean number of previous spine surgeries was 1.29 in NORspine and 1.42 in EPR.
****Type of surgery was graded as decompression or decompression and fusion.
*****Decompression options were unilateral foraminotomy, crossover ("over the top"), or bilateral foraminotomy.

7.3 Results paper 2 (characteristics of non-responders)

In the study group, 140 (30%) of the included patients had not completed 12 months of follow-up (non-responders), and 334 (70%) had completed 12 months of follow-up (responders). Of the 140 non-responders, 17 were not possible to contact (unknown address, moved abroad or deceased); hence only 123 were included in the analyses of clinical outcome. Sixty-four patients (52%) returned our questionnaires ("responsive non-responders"), while 59 (48%) did not return the forms ("resistant non-responders").

The non-responders were younger than the responders, 63.0 (95% CI: 61.0-64.9) vs 67.7 (95% CI: 66.6 – 68.7) years, with a mean difference (95% CI) of 4.7 years (2.59 - 6.74); p=<0.001. Non-responders were more often smokers compared to responders (41 (30%) vs 70 (21%), RR (95%CI) 1.40 (1.01 - 1.95); p = 0.044). Furthermore, non-responders had a lower proportion of surgeon-reported relevant comorbidities compared to responders (93 (69%) vs 243 (78%), RR (95%CI) 0.89 (0.77 - 1.00); p= 0.047). However, we found no statistically significant difference in ASA classification

between the non-responders and responders (the number (%) of ASA grade 1 and 2 was 111 (79%) vs 242 (72%), RR (95% CI) 1.09 (0.98 to 1.22); p = 0.100). There was no statistically significant difference between the non-responders and responders at baseline for the other baseline variables. We found no statistically significant difference in the type of surgery received by non-responders and responders.

The median follow-up time for non-responders in our supplementary cross-sectional study was 50 months, interquartile range of 10 months (min 36 – max 64). We did not find any statistically significant difference in mean (95%CI) ODI score between the non-responders and responders postoperatively 28.2 (23.2-33.2) vs 25.2 (23.2-27.2), mean difference (95% CI) 2.99 (-2.11 to 8.11); p =0.250 (Table 3). We did not find any statistically significant difference between the non-responders and responders in other PROMs or proportions reporting success by GPE (63 (70%) vs 330 (79%), RR (95%CI) 0.89 (0.75 to 1.06); p = 0.183)) (Table 3).

We found no statistically significant differences in baseline variables between non-responders answering our delayed follow-up ("responsive non-responders") and non-responders not answering ("resistant non-responders").

Table 3. Postoperative clinical outcome for a selection of patients operated for LSS comparing responders and non-responders.

Clinical outcome	Non-responders. Mean (95%CI) / n (%)	Responders. Mean (95%CI / n (%)	Mean diff (95%CI) or relative risk (95%CI)	p-value
ODI	28.2 (23.2 to 33.2)	25.2 (23.2 to 27.2)	3.0 (-2.1 to 8.1)	0.250
NRS back pain	4.62 (3.87 to 5.37)	4.09 (3.77 to 4.41)	0.43 (-0.3 to 1.2)	0.271
NRS leg pain	4.00 (3.21 to 4.79)	3.84 (3.50 to 4.18)	0.15 (-0.7 to 1.0)	0.719
Success by GPE*	63 (70%)	263 (79%)	0.89 (0.8 to 1.1)	0.183

- Success defined as "completely recovered" or "much better" according to Global Perceived Effect (GPE) scale.

7.4 Results paper 3 (criteria for failure)

An ODI percentage change of less than 20% displayed the highest accuracy in identifying failure 12 months after surgery. The area under the curve (AUC) (95% CI) was 0.89 (0.88 - 0.90)), sensitivity 82%, and specificity 81%. An ODI final score of 31 points or more, and an ODI absolute change of less than 8 points, also accurately classified failure (AUCs 0.87 and 0.86) (Table 4a).

An ODI final score over 39 points showed excellent accuracy in identifying worsening 12 months after surgery. AUC (95%CI) was 0.91 (0.90-0.92) and sensitivity 83%, and specificity 79%). ODI percentage change of less than 9% and an ODI absolute change (improvement) of less than 4 points also accurately classified worsening (AUCs 0.87 and 0.86) (Table 4b).

Table 4a. PROM accuracy to identify failure (GPE=4-7) and worsening (GPE=6-7) 12 months after surgical treatment of spinal stenosis in 8,258 patients. An area under the curve (AUC) > 0.7 indicates acceptable sensitivity and specificity.

		Failure (GPE 4-7) n= 1683/8258 (20%)			
Outcomes	n	Cut-off	AUC (95% CI)	sensitivity	specificity
Disability					
ODI final score	8,220	31	0.87 (0.86-0.88)	0.79	0.78
ODI absolute change	8,174	-8	0.86 (0.86-0.87)	0.78	0.79
ODI percentage change	8,161	-20%	0.89 (0.88-0.90)	0.82	0.81
Back Pain					
NRS back pain final score	8,174	5,5	0.87 (0.86-0.88)	0.79	0.81
NRS back pain absolute change	7,687	-1.5	0.83 (0.82-0.84)	0,80	0,74
NRS back pain percentage change	7,573	-21%	0.85 (0.84-0.86)	0,81	0,77
Leg Pain					
NRS leg pain final score	8,067	5.5	0.85 (0.84-0.86)	0,73	0.82
NRS leg pain absolute change	7,518	-1.5	0.83 (0.82-0.84)	0,77	0,76
NRS leg pain percentage change	7,398	-24%	0.85 (0.84-0.86)	0,79	0,78
Quality of Life*					
EQ-5D final score	7,098	0.62	0.86 (0.85-0.87)	0,77	0,77
EQ-5D absolute change	6,585	0.06	0.79 (0.78-0.81)	0,71	0,76

The *final score* was the absolute value at 12 months follow up. The *absolute change* was the final score minus the preoperative score (negative values indicate improvement in ODI and NRS; positive values indicate improvement in EQ-5D). The *percentage change* was the absolute change divided by the preoperative score (negative values indicate improvement in ODI and NRS; positive values indicate improvement in EQ-5D).
* EQ-5D percentage change is not meaningful due to a denominator between -0.6 and 1.0

Table 4b PROM accuracy to identify failure (GPE=4-7) and worsening (GPE=6-7) 12 months after surgical treatment of spinal stenosis in 8,258 patients. An area under the curve (AUC) > 0.7 indicates acceptable sensitivity and specificity.					
		Worsening (GPE 6-7) n= 476/8258 (6%)			
Outcomes	n	Cut-off	Cut-off	Cut-off	Cut-off
Disability					
ODI final score	8,220	39	0.91 (0.90-0.92)	0.83	0.79
ODI absolute change	8,174	-4	0.86 (0.85-0.88)	0.77	0.79
ODI percentage change	8,161	-9%	0.87 (0.86-0.88)	0.80	0.80
Back Pain					
NRS back pain final score	8,174	6.5	0.90 (0.89-0.91)	0.86	0.82
NRS back pain absolute change	7,687	-0.5	0.83 (0.81-0.85)	0,78	0,77
NRS back pain percentage change	7,573	-12%	0.84 (0.82-0.85)	0,82	0,77
Leg Pain					
NRS leg pain final score	8,067	6.5	0.87 (0.86-0.89)	0,77	0,82
NRS leg pain absolute change	7,518	-0.5	0.82 (0.81-0.84)	0,72	0,79
NRS leg pain percentage change	7,398	-13%	0.83 (0.82-0.85)	0,80	0,73
Quality of Life*					
EQ-5D final score	7,098	0.53	0.90 (0.89-0.92)	0,86	0,81
EQ-5D absolute change	6,585	0.03	0.81 (0.79-0.83)	0,78	0,74
The *final score* was the absolute value at 12 months follow up. The *absolute change* was the final score minus the preoperative score (negative values indicate improvement in ODI and NRS; positive values indicate improvement in EQ-5D). The *percentage change* was the absolute change divided by the preoperative score (negative values indicate improvement in ODI and NRS; positive values indicate improvement in EQ-5D). * EQ-5D percentage change is not meaningful due to a denominator between -0.6 and 1.0					

7.5 Results paper 4 (dural tear)

Surgeons reported incidental durotomy in 439/8919 cases (4.9%). Patients who suffered an ID more often reported failure (adjusted OR (95%CI) 1.45 (1.12 – 1.87); p=0.005) and worsening (adjusted OR (95%CI) 1.50 (1.01 – 2.23); p=0.045, compared to patients with no ID (Table 5). Patients who suffered an ID during surgery reported a higher ODI score twelve months after surgery than those who did not

suffer an ID (ODI (95%CI) = 27.9 (26.2 – 29.6) vs 23.6 (23.3 – 24.0)). This difference remained significant after adjusting for possible confounders (beta (95 % CI) = 2.29 (0.58 – 4.00); p=0.009).

Furthermore, patients who suffered an ID reported more leg pain after surgery compared to patients without ID: mean NRS leg pain was 4.2 (3.9 – 4.5) vs 3.5 (3.5 – 3.6); this difference remained significant after adjusting for confounders (beta (95%CI) of 0.6 (0.3 – 0.9); p<0.001). Patients with ID had longer hospital stays than patients without ID (mean (95 % CI) 5.7(5.2-6.2) vs 3.3 (3.2 – 3.4); this difference remained significant after adjusting for confounders (beta (95%CI) 1.58 (1.25-1.92) days; p<0.001.

Among responders at a 3-month follow-up, 1259 (14.2 %) patients reported postoperative complications. The corresponding numbers for patients with ID were 105 (23.3 %), and for patients without ID, 1154 (13.7 %). Patients with ID had increased odds of urinary tract infection (UTI) after surgery (OR (95 % CI) 2.42 (1.53 – 2.73); p<0.001).

High age, gender (female), former surgery, and multilevel operations were associated with increased odds of ID.

Table 5. Multiple logistic regression using «failure»* and "worsening"** at 12 months follow-up as the dependent variable and dural tear and potential confounders as covariates.

	Failure		**Worsening**	
variables	OR (95%CI)	p-value	OR (95%CI)	p-value
Dural tear	**1.45 (1.12 – 1.87)**	**0.005**	**1.50 (1.01 – 2.23)**	**0.045**
Age	1.00 (1.00 – 1.01)	0.312	1.00 (0.99 – 1.01)	0.547
Gender (Female)	0.88 (0.78 – 1.00)	0.043	0.97 (0.78 – 1.20)	0.769
Body Mass Index (cont)	1.01 (1.00 – 1.03)	0.028	1.00 (0.97 – 1.02)	0.659
Smoking	1.41 (1.22 – 1.64)	0.000	1.53 (0.93 – 1.54)	0.001
ASA (3+4+5)***	1.13 (0.97 – 1.31)	0.125	1.20 (0.94 – 1.54)	0.153
Preoperative ODI (cont)****	1.01 (1.01 – 1.02)	0.000	1.03 (1.02 – 1.04)	0.000
Preoperative NRS leg pain*****	0.93 (0.89 – 0.96)	0.000	0.96 (0.90 – 1.02)	0.202
Preoperative NRS back pain*****	1.10 (1.05 – 1.14)	0.000	1.15 (1.07 – 1.24)	0.000
Duration leg pain >12mts	1.63 (1.43 – 1.87)	0.000	1.54 (1.22 – 1.94)	0.000
Former surgery at same level	1.92 (1.64 – 2.26)	0.000	1.79 (1.38 – 2.31)	0.000
More than one level operated	0.90 (0.79 – 1.02)	0.092	0.93 (0.74 – 1.15)	0.490
Additional fusion, any type	0.62 (0.51 – 0.75)	0.000	0.61 (0.43 – 0.85)	0.003

- *Defined as Global Perceived Effect (GPE) 4-7 (unchanged or any degree of worsening) at 12 months
- **Defined as Global Perceived Effect (GPE) 6+7 ("much worse" or "worse than ever") at 12 months
- ***American Society of Anesthesiologists classification (1-5) (grade 3 to 5)
- ****Oswestry Disability Index (0 -100), increasing for increasing disability
- *****Numeric Rating Scale (0 -10), increasing for increasing pain

7.6 Results paper 5 (predictors for failure)

Failure: Table 6a displays the predictors for failure. The most substantial independent risk factors for failure identified in the multivariate model were duration of back pain >12 months (OR=2.17 (1.88 – 2.50); p<0.001), former spinal surgery (OR=2.21 (1.94 – 2.51); p<0.001) and age >70 years (OR=1.99 (1.71 – 2.31); p<0.001). Socioeconomic variables, i.e. receiving disability benefits, low educational level, not being a native Norwegian speaker, and living alone, all increased the odds of failure (OR between 1.34 – 1.66). Variables concerning general health, i.e., smoking, BMI >30, and ASA>2, also increased the odds of failure (OR 1.32 – 1.40). Higher preoperative ODI score (spine-related disability increased the odds for failure (OR 1.06 (1.05-1.06; p<0.001)). Of the radiological variables, only the finding of degenerative olisthesis on x-ray affected the odds for failure with decreased odds (OR=0.76 ((0.64 – 0.89); p<0.001)).

Worsening: Table 6b displays the predictors for worsening. The most substantial independent risk factors for worsening identified in the multivariate model were former spinal surgery (OR=2.00 (1.74 – 2.30); p<0.001), duration of back pain >12 months (OR=1.85 (1.47 – 2.32); p<0.001), and age > 70 years (OR=1.93 (1.62 – 2.31); p<0.001). Socioeconomic variables, i.e., receiving a disability benefit, low educational level, and living alone, increased the odds of worsening (OR between 1.26 – 1.66). Variables concerning general health, i.e. BMI >30 and ASA >2, increased the odds of worsening (OR 1.33 – 1.39). High preoperative ODI score and duration of leg pain > 12 months increased the odds for worsening (OR 1.07 – 1.29). None of the preoperative radiological variables influenced the odds of worsening.

The proportion of patients that reported failure increased by the number of spinal levels operated on. Previous spine surgery increased the odds of failure, but if the previous surgery had been done at the same or another level did not matter (48% failures reported by patients who had previously received surgery at the same spinal level, compared to 47% for those who were previously operated at another spinal level).

Table 6a: Logistic regression for 8919 patients operated for lumbar spinal stenosis and registered in NORspine during 2007-2017, using failure (ODI>31) as dependent variables and potential predictors as explanatory variables

	Univariate		Multivariate	
Variables	OR (95%CI)	p-value	OR (95%CI)	p-value
Age >70 years	1.50 (1.37 – 1.64)	<0.001	1.99 (1.71 – 2.31)	<0.001
Gender (female)	1.44 (1.32 – 1.57)	<0.001		
Smoking	1.46 (1.31 - 1.63)	<0.001	1.40 (1.21 – 1.62)	<0.001
Body mass index >30	1.54 (1.39 – 1.70)	<0.001	1.34 (1.18 – 1.53)	<0.001
ASA grade >2 *	2.05 (1.85 – 2.28)	<0.001	1.34 (1.16 – 1.54)	
Education level below college	1.99 (1.79 – 2.21)	<0.001	1.54 (1.35 – 1.75)	<0.001
Civil status, living alone	1.62 (1.46 – 1.78)	<0.001	1.33 (1.17 – 1.52)	<0.001
Not Native Norw speakers	1.58 (1.26 – 2.00)	<0.001	1.66 (1.23 – 2.23)	0.001
Disability benefit (all types)**	1.46 (1.33 – 1.60)	<0.001	1.67 (1.44 – 1.94)	<0.001
Former lumbar spine surgery (any)	2.26 (2.05 – 2.50)	<0.001	2.21 (1.94 – 2.51)	<0.001
MRI central stenosis	1.05 (0.95 – 1.15)	0.358		
MRI lateral stenosis	0.91 (0.83 – 1.00)	0.040		
MRI foraminal stenosis	1.18 (1.02 – 1.36)	0.024		
RF degen olisthesis	0.85 (0.75 – 0.97)	0.013	0.76 (0.64 – 0.89)	0.001
Pre opr ODI (cont)***	1.06 (1.06 – 1.07)	<0.001	1.06 (1.05 – 1.06)	<0.001
Duration leg pain >12months	1.68 (1.52 – 1.86)	<0.001		
Duration backpain >12months	1.87 (1.68 – 2.10)	<0.001	2.17 (1.88 – 2.50)	<0.001
Multilevel surgery ****	1.21 (1.11 – 1.33)	<0.001		

* ASA = American Society of Anesthesiologists classification (1-5)
** All types of disability benefit, both full and partly supported
*** ODI = Oswestry Disability Index (0-100), indicating increasing disability
**** More than one level operated

Table 6b: Logistic regression for 8919 patients operated for lumbar spinal stenosis and registered in NORspine during 2007-2017, using worsening (ODI>39) as dependent variables and potential predictors as explanatory variables				
	Univariate		Multivariate	
Variables	OR (95%CI)	p-value	OR (95%CI)	p-value
Age >70 years	1.50 (1.36 – 1.66)	<0.001	1.93 (1.62 – 2.31)	<0.001
Gender (female)	1.36 (1.23 – 1.51)	<0.001		
Smoking	1.52 (1.35 – 1.71)	<0.001	1.53 (1.31 – 1.80)	<0.001
Body mass index >30	1.53 (1.36 – 1.71)	<0.001	1.33 (1.15 – 1.54)	<0.001
ASA grade >2 *	2.14 (1.91 – 2.40)	<0.001	1.39 (1.19 – 1.62)	<0.001
Education level below college	1.95 (1.72 – 2.21)	<0.001	1.51 (1.29 – 1.76)	<0.001
Civil status, living alone	1.52 (1.37 – 1.71)	<0.001	1.26 (1.09 – 1.45)	0.002
Not Native Norw speakers	1.49 (1.16 – 1.92)	<0.001		
Disability benefit (all types)**	1.47 (1.32 – 1.63)	<0.001	1.66 (1.40 – 1.98)	<0.001
Former lumbar spine surgery (any)	2.19 (1.96 – 2.44)	<0.001	2.00 (1.74 – 2.30)	<0.001
MRI central stenosis	1.05 (0.94 – 1.17)	0.428		
MRI lateral stenosis	0.90 (0.81 – 1.00)	0.044		
MRI foraminal stenosis	1.14 (0.97 – 1.34)	0.120		
RF degen olisthesis	0.92 (0.80 – 1.06)	0.255		
Pre opr ODI (cont)***	1.07 (1.07 – 10.7)	<0.001	1.07 (1.06 – 1.07)	<0.001
Duration leg pain >12months	1.74 (1.55 – 1.96)	<0.001	1.29 (1.06 – 1.56)	0.010
Duration backpain >12months	1.95 (1.70 – 2.24)	<0.001	1.85 (1.47 – 2.32)	<0.001
Multilevel surgery ****	1.19 (1.07 – 1.32)	0.001		
* ASA = American Society of Anesthesiologists classification (1-5) ** All types of disability benefit, both full and partly supported *** ODI = Oswestry Disability Index (0-100), indicating increasing disability **** More than one level operated				

8 DISCUSSION

8.1 Study design and registry data

Several national registries for spine surgery have been developed over the last 20 years. A systematic review from 2015 identified 25 different spine registries (88). The Scandinavian countries have adapted registries early (2). Registry studies rely on prospectively collected data even if the study idea and design are retrospective. Medical registries have grown in popularity over the years and have some significant advances; they provide large populations that reflect daily practice, results are generalizable, and data collection is cheap and quick as data already exists (103, 113, 114). Data are collected prospectively (ideally with no recall bias) and by clinicians and patients independent from the actual study.
Furthermore, registry-based studies supplement results from more narrow studies, i.e., RCTs (114, 115). Even if registry studies are observational in design, registries can collect many variables allowing to adjust for confounders.

The main reasons for using registry data to answer our research questions were the possibility of studying large populations, as we planned to study complications, and focused on failure and worsening with predictors. We thought achieving this in a traditional prospective study would have been too difficult; the Norwegian population is spread over a sparsely populated area, and most of the treatment centres are small. Additionally, registry design has practical and economic advantages. Furthermore, there were already ongoing prospective studies on LSS patients, and we considered adding one more project, including more questionnaires, would be a too significant burden on this patient group.

There are critics of research on registry data. A systematic review from 2015 concluded that spine registries had not improved spinal care (88). The authors pointed out recommendations and that registries can show trends, monitor quality and ultimately improve care. The sample size in registry studies is essential to have in mind when interpreting the results, as large populations can provide statistically significant findings that are not clinically relevant; the effect estimate is more important than the p-value.

Other concerns when designing registries are how extensive the registry should be, what types of data should be recorded and which data sources are appropriate. The wish for data must be balanced against the risk of reducing completeness and introducing missing data and the risk of non-accurate data collection. NORspine has chosen to let the patients complete the form concerning the preoperative clinical status and socioeconomic data. The surgeon completes the topics on spinal diagnosis, radiological findings and indication for surgery, and the surgical details. The patients complete the two follow-ups, which include clinical outcomes (PROMs) and postoperative complications. One can question if this is the best way to collect these types of data – surgeons might not be sufficiently aware of comorbidity, patients might not be competent to report complications, and the patient might be biased by the behaviour of the surgeon and general impression of the treatiment centre when scoring the clinical result.

Cybersecurity is an area of increasing interest. Health information can be hacked, lost, stolen or changed. A study from 2020 reported that 94% of healthcare organizations had experienced one or more cyber attacks (127). Holding health information in two parallel databases (medical registries and EPR) and transferring medical information between different sources can increase the risk of displaced health information. Ownership of health data registered in medical registries can also be discussed. According to the legal considerations, the patient owns her/his information and can demand information in EPR changed and information in registries deleted. These are crucial facets of research, as trust in society and among patients is paramount to maintaining completeness and improving the response rate. Hence, stable and trustworthy organization of medical registries and reliable handling of registry data in hospitals are essential.

8.2 Study population and overall results

The study population was created with a combination of the diagnosis of LSS and surgical procedure decompression. Additional diagnoses and procedures were included, which introduces heterogeneity in our population. A heterogenic population could be problematic, but it is also an advantage, as it reflects the common everyday practice and improves the external validity of our studies. The inclusion period (2007-2017) was set to have 12 months follow-up.

Baseline data from the study population was a standard LSS population similar to other studies (6, 7, 11, 37). The preoperative ODI of 40 and the clinical outcome with a mean ODI score of 23.9 and a mean change in ODI score of 16 points is also in line with previously reported results (7).

Preoperative ODI had a normal distribution and no floor or ceiling effect (Figure 7). There was a tendency to floor effect 12 months after surgery (Figure 8); this is a known challenge. ODI is reported to have difficulty stratifying patients with low functional decline due to a floor effect (128). However, the ODI score is a validated and widely used PROM. As we aimed to study failure and worsening, the floor effect representing patients with good effects of treatment and low disabilities was not considered relevant.

During our study period, an evaluation of which PROMs are best suited for measuring patients with LSS was published; this study preferred the Zurich claudication questionnaire (129). However, we used ODI as NORspine recorded this, but we will consider adding ZCQ in future studies on LSS.

8.3 Paper 1 (data accuracy in NORspine)

Data quality is paramount in all research; prospective controlled trials (i.e. RCTs) and medical registries usually have procedures for ensuring data quality. However, in medical registries, many different treatment centres and medical staff contribute to data collection. Hence, data quality may vary. Furthermore, patients (non-medical staff) also often contribute to registry data collection, which may further impair the data quality. The development of artificial intelligence (AI) emphasizes the importance of data quality, as AI totally relies on data, and no humans can control the algorithms or results. Data quality affects not only medical research; a business study from 2017 tested the data quality in various companies and found that only 3% of the companies had a data quality over the chosen cut-off at 97% accuracy (130).

Our data quality assessment found an almost perfect agreement for patients' demographics and a strong agreement for surgical details but a weak agreement for perioperative complications and comorbidities; the registry underreported complications and comorbidities. Our findings are in line with previous studies (119, 131). Registration of data closely related to the surgeons' speciality seems to be easier than data far from the surgeons' speciality. Previously published studies reported high accuracy when orthopaedic surgeons coded surgical procedures and classifying x-rays and low accuracy when the surgeons coded diagnoses, assessed cognitive function and registered the antibiotics (132 133, 134).

The level of data quality in the NORspine registry could be illustrated by the difference in agreement between previous surgery; the dichotomous variable previous surgery yes/no had excellent agreement (kappa 0.93), while the number of previous surgeries had a considerably inferior agreement (kappa 0.62).

Spine registries have developed to become essential contributors to science. We think focusing on data quality and knowledge about challenges in recording clinical data is essential when interpreting existing research and planning new research projects. Another Spine registry in Germany published a validation study after we had planned our study and just before our publication (119); this illustrates the international focus on this field.

Data quality should be a future focus for medical registries, and revising the questionnaires to ease understanding and completion may contribute to better data quality. The extent of the questionnaires should be discussed within each registry; one should only record relevant data of high quality. Making the forms shorter could be a way to improve concordance and increase the coverage

and follow-up rates, as shorter forms may decrease the workload and burden for patients and health care personnel completing them. Furthermore, combining patient - and surgeon-registered data into a combined construct could also help increase data quality (135). Artificial intelligence (AI) could be used to extract information to the registry from unstructured EPRs or to control information in medical registries. A study from 2022 reported promising results for a language algorithm to identify medical information in unstructured data (136).

Limitations:

Our study population was made by selecting four different hospitals over two years, resulting in 474 patients. The selection of treating centres was not random but done for practical and legal reasons; we had to limit the data collection to treating centres where the authors worked. This introduced a possible selection bias as data agreement and accuracy can differ between treating centres. The study sample differed somewhat from the entire NORspine population at baseline; the included patients had more comorbidity, higher BMI, and higher disability (ODI) and pain scores (NRS) for back and leg pain. In addition, the study population had more smokers and fewer perioperative complications than the total LSS population registered in NORspine. However, the differences between the groups were minor, and we consider the sample representative.

When controlling data quality, one needs an alternative data source with corresponding data as a reference set. Ideally, the alternative data source is correct and considered a “gold standard”. The variables should be in the same form (continuous or categorical) and use the same scales or classification systems. However, gold standards hardly exist in the medical world, and variables are sometimes recorded differently. There are numerous possibilities for misrecordings in all data sources; wrong measurement, misinterpretation, misspellings or incorrect plotting. These challenges limited our choice of variables possible to compare.

We chose electronic patient records (EPR) as the comparable data source. EPR is compulsory in Norway and has a solid legal stand. EPRs have also been used in previous validation studies (119, 135, 137, 138). Recording data from EPR is challenging as EPR does not consist of structured data. We had to read a free text and interpret it into the actual variables /categories. To ensure reliable registration when collecting data from EPR, two authors independently recorded data from 22 of the same patients, and the interrater was almost perfect. Twenty-two patients might be too small a sample to estimate interrater reliability; this was done as a practical solution.

We considered some of the variables collected from EPR as “gold standards”, while others were considered only as data from an alternative source. We calculated agreement for all variables, and in addition, we calculated accuracy only when the corresponding data set could be regarded as a “gold standard”. We would not pretend to have a better reference data source than we had. When presenting accuracy, we chose to report proportion correctly classifies (PCC) and sensitivity as these are commonly used and easily understood.

PROMs are essential in evaluating clinical outcomes in spine surgery. Unfortunately, our EPR did not include PROM or any systematic evaluation of the outcome. The surgeon or physiotherapist had written a few words about the clinical outcome for some patients. However, this was insufficient to categorize or translate into any variable similar to what NORspine had recorded. Hence, we could not evaluate any variables concerning the clinical result.

The surgeons recorded comorbidities in NORspine as “*relevant* comorbidities”, while EPR ideally recorded all comorbidities. We think these data sources are difficult to compare as they measure comorbidity somewhat differently. Therefore we chose to present comorbidity only as prevalences in each data source rather than agreement or accuracy.

Postoperative complications are recorded in the NORspine by the patients and in EPR by medical staff, and *only if* the patient contacts the same treating centre again. We found these two data sets too different to compare concerning agreement, accuracy or prevalence. We will consider analyzing data agreement regarding postoperative complications in a later study.

We evaluated patients who operated for LSS; this is often easy surgery with few technical variations and a low risk for complications. Hence, our study sample might be a “best-case scenario” for assessing data quality regarding surgical details and complications.

8.4 Paper 2 (characteristics of non-responders)

In our population, there was a loss to follow-up at 30%. This is within the recommended follow-up rate of 60-80% recommended by van Hoof; however, no consensus exists as an ultimate limit (88, 139). Our main finding was that patients lost to follow-up had only minor differences in patient characteristics, and they had similar clinical outcomes as responders. Our findings are supported by previously published data (116, 117, 118, 140).

Missing data can be classified as missing at random (MAR), missing completely at random (MCAR), and missing at non-random (MNAR) (141). The classification of missing data is essential as it has implications for analyzing the data. Random dropout is easier to handle than a systematic loss to follow-up (attrition bias). In MAR, the non-responders may differ at baseline from responders but still report similar clinical outcomes; MCAR means the groups are similar at baseline and follow-up; in MNAR, the two groups report different clinical outcomes. MNAR represents an attrition bias risk as the results are based only on respondents. There are methods to handle MNAR; multiple imputations and mixed linear models (142).

Previous studies conclude the loss to follow-up to be of MAR type in the Norwegian and Danish spine registries (116, 117). We also concluded that data was missed at random since only baseline characteristics had some minor differences, while clinical outcomes were similar. We recognize that this conclusion might have been too light as multiple imputations are advised in registry analysis (88).

NORspine has three and twelve-12 months follow-up. A literature review from 2020 showed decreasing FU rates with increasing follow-up time; in addition, a review from 2011 reported that more questions also increased loss to follow-up (143, 144). Reducing the follow-up time to 3 months and the number of questions could increase the follow-up rate. Another possible way to increase follow-up rates could be to randomize follow-up for a smaller group and focus on high follow-up rates among the randomly selected subpopulation. The New Zealand joint registry randomly selects 20% of hip and knee replacements for follow-up; this may also cut costs related to a complete follow-up (145).

Limitations.

We used simple statistics comparing the two groups. T-test presumes normality, and this was tested prior to the analyses. More sophisticated methods could have been used, for instance, logistic

regression analysis using loss to follow up as a dependent variable. This could allow us to build a complete model adjusting the different predictors for being lost to follow-up.

Furthermore, our study was vulnerable to multiple testing; we tested 13 baseline variables and four outcome variables, and as every test has a 5% risk of discovering a "significant" difference by chance (p-value chosen at 0.05), some of our findings could be a result of multiple testing (type 1 error). Furthermore, the study group was only 474 patients; we had no power analysis and might have missed relevant differences (type 2 error).

We applied to The Norwegian Regional Committee for medical and health research ethics and were allowed to contact the patients who had not responded to NORspine with two letters and one SMS. We were not allowed to phone or in any other way make contact with the non-responders. This limitation leads to a follow-up rate among non-responders of only 52%. In a former non-responder study in NORspine, they phoned the patients and made contact with 97% (116).

We contacted the non-responders several years after the operation, median time was 50 months; this may affect the results as we compared the outcome after 12 months for responders with the outcome for non-responders after 50 months. However, clinical results after surgery for LSS are stable; hence, this limitation may not be essential (118, 146, 147, 148).

8.5 Paper 3 (criteria for failure)

To define failure, we found an ODI final score of 31 points or more, a percentage change in ODI of less than 20%, or an absolute change of 8 points or less to be the most accurate cut-offs. Furthermore, to define worsening, we found an ODI final score of 39 points or more, a percentage change of less than 9%, or an absolute change of 4 points or less to be the most accurate cut-offs. We also found cut-offs for NRS back – and leg pain (final score, absolute change and percentage change). The ROC analyses showed that these cut-offs have good or excellent ability to identify failure and worsening; these cut-offs can be used to categorize clinical outcomes. Our findings for the final ODI score align with previous studies on PASS score and success criteria which found cut-offs for success at 22-24 ODI points (109, 110).

Our finding of failure, defined as 8 points improvement in ODI and worsening as 4 points improvement, may be hard to understand. ODI is a validated PROM with good intra-person psychometric capability. One may think that the threshold between improvement and failure is an ODI change of zero because ODI improvement should reflect clinical improvement; furthermore, clinical worsening should increase the ODI score. However, the direction of change has shown to be important, as clinically meaningful change has been reported differently for improvement and deterioration (149, 150). We believe patients expect a certain improvement after surgical treatment, and when this expectation is not met, patients feel the treatment as failure. Another possibility is that when a patient knows there is a possibility for improvement in future surgical care, this has some relieving effect per se and that when surgery is performed and "expended" without a significant clinical improvement, patients have lost one possibly relieving asset; hence they might perceive their symptoms as worse.

Different methods for categorizing a continuous clinical outcome (PROM) into success (or failure/worsening) exist. One can use purely statistical methods or anchor-based methods. The statistical methods are **distribution based**, using the variance in the data set to calculate cut-offs. *Effect sizes* are a method using the mean change divided by the standard error described by Kazai, and reliable *Change index* is another method (151, 152). *Standards error of measurement (SEM)* quantifies the amount of error and random variation when the measurement is repeated. The main limitations of the distribution-based methods are that they do not relate to clinical improvement, are sample-specific, and are not suited to generalize (153, 154).

Anchor-based methods are traditionally viewed as the gold standard for defining a specific clinical outcome (154). There are several ways to use an external anchor; one can use the threshold category (i.e. “somewhat better”) and calculate the mean PROM value within this subgroup. This value will represent the cut-off for patients perceiving themselves as better. An advantage of this method is that it might be more robust when the continuous clinical outcome variable has a ceiling effect, as only the patients experiencing a minimum improvement are used.
Receiver operating characteristics curve (ROC) analysis is another anchor-based method. This method parallels diagnostics, using the external anchor as a “gold standard” (diagnosis). To find the PROM cut-off value, one must search for the best combination of sensitivity and specificity to “diagnose” the categorized outcome (i.e. failure). The ROC method is widely used in this manner, and the area under the curve (AUC) displays the “quality/accuracy” of the chosen PROM. Hence, ROC analysis can compare the properties of different continuous variables (PROMs).

The AUC value is a measure of the test properties. The AUC is the probability that the PROM in a randomly selected failure patient is higher than in a non-failure patient. An AUC of 0.5 means the test is no better than flipping a coin (155). In our study, the AUCs were over 0.80 for all PROM-derivates; this is considered as good or excellent (123).

Different methods are used to precisely identify the best combination of sensitivity and specificity. One can find the point on the ROC-curve closest to the upper left corner (Figure 9), one can calculate the Youden Index, the sum of sensitivity and (specificity) and choose the highest combination, or use the point on the curve where sensitivity and specificity are equal (155).
In certain diagnostic or screening circumstances, misclassifying patients as either positive or negative could be more critical. In our study, we had no priority regarding what was most important. Hence, we think the method to identify the cut off-point was less important; we chose to find the point on the curve closest to the upper left corner. However, this method introduces a potential error as one must transfer this point to the table describing the curve to find the cut off-value and corresponding sensitivity and specificity.

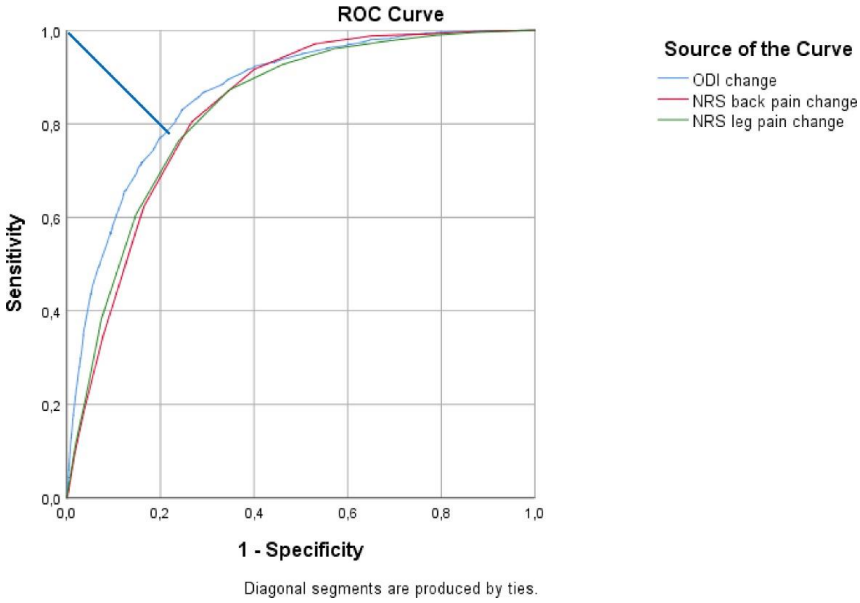


Figure 9. ROC curves for change scores for ODI (blue), NRS back pain (red), and NRS leg pain (green). The straight blue line displays the closest point to the upper left corner. The right side of each curve describes the area under the curve and is between 0 and 1.00.

Limitations:

The choice of external anchor (GPE) can be discussed. The patients answer the GPE simultaneously as they complete the ODI and NRS scores; hence, the GPE might not be entirely external. Furthermore, GPE might be subject to recall bias as the current clinical state has shown to dominate when patients determine their GPE change category (156, 157, 158, 159). The reliability of transition scales can also be questioned. One study reported moderate to substantial reliability for these scales (160). Since ODI, NRS and GPE are recorded simultaneously (by the same questionnaire) after surgery; one could argue that one of the other PROMs could have been the anchor. However, GPE is recommended as anchor according to an article from 2008 and is stated to provide a reliable assessment of health transition (102, 161).

We used the cut-off from this study in paper 5, and these cut-offs defined about 30% of the LSS patients as failures and 20% as worse. These proportions are higher than what the GPE instrument displays (21% failure and 6% worse). Higher proportions of failure and worsening may be a

consequence of the method for defining the cut-offs (ROC analyses with the best *combination* of sensitivity and specificity); hence, the cut-off might overestimate the true proportion of failure and worsening. This is illustrated by the distributions of failures and "non-failure" in figures 10 and 11, as the proportion of failure is far smaller than the proportion of "non-failure": The tails of the false positive "non-failure" overlap and is larger than the tail of false negative failures. Hence, these cut-offs overestimate the proportion of failures (and worsening).

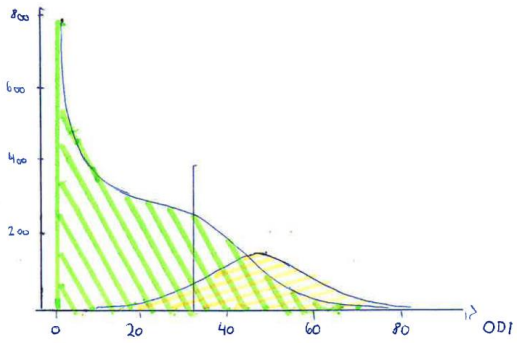


Figure 10. Distributions of failure (yellow) and non-failure (green) according to GPE score; ODI final score. Cut-off of 31 points marked.

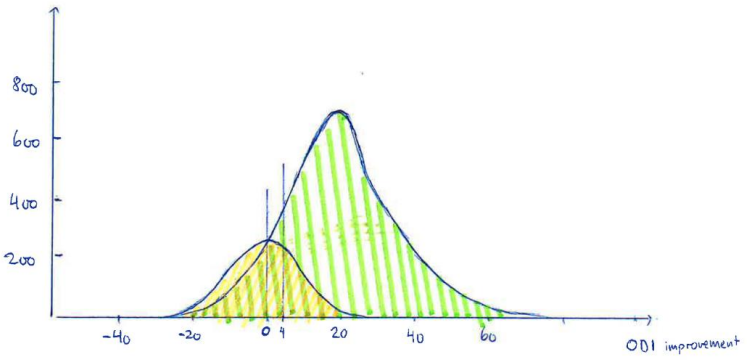


Figure 11. Distributions of failure (orange) and "non-failure" (green) according to GPE; ODI improvement. Cut-off of 4 points and 0 points marked.

Using the final PROM score to define failure (and worsening) raises another significant criticism, could cut-offs be different in patients with an extensive disability compared to patients suffering only mild disability before surgery? One study on patients with back pain reported that patients with more severe pain required a greater change to perceive the treatment as successful than patients with less severe pain (162). Using percentage change derivates is one way to handle this concern.

There are also critics of the specific ROC method. ROC analyses are sensitive to random sample variation, the non-parametric methods make constructing confidence intervals difficult, and ROC analyses do not adjust for factors modifying the clinical outcome (other than analyzing subgroups) 163). *The predictive modelling* approach uses logistic regression analysis to solve these concerns. Predictive modelling has been shown to have more accuracy in populations with skewed outcomes and if a ceiling or floor effect exists. A Mannion recently presented this method on a spine population at the Eurospine 2022 congress (164).

Using data recorded only after the surgery (ODI final score and GPE) to categorize the clinical result may be too uncertain. Recall bias and assessing the two PROMs by the same form might weaken the anchor. Figures 10 and 11 display a considerable overlap between the failure and “non-failure” populations. This is a drawback of the cut-off concept; wherever we set the cut-off, large proportions of either failure or “non-failure” patients will be misclassified. One can also imagine patients improving in only one domain (i.e. less pain) and not improving in other domains (i.e. disability) or the opposite. Is this failure? Designing a construct cut-off, combining the different PROMS (ODI + NRS back pain and leg pain + EQ-5D) in an algorithm (i.e. failure defined as reaching three out of four cut-offs), could help categorize such ambivalent clinical outcomes. Another idea could be to use qualitative methods to define failure (or worsening) and as an anchor. A third way could be to define individual goals for each patient before surgery and use these goals as definitions for success. Pain and disability are subjective symptoms, and the effect of treatment might best be measured related to the subjective expectations each patient has to symptom relieve before surgery. A preoperatively recorded PROM where the patients define their individual minimal acceptable change before surgery using modified ODI and NRS forms could be used as the anchor. This has been tested before with exciting results, i.e. patients' expectations exceeding actual outcomes (165, 166, 167, 168).

8.6 Paper 4 (dural tear)

We found that dural tears were associated with increased odds of failure and worsening. When exploring the effect of ID on ODI score, the effect estimate was minor and far under the MCID values. Our findings align with previously published data (55. 58. 124, 169, 170). However, adapting MCID criteria to group differences may mislead the conclusions. This method may oversee an inferior result in parts of one group, and categorized outcomes are advised in a review article (171).

We are aware that confounding factors may affect the risk for both dural tears and certain clinical results (failure). For example, previous surgery may increase the risk of ID and the risk of failure (unfavourable clinical result); this is displayed in figure 12. To avoid measuring the effect of confounders, there are several statistical methods possible. To adjust for one confounder, one can analyze subgroups, i.e. split the population into previously operated or first operation, and analyze each subgroup separately. To account for many confounders, one can calculate a propensity score, a probability for dural tears, for each patient based on the confounding factors and match patients from each group with equal propensity scores. This result in a matched control design, where simple statistical methods can be used to compare the groups. A drawback with this method is that one may miss some patients that do not match, decreasing the total population and losing statistical power.

We adjusted for confounders by multiple logistic regression, using known confounding factors as covariates. Logistic regression allowed us to keep more patients and avoid decreased statistical power (6, 78, 172, 173). One could have considered a more advanced method; mixed modelling would have kept all patients and used the available data (3 and 12 months follow-up).

Limitations:

We used GPE to categorize clinical outcomes. GPE may be subject to recall bias; still, it is recommended as an outcome (156, 102). Why did we not use the ODI cut-offs presented in paper 3? We planned to use ODI as a secondary outcome to assess the exact effect of ID and chose GPE as the primary outcome to keep the primary and secondary outcomes apart.

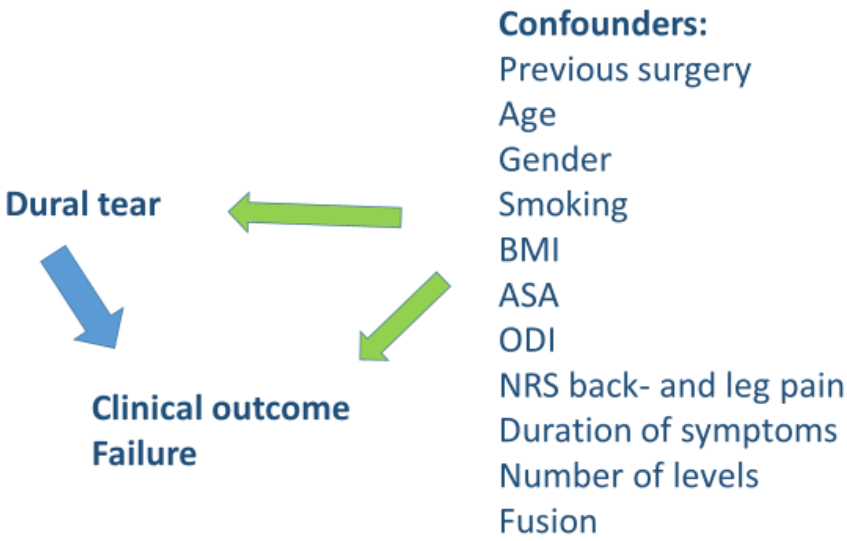


Figure 12

Causal relation between dural tear and clinical outcome. Green arrows shows the effect of confounding factors (list on the right).

Dural tears vary from small punctures to large defects, damage to the neural structures may vary, and the treatment will also vary from neglectance to watertight suture. These variations are not well captured in a registry. We have shown that complications such as dural tears are not always recorded (paper 1). Hence, the trustworthiness of registry data used to analyze complications like dural tears can be questioned. This is important when reading the published literature on this field. All the large population studies are registry-based.

Even if complications such as dural tears are rare and a large population is mandatory, our primary idea of using registry data might not be optimal for detecting these complications and the variety among them. Furthermore, the NORspine registry might not be suited to detect postoperative complications related to dural tears (re-operations, infections, etc.). We might have planned the dural tear study differently, knowing this. A prospective multicenter study focusing on data quality may have been more suitable. An alternative design could have been to supplement NORspine with detailed data from EPR on ID for a period, using the NORspine registry to look closer into a specific topic.

8.7 Paper 5 (predictors for failure)

Knowledge of predictors of the outcome of LSS surgery is central to the proper selection of patients. Certain predictors have been identified before; however, most focus has been on success. Failure or worsening after surgery are known outcomes after spine surgery and could be challenging to cope with for patients (and surgeons). Hence, information and advising on predictors for failure and worsening are essential. One recent systematic review on LBP found a significant association between socioeconomic factors such as low educational level and low income and clinical outcomes after surgery and emphasized the importance of understanding predictive factors for poor outcomes in research (174).

Our study identified several predictors of failure and worsening. The strongest predictors were previous spine surgery, duration of back pain of more than 12 months and age over 70 years. Additionally, we identified several socioeconomic factors associated with increased odds of failure and worsening (low education, living alone, not Norwegian speaking, and receiving disability benefits). Our findings are in line with previous literature, except for age, as there are divergent findings on whether age affects clinical outcomes (7, 37, 78, 174, 175, 176, 177, 178, 179, 180, 181). Our findings were consistent as the sensitivity analyses (repeated regression analyses using other definitions for failure and worsening) revealed almost identical results.

Interestingly preoperative MRI findings were not associated with failure or worsening. Previous studies confirm no associations between radiologic findings and symptoms or clinical outcomes; however, one study from 2017 found an association between the grade of LSS on MRI and clinical outcomes (7, 29, 181, 182). A recent Norwegian study on LSS patients supports this; they found no association between MRI findings (other than severe disk degeneration) and clinical outcome (183). The lack of association between preoperative MRI findings and clinical outcome does not mean one should not consider MRI before surgery. The above-mentioned studies are performed on populations already selected for surgery based partly on MRI findings.

Socioeconomic variables were associated with failure and worsening. According to the effect estimate, a combination of several socioeconomic variables seems to have a significant impact on clinical outcomes. NORspine records many socioeconomic factors, we included variables that are independent, and the selection of variables was tested for possible correlations. However, the

chosen variables can be mediators of a more basic patient-related trait or condition that are hard to define or measure.
Even if socioeconomic factors are associated with an increased risk of failure and worsening, it is crucial to keep in mind that all patients should have equal rights to health care. However, if the expected utility of the planned surgery is low, the best individual decision might be a non-surgical treatment for certain patients.

Understanding the properties of the chosen dependent outcome variable is essential when interpreting the results. We used the final ODI score over 31 (39) to define failure (worsening), and in the sensitivity analyses, we used an ODI improvement of less than 8 (4) points. We chose ODI final score as the outcome based on our findings in paper 3 and because a final score is easy to understand and use. Additionally, there are parallels between ODI final score and the PASS score assessing success reported by van Hooff et al. (ODI=22) and a criteria for success reported by Austevoll et al. (ODI=24) (109, 110). One Norwegian study from 2015 assessed predictors of failure; this study used an *increase* in ODI of 8 points (MCID) as the definition of failure (78). We question this definition, and this way of using MCID as the direction of change has shown to be relevant (149, 150).

Previous spine surgery was a strong predictor of failure and worsening, a recent Norwegian study on success rates after spinal reoperations confirmed this (184). We used the final ODI score as the dependent variable. Patients undergoing repeated spine surgery might have less chance of achieving a particular ODI final score than patients undergoing surgery for the first time. The recent study on outcomes after reoperations reported declining success rates from 66% after the first surgery to 22% after four (or more) spinal surgeries (184). A success rate of 22% is surprisingly low and does not fit well with our clinical judgement. Patients undergoing repeated spine surgery could have different expectations than patients undergoing first-time surgery. Hence, the threshold of patient-reported success might have to be redefined for patients undergoing multiple spinal surgeries.

Patients seeking help for spine-related symptoms might have tried conservative treatment before. They might be out of work and be physically and mentally compromised. Balancing considerations may be complex when faced with a possible solution (surgery). The patients translate all information provided, and this translation could be affected by the situation the patients are in. Negative information (risks of failure and worsening and complications) might be underweighted, and the patient's expectations of improvement might be overweighted. A parallel can be found in prospect

theory; people can weigh probabilities wrong (185). We emphasize the importance of communication with patients considering spinal surgery.

Limitations:

It can be tempting to analyze predictive factors by assessing how they affect the clinical result or how they result of selection; i.e. smoking may affect the biological healing process, hence the result. On the other hand, smoking can be associated with certain personality traits and mechanisms of coping with chronic pain (or chronic pain) and hence, a confounding factor. Our study was observational and can only assess associations rather than discover causality. Predictive factors are only associated with failure, not necessarily causing failure. The purpose of a predictor analysis is to improve patient selection and information. Even if one could intervene and amend some modifiable factors at baseline, we do not know if this would change the clinical outcomes. Furthermore, the potential predictive factors are limited to variables adapted by the NORspine registry, and they may be mediators or confounders. There could still be unrecorded variables with a substantial effect on clinical outcomes.

Even if logistic regression analysis and building a multiple logistic model is a standard method in predictor analysis, the overall importance of the variables included in our model was limited (186). Negelkarke R^2=0.292, cox and Snell R^2=0.209, meaning that only 20-30% of the variance seen is explained by the explanatory variables included in the model. We did not perform analyses to determine the relative importance of the different types of variables. We could have sequentially excluded groups of variables (i.e. socioeconomic or pain-related variables) and tested how this process affected the R^2 measures. Such an approach could be interesting in future predictor analyses and help point out which variables are more relevant than others.

We were aware of the pitfall of including all the NORspine variables in our prediction model. To reduce the possibility of accidental findings due to multiple testing, we used a p-value of 0.01. We paid attention to only including explanatory variables known from the literature or clinic. Our population was large, and according to a “rule of thumb” of ten events per possible predictor, our study was powered to identify predictors (186, 187). A recent study from Mannion presented in Eurospine in October 2022 reported that adding more than 20 predictors did not increase the explained variance (R^2) (164).

We focused on LSS diagnosis and decompression irrespective of surgical technique and additional fusion; our population was somewhat heterogenous but reflected everyday practice. Our focus was to identify predictors prior to surgery, irrespective of different surgical techniques. We performed subgroup analyses on patients that received additional fusion surgery with similar findings.
We did not impute missing data. We regarded missing data as missing at random (MAR); hence no need for imputing data (paper 2). We are aware of this limitation and will consider imputing data in future research, according to a previously published recommendation (186).

Our predictor analysis focuses on failure and worsening after LSS surgery. After planning this study (2016-2017), artificial intelligence (AI), machine learning and more complex predictive models have advanced. The Swedish spine registry published **Dialog Support** in 2021 and The Swiss Shulthess clinic **Prognostic tool** in 2022, both available as free internet services (177, 188, 189). These predictive tools are valuable assets in patient selection and information. They confirm our findings but have advanced some steps further.

8.8 Future implications

Based on the work in this thesis, some new research questions and ideas have evolved. Concerning data quality and concordance, we concluded that some variables had low agreement and accuracy. NORspine also collects data on MRI findings. We question the concordance of radiological data and plan a validation study where we assess the reliability of surgeon-reported MRI findings.

Data on postoperative complications were not applicable to validate in terms of agreement and accuracy. We plan to explore the prevalence of postoperative complications in NORspine compared to EPR and to explore the association between postoperative complications and clinical outcome and patient satisfaction.

The NORspine registry has a coverage of 70%; this opens questions about the patients never registered, which may introduce selection biases. Studies have been performed using a national patient registry to explore this topic, but this area could be supplemented by a study using clinical data from treating centres.

We are strengthened in the view that categorical outcome measures are helpful but still uncertain about the criteria for success and failure. However, the patient's expectations may vary according to baseline data and the patient's situation before surgery. Previous studies have reported a gap between expectations and actual outcomes (165, 166, 167, 168). We plan a study to explore patients' expectations of spine surgery using a modified NRS and ODI survey before surgery. The goal is to quantify how much pain (NRS) and disability (ODI) the patients think they will accept after surgery and if achieving these predetermined threshold scores will correlate with success and satisfaction.

Previous surgery is associated with increased odds of failure (184). However, clinical experience questions the impact of previous surgery. Defining stratified criteria for success and failure related to the number of previous spinal operations could be of interest. This could be performed by repeating our cut-off study (paper 3) using subgroups based on the number of previous spine surgery.

We find the identified predictors in paper 5 interesting. Constructing an algorithm to assess the risks for certain outcomes using our predictors for failure and worsening combined with predictors for success can be a future goal. Exploring the relative importance of the different categories of predictors could be one step towards this future goal.

9 Conclusions

NORspine data had varying accuracy; patient-provided data and surgical details showed a good concordance, while surgeon-provided medical information and complications showed a low concordance. Hence, the NORspine registry may not be optimal for analyzing certain aspects, such as comorbidities and complications. Patients lost to follow-up are somewhat different from patients completing follow-up, but the clinical result did not differ after LSS surgery. Research using registry data is valuable and necessary but must be planned thoroughly and interpreted carefully.

Clinical outcomes after surgery for LSS can be challenging to assess. PROMs are essential in disorders dominated by pain and disability. Interpreting PROMs is challenging, and a combination of continuous and categorical outcome variables may be helpful. We found that an ODI final score of >31 and an ODI percentage change of <20% defined failure best. For worsening, the corresponding cut-offs were an ODI final score of>39 and an ODI percentage change of <9%. Finally, 20-30% of patients reported failure after surgery for LSS, and 6-20% reported worsening, depending on the PROM used to define failure and worsening.

The most common perioperative complication was a dural tear, with an incidence of 5%. Patients who suffered from a dural tear had increased odds of failure and worsening.

Several patient characteristics were associated with clinical outcomes; back pain lasting longer than 12 months before surgery, previous spine surgery, and age over 70 increased the odds of failure and worsening. Socioeconomic factors were also associated with the clinical outcome.

The best treatment for spinal stenosis may include a proper selection of patients, evidence-based patient information and a safe surgical technique with minimized risk of complications. We hope that the results of the work in this thesis have contributed to the process of defining failure and worsening after surgery for LSS and to the identification of predictors in a pragmatic clinical setting that may aid surgeons in informing and selecting patients for surgery for LSS.

10 References

1. Weinstein JN, Lurie JD, Olson PR, Bronner KK, Fisher ES. United States' trends and regional variations in lumbar spine surgery: 1992-2003. Spine (Phila Pa 1976). 2006 Nov 1;31(23):2707-14. doi: 10.1097/01.brs.0000248132.15231.fe. PMID: 17077740; PMCID: PMC2913862.
2. Strömqvist B, Fritzell P, Hägg O, Jönsson B; Swedish Society of Spinal Surgeons. The Swedish Spine Register: development, design and utility. Eur Spine J. 2009 Aug;18 Suppl 3(Suppl 3):294-304. doi: 10.1007/s00586-009-1043-4. Epub 2009 Jun 4. PMID: 19495812; PMCID: PMC2899320
3. Jensen RK, Jensen TS, Koes B, Hartvigsen J. Prevalence of lumbar spinal stenosis in general and clinical populations: a systematic review and meta-analysis. Eur Spine J. 2020 Sep;29(9):2143-2163. doi: 10.1007/s00586-020-06339-1. Epub 2020 Feb 24. PMID: 32095908.
4. Kalichman L, Cole R, Kim DH, Li L, Suri P, Guermazi A, Hunter DJ. Spinal stenosis prevalence and association with symptoms: the Framingham Study. Spine J. 2009 Jul;9(7):545-50. doi: 10.1016/j.spinee.2009.03.005. Epub 2009 Apr 23. PMID: 19398386; PMCID: PMC3775665.
5. Nerland US, Jakola AS, Solheim O, Weber C, Rao V, Lønne G, et. al.: Minimally invasive decompression versus open laminectomy for central stenosis of the lumbar spine: pragmatic comparative effectiveness study. BMJ 2015; 350: pp. h1603
6. Aalto T, Sinikallio S, Kröger H, et. al.: Preoperative Predictors for Good Postoperative Satisfaction and Functional Outcome in Lumbar Spinal Stenosis Surgery — A Prospective Observational Study with a Two-Year Follow-up. Scandinavian Journal of Surgery December 2012; pp. 255-260
7. Sigmundsson Freyr G, Kang Xiao P, Jönsson Bo, Strömqvist Björn: Prognostic factors in lumbar spinal stenosis surgery. Acta Orthopaedica 2012; 83: pp. 536-542. To link to this article: 10.3109/17453674.2012.73391
8. Ogura Y, Kobayashi Y, Shinozaki Y, et. al.: Factors Influencing Patient Satisfaction After Decompression Surgery Without Fusion for Lumbar Spinal Stenosis. Global Spine Journal 2020; 10: pp. 627-632
9. Gunzburg R, Keller T, Szpalski M, et. al.: Clinical and psychofunctional measures of conservative decompression surgery for lumbar spinal stenosis: a prospective cohort study. Eur Spine J 2003; 12: pp. 197-204.
10. Kleinstück FS, Grob D, Lattig F, Bartanusz V, Porchet F, Jeszenszky D, O'Riordan D, Mannion AF.: The Influence of Preoperative Back Pain on the Outcome of Lumbar Decompression Surgery. Spine 2009; 34: pp. 1198-1203.
11. Hébert JJ, Abraham E, Wedderkopp N, et. al.: Preoperative Factors Predict Postoperative Trajectories of Pain and Disability Following Surgery for Degenerative Lumbar Spinal Stenosis. Spine (Phila Pa 1976) 2020; 45: pp. E1421-E1430
12. Chad DA. Lumbar spinal stenosis. Neurol Clin. 2007 May;25(2):407-18. doi: 10.1016/j.ncl.2007.01.003. PMID: 17445736.
13. Kreiner DS, Shaffer WO, Baisden JL, Gilbert TJ, Summers JT, Toton JF, Hwang SW, Mendel RC, Reitman CA; North American Spine Society. An evidence-based clinical guideline for the diagnosis and treatment of degenerative lumbar spinal stenosis (update). Spine J. 2013 Jul;13(7):734-43. doi: 10.1016/j.spinee.2012.11.059. PMID: 23830297.
14. www.spine.org/Research-Clinical-Care/Quality-Improvement/Clinical-Guidelines (accessed 4. Th Jan 2023)
15. Lee CK, Rauschning W, Glenn W. Lateral lumbar spinal canal stenosis: classification, pathologic anatomy and surgical decompression. Spine (Phila Pa 1976). 1988 Mar;13(3):313-20. doi: 10.1097/00007632-198803000-00015. PMID: 3388117.
16. Jezek J, Sepitka J, Daniel M, Kujal P, Blankova A, Waldauf P, Krbec M, Dousa P, Skala-Rosenbaum J, Samal F, Jirasek T. The role of vascularization on changes in ligamentum flavum mechanical properties and development of hypertrophy in patients with lumbar spinal stenosis. Spine J. 2020 Jul;20(7):1125-1133. doi: 10.1016/j.spinee.2020.03.002. Epub 2020 Mar 13. PMID: 32179155.

17. Yoshiiwa T, Miyazaki M, Kawano M, Ikeda S, Tsumura H. Analysis of the Relationship between Hypertrophy of the Ligamentum Flavum and Lumbar Segmental Motion with Aging Process. Asian Spine J. 2016 Jun;10(3):528-35. doi: 10.4184/asj.2016.10.3.528. Epub 2016 Jun 16. PMID: 27340534; PMCID: PMC4917773.
18. Tomkins-Lane C, Melloh M, Lurie J, Smuck M, Battié MC, Freeman B, Samartzis D, Hu R, Barz T, Stuber K, Schneider M, Haig A, Schizas C, Cheung JPY, Mannion AF, Staub L, Comer C, Macedo L, Ahn SH, Takahashi K, Sandella D. ISSLS Prize Winner: Consensus on the Clinical Diagnosis of Lumbar Spinal Stenosis: Results of an International Delphi Study. Spine (Phila Pa 1976). 2016 Aug 1;41(15):1239-1246. doi: 10.1097/BRS.0000000000001476. PMID: 26839989; PMCID: PMC4966995.
19. Olmarker K, Holm S, Rosenqvist AL, Rydevik B. Experimental nerve root compression. A model of acute, graded compression of the porcine cauda equina and an analysis of neural and vascular anatomy. Spine (Phila Pa 1976). 1991 Jan;16(1):61-9. PMID: 2003239.
20. Olmarker K. Spinal nerve root compression. Nutrition and function of the porcine cauda equina compressed in vivo. Acta Orthop Scand Suppl. 1991;242:1-27. PMID: 1645923.
21. Katz JN, Harris MB. Lumbar Spinal Stenosis. N Engl J Med 2008; 358:818-825 DOI: 10.1056/NEJMcp0708097
22. Sekiguchi M, Kikuchi S, Myers RR. Experimental spinal stenosis: relationship between degree of cauda equina compression, neuropathology, and pain. Spine (Phila Pa 1976). 2004 May 15;29(10):1105-11. doi: 10.1097/00007632-200405150-00011. PMID: 15131438.
23. Solberg, T. K., Ingebrigtsen, T., Olsen, L. R., & Thyrhaug, A. M. (2022). Årsrapport 2021: Nasjonalt kvalitetsregister for ryggkirurgi, resultater og forbedringstiltak. *Septentrio Reports*, (1). https://doi.org/10.7557/7.6865
24. Ulrich NH, Burgstaller JM, Held U, Winklhofer S, Farshad M, Pichierri G, Steurer J, Porchet F; LSOS Study Group. The Influence of Single-level Versus Multilevel Decompression on the Outcome in Multisegmental Lumbar Spinal Stenosis: Analysis of the Lumbar Spinal Outcome Study (LSOS) Data. Clin Spine Surg. 2017 Dec;30(10):E1367-E1375. doi: 10.1097/BSD.0000000000000469. PMID: 28059949.
25. Katz JN, Dalgas M, Stucki G, Katz NP, Bayley J, Fossel AH, Chang LC, Lipson SJ. Degenerative lumbar spinal stenosis. Diagnostic value of the history and physical examination. Arthritis Rheum. 1995 Sep;38(9):1236-41. doi: 10.1002/art.1780380910. PMID: 7575718.
26. de Schepper EI, Overdevest GM, Suri P, Peul WC, Oei EH, Koes BW, Bierma-Zeinstra SM, Luijsterburg PA. Diagnosis of lumbar spinal stenosis: an updated systematic review of the accuracy of diagnostic tests. Spine (Phila Pa 1976). 2013 Apr 15;38(8):E469-81. doi: 10.1097/BRS.0b013e31828935ac. PMID: 23385136.
27. Aaen, J., Austevoll, I.M., Hellum, C. *et al.* Clinical and MRI findings in lumbar spinal stenosis: baseline data from the NORDSTEN study. *Eur Spine J* **31**, 1391–1398 (2022). https://doi.org/10.1007/s00586-021-07051-4
28. 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
29. Ogikubo O, Forsberg L, Hansson T. The relationship between the cross-sectional area of the cauda equina and the preoperative symptoms in central lumbar spinal stenosis. Spine (Phila Pa 1976). 2007 Jun 1;32(13):1423-8; discussion 1429. doi: 10.1097/BRS.0b013e318060a5f5. PMID: 17545910.
30. de Graaf I, Prak A, Bierma-Zeinstra S, et al. Diagnosis of lumbar spinal stenosis: a systematic review of the accuracy of diagnostic tests. 2006. In: Database of Abstracts of Reviews of Effects (DARE): Quality-assessed Reviews [Internet]. York (UK): Centre for Reviews and Dissemination (UK); 1995-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK73306/
31. 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
32. Schönström N, Willén J. Imaging lumbar spinal stenosis. Radiol Clin North Am. 2001 Jan;39(1):31-53, v. doi: 10.1016/s0033-8389(05)70262-1. PMID: 11221505

33. Schizas, Constantin MD, PD, FRCS*; Theumann, Nicolas MD, PD†; Burn, Alexandre MD*; Tansey, Rosamond MD*; Wardlaw, Douglas FRCS‡; Smith, Francis W. MD§; Kulik, Gerit PhD*. Qualitative Grading of Severity of Lumbar Spinal Stenosis Based on the Morphology of the Dural Sac on Magnetic Resonance Images. Spine 35(21):p 1919-1924, October 1, 2010. | DOI: 10.1097/BRS.0b013e3181d359bd
34. Katz JN, Dalgas m; Stucki G, Lipson SJ. Diagnosis of lumbar spinal stenosis. Rheumatic disease clinics of North America 1994
35. Zaina F, Tomkins-Lane C, Carragee E, Negrini S. Surgical versus non-surgical treatment for lumbar spinal stenosis. Cochrane Database Syst Rev. 2016 Jan 29;2016(1):CD010264. doi: 10.1002/14651858.CD010264.pub2. PMID: 26824399; PMCID: PMC6669253.
36. Amundsen T, Weber H, Nordal HJ, Magnaes B, Abdelnoor M, Lilleâs F. Lumbar spinal stenosis: conservative or surgical management?: A prospective 10-year study. Spine (Phila Pa 1976). 2000 Jun 1;25(11):1424-35; discussion 1435-6. doi: 10.1097/00007632-200006010-00016. PMID: 10828926.
37. Weinstein JN, Tosteson TD, Lurie JD, Tosteson ANA, 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
38. Matz PG, Meagher RJ, Lamer T, Tontz WL Jr, Annaswamy TM, Cassidy RC, Cho CH, Dougherty P, Easa JE, Enix DE, Gunnoe BA, Jallo J, Julien TD, Maserati MB, Nucci RC, O'Toole JE, Rosolowski K, Sembrano JN, Villavicencio AT, Witt JP. Guideline summary review: An evidence-based clinical guideline for the diagnosis and treatment of degenerative lumbar spondylolisthesis. Spine J. 2016 Mar;16(3):439-48. doi: 10.1016/j.spinee.2015.11.055. Epub 2015 Dec 8. PMID: 26681351.
39. Lurie J, Tomkins-Lane C. Management of lumbar spinal stenosis. BMJ. 2016 Jan 4;352:h6234. doi: 10.1136/bmj.h6234. PMID: 26727925; PMCID: PMC6887476.
40. Ammendolia C, Hofkirchner C, Plener J, et al. Non-operative treatment for lumbar spinal stenosis with neurogenic claudication: an updated systematic review. BMJ Open 2022;12:e057724. doi:10.1136/bmjopen-2021-057724
41. Armon C, Argoff CE, Samuels J, Backonja MM; Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology. Assessment: use of epidural steroid injections to treat radicular lumbosacral pain: report of the Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology. Neurology. 2007 Mar 6;68(10):723-9. doi: 10.1212/01.wnl.0000256734.34238.e7. PMID: 17339579.
42. Ammendolia C, Stuber KJ, Rok E, Rampersaud R, Kennedy CA, Pennick V, Steenstra IA, de Bruin LK, Furlan AD. Nonoperative treatment for lumbar spinal stenosis with neurogenic claudication. Cochrane Database of Systematic Reviews 2013, Issue 8. Art. No.: CD010712. DOI: 10.1002/14651858.CD010712
43. Physical Therapy vs Surgical Decompression for Lumbar Spinal Stenosis - Full Text View - ClinicalTrials.gov, (accessed 5.th Jan 2023)
44. Verbiest H. A radicular syndrome from developmental narrowing of the lumbar vertebral canal. J Bone Joint Surg Br. 1954 May;36-B(2):230-7. doi: 10.1302/0301-620X.36B2.230. PMID: 13163105.
45. Oppenheim Ueber Einklemmung bzw. Strangualtion der Cauda equine, deutsche medizinishe wochenschrift 1909
46. Cauchoix, J., M. Benoist, and V. Chassaing, Degenerative spondylolisthesis. Clin Orthop Relat Res, 1976(115): p. 122-9
47. Yasargil Microsurgical operation of herniated lumbar disc, Advances in neurosurgery 1977
48. Caspar A new surgical procedure for lumbar disc herniation causing less tissue damage through a microsurgical approach, Advances in neurosurgery 1977
49. Hermansen E, Austevoll IM, Hellum C, Storheim K, Myklebust TÅ, Aaen J, Banitalebi H, Anvar M, Rekeland F, Brox JI, Franssen E, Weber C, Solberg TK, Furunes H, Grundnes O, Brisby H, Indrekvam K. Comparison of 3 Different Minimally Invasive Surgical Techniques for Lumbar Spinal Stenosis: A Randomized Clinical Trial. JAMA Netw Open. 2022 Mar 1;5(3):e224291. doi: 10.1001/jamanetworkopen.2022.4291. PMID: 35344046; PMCID: PMC8961320.

50. Thomé C, Zevgaridis D, Leheta O, Bäzner H, Pöckler-Schöniger C, Wöhrle J, Schmiedek P. 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. doi: 10.3171/spi.2005.3.2.0129. PMID: 16370302.
51. H:M: Meyer A history of endoscopic lumbar spine surgery: what have we learned? Hindawi 2019). Park, H.K. and J.C. Chang, Microdecompression in spinal stenosis: a review. J Neurosurg Sci, 2014. 58(2): p. 57-64
52. Machado GC, Ferreira PH, Yoo RI, Harris IA, Pinheiro MB, Koes BW, van Tulder MW, Rzewuska M, Maher CG, Ferreira ML. Surgical options for lumbar spinal stenosis. Cochrane Database Syst Rev. 2016 Nov 1;11(11):CD012421. doi: 10.1002/14651858.CD012421. PMID: 27801521; PMCID: PMC6464992.
53. Anderson DB, Ferreira ML, Harris IA, Davis GA, Stanford R, Beard D, Li Q, Jan S, Mobbs RJ, Maher CG, Yong R, Zammit T, Latimer J, Buchbinder R. SUcceSS, SUrgery for Spinal Stenosis: protocol of a randomised, placebo-controlled trial. BMJ Open. 2019 Feb 13;9(2):e024944. doi: 10.1136/bmjopen-2018-024944. PMID: 30765407; PMCID: PMC6398750
54. Buck JS, Yoon ST. (2015) The Incidence of Durotomy and its Clinical and Economic Impact in Primary, Short-segment Lumbar Fusion: An Analysis of 17,232 Cases. Spine (Phila Pa 1976) 40(18):1444-50. doi: 10.1097/BRS.0000000000001025. PMID: 26426713.
55. Desai A, Ball PA, Bekelis K, Lurie J, Mirza SK, Tosteson TD, Weinstein JN. (2015) SPORT: Does incidental durotomy affect longterm outcomes in cases of spinal stenosis? Neurosurgery 76 Suppl 1(0 1):S57-63; discussion S63. doi: 10.1227/01.neu.0000462078.58454.f4. PMID: 25692369; PMCID: PMC4517439
56. Herren C, Sobottke R, Mannion AF, Zweig T, Munting E, Otten P, Pigott T, Siewe J, Aghayev E; Spine Tango Contributors.(2017) Incidental durotomy in decompression for lumbar spinal stenosis: incidence, risk factors and effect on outcomes in the Spine Tango registry. Eur Spine J. 26(10):2483-2495. doi: 10.1007/s00586-017-5197-1. Epub 2017 Jun 20. PMID: 28634709
57. Papavero L, Engler N, Kothe R. (2015) Incidental durotomy in spine surgery: first aid in ten steps. Eur Spine J. 24(9):2077-84. doi: 10.1007/s00586-015-3837-x. Epub 2015 Mar 4. PMID: 25735610
58. Strömqvist F, Sigmundsson FG, Strömqvist B, Jönsson B, Karlsson MK. (2019) Incidental durotomy in degenerative lumbar spine surgery - a register study of 64,431 operations. Spine J. 19(4):624-630. doi: 10.1016/j.spinee.2018.08.012. Epub 2018 Aug 30. PMID: 30172899
59. Ulrich NH, Burgstaller JM, Brunner F *et al.* (2016) The impact of incidental durotomy on the outcome of decompression surgery in degenerative lumbar spinal canal stenosis: analysis of the Lumbar Spinal Outcome Study (LSOS) data—a Swiss prospective multi-center cohort study. BMC Musculoskelet Disord. 17, 170. https://doi.org/10.1186/s12891-016-1022-y
60. Weiss H, Garcia RM, Hopkins B, Shlobin N, Dahdaleh NS. A Systematic Review of Complications Following Minimally Invasive Spine Surgery Including Transforaminal Lumbar Interbody Fusion. Curr Rev Musculoskelet Med. 2019 Jul 13;12(3):328-339. doi: 10.1007/s12178-019-09574-2. Epub ahead of print. PMID: 31302861; PMCID: PMC6684700.
61. Takai K, Matsumoto T, Yabusaki H, Yokosuka J, Hatanaka R, Taniguchi M. Surgical complications associated with spinal decompression surgery in a Japanese cohort. J Clin Neurosci. 2016 Apr;26:110-5. doi: 10.1016/j.jocn.2015.06.029. Epub 2016 Jan 12. PMID: 26791473.
62. 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. 2017 Feb;26(2):404-413. doi: 10.1007/s00586-016-4683-1. Epub 2016 Jul 15. PMID: 27421276
63. Försth P, Michaëlsson K, Sandén B. 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 Jul;95-B(7):960-5. doi: 10.1302/0301-620X.95B7.30776. PMID: 23814250.

64. Försth P, Ólafsson G, Carlsson T, et al. A randomized, controlled trial of fusion surgery for lumbar spinal stenosis. N Engl J Med 2016;374:1413-1423.
DOI: 10.1056/NEJMoa1513721
65. Austevoll IM, Hermansen E, Fagerland MW et al. Decompression with or without fusion in degenerative lumbar spondylolisthesis N Engl J Med 2021; 385:526-538
DOI: 10.1056/NEJMoa2100990
66. Lønne G, Fritzell P, Hägg O, Nordvall D, Gerdhem P, Lagerbäck T, Andersen M, Eiskjaer S, Gehrchen M, Jacobs W, van Hooff ML, Solberg TK. Lumbar spinal stenosis: comparison of surgical practice variation and clinical outcome in three national spine registries. Spine J. 2019 Jan;19(1):41-49. doi: 10.1016/j.spinee.2018.05.028. Epub 2018 May 21. PMID: 29792994
67. Lønne G, Schoenfeld AJ, Cha TD et al. Variation in selection criteria and approaches to surgery for lumbar spinal stenosis amongients treated in Boston And Norway. Clin Neurol nurosurg (2017), 156;77-82.
68. Hijikata, S. A. "A method of percutaneous nuclear extraction." *J Toden Hosp* 5.5 (1975): 39.
69. Schreiber A, Suezawa Y, Leu H. Does Percutaneous Nucleotomy With Discoscopy Replace Conventional Discectomy?: Eight Years of Experience and Results in Treatment of Herniated Lumbar Disc. Clinical Orthopaedics and Related Research 238():p 35-42, January 1989.
70. Yeung AT, Yeung CA. Advances in endoscopic disc and spine surgery: foraminal approach. Surgical Technology International. 2003 ;11:255-263. PMID: 12931309.
71. Ruetten S,Komp M, Merk H, Godolias G., "Surgical treatment for lumbar lateral recess stenosis with the full-endoscopic interlaminar approach versus conventional microsurgical technique: a prospective, randomized, controlled study," *Journal of Neurosurgery: Spine*, vol. 10, no. 5, pp. 476–485, 2009.
72. Ransom NA, Gollogly S, Lewandrowski KU, Yeung A. Navigating the learning curve of spinal endoscopy as an established traditionally trained spine surgeon. J Spine Surg. 2020 Jan;6(Suppl 1):S197-S207. doi: 10.21037/jss.2019.10.03. PMID: 32195428; PMCID: PMC7063306
73. Simpson AK, Lightsey HM, Xiong GX, Crawford AM, Minamide A, Schoenfeld AJ: Spinal endoscopy: evidence techniques, global trends, and future projections. The Spine Journal (2022), 22;64-74.
74. Wu PH, Kim HS, Jang IT. A Narrative Review of Development of Full-Endoscopic Lumbar Spine Surgery. Neurospine. 2020 Jul;17(Suppl 1):S20-S33. doi: 10.14245/ns.2040116.058. Epub 2020 Jul 31. PMID: 32746515; PMCID: PMC7410380
75. Lønne, Greger MD[*,†,‡]; Johnsen, Lars Gunnar MD, PhD[†,‡,§]; Rossvoll, Ivar MD, PhD[†,‡,§]; Andresen, Hege MSc[†,‡]; Storheim, Kjersti PhD[¶]; Zwart, John Anker MD, PhD[¶]; Nygaard, Øystein P. MD, PhD[†,‡,‖]. Minimally Invasive Decompression Versus X-Stop in Lumbar Spinal Stenosis: A Randomized Controlled Multicenter Study. Spine 40(2):p 77-85, January 15, 2015. | DOI: 10.1097/BRS.0000000000000691
76. Lønne, Greger MD[*,†,‡]; Johnsen, Lars Gunnar MD, PhD[†,‡,§]; Aas, Eline PhD[¶]; Lydersen, Stian PhD[‖]; Andresen, Hege MSc[†,‡]; Rønning, Roar MD[*]; Nygaard, Øystein P. MD, PhD[†,‡,‖,**]. Comparing Cost-effectiveness of X-Stop With Minimally Invasive Decompression in Lumbar Spinal Stenosis: A Randomized Controlled Trial. Spine 40(8):p 514-520, April 15, 2015. | DOI: 10.1097/BRS.0000000000000798
77. Burkhard M, Farshad M, Suter D, Cornaz F, Leoty L, Furnstahl P, Spirig JM. Spinal decompression with patient-specific guides, The Spine Journal (2022), 22;1160-68.
78. Nerland US, Jakola AS, Giannadakis C, Solheim O, Weber C, Nygaard ØP, Solberg TK, Gulati S. The Risk of Getting Worse: Predictors of Deterioration After Decompressive Surgery for Lumbar Spinal Stenosis: A Multicenter Observational Study. World Neurosurg. 2015 Oct;84(4):1095-102. doi: 10.1016/j.wneu.2015.05.055. Epub 2015 Jun 3. PMID: 26049114
79. Jones A, Zachariae R. Investigation of the interactive effects of gender and psychological factors on pain response. Bristish journal of health psychology 2010.
80. Kim, Ho-Joong, et al. GENDER DIFFERENCE OF SYMPTOM SEVERITY IN LUMBAR SPINAL STENOSIS: ROLE OF PAIN SENSITIVITY Pain Physican(2013) 16;E715-723.

81. Racine M, Tousignant-Laflamme Y, Kloda LA, Dion D, Dupuis G, Choiniere M. A systematic literature review of 10 years of research on sex/gender and experimental pain perception – Pain (2012), 153;602-618.
82. Sharon B. Young, PT, DSc, Charles Aprill, MD, Jennifer Braswell, PT, PhD, William K. Ogard, PT, PhD, J. Scott Richards, PhD, John P. McCarthy, PT, PhD, Psychological Factors and Domains of Neck Pain Disability, *Pain Medicine*, Volume 10, Issue 2, March 2009, Pages 310–318, https://doi.org/10.1111/j.1526-4637.2009.00571.x
83. Lunde IM. Patients' perceptions-a shift in medical perspective. Scand J Prim Health Care. 1993;11:98–104.
84. Fairbank JC, Pysent PB. The oswestry disability index. Spine 2000;25:2940-52
85. McCaffery M. Pain, clinical manual for nursing practice. Mosby;1989
86. Wells GA, Shea B, O\Connell J, Welch V, Losos M, Tugwell P. The Newcastle Ottawa scale (NOS) for assessing the quality of nonrandomized studies in meta-analyses, www.ohri.ca/programs
87. McCormick JD, Werner BC, Shimer AL. Patient-reported outcome measures in spine surgery. J Am Acad Orthop Surg. 2013 Feb;21(2):99-107. doi: 10.5435/JAAOS-21-02-99. PMID: 23378373.
88. 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(5):534-44. doi: 10.3109/17453674.2015.1043174. PMID: 25909475; PMCID: PMC4564774.
89. ICHOM. (International Consortium for Health Outcomes Measurements). Standard Set for Low Back Pain. 2014. Available from: http://www.ichom.org/project/low-back-pain/. Accessed on: 9.th February 2022.
90. EuroQolGroup: EuroQol–a new facility for the measurement of health-related quality of life. Health Policy 1990; 16: pp. 199-208.
91. 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 Feb;9(2):105-21. doi: 10.1016/j.jpain.2007.09.005. Epub 2007 Dec 11. PMID: 18055266.
92. Grotle M, Brox I. CROSS-CULTURAL ADAPTATION OF THE NORWEGIAN VERSIONS OF THE ROLAND-MORRIS DISABILITY QUESTIONNAIRE AND THE OSWESTRY DISABILITY INDEX *Journal of Rehabilitation Medicine* Vol. 35 Issue 5, pp. 241–247, 2003.
93. Godil SS, Parker SL, Zuckerman SL, Mendenhall SK, Glassman SD, McGirt MJ. Accurately measuring the quality and effectiveness of lumbar surgery in registry efforts The Spine J (2014), 14;2885-2891.
94. Djurasovic M, Glassman SD, Dimar JR 2nd, Crawford CH 3rd, Bratcher KR, Carreon LY. Changes in the Oswestry Disability Index that predict improvement after lumbar fusion. J Neurosurg Spine. 2012 Nov;17(5):486-90. doi: 10.3171/2012.8.SPINE12614. Epub 2012 Sep 21. PMID: 22998402.
95. Tonosu, J., Takeshita, K., Hara, N. *et al.* The normative score and the cut-off value of the Oswestry Disability Index (ODI). *Eur Spine J* **21**, 1596–1602 (2012). https://doi.org/10.1007/s00586-012-2173-7)
96. Roland M, Morris R. A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low-back pain. Spine (Phila Pa 1976). 1983 Mar;8(2):141-4. doi: 10.1097/00007632-198303000-00004. PMID: 6222486
97. Stucki G, Daltroy L, Liang MH, Lipson SJ, Fossel AH, Katz JN. Measurement properties of a self-administered outcome measure in lumbar spinal stenosis. Spine. 1996;21: 796–803. pmid:8779009
98. Abou-Al-Shaar H, Adogwa O, Mehta AI. Lumbar Spinal Stenosis: Objective Measurement Scales and Ambulatory Status. Asian Spine J. 2018 Aug;12(4):765-774. doi: 10.31616/asj.2018.12.4.765. Epub 2018 Jul 27. PMID: 30060388; PMCID: PMC6068421.
99. Ammendolia C, Côté P, Southerst D, Schneider M, Budgell B, Bombardier C, Hawker G, Rampersaud YR. Comprehensive Nonsurgical Treatment Versus Self-directed Care to Improve Walking Ability in

Lumbar Spinal Stenosis: A Randomized Trial. Arch Phys Med Rehabil. 2018 Dec;99(12):2408-2419.e2. doi: 10.1016/j.apmr.2018.05.014. Epub 2018 Jun 20. PMID: 29935152.
100. Solberg TK, Olsen JA, Ingebrigtsen T, Hofoss D, Nygaard OP. Health-related quality of life assessment by the EuroQol-5D can provide cost-utility data in the field of low-back surgery. Eur Spine J. 2005 Dec;14(10):1000-7. doi: 10.1007/s00586-005-0898-2. Epub 2005 Apr 21. PMID: 15843969.
101. Jensen, M. P., & Karoly, P. (2011). Self-report scales and procedures for assessing pain in adults. In D. C. Turk & R. Melzack (Eds.), *Handbook of pain assessment* (pp. 19–44). The Guilford Press.
102. Kamper SJ, Ostelo RW, Knol DL, Maher CG, de Vet HC, Hancock MJ.: 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: pp. 760-766. e1
103. Thygesen LC, Ersbøll AK. When the entire population is the sample: strengths and limitations on register-based epidemiology. Eur J epidemiol (2013) DOI 10.1007/s10654-013-9873-0
104. Copay AG, Subach BR, Glassman SD, Polly DW Jr, Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. Spine J. 2007 Sep-Oct;7(5):541-6. doi: 10.1016/j.spinee.2007.01.008. Epub 2007 Apr 2. PMID: 17448732.
105. Copay A, Martin M, Subach B, Carreon LY, Glassman SD, Schuler TC, et. al.: Assessment of spine surgery outcomes: inconsistency of change amongst outcome measurements. The Spine Journal 2010; 10: pp. 291-296. Iss
106. Solberg T, Johnsen LG, Nygaard ØP, 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: pp. 196-201.
107. Ostelo RW, Deyo RA, Stratford P, Waddell G, Croft P, Von Kroff M, 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: pp. 90-94.
108. Glassman SD, Copay AG, Berven SH, Polly DW, Subach BR, Carreon LY.: Defining substantial clinical benefit following lumbar spine arthrodesis. J Bone Joint Surg Am 2008; 90: pp. 1839-1847
109. Austevoll IM, Gjestad R, Grotle M, Solberg T, Brox JI, Hermansen E, Rekeland F, Indrekvam K, Storheim K, Hellum C. 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 Jan 18;20(1):31. doi: 10.1186/s12891-018-2386-y. PMID: 30658613; PMCID: PMC6339296.
110. 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 Oct;16(10):1221-1230. doi: 10.1016/j.spinee.2016.06.010. Epub 2016 Jun 22. PMID: 27343730.
111. Terwee CB, Peipert JD, Chapman R, Lai JS, Terluin B, Cella D, Griffiths P, Mokkink LB. Minimal important change (MIC): a conceptual clarification and systematic review of MIC estimates of PROMIS measures. Qual Life Res. 2021 Oct;30(10):2729-2754. doi: 10.1007/s11136-021-02925-y. Epub 2021 Jul 10. PMID: 34247326; PMCID: PMC8481206.
112. Werner DAT, Grotle M, Gulati S, Austevoll IM, Lønne G, Nygaard ØP, Solberg TK. Criteria for failure and worsening after surgery for lumbar disc herniation: a multicenter observational study based on data from the Norwegian Registry for Spine Surgery. Eur Spine J. 2017 Oct;26(10):2650-2659. doi: 10.1007/s00586-017-5185-5. Epub 2017 Jun 14. PMID: 28616747.
113. 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
114. 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

115. Kaye ID, Butler JS, Morrissey PB, Sebastian AS, Wagner SC, Vaccaro AR. Spine Registries: Where Do We Stand? Clin Spine Surg. 2018 Nov;31(9):389-394. doi: 10.1097/BSD.0000000000000589. PMID: 29064835.
116. Solberg TK, Sørlie A, Sjaavik K, Nygaard ØP, Ingebrigtsen T. Would loss to follow-up bias the outcome evaluation of patients operated for degenerative disorders of the lumbar spine? Acta Orthop. 2011 Feb;82(1):56-63. doi: 10.3109/17453674.2010.548024. Epub 2010 Dec 29. PMID: 21189113; PMCID: PMC3229998.
117. Højmark K, Støttrup C, Carreon L, Andersen MO. Patient-reported outcome measures unbiased by loss of follow-up. Single-center study based on DaneSpine, the Danish spine surgery registry. Eur Spine J. 2016 Jan;25(1):282-286. doi: 10.1007/s00586-015-4127-3. Epub 2015 Jul 25. PMID: 26208938
118. Endler P, Ekman P, Hellström F, Möller H, Gerdhem P. Minor effect of loss to follow-up on outcome interpretation in the Swedish spine register. European Spine Journal : Official Publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society. 2020 Feb;29(2):213-220. DOI: 10.1007/s00586-019-06181-0. PMID: 31781864
119. Meyer B, Shiban E, Albers LE, Krieg SM. Completeness and accuracy of data in spine registries: an independent audit-based study. Eur Spine J. 2020 Jun;29(6):1453-1461. doi: 10.1007/s00586-020-06342-6. Epub 2020 Mar 4. PMID: 32130526.
120. Solberg T, Olsen RR, Berglund ML (2018) Annual report for NORspine 2018. 7.Oct 2019. (Accsessed 11th November 2021).
121. McHugh ML (2012) Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 22(3):276–282 (PMID: 23092060; PMCID: PMC3900052
122. Koo TK, Li MY (2016) A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research [published correction appears in J Chiropr Med. 2017 Dec;16(4):346]. J Chiropr Med 15(2):155–163. Doi:https://doi.org/10.1016/j.jcm.2016.02
123. Tape TG, University of Nebraska Medical Cenrter. Interpreting diagnostic tests. http://gim.unmc.edu/dxtests/ROC3.htm . (accessed 17.th August 2020)
124. Alshameeri ZAF, Ahmed E-N, Jasani V. Clinical Outcome of Spine Surgery Complicated by Accidental Dural Tears: Meta-Analysis of the Literature. *Global Spine Journal*. 2021;11(3):400-409. doi:10.1177/2192568220914876
125. van Hoof ML, van Loon J, van Limbeek J, de Kleuver M. The Nijmegen descision tool for chronic low back pain. Development of a clinical decision tool for secondary or tertiary spine care specialists. Plos one 2014.
126. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. Ann Intern Med 2007; 147(8):573-577. PMID: 17938396
127. Bhuyvan,SS et al. Transforming Healthcare Cybersecurity from reactive to proactive: current status and future recommendations. J Med Syst 2020, 44(5): 98
128. Brodke DS, Goz V, Lawrence BD, Spiker WR, Neese A, Hung M. Oswestry Disability Index: a psychometric analysis with 1,610 patients. Spine J. 2017 Mar;17(3):321-327. doi: 10.1016/j.spinee.2016.09.020. Epub 2016 Sep 29. PMID: 27693732.
129. Fujimori T, Ikegami D, Sugiura T, Sakaura H. Responsiveness of the Zurich Claudication Questionnaire, the Oswestry Disability Index, the Japanese Orthopaedic Association Back Pain Evaluation Questionnaire, the 8-Item Short Form Health Survey, and the Euroqol 5 dimensions 5 level in the assessment of patients with lumbar spinal stenosis. Eur Spine J. 2022 Jun;31(6):1399-1412. doi: 10.1007/s00586-022-07236-5. Epub 2022 May 7. PMID: 35524825.
130. Only 3% of Companies' Data Meets Basic Quality StandardsTECHNOLOGY & OPERATIONS DIGITAL ARTICLE Thomas C. RedmanDavid SammonTadhg NagleThomas C RedmanData is in far worse shape than most managers realize.SEPTEMBER 11, 2017

131. Öhrn A, Elfström J, Liedgren C, Rutberg H (2011) Reporting of sentinel events in Swedish hospitals: a comparison of severe adverse events reported by patients and providers. Jt Comm J Qual Patient Saf 37(11):495–501. https://doi.org/10.1016/S1553-7250(11)37063.ISSN1553-7250)
132. Kazberouk ABS, Martin BI, Stevens JP, McGuire KJ (2015) Validation of an administrative coding algorithm for classifying surgical indication and operative features of spine surgery. Spine 40(2):114–120.
133. Kristoffersen MH, Dybvik E, Steihaug OM et al (2019) Validation of orthopaedic surgeons' assessment of cognitive function in patients with acute hip fracture. BMC Musculoskelet Disord. https://doi.org/10.1186/s12891-019-2633-x
134. Horner NS, Grønhaug-Larsen KM, Svanteson E, Samuelsson K, Aveni OR et al (2020) Timing of hip hemiarthroplasty and the influence on prosthetic joint infection. PLOS ONE 15(3):e0229947
135. Varmdal T, Bakken IJ, Janszky I, Wethal T, Ellekjær H, Rohweder G, Fjærtoft H, Ebbing M, Bønaa KH (2016) Comparison of the validity of stroke diagnoses in a medical quality register and an administrative health register. Scand J Public Health 44(2):143–149. https://doi.org/10.1177/1403494815621641
136. Bjørnstad-Tuveng Identifying autopsy-reports in unstructured, digitized patient records: A pilot-study from the Norwegian Health Archives Registry. hAIst Conference – Health-related Artificial Intelligence in Stavanger (accessed 14.th Nov.2022)
137. Löfgren L, Eloranta S, Krawiec K, Asterkvist A, Lönnqvist C, Sandelin K, Steering group of the National Register for Breast Cancer (2019) Validation of data quality in the Swedish national register for breast cancer. BMC Public Health 19(1):495
138. Landberg A, Bruce D, Lindblad P, Ljungberg B, Lundstam S, Thorstenson A, Sundqvist P (2021) Validation of data quality in the National Swedish kidney cancer register. Scand J Urol 55(2):142–148. https://doi.org/10.1080/21681805.2021.1885485
139. Kristman V, Manno M, Côté P. Loss to follow-up in cohort studies: how much is too much? Eur J Epidemiol. 2004;19(8):751-60. doi: 10.1023/b:ejep.0000036568.02655.f8. PMID: 15469032.
140. Parai C, Hägg O, Willers C, Lind B, Brisby H. Characteristics and predicted outcome of patients lost to follow-up after degenerative lumbar spine surgery. Eur Spine J. 2020 Dec;29(12):3063-3073. doi: 10.1007/s00586-020-06528-y. Epub 2020 Jul 14. PMID: 32666273.
141. Little RJ RD. Statistical analysis with missing data. John Wiley & Sons. 2019
142. Twisk J, de Vente W. Attrition in longitudinal studies. How to deal with missing data. J Clin Epidemiol. 2002;55(4):329–37.
143. Wang K, Eftang CN, Jakobsen RB, Årøen A. Review of response rates over time in registry-based studies using patient-reported outcome measures. BMJ open 2020;**10:**e030808. doi: 10.1136/bmjopen-2019-030808
144. Rolstad S. PhLic JA, Ryden A. Response burden and questionnaire length: Is shorter better? A review and meta-analysis. Value in health (2011(, 14;1101-1108.
145. DH8426_NZJR_2020_Report_v5_30Sep.pdf (nzoa.org.nz) (accessed 11.th Jan 2023)
146. Parai C, Hägg O, Lind B, Brisby H.: Follow-up of degenerative lumbar spine surgery-PROMs stabilize after 1 year: an equivalence study based on Swespine data. Eur Spine J 2019; 28: pp. 2187-2197.
147. Weinstein JN, Lurie JD, Tosteson TD, Zhao W, Blood EA, Tosteson ANA, 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: pp. 1295-1304.
148. Weinstein JN, Tosteson TD, Lurie JD, Tosteson A, Blood E, Herkowitz H, et. al.: Surgical versus nonoperative treatment for lumbar spinal stenosis four-year results of the Spine Patient Outcomes Research Trial. Spine (Phila Pa 1976) 2010; 35: pp. 1329-1338
149. Ware JE.SF-36 health survey: manual and interpretation guide Boston The Health Institue: 1993, 6:1-6.
150. Guyatt GH, Jaeschke RRJ. Reassessing quality of life instruments in theevaluation of new drugs. Pharmacoeconomics 1997;12:616-26

151. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. Med Care. 1989 Mar;27(3 Suppl):S178-89. doi: 10.1097/00005650-198903001-00015. PMID: 2646488.)
152. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol. 1991 Feb;59(1):12-9. doi: 10.1037//0022-006x.59.1.12. PMID: 2002127.
153. Wyrwich KW, Nienaber NA, Tierney WM, Wolinsky FD. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. Med Care. 1999 May;37(5):469-78. doi: 10.1097/00005650-199905000-00006. PMID: 10335749wyrwich
154. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life , Journal of clinical epidemilogoly 2003;56:395-407
155. Habibzadeh F, Habibzadeh P, Yadollahie M. On determining the most appropriate test cut-off value: the case of tests with continuous results, Biochemia Medica (2016, 26(3):297-307.
156. Grøvle L, Haugen AJ, Hasvik E, Natvig B, Brox JI, Grotle M, et. al.: Patients' ratings of global perceived change during 2 years were strongly influenced by the current health status. J Clin Epidemiol 2014; 67: pp. 508-515
157. Gatchel RJ, Mayer TG.: Testing minimal clinically important difference: additional comments and scientific reality testing. Spine J 2010; 10: pp. 330-332
158. Arpey N, Barett J, Gerlach EB, Alvandi B, Divi S, Hsu WK, Patel AA. Does recall bias exist in lumbar spine surgery patients as reported through PROMIS quesdtionnaires?, Spine Journal, The, 2022-09-01, Volume 22, Issue 9, Pages S22-S23
159. Schwartz N, Sudman S. Autobiographical memory and the validity of retrospective reports. New York Springer verlag 1994
160. Wyrwich KW et al. The reliability of retrospective change assessments Quality of Life Research (2002);11: 636-636.
161. Revicki D, Hays RD, Cella D, Sloan J.: Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol 2008; 61: pp. 102-109
162. Stratford PW, Binkley JM, Riddle DL, Guyatt G. Sensitivity to change of the Roland Morris BACK pain Questionnaire: part 1.Phys Ther 1998;78:1186-96. https://doi.org/10.1093/ptj/78.11.1186
163. Terluin B, Eekhout I, Terwee CB, de Vet HC. Minimal important change (MIC) based on a predictive modeling approach was more precise than MIC based on ROC analysis. J Clin Epidemiol. 2015 Dec;68(12):1388-96. doi: 10.1016/j.jclinepi.2015.03.015. Epub 2015 Mar 28. PMID: 25913670
164. One size does not fit all when it comes to the minimal clinically important change score, Mannion A, Eurospine Oct 2022
165. Carragee EJ, Cheng I. Minimun acceptable outcomes after lumbar spinal fusion. The Spine J (2010);10(4):313-320.
166. Taso, M., Sommernes, J.H., Bjorland, S. *et al.* What is success of treatment? Expected outcome scores in cervical radiculopathy patients were much higher than the previously reported cut-off values for success. *Eur Spine J* **31**, 2761–2768 (2022). https://doi.org/10.1007/s00586-022-07234-7
167. Arpey N, Barett J, Gerlach EB, Divi S, Hsu WK, Patel AA. Quantitative prospective patient-repoprted expectations challenge commonly used thersholds fro minimal clinically importanmd defference en elective spine surgery, The Spine Journal (2022);22(9): S67-S67.
168. Witiw, C.D., Mansouri, A., Mathieu, F. *et al.* Exploring the expectation-actuality discrepancy: a systematic review of the impact of preoperative expectations on satisfaction and patient reported outcomes in spinal surgery. *Neurosurg Rev* **41**, 19–30 (2018). https://doi.org/10.1007/s10143-016-0720-0
169. Jones AA, Stambough JL, Balderston RA, Rothman RH, Booth RE Jr. (1989) Long-term results of lumbar spine surgery complicated by unintended incidental durotomy. Spine (Phila Pa 1976) 14(4):443-6. doi: 10.1097/00007632-198904000-00021. PMID: 2718049

170. Strömqvist F, Jönsson B, Strömqvist B; Swedish Society of Spinal Surgeons. (2010) Dural lesions in lumbar disc herniation surgery: incidence, risk factors, and outcome. *Eur Spine J*. 19(3):439-442. doi:10.1007/s00586-009-1236-x
171. Katz NP, Paillard FC, Ekman E. (2015) Determining the clinical importance of treatment benefits for interventions for painful orthopedic conditions. J Orthop Surg. 10:24. doi: 10.1186/s13018-014-0144-x. - DOI - PMC - PubMed
172. Alshameeri ZAF, Jasani V. (2021)Risk Factors for Accidental Dural Tears in Spinal Surgery. Int J Spine Surg. 15(3):536-548. doi: 10.14444/8082. Epub 2021 May 13. PMID: 33986000; PMCID: PMC8176817
173. Meghan E. Murphy, Panagiotis Kerezoudis, Mohammed Ali Alvi, Brandon A. McCutcheon, Patrick R. Maloney, Lorenzo Rinaldo, Daniel Shepherd, Daniel S. Ubl, William E. Krauss, Elizabeth B. Habermann & Mohamad Bydon. (2017) Risk factors for dural tears: a study of elective spine surgery, Neurological Research. 39:2, 97-06, DOI: 10.1080/01616412.2016.1261236
174. Yap ZL, Summers SJ, Grant AR, Moseley GL, Karran EL.: The role of the social determinants of health in outcomes of surgery for low back pain: a systematic review and narrative synthesis. The Spine journal 2022; 22: pp. 793-809
175. Claus CF, Tong D, Lytle E, Bahoura M, Garmo L, Li C, Park P, Carr DA, Easton R, Abdulhak M, Chang V, Houseman C, Bono P, Richards B, Soo TM.: Age as a Predictor for Complications and Patient-reported Outcomes in Multilevel Transforaminal Lumbar Interbody Fusions: Analyses From the Michigan Spine Surgery Improvement Collaborative (MSSIC). Spine (Phila Pa 1976) 2021; 46: pp. 356-365. 15PMID: 33620179
176. Giannadakis C, Solheim O, Jakola AS, Nordseth T, Gulati AM, Nerland US, Nygaard ØP, Solberg TK, Gulati S.: Surgery for Lumbar Spinal Stenosis in Individuals Aged 80 and Older: A Multicenter Observational Study. J Am Geriatr Soc 2016; 64: pp. 2011-2018. Epub 2016 Sep 9. PMID: 27611928
177. Fritzell P, Mesterton J, Hagg O.: Prediction of outcome after spinal surgery-using The Dialogue Support based on the Swedish national register. Eur Spine J 2021; pp. 27. Epub ahead of print. PMID: 34837113
178. Onyekwelu I, Glassman SD, Asher A, Shaffrey CI, Mummaneni PV, Carreon LY.: Impact of obsiety on complications and outcomes: a comparison of fusion and nonfusion lumbar spine surgery. Journal of neurosurgery: Spine 2017; 26: pp. 158-162
179. Radcliff KE, Rihn J, Hilibrand A, DiIorio T, Tosteson T, Lurie JD, Zhao W, Vaccaro AR, Albert TJ, Weinstein JN. Does the duration of symptoms in patients with spinal stenosis and degenerative spondylolisthesis affect outcomes?: analysis of the Spine Outcomes Research Trial. Spine 2011; 36: pp. 2197-2210
180. ChaoT, Pham K, Steward O, Gupta R: Chronic nerve compression injury induces a phenotypic switch of neurons within the dorsal root ganglia. The journal of comparative neurology 2008; 506: pp. 180-193
181. Weber C, Giannadakis C, RV, Jakola AS., Nerland U, Nygaard ØP., Solberg TK., Gulati S, Solheim O. Is There an Association Between Radiological Severity of Lumbar Spinal Stenosis and Disability, Pain, or Surgical Outcome?. SPINE 2016; 41: pp. E78-E83
182. Mannion, A.F., Fekete, T.F., Pacifico, D. *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* **26**, 2552–2564 (2017). https://doi.org/10.1007/s00586-017-5280-7
183. Aaen, J., Banitalebi, H., Austevoll, I.M. *et al.* The association between preoperative MRI findings and clinical improvement in patients included in the NORDSTEN spinal stenosis trial. *Eur Spine J* **31**, 2777–2785 (2022). https://doi.org/10.1007/s00586-022-07317-5
184. hostmoteboka-2022.pdf (legeforeningen.no) (accessed 12. th Jan 2023)
185. Tversky A, Kahneman D. Advances in prospect theory: cumulative representation of incertainty, Journal of Risk and Uncertainty, Vol 5, 1992, pp 297-323.
186. Shipe ME, Deppen SA, Farjah F, Grogan EL. Developing prediction models for clinical use using logistic regression: an overview. J Thorac Dis. 2019 Mar;11(Suppl 4):S574-S584. doi: 10.21037/jtd.2019.01.25. PMID: 31032076; PMCID: PMC6465431.

187. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol. 1996 Dec;49(12):1373-9. doi: 10.1016/s0895-4356(96)00236-3. PMID: 8970487.
188. Dialogue support for spine surgery – based on Swespine, the Swedish Spine Register - Molnify (accessed 15.th Feb 2023)
189. PrognosticTool: degenerative spinal disorders (kws.ch) (accessed 15.th Feb 2023)

11 Appendix

11.1 Appendix 1

Norwegian version of Oswestry Disability Index

OSWESTRY LISTE FOR FUNKSJONSBEGRENSNINGER, versjon 2.0

*Vennligst les: Dette spørreskjemaet er utformet for å gi behandleren opplysninger om hvordan ryggsmertene dine har påvirket din evne til å klare deg i dagliglivet. Vennligst svar på hvert avsnitt, og marker bare **det ene feltet** i hvert avsnitt som gjelder for deg. Vi forstår at du kanskje synes at to av utsagnene i hvert avsnitt gjelder deg, men vennligst **marker bare feltet som best beskriver ditt nåværende problem**.*

Del 1 – Smerteintensitet

1. Jeg har ingen smerter for øyeblikket
2. Smertene er veldig svake for øyeblikket
3. Smertene er moderate for øyeblikket
4. Smertene er temmelig sterke for øyeblikket
5. Smertene er veldig sterke for øyeblikket
6. Smertene er de verste jeg kan tenke meg for øyeblikket

Del 2 – Personlig stell (vaske seg, kle på seg, osv.)

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

Del 3 – Løfte

1. Jeg kan løfte tunge ting uten å få mer smerter
2. Jeg kan løfte tunge ting, men får mer smerter
3. Smertene hindrer meg i å løfte tunge ting opp fra gulvet, men jeg greier det hvis det som skal løftes er gunstig plassert, f.eks. på et bord
4. Smertene hindrer meg i å løfte tunge ting, men jeg kan klare lette eller middels tunge ting, hvis det er gunstig plassert
5. Jeg kan bare løfte noe som er veldig lett
6. Jeg kan ikke løfte eller bære noe i det hele tatt

Del 4 – Gå

1. Smerter hindrer meg ikke i å gå i det hele tatt
2. Smerter hindrer meg i å gå mer enn 1 ½ km
3. Smerter hindrer meg i å gå mer enn ¾ km
4. Smerter hindrer meg i å gå mer enn 100 m
5. Jeg kan bare gå med stokk eller krykker
6. Jeg ligger for det meste i sengen og jeg må krabbe til toalettet

Del 5 – Sitte

1. Jeg kan sitte så lenge jeg vil i en hvilken som helst stol
2. Jeg kan sitte så lenge jeg vil i min favorittstol
3. Smerter hindrer meg i å sitte i mer enn en time
4. Smerter hindrer meg i å sitte i mer enn en halv time
5. Smerter hindrer meg i å sitte i mer enn ti minutter
6. Smerter hindrer meg i å sitte i det hele tatt

Forts.OSWESTRY LISTE FOR FUNKSJONSBEGRENSNINGER

Del 6 – Stå

1. Jeg kan stå så lenge jeg vil uten å få mer smerter
2. Jeg kan stå så lenge jeg vil, men får mer smerter
3. Smerter hindrer meg i å stå i mer enn en time
4. Smerter hindrer meg i å stå i mer enn en halv time
5. Smerter hindrer meg i å stå i mer enn ti minutter
6. Smerter hindrer meg i å stå i det hele tatt

Del 7 – Sove

1. Søvnen min forstyrres aldri av smerter
2. Søvnen min forstyrres av og til av smerter
3. På grunn av smerter får jeg mindre enn seks timers søvn
4. På grunn av smerter får jeg mindre enn fire timers søvn
5. På grunn av smerter får jeg mindre enn to timers søvn
6. Smerter hindrer all søvn

Del 8 – Seksualliv

1. Seksuallivet mitt er normalt og forårsaker ikke mer smerter
2. Seksuallivet mitt er normalt, men forårsaker noe mer smerter
3. Seksuallivet mitt er normalt, men svært smertefullt
4. Seksuallivet mitt er svært begrenset av smerter
5. Seksuallivet mitt er nesten borte på grunn av smerter
6. Smerter forhindrer alt seksualliv

Del 9 – Sosialt liv

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

Del 10 – Reising

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

Skåring; kode om spørsmålene til 0-5 (1=0….6=5). Summer hvert spørsmål, deles med antall besvarte spørsmål, multipliseres med 0,2 og 100 for å få en prosentskår.

The Modified Oswestry Disability Index (Baker et al 1990)
Oversatt av Margreth Grotle og Nina K:Vøllestad 2001,
Seksjon for Helsefag, Universitetet I Oslo

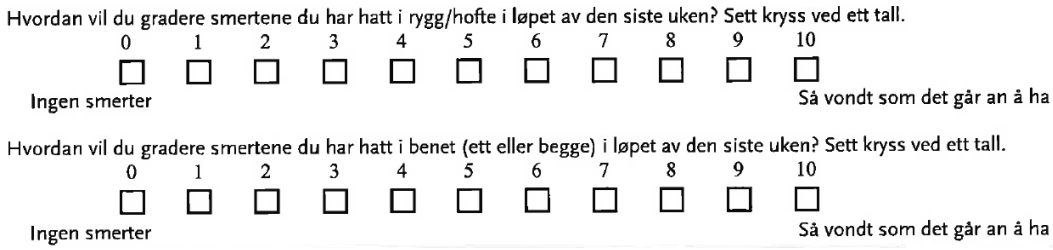
11.2 Appendix 2

Norwegian version of EQ-5D

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

11.3 Appendix 3

Norwegian version of numeric rating scale (NRS) back pain and leg pain



11.4 Appendix 4

Norwegian version of the Global Perceived Effect (GPE) scale

11.5 Appendix 5

NORspine form, patient-completed, before 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

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, familie- 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

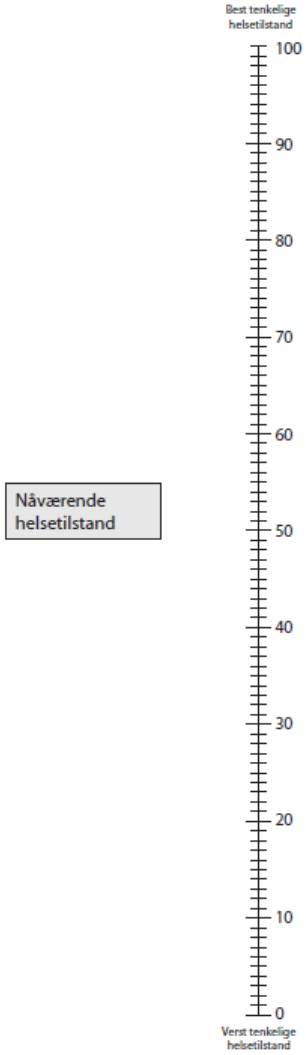
Hvis du har svart ja: Hvor ofte bruker du smertestillende medisiner? (Sett kun ett kryss)

- [ ] 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.



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

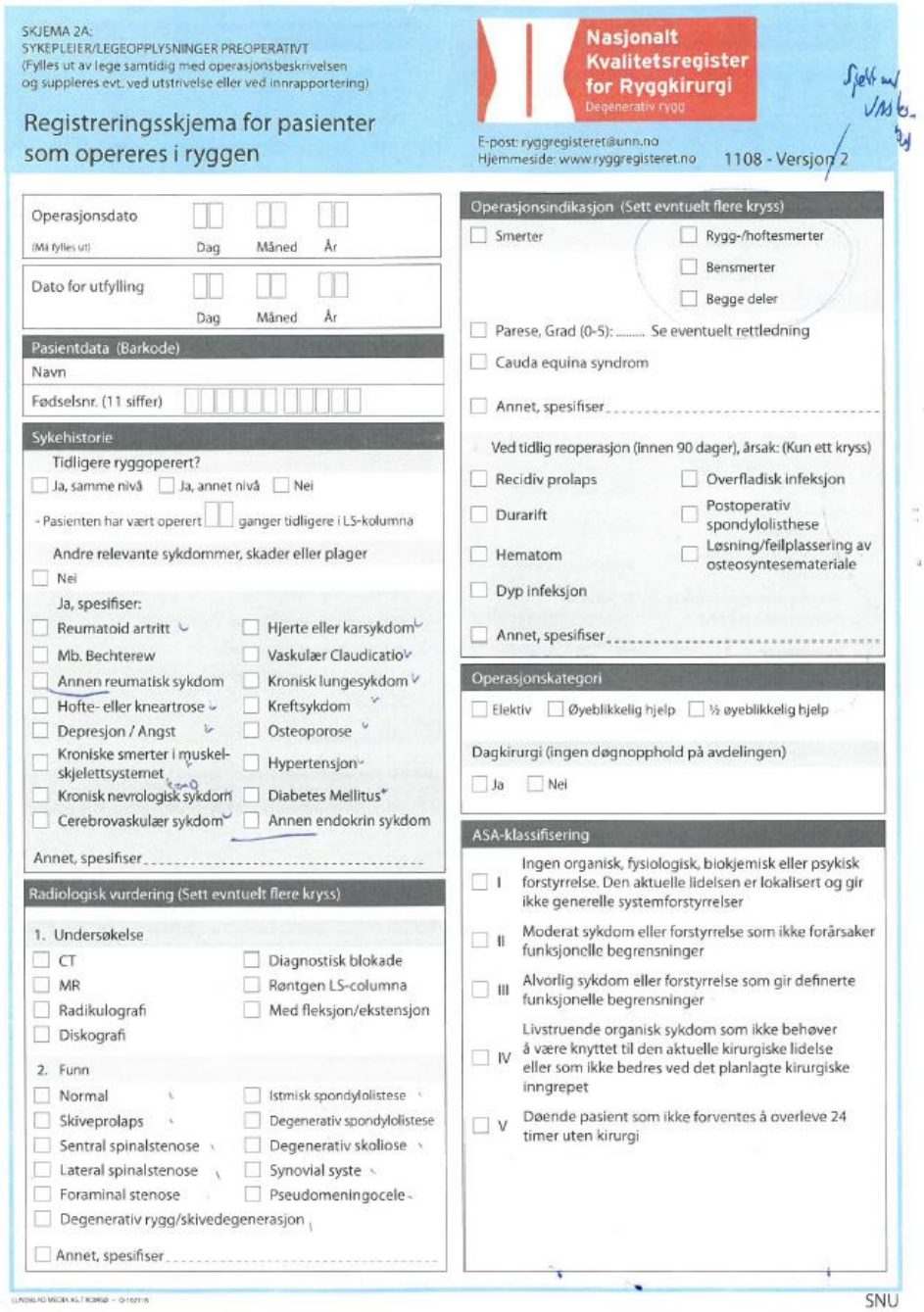
Har du søkt om erstatning fra forsikringsselskap eller folketrygden (eventuelt yrkesskadeerstatning)?

(Sett kun ett kryss)

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

11.6 Appendix 6

NORspine form, surgeon-completed, after surgery.



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

Nasjonalt Kvalitetsregister for Ryggkirurgi
Degenerativ rygg

Registreringsskjema for pasienter som opereres i ryggen

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

Annet, spesifiser

Radiologisk vurdering (Sett evntuelt flere kryss)

1. Undersøkelse

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

2. Funn

☐ Normal	☐ Istmisk spondylolistese
☐ Skiveprolaps	☐ Degenerativ spondylolistese
☐ Sentral spinalstenose	☐ Degenerativ skoliose
☐ Lateral spinalstenose	☐ Synovial syste
☐ Foraminal stenose	☐ Pseudomeningocele

☐ Degenerativ rygg/skivedegenerasjon
☐ 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	☐ Overfladisk infeksjon
☐ Durarift	☐ Postoperativ spondylolisthese
☐ Hematom	☐ Løsning/feilplassering av osteosyntesemateriale
☐ Dyp infeksjon	

☐ 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

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:

- ☐ 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

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

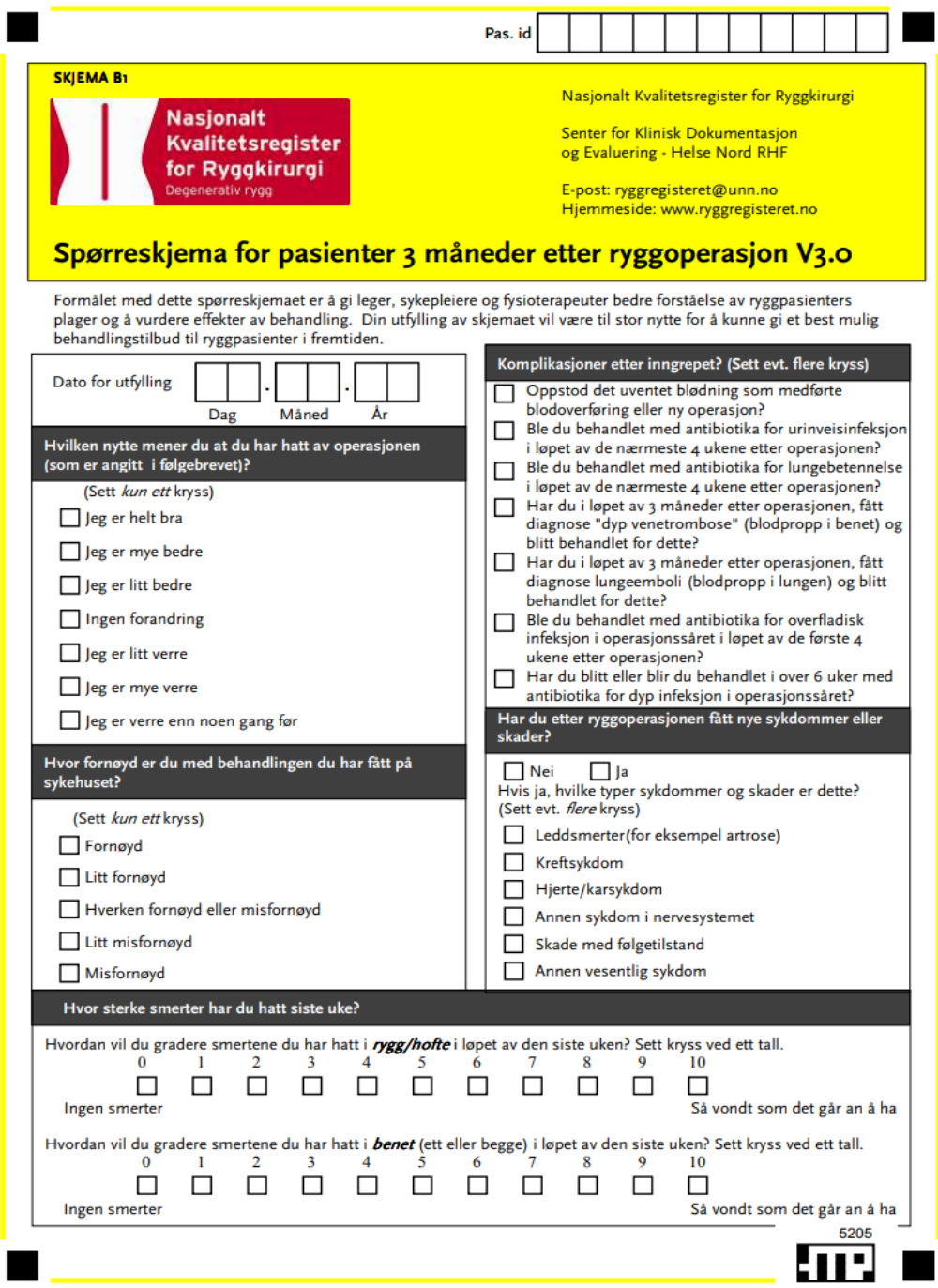
☐☐☐ (dager)

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

- ☐ Cardiogen årsak
- ☐ Lungeemboli
- ☐ Pneumoni
- ☐ Annen infeksjon
- ☐ Anafylaksi
- ☐ Cerebrovaskulær årsak
- ☐ Blødning
- ☐ Annet, spesifiser _____

11.7 Appendix 7

NORspine 3 months follow-up form, patient-completed.



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 V3.0

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.

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

Hvilken nytte mener du at du har hatt av operasjonen (som er angitt i følgebrevet)?

(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

Komplikasjoner etter inngrepet? (Sett evt. flere kryss)

- ☐ Oppstod det uventet blødning som medførte blodoverføring eller ny operasjon?
- ☐ Ble du behandlet med antibiotika for urinveisinfeksjon i løpet av de nærmeste 4 ukene etter operasjonen?
- ☐ Ble du behandlet med antibiotika for lungebetennelse i løpet av de nærmeste 4 ukene etter operasjonen?
- ☐ Har du i løpet av 3 måneder etter operasjonen, fått diagnose "dyp venetrombose" (blodpropp i benet) og blitt behandlet for dette?
- ☐ Har du i løpet av 3 måneder etter operasjonen, fått diagnose lungeemboli (blodpropp i lungen) og blitt behandlet for dette?
- ☐ Ble du behandlet med antibiotika for 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 etter ryggoperasjonen fått nye sykdommer eller skader?

☐ Nei ☐ Ja

Hvis ja, hvilke typer sykdommer og skader er dette?
(Sett evt. *flere* kryss)

- ☐ Leddsmerter(for eksempel artrose)
- ☐ Kreftsykdom
- ☐ Hjerte/karsykdom
- ☐ Annen sykdom i nervesystemet
- ☐ Skade med følgetilstand
- ☐ Annen vesentlig sykdom

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

5205

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

5205

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 har middels store problemer med å gå omkring
- ☐ Jeg har store problemer med å gå omkring
- ☐ Jeg er ute av stand til å gå omkring

2. Personlig stell

- ☐ Jeg har ingen problemer med å vaske meg eller kle meg
- ☐ Jeg har litt problemer med å vaske meg eller kle meg
- ☐ Jeg har middels store problemer med å vaske meg eller kle meg
- ☐ Jeg har store 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, familie- eller fritidsaktiviteter)

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

4. Smerte og ubehag

- ☐ Jeg har verken smerter eller ubehag
- ☐ Jeg har litt smerter eller ubehag
- ☐ Jeg har middels sterke smerter eller ubehag
- ☐ Jeg har sterke smerter eller ubehag
- ☐ Jeg har svært sterke smerter eller ubehag

5. Angst og depresjon

- ☐ Jeg er verken engstelig eller deprimert
- ☐ Jeg er litt engstelig eller deprimert
- ☐ Jeg er middels engstelig eller deprimert
- ☐ Jeg er svært engstelig eller deprimert
- ☐ Jeg er ekstremt 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)

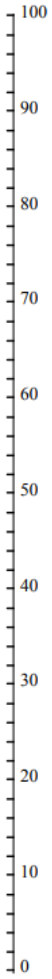
- ☐ Sjeldnere enn hver måned
- ☐ Hver måned
- ☐ Hver uke
- ☐ Daglig
- ☐ Flere ganger daglig

5205

Helsetilstand

Vi vil gjerne vite hvor god eller dårlig helsen din er I DAG.
- Denne skalaen er nummerert fra 0 til 100
- 100 betyr den beste helsen du kan tenke deg.
- 0 betyr den dårligste helsen du kan tenke deg.
- Sett et X på skalane for å angi hvordan helsen din er I DAG.
- Skriv deretter tallet du merket av på skalaen inn i boksen nedenfor.

Den beste helsen du kan tenke deg



Helsen din i dag = ☐☐☐

Den dårligste helsen du kan tenke deg

Pas. id ☐☐☐☐☐☐☐☐☐☐☐

Arbeidsstatus

☐ Fulltidsjobb
☐ Deltidsjobb
☐ Student/skoleelev
☐ Alderspensjonist
☐ Arbeidsledig
☐ Sykemeldt
☐ Delvis sykemeldt
☐☐ % sykemeldt
☐ Arbeidsavklaringspenger
☐ Uføretrygdet
evt. ☐☐ % uføretrygdet

Føler du at din arbeidsgiver ønsker deg tilbake i jobb?

☐ Ja ☐ Nei ☐ Vet ikke

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

Har du vært operert i ryggen etter ryggoperasjonen? (Dato angitt på forsiden)

☐ Nei ☐ Ja

Hvis ja, skriv antall operasjoner: ☐☐

Hvis ja, ble du operert i samme område («etasje», nivå) i ryggen? (Sett *kun ett* kryss)

☐ Ja, samme område
☐ Nei, i annet område
☐ I samme og annet område
☐ Vet ikke

5205

11.8 Appendix 8

NORspine 12 months follow-up form, patient-completed.

Pas. id

SKJEMA B2

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 12 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

7348

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

7348

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

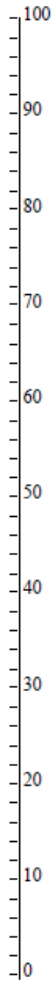
7348

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



Nåværende helsetilstand

Verst tenkelige 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

7348

12 Original papers

12.1 Paper 1

European Spine Journal (2022) 31:801–811
https://doi.org/10.1007/s00586-021-07093-8

ORIGINAL ARTICLE

Accuracy and agreement of national spine register data for 474 patients compared to corresponding electronic patient records

Ole Kristian Alhaug[1,2,3] · Simran Kaur[4] · Filip Dolatowski[4] · Milada Cvancarova Småstuen[5] · Tore K. Solberg[6,7] · Greger Lønne[1,3]

Received: 6 July 2021 / Revised: 12 November 2021 / Accepted: 18 December 2021 / Published online: 6 January 2022

Abstract

Purpose Data quality is essential for all types of research, including health registers. However, data quality is rarely reported. We aimed to assess the accuracy of data in a national spine register (NORspine) and its agreement with corresponding data in electronic patient records (EPR).

Methods We compared data in NORspine registry against data in (EPR) for 474 patients operated for spinal stenosis in 2015 and 2016 at four public hospitals, using EPR as the gold standard. We assessed accuracy using the proportion correctly classified (PCC) and sensitivity. Agreement was quantified using Kappa statistics or interaclass correlation coefficient (ICC).

Results The mean age (SD) was 66 (11) years, and 54% were females. Compared to EPR, surgeon-reported perioperative complications displayed weak agreement (kappa (95% CI) = 0.51 (0.33–0.69)), PCC of 96%, and a sensitivity (95% CI) of 40% (23–58%). ASA classification had a moderate agreement (kappa (95%CI) = 0.73 (0.66–0.80)). Comorbidities were underreported in NORspine. Perioperative details had strong to excellent agreements (kappa (95% CI) ranging from 0.76 (0.68–0.84) to 0.98 (0.95–1.00)), PCCs between 93% and 99% and sensitivities (95% CI) between 92% (0.84–1.00%) and 99% (0.98–1.00%). Patient-reported variables (height, weight, smoking) had excellent agreements (kappa (95% CI) between 0.93 (0.89–0.97) and 0.99 (0.98–0.99)).

Conclusion Compared to electronic patient records, NORspine displayed weak agreement for perioperative complications, moderate agreement for ASA classification, strong agreement for perioperative details, and excellent agreement for height, weight, and smoking. NORspine underreported perioperative complications and comorbidities when compared to EPRs. Patient-recorded data were more accurate and should be preferred when available.

Keywords Validation · Accuracy · Agreement · Registry · Lumbar spinal stenosis

✉ Ole Kristian Alhaug
olekralhaug@hotmail.com

[1] Innlandet Hospital Trust, Brumunddal, Norway

[2] Akershus University Hospital, Nordbyhagen, Norway

[3] Norwegian University of Science and Technology, Trondheim, Norway

[4] Division of Orthopaedic Surgery, Oslo University Hospital, Oslo, Norway

[5] OsloMet University, Oslo, Norway

[6] University Hospital of North Norway, Tromsø, Norway

[7] Department of Clinical Medicine, The Arctic University of Norway (UiT), Tromsø, Norway

Introduction

In clinical research, it is crucial to question how true and accurate data are; however, data validity and accuracy assessments are rarely published explicitly. National medical registries collect large-scale data during the dynamic workflow of daily clinical practice and have become essential sources of evidence-based medicine and health care policies. Register-based studies reflect everyday practice and have high external validity, and complement randomized control trials (RCTs) that assess smaller populations with lower external validity. Register data are collected and recorded by healthcare personnel, and not by dedicated research assistants. Therefore, it is essential to periodically assess the quality of register data reported by healthcare personnel and patients by validating it against other sources of data

[1–3]. Because systematic errors can lead to bias, register validations may impact the robustness of medical and political conclusions based on register data. The literature on the validity of medical register data is sparse. Some studies are reporting good validity of medical and cancer registries [4–6]. However, a recent validation study of a German spine registry (DWG) showed high inaccuracy [7] and the authors recommended against using these register data.

Our study aimed to assess the accuracy and agreement of NORspine data by comparing it to electronic patient records (EPR). Such information can aid in identifying pitfalls and conceptual problems related to data collection, not only relevant for other spine registers but also others, routinely recording clinical data.

Patients and methods

In this cross-sectional study, we reviewed electronic patient records (EPRs) of patients operated for lumbar spinal stenosis (LSS) who consented and responded to NORspine between January 1, 2015, and December 31, 2016**.** The authors were authorized to access data from four public hospitals within one health region (South-Eastern Norway Regional Health Authority) in Norway. To assess the representativity of our sample, we compared the study population to those treated at the remaining hospitals.

In Norway, all 39 hospitals (coverage = 100%) that offer surgery for degenerative spinal disorders are obliged to report data to NORspine. Seventy percent of all patients that undergo elective spine surgery in Norway are included in NORspine, and the proportion that responds one year after surgery is seventy-four percent [8].

A NORspine data set consists of a preoperative form completed by the patient at admission for surgery. This form covers items related to sociodemographic and lifestyle variables (e.g., smoking, height, and weight) and a standard battery of questionnaires assessing pain and disability (Table 5). Immediately after completing surgery, and optimally while still in the operating theater, the surgeon completes a standardized form and reports clinical and radiological diagnosis, relevant comorbidities, ASA classification—usually as graded by the anesthetist, and details about the surgery, e.g., previous surgery, surgical access, surgical methods, and level(s) operated. The surgeon also reports perioperative complications by a predefined list (Table 6).

Patients report the clinical outcome at 3 and 12 months after surgery as assessed by standard Patient-Reported Outcome Measures (PROMs).

Electronic patient records (EPRs) consist of non-structured text documents (free text) recorded by DIPS® software within predetermined headings. We reviewed the EPRs using a standard empty NORspine form, and the investigators (OKA and SK) had no access to the corresponding data previously recorded in the NORspine. The study group selected a set of NORspine variables that could be recaptured from EPRs. Furthermore, we reviewed EPR documents (e.g., admission and surgeon's notes) at the same time point as the time of surgery recorded in NORspine. We did not assess variables that were not registered routinely or consistently in EPRs, such as PROMs, symptom duration, marital status, education level, mother tongue, and working capability. The clinical follow-up at the treating centers was not standardized, and it was performed at different time points at the hospitals without structural recording in EPR. Hence, follow-up data (including reoperations) in NORspine were not evaluated against EPRs in this study.

The EPRs of 22 patients were independently reviewed by two raters (OA and SK) to estimate interobserver reliability.

We calculated concordance in terms of agreement when comparing the structured NORspine data with EPR data; we also calculated accuracy for dichotomous variables, using EPR as the gold standard. We chose to report both accuracy and agreement because the use of certain EPR variables as a reference could be questioned (e.g., smoking and comorbidity).

The NORspine form requires the surgeon to report *relevant* comorbidities from a list, such as cardiovascular disease, diabetes, and osteoarthritis. In the EPR, comorbidity is recorded irrespective of its relevance to the planned spinal surgery. Consequently, agreement and accuracy were not evaluated for comorbidities. We only compared frequencies of relevant comorbidities recorded in NORspine vs. the corresponding comorbidities recorded in EPRs. Furthermore, we assessed the agreement for ASA classification between the two data sources.

Statistical analyses

Baseline data were described using means (95%CI) (continuous data) and proportions (categorical data). Accuracy was assessed by proportion correctly classified (PCC) and sensitivity. Perioperative complications were categorized by eight categories (Table 6), and the accuracy of complication recording was assessed by class average accuracy (CAA) using the micro-averaged method. Agreement between NORspine and EPRs was assessed by Cohen's kappa (ƙ) or Fleiss weighted kappa (ƙ) for categorical variables (dichotomous and ordinal variables). (ASA classification was analyzed as an ordinal variable, ranging from 1 to 5, in the agreement analysis.) For continuous variables, we

calculated the intraclass correlation coefficient (ICC) using a two-way mixed model to assess absolute agreement [9]. We classified agreement (ƙ-value) as minimal (0.21–0.39), weak (0.40–0.59), moderate (0.60–0.79), strong (0.80–0.90), and almost perfect (> 0.90) [10]. Agreement according to ICC (values) was classified as poor (< 0.50), moderate (0.50–0.75), strong (0.75–0.90), and excellent (< 0.90) [11]. Finally, we calculated the prevalence of missing values for each variable. The results are presented as point estimates with 95% confidence intervals (CI).

We used SPSS, version 26 (IBM Corp., Armonk, N.Y., USA) and STATA version 16 (StataCorp. 2019. *Stata Statistical Software: Release 16*. College Station, TX: StataCorp LLC.)

Ethical considerations

The Norwegian Regional Committee for medical and health research ethics approved this study (reference no. 2017(2157)), as did the data protection officers at the four hospitals. All patients had provided informed consent, and the study was conducted in compliance with the Helsinki declaration.

Results

NORspine recorded 3,843 patients operated for LSS during 2015 and 2016. The investigators were authorized to access EPRs at four hospitals and reviewed the EPRs of 474 consecutive operated patients (12.3% of the NORspine population). Mean age (95%CI) was 66 (65.3–67.2) years, and 254 (54%) were females. The total of missing data were 0.9% in NORspine (completeness 99.1%) and 2.8% (completeness 97.2%) in EPRs (Table 7).

Patient characteristics, including data on the rest of the NORspine patients operated for lumbar spinal stenosis, are shown in Table 1. Our sample differed somewhat from the rest of the NORspine population at baseline. The included patients had more comorbidity, higher BMI, and higher disability (ODI) and pain scores (NRS = numeric rating scales) for leg and back pain. In addition, the study population had more smokers and had fewer perioperative complications than the total spinal stenosis population registered in NORspine (Table 1). For a sample of 22 patients, the interrater reliability for the two authors that reviewed EPR variables was almost perfect.

Table 1 Patient characteristics and perioperative details of 474 NORspine patients operated for spinal stenosis at four hospitals compared to 3369 from the remaining hospitals

	NORspine data—4 hospitals $n = 474$	NORspine data—remaining hospitals $n = 3369$
	Mean (95% CI) or *n* (%)	Mean (95% CI) or *n* (%)
Age	66.3 (65.3–67.2)	65.7 (65.4–66.1)
Female	254 (53.6%)	1,743 (51.7%)
Civil status—single	115 (24.3%)	899 (26.7%)
ASA* grade 1 and 2	353 (74.5%)	2,582 (77.1%)
Body mass index (BMI)	28.3 (27.9–28.7)	27.7 (27.6–27.9)
Smoking	111 (23.7%)	628 (18.8%)
Comorbidity, any	336 (75.5%)	2,208 (69.6%)
Previous spinal surgery, any	131 (27.8%)	866 (26.0%)
Preoperative ODI**	40.9 (15.9)	40.0 (15.4)
Preoperative NRS*** back pain	6.8 (2.1)	6.5 (2.2)
Preoperative NRS leg pain	7.0 (2.1)	6.5 (2.2)
Preoperative EQ-5D index****	0.332 (0.319)	0.369 (0.323)
Additional fusion	51 (10.8%)	355 (10.5%)
Perioperative complications	15 (3.2%)	184 (5.5%)
Missing data*****	58 (0.9%)	515 (1.1%)

*ASA: American society of anesthesiologists—classification of physical status (1–5)

**ODI: Oswestry Disability Index (0–100)

***NRS Numeric rating scale (0–10)

****EQ-5D: EuroQol five-dimensional index: quality of life (−0.59–1.00), values under 0.00 are considered "worse than dead"

***** Appendix Table 1 shows that five variables missed some data (BMI, smoking, comorbidity (any), previous spinal surgery (any), and surgical access.

Table 2 Accuracy and agreement of NORspine data for 474 spinal stenosis patients compared to their electronic patient records

Variable (missing, *n*)	Prevalence * *n* (%)	Proportion correctly classified, PCC (%)	Sensitivity (95%CI)	Kappa (95%CI)**
Perioperative complications (11)	30 (6.4%)	96	40% (23–58)	0.51 (0.33–0.69)
Previous spinal surgery (14)	120 (26.1%)	97	96% (92–99)	0.93 (0.89–0.97)
Additional fusion	51 (10.8%)	99	94% (88–100)	0.93 (0.88–0.99)
Access, posterior midline (26)	414 (92.4%)	93	93% (91–96)	0.19 (0.03–0.35)
Level L2-3 (14)	74 (16.1%)	99	99% (96–100)	0.98 (0.95–1.00)
Level L3-4 (13)	193 (41.9%)	99	98% (96–100)	0.97 (0.95–0.99)
Level L4-5 (13)	312 (67.7%)	98	99% (98–100)	0.95 (0.92–0.98)
Level L5-S1 (13)	48 (10.4%)	99	92% (84–100)	0.92 (0.86–0.98)
Smoking (43)***	112 (26.0%)	97	92% (87–97)	0.93 (0.89–0.97)

*prevalence according to EPR

**Cohens Kappa

***Smoking was registered by patients on the preoperative form; the remaining variables were registered by the surgeon on the postoperative form

Perioperative complications were recorded for 15 (3.2%) patients in NORspine, and 30 (6.5%) patients in the EPRs. The agreement between NORspine and EPR was weak (ƙ (95%CI) = 0.51 (0.33–0.69)). The class average accuracy for all perioperative complications was 99.4% (eight different categories combined), and for dural tears isolated, 97.0% were classified correctly (PCC). The sensitivity for recording a complication (95%CI) was 40% (23–58%) (Table 2).

As shown in Table 3, ASA classification (1–5) showed moderate agreement (ƙ (95%CI) = 0.73 (0.66–0.80)). Table 4 shows the differences in the prevalence of comorbidities. NORspine underreported comorbidities compared to EPRs.

As shown in Table 2, previous surgery (yes or no) had an almost perfect agreement (ƙ (95%CI) = 0.93 (0.89–0.97)), a proportion classified correctly of 97.2%, and a sensitivity of 95.8%. The number of previous surgeries showed moderate agreement (ƙ (95%CI) = 0.62 (0.48–0.75)), as shown in Table 3.

Perioperative details (method of decompression, fusion, surgical access, spinal level operated) recorded by the surgeon showed moderate to excellent agreement between NORspine and EPR (ƙ = 0.76 to 0.98), and high proportions were classified correctly (93–99%). The sensitivity for the recording of perioperative details was high (92–99%).

Table 3 Agreement for NORspine data for 474 spinal stenosis patients compared to their electronic patient records, ordinal or continuous variables

Data source	Variable	Agreement* (95%CI)
Surgeon, postoperative form	ASA classification**	0.73 (0.66–0.80)
	Number of previous surgeries	0.62 (0.48–0.75)***
	Number of levels operated	0.91 (0.84–0.99)
	Type of surgery****	0.90 (0.82–0.98)
	Method of decompression*****	0.76 (0.68–0.84)
Patient, preoperative form	Height (centimeters)	0.99 (0.98–0.99)
	Weight (kilograms)	0.99 (0.99–0.99)
	BMI (calculated)	0.99 (0.98–0.99)

*Fleiss weighted kappa for ordinal data, intraclass correlation coefficient (ICC) for continuous data. ICC was calculated using a two-way mixed model and absolute agreement (average measures)

** ASA: American Society of Anesthesiologists—classification of physical status (1–5). Mean ASA score was 2.17 in NORspine and 2.14 in EPR

***Mean number of previous spine surgeries was 1.29 in NORspine and 1.42 in EPR

****Type of surgery was graded as decompression or decompression and fusion

*****Decompression options were unilateral foraminotomy, crossover ("over the top"), or bilateral foraminotomy

Table 4 Prevalence of relevant comorbidities reported by NORspine compared to relevant comorbidities reported in EPRs for 474 patients operated for LSS

Diagnosis	NORspine, *n* (%)	EPR, *n* (%)
Hypertension	161 (34.0%)	243 (51.3%)
Heart disease	119 (25.1%)	107 (22.6%)
Hip or knee osteoarthritis	31 (6.5%)	67 (14.1%)
Diabetes	49 (10.3%)	67 (14.1%)
Depression or anxiety	12 (2.5%)	54 (11.4%)
Rheumatoid arthritis	7 (1.5%)	15 (3,2%)
Ankylosing spondylitis	1 (0.2%)	3 (0.6%)
Other rheumatic disorder	19 ((4.0%)	25 (5.3%)
Chronic pain	16 (3.4%)	40 (8.4%)
Chronic neurologic disorder	13 (2.7%)	34 (7.2%)
Peripheral vascular disease	14 (3.0%)	32 (6.8%)
Chronic lung disease	52 (11.0%)	89 (18.8%)
Cancer	17 (3.6%)	47 (9.9%)
Osteoporosis	6 (1.3%)	23 (4.9%)
Endocrine disorder	24 (5.1%)	49 (10.3%)
Other	75 (15.8%)	218 (46.0%)

Smoking status had an almost perfect agreement (κ (95%CI) = 0.93 (0.89–0.97)), a proportion correctly classified of 97.2%, and a sensitivity of 92.0%. Furthermore, as shown in Table 3, the patients' height, weight, and BMI showed excellent agreement between NORspine and EPRs (ICC = 0.99 to 0.99).

Discussion

This cross-sectional study compared Norwegian spine registry (NORspine) data to corresponding EPR data. We found a weak agreement for perioperative complications, a moderate agreement for ASA classification, a moderate to strong agreement for perioperative details, and almost perfect agreement for demographics. NORspine underreported perioperative complications and comorbidity.

Perioperative complications had a weak agreement and were underreported (sensitivity of only 40%) in NORspine. For example, dural tears were recorded in 13 patients (2.7%) in NORspine and 25 patients (5.3%) in EPR. Physicians' underreporting of surgical complications has been previously reported [12–17]. In line with our findings, a Swedish study of medical registers by Øhrn et al. from 2011 showed that only 74 of 210 (35%) of complications registered in a patient claim database had been recorded in the Swedish spine register [18]. Furthermore, a study validating German spine register data found wrong entries ranging from 10 to 50% for variables describing complications and reoperations [7]. Still, a sensitivity of 40% for surgeon-reported perioperative complications in the present study was unexpectedly low. We found a class average accuracy (CAA) for all perioperative complications of 99.4%; however, some of the complications listed are extremely rare, and CAA may, therefore, overestimate the accuracy of complication reporting. Previously published data on the prevalence of perioperative complications range between 3 and 16% [19–22]. The corresponding number in NORspine was 3.2%, also indicating an underreporting. EPRs documented 6.5% perioperative complications – a number more concordant with previous studies. Perioperative complications are recorded in NORspine and EPR at the same time point, and these data sources should match. Possible explanations for the discrepancy between the frequencies of complications recorded in NORspine and EPRs can be different definitions; for example, a minor repaired dural tear may not be graded as a complication by some surgeons.

ASA classification showed a moderate agreement, and the means between the two data sources were similar (2.17 vs. 2.14), illustrating no tendency to either under- or over-classification. The German spine register validation study reported wrong entries for ASA classification in 25% of cases and showed that a relatively simple classification system might be reported inaccurately [7]. However, all classification systems are subject to interpretation and inherent disagreement. We considered the ASA classification recorded in EPRs by anesthetists as the gold standard. However, the surgeon completing the NORspine form could either miss or disagree with the ASA classification provided by the anesthetist or use an ASA score recorded elsewhere in the EPR.

Each comorbidity was underreported in NORspine; this may be because surgeons could have different definitions of comorbidity they considered relevant, which illustrates a problem with the concept validity of this item in the NORspine questionnaire. Carreon et al. studied the comorbidity in patients with spinal stenosis in 2003 [21]. They found prevalence on the same level as we did in EPR, which supports our conclusion that comorbidity was underreported in NORspine. Moreover, previous studies have found low accuracy for orthopedic surgeons performing coding of diagnoses and indications for surgery, assessing cognitive function, and registering antibiotic use [23–25]. The discrepancy in the recorded prevalence of depression and anxiety in NORspine vs. EPRs may indicate that spine surgeons are not sufficiently aware of patients' mental health and how mental health may influence the clinical results (PROMs) after spinal surgery.

One should consider alternative ways of assessing comorbidity. However, other comorbidity scoring systems as frailty score and comorbidity indices (Charlson comorbidity index (CCI) and Elixhauser comorbidity index) [26, 27] are more complex, possibly affecting response rates and accuracy. We found ASA classification to be the most feasible comorbidity measure, and it displayed moderate agreement in our study. Mannion et al. found that ASA was a strong predictor of complications after hip surgery, and adding a more complex score (CCI) was not superior in predicting postoperative complications [28]. Hence, we recommend using ASA classification over more complex measures despite its limitations.

There was a discrepancy in accuracy between the different variables concerning previous surgery. Previous spinal surgery (yes/no) had an agreement of 0.93, and the number of previous surgeries had an agreement of 0.62; this indicates that NORspine is more precise in recording patients who had any previous surgery than the exact number of previous surgeries.

Perioperative details were accurately registered, with the proportion correctly classified above 93%. There was a strong to excellent agreement between NORspine data and the EPR data, with kappa values above 0.90; this is also in line with the literature; orthopedic surgeons coded surgical procedures and classified x-rays accurately in previous studies [23, 24]. However, surgical access reported by the surgeon showed minimal agreement between NORspine and EPR. Defining surgical accesses in NORspine may have been subject to interpretation, as surgeons may have misinterpreted the "lateral/Wiltzes'" choice as the direct lateral approach. Therefore, the NORspine board plans to clarify and amend options for surgical accesses in the next version of the surgeon-reported questionnaire.

Smoking status is recorded in the EPR as a direct question to the patient and in the NORspine as a simple yes or no question. The source of these two variables was the same, the patient. However, there was an error rate of 2,5% (PCC 97,5%) and an agreement of 0.93. This variable can indicate the rate of random error in NORspine. Patients' height, weight, and BMI displayed excellent agreement. The patients themselves report these variables to NORspine, and their accuracy and agreement could serve as an aim for surgeon-recorded variables. It is questionable to define EPR as a gold standard because some variables could be more correctly reported by patients than healthcare personnel. A further step to improve data quality could be to use a combined construct of patient- and physician-recorded variables [4].

About 1% of NORspine data values were missing values, as compared to 3% in the EPR. This is in line with a literature review of data quality from 2002 [29]; they found 2% missing data in automatically collected and 5% in manually collected register data.

Our study has several limitations: We used EPRs as an external data source, although they may lack relevant information. EPR data might not be appropriate for some variables as a reference, so we chose to report both accuracy and agreement. Agreement would be a more appropriate measure when no clear reference standard exists. The EPRs at the four hospitals were not standardized (free text format) and could miss or misinterpret relevant information. On the other hand, every patient has an EPR, and it has been defined as a gold standard in other validation studies [4–7] and has a high medical and legal status. Ideally, to be defined as a complete gold standard, the EPR should record PROMs.

Another limitation was potential selection bias due to the non-randomized selection of hospitals. The accuracy of NORspine and EPR data registration could differ between hospitals, limiting the generalizability of our findings. However, most of the differences in patient characteristics between the four selected and the remaining hospitals reporting to NORspine were small, and some of them might be incidental findings. Therefore, the authors consider the patient sample representative for the broader population of the NORspine. Patients analyzed in the present study were operated on and included during 2015–2016, and no relevant changes have been made in NORspine since 2015. Therefore, we believe that our findings are still relevant.

The selection of variables had to be limited to those available and suitable for comparison in both data sources. Therefore, the concordance of some relevant variables could not be assessed (e.g., patient-reported disability and pain).

We only assessed patients who underwent decompression due to spinal stenosis, who were treated with a limited number of simple procedures and surgical accesses. Our results may, therefore, represent a "best-case scenario" regarding the quality of NORspine data.

The strength of this study was a comprehensive and systematic review of a large number of EPRs at four hospitals. We assessed both accuracy (PCC and sensitivity) and agreement (kappa or ICC) of patient—and surgeon-reported data to validate different NORspine variables.

Future perspectives and implications

A long-term goal could be the inclusion of clinical registry data in a structured EPR. Structured EPRs have been

implemented in Norway for hip fracture patients, and data from a structured EPR are sent directly to the national hip fracture audit. Structured EPRs can improve the quality of the EPR and the quality and completeness of registry data. Furthermore, structured EPRs could make valuable data more accessible to clinical research. A future perspective would be to integrate spine registers into a structured EPR.

Conclusions

This cross-sectional validation study showed that the Norwegian Registry for Spine Surgery (NORspine) tended to underreport perioperative complications to spine surgery compared to corresponding EPRs. This finding may represent a systematic error (information bias), and future register studies on complications after spinal surgery could cross-reference perioperative complications with other data sources to reduce the risk of underreporting. Comorbidities were also underreported in NORspine; the ASA classification seems the simplest and most reliable way to assess comorbidity. Perioperative details and patient-reported data had moderate to excellent agreement.

Appendix

See Tables 5, 6 and7.

Table 5 Patient-reported preoperative questionnaire (2015–2016 edition, authors translation)

Variables	NORspine question/alternatives	Recorded in EPR review
Do you smoke?	Yes/No	Yes/No
Height	Centimeters	Centimeters
Weight	Kilograms	Kilograms
Education, highest level	Five options	Not evaluated
Family and children		
Civil status	Married/cohabitant/Single	Not evaluated
How many children do you have?	Number	Not evaluated
First language	Norwegian/Sami (indigenous)/Other	Not evaluated
Pain level during last week	Numeric rating scale back pain 0–10	Not evaluated
	Numeric rating scale leg pain 0–10	Not evaluated
	Oswestry Disability Index	Not evaluated
Quality of life	EuroQol-5Dimensions-3Level	Not evaluated
	EuroQol VAS (0 worst – 100 best)	Not evaluated
Analgesics	Do you use analgesics due to back and/or leg pain? Yes/No	Not evaluated
	How often? (Four options)	Not evaluated
Duration of back symptoms	Four options	Not evaluated
Duration of leg symptoms	Four options	Not evaluated
Occupational status	Ten options	Not evaluated
Have you applied for a disability benefit?	Four options	Not evaluated
Applied for compensation from insurance co. or workers' comp?	Four options	Not evaluated

Table 6 Surgeon-reported perioperative questionnaire (2015–2016 edition, authors' translation)

Variable	NORspine question/alternative(s)	Recorded from EPRs
Previous back surgery	Yes, same level/Yes, different level/No	Yes (any)/No
Previous surgery, number	Number of previous lumbar surgeries (count)	Number of previous lumbar surgeries (count)
Other relevant diseases, injuries, or complaints	No	No
Yes, specify	Rheumatoid arthritis	Rheumatoid arthritis
	Ankylosing spondylitis	Ankylosing spondylitis
	Other rheumatic disorder	Other rheumatic disorder
	Hip or knee osteoarthritis	Hip or knee osteoarthritis
	Depression or anxiety	Depression or anxiety
	Chronic musculoskeletal pain	Chronic musculoskeletal pain
	Chronic neurologic disorder	Chronic neurologic disorder
	Cerebrovascular disease	Cerebrovascular disease
	Cardiac or vascular disease	Cardiac or vascular disease
	Peripheral vascular disease	Peripheral vascular disease
	Chronic pulmonary disease	Chronic pulmonary disease
	Cancer	Cancer
	Osteoporosis	Osteoporosis
	Hypertension	Hypertension
	Diabetes mellitus	Diabetes mellitus
	Other endocrine disorder	Other endocrine disorder
	Other, specify	Other, specify
Radiologic examination	CT	Not evaluated
	MRI	Not evaluated
	Radiculography	Not evaluated
	Discography	Not evaluated
	Diagnostic block	Not evaluated
	Lumbosacral spine X-ray	Not evaluated
	With flexion/extension	Not evaluated
Radiologic assessment	Normal	Not evaluated
	Disk herniation	Not evaluated
	Spinal stenosis, central	Not evaluated
	Spinal stenosis, lateral	Not evaluated
	Foraminal stenosis	Not evaluated
	Degenerative spine/disk	Not evaluated
	Isthmic spondylolisthesis	Not evaluated
	Degenerative spondylolisthesis	Not evaluated
	Degenerative scoliosis	Not evaluated
	Synovial cyst	Not evaluated
	Pseudomeningocele	Not evaluated
	Other, specify	Not evaluated
Indication for surgery	Back/hip pain/Leg pain/Both	Not evaluated
	Palsy Yes/No	Not evaluated
	Palsy grade (0–5)	Not evaluated
	Cauda equina syndrome	Not evaluated
	Other, specify	Not evaluated

Table 6 (continued)

Variable	NORspine question/alternative(s)	Recorded from EPRs
Early reoperation (90 days), cause	Recurrent disk herniation	Not evaluated
	Dural tear	Not evaluated
	Hematoma	Not evaluated
	Deep infection	Not evaluated
	Superficial infection	Not evaluated
	Post-op spondylolisthesis	Not evaluated
	Loosening/malposition of implants	Not evaluated
Type of surgery	Planned/emergency/in between	Not evaluated
	Outpatient surgery (yes/No)	Not evaluated
ASA classification	1–5 (explained in text)	1–5
Surgical method	Microscope/magnifying glasses (yes/no)	Not evaluated
	Extirpation of disk herniation No/Yes with complete discectomy/Yes without complete discectomy	Not evaluated
	Midline preserving decompression (yes/no)	Midline preserving decompression (yes/no)
	Unilateral (Yes/No)	Unilateral (Yes/No)
	Crossover ("over the top") (Yes/No)	Crossover ("over the top") (Yes/No)
	Bilateral laminotomy yes/No	Bilateral laminotomy (Yes/No)
	Laminectomy Yes/No	Laminectomy (Yes/No)
	Removal of facet joint yes/no	Not evaluated
	Removal of facet joint uni-/bilateral	Not evaluated
Other surgical methods	Endoscopic	Not evaluated
	Minimally invasive (tube)	Not evaluated
	Interspinous implant	Not evaluated
	Removal of interspinous implant	Not evaluated
	Disk prosthesis	Not evaluated
	Nucleus implant	Not evaluated
	Nucleotomy	Not evaluated
	Chemical nucleolysis	Not evaluated
	Revision of implants	Not evaluated
	Removal of implants	Not evaluated
	Other, specify	Not evaluated
Surgical access	Posterior midline, lateral (Wiltze), anterior	Posterior midline, lateral (Wiltze), anterior
Fusion	Posterolateral (instrumented/bone graft)	Fusion (Yes/No)
	ALIF (cage/bone block)	
	PLIF (cage/bone block)	
	TLIF (cage/bone block)	
	Other, specify	
Type of bone graft	Autograft/bone substitute/allograft	Not evaluated
Spinal level(s) and side(s) treated	L2-3 (right/left)	L2-3 (right/left)
	L3-4 (right/left)	L3-4 (right/left)
	L4-5 (right/left)	L4-5 (right/left)
	L5-S1 (right/left)	L5-S1 (right/left)
	Other, specify	Other, specify
Antibiotic prophylaxis	Yes/No	Not evaluated
Wound drainage	Yes/No	Not evaluated
Surgical time	Start/stop (time)	Not evaluated

Table 6 (continued)

Variable	NORspine question/alternative(s)	Recorded from EPRs
Perioperative complications	Dural tear/CSF leakage	Dural tear/CSF leakage
	Nerve root damage	Nerve root damage
	Operated wrong level/side	Operated wrong level/side
	Malpositioned implant	Malpositioned implant
	Bleeding (necessitating transfusion)	Bleeding (necessitating transfusion)
	Respiratory complications	Respiratory complications
	Cardiovascular complications	Cardiovascular complications
	Anaphylactic reaction	Anaphylactic reaction
	Other, specify	Other, specify
Surgical code	NOMESCO classification of Surg Proc	Not evaluated
Length of stay	Number of days	Not evaluated
Death, in hospital	Cause of death	Not evaluated

Table 7 Completeness of NORspine compared to electronic patient records (EPR) for 474 patients operated for spinal stenosis

Variable	Missing data in NORspine, *n* (%)	Missing data in EPR review, *n* (%)
Age	None	None
Sex	None	None
BMI	9 (1.9%)	20 (4.2%)
Smoking	5 (1.1%)	39 (8.2%)
Comorbidity, any	29 (6.1%)	10 (2.1%)
Previous surgery, any	3 (0.6%)	11 (2.3%)
Access, posterior midline	12 (2.5%)	13 (2.7%)
Level L2-3	None	14 (3.0%)
Level L3-4	None	13 (2.7%)
Level L4-5	None	13 (2.7%)
Level L5-S	None	13 (2.7%)
Number of levels operated	None	13 (2.7%)
Type of surgery	None	15 (3.2%)
Perioperative complications	None*	11 (2.3%)
Total ($n=6{,}636$)**	58 of 6,636 (0.9%)	185 of 6,636 (2.8%)

*Perioperative complications were registered as either "yes" or "no"; there were no missing data regarding complications in NORspine, but NORspine underreported complications compared to EPR

**The total number is number of variables (14) multiplied with number of patients (474)

Author's contribution All authors have contributed to some or all significant parts of the study.

Funding Open access funding provided by NTNU Norwegian University of Science and Technology (incl St. Olavs Hospital - Trondheim University Hospital). No funding was received for conducting this study.

Availability of data and material Data available on request.

Declarations

Conflicts of interest Financial interests: The authors have no relevant financial interests to disclose. Non-financial interests: Author Tore Solberg is scientific leader of the Norwegian Registry of Spine Surgery, and author Greger Lønne is member of the board of the Norwegian Registry of Spine Surgery.

Ethics approval The study was approved by the Norwegian Regional Committee for medical and health research ethics approved this study (reference no. 2017(2157), as well as the data protection officers at the four hospitals. The study was conducted in compliance with the Helsinki Declaration.

Consent to participate All patients have provided informed consent, and the study was conducted in compliance with the Helsinki Declaration

References

1. Brooke EM, World Health Organization (1974) The current and future use of registers in health information systems / Eileen M. World Health Organization, Brooke

2. Kodra Y, de la Paz MP, Coi A, Santoro M, Bianchi F, Ahmed F, Rubinstein YR, Weinbach J, Taruscio D (2017) Data quality in rare diseases registries. Adv Exp Med Biol 1031:149–164. https://doi.org/10.1007/978-3-319-67144-4_8
3. Derakhshan P, Azad Z, Naghdi K, Safdarian M, Zarei M, Jazayeri SB, Sharif-Alhoseini M, Zendehdel K, Amirjamshidi A, Ghodsi Z, Faghih M, Mohammadzadeh M, Khazaie Z, Zadegan S, Abedi A, Sadeghian F, Rahimi-Movaghar V (2018) P206 -The impact of data quality assurance and control solutions on the completeness, accuracy and consistency of data in a national spine registry (NSCIR). Global Spine Conferance 2018
4. Varmdal T, Bakken IJ, Janszky I, Wethal T, Ellekjær H, Rohweder G, Fjærtoft H, Ebbing M, Bønaa KH (2016) Comparison of the validity of stroke diagnoses in a medical quality register and an administrative health register. Scand J Public Health 44(2):143–149. https://doi.org/10.1177/1403494815621641
5. Löfgren L, Eloranta S, Krawiec K, Asterkvist A, Lönnqvist C, Sandelin K, Steering group of the National Register for Breast Cancer (2019) Validation of data quality in the Swedish national register for breast cancer. BMC Public Health 19(1):495. https://doi.org/10.1186/s12889-019-6846-6
6. Landberg A, Bruce D, Lindblad P, Ljungberg B, Lundstam S, Thorstenson A, Sundqvist P (2021) Validation of data quality in the National Swedish kidney cancer register. Scand J Urol 55(2):142–148. https://doi.org/10.1080/21681805.2021.1885485
7. Meyer B, Shiban E, Albers LE, Krieg SM (2020) Completeness and accuracy of data in spine registries: an independent audit-based study. Eur Spine J 29:1453–1461. https://doi.org/10.1007/s00586-020-06342-6
8. Solberg T, Olsen RR, Berglund ML (2018) Annual report for NORspine 2018. Accsessed 11th Novbember 2021
9. Altmann DG (1991) Practical statistics for medical research, Chap 14. Chapmall & Hall, USA, pp 396–439
10. McHugh ML (2012) Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 22(3):276–282 (**PMID: 23092060; PMCID: PMC3900052**)
11. Koo TK, Li MY (2016) A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research [published correction appears in J Chiropr Med. 2017 Dec;16(4):346]. J Chiropr Med 15(2):155–163. Doi:https://doi.org/10.1016/j.jcm.2016.02
12. Healey MA, Shackford SR, Osler TM, Rogers FB, Burns E (2002) Complications in surgical patients. Arch Surg 137(5):611–618. https://doi.org/10.1001/archsurg.137.5.611 (**PMID: 11982478**)
13. Cooper GS, Kou TD, Rex DK (2013) Complications following colonoscopy with anesthesia assistance: a population-based analysis. JAMA Intern Med 173(7):551–556. https://doi.org/10.1001/jamainternmed.2013.2908
14. Litwin MS, Lubeck DP, Henning JM, Carrol PR (1998) Differences in urologist and patient assessment of health related quality of life in men with prostate cancer: results of the capsure database. J Urol 159:1988–1992. https://doi.org/10.1016/S0022-5347(01)63222-1
15. Hutter MM, Rowell KS, Devaney LA, Sokal SM, Warshaw AL, Abbott WM, Hodin RA (2006) Identification of surgical complications and deaths: an assessment of the traditional surgical morbidity and mortality conference compared with the american college of surgeons-national surgical quality improvement program. J Am Coll Surg 203(5):618–624. https://doi.org/10.1016/j.jamcollsurg.2006.07.010 (**ISSN 1072-7515**)
16. Fromme EK, Eilers KM, Mori M, Hsieh Y-C, Beer TM (2004) How Accurate Is clinician reporting of chemotherapy adverse effects? A comparison with patient-reported symptoms from the quality-of-life questionnaire C30. J Clin Oncol 22(17):3485–3490
17. Grossman SA, Sheidler VR, Swedeen K, Mucenski J, Piantadosi S (1991) Correlation of patient and caregiver ratings of cancer pain. J Pain Symptom Manag 6(2):53–57. https://doi.org/10.1016/0885-3924(91)90518-9 (**PMID: 2007792**)
18. Öhrn A, Elfström J, Liedgren C, Rutberg H (2011) Reporting of sentinel events in Swedish hospitals: a comparison of severe adverse events reported by patients and providers. Jt Comm J Qual Patient Saf 37(11):495–501. https://doi.org/10.1016/S1553-7250(11)37063.ISSN1553-7250
19. Deyo RA, Hickam D, Duckart JP, Piedra M (2013) Complications after surgery for lumbar stenosis in a veteran population. Spine (Phila Pa 1976) 38(19):1695–1702. https://doi.org/10.1097/BRS.0b013e31829f65c1
20. Cassinelli EH, Eubanks J, Vogt M, Furey C, Yoo J, Bohlman HH (2007) Risk Factors for the development of perioperative complications in elderly patients undergoing lumbar decompression and arthrodesis for spinal stenosis. Spine 32(2):230–235. https://doi.org/10.1097/01.brs.0000251918.19508.b3
21. Carreon LY, Puno RM, Dimar JR, Glassman SD, Johnson JR (2003) Perioperative complications of posterior lumbar decompression and arthrodesis in older adults. J Bone Jt Surg 85(11):2089–2092
22. Papavero L, Engler N, Kothe R (2015) Incidental durotomy in spine surgery: first aid in ten steps. Eur Spine J 24:2077–2084. https://doi.org/10.1007/s00586-015-3837-x
23. Kazberouk ABS, Martin BI, Stevens JP, McGuire KJ (2015) Validation of an administrative coding algorithm for classifying surgical indication and operative features of spine surgery. Spine 40(2):114–120. https://doi.org/10.1097/BRS.0000000000000682
24. Kristoffersen MH, Dybvik E, Steihaug OM et al (2019) Validation of orthopaedic surgeons' assessment of cognitive function in patients with acute hip fracture. BMC Musculoskelet Disord. https://doi.org/10.1186/s12891-019-2633-x
25. Horner NS, Grønhaug-Larsen KM, Svanteson E, Samuelsson K, Aveni OR et al (2020) Timing of hip hemiarthroplasty and the influence on prosthetic joint infection. PLOS ONE 15(3):e0229947. https://doi.org/10.1371/journal.pone.0229947
26. Charlson ME, Pompei P, Ales KL, MacKenzie CR (1987) A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 40(5):373–383. https://doi.org/10.1016/0021-9681(87)90171-8 (**PMID: 3558716**)
27. Elixhauser A, Steiner C, Harris DR, Coffey RM (1998) Comorbidity measures for use with administrative data. Med Care 36:8–27
28. Mannion AF, Nauer S, Arsoy D, Impellizzeri FM, Leunig M (2020) The association between comorbidity and the risks and early benefits of total hip arthroplasty for hip osteoarthritis. J Arthroplasty 35(9):2480–2487. https://doi.org/10.1016/j.arth.2020.04.090
29. Arts DGT, de Keizer N, Scheffer GJ (2002) Defining and improving data quality in medical registries: a literature review, case study, and generic framework. J Am Med Inform Assoc 9(6):600–611. https://doi.org/10.1197/jamia.M1087

12.2 Paper 2

Kaur *et al. BMC Musculoskeletal Disorders* (2023) 24:164
https://doi.org/10.1186/s12891-023-06267-3

BMC Musculoskeletal Disorders

RESEARCH Open Access

Characteristics and outcomes of patients who did not respond to a national spine surgery registry

Simran Kaur[1*], Ole Kristian Alhaug[2,3,4], Filip C. Dolatowski[5], Tore K. Solberg[6,7] and Greger Lønne[2,4]

Abstract

Background Loss to follow-up may bias outcome assessments in medical registries. This cohort study aimed to analyze and compare patients who failed to respond with those that responded to the Norwegian Registry for Spine Surgery (NORspine).

Methods We analyzed a cohort of 474 consecutive patients operated for lumbar spinal stenosis at four public hospitals in Norway during a two-year period. These patients reported sociodemographic data, preoperative symptoms, and Oswestry Disability Index (ODI), numerical rating scales (NRS) for back and leg pain to NORspine at baseline and 12 months postoperatively. We contacted all patients who did not respond to NORspine after 12 months. Those who responded were termed responsive non-respondents and compared to 12 months respondents.

Results One hundred forty (30%) did not respond to NORspine 12 months after surgery and 123 were available for additional follow-up. Sixty-four of the 123 non-respondents (52%) responded to a cross-sectional survey done at a median of 50 (36–64) months after surgery. At baseline, non-respondents were younger 63 (SD 11.7) vs. 68 (SD 9.9) years (mean difference (95% CI) 4.7 years (2.6 to 6.7); $p = < 0.001$) and more frequently smokers 41 (30%) vs. 70 (21%) RR (95%CI) = 1.40 (1.01 to 1.95); $p = 0.044$. There were no other relevant differences in other sociodemographic variables or preoperative symptoms. We found no differences in the effect of surgery on non-respondents vs. respondents (ODI (SD) = 28.2 (19.9) vs. 25.2 (18.9), MD (95%CI) = 3.0 (-2.1 to 8.1); $p = 0.250$).

Conclusion We found that 30% of patients did not respond to NORspine at 12 months after spine surgery. Non-respondents were somewhat younger and smoked more frequently than respondents; however, there were no differences in patient-reported outcome measures. Our findings suggest that attrition bias in NORspine was random and due to non-modifiable factors.

Keywords Loss to follow-up, Spine surgery, Attrition bias, Registry, Non-response, Lumbar spinal stenosis

*Correspondence:
Simran Kaur
Simran.85.k@gmail.com
[1] Department of Orthopedic Surgery, Martina Hansens Hospital, Sandvika, Norway
[2] Innlandet Hospital Trust, Brumunddal, Norway
[3] Akershus University Hospital, Nordbyhagen, Norway
[4] Norwegian University of Science and Technology, Trondheim, Norway
[5] Division of Orthopedic Surgery, Oslo University Hospital, Oslo, Norway
[6] Department of Neurosurgery & Norwegian Registry for Spine surgery, University Hospital of North Norway, Tromsø, Norway
[7] Department of Clinical Medicine, The Arctic University of Norway (UiT), Tromsø, Norway

Background

Medical registries provide clinicians with large data sets of high external validity and complement randomized controlled trials that examine more targeted populations and treatments [1, 2]. Medical registries can guide decision-making and improve the quality of care by monitoring patient-reported outcome measures (PROMs) stratified by different populations, diagnoses, and treatments [3, 4]. Medical registries face higher attrition rates compared to clinical trials—rigorous attempts to attain data are costly and impractical in a registry setting [5]. Still, sufficient follow-up rates are crucial for the quality of registries, and awareness of follow-up rates is important when interpreting register data.

Non-respondents may systematically differ from respondents and introduce attrition bias that compromises the validity of register data [1, 6–8]. However, some studies suggest that non-response occurs at random [2, 4, 5]. The last assessment of non-respondents in NORspine was conducted in 2007 and reported a loss to follow-up of 22% at two years postoperatively and did not reveal any differences in outcomes between non-respondents and respondents [5]. This study was conducted before NORspine expanded to a national registry, and a reassessment is warranted. In order to assess the impact of attrition on NORspine data, we aimed to assess baseline characteristics and clinical outcomes for patients who responded at 12 months after surgery compared to those who did not.

Methods

This cohort study was based on retrospective analyses of prospectively collected NORspine data. We compared baseline variables for patients who did not respond to NORspine at 12 months after surgery with those who had responded. We reached out to those who did not respond to NORspine at 12 months after surgery and performed an additional cross-sectional survey at a median of 50 (36–64) months after surgery. We assessed clinical outcomes for those who finally responded to our additional questionnaire. As an additional analysis, we also compared the baseline variables of the subgroup that never responded compared to those who responded to the additional cross-sectional survey.

NORspine

All Norwegian hospitals that offer spine surgery are obliged to report to NORspine. Currently, 70% of all degenerative spine surgeries done in Norway are registered in NORspine [9]. NORspine is a consent-based register. Patients with primary infections of the spine, fractures of the spine, and patients who are unable to comprehend questionaries in Norwegian, are not invited to participate.

A NORspine dataset consists of both patient- and surgeon-reported variables. Patients complete a standardized questionnaire preoperatively on sociodemographic data such as age, sex, native language, level of education, and marital status. Patients also report preoperative symptoms, as assessed by validated PROMs: Oswestry Disability Index (ODI) ranging from 0 (minimal disability) to 100 (bedbound), Numeric Rating Scales (NRS) ranging from 0 (no pain) to 10 (worst imaginable pain) for back and leg pain, and quality of life as assessed by EuroQol 5 Dimension 3 level—0.59 (worse than dead) to 1,0 (perfect health) [10–13].

Surgeons report directly after the surgery on diagnoses, relevant comorbidities, and perioperative details such as the type of surgery. The NORspine sends follow-up questionnaires to patients at 3 and 12 months after surgery by regular mail, including one reminder if the patient does not reply. Patients report directly to NORspine at follow-ups using PROMs (ODI, NRS back and leg pain, EQ5D, and Global Perceived Effect (GPE)—a seven-point Likert scale (1 = completely recovered, 2 = much improved, 3 = slightly improved, 4 = unchanged, 5 = slightly worse, 6 = much worse, 7 = worse than ever) [14].

Data collection

We analyzed prospectively collected NORspine data on patients operated for lumbar spinal stenosis (LSS) at four hospitals between January 1st, 2015 and December 31st, 2016. Patients who consented to participate in NORspine completed questionnaires at baseline. The NORspine registry then mailed similar questionnaires to patients at 12 months postoperatively. Patients responded directly to NORspine without the involvement of the treating center. NORspine routinely sends one postal reminder to those who do not respond before they are considered non-respondents. We engaged the NORspine office to reach out to those who did not respond at 12 months after surgery. The 12 months postoperative questionnaire was sent once again. We also sent one reminder by mail and one by SMS to those who still did not respond. Patients that responded at 12 months postoperatively are termed respondents, while those who did not respond are termed non-respondents. Those who finally responded are termed "responsive non-respondents", and those who never responded to any contact are termed "resistant non-respondents".

Baseline and outcome measures

At baseline, we compared 140 non-respondents with 334 respondents. In our cross-sectional analysis, we were able to contact 123 of the 140 non-respondents—17 were classified as "unknown address", "moved abroad", or "deceased" (Fig. 1). We then compared clinical outcomes

assessed by PROMs between responsive non-respondents (median 50 months after surgery) and respondents (12 months after surgery). We also dichotomized clinical outcome using the GPE scale, defining success as "completely recovered" and "much improved", and compared the proportions of successfully treated non-respondents versus respondents.

Finally, we compared the baseline characteristics of the responsive non-respondents and the resistant non-respondents.

Statistics

We used descriptive statistics presented by means (SD) for continuous variables and numbers (percentages) for categorical variables. We analyzed between-group differences by mean difference (95%CI) and Student's T-test for continuous variables, or relative risk (95%CI) and z-statistics for categorical variables. Statistical analyses were performed using SPSS, version 26 (IBM Corp., Armonk, N.Y. USA) and MedCalc Software Ltd. Relative risk calculator. https://www.medcalc.org/calc/relative_risk.php (Version 20.027; accessed March 14, 2022).

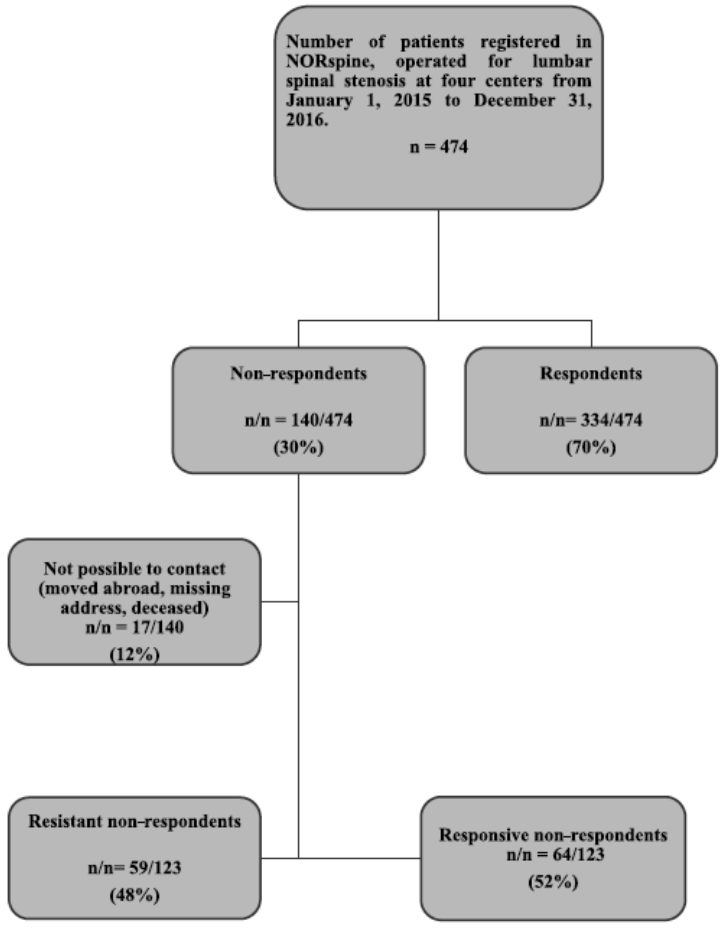


Fig. 1 Study flowchart

Ethical considerations

All patients provided an informed consent when entering the registry. The Norwegian national ethical board (Regional Committee for medical and health research ethics, reference number 2017/2157) approved this study, as did the data protection officers at the four participating hospitals. This study was conducted in accordance with the Helsinki declaration [15].

Results

As seen in Fig. 1, of the 474 consenting patients, 140 (30%) patients did not return the questionnaire at 12 months postoperatively. At the time of cross-sectional data collection, 17 patients were not possible to contact, leaving 123 for analysis. Of the 123 non-respondents, 64 (52%) patients returned questionnaires ("responsive non-respondents"), while 59 (48%) failed to respond ("resistant non-respondents").

Baseline characteristics

The non-respondents were younger than the respondents, 63 (SD 11.7) vs. 68 (SD 9.9) years, mean difference (95% CI) 4.7 years (2.59 to 6.74); $p = < 0.001$. Non-respondents were more frequently smokers compared to respondents: 41 (30%) vs. 70 (21%), RR (95%CI) 1.40 (1.01 to 1.95); $p = 0.044$. Furthermore, non-respondents had a lower proportion of surgeon-reported relevant comorbidities compared to respondents 93 (69%) vs. 243 (78%), RR (95%CI) 0.89 (0.77 to 1.00); $p = 0.047$. However, we found no difference in ASA classification between non-respondents and respondents: the number (%) of ASA grades 1 and 2 was 111 (79%) vs. 242 (72%) RR (95% CI) 1.09 (0.98 to 1.22); $p = 0.100$. As shown in Table 1, there were no other differences between the non-respondents and respondents at baseline. Also, we found no differences in the type of surgery (decompression only vs. decompression and additional fusion) among the non-respondents and respondents.

Table 1 Baseline characteristics and perioperative data of 474 patients with lumbar stenosis who reported to NORspine

	N Missing	Non-respondents (SD, %)	N Missing	Respondents (SD, %)	Mean difference (95% CI) or relative risk (95% CI)	*P* value
Age (years)	N = 140 Missing = 0	63.0 (11.7)	N = 334 Missing = 0	67.7 (9.9)	-4.7 (-6.74 to -2.59)	< 0.001
Female	N = 140 Missing = 0	71 (51%)	N = 334 Missing = 0	183 (55%)	0.93 (0.77 to 1.12)	0.426
BMI	N = 136 Missing = 4	28.6 (4.5)	N = 329 Missing = 5	28.2 (4.5)	0.4 (-0.48 to 1.33)	0.362
Comorbidities*	N = 135 Missing = 5	93 (69%)	N = 310 Missing = 24	243 (78%)	0.89 (0.77 to 1.00)	0.047
ASA grade I and II	N = 140 Missing = 0	111 (79%)	N = 334 Missing = 0	242 (72%)	1.09 (0.98 to 1.22)	0.100
Smokers	N = 138 Missing = 2	41 (30%)	N = 331 Missing = 3	70 (21%)	1.40 (1.01 to 1.95)	0.044
Norwegian as first language	N = 138 Missing = 2	130 (94%)	N = 331 Missing = 3	324 (98%)	0.96 (0.92 to 1.01)	0.090
University or college education > 4 years	N = 136 Missing = 4	30 (22%)	N = 328 Missing = 6	83 (25%)	0.87 (0.60 to 1.26)	0.463
Single civil status	N = 139 Missing = 1	30 (22%)	N = 332 Missing = 2	85 (26%)	0.84 (0.58 to 1.22)	0.361
Preoperative ODI	N = 136 Missing = 4	42.3 (16.1)	N = 329 Missing = 5	40.4 (15.8)	1.87 (-1.31 to 5.06)	0.248
Preoperative NRS back pain	N = 127 Missing = 13	6.9 (2.0)	N = 312 Missing = 22	6.8 (2.1)	0.17 (-0.26 to 0.61)	0.430
Preoperative NRS leg pain	N = 123 Missing = 17	6.9 (2.2)	N = 311 Missing = 23	7.0 (2.1)	-0.04 (-0.49 to 0.40)	0.844
Decompression only type surgery	N = 140 Missing = 0	122 (87%)	N = 334 Missing = 0	301 (90%)	0.98 (0.84 to 1.15)	0.820
Fusion type surgery	N = 140 Missing = 0	18 (13%)	N = 334 Missing = 0	33 (10%)	1.27 (0.74 to 2.18)	0.393

* Comorbidities that were assessed as relevant by the reporting surgeon

Clinical outcomes

As presented in Table 2, we did not find any differences in mean (SD) ODI scores between the responsive non-respondents and respondents postoperatively 28.2 (19.9) vs. 25.2 (18.9), mean difference (95% CI) = 3.0 (-2.1 to 8.1); $p = 0.250$. Nor did we find any differences between responsive non-respondents versus respondents for NRS back pain, 4.6 (3.0) vs. 4.1 (2.9), mean difference (95% CI) 0.43 (-0.3 to 1.2); $p = 0.271$ or NRS leg pain score 4.0 (3.2) vs. 3.9 (3.1) mean difference (95% CI) 0.15 (-0.7 to 1.0); $p = 0.719$. Finally, we found similar proportions of successively treated patients among non-respondents and respondents, as assessed by GPE (63 (70%) vs. 330 (79%), RR (95%CI) 0.89 (0.75 to 1.06); $p = 0.183$).

Resistant non-respondents

Appendix Table 1 compares the responsive non-respondents (64 (52%)) to resistant non-respondents (59 (48%)). We did not find any age difference; however, resistant non-respondents were more frequently smokers (22 (38%) vs. 13 (20%), RR (95% CI) 1.87 (1.04 to 3.36); p = 0.037). As shown in Appendix Table 1, we did not find differences in other baseline characteristics such as sex, marital status, level of education, native language, ASA grade, or preoperative PROM levels.

Discussion

The main findings from this register-based cohort study of patients who had spinal surgery due to lumbar spinal stenosis were that non-respondents were somewhat younger and tended to smoke more often than those who responded. Moreover, we found no differences in PROM scores between non-respondents compared to respondents, neither at baseline nor after surgery.

Several studies have demonstrated that non-respondents are younger than respondents [2, 4, 16–19]. Completing and posting questionnaires consumes time, and younger patients may be busier due to work and family obligations. Our finding that non-respondents were more frequently smokers has also been supported by others [2, 4, 6, 17, 18]. Also, we found that surgeons reported fewer relevant comorbidities for non-respondents than respondents. However, the variable "relevant comorbidity" is subject to interpretation by the treating surgeon. Therefore, the registration of relevant comorbidities by the treating surgeon may be questioned. A validation study of NORspine data found that surgeons tended to underreport relevant comorbidities and that ASA grading done by the anesthetist could be more reliable in assessing comorbidity [20]. In our study, there was no difference in the proportions of ASA grades 1 and 2 patients among non-respondents compared to respondents.

In addition to young age and smoking, previous studies of non-respondents also reported a predominance of the male gender, living alone, higher anxiety levels, and worse PROM scores [2, 4–6, 16–19]. Two observational spine studies found that non-respondents had higher ODI scores, lower quality of life (EuroQol 5D), and lower function (Short form health survey—SF-36) preoperatively compared to those who responded [2, 19]. The aforementioned studies implied that non-respondents had a worse starting point and were not quite representative of the entire register population. However, these findings were not reproduced in our study. Neither at baseline nor at follow-up did we find any differences in ODI between the non-respondents and respondents (Tables 1 and 2).

Another Swedish spine register study reported that non-respondents had inferior clinical outcomes [6], while other studies support our findings of similar postoperative outcomes for non-respondents versus respondents [2, 4, 5, 16–18]. Minor differences in PROMs have been

Table 2 Postoperative clinical outcomes for responsive non-respondents and respondents operated for lumbar spinal stenosis

	N Missing	Responsive non-respondents* Mean (SD) / n (%)	N Missing	Respondents** Mean (SD)/ n (%)	Mean diff (95% CI) or Relative risk (95% CI)	*P*-value
ODI	N = 64 Missing = 0	28.2 (19.9)	N = 333 Missing = 1	25.2 (18.9)	2.99 (-2.1 to 8.1)	0.250
NRS back pain	N = 64 Missing = 0	4.6 (3.0)	N = 328 Missing = 5	4.1 (2.9)	0.43 (-0.3 to 1.2)	0.271
NRS leg pain	N = 63 Missing 1	4.0 (3.2)	N = 321 Missing = 12	3.9 (3.1)	0.15 (-0.7 to 1.0)	0.719
Success by GPE***	N = 63 Missing 1	63 (70%)	N = 330 Missing = 3	330 (79%)	0.89 (0.8 to 1.1)	0.183

* PROM scores collected retrospectively at a median of 50 months after surgery

** PROM scores collected prospectively at 12 months after surgery

*** Success defined as "completely recovered" or "much recovered" on the GPE scale

reported between non-respondents and respondents, but the magnitudes of these differences were assessed as clinically irrelevant [21].

Some studies suggest that loss to follow-up of as little as 5% [22, 23] may cause bias, while rates above 20% [24] could potentially lead to serious bias. There is a variation in loss to follow-up rates in spine register studies ranging from 12% [4] to 42% [2]. The loss to follow-up at 12 months after surgery in our study was 30%. Moreover, previous studies have implied that it is not the extent of loss to follow-up but the type of attrition that is relevant for the assessment of bias [1, 7, 25]. Classification of missing data based on Rubin´s and Little´s work differentiates between data missing at random (MAR), missing completely at random (MCAR), and missing at non-random (MNAR) [26]. In cases of MAR, the non-respondents and respondents differ at baseline but report similar clinical outcomes after treatment; in cases of MCAR, the groups are similar at baseline and report similar outcomes; in cases of MNAR, the two groups compared report different outcomes. The largest risk of bias in a registry setting arises in cases of MNAR—the results are based on respondents only [1]. The use of multiple imputations and mixed linear models are used to manage MNAR [25]. Parai et al. found the loss to follow-up in the Swedish spine registry to be of the MNAR type [6], while Solberg et al. and Højmark et al. found MAR as the mechanism of loss to follow-up in the Norwegian and Danish spine registries [4, 5]. In our study, data seem to be missing at random since baseline characteristics differ somewhat between non-respondents and respondents, but the two groups report similar outcomes.

The methods used by registries to collect data may influence patient response. Reasons for patients not responding can be related to forgetfulness, lack of interest, and questionnaires being too time demanding. Clinical visits and telephone interviews have been shown to increase response rates [5], but they are time-consuming, costly, and not practical in a register setting. A web-based registry has shown a high loss to follow-up (59%) [17]. A combination of postal and web-based methods could complement each other and increase response rates. NORspine plans to implement a combination of methods to increase the follow-up rate.

Strengths and limitations

The main weaknesses of our study are that we reached out to a sample of all potential register patients and that responsive non-respondents were compared to respondents at different time points, i.e., 12 months vs. 50 (36–64) months after surgery. However, previously published data have shown that patients who are followed longer than one year after spinal surgery keep reporting stable symptoms [27].

Conclusion

In this observational study based on data from a national spine registry, we found a 30% loss to follow-up at 12 months after surgery for lumbar spinal stenosis. We reached out to non-respondents after surgery and found that non-respondents were somewhat younger and more frequently smokers. However, non-respondents reported similar clinical outcomes compared to those who responded. Our findings suggest that attrition bias in NORspine was random and due to non-modifiable factors.

Abbreviations

NORspine	Norwegian registry for spine surgery
ODI	Oswestry Disability Index
NRS	Numerical rating scale
PROM	Patient-reported outcome measure
GPE	Global Perceived Effect
EQ-5D	European quality of life 5-dimension questionnaire
SF-36	Short form- 36 health survey
MAR	Missing at random
MNAR	Missing not at random
MCAR	Missing completely at random

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12891-023-06267-3.

Additional file 1.

Acknowledgements

The authors would like to thank the Norwegian Registry for Spine Surgery (NORspine).

Authors' contributions

All authors read and approved the manuscript. GL is the guarantor and gave the original concept of the study. GL and OKA were involved in the study design. OKA supervised the study, was involved in statistics, and took part in the writing the manuscript. SK was involved in collecting of the data and statistics, and writing of the manuscript. FD and TS contributed to the study design and writing of the manuscript.

Funding

None.

Availability of data and materials

The datasets generated and analysed during the current study are not publicly available due to the Norwegian data protection law but are available from the corresponding author at reasonable request.

Declarations

Ethics approval and consent to participate

The Norwegian national ethical board (Regional committees for medical and health research ethics, reference number 2017/2157) approved this study, as did the data protection officers at the four participating hospitals.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Received: 29 August 2022 Accepted: 23 February 2023
Published online: 04 March 2023

References

1. Hooff MLv, Jacobs WCH, Willems PC, Wouters MWJM, Kleuver Md, Peul WC, et al. Evidence and practice in spine registries. Acta Orthopaedica. 2015;86(5):534–44.
2. Endler P, Ekman P, Hellström F, Möller H, Gerdhem P. Minor effect of loss to follow-up on outcome interpretation in the Swedish spine register. Eur Spine J. 2020;29(2):213–20.
3. Wang K, Eftang CN, Jakobsen RB, Aroen A. Review of response rates over time in registry-based studies using patient-reported outcome measures. BMJ Open. 2020;10(8): e030808.
4. Højmark K, Støttrup C, Carreon L, Andersen MJESJ. Patient-reported outcome measures unbiased by loss of follow-up. Single-center study based on DaneSpine, the Danish spine surgery registry. 2016;25(1):282–6.
5. Solberg TK, Sørlie A, Sjaavik K, Nygaard ØP, Ingebrigtsen T. Would loss to follow-up bias the outcome evaluation of patients operated for degenerative disorders of the lumbar spine? Acta Orthop. 2011;82(1):56–63.
6. Paral C, Hägg O, Willers C, Lind B, Brisby H. Characteristics and predicted outcome of patients lost to follow-up after degenerative lumbar spine surgery. Eur Spine J. 2020;29(12):3063–73.
7. Kristman V, Manno M, Cote P. Loss to follow-up in cohort studies: how much is too much? Eur J Epidemiol. 2004;19(8):751–60.
8. Graham JW. Missing data analysis: making it work in the real world. Annu Rev Psychol. 2009;60:549–76.
9. Solberg T. Norwegian Registry for Spine Surgery (NORspine). Annual Report 2020. 2020. Available from: [https://www.kvalitetsregistre.no/sites/default/files/2021-06/Ryggkirurgi%20Årsrapport%202020.pdf].
10. Fairbank JC, Pynsent PB. The Oswestry Disability Index. Spine (Phila Pa 1976). Spine. 2000;25(22):2940–52 discussion 52.
11. Hawker GA, Mian S, Kendzerska T, French M. Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP). Arthritis Care Res (Hoboken). 2011;63(Suppl 11):S240–52.
12. EuroQol Research Foundation. EQ-5D-5L User Guide. 2019.
13. Grotle M, Brox JI, Vollestad NK. Cross-cultural adaptation of the Norwegian versions of the Roland-Morris Disability Questionnaire and the Oswestry Disability Index. J Rehabil Med. 2003;35(5):241–7.
14. Kamper SJ, Ostelo RW, Knol DL, Maher CG, de Vet HC, Hancock MJ. 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):760-6 e1.
15. World Medical Association Declaration of Helsinki. ethical principles for medical research involving human subjects. JAMA. 2000;284(23):3043–5.
16. Randsborg PH, Adamec D, Cepeda NA, Pearle A, Ranawat A. Differences in Baseline Characteristics and Outcome Among Responders, Late Responders, and Never-Responders After Anterior Cruciate Ligament Reconstruction. Am J Sports Med. 2021;49(14):3809–15.
17. Schröder ML, de Wispelaere MP, Staartjes VE. Predictors of loss of follow-up in a prospective registry: which patients drop out 12 months after lumbar spine surgery? The Spine Journal. 2019;19(10):1672–9.
18. Elkan P, Lagerbäck T, Möller H, Gerdhem P. Response rate does not affect patient-reported outcome after lumbar discectomy. Eur Spine J. 2018;27(7):1538–46.
19. Bisson EF, Mummaneni PV, Knightly J, Alvi MA, Goyal A, Chan AK, Guan J, Blase M, Strauss A, Glassman S, Foley K, Slotkin JR, Potts E, Shaffrey M, Shaffrey CI, Haid RW, Fu KM, Wang MY, Park P, Asher AL, Bydon M. Assessing the differences in characteristics of patients lost to follow-up at 2 years: results from the Quality Outcomes Database study on outcomes of surgery for grade I spondylolisthesis. J Neurosurg Spine. 2020:1-9. https://doi.org/10.3171/2019.12.SPINE191155. Epub ahead of print.
20. Alhaug OK, Kaur S, Dolatowski F, Smastuen MC, Solberg TK, Lonne G. Accuracy and agreement of national spine register data for 474 patients compared to corresponding electronic patient records. European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society. 2022;31(3):801–11.
21. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials. 1989;10(4):407-15. https://doi.org/10.1016/0197-2456(89)90005-6.
22. Haynes RB, Sackett DL, Richardson WS, Rosenberg W, Langley GR. Evidence-based medicine: How to practice & teach EBM. Can Med Assoc J. 1997;157(6):788.
23. Dettori JR. Loss to follow-up. Evid Based Spine Care J. 2011;2(1):7–10.
24. Bhandari M, Guyatt GH, Swiontkowski MF. User's Guide to the Orthopaedic Literature: How to Use an Article About a Surgical Therapy. JBJS. 2001;83(6):916–26.
25. Twisk J, de Vente W. Attrition in longitudinal studies. How to deal with missing data. J Clin Epidemiol. 2002;55(4):329-37. https://doi.org/10.1016/s0895-4356(01)00476-0.
26. Little RJ, Rubin DB. Statistical Analysis with Missing Data (Vol. 793). Hoboken: Wiley; 2019. https://doi.org/10.1002/9781119482260.
27. Paral C, Hägg O, Lind B, Brisby H. Follow-up of degenerative lumbar spine surgery—PROMs stabilize after 1 year: an equivalence study based on Swespine data. Eur Spine J. 2019;28(9):2187–97.

Publisher's Note

12.3 Paper 3

ELSEVIER

The SPINE JOURNAL

The Spine Journal 21 (2021) 1489–1496

Clinical Study

Criteria for failure and worsening after surgery for lumbar spinal stenosis: a prospective national spine registry observational study

Ole Kristian Alhaug, MD[a,b,c,*], Filip C. Dolatowski, MD, PhD[c], Tore K. Solberg, MD, PhD[d], Greger Lønne, MD, PhD[a]

[a] *Innlandet Hospital Trust, The Research Center for Age-Related Functional Decline and Disease, PO Box 68, N-2313, Ottestad, Norway*

[b] *Orthopedic department, Akershus University Hospital, PO Box 1000, N-1478, Loerenskog, Norway*

[c] *Orthopedic department, Oslo University Hospital, PO Box 4956, N-0424, Oslo, Norway*

[d] *Neurosurgical department, University hospital of North Norway, N-9038, Tromsoe, Norway*

[e] *Norwegian University of Science and Technology, NTNU PO Box 191, N-7491 Trondheim, Norway*

Received 31 August 2020; revised 25 March 2021; accepted 6 April 2021

Abstract

BACKGROUND CONTEXT: Criteria for success after surgical treatment of lumbar spinal stenosis (LSS) have been defined previously; however, there are no clear criteria for failure and worsening after surgery as assessed by patient-reported outcome measures (PROMs).

PURPOSE: We aimed to quantify changes in standard PROMs that most accurately identified failure and worsening after surgery for LSS.

STUDY DESIGN /SETTING: Retrospective analysis of prospective national spine registry data with 12-months follow-up.

PATIENT SAMPLE: We analyzed 10,822 patients aged 50 years and older operated in Norway during a decade, and 8,258 (76%) responded 12 months after surgery.

OUTCOME MEASURES (PROMS): We calculated final scores, absolute changes, and percentage changes for Oswestry Disability Index (ODI), Numeric Rating Scale (NRS) for back and leg pain (0-10), and EuroQol-5D (EQ-5D). These 12 PROM derivates were compared to the Global Perceived Effect (GPE), a 7-point Likert scale.

METHODS: We used ODI, NRS back and leg pain, and EQ-5D 12 months after surgery to identify patients with failure (no effect) and worsening (clinical deterioration). The corresponding GPE at 12-months was graded as failure (GPE=4-7) and worsening (GPE=6-7) and used as an external criterion. To quantify the most accurate cut-off values corresponding to failure and worsening, we calculated areas under the curves (AUCs) of receiver operating characteristics (ROC) curves for the respective PROM derivates.

RESULTS: Mean (95% CI) age was 68.3 (68.1 − 68.5) years, and 52% were females. There were 1,683 (20%) failures, and 476 (6%) patients were worse after surgery. The mean (95% CI) pre- and postoperative ODIs were 39.8 (39.5 − 40.2) and 23.7 (23.3 − 24.1), respectively. At 12 months, the mean difference (95% CI) in ODI was 16.1 (15.7 − 16.4), and the mean (95% CI) percentage improvement 38.8% (37.8 − 38.8).

The PROM derivates identified failure and worsening accurately (AUC>0.80), except for the absolute change in EQ-5D. The ODI derivates were most accurate to identify both failure and worsening. We found that less than 20% improvement in ODI most accurately identified failure (AUC=0.89 [95% CI: 0.88 to 0.90]), and an ODI final score of 39 points or more most accurately identified worsening (AUC =0.91 [95% CI: 0.90 − 0.92]).

FDA. Device/drug status: Not applicable

Author disclosure: ***OKA:*** Nothing to disclose. ***FCD:*** Nothing to disclose. ***TKS:*** Nothing to disclose. ***GL:*** Nothing to disclose.

*Corresponding author: Sykehuset Innlandet, The Research Center for Age-Related Functional Decline and Disease, PO Box 68, N-2313 Ottestad, Norway, Tel.: +(47) 41127443

E-mail address: ole.kristian.alhaug@sykehuset-innlandet.no (O.K. Alhaug).

https://doi.org/10.1016/j.spinee.2021.04.008

CONCLUSIONS: In this national register study, ODI derivates were most accurate to identify both failure and worsening after surgery for degenerative lumbar spinal stenosis. We recommend use of ODI percentage change and ODI final score for further studies of failure and worsening in elective spine surgery.

Keywords: Spinal Stenosis; Spine registry; Failure; Worsening; Cut-off; PROM

Introduction

Patients operated for lumbar spinal stenosis (LSS) are more likely to improve than those treated conservatively [1-3]. However, about 20% report persisting back and leg pain after surgery [4].

Success after surgical treatment of degenerative LSS has previously been defined as a substantial clinical improvement ("completely recovered" or "much improved") [4]. In contrast, there are no clear definitions of failure and worsening after surgery. Failure can be defined as unchanged or worsening of symptoms and worsening as a clear deterioration of symptoms after treatment [5]. The term "non-success" includes a small improvement and cannot be classified as neither failure nor worsening. Hence "non-success" and failure are different concepts.

Patients may accept a lack of improvement after surgery, but worsening, indicating a potentially harmful treatment effect, is not well tolerated [5]. Therefore, it is important to distinguish between these concepts and to define specific cut-off criteria for both failure and worsening for common patient-reported outcome measures (PROMs). Such criteria could be used in patient selection [6] and further research.

In this national spine registry study, we aimed to define changes in Oswestry Disability Index (ODI) [7], numeric rating scales (NRS) for back and leg pain, and quality of life (EQ-5D index) that most accurately described failure and worsening after operative treatment for LSS.

Method

We conducted a retrospective observational study using prospectively collected data from the Norwegian national spine registry (NORspine). We report data according to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) recommendations [8].

Patient population

Eligible were 10,822 patients reported to NORspine, aged 50 years or older, operated for Lumbar Spinal Stenosis in Norway between January 1, 2007, and April 1 , 2017 (Fig. 1).

The NORspine registry

All private and public hospitals that perform spine surgery in Norway (100 %) report to NORspine, which is a comprehensive clinical registry, currently covering 70 % of all operations for degenerative spine done in Norway [9]. Patients unable to give informed consent, with severe psychiatric diagnoses, or drug problems, as well as patients treated for spinal tumors, fractures, or primary infections, are not included in NORspine.

At admission for surgery (baseline), patients signed an informed consent and completed a questionnaire that included PROMs and questions about the duration of leg and back pain, socio-demographics, and lifestyle issues. The surgeon recorded information about the diagnosis, indication for surgery (radiologic findings and symptoms), comorbidity, treatment, and perioperative complications on a standardized form. At 3 and 12 months after the operation, the patient completed follow-up questionnaires, including repetitive PROMs. Patients received and returned the 3- and 12-month follow-up questionnaires directly to NORspine by mail without the treating hospital's involvement. Non-responders got one reminder questionnaire by mail.

Patient-reported outcome measures (PROMs) and reference

Oswestry Disability Index (ODI) is a validated measure of back pain-related disability [7]. It consists of ten questions related to activities of daily living, each with five response

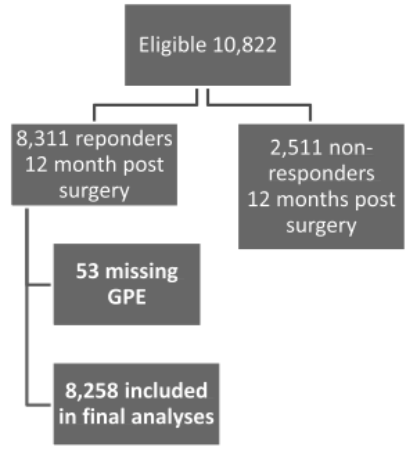


Fig. 1. Flowchart showing eligible patients, responders, non-responders, and those that could not be analyzed due to missing Global Perceived Effect (GPE) score. 8,258 patients were available for final analyses.

alternatives (0–5), which are summarized into a percentage score ranging from 0 (minimal disability) to 100 (bed bound).

The Numeric Rating Scales for back and leg pain range from 0 (no pain) to 10 (worst imaginable pain). NRS is easy to use, correlates well with other pain measuring tools, and is recommended for measuring chronic pain [10,11].

EuroQol-5-Dimension-3-Level (EQ-5D) is a validated non-disease specific health-related quality of life measure. Patients report five dimensions: mobility, self-care, the activity of daily living, pain, and anxiety/depression. Each dimension is graded by three levels (no, moderate, or severe problems). The index score varies between minus 0,59 to 1, 0 ("worse than dead" to "perfect health" [12-14].

At 12-month follow-up, patients also rated their perceived effect of surgery by a Global Perceived Effect scale (GPE) [15]. We used GPE as a reference to study PROMs mentioned above. The seven response alternatives were: 1= completely recovered, 2= much better, 3= somewhat better, 4= unchanged, 5= somewhat worse, 6= much worse, and 7= worse than ever. We graded patients who perceived themselves as unchanged or any degree of worsening (GPE 4-7) as "failures." Patients who perceived themselves as "much worse" or "worse than ever" (GPE 6 and 7) were grades as "worsening."

We calculated three different derivates for each of the PROMs; final score (12 months after surgery), the absolute change, and the percentage change. We assessed the accuracy of these 12 PROM derivates to identify failure and worsening, using the GPE as an external criterion, as explained above [16-18].

Statistical analyses

We analyzed differences within or between groups with student T-test for continuous data (reported as mean, 95% confidence interval (CI), and mean difference). We used relative risk (RR) with 95% CI and z-statistics when comparing categorical data.

We used Receiver Operating Characteristics (ROC) curves for each PROM outcome to identify cut-off values for failure (GPE = 4–7) and worsening (GPE = 6–7) after LSS surgery. We used the closest point to the upper left corner of the ROC curve (Fig. 2) to determine the cut-off with the highest sensitivity and specificity. We calculated the areas under the respective curves (AUC) to determine how accurate the PROM derivates classified the outcomes as failure vs. non-failure and worsening vs. non-worsening. AUC values and corresponding grades of accuracy were interpreted as follows: < 0.7 = poor, 0.7 - 0.8 = fair, 0.8 - 0.9 = good, and ≥0.9 = excellent accuracy [19].

To evaluate the consistency of our results across subgroups, we performed ancillary analyses for age, preoperative ODI score quartiles, and type of surgery (decompression vs. decompression and fusion). We performed the subgroup analysis only for the failure group, as the worsening group was considered too small. Patients with a missing variable were excluded only in the analyses for that missing variable, and we did not perform any imputation.

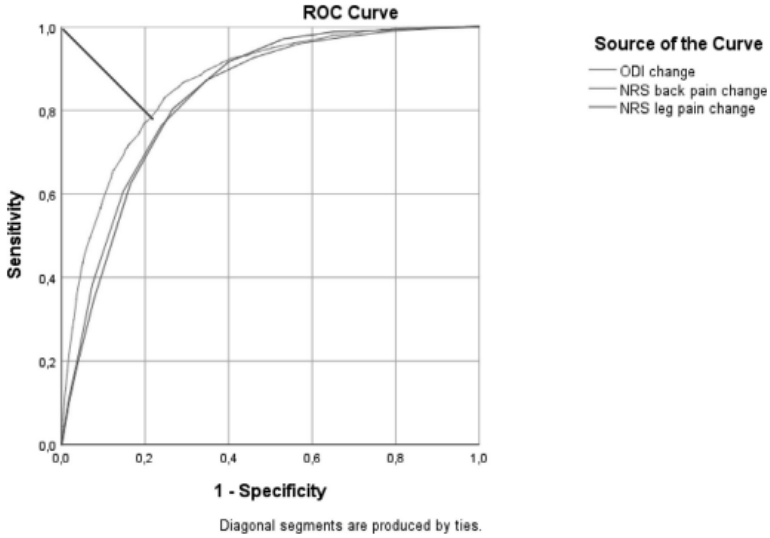


Fig. 2. ROC curves for absolute change in ODI, NRS back pain, and NRS leg pain vs. "failure" (GPE 4–7). The oblique blue line demonstrates "distance to corner" - a method to identify the highest sensitivity and specificity for each curve. The ROC curves also demonstrate "area under the curve" (AUC) - a measurement for the accuracy to identify failure.

We performed statistical analyses using SPSS version 25. (IBM Corp. released in 2017. IBM SPSS Statistics for Windows, Version 25. Armonk, NY)

Ethical considerations

All patients provided written informed consent before entering the registry. The study was approved by the Norwegian national research ethics committee (reference: 2017/2157, May 15 2018). The study was conducted in accordance with the Helsinki declaration [20].

Results

Of 10,822 patients enrolled in the registry, 8,311 (77%) responded at 12 months follow-up. Fifty-three of the responders (0.6%) did not report the GPE score. Hence, 8,258 were included in the final analyses (Fig. 1).

The overall mean age (95% CI) was 68.0 (67.8-68.1) years, and 5,690 (53%) were women. Table 1 shows patient characteristics at baseline. We found several small, but statistically significant, differences between the responders and non-responders. Non-responders were somewhat younger and more often smokers, single, and disability benefit receivers, and had more comorbidities (Table 1). Responders had less pain and disability at baseline, ODI mean difference (95% CI) was 2.7 points (2.0 to 3.4); p<.001. There were no relevant differences between the responders and the non-responders for the remaining patient characteristics and diagnoses (Appendix Table 1).

Overall, 73 % had MRI findings of central spinal stenosis, and 16% had degenerative spondylolisthesis. The main surgical techniques were foraminotomy (70%), and laminectomy (12%). Also, 80 % of the surgeries were done microscopically assisted, and 12% of the patients underwent an additional fusion procedure. Table 2 shows the effect of surgery as assessed by ODI, NRS back pain, NRS leg pain, and EQ-5D derivates: final score, absolute change, and percentage change.

At 12 months follow-up, the outcomes of 1,683 patients (20%) were classified as failures according to the GPE (GPE 4-7) and the outcomes of 476 patients (6%) as worsening (GPE 6–7) (table 3). Table 3 also shows the PROM derivates with corresponding cut-off values and accuracies to identify failure and worsening. *ODI percentage change* had the highest accuracy (AUC (95% CI) = 0.89 (0.88–0.90)), with a cut-off value of less than 20%, to identify failure after surgery (sensitivity/specificity: 82%/ 81%) at 12-months follow-up. An ODI final score of 31 points or more, and an ODI absolute change of less than 8 points, also accurately classified failure.

ODI final score showed excellent accuracy (AUC (95%CI) = 0.91 (0.90–0.92)) with a cut-off value of more than 39 points to identify worsening after surgery (sensitivity/specificity: 83%/ 79%), followed by ODI percentage change of less than 9% and an ODI absolute change (improvement) of less than 4 points.

NRS back and leg pain derivates showed good accuracy identifying both failure and worsening (AUC > 0.80). EQ-5D final score showed excellent accuracy to identify worsening (AUC=0.90), but EQ-5D absolute change showed only fair accuracy to identify failure (AUC=0.79).

Ancillary analyses (Appendix Table 2) showed that the cut-offs for PROM derivates to identify failure did not change across age quartiles (<25%, 25%-75%, and >75%).

Table 1
Patient characteristics of 10,822 Norwegian patients, 50 years and older, with surgically treated spinal stenosis (broken down by responders versus non-responders)

	Responders n = 8,311 (76,8%) Mean (95% CI) or n (%)	Non-responders n = 2,511 (23,2%) Mean (95% CI) or n (%)	Mean diff (95%CI) or Relative Risk (95% CI)	p-value
Age	68.3 (68.1 - 68.5)	66.8 (66.4 - 67.1)	1.5 (1.1 - 1.9)	<.001
Female	4,335 (52.2%)	1,355 (54.0%)	1.03 (0.99 - 1.08)	.109
Civil status - single	2,145 (25.9%)	783 (31.4%)	1.21 (1.13 - 1.30)	.001
Norwegian as 1st language	8,014 (96.9%)	2,384 (95.2%)	0.98 (0.97 - 0.99)	<.001
ASA* grade 1 and 2	6,409 (77.9%)	1,873 (75.5%)	0.97 (0.95 - 0.99)	.015
Body Mass Index	27.5 (27.4 - 27.6)	27.7 (27.5 - 27.9)	0.2 (0.0 - 0.4)	.045
Smoking	1,521 (18.5%)	664 (26.7%)	1.45 (1.34 - 1.57)	<.001
University or college education > 4 years	2,439 (29.3%)	679 (27.0%)	0.92 (0.86 - 0.99)	.023
Comorbidity, any	5,228 (69.2%)	1,692 (73.8%)	1.07 (1.04 - 1.10)	<.001
Receives Disability benefit	2,368 (28.5%)	902 (35.9%)	1.26 (1.18 - 1.34)	<.001
Previous spinal surgery, any	1,994 (24.3%)	694 (26.1%)	1.07 (0.99 - 1.16)	.070
Back pain >12 months before surgery	5,851 (70.4%)	1,833 (73.0%)	1.03 (1.01 - 1.07)	.010
Leg pain >12 months before surgery	4,950 (59.6%)	1,567 (62.4%)	1.05 (1.01 - 1.09)	.009
Pre-operative ODI**	39.8 (39.5 - 40.2)	42.6 (41.9 - 43.2)	2.7 (2.0 - 3.4)	<.001
Pre-operative NRS*** back pain	6.5 (6.5 - 6.6)	6.7 (6.6 - 6.7)	0.1 (0.0 - 0.2)	0,007
Pre-operative NRS leg pain	6.6 (6.5 - 6.6)	6.7 (6.6 - 6.8)	0.1 (0.0 - 0.2)	0,015
Pre-operative EQ-5D****	0.38 (0.37 - 0.38)	0.32 (0.31 - 0.33)	0.06 (0.04 - 0.07)	<.001

Table abbreviations explained: ASA = American Society of Anesthesiologists classification of physical status (1–5). ODI = Oswestry Disability Index (0–100). NRS = Numeric Rating Scale 0-10. EQ-5D = EuroQol 5-Dimension 3-Level.

Table 2
Effect of surgical treatment for spinal stenosis reported by 8,311 patients at 12 months follow-up.

PROM	Final score		Absolute change (improvement)		Percentage change (improvement)	
	Mean	95%CI	Mean	95%CI	Mean	95%CI
ODI	23.7	23.3 – 24.1	16.1	15.7 – 16.4	38.8%	37.8% – 38.8%
NRS back pain	3.8	3.8 – 3.9	2.7	2.6 – 2.7	37.5%	36.3% – 38.7%
NRS leg pain	3.6	3.5 – 3.6	3.0	3.0 – 3.1	41.9%	40.5% – 43.3%
EQ-5D index*	0.64	0.64 – 0.63	0.26	0.26 – 0.27	-	-

* Percentage change of the EQ-5D index is not meaningful due to a denominator between -0.6 and 1.0.

Table 3
PROM accuracy to identify failure (GPE=4–7) and worsening (GPE=6–7) 12 months after surgical treatment of spinal stenosis in 8,258 patients. An area under the curve (AUC) > 0.7 indicates acceptable sensitivity and specificity.

		Failure (GPE 4-7) n= 1683/8258 (20%)				Worsening (GPE 6-7) n= 476/8258 (6%)			
Outcomes	n	Cut-off	AUC (95% CI)	sensitivity	specificity	Cut-off	AUC (95%CI)	sensitivity	specificity
Disability									
ODI final score	8,220	31	0.87 (0.86-0.88)	0.79	0.78	39	0.91 (0.90-0.92)	0.83	0.79
ODI absolute change	8,174	-8	0.86 (0.86-0.87)	0.78	0.79	-4	0.86 (0.85-0.88)	0.77	0.79
ODI percentage change	8,161	-20%	0.89 (0.88-0.90)	0.82	0.81	-9%	0.87 (0.86-0.88)	0.80	0.80
Back Pain									
NRS back pain final score	8,174	5,5	0.87 (0.86-0.88)	0.79	0.81	6.5	0.90 (0.89-0.91)	0.86	0.82
NRS back pain absolute change	7,687	-1.5	0.83 (0.82-0.84)	0,80	0,74	-0.5	0.83 (0.81-0.85)	0,78	0,77
NRS back pain percentage change	7,573	-21%	0.85 (0.84-0.86)	0,81	0,77	-12%	0.84 (0.82-0.85)	0,82	0,77
Leg Pain									
NRS leg pain final score	8,067	5.5	0.85 (0.84-0.86)	0,73	0.82	6.5	0.87 (0.86-0.89)	0,77	0,82
NRS leg pain absolute change	7,518	-1.5	0.83 (0.82-0.84)	0,77	0,76	-0.5	0.82 (0.81-0.84)	0,72	0,79
NRS leg pain percentage change	7,398	-24%	0.85 (0.84-0.86)	0,79	0,78	-13%	0.83 (0.82-0.85)	0,80	0,73
Quality of Life*									
EQ-5D final score	7,098	0.62	0.86 (0.85-0.87)	0,77	0,77	0.53	0.90 (0.89-0.92)	0,86	0,81
EQ-5D absolute change	6,585	0.06	0.79 (0.78-0.81)	0,71	0,76	0.03	0.81 (0.79-0.83)	0,78	0,74

The *final score* was the absolute value at 12 months follow up. The *absolute change* was the final score minus the preoperative score (negative values indicate improvement in ODI and NRS; positive values indicate improvement in EQ-5D). The *percentage change* was the absolute change divided by the preoperative score (negative values indicate improvement in ODI and NRS; positive values indicate improvement in EQ-5D).

* EQ-5D percentage change is not meaningful due to a denominator between -0.6 and 1.0.

However, the cut-offs varied between the quartiles of baseline ODI scores. For the highest and lowest preoperative ODI quartiles, an ODI final score of 46 and 19 points, respectively, indicated failure. The ODI absolute change and ODI percentage change also displayed considerable differences across the highest and lowest quartiles of baseline ODI score (15 points vs. 2 points and 25% vs. 10%, respectively). At 12 months follow-up, there were no relevant differences in follow-up rates (76.7% vs. 77.6%) and cut-off values defining failure, comparing those who underwent decompression vs. those who underwent decompression and fusion (ODI final scores of 31 vs. 32, absolute ODI changes of -8 vs. -9, and ODI percentage changes of -20% vs. -24%).

Discussion

In this national spine registry study of Norwegian patients aged 50 years and older, derivates of Oswestry Disability Index were the most accurate tools to identify both failure and worsening after surgery for degenerative lumbar spinal stenosis. The patients reported a clinically relevant improvement in ODI, NRS back and leg, and quality of life (EQ-5D) 12 months after surgery. Of the different ODI derivates, a post-operative ODI percentage change of less than 20% (improvement) most accurately classified outcome as failure. ODI final score was the most accurate derivate to identify worsening, with a cut-off at 39 points.

The NRS back and leg pain derivates also displayed good accuracy in identifying failure and worsening after surgery, albeit with lower AUCs, sensitivities, and specificities than the ODI derivates (Table 3). These findings are in line with previously published data on disc herniations [5].

Surprisingly, EQ-5D final score also displayed excellent accuracy for the classification of worsening (Table 3). The other EQ-5D derivates showed lower accuracy. EQ-5D is a generic instrument designed to assess cost-benefit rather than the clinical effect of treatment [21].

Previously published data have estimated the minimal clinically important change (MCIC) for ODI between 8 to 20 points [7, 10, 22-24]. Nerland et al. defined worsening as an 8-point increase on the ODI scale [25]. We found that even patients with a minor ODI *improvement* (cut-off at 4 points, Table 3), can perceive the result as a worsening after

surgery. This result may be explained by patients being exhausted due to severe disability and persisting symptoms, and due to recall bias when patients report GPE 12 months after surgery [18].

The concept of a patient acceptable symptom status (PASS) was developed by Van Hoof et al. in 2016 [26]. They estimated a PASS for ODI at 22 points final score after surgery for degenerative spinal disorders. Austevoll et al found a cut-off for success after surgery for spinal stenosis at 24 points for ODI final score [4]. As expected, these values are lower than the 31 points we found for cut-off for failure, emphasizing that non-success, or not reaching PASS, is not the same as failure. This means that there is a grey zone of outcomes between thresholds for non-success and failure that are difficult to classify [5].

We defined failure with reference to GPE categories of "unchanged" and any degree of worsening after surgery. In this cohort, 20 % of patients classified their outcomes after surgery as failure, and 6 % were worse after surgery. Nerland found that 8.7 % of LSS patients reported worsening after decompression [25]. Previously published data from the SPORT study described a success rate of 65% after surgical decompression of LSS [27]; however, neither failure nor worsening were explicitly reported.

In a shared decision-making process before surgery, the surgeon should inform the patient about the risk of failure and worsening. A dichotomous outcome may be understood more easily than a PROM number and can be used to estimate the risk of failure. However, the most intuitive PROM derivate is probably the ODI final score, because it indicates if a patient has reached an unfavorable outcome or not.

We expected that cut-off values could vary by age groups with different expectations and demands concerning physical performance. However, our findings do not support this hypothesis (Appendix Table 2). Furthermore, our ancillary analyses indicated that ODI cut-off values varied with preoperative ODI levels, but not with the type of surgery (decompression vs. decompression and fusion). Hence, analyses of failure and worsening should be performed with adjustment for the baseline values of the PROMs investigated.

This nation-wide observational spine registry study is based on prospectively collected data reported by thousands of patients operated in many hospitals, indicating that data are robust with high external validity. Previous studies have shown that the indications for surgery in the Scandinavian LSS population are similar to those used in the US, although the surgical techniques may differ between countries [28,29]. Furthermore, the patient-reported outcomes after surgery are similar to the results reported in previous studies [27,28].

Limitations are that NORspine covers about 70% of all surgeries done in Norway [9], and our loss to follow-up was 23 % at 12 months. Baseline characteristics of responders and non-responders displayed some statistically significant differences and could indicate that non-responders would be at higher risk for inferior outcomes [25,30]. This could represent a selection bias when evaluating treatment effects. However, the main purpose of this study was rather to evaluate cut-offs for four common PROMs used to assess the effect of spinal surgery. Moreover, previous cohort studies reported similar clinical outcomes for non-responders compared to responders and lost-to-follow-up rates of 12% to 42%. [31,32,33,34]. The authors of one systematic review of spine register data recommended a follow-up rate of 60% −80% to ensure sufficient quality in spine registers [35].

Another possible limitation of the NOR. spine spine register is that it does not extend beyond 12 months follow-up. However, several studies have shown that the effect of surgical treatment of the degenerative spine stabilizes after 12 months [27,34,36-38].

Selecting the GPE as an external criterion may have weaknesses due to lack of objectivity and potentially a recall bias [18,39]. However, GPE has been recognized as an acceptable tool to measure the effect of lumbar degenerative spinal surgery [40] and is a recommended clinical anchor [41].

Finally, our patients underwent different surgical procedures. Still, we found no relevant differences in follow-up rates and ODI cut-off values defining failure, comparing those who underwent decompression vs. those who underwent decompression and fusion.

Despite these limitations, the authors believe that our cut-off criteria for failure and worsening may facilitate clinical guidance and be used as a common language in future research on the effects of surgical treatment of the degenerative spine.

Conclusions

In this national spine registry study, ODI derivates were most accurate to identify both failure and worsening after surgery for degenerative lumbar spinal stenosis. We found that less than 20% improvement in ODI most accurately identified failure and that an ODI final score of 39 points, or more, most accurately identified worsening. We recommend using ODI percentage change and ODI final score for further studies of failure and worsening after spine surgery.

Acknowledgements

We would like to thank Milada Cvancarova Småstuen, Associate Professor at Oslo Metropolitan University, for her invaluable help with statistics and proofreading of the manuscript draft.

The funding for this study was granted by Innlandet Hospital Trust to the corresponding author. None of the authors declare any potential conflicts of interest.

Declarations of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix Table 1
Diagnosis and treatment for 10,822 Norwegian patients, 50 years and older, with surgically treated spinal stenosis (broken down by responders versus non-responders)

	Responders n = 8,311	Non-responders n = 2,511	Relative Risk (95% CI)	p-value
Diagnoses				
Central spinal stenosis	6,081 (73.2%)	1,825 (72.7%)	0.99 (0.97-1.02)	.631
Lateral spinal stenosis	4,569 (55.0%)	1,428 (56.9%)	1.03 (0.99-1.08)	.091
Disc herniation	460 (5.5%)	140 (5.6%)	1.01 (0.84-1.21)	.938
Degenerative disk	1,304 (15.7%)	426 (17.0%)	1.08 (0.98-1.19)	.125
Foraminal stenosis	874 (10.5%)	260 (10.4%)	0.98 (0.86-1.12)	.817
Spondylolisthesis, degenerative	1,334 (16.1%)	437 (17.4%)	1.08 (0.98-1.20)	.107
Synovial cyst	199 (2.4%)	61 (2.4%)	1.01 (0.76-1.35)	.920
Degenerative scoliosis	354 (4.3%)	106 (4.2%)	0.99 (0.80-1.23)	.934
Pseudomeningocele	0 (0%)	0 (0%)	-	-
Spondylolysis	0 (0%)	0 (0%)	-	-
Treatment				
Decompression with microscope	6,577 (79.1%)	2,102 (83.7%)	1.05 (1.04-1.08)	<.001
Decompression with fusion	1,022 (12.3%)	295 (11.7%)	0.96 (0.85-1.07)	.462
Complication(s) peri-operatively	486 (5.8%)	171 (6.8%)	1.16 (0.98-1.38)	.076
Operated > 1 level	3,129 (38.0%)	888 (35.7%)	0.94 (0.88-1.00)	.040

Appendix Table 2
ODI cut-off values with corresponding AUCs indicating the highest accuracy to identify failure, broken down by quartiles of pre-operative ODI and age.

Subgroups	ODI final score (AUC)	ODI absolute change (AUC)	ODI % change (AUC)
Pre-op ODI >51,1	46 (0.89)	-16 (0.89)	-27% (0.89)
Pre-op ODI 28,9-51,1	32 (0.90)	-8 (0.90)	-20% (0.90)
Pre-op ODI <28,9	19 (0.87)	-2 (0.90)	-9% (0.90)
Age ≥74	31 (0.84)	-8 (0.84)	-20% (0.87)
Age 62 – 73	29 (0.88)	-8 (0.88)	-21% (0.90)
Age ≤61	29 (0.88)	-6 (0.88)	-17% (0.90)

References

[1] Levin K. Lumbar spinal stenosis: Treatment and prognosis. . UpToDateThis topic last updated: Dec 12, 2018 https://www.uptodate.com/contents/lumbar-spinal-stenosis-treatment-and-prognosis?source=related_link. Accessed August 17, 2020.

[2] Weinstein JN, Tosteson TD, Lurie JD, Tosteson ANA, Blood E, Hanscom B, et al. Surgical versus nonsurgical therapy for lumbar spinal stenosis. N Engl J Med 2008;358(8):794–810. https://doi.org/10.1056/NEJMoa0707136.

[3] Atlas SJ, Delitto A. Spinal stenosis: surgical versus nonsurgical treatment. Clin Orthop Relat Res 2006;443:198–207. https://doi.org/10.1097/01.blo.0000198722.70138.96.

[4] Austevoll IM, Gjestad R, Grotle M, Solberg T, Brox JI, Hermansen E, 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;20:31. https://doi.org/10.1186/s12891-018-2386-y.

[5] Werner DAT, Grotle M, Gulati S, Austevoll IM, Lønne, Nygaard ØP, et al. Criteria for failure and worsening after surgery for lumbar disc herniation: a multicenter observational study based on data from the Norwegian Registry for Spine Surgery. Eur Spine J 2017;26(10):2650–9. https://doi.org/10.1007/s00586-017-5185-5.

[6] Deyo RA, Mirza SK, Turner JA, Martin BI. Overtreating chronic back pain: time to back off? J Am Board Fam Med 2009;22(1):62–8. https://doi.org/10.3122/jabfm.2009.01.080102.

[7] Fairbank JC, Pynsent PB. The oswestry disability index. Spine (Phila Pa 1976) 2000;25(22):2940–52. https://doi.org/10.1097/00007632-200011150-00017.

[8] von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol 2008;61(4):344–9. https://doi.org/10.1016/j.jclinepi.2007.11.008.

[9] Norwegian Registry for Spine Surgery (NORspine). Annual report 2017. Tore K Solberg, Lena Ringstad Olsen,Universitetssykehuset Nord Norge (UNN) 2SKDE; 2018. October 23.th https://www.kvalitetsregistre.no/sites/default/files/30_arsrapport_2017_ryggkirurgi_1.pdf. Accessed August 17, 2020.

[10] Copay A, Martin M, Subach B, Carreon LY, Glassman SD, Schuler TC, et al. Assessment of spine surgery outcomes: inconsistency of change amongst outcome measurements. The Spine Journal 2010;10(4):291–6. Iss.

[11] Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain 2005;113(1–2):9–19. https://doi.org/10.1016/j.pain.2004.09.012.

[12] EuroQolGroup. EuroQol—a new facility for the measurement of health-related quality of life. Health Policy 1990;16(3):199–208. https://doi.org/10.1016/0168-8510(90)90421-9.

[13] Dolan P, Gudex C, Kind P, Williams A. The time trade-off method: results from a general population study. Health Econ 1996;5(2):141–54. https://doi.org/10.1002/(SICI)1099-1050(199603)5:2<141::AID-HEC189>3.0.CO;2-N.

[14] Solberg TK, Olsen JA, Ingebrigtsen T, Hofoss D, Nygaard OP. Health-related quality of life assessment by the EuroQol-5D can provide cost-utility data in the field of low-back surgery. Eur Spine J 2005;14(10):1000–7. https://doi.org/10.1007/s00586-005-0898-2.

[15] Kamper SJ, Ostelo RW, Knol DL, Maher CG, de Vet HC, Hancock MJ. 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):760–6 e1. https://doi.org/10.1016/j.jclinepi.2009.09.009.

[16] Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. J Clin Epidemiol 2003;56 (5):395–407. https://doi.org/10.1016/s0895-4356(03)00044-1.

[17] Clinical Significance Consensus Meeting Group, Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR. Methods to explain the clinical significance of health status measures. Mayo Clin Proc 2002;77 (4):371–83. https://doi.org/10.4065/77.4.371.

[18] Grøvle L, Haugen AJ, Hasvik E, Natvig B, Brox JI, Grotle M, 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):508–15. https://doi.org/10.1016/j.jclinepi.2013.12.001.

[19] Tape TG, University of Nebraska Medical Cenrter. Interpreting diagnostic tests. http://gim.unmc.edu/dxtests/ROC3.htm. (accessed 17. August 2020)

20] World Medical Association. World medical association declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2013;310(20):2191–4. https://doi.org/10.1001/jama.2013.281053.

[21] Ranstam J, Robertssen O, W-Dahl A. EQ-5D—a difficult-to-interpret tool for clinical improvement work. Lakartidningen 2011;108 (36):1707–8.

[22] Solberg T, Johnsen LG, Nygaard ØP, 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(2):196–201. https://doi.org/10.3109/17453674.2013.786634.

[23] Ostelo RW, Deyo RA, Stratford P, Waddell G, Croft P, Von Kroff M, 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):90–4. https://doi.org/10.1097/BRS.0b013e31815e3a10.

[24] Glassman SD, Copay AG, Berven SH, Polly DW, Subach BR, Carreon LY. Defining substantial clinical benefit following lumbar spine arthrodesis. J Bone Joint Surg Am 2008;90(9):1839–47. https://doi.org/10.2106/JBJS.G.01095.

[25] Nerland U, Jakola A, Giannadakis C, Solheim O, Weber C, Nygaard ØP, et al. The risk of getting worse: predictors of deterioration after decompressive surgery for lumbar spinal stenosis: a multicenter observational study. World Neurosurg 2015;84(4):1095–102.

[26] 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(10):1221–30. https://doi.org/10.1016/j.spinee.2016.06.010.

[27] Weinstein JN, Tosteson TD, Lurie JD, Tosteson A, Blood E, Herkowitz H, et al. Surgical versus nonoperative treatment for lumbar spinal stenosis four-year results of the Spine Patient Outcomes Research Trial. Spine (Phila Pa 1976) 2010;35(14):1329–38. https://doi.org/10.1097/BRS.0b013e3181e0f04d.

[28] Lønne G, Fritzell P, Hägg O, Nordvall D, Gerdheim P, Lagerback T, et al. Lumbar spinal stenosis: comparison of surgical practice variation and clinical outcome in three national spine registries. Spine J 2019;19(1):41–9. https://doi.org/10.1016/j.spinee.2018.05.028.

[29] Lønne G, Schoenfeld AJ, Cha TD, Nygaard ØP, Zwart JAH, Solberg T, et al. Variation in selection criteria and approaches to surgery for lumbar spinal stenosis among patients treated in Boston and Norway. Clinl Neurol Neurosurg 2017;156:77–82. https://doi.org/10.1016/j.clineuro.2017.03.008.

[30] Aalto TJ, Malmivaara A, Kovacs F, Herno A, Alen M, Salmi L, et al. Preoperative predictors for postoperative clinical outcome in lumbar spinal stenosis: systematic review. Spine (Phila Pa 1976) 2006;31 (18):E648–63. https://doi.org/10.1097/01.brs.0000231727.88477.da.

[31] Solberg TK, Sørlie A, Sjaavik K, Nygaard ØP, Ingebrigtsen T. Would loss to follow-up bias the outcome evaluation of patients operated for degenerative disorders of the lumbar spine? Acta Orthop 2011;82 (1):56–63. https://doi.org/10.3109/17453674.2010.548024.

[32] Højmark K, Støttrup C, Carreon L, Andersen MO. Patient-reported outcome measures unbiased by loss of follow-up. single-center study based on danespine, the danish spine surgery registry. Eur Spine J 2016;25(1):282–6. https://doi.org/10.1007/s00586-015-4127-3.

[33] Elkan P, Lagerbäck T, Möller H, Gerdhem P, et al. Response rate does not affect patient-reported outcome after lumbar discectomy. Eur Spine J 2018;27:1538–46. https://doi.org/10.1007/s00586-018-5541-0.

[34] Endler P, Ekman P, Hellström F, Møller H, Gerdhem P, et al. Minor effect of loss to follow-up on outcome interpretation in the Swedish spine register. Eur Spine J 2020;29:213–20. https://doi.org/10.1007/s00586-019-06181-0.

[35] van Hooff ML, Jacobs WCH, Willems PC, Wouters MWJM, de Kleuver M, Peul WC, et al. Evidence and practice in spine registries. Acta Orthop 2015;86(5):534–44. https://doi.org/10.3109/17453674.2015.1043174.

[36] Weinstein JN, Lurie JD, Tosteson TD, Zhao W, Blood EA, Tosteson ANA, 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):1295–304. https://doi.org/10.2106/JBJS.H.00913.

[37] Parai C, Hägg O, Lind B, Brisby H. Follow-up of degenerative lumbar spine surgery-PROMs stabilize after 1 year: an equivalence study based on Swespine data. Eur Spine J 2019;28(9):2187–97. https://doi.org/10.1007/s00586-019-05989-0.

[38] Lønne G, Johnsen LG, Rossvoll I, Andresen H, Storheim K, Zwart JA, et al. Minimally invasive decompression versus x-stop in lumbar spinal stenosis: a randomized controlled multicenter study. Spine (Phila Pa 1976) 2015;40(2):77–85. https://doi.org/10.1097/BRS.0000000000000691.

[39] Gatchel RJ, Mayer TG. Testing minimal clinically important difference: additional comments and scientific reality testing. Spine J 2010;10(4):330–2. https://doi.org/10.1016/j.spinee.2010.01.019.

[40] Parai C, Hägg O, Lind B, Brisby H, et al. The value of patient global assessment in lumbar spine surgery: an evaluation based on more than 90,000 patients. Eur Spine J 2018;27:554–63. https://doi.org/10.1007/s00586-017-5331-0.

[41] Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol 2008;61(2):102–9. https://doi.org/10.1016/j.jclinepi.2007.03.012.

12.4 Paper 4

Acta Neurochirurgica (2023) 165:99–106
https://doi.org/10.1007/s00701-022-05421-5

ORIGINAL ARTICLE - SPINE DEGENERATIVE

Incidental dural tears associated with worse clinical outcomes in patients operated for lumbar spinal stenosis

Ole Kristian Alhaug[1,2,3] · Filip Dolatowski[4] · Ivar Austevoll[5] · Sverre Mjønes[2] · Greger Lønne[1,3]

Received: 13 September 2022 / Accepted: 2 November 2022 / Published online: 18 November 2022

Abstract

Study design Retrospective cohort study.

Objective Incidental dural (ID) tear is a common complication of spine surgery with a prevalence of 4–10%. The association between ID and clinical outcome is uncertain. Former studies found only minor differences in Oswestry Disability Index (ODI). We aimed to examine the association of ID with treatment failure after surgery for lumbar spinal stenosis (LSS).

Methods Between 2007 and 2017, 11,873 LSS patients reported to the national Norwegian spine registry (NORspine), and 8,919 (75.1%) completed the 12-month follow-up. We used multivariate logistic regression to study the association between ID and failure after surgery, defined as no effect or any degrees of worsening; we also compared mean ODI between those who suffered a perioperative ID and those who did not.

Results The mean (95% CI) age was 66.6 (66.4–66.9) years, and 52% were females. The mean (95% CI) preoperative ODI score (95% CI) was 39.8 (39.4–40.1); all patients were operated on with decompression, and 1125 (12.6%) had an additional fusion procedure. The prevalence of ID was 4.9% (439/8919), and the prevalence of failure was 20.6% (1829/8919). Unadjusted odds ratio (OR) (95% CI) for failure for ID was 1.51 (1.22–1.88); $p < 0.001$, adjusted OR (95% CI) was 1.44 (1.11–1.86); $p = 0.002$. Mean postoperative ODI 12 months after surgery was 27.9 for ID vs. 23.6 for no ID.

Conclusion We demonstrated a significant association between ID and increased odds for patient-reported failure 12 months after surgery. However, the magnitude of the detrimental effect of ID on the clinical outcome was small.

Keywords Dural tear · Failure · Worsening · Spine registry · Lumbar spinal stenosis

Introduction

Incidental durotomy (ID) is the most common perioperative complication of spinal surgery. A dural tear usually leads to cerebrospinal fluid leakage (CSF), and nerve filaments may erupt and become damaged. The prevalence of ID in spinal stenosis surgery ranges from 4 to 10% [7, 8, 14, 24, 29, 31], and ID is associated with an increased risk for neurological deficit, revision surgery, longer hospital stay, and increased treatment costs [7, 14, 30]. However, the effect of ID on clinical outcomes is debated, and previously published data is somewhat conflicting. A systematic review found minor differences in Oswestry Disability Index (ODI) scores between patients with and without ID [5]. Interpreting differences in patient-reported outcome measures (PROMs) between groups may be challenging, and few studies have used dichotomous endpoints. However, two register studies reported dichotomous patient satisfaction and found that patients who suffered an ID were less satisfied despite minor differences in PROM scores [14, 29].

Factors such as previous surgery may confound the effect of ID on outcome measures by affecting both the risk for ID and the clinical outcome. However, confounding factors are not always adjusted for in studies of ID.

This article is part of the Topical Collection on *Spine degenerative*.

✉ Ole Kristian Alhaug
olekralhaug@hotmail.com

[1] Innlandet Hospital Trust, Brumunddal, Norway

[2] Akershus University Hospital, Nordbyhagen, Norway

[3] Norwegian University of Science and Technology, Trondheim, Norway

[4] Division of Orthopaedic Surgery, Oslo University Hospital, Oslo, Norway

[5] Helse Bergen HF, Bergen, Norway

The aim of this retrospective observational study on prospectively collected national spine register data was to assess the association between ID and failure and worsening after surgery for lumbar spinal stenosis (LSS).

Methods

This retrospective study on prospectively collected data included adult patients operated on for lumbar spinal stenosis (LSS) between 2007 and 2017 in Norway. Patients were included in the Norwegian Registry of Spine Surgery (NORspine), a comprehensive national registry for quality control and research. The coverage is 70%, and the 12-month loss to follow-up is 26% [27]. The registration includes informed consent and patient and surgeon-reported data.

Before surgery, patients report symptoms of their spinal disease by standardized PROMs and general health status, quality of life, and socioeconomic status. Immediately after surgery, surgeons report details concerning the spinal diagnosis, relevant comorbidities, and surgical details, including any perioperative complications. Three months after surgery, patients report directly to NORspine by regular mail on the effect of surgery by common PROMs. The PROMs are the Norwegian translation of the Oswestry Disability Index (ODI), a pain-related disability score ranging from 0 (no impairment) to 100 (bedbound) [10, 11, 13], Numerical Rating Scale (NRS) for back and leg pain (ranging from 0 = no pain to 10 = worst pain imaginable) [19], and global perceived effect (GPE) scale — a seven-step transition scale (1 = completely recovered, 2 = much improved, 3 = somewhat improved, 4 = unchanged, 5 = somewhat worse, 6 = much worse, 7 = worse than ever) [17]. Patients also report any postoperative complications at 3 months. Finally, 12 months after surgery, patients repeatedly grade the effect of surgery on their symptoms by the PROMs mentioned above.

Primary outcome Failure was defined as patients who perceived themselves as unchanged or any degree of worse after surgery (GPE 4–7). Worsening was defined as patients who perceived themselves as "much worse" or "worse than ever" after surgery (GPE 6–7).

Secondary outcomes Failure and worsening were also defined by cutoffs for ODI final score, ODI absolute difference (postoperative minus preoperative), and ODI percentage change according to previously published definitions [2]. The ODI cutoffs for failure used in this study were ODI final score > 31, ODI absolute improvement < 8 points, and ODI percentage change < 20%. The corresponding cutoffs used to define worsening were ODI final score > 39, ODI absolute improvement < 4 points, and ODI percentage change < 9% [2]. We also used mean ODI final score to assess the impact of ID on patient-reported outcomes after surgery and mean NRS leg as an indirect measure of the association between ID and neurologic symptoms after surgery. Finally, we registered the length of hospital stay and patient-reported postoperative complications.

Statistics

Descriptive statistics The study population was described using means (95% CIs) for continuous data and numbers and proportions for categorical data.

Primary outcome To estimate the association between ID and clinical outcomes, we used multiple logistic regression with failure and worsening (defined by GPE) as dependent variables, ID (yes/no), and potential confounders as independent variables. Based on previously published data, we adjusted the primary analysis by the following potential confounders: age, gender, BMI, smoking, ASA (dichotomized as grades 1 and 2 vs. grades 3, 4, and 5), preoperative PROMs, duration of leg pain before surgery, previous surgery (at the same lumbar level), multilevel surgery, and fusion (in addition to decompression) [1, 4, 21, 22]. The potential confounders were decided a priori and not by statistical testing. We provide unadjusted and adjusted estimates for odds ratios (OR) with 95% CIs.

Secondary outcomes To examine the secondary outcomes, we repeated the regression analysis using the different dichotomous outcomes (ODI final score, ODI absolute change, and ODI percentage change). To quantify the association between ID and the mean ODI final score and NRS leg pain score, we used multiple linear regression with ODI final score and NRS leg pain as dependent variables, adjusting for the aforementioned possible confounders. We also analyzed the association between ID and length of hospital stay and patient-reported postoperative complications, using multiple linear regression and multiple logistic regression, adjusting for possible confounders. We did not impute any missing data.

We used SPSS, version 26 (IBM Corp., Armonk, NY, USA) for the statistical analyses.

Ethics Participation in NORspine is voluntary and presumes written consent. The study was also approved by The Norwegian Regional Committee for medical and health research ethics (2017/2157). The study was conducted in accordance with the Helsinki declaration, and we have reported the results in line with the STROBE guidelines.

Results

We identified 11,873 NORspine patients operated for LSS during 10 years (January 2007 to April 2017). A total of 8863 (74.6%) patients completed the 3-month follow-up, and 8919 (75.1%) responded 12 months after surgery. The mean (95% CI) age of the study population was 66.6 (66.4–66.9) years, and 4384 (52%) were females. The mean (95% CI) preoperative ODI was 39.8 (39.4–40.1). All patients were operated on with decompression, and 1125 (12.6%) had an additional fusion procedure. Patient characteristics and baseline data for PROMS are shown in Table 1. Table 1 also shows that 1829 (20.6%) reported a GPE score corresponding to failure, and 521 (5.9%) reported a GPE corresponding to worsening after surgery. Surgeons reported incidental durotomy in 439 cases (4.9%).

A total of 1708 (20.3%) patients without ID perceived treatment as failure, compared to 121 (27.8%) patients with ID. A total of 480 (5.7%) patients without ID reported worsening, compared to 41 (9.4%) patients with ID. Compared to patients with no perioperative ID, patients who suffered an ID more often reported failure (adjusted OR (95% CI) 1.45 (1.12–1.87); $p = 0.005$) and worsening (adjusted OR (95% CI) 1.50 (1.01–2.23); $p = 0.045$ (Table 2). Patients who suffered an ID during spinal surgery more often reported failure and worsening assessed by other outcome measures (Table 3).

Patients who suffered an ID during surgery reported lower ODI 12 months after surgery than those who did not suffer an ID (23.6 (23.3–24.0)) vs. 27.9 (26.2–29.6). This difference remained significant after adjusting for possible confounders (beta (95% CI) = 2.29 (0.58–4.00); $p = 0.009$) (Tables 1 and 4; Appendix Table 7). Also, patients who suffered an ID reported more leg pain after surgery compared to patients without ID: mean NRS leg pain was 4.2 (3.9–4.5) vs. 3.5 (3.5–3.6); this difference

Table 1 Patient characteristics and clinical outcome for 8908 Norwegian patients with surgically treated lumbar spinal stenosis and completed the 12-month follow-up, broken down by no incidental dural tear and incidental dural tear

	No incidental dural tear ($n = 8427$) Mean (95% CI) or n (%)	Incidental dural tear ($n = 436$) Mean (95% CI) or n (%)
Age	66.5 (66.3–66.7)	69.2 (68.2–70.1)
Female gender	4384 (51.7%)	260 (59.2%)
Civil status, single	2164 (25.5%)	118 (26.9%)
Norwegian as 1st language	8201 (96.7%)	421 (95.9%)
ASA (grade 3 to 5)*	1732 (20.7%)	108 (24.7%)
Body mass index	27.5 (27.4–27.6)	27.8 (27.4–28.4)
Smoking	1629 (19.4%)	68 (15.6%)
University or college education	2506 (29.6%)	120 (27.3%)
Receives disability benefit	1198 (14.1%)	67 (15.3%)
Previous spinal surgery (same level)	1133 (13.5%)	100 (23.1%)
Leg pain > 12 months before surgery	5023 (64.2%)	261 (65.3%)
Additional fusion (any type)	1050 (12.4%)	75 (17.1%)
Preoperative ODI**	39.6 (39.3–40.0)	42.1 (40.6–43.6)
Preoperative NRS leg pain***	6.56 (6.51–6.61)	6.68 (6.45–6.90)
Preoperative NRS back pain***	6.49 (6.44–6.54)	6.70 (6.49–6.92)
Preoperative EQ-5D****	0.377 (0.370–0.384)	0.355 (0.323–0.388)
More than one level operated	3024 (36.0%)	231 (53.2%)
Failure 12 months (missing 56)*****	1708 (20.3%)	121 (27.8%)
Worsening 12 months (missing 56)******	480 (5.7%)	41 (9.4%)
ODI final score	23.6 (23.3–24.0)	27.9 (26.2–29.6)
Length of stay (days)	3.31 (3.24–3.39)	5.68 (5.17–6.20)
Postoperative complications, any ($n = 8882$)	1154 (13.7%)	105 (23.3%)

*American Society of Anesthesiologists classification (1–5), increasing for worse health

**Oswestry Disability Index (0–100), increasing for increasing disability

***Numeric Rating Scale (0–10), increasing for increasing pain

****EuroQol's quality of life (−0.60 to 1.00), increasing for better quality of life

*****Defined as global perceived effect (GPE) 4 + 5 + 6 + 7 (unchanged or any degree of worsening) at 12 months

******Defined as global perceived effect (GPE) 6 + 7 ("much worse" or "worse than ever") at 12 months

Table 2 Multiple logistic regression using "failure"* and "worsening"** at 12-month follow-up as the dependent variable and dural tear and potential confounders as covariates

Variables	Failure		Worsening	
	OR (95% CI)	*p*-value	OR (95% CI)	*p*-value
Dural tear	1.45 (1.12–1.87)	0.005	1.50 (1.01–2.23)	0.045
Age	1.00 (1.00–1.01)	0.312	1.00 (0.99 –1.01)	0.547
Gender (female)	0.88 (0.78–1.00)	0.043	0.97 (0.78–1.20)	0.769
Body mass index (cont)	1.01 (1.00–1.03)	0.028	1.00 (0.97–1.02)	0.659
Smoking	1.41 (1.22–1.64)	0.000	1.53 (0.93–1.54)	0.001
ASA (3+4+5)***	1.13 (0.97–1.31)	0.125	1.20 (0.94–1.54)	0.153
Preoperative ODI (cont)****	1.01 (1.01–1.02)	0.000	1.03 (1.02–1.04)	0.000
Preoperative NRS leg pain*****	0.93 (0.89–0.96)	0.000	0.96 (0.90–1.02)	0.202
Preoperative NRS back pain*****	1.10 (1.05–1.14)	0.000	1.15 (1.07–1.24)	0.000
Duration leg pain > 12 months	1.63 (1.43–1.87)	0.000	1.54 (1.22–1.94)	0.000
Former surgery at same level	1.92 (1.64–2.26)	0.000	1.79 (1.38–2.31)	0.000
More than one level operated	0.90 (0.79–1.02)	0.092	0.93 (0.74–1.15)	0.490
Additional fusion, any type	0.62 (0.51–0.75)	0.000	0.61 (0.43–0.85)	0.003

*Defined as global perceived effect (GPE) 4–7 (unchanged or any degree of worsening) at 12 months

**Defined as global perceived effect (GPE) 6+7 ("much worse" or "worse than ever") at 12 months

***American Society of Anesthesiologists classification (1–5) (grades 3 to 5)

****Oswestry Disability Index (0–100), increasing for increasing disability

Table 3 Associations between dural tear and various dichotomous clinical outcomes in repeated multiple logistic regression analyses, adjusted for the same variables as in the main model (Table 2), displaying only the main results (effects of dural tear)

Clinical outcome, dichotomized		OR (95% CI)
GPE	Failure (GPE 4–7)*	1.45 (1.12–1.87)
	Worsening (GPE 6+7)**	1.50 (1.01–2.23)
ODI final score (12 months)	failure (ODI > 31)***	1.29 (1.00–1.67)
	worsening (ODI > 39)***	1.43 (1.08–1.88)
ODI absolute difference	Failure (<8 points improvement)***	1.15 (0.91–1.47)
	Worsening (<4 points improvement)***	1.28 (1.00–1.65)
ODI percentage difference	Failure (<20% improvement)***	1.17 (0.93–1.49)
	Worsening (<9% improvement)***	1.25 (0.96–1.61)

*Global perceived effect scale 4–7 (unchanged or any degree of worsening)

**Global perceived effect scale 6+7 ("much worse" or "worse than ever")

***Oswestry Disability Index (ODI) cutoffs selected from previous published study (15)

Table 4 Linear regression analyses displaying the effect of dural tear on "length of stay," "ODI final score," and "NRS leg pain," adjusted for the same potential confounders as Table 2, displaying only the main results (see Appendix Tables 7, 8, and 9 for complete results)

Variable	Beta	95% CI	*p*-value
Length of stay*	1.58	1.25–1.92	0.000
ODI final score**	2.29	0.58–4.00	0.009
NRS leg pain***	0.6	(0.3–0.9)	0.000

*Length of stay, days

**Oswestry Disability Index (0–100), increasing for increasing disability

***Numeric Rating Scale (0–10), increasing for increasing leg pain

remained significant after adjusting for confounders (beta (95% CI) of 0.6 (0.3–0.9); $p < 0.001$) (Table 4; Appendix Table 8). Patients with ID had longer hospital stays than patients without ID (mean (95% CI) 5.7(5.2–6.2) vs. 3.3 (3.2–3.4) (Table 1). This difference remained significant after adjusting for confounders (beta (95% CI) 1.58 (1.25–1.92) days; $p < 0.001$ (Table 4; Appendix Table 9).

Among respondents at a 3-month follow-up, 1259 (14.2%) patients reported any postoperative complication. The corresponding numbers for patients without ID were 1154 (13.7%), and for patients with ID, 105 (23.3%) (Table 1). The multiple logistic regression showed that patients with ID had increased odds of urinary tract infection (UTI) after surgery (OR (95% CI) 2.42 (1.53–2.73); $p < 0.001$). However, the odds for other postoperative

Table 5 Main results from multiple logistic regression analyses showing odds ratios for the association between incidental dural tear and postoperative complications, $n = 8882$ (patients answered at 3-month follow-up), adjusted for the same potential confounders as in Table 2

	OR	95% CI	*p*-value
Postoperative complications, in total	1.42	0.96–2.10	0.078
Infection, superficial	0.45	0.14–1.47	0.187
Infection, deep	0.62	0.78–4.91	0.650
Deep venous thrombosis	0.00	0.00–	0.995
Pulmonary embolism	0.00	0.00–	0.996
Pneumonia	0.62	0.15–2.69	0.527
Urinary tract infection	2.42	1.53–2.73	0.000
Micturition problems	1.46	0.78–2.73	0.232

All postoperative complications are patient reported only; there is no medical confirmation of the postoperative complications in the NORspine registry

Table 6 Multiple logistic regression displaying the effect of the potential confounders on the odds for dural tear

Variables	OR (95% CI)	*p*-value
Age	1.02 (1.01–1.04)	0.000
Gender (female)	1.36 (1.08–1.72)	0.010
Body mass index (cont)	1.02 (1.00–1.05)	0.075
Smoking	0.93 (0.68–1.27)	0.652
ASA (3 + 4 + 5)*	1.08 (0.82–1.43)	0.567
Preoperative ODI (cont)**	1.01 (1.00–1.02)	0.159
Preoperative NRS leg pain***	0.98 (0.92–1.04)	0.425
Preoperative NRS back pain***	1.02(0.95–1.09)	0.616
Duration leg pain > 12 months	0.92 (0.73–1.16)	0.458
Former surgery at same level	1.94 (1.48–2.54)	0.000
More than one level operated	1.84 (1.47–2.31)	0.000
Additional fusion	1.12 (0.82–1.53)	0.476

*American Society of Anesthesiologists (ASA) classification (1–5) (grade 3 to 5)

**Oswestry Disability Index (0–100), increasing for increasing disability

***Numeric Rating Scale (0–10), increasing for increasing pain

complications were not increased for patients with ID (Table 5).

We found that the following covariates affected the odds for failure (GPE 4–7) and worsening (GPE 6–7) after surgery for lumbar stenosis: gender, BMI, preoperative PROMs, preoperative duration of leg pain, former surgery at the same lumbar level, a fusion procedure, and smoking (Table 2). Table 6 displays possible risk factors for ID: age, gender (female), former surgery, and multilevel operations increased the odds of ID.

Discussion

Main findings This retrospective study on prospectively collected data from a nationwide register of nearly nine thousand patients operated for lumbar spinal stenosis demonstrated a significant association between incidental dural tear (ID) and increased odds for patient-reported failure and worsening 12 months after surgery. However, the magnitude of the detrimental effect of ID on the outcome of spinal surgery was small. The main finding was supported by analyses of secondary endpoints, including using different PROM derivates that defined treatment failure and worsening after surgical treatment of lumbar spinal stenosis.

A dural tear may vary from a small partial puncture of the dura, with an intact arachnoidea, to a large defect with gross leakage of cerebrospinal fluid (CSF) and damaged nerve filaments. Also, the repair of the dural tear may vary from a waterproof dural suture to an incomplete repair with continuous leakage of CSF. A surgical repair of a dural tear may also potentially traumatize nerve filaments. Clinical outcomes are expected to differ between the extremes mentioned above of dural tear and repair; however, NORspine data on IDs do not differentiate between different grades of ID or various surgical repairs. We cannot rule out that the increased risk for failure and worsening associated with ID may be attributed to the most severe IDs and incomplete repairs.

Several studies have reported minor detrimental effects of ID on PROMs, with uncertain clinical importance effect sizes [5, 8, 16, 25, 28, 29]. One large register study from Sweden [29] found only a minor increase in postoperative ODI associated with ID but still a significantly inferior patient satisfaction (a categorical variable) associated with ID. Similarly, a Tango spine register study found no effect of ID on core outcome measure index (COMI) scores but a trend toward lesser patient-reported satisfaction among patients with an ID [14]. Small to moderate between-groups differences in proportions of failures are often associated with only minor differences in absolute PROM scores. Our partly conflicting results between dichotomous and continuous outcome measures reflect the small effect ID has on ODI scores and are in line with the studies mentioned above. This illustrates the importance of using categorized outcomes in addition to mean PROM values. Categorized outcomes are also emphasized in a review article on clinical important change [18].

Secondary outcomes Patients who suffered a dural tear needed longer hospitalization, a finding consistent with previously published data [7, 8, 14]. Current practice is to advise (one or) a few days of bed rest after IDs with CSF leakage, especially when patients report spinal headaches. Bedrest after IDs may be a primary reason for prolonged

hospitalization. Observation, further diagnostic tests, and reoperations may explain the prolonged stay in patients with IDs.

In this (cohort) study, IDs were associated with increased odds of postoperative urinary tract infection, however not for other patient-reported postoperative complications. Previous studies have shown that IDs may be associated with increased risk for other postoperative complications such as wound infections, neural deficits, postoperative delirium, and perioperative blood loss [7, 8, 30]. NORspine records patient-recorded complications at 3 months after surgery. Patients may define complications differently from health care personnel. The NORspine design and complication recording probably explain the difference between our findings and previously published data on the impact of ID on postoperative complications.

This study was specifically designed to assess the effect of ID on the odds of failure and worsening after surgical treatment of lumbar stenosis. However, as presented in Table 2, other baseline variables such as preoperative ODI, duration of leg pain, and previous lumbar surgery also affected the odds of failure and worsening — in line with previous studies [1, 22]. However, our study was not designed to numerically assess the predictors for failure and worsening. Our finding of increased odds for ID with increasing age and previous surgery is also in line with an earlier Tango register study [14].

Limitations NORspine data do not differentiate between different grades of ID or different managements of ID. This may obscure essential factors that may affect the impact of ID on clinical outcome. Furthermore, surgeons are likely to underreport dural tears. In a former validation study of NORspine data against electronic patient records, the authors demonstrated a sensitivity of 40% for the actual reporting of perioperative complications [3]. Underreporting of complications has been shown in other registers [20, 23] and could have affected conclusions regarding the impact of ID on clinical outcomes in other studies. Although the NORspine frequency of IDs (4.9%) is comparable to some reports [7, 8, 14, 24, 29, 31], it is inferior to the prevalence of IDs (9–11%) in two large RCTs on similar patient groups [6, 12]. The underreporting of ID may contribute to underestimating the effect of ID on clinical outcome.

ID may lead to neurological damage. NORspine does not record postoperative neurological sequelae or liquorrhea. However, we used NRS leg pain as a surrogate variable to assess neurological sequelae.

Furthermore, *postoperative* complications registered in NORspine are only reported by patients 3 months after surgery, and this register design may miss some postoperative complications. Postoperative complications may have been treated at different medical centers or in primary care, and complete complication data are unavailable in NORspine.

At 12 months after surgery, NORspine demonstrates a loss to follow-up of 26% [27]. According to former validation studies of medical registers, loss to follow-up does not systematically bias clinical outcomes [9, 15, 26]. One of these studies [26] specifically examined the impact of loss to follow-up in the NORspine registry.

Our choice of GPE as the primary outcome can be discussed. GPE may be susceptible to recall bias. However, GPE has shown excellent test re-test reliability in studies of musculoskeletal disorders [17].

Conclusion

This prospective nationwide spine register study found that incidental dural tear was associated with increased odds of patient-reported failure and worsening after surgery for lumbar spinal stenosis. The detrimental association to the clinical outcome was small and could be attributed to the most severe dural tears. Incidental dural tear was also associated with increased length of stay.

Appendix

Table 7 Quantifying the association between dural tear and final ODI score using multiple linear regression, "ODI 12 months" as the dependent variable, and dural tear and potential confounders as independent variables

Variables	*B* (95% CI)	*p*-value
Dural tear	2.29 (0.58 to 4.00)	0.009
Age	0.05(0.01 to 0.08)	0.019
Gender (female)	1.05 (0.29 to 1.81)	0.007
Body mass index (cont)	0.28 (0.19 to 0.36)	0.000
Smoking	3.56 (2.61 to 4.51)	0.090
ASA (3+4+5)*	2.80 (1.83 to 3.77)	0.000
Preoperative ODI (cont)**	0.50 (0.47 to 0.53)	0.000
Preoperative NRS leg pain***	−0.80 (−1.02 to −0.59)	0.000
Preoperative NRS back pain***	0.89 (0.66 to 1.11)	0.000
Duration leg pain > 12 months	4.98 (4.21 to 5.74)	0.000
Former surgery at same level	6.40 (5.33 to 7.46)	0.000
More than one level operated	−0.04 (−0.82 to 0.73)	0.917
Additional fusion	−2.52 (−3.65 to −1.39)	0.000

*American Society of Anesthesiologists (ASA) classification (1–5) (grades 3 to 5)

**Oswestry Disability Index (0–100), increasing for increasing disability

***Numeric Rating Scale (0–10), increasing for increasing pain

Table 8 Quantifying the association between dural tear and NRS leg pain score using multiple linear regression, "NRS leg pain" as the dependent variable, and dural tear and potential confounders as independent variables

Variables	*B* (95% CI)	*p*-value
Dural tear	0.57 (0.26 to 0.89)	0.000
Age	0.01(0.0 to 0.02)	0.023
Gender (female)	0.62 (−0.8 to 0.20)	0.377
Body mass index (cont)	0.03 (0.02 to 0.05)	0.000
Smoking	0.44 0.28 to 0.62)	0.000
ASA (3 + 4 + 5)*	0.28 (0.11 to 0.46)	0.000
Preoperative ODI (cont)**	0.03 (0.02 to 0.03)	0.000
Preoperative NRS leg pain***	0.11 (0.08 to 0.15)	0.000
Preoperative NRS back pain***	0.11 (0.07 to 0.15)	0.000
Duration leg pain > 12 months	0.79 (0.65 to 0.93)	0.000
Former surgery at same level	0.94 (0.75 to 1.13)	0.000
More than one level operated	−0.13 (−0.27 to 0.01)	0.078
Additional fusion	−0.62 (−0.83 to −0.42)	0.000

*American Society of Anesthesiologists (ASA) classification (1–5) (grade 3 to 5)

**Oswestry Disability Index (0–100), increasing for increasing disability

***Numeric Rating Scale (0–10), increasing for increasing pain

Table 9 Quantifying the association between dural tear and length of stay, using multiple linear regression, "length of stay" as the dependent variable, and dural tear and potential confounders as independent variables

Variables	*B* (95% CI)	*p*-value
Dural tear	1.58 (1.25 to 1.92)	0.000
Age	0.03(0.02 to 0.04)	0.000
Gender (female)	0.50 (0.35 to 0.64)	0.000
Body mass index (cont)	0.02 (0.00 to 0.04)	0.000
Smoking	−0.16 (−0.35 to 0.03)	0.090
ASA (3 + 4 + 5)*	0.35 (0.17 to 0.54)	0.000
Preoperative ODI (cont)**	0.03 (0.02 to 0.02)	0.000
Preoperative NRS leg pain***	−0.07 (−0.12 to −0.3)	0.001
Preoperative NRS back pain***	0−01 (−0.04 to 0.05)	0.691
Duration leg pain > 12 months	0.25 (0.10 to 0.40)	0.001
Former surgery at same level	0.22 (0.01 to 0.43)	0.042
More than one level operated	0.81 (0.65 to 0.96)	0.000
Additional fusion	3.54 (3.31 to 3.76)	0.000

*American Society of Anesthesiologists (ASA) classification (1–5) (grade 3 to 5)

**Oswestry Disability Index (0–100), increasing for increasing disability

***Numeric Rating Scale (0–10), increasing for increasing pain

Funding Open access funding provided by NTNU Norwegian University of Science and Technology (incl St. Olavs Hospital - Trondheim University Hospital)

Declarations

Ethical approval and consent to participate All procedures performed in studies involving human participants were in accordance with the ethical standards of the Norwegian National Research Committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study.

Conflict of interest The authors declare no competing interests.

References

1. Aalto T, Sinikallio S, Kröger H, Viinamäki H, Herno A, Leinonen V, Turunen V, Savolainen S, Airaksinen O (2012) Preoperative predictors for good postoperative satisfaction and functional outcome in lumbar spinal stenosis surgery–a prospective observational study with a two-year follow-up. Scand J Surg 101(4):255–260. https://doi.org/10.1177/145749691210100406
2. Alhaug OK, Dolatowski FC, Solberg TK, Lønne G (2021) Criteria for failure and worsening after surgery for lumbar spinal stenosis: a prospective national spine registry observational study. Spine J 21(9):1489–1496. https://doi.org/10.1016/j.spinee.2021.04.008
3. Alhaug OK, Kaur S, Dolatowski F et al (2022) Accuracy and agreement of national spine register data for 474 patients compared to corresponding electronic patient records. Eur Spine J 31:801. https://doi.org/10.1007/s00586-021-07093-8
4. Alshameeri ZAF, Jasani V (2021) Risk factors for accidental dural tears in spinal surgery. Int J Spine Surg 15(3):536–548. https://doi.org/10.14444/8082
5. Alshameeri ZAF, Ahmed E-N, Jasani V (2021) Clinical outcome of spine surgery complicated by accidental dural tears: meta-analysis of the literature. Global Spine J 11(3):400–409. https://doi.org/10.1177/2192568220914876
6. Austevoll IM, Hermansen E, Fagerland MW et al (2021) Decompression with or without fusion in degenerative lumbar spondylolisthesis. N Engl J Med 385:526–538
7. Buck JS, Yoon ST (2015) The incidence of durotomy and its clinical and economic impact in primary, short-segment lumbar fusion:

an analysis of 17,232 cases. Spine (Phila Pa 1976) 40(18):1444–50. https://doi.org/10.1097/BRS.0000000000001025
8. Desai A, Ball PA, Bekelis K, Lurie J, Mirza SK, Tosteson TD, Weinstein JN (2015) SPORT: Does incidental durotomy affect longterm outcomes in cases of spinal stenosis? Neurosurgery 76(Suppl 1 (0 1)):S57-63. https://doi.org/10.1227/01.neu.0000462078.58454.f4
9. Endler P, Ekman P, Hellström F, Möller H, Gerdhem P (2020) Minor effect of loss to follow-up on outcome interpretation in the Swedish spine register. Eur Spine J 29(2):213–220. https://doi.org/10.1007/s00586-019-06181-0
10. Fairbank JC, Pynsent PB (2000) The Oswestry Disability Index. Spine (Phila Pa 1976) 25(22):2940–52; discussion 2952. https://doi.org/10.1097/00007632-200011150-00017
11. Fairbank JC, Couper J, Davies JB, O'Brien JP (1980) The Oswestry Low Back Pain Disability Questionnaire. Physiotherapy 66(8):271–273
12. Försth P, Ólafsson G, Carlsson T et al (2016) A randomized, controlled trial of fusion surgery for lumbar spinal stenosis. N Engl J Med 374:1413–1423
13. Grotle M, Brox JI, Vøllestad NK (2003) Cross-cultural adaptation of the Norwegian versions of the Roland-Morris Disability Questionnaire and the Oswestry Disability Index. J Rehabil Med 35(5):241–247. https://doi.org/10.1080/16501970306094
14. Herren C, Sobottke R, Mannion AF, Zweig T, Munting E, Otten P, Pigott T, Siewe J, Aghayev E, Contributors ST (2017) Incidental durotomy in decompression for lumbar spinal stenosis: incidence, risk factors and effect on outcomes in the Spine Tango registry. Eur Spine J 26(10):2483–2495. https://doi.org/10.1007/s00586-017-5197-1
15. Høimark K, Støttrup C, Carreon L, Andersen MO (2016) Patient-reported outcome measures unbiased by loss of follow-up. Single-center study based on DaneSpine, the Danish spine surgery registry. Eur Spine J 25(1):282–286. https://doi.org/10.1007/s00586-015-4127-3
16. Jones AA, Stambough JL, Balderston RA, Rothman RH, Booth RE Jr (1989) Long-term results of lumbar spine surgery complicated by unintended incidental durotomy. Spine (Phila Pa 1976) 14(4):443–6. https://doi.org/10.1097/00007632-198904000-00021
17. 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(7):760-766.e1. https://doi.org/10.1016/j.jclinepi.2009.09.009
18. Katz NP, Paillard FC, Ekman E (2015) Determining the clinical importance of treatment benefits for interventions for painful orthopedic conditions. J Orthop Surg 10:24. https://doi.org/10.1186/s13018-014-0144-x
19. McCaffery M, Beebe A (1989) Pain: clinical manual for nursing practice. Mosby, St. Louis
20. Meyer B, Shiban E, Albers LE, Krieg SM (2020) Completeness and accuracy of data in spine registries: an independent audit-based study. Eur Spine J 29:1453–1461
21. Murphy ME, Kerezoudis P, Alvi MA, McCutcheon BA, Maloney PR, Rinaldo L, Shepherd D, Ubl DS, Krauss WE, Habermann EB, Bydon M (2017) Risk factors for dural tears: a study of elective spine surgery. Neurol Res 39(2):97–106. https://doi.org/10.1080/01616412.2016.1261236
22. Nerland US, Jakola AS, Giannadakis C, Solheim O, Weber C, Nygaard ØP, Solberg TK, Gulati S (2015) The risk of getting worse: predictors of deterioration after decompressive surgery for lumbar spinal stenosis: a multicenter observational study. World Neurosurg 84(4):1095–1102. https://doi.org/10.1016/j.wneu.2015.05.055
23. Öhrn A, Elfström J, Liedgren C, Rutberg H (2011) Reporting of sentinel events in Swedish hospitals: a comparison of severe adverse events reported by patients and providers. Jt Comm J Qual Patient Saf 37(11):495–501. https://doi.org/10.1016/S1553-7250(11)37063.ISSN1553-7250
24. Papavero L, Engler N, Kothe R (2015) Incidental durotomy in spine surgery: first aid in ten steps. Eur Spine J 24(9):2077–2084. https://doi.org/10.1007/s00586-015-3837-x
25. Saxler G, Krämer J, Barden B, Kurt A, Pförtner J, Bernsmann K (2005) The long-term clinical sequelae of incidental durotomy in lumbar disc surgery. Spine (Phila Pa 1976) 15;30(20):2298–302. https://doi.org/10.1097/01.brs.0000182131.44670.f7
26. Solberg TK, Sørlie A, Sjaavik K, Nygaard ØP, Ingebrigtsen T (2011) Would loss to follow-up bias the outcome evaluation of patients operated for degenerative disorders of the lumbar spine? Acta Orthop 82(1):56–63. https://doi.org/10.3109/17453674.2010.548024
27. Solberg T, Olsen RR, Berglund ML (2018) Annual report for NORspine 2018
28. Strömqvist F, Jönsson B, Strömqvist B, Swedish Society of Spinal Surgeons (2010) Dural lesions in lumbar disc herniation surgery: incidence, risk factors, and outcome. Eur Spine J 19(3):439–442. https://doi.org/10.1007/s00586-009-1236-x
29. Strömqvist F, Sigmundsson FG, Strömqvist B, Jönsson B, Karlsson MK (2019) Incidental durotomy in degenerative lumbar spine surgery - a register study of 64,431 operations. Spine J 19(4):624–630. https://doi.org/10.1016/j.spinee.2018.08.012
30. Takenaka S, Makino T, Sakai Y, Kashii M, Iwasaki M, Yoshikawa H, Kaito T (2019) Dural tear is associated with an increased rate of other perioperative complications in primary lumbar spine surgery for degenerative diseases. Medicine (Baltimore) 98(1):e13970. https://doi.org/10.1097/MD.0000000000013970
31. Ulrich NH, Burgstaller JM, Brunner F et al (2016) The impact of incidental durotomy on the outcome of decompression surgery in degenerative lumbar spinal canal stenosis: analysis of the Lumbar Spinal Outcome Study (LSOS) data—a Swiss prospective multicenter cohort study. BMC Musculoskelet Disord 17:170. https://doi.org/10.1186/s12891-016-1022-y

12.5 Paper 5

ELSEVIER

The SPINE JOURNAL

The Spine Journal 23 (2023) 261–270

Clinical Study

Predictors for failure after surgery for lumbar spinal stenosis: a prospective observational study

Ole Kristian Alhaug, MD[a,b,c,*], Filip C. Dolatowski, MD, PhD[d], Tore K. Solberg, MD, PhD[e], Greger Lønne, MD, PhD[a,b]

[a] *Innlandet Hospital Trust, The Research Center for Age-Related Functional Decline and Disease, PO Box 68, N-2313 Ottestad, Norway*
[b] *Norwegian University of Science and Technology, NTNU PO Box 191, N-7491 Trondheim, Norway*
[c] *Orthopedic Department, Akershus University Hospital, PO Box 1000, N-1478 Loerenskog, Norway*
[d] *Orthopedic Department, Oslo University Hospital, PO Box 4956, N-0424, Oslo, Norway*
[e] *Neurosurgical Department, University Hospital of North Norway, N-9038, Tromsoe, Norway*

Received 2 August 2022; revised 20 October 2022; accepted 21 October 2022

Abstract

BACKGROUND/CONTEXT: Some patients do not improve after surgery for lumbar spinal stenosis (LSS), and surgical treatment implies a risk for complications and deterioration. Patient selection is of paramount importance to improve the overall clinical results and identifying predictive factors for failure is central in this work.

PURPOSE: We aimed to explore predictive factors for failure and worsening after surgery for LSS.

STUDY DESIGN /SETTING: Retrospective observational study on prospectively collected data from a national spine registry with a 12-month follow-up.

PATIENT SAMPLE: We analyzed 11,873 patients operated for LSS between 2007 and 2017 in Norway, included in the Norwegian registry for spine surgery (NORspine). Twelve months after surgery, 8919 (75.1%) had responded.

OUTCOME MEASURES: Oswestry Disability Index (ODI) 12 months after surgery.

METHODS: Predictors were assessed with uni- and multivariate logistic regression, using backward conditional stepwise selection and a significance level of 0.01. Failure (ODI>31) and worsening (ODI>39) were used as dependent variables.

RESULTS: Mean (95%CI) age was 66.6 (66.4−66.9) years, and 52.1% were females. The mean (95%CI) preoperative ODI score was 39.8 (39.4−40.1). All patients had decompression, and 1494 (12.6%) had an additional fusion procedure. Twelve months after surgery, the mean (95%CI) ODI score was 23.9 (23.5−24.2), and 2950 patients (33.2%) were classified as failures and 1921 (21.6%) as worse. The strongest predictors for failure were duration of back pain > 12 months (OR [95%CI]=2.24 [1.93−2.60]; p<.001), former spinal surgery (OR [95%CI]=2.21 [1.94−2.52]; p<.001) and age>70 years (OR (95%CI)=1.97 (1.69−2.30); p<.001). Socioeconomic variables increased the odds of failure (ORs between 1.36 and 1.62). The strongest predictors for worsening were former spinal surgery (OR [95%CI]=2.04 [1.77−2.36]; p<.001), duration of back pain >12 months (OR [95%CI]=1.83 [1.45−2.32]; p<.001) and age >70 years (OR [95%CI]=1.79 [1.49−2.14]; p<.001). Socioeconomic variables increased the odds of worsening (ORs between 1.33−1.67).

CONCLUSIONS: After surgery for LSS, 33% of the patients reported failure, and 22% reported worsening as assessed by ODI. Preoperative duration of back pain for longer than 12 months, former spinal surgery, and age above 70 years were the strongest predictors for increased odds of

FDA device/drug status: Not applicable.

Author disclosures: ***OKA:*** Nothing to disclose. ***FCD:*** Nothing to disclose. ***TKS:*** Nothing to disclose. ***GL:*** Nothing to disclose.

*Corresponding author. Innlandet Hospital Trust, The Research Center for Age-Related Functional Decline and Disease, PO Box 68, N-2313 Ottestad, Norway. Tel.: +47 41127443

E-mail address: olekralhaug@hotmail.com (O.K. Alhaug).

https://doi.org/10.1016/j.spinee.2022.10.010

Keywords: Failure; Patient selection; Predictors; Spinal stenosis; Spine registries; Worsening

Introduction

Lumbar spinal stenosis (LSS) is a common condition and represents an increasing burden for our health care system. Surgical treatment for LSS is a good option for patients failing nonoperative care [1–6].

Although the results after surgical treatment can vary, many studies report good results in 62%–75% of patients [7–13]. Furthermore, even though the surgical techniques constantly develop toward less invasiveness, the clinical results seem stable [7,14].

Of note, surgery also implies a risk for complications and deterioration, which underlines the importance of proper patient selection for surgery. It is tempting to consider a clinician's experience and judgment as the best prerequisite to selecting suitable patients for surgery. However, complex and subtle clinical pictures are difficult to perceive, and expert surgeons may overestimate the benefit of surgery and underestimate the risk of unfavorable outcomes [15]. Furthermore, risk factors may be positively or negatively associated with results, interact, and be subject to confounding. It may be challenging to overview a range of predictors in a clinical setting and assess their combined clinical relevance. However, previous studies have shown that a combined set of predictors perform better than one single in predicting outcomes and that predictors are superior to clinical judgment. Hence predictor analyses are central for informing and improving patient selection and clinical decision-making before surgery [16–19].

Many patients seem to understand the uncertainty regarding clinical outcomes and that not everybody improves. The risk of getting worse may be harder to accept. Former predictor studies are based on different databases (mandatory registers, voluntary registers, or a few treating centers) and include different variables or definitions of variables. Hence, the results are not necessarily transferable to other settings. Some former studies focus on improvement rather than failure or worsening. Risk assessment concerning unfavorable outcomes is crucial for informing patients and making clinical decisions and could aid in reducing the adverse effects of spine surgery [20].

We aimed to explore predictors for failure and worsening after surgery for LSS in a national comprehensive spine registry.

Material/method

This retrospective study on prospectively collected data includes adult patients operated on for lumbar spinal stenosis (LSS) between 2007 and 2017 in Norway. The Norwegian registry for spine surgery (NORspine) is a mandatory register with a coverage of 70% at the case level and 100% at the surgical unit level [21]. The registration process includes a preoperative form on socio-economical items and standard PROMs filled by the patients at admission for surgery (baseline). The surgeon fills out one postoperative form on diagnosis and surgical details. The patients complete two follow-up forms, one at three months and one at 12 months. They include common PROMs and a global perceived effect (GPE) transition scale. The PROMs are the Norwegian translation of the Oswestry Disability Index (ODI), a pain-related disability score ranging from 0 (no impairment) to 100 (bedbound) [22–23], and Numerical Rating Scales (NRS) for back and leg pain (ranging from 0 = no pain to 10 = worst pain imaginable) [24]. The GPE scale have seven steps (1=completely recovered, 2=much improved, 3=somewhat improved, 4=unchanged, 5=somewhat worse, 6=much worse, 7=worse than ever) [25].

Primary outcome: We defined failure as ODI final score >31 points and worsening as ODI >39 points, in accordance with a recent study [26].

Sensitivity analysis: To evaluate the robustness of the prediction we defined an ODI change of less than 8 points as failure and less than 4 points as worsening [26]. Finally, we also used GPE to assess the effect after surgery.

Statistics

We report central tendency in terms of mean (95% CI) for continuous data with normal distribution and number and proportions (%) for categorical data. We assessed predictors using uni- and multivariate logistic regression, with backward conditional stepwise selection with an entry and removal threshold of 0.01.

ODI score 12 months after surgery of 31 for failure and 39 for worsening were used as dependent variables (outcome). Covariates in the predictor analyses were chosen according to previous literature: age, gender, smoking, ASA classification, BMI, educational level, civil status, Norwegian speakers, disability benefit, former spinal surgery, MRI findings, preoperative ODI score, duration of symptoms, multilevel surgery [27–29]. Among the covariate variables, some were dichotomized to improve the data-to-model fit and facilitate interpretation of the analyses (age, BMI, ASA classification, and educational level). There was no strong (<0.7) correlation between the covariates, and linearity between continuous variables was checked against logit failure. Only preoperative ODI had nonlinearity, as displayed in Figure. Covariates were tested

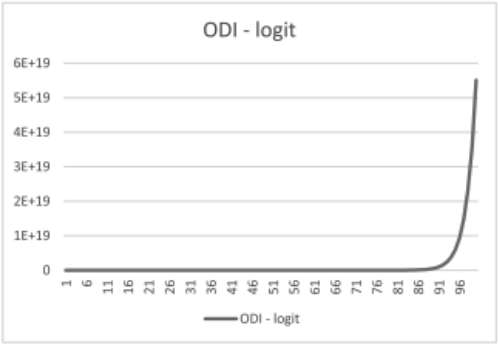


Figure. Non linear realtionship between preoperaitve ODI and logit failure.

for relevant interactions using multivariate logistic regression, and no interactions had a statistically significant association with the outcomes.

Subgroup analyses

To explore the role of preoperative back pain, we dichotomized the population into those who reported more back than leg pain (yes/no). We reported the number (%) of failures in patients who received decompression and fusion versus those who received decompression only. We also explored the role of the number of levels operated by analyzing the number (%) of patients reporting failure who were operated on at one, two, three, and four levels. We performed subgroup analyses of patients with previous surgery at the same or another level. For secondary explorative analyses, we used simple crosstabulations.

We did not impute any missing data. All statistical analyses were done with SPSS version 26 (IBM Corp. released in 2017. IBM SPSS Statistics for Windows, Version 25. Armonk, NY, USA).

Ethics

Participation in the registry is voluntary and includes written consent. The study was also approved by The Norwegian Regional Committee for medical and health research ethics ((2017/2157). The study was conducted in accordance with the Helsinki declaration and is presented according to the STROBE statement [30].

Results

Baseline

We identified 11,873 patients operated for LSS between January 2007 and April 2017, 8,863 (74.6%) had completed three months follow-up, and 8,919 (75.1%) had completed 12 months follow-up. Table 1 displays patient characteristics at baseline for all patients and patients with treatment categorized as "failure" and "worsening" subgroups. The mean (95%CI) age was 66.6 (66.4–66.9) years, and 4,644 (52.1%) were females. The mean (95%CI) ODI was 39.8 (39.4–40.1). Patients with failure and worsening were elder, more often ASA >2, and had higher BMI and preoperative ODI. In addition, they more often had low education, comorbidities, disability benefit, and former surgery (Table 1). Patients lost to follow-up at 12 months were younger, more often smokers, and had higher preoperative ODI scores (Appendix Table 1). Table 2 displays the type of surgical treatment given. All patients had some kind of decompression, bilateral foraminotomy was the largest group, and 1,494 (12.6%) patients had an additional fusion procedure.

Clinical results

Twelve months after the operation, 2950 (33.2%) patients were categorized as "failures," including 1,921 (21.6%) classified as "worse" according to the ODI final score cut-offs. When we used the ODI change score cut –offs, , (32.8%) reported failure, and 2,132 (24.2%) reported worse. The mean (95%CI) ODI 12 months after surgery was 23.9 (23.5–24.2) and the mean (95%CI) improvement in ODI was 15.9 (15.5–16.3) points. When patients graded the effect of surgery by GPE, 1,829 (20.6%) perceived themselves as "unchanged" or any degree of worsening, and 521 (5.9%) reported "much worse" or "worse than ever."

Predictors

Table 3 shows the results of the uni- and multivariate logistic regression analyses.

Failure

The strongest independent risk factors for failure identified in the multivariate model were duration of back pain >12 months (OR=2.24 [1.93–2.60]; p<.001), former spinal surgery (OR=2.21 [1.94–2.52]; p<.001) and age >70 years (OR=1.97 [1.69–2.30]; p<.001). Socioeconomic variables, that is, receiving disability benefits, low educational level, not being a native Norwegian speaker, and living alone, all increased the odds of failure (OR between 1.36–1.62). Variables concerning general health, that is, smoking, BMI >30, and ASA>2, also increased the odds of failure (OR 1.32 –1.40). The spine-related disability (ODI) and pain medication increased the odds (OR 1.06–1.29). Of the radiological variables, only the finding of degenerative olisthesis showed an effect on the odds for failure with decreased odds (OR=0.75).

Worsening

The strongest independent risk factors for worsening identified in the multivariate model were former spinal

Table 1
Patient characteristics of 11,873 Norwegian patients with surgically treated lumbar spinal stenosis in a 10 years period (2007–2017)

	All patients, n=11,873 Mean (95%CI), or n (%)	Failure (ODI>31), n=2,950 Mean (95%CI), or n (%)	S
Age (cont)	65.8 (65.6–66.0)	67.8 (67.2–68.0)	67.8 (67.3–68.3)
Age > 70 years	4,442 (37.5%)	1,352 (45.8%)	1,016 (52.9%)
Gender female	6,204 (52.3%)	1,714 (58.1%)	1,115 (58.0%)
Civil status, living alone	3,169 (26.8%)	937 (31.9%)	619 (32.4%)
Native Norwegian speaker	11,353 (96.0%)	2,796 (95.4%)	1,910 (95.3%)
ASA grade >2*	2,462 (21.0%)	848 (29.1%)	601 (31.7%)
Body mass index (cont)	27.6 (27.5–27.7)	28.0 (27.9–28.2)	28.1 (27.9–28.3)
Body mass index >30	2,920 (26.2%)	853 (31.1%)	573 (32.1%)
Smoking	2,518 (21.4%)	682 (23.3%)	470 (24.7%)
Level of education below college	8,209 (70.4%)	2,281 (79.1%)	1,501 (80.1%)
Any comorbidity	7,243 (67.2%)	2,031 (75.2%)	1,347 (76.0%)
Receives disability benefit (all types)	4,007 (34.8%))	1,082 (38.1%)	726 (39.3%)
Previous lumbar spine surgery	2,968 (25.3%)	1,025 (35.3%)	703 (37.1%)
MRI central stenosis	8,288 (69.8%)	2,104 (71.3%)	1,372 (71.4%)
MRI lateral stenosis	6,796 (57.2%)	1,616 (54.8%)	878 (45.7%)
MRI foraminal stenosis	1,225 (10.3%)	337 (11.4%)	218 (11.3%)
X-ray degenerative olisthesis	1,854 (15.6%)	416 (14.1%)	281 (14.6%)
Leg pain > 12 months duration	7,115 (65.1%)	1,940 (72.9%)	1,295 (73.8%)
Back pain > 12 months duration	8,415 (75.4%)	2,267 (82.3%)	1,507 (83.7%)
Preoperative ODI†	40.3 (40.1–40.6)	48.1 (47.6–48.6)	50.8 (50.2–51.4)
Preoperative leg pain (NRS)‡	6.59 (6.55–6.63)	7.05 (6.97–7.12)	7.22 (7.12–7.32)
Preoperative back pain (NRS)‡	6.53 (5.49–6.57)	7.24 (7.17–7.31)	7.44 (7.35–7.52)
Preoperative EQ-5D§	0.363 (0.357–0.369)	0.253 (0.241–0.265)	0.205 (0.191–0.219)

All patients, and patients reported as failure and worse.
* ASA = American Society of Anesthesiologists classification (1–5).
† ODI = Oswestry Disability Index (0–100), indicating increasing disability.
‡ NRS = Numeric Rating Scale (0–10), indicating increasing pain.
§ ED-5D = EuroQol's quality of life, (-0.60 to 1.00), indicating increasing quality of life.

surgery (OR=2.04 [1.77–2.36]; p<.001), duration of back pain >12 months (OR=1.83 [1.45–2.32]; p<.001), and age >70 years (OR=1.79 (1.49–2.14); p<.001). Socioeconomic variables, that is, receiving a disability benefit, low educational level, and living alone, increased the odds of worsening (OR between 1.33 and 1.67). Variables concerning general health, that is, as BMI >30 and ASA >2 increased the odds for worsening (OR 1.28–1.38), and spine-related disability (ODI) and duration of leg pain > 12 mths increased the odds for worsening (OR 1.07–1.30). None of the preoperative radiological variables influenced the odds of worsening, except the finding of a degenerative olisthesis, which decreased the odds of failure (OR=0.76 [0.64–0.89]; p<.001)).

Table 2
Surgical treatment for 11,873 Norwegian patients with lumbar spinal stenosis in a 10 years period (2007–2017)

	Completed 12 months follow up (n=8,919) Mean (95%CI), or n (%)	Lost to follow up (n=2,954) Mean (95%CI), or n (%)
Fusion surgery	1,125 (12.6%)	369 (12.5%)
Fusion, TLIF*	309 (3.5%)	120 (4.1%)
Fusion, PLIF†	38 (0.4%)	6 (0.2%)
Fusion, PLF‡	769 (8.6%)	241 (8.2%)
Fusion, other	9 (0.2%)	2 (0.1%)
Decompression		
Unilateral foramenotomy	1,973 (22.1%)	732 (24.8%)
Bilateral foraminotomy	3,485 (39.1%)	1,120 (37.9%)
Cross over / "over the top"	1,388 (15.6%)	544 (18.4%)
Laminectomy	2,199 (24.7%)	622 (21.1%)
More than one level operated	3,255 (36.9%)	975 (33.3%)

Patients completed 12 months follow-up, and patients lost to follow-up.
* Transforaminal Lumbar Interbody Fusion.
† Posterior Lumbar Interbody Fusion.
‡ Posterolateral Lumbar Fusion.

Table 3
Logistic regression for 8,919 patients operated for lumbar spinal stenosis and registered in NORspine during 2007−2017, using failure (ODI>31) and worsening (ODI>39) as dependent variables and potential predictors as explanatory variables

	Failure (ODI>31 points)				Worsening (ODI>39 points)			
Variables	Univariate OR (95%CI)	p value	Multivariate OR (95%CI)	p value	Univariate OR (95%CI)	p value	Multivariate OR (95%CI)	p value
Age >70 years	1.50 (1.37−1.64)	<.001	1.99 (1.71−2.31)	<.001	1.50 (1.36−1.66)	<.001	1.93 (1.62−2.31)	<.001
Gender (female)	1.44 (1.32−1.57)	<.001			1.36 (1.23−1.51)	<.001		
Smoking	1.46 (1.31−1.63)	<.001	1.40 (1.21−1.62)	<.001	1.52 (1.35−1.71)	<.001	1.53 (1.31−1.80)	<.001
Body mass index >30	1.54 (1.39−1.70)	<.001	1.34 (1.18−1.53)	<.001	1.53 (1.36−1.71)	<.001	1.33 (1.15−1.54)	<.001
ASA grade >2 *	2.05 (1.85−2.28)	<.001	1.34 (1.16−1.54)		2.14 (1.91−2.40)	<.001	1.39 (1.19−1.62)	<.001
Education level below college	1.99 (1.79−2.21)	<.001	1.54 (1.35−1.75)	<.001	1.95 (1.72−2.21)	<.001	1.51 (1.29−1.76)	<.001
Civil status, living alone	1.62 (1.46−1.78)	<.001	1.33 (1.17−1.52)	<.001	1.52 (1.37−1.71)	<.001	1.26 (1.09−1.45)	.002
Not Native Norw speakers	1.58 (1.26−2.00)	<.001	1.66 (1.23−2.23)	.001	1.49 (1.16−1.92)	<.001		
Disability benefit (all types)†	1.46 (1.33−1.60)	<.001	1.67 (1.44−1.94)	<.001	1.47 (1.32−1.63)	<.001	1.66 (1.40−1.98)	<.001
Former lumbar spine surgery (any)	2.26 (2.05−2.50)	<.001	2.21 (1.94−2.51)	<.001	2.19 (1.96−2.44)	<.001	2.00 (1.74−2.30)	<.001
MRI central stenosis	1.05 (0.95−1.15)	.358			1.05 (0.94−1.17)	.428		
MRI lateral stenosis	0.91 (0.83−1.00)	.040			0.90 (0.81−1.00)	.044		
MRI foraminal stenosis	1.18 (1.02−1.36)	.024			1.14 (0.97−1.34)	.120		
RF degen olisthesis	0.85 (0.75−0.97)	.013	0.76 (0.64−0.89)	.001	0.92 (0.80−1.06)	.255		
Pre opr ODI (cont)‡	1.06 (1.06−1.07)	<.001	1.06 (1.05−1.06)	<.001	1.07 (1.07−10.7)	<.001	1.07 (1.06−1.07)	<.001
Duration leg pain >12 months	1.68 (1.52−1.86)	<.001			1.74 (1.55−1.96)	<.001	1.29 (1.06−1.56)	.010
Duration backpain >12months	1.87 (1.68−2.10)	<.001	2.17 (1.88−2.50)	<.001	1.95 (1.70−2.24)	<.001	1.85 (1.47−2.32)	<.001
Multilevel surgery§	1.21 (1.11−1.33)	<.001			1.19 (1.07−1.32)	.001		

* ASA = American Society of Anesthesiologists classification (1−5).
† All types of disability benefit, both full and partly supported.
‡ ODI = Oswestry Disability Index (0−100), indicating increasing disability.
§ more than one level operated.

Sensitivity analyses

Appendix Table 2 displays the multiple regression using ODI change score to define failure and worsening; there were minor differences from the primary analysis.

Subgroup analyses

Predominant preoperative back pain was reported by 1968 patients, of which 307 (16%) received decompression and fusion and 1,661 (84%) received decompression only. In the decompression and fusion group, 125 (41%) reported failure compared to 581 (35%) in the decompression only group. In the predominant leg pain group, 292 (36%) reported failure in the decompression and fusion group versus 1,921 (32%) in the decompression only group. Patients with predominant back pain had an increased risk ((RR) of 1.11 (1.04−1.19); p=.002) of failure.

Appendix table 3 displays failure rates according to the number of levels operated. The proportion of patients that reported treatment failure increased by numbers of spinal levels operated. There were 48% failures reported by patients who had previously received surgery at the same spinal level, compared to 47% for those who were previously operated at another spinal level.

Discussion

In this register study, 33% of patients operated for lumbar spinal stenosis, were classified as failure after surgery, including 22% classified as worse. The strongest predictors for failure were preoperative duration of back pain for at least 12 months, previous spinal surgery, and age above 70 years. Both socioeconomic variables, general health variables, and spine-related variables affected the odds for failure. The same patterns were seen regarding the odds for worsening.

The proportion of patients reported as failure and worsening seemed relatively high and may be partially explained by different outcome measures. For instance, the proportion of patients that perceived themselves as unchanged or worse was lower when patients used GPE, rather than ODI, to assess the effect of surgery. Similar results are reported in the literature. This is not surprising as GPE is conceptually different from a disease specific outcome measure. Previous studies reported success rates of about 62%−75% or failure rates of 25%−31% [7−13,31]. Moreover, the effect of surgery in our study with a mean ODI final score of 24, and a mean ODI improvement of 16 points is also in line with other studies [7−8].

Socioeconomics

A short education, living alone, not being a native Norwegian speaker, and receiving disability benefits increased the odds of failure. The findings of associations between socioeconomic factors and odds for failure and worsening are known from the literature [20]. The effect of these factors in our study was moderate (ORs between 1.33 and 1.67). One former study reported socioeconomic factors as more important than factors related to spine surgery and

general health regarding return to work after spine surgery [32]. The impact of socioeconomics on the results of surgical treatment may seem surprising, but pain and disability are subjective feelings and functions. Hence, they may be affected by patient-related factors. Furthermore, communication is crucial in deciding on surgical treatment for a condition of pain and during clinical follow-up. Consequently, socioeconomic factors may impact the shared decision-making process between patients and health care personnel [33].

Our findings of associations between socioeconomic factors and failure and worsening may contribute to a higher threshold to receive surgical treatment for some patients with socioeconomic challenges. Nevertheless, it is essential to consider equal rights to health care for all patients.

General health

Age >70 years was associated with almost doubled odds for failure and worsening (ORs between 1.93 and 1.99). However, the literature on the effect of age on clinical results after surgery for LSS is conflicting. Some studies find no or minimal association between age and clinical results [34–37]. In contrast, others find decreased risk for success with increased age, or age <75 as a predictor of satisfaction [8,38]. Possible reasons for conflicting results include different outcomes, different ways of defining age groups (continuous data, age groups, or specific cut-offs), or differences in study populations. In addition, high age may be correlated with increased prevalence of other illnesses (ie, osteoarthritis), contributing to the association between age >70 years and increased odds for failure.

Smoking, ASA>2, and BMI>30 also showed associations with failure and worsening (ORs between 1.33 and 1.53). Other studies have found ASA>2 more likely to have poor outcomes [13]. The effect of BMI on the results is more uncertain in the literature. Mauro et al. reported worse outcomes with high BMI, while Onyekwelu et al. reported similar results for patients with BMI > 30 and BMI <30 [39–40].

Using the final ODI score as the outcome, it seems natural that variables concerning general health (smoking, ASA>2, BMI>30) affect the outcome as they presumably reduce function.

Some studies report frailty as a composite variable on general health, and frailty has shown an apparent effect on clinical results and complications after spine surgery and surgery in general [41–42]. We believe the general health condition impacts the clinical outcome and disability after surgery for LSS. However, grading and recording this can be done in different ways. Hence, detecting it can be challenging, especially in a registry setting.

Disease-related factors

Duration of symptoms of >12 months strongly predicted both failure and worsening. Still, long-lasting back pain had a more negative impact than prolonged leg pain. In the multivariate analyses, leg pain > 12 months was either not detectable or moderated. We found an increased risk for failure for patients with preoperative predominant back pain. Surgery for LSS aims to increase the cross-sectional area of the spinal canal to relieve leg pain and improve walking capability. Patients who reporting predominately back prior to surgery could therefore be expected to benefit less of surgery. Previous studies have shown better outcomes among patients with short symptom duration [8,43]. One possible explanation is that prolonged symptoms may lead to biochemical differences in the nerve cells and chronic pain, hence poorer treatment effect [44–45].

Former spinal surgery was a significant predictor of failure and worsening, doubling the odds for these outcomes (ORs 2.00–2.21), dichotomized irrespective of level, that is, including surgery at the same or another segment. In the subgroup analysis, we analyzed previous surgery at the same spinal level versus previous surgery at another spinal level. Interestingly, we found no differences in failure rates between these subgroups. Nerland et al. reported similar ORs for worsening associated with previous surgery in the same segment [36?]. Furthermore, Aalto et al. reported an association between no previous surgery and increased odds for good results (OR=3.65) [8]. Possible explanations for this association can be that previous surgery may lead to scar tissue resulting in technical difficulties in surgery. In addition, patients undergoing repeated surgery may be non-responsive to surgical treatment, often achieving failure and worsening.

Preoperative ODI showed increased OR for failure and worsening. The effect may seem small (OR=1.06–1.07 per ODI point). However, marked differences in preoperative ODI will significantly affect the odds of failure and worsening. Association between preoperative ODI scores and clinical outcomes have been reported before [9,46]. As we define failure and worsening by final ODI scores, the preoperative ODI score seems as a natural predictive factor; it is less likely to achieve a postoperative cut-off with a higher preoperative disability.

Patients who had surgery in more than one level had slightly increased odds for failure and worsening (OR = 1.19–1.21), although not significant in the multivariate analyses. The proportion of patients reporting treatment failure increased by number of spinal levels operated. Two former studies reported no statistically significant differences in outcomes for one and multilevel LSS treated surgically [36,47]. If one level operation has a specific risk for failure, adding the chance for failure per level could be a reasonable way to estimate the risk for failure in multilevel surgery. However, our study did not support such findings.

In our study, radiological findings showed no or negligible associations with failure or worsening. That is in line with previous studies, showing no clinically relevant association between MRI findings and preoperative disability and no or minor association between MRI findings and clinical

outcome [9,48]. Radiologic findings were only recorded as yes /no, and no grading of the radiologic findings was recorded. The validity of radiological data in NORspine has not been reported.

When we examined failure and worsening as defined by ODI change score in the sensitivity analysis, we found that previous surgery, preoperative back pain lasting longer than 12 months, and age above 70 years were the strongest predictors of failure and worsening (Appendix Table 2)

In patients with predominant back pain, fusion, in addition to decompression, did not improve the results.

Future perspectives

There is a need for better instruments predicting outcomes after surgery for LSS. Prediction models have been developed to assist in patient selection. However, Staartjes et al. reported only a moderate ability to identify patients likely to benefit from surgery for degenerative spine disorder. Therefore, they concluded that prediction models should only play a minor role in decision-making [49]. The Swedish spine registry (Swespine) has also developed a prediction tool based on a prediction model to aid in patient selection for spine surgery [39]. From a future perspective, it could be interesting to develop a parallel prediction model based on NORspine data. The prediction model might help select suitable patients for surgery.

Limitations

There are several limits to this study. One is whether the NORspine registry records relevant variables to predict clinical outcomes. The study design does not allow conclusions regarding causality; only associations can be discovered. Some associations might be confounders or mediators connected to causal variables left unobserved. For instance, the NORspine did not record data on spinal alignment during the study period.

Our study had a loss to follow-up of 24.9%; Appendix Table 1 and Table 2 display no significant differences between responders and nonresponders in baseline data or surgical treatment. Although according to former studies, loss to follow-up in national spine registers does not affect the clinical outcome, one of these studies examined the NORspine population [50–52].

One can discuss the choice of the cut-offs for failure and worsening. The cut-offs were assessed in a former study on LSS patients using a transitional scale as an anchor and are in concordance with another survey of cut-offs for success and with the PASS score of 22 proposed by van Hooff [26,53–54]. Our cut-offs result in proportions classified as failure and worsening, similar to other studies [7–13,31]. Failure and worsening have been defined differently in other studies, but the use of MCID regarding increasing ODI score and worsening is not well supported [36]. In the sensitivity analysis, our main findings were confirmed. There are, however, no explicit definitions of failure and worsening after spinal surgery.

Twelve months follow-up might be short for a chronic illness. On the other hand, several studies show no clinically significant differences between 12 and 24 months. We, therefore, consider 12 months as sufficient in LSS patients [50,55–58]. Twelve months follow-up is also recommended in a systematic review with recommendations for spine registries in 2015 [28].

Different surgical techniques were used in this study; both decompression methods and fusion methods varied. Naturally, this introduces heterogeneity in our material, but on the other hand, it reflects the everyday practice and increases the external validity.

Different findings between predictor studies can result from differences in patient selection, surgical techniques, and the selection and recording of possible predictive variables.

Strengths

The study population is large and recruited in a national register obligatory to all treating centers in Norway. Therefore, the patient population reflects everyday practice, and we consider the external validity good. The large sample size also allowed for strict thresholds for entry and removal of covariates into the model of 0.01, improving the power of the analyses.

Interpreting changes in PROM scores, that is, ODI, in groups is not straightforward, and especially in large datasets, one can find statistically significant findings that do not reflect clinical importance. Therefore, we chose to use dichotomous outcomes to define failure and worsening, emphasized in a review article as favorable regarding clinical important change [59].

Conclusion

In this prospective observational spine register study, 33% of patients reported treatment failure, including a worsening rate of 22%, after surgery for LSS. Associated with increased odds for failure and worsening were duration of back pain of more than 12 months, former spinal surgery, and age >70 years. This information can assist in patient information and patient selection for surgery.

Declaration of Competing Interest

The authors declare that they have no known potential conflicts of interest influencing the work with this paper. None of the authors have received financial support to complete this study.

Acknowledgments

We would like to thank Milada Cvancarova Småstuen, Professor at Oslo Metropolitan University, for her invaluable help with the statistics in the study.

Appendix table 1
Patient characteristics of 11,873 Norwegian patients with surgically treated lumbar spinal stenosis in a 10 years period (2007-2017)

	Completed 12 months follow up (n=8919) Mean (95%CI), or n (%)	Lost to follow up (n=2954) Mean (95%CI), or n (%)
Age (cont)	66.6 (66.4–66.9)	63.2 (62.7–63.6)
Female sex	4,644 (52.1%)	1,560 (52.8%)
Civil status, living alone	2,282 (25.7%)	887 (30.2%)
Native Norwegian speaking	8,565 (96.5%)	2,788 (94.7%)
ASA grade >2*	1,840 (20.8%)	622 (21.3%)
Body Mass Index (cont)	27.5 (27.4–27.6)	27.8 (27.6–28.0)
Smoking	1,697 (19.2%)	2,100 (28.1%)
Education level below college	6,145 (70.1%)	2,064 (71.4%)
Comorbidity, any	5,410 (66.9%)	1,833 (68.0%)
Patient not working	7,443 (86.0%)	2,425 (85.1%)
Receives Disability benefit (uføret)	1,421 (15.9%)	511 (17.9%)
Previous spinal surgery, any level	2,173 (24.7%)	795 (27.3%)
Leg pain > 12 months duration	5,284 (64.3%)	1,831 (67.5%)
Back pain > 12 months duration	6,280 (74.9%)	2,135 (76.8%)
Preoperative ODI *	39.8 (39.4–40.1	42.1 (41.6–42.7)
Presoperative NRS leg pain†	6.57 (6.52–6.61)	6.7 (6.6–6.8)
Pre operative NRS back pain	6.50 (6.45–6.55)	6.6 (6.5–6.7)
Pre operative EQ-5D‡	0.376 (0.369–0.383)	0.323 (0.311–0.335)

Patients completed 12 months follow-up, and patients lost to follow-up.
* ASA = American Society of Anesthesiologists classification (1–5).
† ODI = Oswestry Disability Index (0–100).
‡ NRS = Numeric Rating Scale (0–10)****ED-5D = EuroQol's quality of life, (-0.60–1.00)

Appendix table 2
Sensitivity analysis, multivariable logistic regression for 8,919 patients operated for lumbar spinal stenosis and registered in NORspine during 2007–2017, using failure (ODI change <8 points) and worsening (ODI change <4 points) as dependent variables and potential predictors as explanatory variables

	Failure (ODI change <8 points)			Worsening (ODI change <4 points)		
	OR	95%CI	p	OR	95%CI	p
Age >70	1.68	1.47–1.93	<.001	1.61	1.39–1.87	<.001
Smoking	1.39	1.22–1.60	<.001	1.49	1.29–1.72	<.001
BMI>30	1.25	1.11–1.42	<.001			
ASA>2	1.37	1.19–1.56	<.001	1.49	1.22–1.63	<.001
Education level below college	1.45	1.28–1.63	<.001	1.53	1.34–1.75	<.001
Civil status, living alone	1.26	1.11–1.42	<.001	1.26	1.20–1-44	<.001
Disability benefit (all types)	1.49	1.30–1.71	<.001	1.50	1.29–1.75	<.001
Previous surgery,any	1,90	1.68–2.14	<.001	2.01	1.77–2.29	<.001
MRI: central stenosis	0.79	0.71–0.89	<.001	0.81	0.71–0.92	<.001
Degenerative olisthesis (x-ray)	0.69	0.59–0.81	<.001	0.70	0.59–0.84	<.001
Preoprerative ODI (cont)	0.96	0.96–0.97	<.001	0.96	0.96–0.97	<.001
Back pain >12 mnts	1.89	1.65–2.15	<.001	1.71	1.06–1.49	<.001
Leg pain >12 mhts				1.26	1.06–1.49	.008

Appendix table 3
Cross tabulation of number of operated levels and rates of treatment failure defined by ODI final score >31

Number of levels operated	Failure	Nonsssfailure	Total	Proportion failure
1	1,747	3,796	5,543	32%
2	941	1,740	2,681	35%
3	199	301	500	39%
4	23	35	58	39%

References

[1] Amundsen T*, Weber H, Nordal HJ, Magnaes B, Abdelnoor M, Lilleås F. Lumbar spinal stenosis: conservative or surgical management?: a prospective 10-year study. Spine 2000;25:1424–36.

[2] Atlas SJ, Keller RB, Robson DBS, Deyo RA, Singer DE. Surgical and nonsurgical management of lumbar spinal stenosis: four-year outcomes from the maine lumbar spine study. Spine 2000;25:556–62.

[3] Atlas SJ, Keller RB, Wu YA, Deyo RA, Singer DE. Long-term outcomes of surgical and nonsurgical management of lumbar spinal

stenosis: 8 to 10 year results from the maine lumbar spine study. Spine 2005;30:936–43. https://doi.org/10.1097/01.brs.0000158953.57966.c0.

[4] Malmivaara A, Slätis P, Heliövaara M, Sainio P, Kinnunen H, Kankare J, et al. 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.

[5] Weinstein JN, Tosteson TD, Lurie JD, Tosteson ANA, Blood E, Hanscom B, et al. Surgical versus nonsurgical therapy for lumbar spinal stenosis. N Engl J Med 2008;358:794–810. https://doi.org/10.1056/NEJMoa0707136.

[6] Kreiner DS, Shaffer WO, Baisden JL, Gilbert TJ, Summers JT, Toton JF, et al. An evidence-based clinical guideline for the diagnosis and treatment of degenerative lumbar spinal stenosis (update). Spine J 2013;13:734–43.

[7] Nerland US, Jakola AS, Solheim O, Weber C, Rao V, Lønne G, et al. Minimally invasive decompression versus open laminectomy for central stenosis of the lumbar spine: pragmatic comparative effectiveness study. BMJ 2015;350:h1603. https://doi.org/10.1136/bmj.h1603.

[8] Aalto T, Sinikallio S, Kröger H, et al. Preoperative predictors for good postoperative satisfaction and functional outcome in lumbar spinal stenosis surgery — a prospective observational study with a two-year follow-up. Scandinavian J Surg 2012:255–60. https://doi.org/10.1177/145749691210100406.

[9] Sigmundsson Freyr G, Kang Xiao P, Jönsson Bo, Strömqvist Björn. Prognostic factors in lumbar spinal stenosis surgery. Acta Orthopaedica 2012;83(5):536–42 To link to this article: 10.3109/17453674.2012.73391. https://doi.org/10.3109/17453674.2012.733915.

[10] Ogura Y, Kobayashi Y, Shinozaki Y, et al. Factors influencing patient satisfaction after decompression surgery without fusion for lumbar spinal stenosis. Global Spine J 2020;10:627–32. https://doi.org/10.1177/2192568219868205.

[11] Gunzburg R, Keller T, Szpalski M, et al. Clinical and psychofunctional measures of conservative decompression surgery for lumbar spinal stenosis: a prospective cohort study. Eur Spine J 2003;12(2003):197–204. https://doi.org/10.1007/s00586-002-0479-6.

[12] Kleinstück FS, Grob D, Lattig F, Bartanusz V, Porchet F, Jeszenszky D, et al. The influence of preoperative back pain on the outcome of lumbar decompression surgery. Spine 2009;34:1198–203. https://doi.org/10.1097/BRS.0b013e31819fcf35.

[13] Hébert JJ, Abraham E, Wedderkopp N, et al. Preoperative factors predict postoperative trajectories of pain and disability following surgery for degenerative lumbar spinal stenosis. Spine (Phila Pa 1976) 2020;45:E1421–30. https://doi.org/10.1097/BRS.0000000000003587.

[14] Simpson AK, Lightsey IV, >Xiong GX, Crawford AM, Minamide A, Schoenfeld AJ. Spinal endoscopy: evidence, techniques, global trends, and future projections. Spine J 2022;22:64–74.

[15] Sacks GD, Dawes AJ, Ettner SL, Brook RH, Fox CR, Maggard-Gibbons M, et al. Surgeon perception of and benefit in the decision to operate. Ann Surg 2016:896–903. https://doi.org/10.1097/SLA.0000000000001784.

[16] Kahneman D, Sibony O, Sunstein CR. Noise: a flaw in human judgment. London William Collins; 2021.

[17] Meehl PE. Clinical versus statistical prediction: a theoretical analysis and a review of the evidence. University of Minnesota Press; 1954. https://doi.org/10.1037/11281-000.

[18] Grove WM, Zald DH, Lebow BS, Snitz BE, Nelson C. Clinical versus mechanical prediction: a meta-analysis. Psycholo Assess 2000;12:19–30. https://doi.org/10.1037/1040-3590.12.1.19.

[19] Yu MC, Kuncel NR. Pushing the limits for judgmental consistency: comparing random weighting schemes with expert judgments. Personnel Assess Decisions 2020;Number 6(Iss. 2) Article 2. DOI. Available at:. https://doi.org/10.25035/pad.2020.02.002. https://scholarworks.bgsu.edu/pad/vol6/iss2/2. Accessed August 1, 2022.

[20] Yap ZL, Summers SJ, Grant AR, Moseley GL, Karran EL. The role of the social determinants of health in outcomes of surgery for low back pain: a systematic review and narrative synthesis. Spine J 2022;22:793–809. https://doi.org/10.1016/j.spinee.2021.11.013.

[21] Norwegian registry for Spine Surgery (NORspine) Annual report 2017, 2SKDE. Tore K Solberg, Lena Ringstad olsen Universtitetssykehuset I Nord Norge (UNN; 2018. Available at: https://www.kvalitetsregistre.no/sites/default/files/30_arsrapport_2017_ryggkirurgi_1.pdf.

[22] Fairbank JC, Pynsent PB. The oswestry disability index. Spine 2000;25:2940–52.

[23] Grotle M, Brox JI, Vøllestad NK. Cross-cultural adaption of the Norwegian versions of the roland-morris disability questionnaire and the oswestry disability index. J Rehabil Med 2003;35:241–7.

[24] McCaffery M. Pain, clinical manual for nursing practice. Mosby; 1989.

[25] Kamper SJ, Ostelo RWJG, Knol DL, Maher CG, de Vet HCW, Hancock MJ. 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;63:760–6. https://doi.org/10.1016/j.jclinepi.2009.09.009.

[26] Alhaug OK, Dolatowski FC, Solberg TK, Lønne G. Criteria for failure and worsening after surgery for lumbar spinal stenosis: a prospective national registry observational study. Spine J 2021;21:1489–96 criteria DOI:. https://doi.org/10.1016/j.spinee.2021.04.008.

[27] ICHOM. (International Consortium for Health Outcomes Measurements). Standard Set for Low Back Pain. 2014 Available at: http://www.ichom.org/project/low-back-pain/. Accessed July 17, 2022. [Google Scholar].

[28] van hoof ML, Jacobs WC, Willems PC, Wouters MWJM, de Kleuver M, Peul WC, et al. Evidence and practice in spine registries . A systematic review, and recommendations for future design of registries. Acta orthopaedic 2015;86:534–44. https://doi.org/10.3109/17453674.2015.1043174.

[29] van Hooff ML, van Loon J, van Limbeek J, de Kleuver M. The Nijmegen decision tool for chronic low back pain. Development of a clinical decision tool for secondary or tertiary spine care specialists. Plos one 2014. https://doi.org/10.1371/journal.pone.0104226.

[30] von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. Ann Intern Med 2007;147:573–7. PMID: 17938396.

[31] McIlroy S, Jadhakhan F, Bell D, Rushton A. Prediction of walking ability following posterior decompression for lumbar spinal stenosis. Eur Spine J 2021;30:3307–18. https://doi.org/10.1007/s00586-021-06938-6.

[32] Asher AL, Devin CJ, Archer KR, Chotai S, Parker SL, Bydon M, et al. An analysis from the quality outcomes database, part 2. predictive model for return to work after elective surgery for lumbar degenerative disease. J Neurosurg 2017;27:370–81. Available at: https://doi.org/10.3171/2016.8.SPINE16527. https://thejns.org/spine/view/journals/j-neurosurg-spine/27/4/article-p370.xml.

[33] Stewart MA. Effective physician-patient communication and health outcomes: a review. CMAJ 1995;152:1423–33. PMID: 7728691.

[34] Sobottke R, Herren C, Siewe J, Mannion AF, Röder C, Aghayev E. Predictors of improvement in quality of life and pain relief in lumbar spinal stenosis relative to patient age: a study based on the Spine Tango registry. Eur Spine J 2017;26:462–72. https://doi.org/10.1007/s00586-015-4078-8. Epub 2015 Jul 3. PMID: 26138216.

[35] Claus CF, Tong D, Lytle E, Bahoura M, Garmo L, Li C, et al. Age as a predictor for complications and patient-reported outcomes in multilevel transforaminal lumbar interbody fusions: analyses from the michigan spine surgery improvement collaborative (MSSIC). Spine (Phila Pa 1976) 2021;46:356–65 15PMID: 33620179. https://doi.org/10.1097/BRS.0000000000003792.

[36] Nerland US, Jakola AS, Giannadakis C, Solheim O, Weber C, Nygaard ØP, et al. The risk of getting worse: predictors of deterioration after decompressive surgery for lumbar spinal stenosis: a multicenter observational study. World Nerosrug 2015;84:1095–102. https://doi.org/10.1016/j.wneu.2015.05.055.

[37] Giannadakis C, Solheim O, Jakola AS, Nordseth T, Gulati AM, Nerland US, et al. Surgery for lumbar spinal stenosis in individuals aged 80 and older: a multicenter observational study. J Am Geriatr Soc 2016;64:2011–8. https://doi.org/10.1111/jgs.14311. Epub 2016 Sep 9. PMID: 27611928.

[38] Fritzell P, Mesterton J, Hagg O. Prediction of outcome after spinal surgery-using the dialogue support based on the Swedish national register. Eur Spine J 2021:27. https://doi.org/10.1007/s00586-021-07065-y. Epub ahead of print. PMID: 34837113.

[39] Onyekwelu I, Glassman SD, Asher A, Shaffrey CI, Mummaneni PV, Carreon LY. Impact of obsiety on complications and outcomes: a comparison of fusion and nonfusion lumbar spine surgery. J Neurosurg 2017;26:158–62.

[40] Radcliff KE, Rihn J, Hilibrand A, DiIorio T, Tosteson T, Lurie JD, et al. Does the duration of symptoms in patients with spinal stenosis and degenerative spondylolisthesis affect outcomes?: analysis of the Spine Outcomes Research Trial. Spine 2011;36:2197–210. https://doi.org/10.1097/BRS.0b013e3182341edf.

[41] Panayi AC, Orkaby AR, Sakthivel D, Endo Y, Varon D, Roh D, et al. Impact of frailty on outcomes in surgical patients: a systematic review and meta-analysis. Am J Surg 2019;218:393–400. https://doi.org/10.1016/j.amjsurg.2018.11.020.

[42] Veronesi F, Borsari V, Martini L, Visani A, Gasbarrini A, Brodano GB, et al. The impact of frailty on spine surgery: systematic review on 10 years clinical studies. Aging Dis 2021;12:625–45. Published 2021 Apr 1. https://doi.org/10.14336/AD.2020.0904.

[43] ChaoT PK, Steward O, Gupta R. Chronic nerve compression injury induces a phenotypic switch of neurons within the dorsal root ganglia. J Comparat Neurol 2008;506:180–93. https://doi.org/10.1002/cne.21537.

[44] Gupta R, Rowshan K, Chao T, Mozaffar T, Steward O. Chronic nerve compression induces local demyeliniation and remyelination in a rat model of carpat tunnel syndrome. Experimental Neuro 2004;187:500–8. https://doi.org/10.1016/j.expneurol.2004.02.009.

[45] Fauzi A, Gatham R, Phedy P, Ajiantoro A, Gatam L, Mahadipta H, et al. Multivariate analysis of factors that influence the clinical outcome oswestry disability index in patients with lumbar spinal stenosis post percutaneous stenoscopic lumbar decompression. Biomed Pharmacol J 2021;14. https://doi.org/10.13005/bpj/2125.

[46] Park DK, An HS, Lurie JD, Zhao W, Tosteson A, Tosteson TD, et al. Does multilevel lumbar stenosis lead to poorer outcomes?: a subanalysis of the Spine Patient Outcomes Research Trial (SPORT) lumbar stenosis study. Spine 2010;35:439–46. https://doi.org/10.1097/BRS.0b013e3181bdafb9.

[47] Aaen J, Austevoll IM, Hellum C, Storheim K, Myklebust TÅ, Banitalebi H, et al. Clinical and MRI findings in lumbar spinal stenosis: baseline data from the NORDSTEN study. Eur Spine J 2022;31:1391–8. https://doi.org/10.1007/s00586-021-07051-4.

[48] Weber C, Giannadakis C, Rao V, Jakola A, Nerland U, Nygaard Ø, et al. Is there an association between radiological severity of lumbar spinal stenosis and disability, pain, or surgical outcome? Spine 2016;41:E78–83. https://doi.org/10.1097/BRS.0000000000001166.

[49] Staartjes VE, Stumpo V, Ricciardi L, Maldamer N, Eversdijk HAJ, Vieli M, et al. FUSE-ML: development and external validation of a clinical prediction model for mid-term outcomes after lumbar spinal fusion for degenerative disease. Eur Spine J 2022. https://doi.org/10.1007/s00586-022-07135-9.

[50] Endler P, Ekman P, Hellström F, Möller H, Gerdhem P, et al. Minor effect of loss to follow-up on outcome interpretation in the Swedish spine register. Eur Spine J 2020;29:213–20. https://doi.org/10.1007/s00586-019-06181-0.

[51] Højmark K, Støttrup C, Carreon L, Andersen MJESJ. Patient-reported outcome measures unbiased by loss of follow-up. Single-Center study based on Dane Spine. Danish Spine Surg Registry 2016;25:282–6. https://doi.org/10.1007/s00586-015-4127-3.

[52] Solberg TK, Sørlie A, Sjaavik K, Nygaard ØP, Ingebrigtsen T. Would loss to follow-up bias the outcome evaluation of patients operated for degenerative disorders of the lumbar spine? Acta Orthopaedica 2011;82:56–63. https://doi.org/10.3109/17453674.2010.548024.

[53] Austevoll IM, Gjestad R, Grotle M, Solberg T, Brox JI, Hermansen E, 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;20:31. https://doi.org/10.1186/s12891-018-2386-y.

[54] van Hoof ML, Mannion AM, Staub LP, Osteolo RWJG, Fairbank JCT. 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.

[55] Weinstein JN, Tosteson TD, Lurie JD, Tosteson A, Blood E, Herkowitz H, et al. Surgical versus nonoperative treatment for lumbar spinal stenosis four-year results of the Spine Patient Outcomes Research Trial. Spine (Phila Pa 1976) 2010;35:1329–38. https://doi.org/10.1097/BRS.0b013e3181e0f04d.

[56] Parai C, Hägg O, Lind B, Brisby H. Follow-up of degenerative lumbar spine surgery-PROMs stabilize after 1 year: an equivalence study based on Swespine data. Eur Spine J 2019;28:2187–97. https://doi.org/10.1007/s00586-019-05989-0.

[57] Weinstein JN, Lurie JD, Tosteson TD, Zhao W, Blood A, Tosteson ANA, 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;91:1295–304. https://doi.org/10.2106/JBJS.H.00913.

[58] Lønne G, Johnsen LG, Rossvoll I, Andresen H, Storheim K, Zwart JA, et al. Minimally invasive decompression versus x-stop in lumbar spinal stenosis: a randomized controlled multicenter study. Spine (Phila Pa 1976) 2015;40:77–85. https://doi.org/10.1097/BRS.0000000000000691.

[59] 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.

ISBN 978-82-326-7122-9 (trykt utg.)
ISBN 978-82-326-7121-2 (elektr. utg.)
ISSN 1503-8181 (trykt utg.)
ISSN 2703-8084 (online ver.)

NTNU
Kunnskap for en bedre verden