Enhancing Quality in Surgery for Cervical Degenerative Disorders: Benchmarks for Clinical Improvement and Prognostic Models for Nonsuccess

Doctoral thesis by
Christer Mjåset

Institute of Clinical Medicine, Faculty of Medicine, University of Oslo

Department of Research and Innovation, Division of Clinical Neuroscience, Oslo University Hospital

Department of Neurosurgery, Oslo University Hospital

June 2023

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

Contents

Acknowledgements

This thesis was performed during my stay at the Department of Research and Innovation, Division of Clinical Neuroscience, and the Department of Neurosurgery at Oslo University Hospital (OUS). First, I would like to specifically thank my main supervisor, professor Margreth Grotle, who has given me invaluable support during the work process. Her sincere and outmost interest in my work, her empathy and patience, as well as her knowledge in the field of musculoskeletal health have all been crucial factors in the making of this doctorial work.

Second, I would like to thank my co-supervisors. As former chief of spine surgery and now head of the Department of Neurosurgery at OUS, Dr. Frode Kolstad was the one that encouraged me to enter the field of musculoskeletal research. During the working process, he contributed with his invaluable clinical knowledge in neurosurgery, and his many motivational remarks helped me stay focused and not lose faith that I would someday finish my thesis. The encouragement and support of professor and head of research at the Division of Clinical Neuroscience, Dr. John-Anker Zwart, was also extremely important during the working process. Dr. Zwart always kept the door open for shorter or longer discussions about the project and warmly included me in the working environment in his division.

Third, my conversations with and guidance from Dr. Tore Solberg, professor at The Arctic University of Norway and consultant neurosurgeon at The University Hospital at Northern Norway, have been of upmost importance for this thesis. Although based in Tromsø, Dr. Solberg has always been available for advice and discussion and has been involved in every strategic decision, especially the methodological considerations related to this doctoral thesis.

the doctoral thesis – has been an invisible member of the family for far too long now, and I believe we are all happy to have it leave the house.

Still, for me this has been a great journey! Thanks again for sharing it with me!

Christer Mjåset

Oslo, 2023

List of papers

PAPER 1

Mjåset C, Zwart JA, Goedmakers CMW, Smith TR, Solberg TK, Grotle M. Criteria
for success after surgery for cervical radiculopathy-estimates for a substantial amount
of improvement in core outcome measures. Spine J. 2020 Sep;20(9):1413-1421. doi:
10.1016/j.spinee.2020.05.549. Epub 2020 Jun 2. PMID: 32502657.

PAPER 2

Mjåset C, Zwart JA, Kolstad F, Solberg T, Grotle M. Clinical improvement after
surgery for degenerative cervical myelopathy; A comparison of Patient-Reported
Outcome Measures during 12-month follow-up. PLoS One. 2022 Mar
8;17(3):e0264954. doi: 10.1371/journal.pone.0264954. PMID: 35259164; PMCID:
PMC8903279.

PAPER 3

Mjåset C, Solberg TK, Zwart JA, Småstuen MC, Kolstad F, Grotle M. Anterior
surgical treatment for cervical degenerative radiculopathy: a prediction model for
non-success. Acta Neurochir (Wien). 2023 Jan;165(1):145-157. doi: 10.1007/s00701-
022-05440-2. Epub 2022 Dec 8. PMID: 36481873; PMCID: PMC9840586.

Abbreviations

ACDA: Anterior Cervical Disc Arthroplasty

ACDF: Anterior Cervical Discectomy and Fusion

AUC: Area under the Receiver Operating Characteristic Curve

BMI: Body Mass Index

BP: Blood pressure

CI: Confidence Interval

COSMIN: Consensus-Based Standards for the Selection of Health Status Measurement Instruments

DRAM: Distress and Risk Assessment Method

EMS: European Myelopathy Score

EQ-5D: EuroQol 5-Dimension-Questionnaire

EQ-VAS: Health-Related Quality-of-Life by EuroQol

GPE scale: Global Perceived Effect Scale

MCID: Minimal Clinically Important Difference

MDC: Minimal Detectable Change

MIC: Minimal Important Difference

mJOA scale: modified Japanese Orthopedic Association scale

MRI: Magnetic Resonance Imaging

NDI: Neck Disability Index

NORspine: Norwegian Registry for Spine Surgery

NRS-AP: Numerical Rating Scale arm pain

NRS-NP: Numerical Rating Scale neck pain

ODI: Oswestry Disability Index

OUS: Oslo University Hospital

PDQ: Pain/Disability Questionnaire

PHQ-9:   Patient Health Questionnaire-9

PROBAST: Prediction Model Risk of Bias Assessment Tool

PROGRESS: Prognosis Research Strategy

REK: Regional Committee for Medical and Health Research Ethics

RCT: Randomized Controlled Trial

ROC curve: Receiver Operating Characteristic curve

SCB: Substantial Clinical Benefit

STROBE: Strengthening the Reporting of Observational studies in Epidemiology.

TRIPOD: Transparent Reporting of a Multivariable Prediction Model for

            Individual Prognosis or Diagnosis

VBHC: Value-Based Health Care

UNN: University Hospital of North Norway

Summary in English

Background

Recent years surgical treatment modalities for cervical degenerative disorders have become increasingly more safe and less invasive, and surgical volumes are rising. Still, how, when and to what extent surgery can contribute to symptom relief and physical recovery for degenerative cervical radiculopathy and myelopathy patients is debated.

By increasing our understanding of criteria for clinical improvement in patients operated for cervical degenerative disorders, better patient counseling and shared decision-making processes prior to surgery can be achieved. With the comprehensive use of prognostic models, improved selection of appropriate candidates for surgery can also be obtained.

Aim

The overall aim of this thesis was to contribute to raise the quality of Norwegian spine surgery through the assessments of benchmarks for clinical improvement and prognostic models. Specific aims were to develop cutoff estimates for success in patients undergoing surgery for cervical radiculopathy, to assess cutoff estimates for Minimal Clinical Important Difference (MCID) in patients operated due to cervical myelopathy and to present proportions of patients achieving a MCID after surgery. Finally, we wanted to develop and internally validate prognostic models for nonsuccess in disability and arm pain in patients operated for cervical radiculopathy.

Methods and results

The analyses in this thesis were based on patient data collected through the Norwegian Neck Registry (NORspine) from 2011 until 2016. We investigated 4229 patients operated for either cervical radiculopathy and/or myelopathy and followed for 3 and 12 months after surgery.

In Paper I, we assessed criteria for success following one- or two-level surgery for cervical radiculopathy using commonly and well-validated Patient-Reported Outcome Measures (PROMs). We also analyzed subgroups of patients operated with either anterior or posterior approach technique and patients operated due to either a disc herniation or foraminal spondylotic changes in one level. Among the collected PROMs, we found that Neck Disability Index (NDI) and Numeric Rating Scale for arm pain (NRS-AP) had the highest discriminative ability. We also found that percentage change and follow-up/absolute scores were more accurate than change scores. There were minor differences in criteria for success between 3- and 12-month follow-up and across subgroups.

In Paper II, we assessed estimates for MCID in patients with cervical myelopathy and for subgroups of patients operated with either anterior or posterior surgical techniques. NDI and Numeric Rating Scale for neck pain (NRS-NP) showed to be the most accurate PROMs for the whole cohort and across subgroups. Again, percentage change scores were more accurate than change scores. MCID estimates were slightly lower for the posterior approach patients compared to those operated with anterior techniques. Analyzing the proportion of patients achieving MCID for percentage change scores, we found that a majority of patients improved after surgery (51-61%).

In Paper III, we wanted to create and internally validate prognostic models for lack of success in patients undergoing anterior surgery for degenerative cervical radiculopathy based on NDI and NRS-AP cutoff estimates generated in Paper I. Approximately 38% of patients receiving surgery did not achieve an improvement in neck disability after 12 months, while the proportion was 35% for arm pain. We developed a prognostic model for neck disability which included physical demanding work, low level of education, pending litigation, previous neck surgery, duration of arm pain>3 months, medium to high levels of baseline disability, as well as anxiety/depression. Area under the Receiver Operating Characteristic Curve (AUC) was 0.78 (95% CI: 0.75, 0.82). The prognostic model for arm pain included physical demanding work, low level of education, pending litigation, previous neck surgery, duration of arm pain>3 months, medium to high levels of baseline disability, foreign mother tongue, smoking, and medium to high levels of baseline arm pain. AUC was 0.68 (95% CI: 0.64, 0.72). The calibration plots showed no signs of over- or underestimation of the prognostic models.

Conclusions

Our investigations showed that cutoff estimates for clinical improvement can be created for patients operated for cervical degenerative disorders and that NDI is the superior PROM in terms of discriminative ability. Percentage change scores are also more accurate than mean change scores. For cervical radiculopathy, cutoff estimates are comparable across subgroups of patients. For cervical myelopathy, cutoff estimates for posterior surgery are slightly lower than those for anterior surgery indicating lower expectations among the former patient group. Two prognostic models for disability and arm pain showed good to acceptable accuracy. Based upon

the models, individualized risk estimates can be made and applied in shared decision-making processes prior to surgery. However, further validation is needed before the models can be applied in a different environment.

Sammendrag på norsk

Bakgrunn

Kirurgiske behandlingsmetoder for degenerative nakkeplager har de senere årene blitt stadig sikrere og mindre invasive, samtidig som de benyttes hyppigere enn før. Selv om det eksisterer flere retningslinjer for både operativ og konservativ behandling, er det behov for mer kunnskap om hvordan, når og i hvilket omfang kirurgi kan bidra til symptomlindring hos pasienter med cervikale degenerative lidelser.

Ved å øke vår kunnskap om kriterier for klinisk forbedring etter kirurgi for degenerative nakkeplager og om hvilke variabler som er assosiert med et dårlig utfall, kan man oppnå bedre forventningsstyring hos pasienter som er kandidater for operative inngrep og forbedre seleksjonen av pasienter som henvises til kirurgi.

Mål

Det overordnede målet med denne avhandlingen var å bidra til å heve kvaliteten på kirurgi for degenerative nakkeplager i Norge ved å frembringe kunnskap om utfallsmål og prognose. Spesifikke mål var å utvikle terskelverdier for suksess for pasienter operert for cervikal radikulopati, fastsette estimater for minste klinisk viktige forskjell – Minimal Clinically Important Difference (MCID) – for pasienter operert for cervikal myelopati, samt angi andelen av pasienter som oppnår MCID etter kirurgi. Til slutt ønsket vi å utvikle og internvalidere prognostiske modeller for funksjon og armsmerte for pasienter operert for cervikal radikulopati.

Metode og resultater

Analysene ble basert på pasientdata fra Norsk Nakke- og Ryggregister fra 2011 til 2016. 4229 pasienter operert for enten cervikal radikulopati og/eller cervikal myelopati fikk oppfølging tre og tolv måneder etter kirurgi.

I artikkel I fastsatte vi kriterier for kirurgisk suksess hos pasienter som ble operert i ett eller to nivåer på grunn av cervikal radikulopati, ved å bruke validerte pasientrapporterte utfallsmål (PROMs). Vi analyserte også subgrupper av pasienter operert med enten fremre eller bakre dekompresjon og pasienter som enten ble operert for skiveprolaps eller foraminale stenotiske forandringer i nakken. Vi fant at Neck Disability Index (NDI) og Numeric Rating Scale for arm smerte (NRS-AP) hadde de beste diskriminerende evnene som suksesskriterier. Prosentvis endringsscore og absoluttscore ved 3 og 12 måneders oppfølging hadde bedre diskriminerende evne enn endringsscore. Det var helt marginale forskjeller mellom suksesskriterier tre og tolv måneder etter kirurgi og på tvers av subgrupper.

Ved bruk av PROMdata fant vi i artikkel II estimater for MCID hos pasienter operert for cervikal myelopati og for subgrupper av pasienter som ble operert med fremre og bakre tilgang. NDI og Numeric Rating Scale for nakkesmerte (NRS-NP) var best til å diskriminere mellom pasienter som oppnådde MCID og de som ikke gjorde det. Prosentvis endringsscore hadde også bedre diskriminerende evne enn endringsscore. MCID-estimatene var noe lavere for pasienter operert med bakre tilgang enn de som ble operert med fremre tilgang. Majoriteten av pasientene oppnådde MCID-terskelverdier (51-61%) og opplevde dermed en forbedring etter kirurgi.

I artikkel III ble prognostiske modeller for mangel på suksess tolv måneder etter nakkekirurgi for cervikal radikulopati utviklet og internvalidert. Utfallsmålene

var basert på terskelverdier fra artikkel I for NDI og NRS-AP. 38% av pasientene som ble operert oppnådde ikke forbedring tilsvarende terskelverdien for NDI, mens andelen var 35% for armsmerte. Den prognostiske modellen for NDI etter tolv måneder inkluderte fysisk krevende arbeid, lavt utdanningsnivå, utestående rettstvist eller forsikringssak, tidligere degenerativ nakkekirurgi, armsmerte over tre måneder, middels til høye basalnivåer av NDI og angst/depresjon. Den prognostiske modellen for armsmerte inkluderte fysisk krevende arbeid, lavt utdanningsnivå, utestående rettstvist eller forsikringssak, tidligere degenerativ nakkekirurgi, armsmerte over tre måneder, middels til høye basalmålinger av NDI, fremmedspråklighet, røyking og middels til høye basalnivåer av NRS armsmerte. Presisjonen til hver av de to modellene ble målt med arealet under kurven (Receiver Operating Characteristic curve), og over/underestimering av modellene ble undersøkt med kalibreringsplott. Arealet under kurven for NDI var 0.78 (95% konfidensintervall: 0.75, 0.82), mens det var 0.68 (95% konfidensintervall: 0.64, 0.72) for armsmerte. Kalibreringsplottene viste ingen over- eller underestimering av de to prognostiske modellene.

Konklusjon

Våre studier viste at grenseverdier for klinisk forbedring basert på flere godt validerte PROMs kan fremstilles for pasienter operert for cervikale degenerative lidelser og at NDI hadde de klart beste diskriminerende evnene for både cervikal radikulopati og myelopati. Prosentvis endringsscore og absoluttscore hadde også bedre diskriminerende evne enn endringsscore. Terskelverdiene for kirurgisk suksess var sammenlignbare på tvers av subgrupper av pasienter med radikulopati. Myelopatipasienter som ble operert med bakre tilgang, hadde noe lavere terskelverdier for MCID enn de som ble operert med fremre tilgang, noe som

indikerer lavere forventninger hos den førstnevnte gruppen med pasienter. To prognostiske modeller for NDI og NRS armsmerte viste god til akseptabel presisjon med tanke på å predikere mangel på suksess tolv måneder etter kirurgi. Basert på disse modellene kan man lage individualiserte risikoestimater for pasienter som vurderes for kirurgi, til bruk i klinisk praksis. Videre valideringsarbeid er nødvendig for at modellene skal kunne brukes i en annen kontekst eller et annet miljø.

Preface

What are the outcomes of my delivered treatments and why do some patients improve, and some do not? The very ambition to start this project was grounded in these questions. I had for several years seen patients suffering from degenerative neck disorders in both the neurology and neurosurgery wards of Norwegian hospitals. It's fair to say that the patients were a mixed cohort, constituted by everything from young males in their thirties experiencing their first hospitalization, to old women in their nineties with a complex set of diseases wondering what they were doing in the emergency room in the first place. Also, the symptoms that patients presented varied greatly – from intense pain and limb paralysis to only slight imbalance or mild neck pain. Apart from the forever important neurological examination, the MRI could sometimes feel like a heaven-sent diagnostic tool providing a delicate and strikingly clear visualization of the illness and its degree of seriousness. For a young resident in neurology or neurosurgery, the treatment of choice often looks obvious when gazing through the sagittal or coronal scans of the patient's spine. However, as years passed, it became apparent that surgical success could not be revealed through imaging techniques. A fair number of my patients did not become well from surgery even though the preoperative MRI scan strongly suggested surgical intervention as the treatment of choice. Still, as I advanced to become the responsible surgeon, I found myself struggling to understand which patients would benefit the most from an intervention and which patients would be better off treated conservatively.

Every surgeon that I know is driven by the purpose to deliver excellent care for their patients, and, at the same time, continuously become even better at what they do. However, to become a better surgeon you need to know how good you are in the first place. It is in surgery as it is in bowling, you do recognize when you get a strike

or a hard spare (no pins go down). However, both a surgeon and a bowler would like to know the total amount of points you get each game of ten rounds. And it is vital for both to know how you average over time and how often you hit a gutter ball without striking any pins.

This is why the Norwegian Registry for Spine Surgery (NORspine) is such an important initiative. Collecting data from every neurosurgical institution in Norway, NORspine can help create benchmarks and statistical evidence for use in clinical practice. The data can help us improve the care given, as well as answer questions and provide perspectives that we are not able to obtain from the operating theatre. When you know the outcomes and results of your own treatments, that is when you can take your clinical role to the next level. That is when you become a real surgeon.

Even though I careerwise have moved from the operating table to an administrative position, my focus on improving outcomes and creating transparency has endeavored. "Knowing your outcomes" have become something I often repeat in different circumstances as a mantra. The expression captures the fact that it is vital to be constantly striving for improvement in any field you are pursuing.

For this purpose, I am thankful if this dissertation can contribute to an increase in focus on quality of care than what currently is the case in Norwegian health care. I also hope it can help create more transparency about results and outcomes – for the best of both clinicians and patients.

1. Introduction

1.1 Challenges in care delivery for cervical degenerative disorders

Degenerative neck surgery has with recent advances in technique and equipment become high-volume procedures across clinics and hospitals around the world [1]. Still, there is a lack of knowledge about various outcomes and effects from neck surgery, benchmarks for success, and different predictors for successful and even failed outcomes. In today's healthcare environment, these issues are getting increasingly important to clarify, as the patient population is growing older, and the current care delivery systems are not sustainable in terms of costs and size of work force [2]. With more evidence-based information regarding patient outcomes and effects from neck surgery, improved appropriateness of care can be achieved through better selection of patients prior to medical intervention. Also, by understanding better what outcomes matter to individual patients, healthcare can move to towards a system that incentivizes and rewards improved care quality by tying payments to outcomes and other quality metrics. This is the concept of "value-based health care", where "value" is defined as "the measured improvement in a patient's health outcomes for the cost of achieving that improvement" [3].

This doctoral project is the first of its kind to investigate outcomes after cervical degenerative neck surgery reported to The Norwegian Registry for Spine Surgery (NORspine). Based on data from the day of surgery and at 3- and 12-month follow-up collected between 2011 and 2016, the aim of this thesis is to define thresholds for a clinically important improvement in patients operated for degenerative cervical radiculopathy or myelopathy by using scores from widely used Patient-Reported Outcome Measures (PROMs). Additionally, we wanted to create

prognostic models for nonsuccessful outcomes among patients undergoing surgery for degenerative cervical radiculopathy. The results will be discussed in the light of the current health care situation and the potential for spine surgery to move from "volume" to more "value" for patients. With this in mind, the main aim of this thesis is to contribute to raise the quality of Norwegian spine surgery through the contributions of the presented benchmarks and prognostic models.

1.2 Cervical degenerative disorders – diagnosis and clinical presentation

Cervical degenerative changes can present itself in a multitude of ways. Patients report everything from slight symptoms to severe regional pain in the neck, head, or limbs, and, in more serious occasions, neurological deficits. Indications for surgery are tied to presentation of such neurological symptoms as radiculopathy and myelopathy [4].

Degenerative cervical radiculopathy (abbreviated to cervical radiculopathy) is a result of the cervical nerve root being impinged by space occupying processes resulting from degenerative changes in the spine. The typical clinical presentation is arm pain and sensation loss along the innervation path of the affected nerve root, as well as loss of corresponding deep tendon reflexes. Axial neck pain is also sometimes present. In more serious cases, the motor strength of the innervated muscle is compromised creating a paresis or even paralysis in the affected arm muscle [5].

Degenerative cervical myelopathy (abbreviated to cervical myelopathy) occurs when degenerative osseocartilaginous changes cause compression of the spinal cord. Symptoms can be highly variable between patients despite similar radiological findings. Progression can also vary between a sudden onset to a more subtle, gradual presentation of symptoms. Patients typically report clumsiness in the hands, pain

and/or sensitivity loss in the dermatomes associated with the affected cervical level, as well as axial pain or neck stiffness [6, 7]. As a result of spinal stenosis, extension or flection of the neck may cause electrical sensations radiating down the spine and limbs (Lhermitte's sign) [8]. In the case of more serious spinal cord compression, poor coordination, imbalance and bladder/bowel problems can occur [9]. However, studies have shown that radiological findings coherent with cord compression do not necessarily mean that symptomatic myelopathy is present [10-12], and since initial symptoms can be subtle, delayed diagnosis is common [13].

Although the clinical pattern varies, cervical degenerative myelopathy is considered a progressive disorder [14].

1.3 Pathophysiology

The most clinically relevant changes in cervical degenerative disease are the deterioration of the cervical discs, ligamentum flavum and facet joints (spondylosis). The natural process of degeneration starts with the disc's decreasing ability to retain water from a 90% content level at 30 years to 70% at the age 80 years [15]. In addition, poor nutrition, or damage to the avascular discs, which metabolism is relying on diffusion from cartilaginous vertebral endplates, may cause disc inflammation and further degeneration. Thus, the gel-like nucleus pulposus begins to fibrose resulting in height reduction, disc bulging and herniation. With the lower disc height, greater mechanical stress is inflicted on the local facet joints and ligaments causing formation of osteophytes and folding of the ligamentum flavum [16, 17]. In time, the degenerative changes lead to narrowing of the spinal canal or the vertebral foramina, and, subsequently, mechanical compression and inflammation of neural tissue. In particular, the folding and hypertrophy of the ligamentum flavum, and, in

rarer cases, ossification of the posterior longitudinal ligament, may cause compression of the spinal cord [18, 19].

While mechanical compression of axons creates sensory or motor dysfunction in radiculopathy, the pathophysiological mechanism of pain is thought to be different. Axon compression alone has been shown not to elicit pain, although dorsal root ganglion compression do [20]. Thus, an active inflammation response surrounding the nerve root is thought to contribute to the radicular pain sensations in degenerative cervical radiculopathy [21]. In a herniated disc, cytokines are expressed by nucleus pulposus cells. The cytokines cause a catabolic response with recruitment of macrophages and mast cells leading to phagocytosis and resorption of herniated disc material, as well as infiltration of blood vessels [22, 23]. They also cause increased levels of nerve growth factor, which, again, contribute to nerve ingrowth into the disc. In the disc, loss of structure and abnormal motion may lead to mechanical stimuli and, eventually, cervical discogenic pain [24].

In myelopathy, the symptoms vary depending on which level and structural part of the spinal cord is affected. Compression of the lateral corticospinal tracts compromise skeletal muscle control, while pressure on spinocerebellar tracts affect proprioception. The posterior columns, on the other hand, control the ipsilateral position and vibration sense, i.e., proprioception. Together, these deficits lead to the wide-based spastic gait and upper extremity clumsiness that is a classic symptom of cervical myelopathy. In terms of pain, patients often have symptoms from the disc and facet joints, as well as pain due to radiculopathy as described above. Also, impingement of the spinothalamic tracts can cause contralateral extremity pain and alter temperature sensation in the skin [25, 26].

Withstanding compression on the spinal cord can cause nerve cell death and gliosis and permanent symptoms. An important factor is the chronic reduction of the intraparenchymal blood flow which leads to a secondary ischemic injury of the spinal cord and accumulation of microglia cells and a neuroinflammatory reaction. Also, the mechanical stretching of the cord has been found to activate key biological events and cause neural degeneration. This compression-mediated cascade of events, evidently, leads to the development of cell death and gliosis, and typically atrophy of the anterior horns associated with motoneuronal loss [27].

1.4 Epidemiology

The degeneration of the osseocartilaginous components in the cervical spine – also called cervical spondylosis or degenerative disc disease – increases with age [28]. Halfway through life, 80-90% of people show signs of wear and tear of the intervertebral discs [6], many of them asymptomatic [29]. Clinical presentations of degenerative changes are more common in men than in women [30, 31], and both genetic predispositions and occupational and activity-related factors play a role in the degenerative process [32].

Although only a subset of patients with radiological degenerative findings present symptoms, the global burden of neck disease is substantial. Worldwide more than a third of a billion people have had mechanical neck pain of at least 3 months duration according to a 2015 review [33]. In the global study of the burden of disease for young people, low back and neck pain were among the top causes of years lost to disability [34].

Several population-based studies have investigated the incidence rate of cervical radiculopathy. In one of the most cited epidemiological studies in terms of

cervical degenerative disorders, the authors investigated patients in the state of Minnesota from 1976 to 1990 and estimated the annual incidence rate to be 107.3 per 100,000 (0,1%) among males and 63.5 per 100,000 for females (0.064%) [35]. A 2020 systematic review of nine articles, found the incidence rate to be 83.2 to 179.0 per 100,000 adults (0.083-0.18%) [36].

For myelopathy, prevalence and incidence rates have been difficult to assess. A recent review of literature found no epidemiological data, but estimated the prevalence of surgically treated myelopathy patients to be 1.6 per 100,000 inhabitants (0,0016%) [37]. Another study investigating 12-year data from a national database in Taiwan found the incidence of related hospitalizations to be 4.04 per 100,000 person-years (0,004%) [38]. The incidence rate also seems to be rising along with the increasing elderly population. In the United States, the annual rate of myelopathy patients admitted from the emergency department more than doubled from 1993 to 2002 (3.73 to 7.88 per 100,000 or 0.004-0.008%) [39].

In Norway, surgery rates for cervical spine disorders on a national level have been assessed in two different studies. Kristiansen et al. reported an increase in rates from 16.9 to 29.4 procedures per 100,000 inhabitants per year (0.017-0.030%) in a study investigating data from 2008 to 2014 [40]. In a recent publication, Ingebrigtsen et al. found the rate to be stable from 2014 to 2018 although rates differed across regions [41].

1.5 Management of cervical degenerative disorders

Recent decades, there has been a rapid development in the safety of and access to operative management, and surgical treatment of both cervical radiculopathy and myelopathy is commonly applied [42, 43]. Still, how, when and to what extent

surgery can contribute to symptom relief and physical recovery is debated. While many guidelines exist in terms of operative and nonoperative care, there is still a need for more evidence-base knowledge regarding treatment strategies and surgical indications [6].

1.5.1 Nonoperative treatment options

Nonoperative treatment is often referred to as *conservative treatment.* There are several treatment modalities in use today in relation to cervical degenerative disease. The most common consist of noninvasive interventions (i.e., drug treatment, regular counseling about everyday activities), and physical therapy (i.e., exercise and traction with or without manipulation of the neck), while acupuncture, and invasive therapies like epidural injections are applied to a lesser extent. The rationale behind the different approaches varies. However, natural or spontaneous improvement over time is a main explanation for relief of symptoms and recovery after weeks or months without surgical intervention [5]. Since the focus of this dissertation is surgical treatment, conservative alternatives will only be briefly described.

For radiculopathy patients, nonoperative treatment options are the preferred initial approach since up to 75-83%% of acute episodes end with spontaneous improvement [44, 45]. *Medical drugs* are often first-line treatment. Because of evidence suggesting that inflammatory mediators are released from a herniated disc, there is a rationale for the use of anti-inflammatory drugs (NSAIDs) or even steroid treatment [46]. Antiepileptic drugs and tricyclic antidepressants can also be warranted, especially for patients with longstanding pain symptoms [14]. However, guidelines do conclude that no adequate evidence currently exists regarding the effectiveness of pharmacologic treatment of cervical radiculopathy [47].

Although *physical therapy,* such as exercise treatment, is commonly used in relation to cervical radiculopathy, it is debated whether it has any actual impact on the clinical course. Systematic reviews have shown that patients experience pain/symptom relief and improve physical function but highlights the need for randomized controlled trials (RCTs) and conclude that no current studies adequately address physical therapy in treatment of cervical radiculopathy [47, 48]. Accordingly, in terms of physical therapy following surgical intervention, a recent review found no evidence for a beneficial outcome in patients receiving treatment [49].

As for physical therapy, there is very little evidence supporting the use of *manipulative treatment* for cervical radiculopathy [50]. In fact, guidelines suggests that careful consideration should be given to current evidence stating that manipulation may lead to worsened symptoms and complications, as well as adverse outcomes [47]. Also, adding acupuncture or traction of the cervical spine to a treatment regime has not shown to give any additional benefits, which is also stated in common guidelines [47, 51]. *Nerve root injection* refers to a procedure where a local anesthetic and steroids are injected transforaminally and epidurally in the vicinity of an inflamed nerve root. Injection treatment remains controversial, although some studies have shown long-term relief of radicular pain and neck pain [52-54]. There are a number of potential and serious complications documented in literature, including spinal cord injury and death, and guidelines suggest that this need to be taken into consideration before performing a procedure [47, 55, 56].

For myelopathy, there is a general understanding that the natural course of the disease involves gradual deterioration, and no solid evidence exists for nonoperative treatment. Still, current guidelines suggest that patients with mild myelopathy can be successfully managed through nonoperative care although close monitoring is

necessary [57, 58]. For the same reasons as for radiculopathy, drug treatment is often used. However, no randomized, placebo-controlled trials have been performed in literature to show its effect [59]. Riluzole – a drug thought to diminish neurological cell destruction – have recently been suggested to have a beneficial impact on cervical myelopathy patients but clinical investigations have not been convincing [60, 61]. Different types of physical therapies, and/or collar use have shown to have some effect in smaller studies [62, 63], but since study samples are low and therapeutical impact is not durable, surgical assessment is often underscored [26]. Also, since there are case reports of neurological worsening from manipulative treatment modalities, like traction, or even massage, caution is recommended when prescribing this type of treatment [57].

1.5.2 Surgical treatment options

The goal of a surgical treatment is to decompress neural tissue to stop progression of symptoms and possible enable full remission while maintaining the stability of the cervical spine. This can be obtained through an anterior approach, a posterior approach, or a combined anteroposterior approach. Each technique has its advantages and disadvantages related to the degree of obtained decompression, the complication rate and issues related to the peri- and postoperative management of the patient, i.e., comorbidity and postoperative pain considerations [8, 14, 47, 64].

In an *anterior approach procedure,* the patient is placed in supine position on the operating table. The cervical spine is reached through a skin incision on the front of the neck followed by a standardized dissecting route [65]. Anterior Cervical Discectomy and Fusion (ACDF) involves a discectomy and then placement of a synthetic implant, or more infrequently a bone graft, in the intervertebral disc space to

stabilize the cervical vertebrae. An option is to insert an artificial prosthesis (Anterior Cervical Discectomy Arthroplasty - ACDA) which is expected to preserve the range of motion of the neck to a higher degree [66]. Several studies have compared the outcomes after ACDF and ACDA, but the results are conflicting. Some emphasize that ACDA is reducing the risk of adjacent disc disease, although long-term outcomes seem to be similar, and there is uncertainty tied to prosthesis degeneration [67-69]. With the rising focus on value in health care, several studies have sought to assess the cost-effectiveness of degenerative neck surgery and compare different procedures. Although some investigations have concluded that ACDA may be less costly over time, there is a paucity of evidence in terms of choice of procedure [70].

In cases where there is need for decompression at the vertebral body level, a full corpectomy can be a surgical option. Autografts, like iliac crest or fibula grafts, were in the beginning regarded as the standard material in such instances due to the high fusion rate following surgery [71]. However, recent years the use of allografts have become more common [72]. Today, different cage types and material are used reconstruction of the cervical column with subsidence and lack of fusion as some of the main complications [73].

The operational advantages of an anterior approach are a direct visualization of the disc and the interspinal space, as well as the opportunity to easily remove a herniation, or even a damaged vertebra. Also, there are several perioperative and anesthesiologic advantages, e.g., easier/safer positioning of the patient and less postoperative pain from the operative incision [74].

A *posterior approach procedure* involves placing the patient in a prone position, which compared to supine positioning is associated with a variety of anesthesiologic complications deriving from increased pressure on anterior structures

[74]. As with an anterior approach procedure, the aim of a posterior approach is to decompress neural tissue. This is obtained through removing parts or all the posterior bone structures of the spine. In a *foraminotomy*, the cervical nerve root is decompressed by expanding the intervertebral foramen. In *laminectomy*, the whole lamina is removed to enlarge the spinal canal in relation to a spinal stenosis and myelopathy symptoms. If a laminectomy is performed on several levels, an *internal fixation* of the spine may be needed to stabilize the vertebra. This is typically performed using lateral mass screws and pedicle screws, sometimes with the addition of autologous bone grafts [75]. Fusion can improve long-term neck pain and prevent development of kyphosis, as well as instability which has been associated with poorer neurological outcome in myelopathy patients [76]. However, recent studies have not shown any difference in clinical outcome or effectiveness of treatment between fusion and nonfusion groups [77, 78]. Fusion is also more costly than nonfusion [79].

An alternative to laminectomy is *laminoplasty,* which involves keeping the lamina in place by hinging one side of the lamina open and thereby relieving pressure on the spine. Laminoplasty was originally described as a non-fusion alternative which simultaneously avoids post-laminectomy kyphosis. One disadvantage is that laminoplasty is not intended to treat spondylotic axial neck pain which is presented by many myelopathy patients, and patient selection is crucial in terms of achieving a good outcome. However, laminoplasty has been shown to be a solid alternative to laminectomy with fusion [80, 81], and several different techniques are currently being used [76, 82, 83].

1.5.3 Surgical treatment of cervical radiculopathy

The optimal treatment of cervical radiculopathy has long been debated. High-quality trials comparing surgical intervention with nonoperative management is lacking. A smaller often quoted Swedish study have shown that the short-term outcomes of surgery are superior to conservative treatment, however, the results converge over time, and the benefits of early surgery are more uncertain when compared to conservative treatment on longer terms [84]. This was confirmed by a more recent study, although the time before the outcomes of the two treatment groups converged were longer (2y vs. 1y) [85]. Although rates of patients achieving complete recovery are high, often within 24-36 months of symptom debut [45], up to a quarter of patients has been shown to have persistent symptomatology and, therefore, requiring surgery [35].

Currently, the consensus is that conservative management of cervical radiculopathy should be attempted in patients with novel symptoms, as long as there is no motor deficit or no myelopathy present [5]. Today's guidelines say that surgery should be attempted in patients with long-standing radiculopathy symptoms and with symptoms corresponding to the radiological findings. Also, in case of severe motor deficits, surgery should be performed as an emergency procedure to prevent chronic weakness of the affected limbs [5, 86].

There is no clear evidence in terms of which surgical technique is superior. ACDF has for a long time been the gold standard since it is safe and well tolerated by patients [5]. Also, long-term follow-up reports have found sustained patient improvement [87]. However, patients undergoing posterior foraminotomy experience improvement [88], even on a long term basis [89], and a 2016 systematic review comparing foraminotomy to ACDF found no difference in outcomes between the two

procedures [90]. Treatment guidelines suggest ACDF over posterior foraminotomy for single-level radiculopathy caused by central and paracentral nerve root compression, as well as spondylotic disease [47]. ACDA is often recommended for younger patients with single-level disease [91].

1.5.4 Surgical treatment of cervical myelopathy

Since cervical myelopathy is regarded as a progressive disorder, early recognition before irreversible spinal cord damage occurs is crucial for patient recovery. Surgical intervention involving decompression of the spinal cord and, in many instances, fusion of the affected levels of the cervical spine, have been shown to halt and even improve symptoms, and management is primarily operative [92]. Guidelines suggest surgery as first-line treatment for moderate to severe disease [57] [93]. However, a recent study indicates that also patients with less symptoms may benefit from surgery [94]

In terms of preferred surgical technique, no consensus exists. As for radiculopathy patients, anterior approach procedures are typically offered to patients with myelopathy due to disc herniation or single-level disease, while patients with multi-level disease are often offered posterior approach surgery [76, 95, 96]. There has been shown no difference in outcomes between patients operated with anterior versus posterior approach procedures [93] although there is a need of larger studies to confirm the results.

1.6 Registry studies

1.6.1 The role of registries in clinical research

In epidemiological research, cohorts are frequently used to investigate the relationship between an exposure and an outcome. There are in general two types of cohort studies: etiologic and prognostic. Etiologic study designs are used to identify associations that are causal, while prognostic studies focus on associations in general – noncausal and causal – since the main aim is to establish evidence for factors that can predict an outcome [97].

In etiological research, the Randomized controlled trial (RCT) is regarded by many as the gold standard to produce evidence for causality and to inform treatment decisions [98, 99]. The design of an RCT helps rule out two scientific pitfalls in relation to internal validity: 1) confounding effects – since the only presumably difference between the intervention group and the control group is the studied intervention or exposure; and 2) selection bias – since randomization eliminates any systematic differences between the two groups allowing attribution of differences in outcome to the intervention in the study [100]. A controlled environment with randomization of research participants minimizes study group differences prone to influence outcomes and, thus, provides strong evidence regarding the impact of an exposure on an outcome [98], but not regarding the impact of variation in prognostic factors.

In surgery, however, the use of RCTs has its limitations. First, setting up an RCT can be time-consuming and costly. Second, some effects may be rare and take years to establish themselves based on a presumable cause. In these instances, to obtain controlled environments over a long period of time can be difficult, if not impossible [98]. At the same time, a controlled or fixed environment can pose a

limitation in terms of data generalizability, since one can question whether the RCT results are valid in other circumstances than the one being studied [101]. Third, a general ethical challenge concerning RCT design stems from the issue that the patients gaining from the results of a trial not necessarily are the ones that bear the risk and/or burden of trial participation. Thus, there are several aspects of implementation that needs ethical and practical consideration. For instance, it would be unethical to randomize patients to risk behavior, such as smoking. It can also be controversial to use sham surgery for a control group in a study investigating the effect of a surgical procedure [102]. Because of current guidelines or surgical traditions suggesting a certain regime, it can be found unethical to expose or not expose a patient to a particular treatment and, therefore, making it difficult to justify the use of sham surgery [103]. Also, even if there is a consensus regarding the ethical aspect, lack of patient willingness to consent to randomization may be a major obstacle in terms of implementation [104].

For all these reasons there are relatively few RCT reports of surgical and nonsurgical interventions in relation to degenerative spine disorders, and many of those who do exist [105-108] have a limited number of included patients. Followingly, there is a paucity of evidence in terms of causality [100, 104].

Prospective registry studies, defined as a collection of standardized information about a certain group of patients sharing the same disease or clinical encounter [109], can complement RCTs. Registries offer a vast amount of data and can provide insight into care patterns and trends, quality of care and clinical outcomes of unselected patients treated in daily practice. Such data may also be leveraged to address issues outside clinical care, such as social status, employment, and education.

In this regard, ethical considerations necessary in a RCT design can be bypassed through the retrospective use of prospectively collected data.

1.6.2 The history of national registries and NORspine

The first national registry emerged in 1975 in Sweden and was set up to collect data from total knee arthroplasty [110]. Since then, several countries have established registries for a variety of surgical and medical conditions. One of the main drivers was the growing awareness of unwarranted treatment variations between medical environments that could not be explained by differences in populations [104, 111]. Major early development was also led by the cardiac surgery environment [112].

The first spine registry, Swespine, was launched in Sweden in 1993, first for lumbar disorders, later for degenerative cervical disease (2006) [113-115]. In Norway, the Norwegian Registry for Spine Surgery (NORspine) followed the same pattern with the establishment of lumbar data collection in 2006 and cervical data collection in 2011 [116]. Degenerative neck surgery is mainly performed in the four university hospitals in Norway (Oslo University Hospital (OUS), Haukeland University Hospital, St Olav´s University Hospital and the University Hospital of North Norway (UNN)), but some private clinics offer daytime surgery. Although the private providers showed great interest in submitting data, it took time for the university clinics to establish department wide collection. The reasons for lack of reporting were technical and information security issues (OUS), high rate of foreign language patients (OUS), and lack of implementation across the hospital (Haukeland). Thus, the coverage rate – calculated as the number of patients registered in the database divided by the number of patients being operated [117] – has steadily been increasing

since 2011. The rate increased drastically in 2021 after the implementation of an electronic registration portal which allows patients to complete the NORspine questionnaire prior to the first consultation at one of the eligible policlinics [118].

In the future, with the growing database, there will be endless possibilities to use the collected data for clinical purposes and a achieve more transparency regarding level of quality and outcomes after surgery.

1.6.3 Patient-Reported Outcome Measures

Among the data collected in NORspine, are several Patient-reported outcome measures (*PROMs*) [116]. These standardized and validated comprehensive questionnaires are used to assess a patient's health status at a particular point in time [119]. While changes in objective outcome measures, like radiological findings or return to work, does not always correlate with changes in a patient's perception of his or her own health status, PROMs represent a numerical rating of the patient's own condition [120].

Originally, PROMs were developed for use in pharmacological and health service research. Gradually, doctors started adopting PROMs in their clinical management of patients to improve the quality of care [120, 121]. As such, PROMs have been important in establishing evidence-based practice in spine surgery [122, 123]. Changes in a PROM score can give insight into a patient´s perceived improvement or worsening after a surgical treatment. This, again, can be helpful to assess the quality of care provided for a group of patients, compare the effectiveness of a treatment for a certain condition or set expectations for a patient prior to an intervention.

Different PROMs are often used in combination to cover major aspects of a patient's health status or level of functioning; *generic* PROMS which are applicable for every condition, and *disease-specific* PROMs which are suitable for a defined condition or disease [124]. To improve the selection of health-measurement instruments, a critical appraisal checklist has been recommended: COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN). The checklist provides assessment criteria and standards, so that the methodological quality of studies reporting and evaluating psychometric properties of PROMs can be evaluated as part of a review process [125, 126]. There are several key features mentioned in the COSMIN checklist, which are important to take into consideration when applying a PROM:

The concept of *validity* deals with the relationship between the measure and the actual world [127]. Is the PROM measuring what it is designed to be measuring? This can be an issue in terms of content, concept or construct or criterion/outcome.

A measure – or PROM – should have high *reliability* – meaning it should yield the same score/result on repeated trials. The *measurement error* – the difference between a measured quantity and its true value – should be low. The two types of measurement error are *information bias* and *selection bias*. Information bias – also called *observational bias* – refers to a situation where the observed data captures a biased or incomplete version of the real world. By using a highly valid instrument, information bias can be avoided. Selection bias, on the other hand, refers to the process of selecting data. Loss to follow-up can be a selection bias source when the patients not responding to an enquiry have different characteristics or outcomes than those who do respond. *Attrition bias* is often applied as a term. *Confounding* is also an important issue in terms of bias. The term refers to unmeasured and/or unobserved

factors which are associated with the outcome, but not accounted for in the analyses. Finally, a PROM should have high *responsiveness* – meaning that the measure should be responsive to changes in what is being investigated. To make meaningful use of PROMs in research and in clinical care, one needs to understand which changes in a PROM score represent a relevant change in the perceived health status of a patient. This is the concept behind "Minimal Clinically Important Change" - or MCID, which refers to the smallest change in a score that a patient perceives as clinically important [128]. "Minimal Important Change" - or MIC [129] – is a similar concept, but a different term.

There are two main approaches to assess MCID. The most common is to use an external criterion or anchor, like a patient-reported global assessment scale. A Receiver Operating Characteristic (ROC) curve is then used to find a cutoff value with the highest sensitivity and specificity representing the optimal MCID [130]. The second approach is to base MCID on the statistical characteristics and the distribution of the cohort in question to obtain a standardized metric. Some methods base their calculations on the Standard Error of Measurement (SEM), which quantifies the amount of error in a PROM that is not due to true changes. To be clinically relevant, the value of MCID needs to be higher than the error or variation in an instrument registration (the Minimal Detectable Change – MDC) to represent a meaningful change [129]. The major limitation for all distribution-based methods is that they do not capture the subjective change in the condition of the patient. Therefore, some authors have suggested a combination of anchor-based and distribution-based methods [129, 131].

Based on the PROM data from NORspine, it has so far been possible to create several PROM cutoff estimates for subgroups of patients undergoing lumbar surgery [132-135], useful for prognostic modelling [136].

1.6.4 Value-Based Health Care

A general concern in Western countries has for many years been the continuously rising health care costs. In the United States (US), health care is taking up almost 20% of the national GDP and an increase to 25% is expected within 2025 [137]. Although European systems are using substantially less money on health care than the US, spending in OECD countries is growing faster than the GDP. An ageing population, a rise in the prevalence of chronic conditions, and the acceleration of medical innovations has increased demand for treatment (2)**.** Norway is spending less per capita on health care, if corrected for the national price level, compared to most of its European neighbors. Still, there is ongoing discussion about the sustainability of the current welfare model and the increasing demands for health services (3).

While capitation has been the main health care financing model in the Nordic countries [138], care delivery systems and doctors in the US have traditionally been paid depending on the volume of the care provided. So-called fee-for-service reimbursement has been shown to lead to overproduction of services and increased health care spending [139]. Consequently, in recent years, there has been a growing focus on linking payments to health outcomes rather than a provider's activity. In this landscape, the term "value-based health care" (VBHC) has emerged. Instead of being rewarded for patient volume, VBHC reimbursement strategies aim to reward those who can create more "patient value". A now famous formula defines "patient value"

as "the health outcomes achieved per dollar spent" [140]. Interpretation of the formula points to two strategies to increase value – to cut costs, or to improve outcomes.

The first strategy is already pursued widely in healthcare. How to measure and improve outcomes is more complex. Standard metrics used in hospitals today are mortality, complication and infection rates, length of stay, bed utilization rates, patient wait times etc. None of these metrics expresses anything about the treatment impact experienced by patients, such as functional status and quality of life. Even patient satisfaction rates are something different than outcome measurements, as they merely say something about the experience the patient have or have had, but nothing about the effect of the medical treatment [141]. PROMs, however, capture both the "patient's voice" and the medical aspect of an intervention. At the same time, they seem capable of supporting many of the requirements for person-centered care addressed in the strategy plans of the National Health Service [142] and the Center for Medicare and Medicaid Services [143]. Even more importantly, the concept seems to resonate with both healthcare administrators and clinicians. A qualitative study of US health care leaders in PROM collecting centers in the US showed that the main incentive to collect PROMs on a system-level was that it was ethical, or "the right thing to do". The moral imperative was tied to quality control and the value of provided care [144]. On the other hand, there are numerous reports about the usefulness of PROMs, not only in a research context, but also in care delivery settings [145-147]. Even patients seem to endorse the use of PROMs when administered appropriately and not at the expense of a consultation [148].

In this context, the use of PROMs as performance measurement tools (PRO-PMs), have frequently been mentioned [149, 150]. By using PROM cutoff estimates as quality benchmarks, both clinicians and administrators can assess a perceived

benefit from a treatment on an individual patient level. On a group level, on the other hand, treatment success should be measured by the proportion of patients achieving a benchmark rather than by the average change in a patient group [128].

One relevant type of payment model that lays the ground for such PROM-PMs are bundled payments [139, 151]. The concept behind a bundled payment is that a provider gets a single reimbursement for all elements of care falling within a care cycle or clinical episode. A bundled payment model incentivizes resource moderation but raises the need to ensure that quality of care is not compromised. In a Stockholm bundled payment model for hip replacement, PROMs have been functioning as important PRO-PMs for a decade [152].

One of the reasons for the lagging application of PROMs as performance metrics, is that an improvement in a PROM score depends on many variables (the context, the patient population, the treating physician, etc.). Therefore, a benchmark for success or MCID needs to be internally and externally validated to be applied across centers and hospitals [153].

Bundled payments are already in place in Norwegian healthcare, although in a limited scope and with little use of quality parameters. According to a recent report, Norway is ahead in terms of VBHC measures. However, further implementation and development depends on a government involvement and ambition [154].

As long as PROMs are collected in the NORspine database and many of the other national registries, the potential is there.

1.7 Prognosis research in spine surgery

1.7.1 The PROGRESS initiative

The term "prognosis" refers to the expected course of a disease or an illness. In a clinical setting, the prognosis of a patient is often based on a set of variables – both individual and contextual, i.e., patient characteristics or demographics, patient history, blood parameters, radiological findings or even the physician's own experience of a treatment situation. A prognostic study of patients being treated for a particular condition therefore must take into consideration a variety of relevant factors in order to provide probabilities of a certain outcome [155]. Such outcomes can be absolute targets, like "death" or "survival", or constructed outcomes like achievements of a certain score on a PROM scale.

According to the PROGnosis RESearch Strategy (PROGRESS) initiative, there are four types of prognosis research: First, in *fundamental prognosis research* [156], the natural or clinical course of a disease or condition in the current health care environment is investigated. The typical research question is what the prognosis is for a group of patients with a certain trait. This type of research spans investigations of cancer risk on a population level, patient safety or health inequity issues, and various screening efforts. Endpoints are often described in rates or absolute risk.

A second type is *prognostic factor research* [99] which involves investigations of variables that are associated with a given health outcome. A prognostic variable may be an internal biomarker, like a single gene, a disease or condition, or a measure or aspect outside an individual, like a geographical area or an external exposure. For example, in degenerative spine research, several studies exist regarding the association between Body Mass Index or obesity and different outcomes [157-159]. In prognostic factor research, it is crucial to adjust for potential

43

confounding factors because it allows for the isolation of the effect of the prognostic factor on the outcome. By controlling for potential confounders, one can more accurately estimate the true association between the prognostic factor and the outcome and make more reliable predictions about the patient's prognosis. On the other hand, failure to adjust for an important confounding factor can lead to biased estimates of an association.

The main limitation of prognostic factor research is that it does not capture the full context of a clinical setting where several prognostic variables are at play. It only describes the relationship between the investigated factor and an outcome. In this regard, if the objective is to offer clinical guidance to physicians, *prognostic model research* [160] is a better option, which is the third type of prognosis research in the PROGRESS framework. A prognostic model can generate a singular risk estimate for a patient who is being evaluated for a specific medical treatment by incorporating prognostic values of several different variables through a multivariate analysis. This risk estimate can be applied by physicians to counsel patients about options related to a treatment or intervention. To develop an effective model, it is important to select the most relevant factors based on their availability in a clinical setting and their previously demonstrated relevance in the literature. The optimal design for such an investigation is a prognostic cohort study where the natural or clinical course of a patient can be followed and registered over time, such as in NORspine. To provide clinical relevance, the outcomes should be thresholds that matters to individual patients and have meaning in different clinical settings and environments [155]. Subsequently, PROMs can be ideal model outcomes in terms of health status or more disease-specific concerns.

As for prognostic factor research, prognostic model research does not provide insight into causality relationship or etiology. It merely describes a set of variables that are statistically associated with a treatment outcome. Although every casual factor is a predictor, and a casual factor often have a stronger prognostic value than one which is not casually related to an outcome [155], the aim for a prognostic model is not to establish causation, but association.

A final type of prognosis research deals with predictors of treatment effect, which is crucial in *stratified or personalized medicine research* [161]. Here, the aim is to single out a group of patients with a certain trait that has the most or least benefit from a treatment, and this approach requires an experimental design such as using a RCT.

1.7.2 Cervical degenerative disorders – prognostic models

Several prognostic models for outcome after surgical treatment of degenerative cervical disease have been developed, both for radiculopathy (Table 1), myelopathy (Table 2) and for mixed cohorts of patients (Table 3). The outcomes measured across these studies vary, and although the same PROMs might have been used, the definitions of cutoff thresholds often differ [162-165]. For example, some authors define a *poor* outcome as failure to reach MCID [166], whereas others define it as a larger reduction in a postoperative PROM score compared to a preoperative score [167], or as below an absolute threshold [168]. One study also uses a combination of PROM scores, clinical presentation, and perioperative factors, such as complication rates, to define their outcome of interest [164].

For *cervical radiculopathy*, the most extensive prognostic models were produced by Archer et al. in 2022 [167]. In a retrospective investigation of 4988

patients 12 months following cervical spine surgery, outcomes for NDI, arm pain and neck pain were reported with performances (AUC) ranging from 0.63 to 0.69 and with the NDI model as the most accurate. In terms of disease-specific variables, high initial arm pain and listhesis were significant positive factors in the models, while low neck disability, high initial neck pain and longer symptom duration was found to be negatively associated with the outcomes. Socioeconomic factors, such as higher education, preoperative employment, private insurance, and white race, were associated with better outcomes. Claims or litigation issues, smoking and ambulation assistance, on the other hand, were associated with worsening of symptoms. A high baseline NDI was the most important prognostic factor for a poor outcome.

The last two decades, three other prognostic models of improved outcomes each based on cohorts of approximately 100 patients undergoing ACDF due to cervical radiculopathy have been published [162, 163, 165]. The performances of the models are weak to moderate with a coefficient of determination (R^2) of 0.14-0.30, meaning that 70-86% of the variability in the results cannot be explained by the models. Hermansen et al. [163] identified factors predicting a good outcome in NDI and VAS neck pain 10-13 years following intervention. Contrary to Archer et al., initial high neck pain intensity was found to be positively associated with improvement in neck pain. In terms of NDI improvement, male sex and nonsmoking were found to be prognostic factors. Nonsmoking was also a positive prognostic factor for NDI and pain improvement in two other studies [162, 165], while male sex came out positive only for pain improvement in one of these studies [165] and for both short-term and long-term NDI improvement in the other study [162]. Other prognostic factors of a beneficial outcome coinciding in these two latter studies and in the models of Archer et al. were low neck disability, older age, and higher educational

level. Contrary to the finding of Archer et al., low pain level was found to be positively associated with improvement in one study. Finally, another study investigating risk factors for failure to reach MCID at 12 months following ACDF only found a high comorbidity burden to be associated with the outcome (NDI) [169].

In terms of *cervical myelopathy patients*, there is less of a question "to operate or not to operate" compared to cases of radiculopathy, as the disease is progressive. This aspect is also reflected in the results of prognostic model studies investigating associations with a poor outcome; symptoms and diagnostic markers for severe myelopathy seem to be the most pertinent factors (Table 2). In the 2020 study of Archer et al. [167] based on a large cohort of 2641 patients, a low modified Japanese Orthopedic Association (mJOA) scale baseline score (indicating severe disease) came out as the strongest prognostic factor for a poor outcome. Other important somatic and functional factors of a poor outcome were longer symptom duration, higher baseline NDI, ambulation assistance prior to surgery and higher baseline arm pain. As for radiculopathy, litigation issues and smoking were negatively associated with the outcomes, as well as posterior surgery and anxiety/depression.

Two other available studies [170, 171] producing prognostic models of favorable outcomes for cervical myelopathy patients somewhat mirror the finding of Archer et al. In a 2013 study of patients enrolled in the AOSpine North America study, a set of clinical and radiological variables were investigated at one-year follow-up. The authors found that a lower preoperative mJOA score (indicating greater severity), smoking, older age, psychological comorbidities, longer duration of symptoms, smaller transverse spinal cord area, and presence of impaired gait were associated with a decreased probability of achieving an absolute cutoff score of 16 on the mJOA scale [170]. The model was externally validated a few years later (2015) to

find that the most significant prognostic factors were baseline myelopathy severity, age, smoking status, and impaired gait. Psychiatric distress and duration of symptoms were less important [172]. In a 2016 study of AOSpine study patients from both the North America and International cohort, a two-year prognostic model based on the same mJOA cutoff was produced. Again, the model showed that nonsmoking, young age, less duration, and severity of symptoms were favorable of achieving a mJOA MCID. The performance of the model was not reported [171].

Finally, four studies have produced prognostic models based on mixed cohorts of cervical radiculopathy and myelopathy patients [164, 168, 173, 174]. In a study of 488 single-level ACDF patients, Anderson et al. presented a prognostic model for NDI success and "overall success" (see Table 3 for definition). The findings were similar to those found in previously mentioned studies. Compensations claims and litigation issues were negative prognostic factors, while high preoperative NDI, gainful employment, greater age, and intact preoperative sensory function were found to be predictors of success. In a small study of only 34 patients presenting four different models with moderate to substantial goodness-of-fit (R^2=0.38-0.73), a key finding was that a low Distress and Risk Assessment Method (DRAM) rating was a strong predictor for a good outcome. Other findings (little pain, young age, nonsmoking) are comparable to previously mentioned results for both myelopathy and radiculopathy models. Scerrati et al., on the other hand, found female sex, two-level surgery and the use of postoperative collar to be prognostic factors for worse NDI scores [168]. Finally, in a study of good quality-of-life outcomes, Lubelski et al. investigated a cohort of 952 patients undergoing either anterior or posterior surgery to assess prognostic models for MCID achievement in the EuroQol 5-Dimension-Questionnaire (EQ-5D), Patient Health Questionnaire (PHQ9) and Pain/Disability

Questionnaire (PDQ). The models showed moderate performance (0.35-0.47). Some of the variables found to be predictive of the quality-of-life outcomes were similar to the ones found to be associated with functionality and pain outcomes (race, age, smoking status, surgery type and number of surgery levels), while others are not typically found in functionality and pain models (diabetes, history of cancer, body mass index etc.). This illustrates how the choice of outcome applied in a prognostic model affects the set of characteristics found to be associated with the outcome.

Table 1. Clinical prognostic models for outcomes of cervical radiculopathy surgery.

Article	Follow-up	N	Surgery type	Outcome definitions	Performance	Significant predictors of outcome
Archer et al., 2020 [167]	12 months	4988	Cervical Spine surgery	1. NDI (0-100) 2. NRS arm pain (0-10) 3. NRS neck pain (0-10)	1. AUC= 0.64-0.69 2. AUC= 0.63-0.65 3. AUC= 0.63-0.67	Improvement: 1. NDI: greater age, preoperative employment, private insurance, greater arm pain, higher education, evidence of listhesis and white race 2. NRS-arm pain: greater age, preoperative employment, private insurance, higher education, greater arm pain, evidence of listhesis 3. NRS-neck pain: preoperative employment, private insurance, higher education, white race Worsening: 1. NDI: longer symptom duration, workers' compensation claim, ambulation assistance prior to surgery, higher baseline NDI, posterior surgical decompression, presence of depression, females 2. NRS-arm pain: longer symptom duration, workers' compensation claim, ambulation assistance prior to surgery, higher baseline NDI, higher baseline neck pain, posterior surgical approach 3. NRS-neck pain: longer symptom duration, workers' compensation claim, ambulation assistance prior to surgery, smoking, higher baseline arm pain, posterior surgical approach
Narain et al., 2019 [169]	12 months	84	1-2 level ACDF	Failure to achieve MCID 1. NDI >17.3 (0-100) 2. VAS neck pain >2.6 (0-10) 3. VAS arm pain >4.1 (0-10)	N/A	1. NDI: Comorbidity burden as evidenced by Charlson Comorbidity Index 2. VAS neck pain: No factors identified 3. VAS arm pain: No factors identified
Hermansen et al., 2013 [163]	10-13 years	95	1-3 level ACDF	Improvement in 1. VAS neck pain≥30 (0-100) 2. NDI≥20 (0-100)	1. R²=0.29 2. R²=0.14	1. VAS neck pain: initial high neck-related pain intensity 2. NDI: male sex, nonsmoking
Peolsson et al., 2008 [175]	3 years	95	ACDF	Good outcome for NDI and pain intensity on a 1. short-term (12-24m) 2. long-term (56-94m)	1. Q²Y=0.26 2. Q²Y=0.29	1. Short-term: Low preoperative NDI and pain intensity, horizontal active range of motion, non-smoking status, hand strength (right), male sex, and kyphosis 2. Long-term: Low short-term NDI and pain intensity, pre-operative low NDI and pain intensity, non-smoking status, sagittal and horizontal AROM, male sex, and hand strength
Peolsson et al., 2003 [165]	1 to 2 years	103	ACDF	Improvement in 1. VAS ≥10 ((0-100) 2. NDI ≥20 (0-100)	1. R²=0.30 2. R²=0.28	1. VAS: Male sex, greater kyphosis at the level operated on, nonsmoking, a greater neck mobility in right rotation, low disability on NDI, and older age 2. NDI: Higher educational level, non-smoking, greater kyphosis at the level operated on, a greater flexion mobility, greater right handgrip strength and lower current pain intensity

NDI=Neck Disability Index, NRS=Numeric Rating Scale, AUC=Area under the Curve, ACDF=Anterior Cervical Discectomy and Fusion, MCID=Minimal Clinically Important Difference, VAS=Visual Analogue Scale.

Table 2. Available clinical prognostic models for outcomes of cervical myelopathy surgery.

Article	Purpose	N	Surgical procedures	Outcome definitions	Performance	Significant predictors
Archer et al., 2020 [167]	12 months	2641	Cervical Spine surgery	Outcomes reported at 12 months: 1. NDI (0-100) 2. NRS arm pain (0-10) 3. NRS neck pain (0-10) 4. mJOA (0-100)	1. AUC=0.64-0-0.70 2. AUC=0.62-0.68 3. AUC=0.62-0.69 4. AUC=0.65-0.73	Improvement: 1. NDI: preoperative employment, white race, high baseline mJOA, evidence of listhesis, higher age 2. NRS-AP: high age, preoperative employment, white race 3. NRS-NP: preoperative employment, white race, high baseline mJOA 4. mJOA: preoperative employment, white race, high baseline mJOA, female sex Worsening 1. NDI: longer symptom duration, workers' compensation claim, higher baseline NDI, smoking, anxiety, depression, posterior surgery 2. NRS-AP: workers compensation claim, higher baseline NDI, higher baseline neck pain 3. NRS-NP: longer symptom duration, worker's compensation claim, higher baseline NDI, higher baseline arm pain 4. Worse mJOA: longer symptom duration, ambulation assistance prior to surgery, smoking, higher baseline NDI, posterior surgery;
Tetreault et al., 2016 [171]	2 years	716	Anterior, posterior, or circumferential surgery	1. Achievement of MCID on the mJOA scale (≥16)	N/A	1. Younger age, shorter duration of symptoms, nonsmoking status, and lack of significant gait impairment
Tetreault et al., 2013 [170]	1 year	272	Anterior, posterior, or circumferential surgery	1. A successful outcome: mJOA score of ≥16 2. A failed outcome: mJOA score of <16	AUC=0.79	1. A higher preoperative mJOA score, nonsmoking, no psychological comorbidities, no impaired gait, younger in age, shorter duration of symptoms and had a larger transverse spinal cord area 2. A lower preoperative mJOA score (greater severity), smoking, older age, psychological comorbidities, longer duration of symptoms, smaller transverse spinal cord area, and presence of impaired gait

NDI=Neck Disability Index, NRS=Numeric Rating Scale, mJOA=modified Japanese Orthopedic Association scale, AUC=Area under the Curve, MCID=Minimal Clinically Important Difference.

Table 3. Clinical prognostic models for outcomes of mixed cohorts of patients operated for radiculopathy or/and myelopathy.

Title	Follow-up	N	Surgical procedures	Outcome definitions	Performance	Significant predictors
Scerrati et al., 2021 [168]	1. Early (1-month follow-up) 2. Late (6-12-month follow-up)	234	1- or 2-level ACDF	1. Worse NDI (cutoff: median value: <6.8%) 2. Worse NDI (cutoff: median value: 2.0%)	1. N/A 2. N/A	1. Female sex, two-level surgery, use of postoperative collar 2. Female sex, two-level surgery, use of postoperative collar
Lubelski et al., 2018 [174]	3. 1 year	952	Anterior or posterior surgery	Achieve MCID for 1. EQ-5D 2. PHQ-9 3. PDQ	1. R^2= 0.43 2. R^2=0.35 3. R^2=0.47	1. Presence of herniation, myelopathy, surgical approach used, history of diabetes, preoperative EQ-5D and PDQ scores 2. Age, presence of herniation, surgical approach, diabetes, preoperative PHQ-9 1. BMI, smoking history, spondylosis, surgical approach, preoperative EQ-5D, preoperative PDQ
Anderson et al., 2009 [164]	2 years	488	Single-level ACDF	1. NDI-success: >15-point improvement in NDI 2. "Overall clinical success" based on meeting the following criteria: >15-point improvement in NDI; maintained or improved neurologic examination; no serious adverse event related to the procedure; and no revision of the plate or graft	1. N/A 2. N/A	4. NDI-success: Greater age, higher preoperative NDI score, and gainful employment were positive predictors and spinal litigation was a negative predictor 5. Overall success: preoperative normal sensory function was a positive predictor and worker's compensation claim a negative predictor
Peolsson et al., 2006 [173]	3 years	34	ACDF	Improvement in 1. Arm pain 2. Neck pain 3. NDI 4. General health	1. R^2=0.52 2. R^2=0.73 3. R^2=0.73 4. R^2=0.38	1. Non-smoking, low pain frequency, normal DRAM rating 2. Low pain frequency, low pain distribution, low back pain, normal DRAM rating 3. Low pain intensity, no use of painkillers, normal DRAM 4. Younger age, lower pain intensity, normal DRAM rating

ACDF=Anterior Cervical Discectomy and Fusion, NDI=Neck Disability Index, DRAM= Distress and Risk Assessment Method, EQ-5D=EuroQol 5-dimension-questionnaire, PHQ-9=Patient Health Questionnaire-9, PDQ=Pain/Disability Questionnaire, BMI=Body Mass Index.

2. Aims of the thesis

The overall aim of this thesis is to contribute to raise the quality of Norwegian spine surgery through the assessment of benchmarks for PROMs after degenerative neck surgery and to develop prognostic models for lack of clinical success in disability and arm pain after surgery for cervical radiculopathy. The specific objectives are:

1. to define success criteria after surgery for cervical radiculopathy based on Neck Disability Index (NDI), EuroQol 5-Dimension-Questionnaire (EQ-5D) with visual analogue scale (EQ-VAS), and Numeric Rating Scale for Arm Pain (NRS-AP) and Neck Pain (NRS-NP) at 3- and 12-month follow-up (Paper I).

2. to explore whether criteria for success vary between anterior and posterior surgical procedures and groups based on the etiology for one-level disease (disc herniation and spondylotic foraminal stenosis) (Paper I).

3. to define estimates for minimal clinically important difference (MCID) at 12-month follow-up for NDI, EQ-5D, NRS-AP, NRS-NP, and European Myelopathy Score (EMS) in patients undergoing surgery for cervical myelopathy (Paper II).

4. to assess proportions of patients achieving MCID for NDI, EQ-5D, NRS-AP, NRS-NP and EMS in cervical myelopathy patients undergoing anterior and posterior surgical procedures at 12-month follow-up (Paper II).

5. to develop and internally validate two prognostic models for nonsuccess in disability or arm pain 12 months after surgery for cervical radiculopathy (Paper III).

6. to illustrate the clinical usefulness of the prognostic models by presenting two case examples: one with low and one with high odds for nonsuccess (Paper III).

3. Materials and methods

3.1 Data collection

The research questions were addressed by using data from NORspine covering the period from January 1, 2011, to August 31, 2016. The registry is administered by UNN but is run by an independent steering group which oversees the data processing and storage and evaluates applications to access the data for research or other purposes. Due to potential conflicts of interest, NORspine receives no funding from the pharmaceutical industry or other sources than UNN. All patient data storage in NORspine is based on consent. Every hospital and private clinic consulting patients with degenerative spine disorders report to NORspine. During the first years of data collection, patients were presented with an option to register on the day of surgery (Appendix A) [176].

In this thesis, all three studies had a prospective cohort design with patient follow-up for 12 months. Patients referred to surgical intervention responded to a comprehensive questionnaire prior to surgery (Appendix B), and by mail 3 and 12 months after surgery (Appendix C). A separate form was presented to the operating surgeon after surgery and contains questions regarding indications for surgery, preoperative radiological and clinical findings, patient comorbidity and ASA level, operative technique, per-operative information, length of hospital stays and operation time (Appendix D).

3.2 Study design

An overview of the study design, follow-up period, collected variables, and study population in the three articles in the thesis is provided in table 4.

Table 4. An overview of study design, data source, follow-up period, and study population in this thesis.

	Paper I	Paper II	Paper III
Study design	Prospective cohort study		
Data source	Patients registered in NORspine between January 2011 and august 2016		
Months of follow-up	3 and 12	3 and 12	12
Study population	Patients operated in one or two levels for cervical radiculopathy	Patients operated for cervical myelopathy	Patients operated for cervical radiculopathy
Subgroups	Procedural patient groups (anterior vs. posterior surgery) and diagnostic groups (disc herniation vs. spondylotic root canal stenosis)	Procedural patient groups (anterior surgery vs. posterior surgery)	NDI group, NRS-AP group
Surgical treatment	ACDF, ACDA and PCF, laminectomy without fusion, hemilaminectomy, laminoplasty	ACDF, ACDA, hemilaminectomy, laminectomy with or without fusion, laminoplasty	ACDF and ACDA
Outcome measures	NDI, NRS-NP, NRS-AP, EQ-5D, EQ-VAS	NDI, NRS-NP, NRS-AP, EQ-5D, EMS	NDI, NRS-AP

ACDF=Anterior Cervical Discectomy and Fusion, ACDA=Anterior Cervical Discectomy Arthroplasty, PCF=Posterior Cervical Foraminotomy, NDI=Neck disability index, NRS-NP=Numeric Rating Scale for neck pain, NRS-AP=Numeric Rating Scale for arm pain, EQ-5D=Health-Related Quality-of-Life EuroQol 3L, EQ-VAS=general health status, EMS=European Myelopathy Score.

Since all our investigations were observational, our studies followed the recommendations for observational studies as defined by the Strengthening the Reporting of Observational studies in Epidemiology (STROBE) initiative [177] in order to ascertain an appropriate presentation of the research data. In the production of the prognostic model in paper III, we followed the guidelines given in the PROGRESS framework [160] which outlines strengths and weaknesses in research design and conduct in order to improve current research standards. We also applied the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) checklist [178] in the development of our prediction model to ensure model and reporting quality. Finally, the methodology was

chosen in accordance with the COSMIN initiative [125, 126] to ensure selection of the most relevant outcome measures.

3.3 Study samples

3.3.1 Paper I

In Paper I, we wanted to assess criteria for success in patients operated for cervical radiculopathy. MCID – defined as the "smallest difference in score in the domain of interest which patients perceive as beneficial" [179] – is often applied when assessing cutoffs for beneficial outcomes from surgery. However, the concept of "success" reflects a substantial improvement in a patient and should, thus, represent something more than a "minimal" difference. Followingly, we defined "success" as "much better" or "completely recovered" on the GPE (Global Perceived Effect) scale, which was used as an anchor to create cutoff estimates.

There were 4229 available patients registered in NORspine in the study period, and in order to create representative cutoff estimates for the radiculopathy cohort, we chose rather strict eligibility criteria (Figure 1). In total, 2868 were included for the main investigations. These patients were operated at one or two levels with either anterior or posterior surgical technique due to cervical radiculopathy, as these procedures are suggested by common guidelines [47]. The following patients were excluded:

1. Patients with missing or inconsistent data related to surgical technique or other information relevant to the research investigations.
2. Patients with former operations at index level to ensure that the symptomatology/condition of the patient was related to the current disease.

3. Patients with more severe or complex pathology operated with circumferential surgical technique, corpectomy, or posterior fixation since recent guidelines suggest minimally invasive techniques and this patient group potentially can have different outcomes [47, 180-182].

4. Patients with bone grafts since recent guidelines suggest allograft use [183].

5. Patients with myelopathy symptoms since this is regarded as more serious than having radiculopathy symptoms and, therefore, can alter the indication for surgery [184].

Further, we analyzed subgroups of patients for cutoff estimates. To separate patients suffering from a disc herniation and those diagnosed with spondylotic changes, we investigated patients with one-level disease only. Patients reported to have both a disc herniation and spondylosis were excluded from this subgroup analysis. The total number of patients in these two groups, therefore, does not add up to the total number of patients in the cohort. Finally, we assessed cutoff scores for patients operated with anterior and posterior techniques separately at 12 months. Anterior techniques included fusion and arthroplasty, while posterior techniques included foraminotomy, laminoplasty or hemilaminectomy/laminectomy without fusion.

Figure 1. Exclusion and inclusion criteria for patients in Paper I with follow-up rates.

3.3.2 Paper II

In Paper II, we wanted to assess estimates for MCID in patients with cervical myelopathy. As described in the introduction, surgical intervention is the primary treatment for patients with cervical myelopathy. The aim is to stop progression of symptoms although many patients improve following surgery. In this context, it is less relevant to investigate the concept of "success", which was the aim in Paper I. Instead, we chose to assess cutoff estimates for a minimal improvement through the use of MCID categorized as "slightly better", "much better" or "completely recovered" on the GPE scale.

A physician often places a myelopathy diagnosis based on a set of different criteria, such as a detailed anamnesis, physical, electrophysiological, and radiological examinations and scoring systems (i.e., PROMs). Although results from several such

investigations were registered in the NORspine questionnaire, we chose to include patients that were check-marked for having myelopathy by the operating surgeon. As in Paper I, we excluded patients with former operations at index level and patients operated with circumferential technique to ensure a homogenous cohort. We also investigated subgroups of patients undergoing anterior approach procedures (ACDF/ACDA) and patients undergoing posterior approach procedures (laminectomy with or without fusion, laminoplasty). The main cohort and subgroups are presented together with follow-up rates in Figure 2.
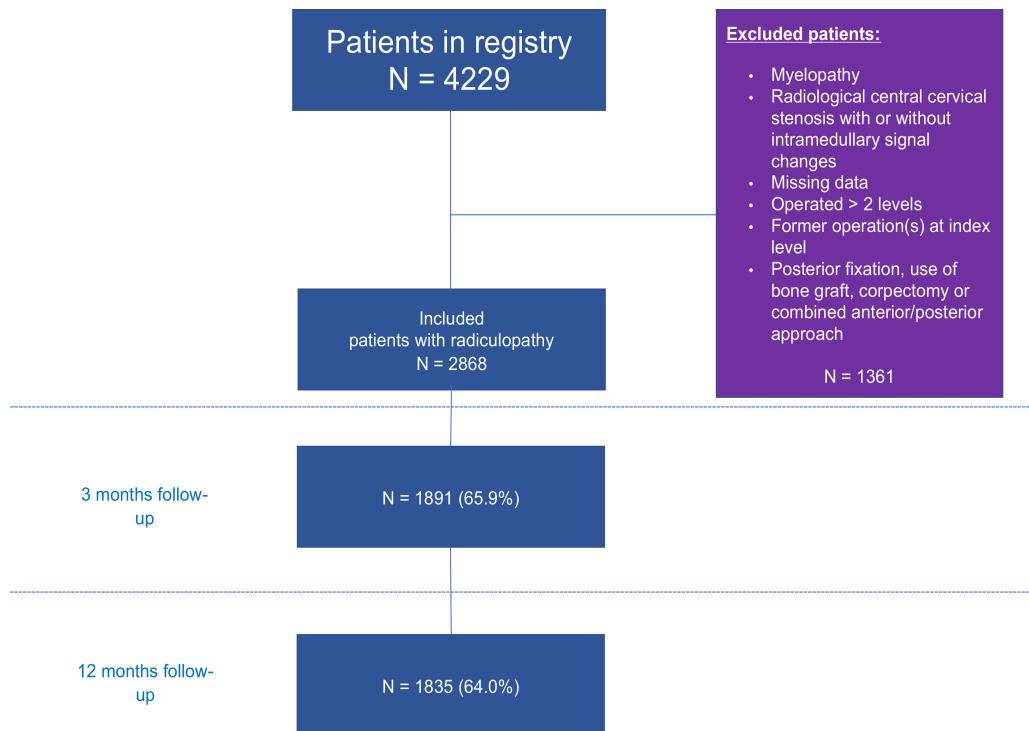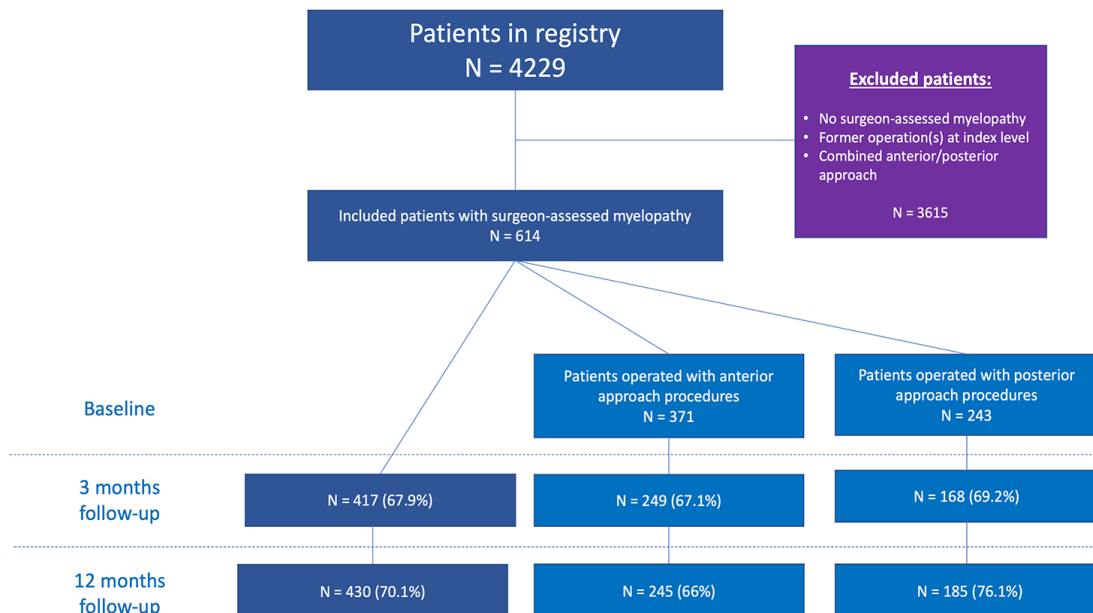

Figure 2. Exclusion and inclusion criteria for patients in Paper II with follow-up rates.

3.3.3 Paper III

In Paper III, we chose to use the same inclusion criteria for patients as in Paper I although we did open up to additionally include registered patients with three-level disease since recent investigations have suggested this to be safe and efficient [185, 186].
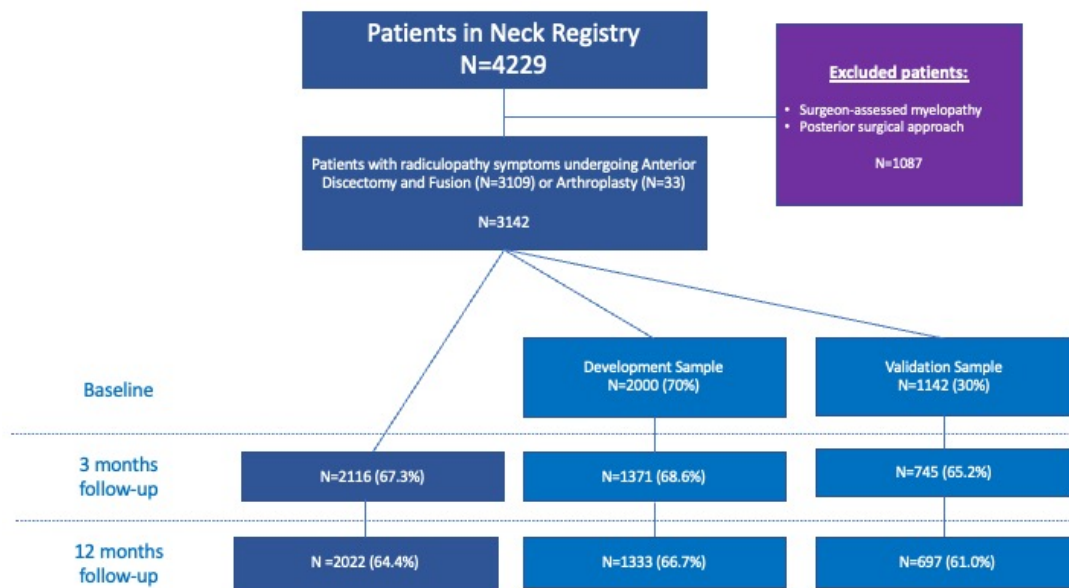
Figure 3. Exclusion and inclusion criteria for patients in Paper III with follow-up rates and with development and validation sample sizes.

Among the 3142 patients, only 33 received a disc arthroplasty (about 1%), while the rest underwent fusion surgery. However, the baseline characteristics and outcomes between the arthroplasty and fusion group were strikingly similar, with only a few significant differences: higher baseline NDI (p=0.02) and NRS neck pain (p=0.002) for the arthroplasty patients, as well as a lower number of operated levels (p<0.001). Finally, we chose to exclude patients undergoing posterior approach procedures. The rationale is based largely on the current routine and more common use of anterior techniques, but it is also based on the Paper I data showing that patients operated with anterior procedures had significantly worse symptomatology than patients operated with posterior procures. As mentioned in the introduction, several prediction models describe PROM baseline levels as important predictors for a beneficial or nonbeneficial outcome. In a clinical setting, it can be natural to tie a prediction model to the procedure at hand rather than a condition. In Figure 3, the patient cohort of Paper III is presented together with follow-up rates.

3.4 Patient-Reported Outcome Measures

The following PROMs were applied in the investigations performed in relation to this thesis:

The *Neck Disability Index* (NDI) [187] is a self-assessment tool used to measure disability associated with degenerative neck disease. The index consists of ten items that assess pain, personal care, lifting, reading, headaches, concentration, work, driving, sleeping, and recreation. Patients rate each item on a six-point ordinal scale ranging from 0 to 5. The ten item scores are then combined and recalculated as a percentage score ranging from 0 (no disability) to 100 (maximum disability). NDI has for many years been widely used, validated and translated into several languages [188] and plays an important role in terms of validation processes across populations.

The *Numeric rating scale for arm pain (*NRS-AP) and *neck pain* (NRS-NP) [189] are two distinct scales used to evaluate the severity of arm and neck pain, respectively. The scales range from 0 to 10, where 0 represents "no pain" and 10 represents "the worst pain imaginable".

The EuroQol-5-Dimension-Questionnaire (EQ-5D) [190] is a generic PROM for capturing health-related quality-of-life. EQ-5D assesses five dimensions: mobility, self-care, usual activity, pain/discomfort, and anxiety/discomfort. Patients rate the level of problems they experience in each dimension on a three-level scale: "none," "mild to moderate," or "severe." The resulting score ranges from -0.59 to 1, with 1 indicating perfect health, 0 representing death, and negative values indicating a health status worse than death. In the *EQ-VAS* part of the assessment, patients rate their overall health on a vertical analog scale from 0 to 100. On the scale, 0 represents "worst imaginable health" and 100 indicates "best imaginable health."

The *European Myelopathy Score* (EMS) [191] is a self-assessment questionnaire mapping spinal cord function through the evaluation of the four major neural systems involved in myelopathy: upper motor neurons, lower motor neurons, the posterior roots and the posterior column. By answering five questions about their health status (gait function, bladder and bowel function, hand function, proprioception and coordination, paresthesia/pain) patients are given a score that ranges from 5 (indicating severe deficit) to 18 (indicating no symptoms).

Global Perceived Effect Scale (GPE) [192] measures the patient perceived benefit of an operation by asking how the situation is for the patient after the procedure. There are seven response categories; (1) "completely recovered," (2) "much improved," (3) "slightly improved," (4) "unchanged," (5) "slightly worse," (6) "much worse", and (7) "worse than ever." In this thesis, GPE is applied as an anchor question to assess cutoff estimates for "success" (Paper I) and "Minimal Important Change" (Paper II).

3.5 Prognostic factors

In order to develop a prognostic model for achieving a clinical important outcome after cervical radiculopathy surgery, we started with a set of putative prognostic factors that could be associated with the outcome. We had access to several variables specifically selected for research purposes throughout the NORspine questionnaire. We also searched the scientific literature for prognostic factors shown to have an impact on outcomes for surgery due to cervical degenerative disorders. The factors selected for the development of a prognostic model in Paper III are listed in Table 5 together with references to relevant literature and categorizations of each

factor. Regarding the selection of putative prognostic factors, we would like to note the following:

Race has been found to be associated with surgical outcome [167, 174]. This variable is not collected in the NORspine questionnaire. However, patients register whether their mother tongue is Norwegian or "other" (Appendix B). We, therefore, found it justified to include "mother tongue" in the analyses.

Preoperative opioid use is a well-documented prognostic factor for both adverse events, poor return to work status and nonbeneficial outcomes after cervical spine surgery [193, 194]. In the NORspine patient questionnaire, the patient is asked to report use of analgesic drugs. However, there are no distinction between different types of drugs. Therefore, "use of analgesics" was applied as a variable in the prognostic model and not "opioid use" which would have been the preferred option.

Posterior approach surgery has been shown to be negatively associated with outcomes [167]. Since we chose to investigate anterior approach procedures, this variable evidently was not included in the analyses.

Comorbidity has been shown to have impact on outcomes in some studies [169]. This factor was covered by ASA level scoring and surgeon-assessed comorbidity in our investigations.

Preoperative pain levels are reported by several studies to have an impact on outcome from surgery [167, 195]. In a 2018 article, Passias et al. introduced the "arm pain vs. neck pain ratio" as a prognostic factor for increased improvement in arm pain [196]. The ratio was included in our analyses.

There are several radiological variables that have been applied in prognostic research, such as evidence of *kyphosis* and *listhesis* [162, 167]. The compliance

regarding documentation of radiological findings in the surgeon's form was low and the registered cases turned out to be not statistically relevant.

3.6 Statistical analyses

All data analyses were performed using SPSS versions 25.0 and 26.0. In Paper III, a medical statistics expert (Milada C. Småstuen) was brought in to assist with the prognostic modelling. In Paper I and II, the statistical work was done by the first and last authors (Christer Mjåset and Margreth Grotle).

3.6.1 Missing data

In terms of missing data, several measures were taken. Patients with missing information in relation to type of surgical procedure and/or complete lack of PROM assessment were excluded from the investigations. PROM scores of patients with at least one obtained PROM baseline or follow-up score were included in the analyses. In terms of PROM questionnaires that were returned incomplete by patients, the missing value for one item was replaced with the mean value of the available items in the patient´s questionnaire.

In Paper III, the proportion of missing data for each prognostic factor was registered. Only for a few factors the missing data rates exceeded 5%, and no imputation technique was used to replace the missing parameters.

Table 5. Prognostic factors analyzed in Paper III together with categorization and references to relevant literature.

Prognostic factor	References	Categories
Gender	Hermansen et al. [163], Archer et al. [167], Peolsson et al. [175], Khan et al. [197]	Male; female
Age	Archer et al. [167], Peolsson et al. [165], Tetreault et al. [170, 171, 198], Merali et al. [199], Anderson et al. [164], Müller et al. [200]	<40 years; 40-60 years; >60 years
Obesity	Lubelski et al. [174], Müller et al. [200]	BMI>30; BMI≤30
Work status	Archer et al. [167], Anderson et al. [164]	Sick leave; retired or disabled; rehabilitation, return-to-work training or out of work; student, in work or househusband/housewife
Physical demands in work	Cagnie et al. [201], Macedo et al. [202], Palmer et al. [203], Kim et al. [204]	Work in front of a computer screen or sitting still; light physical labor; hard physical labor
Educational level	Archer et al. [167], Peolsson et al. [165]	High school or less; less than 4 years of university; 4 or more years of university
Smoking	Archer et al. [167], Peolsson et al. [165, 173, 175], Hermansen et al. [163], Tetreault et al. [170-172, 198], Merali et al. [199]	Yes; no
Mother tongue	Lubelski et al. [174] (race), Archer et al. [167] (race)	Norwegian; other
Comorbidity	Lubelski et al. [174], Narain et al. [166]	Yes; no
Pending litigation	Anderson et al. [164], Archer et al. [167]	A pending or unresolved claim or litigation issue against the Norwegian public welfare agency fund concerning permanent disability pension or compensation claims against private insurance companies or the public Norwegian System of Compensation to Patients; no pending litigation
Duration of arm pain	Archer et al. [167], Merali et al. [199], Tetreault et al. [170, 171], Tarazona et al. [195]	<3 months; 3-12 months; 12 months
Duration of pre-operative paresis	Archer et al. [167], Merali et al. [199], Tetreault et al. [170, 171], Tarazona et al. [195]	No paresis; <3 months; ≥ 3 months
Previous cervical spine surgery	Radcliff et al. [205], Müller et al. [200]	Yes; no
Number of surgical levels	Lubelski et al. [174], Müller et al. [200], Peolsson et al. [165], Scerrati et al. [168]	One-level surgery; two or more levels of surgery
Use of analgesics	Yerneni et al. [193], Faour et al. [194]	Daily use; no daily use
ASA level (1-4)	Kim et al. [204], Lubelski et al. [174], Narain et al. [166]	Level 1 or 2; level 3 or 4
Baseline NDI score	Archer et al. [167]	Low (0-30); medium (31-50); high (50-100)
Baseline NRS arm pain	Archer et al. [167], Peolsson et al. [165, 173, 175]	Low (0-5); medium (6-8); high (<8-10)
Ratio of arm pain vs. neck pain [196]	Archer et al. [167], Hermansen et al. [163]	>1; ≤1
Anxiety or depression	Archer et al. [167], Peolsson et al. [173], Khan et al. [197]	Moderate or severe anxiety or depression; no anxiety or depression

BMI=Body Mass Index; ASA=American Society of Anesthesiologists; NDI=Neck Disability Index; NRS=Numeric Rating Scale; EQ-5D=EuroQol 5-dimension-questionnaire.

3.6.2 General overview of methods and statistical analyses in Paper I and II

Table 6 shows an overview of the statistical analyses performed for the first two papers in this thesis. The different cutoff modalities were produced in accordance with previous literature [206]. The *absolute score* – also called the *raw* score or the *follow-up* score – refers to any PROM score reported by the patient at 3- or 12-month follow-up. The *change score* represents the absolute difference between the baseline score and the 3- or 12-month postoperative score. Finally, the percentage change score was obtained by dividing the change score with the baseline score and multiplying the result by 100.

Table 6. Overview of statistical analyses performed in Paper I-II.

Paper	I	II
Baseline characteristics	Reported as means and standard deviations of continuous variables	
Significance level	$P<0.05$ (two-sided)	
Between group differences	One-way analyses of variance (ANOVA) and Chi-square test	
Cutoff Modalities	Absolute score, change score and percentage change score	Change score and percentage change score
Threshold	Success	MCID
Correlation between PROMs and GPE	Analyzed by the Spearman correlation coefficient, *rho*.	

MCID=Minimal Clinically Important Difference, PROMs=Patient-Reported Outcome Measures, GPE=Global Perceived Effect Scale.

It has been argued that MCID is a floor value and should not be a goal for clinical success following an operation [134, 207]. Therefore, we found the need to assess criteria for success for radiculopathy in Paper I, rather than MCID. For myelopathy, on the other hand, surgery is warranted in most cases to stop progression of symptoms. Therefore, MCID were found to be a better modality than "success".

To produce cutoff estimates for "Success" (Paper I) and MCID (Paper II), the seven-step GPE scale was used as an anchor. "Success" was defined as "completely recovered" or "much better" (1-2), while MCID was defined as "completely recovered", "much better" or "slightly better" (1-3).

We employed Receiver Operating Characteristic (ROC) curves to establish the ideal threshold that provides the highest sensitivity and specificity in terms of correctly classifying an individual estimate according to the anchor [208]. For every possible PROM cutoff value, the ROC curves were generated by plotting the sensitivity against (1-specificity). The sensitivity for each of the cutoffs refers to the probability of correctly classifying a patient as having achieved the cutoff in question. Followingly, the specificity refers to the probability of a patient being correctly classified as not achieving the cutoff estimate. The optimal cutoff estimate is the point on the ROC curve classifying most of the individual patients correctly, i.e., the estimate with the highest sensitivity and specificity. This can be found by identifying the closest point on the curve to the upper left corner in ROC-plane (0,1) [209].

Discriminative ability – or precision in correctly categorizing a case based on the anchor – were evaluated using the Area Under the ROC curves (AUC) with a 95% confidence interval. An AUC of 1 represents a perfect discrimination between patients, while an AUC=0.5 means that there is no distributional difference between the two groups and, therefore, not possible to distinguish between the groups by using the cutoff estimate. According to literature, an AUC of 0.7 to 0.8 is considered acceptable, 0.8 to 0.9 is excellent, and >0.9 is outstanding [210].

Correlation between each investigated PROM and the GPE scale was analyzed by the Spearman correlation coefficient, *rho*. A complete Spearman correlation (-1 or +1) occurs when each of the investigated variables is a perfect monotonic function of

the other variable. A correlation of 0 represents no association. A correlation of >0.8 indicates a very strong association [211].

In Paper I, we investigated patients undergoing surgery for cervical radiculopathy in one or two levels and did sensitivity analyses for patients undergoing anterior surgery and posterior surgery, as well as for patients undergoing one-level surgery due to disc herniation or spondylotic foraminal stenosis. In Paper II, we assessed MCID for all patients undergoing surgery for cervical myelopathy independent of number of levels and did separate analyses for patients undergoing anterior surgery and those undergoing posterior surgery.

3.6.3 Statistical methods used for the prognostic model in Paper III

In paper III, the NDI and NRS-AP cutoff estimates for success were applied as outcomes. For NDI (0-100), nonsuccess was defined as an absolute score of >26. For arm pain intensity assessed by NRS-AP (0-10), nonsuccess was a score of ≥3. We chose to use the absolute scores since the cutoff had the highest sensitivity and specificity compared to the other cutoff modalities (change score and percentage change score). The absolute score is also easy to introduce in a clinical setting as no calculations is needed before explaining the PROM scale to the patient.

With the random sample function in SPSS, cases were selected for a training set to be used to build the two models (70%) and a validation set (30%) to perform an internal validation of the results. Each variable was checked for missing data and then underwent univariate logistic regression analysis to assess the crude association between the prognostic factors and each of the two cutoff estimates. We used Odds Ratios (OR) with a 95% CI to express the association between the factor and the cutoff estimate. A prognostic factor was regarded as significant when $p<0.1$ (lax

criterion), which was the criterion for a factor to be entered into the multivariate analysis for either neck disability and/or arm pain. A standard backward elimination process was started. Factors losing their association to the outcome and showing $p>0.05$ after entering the regression analysis were removed. For each step, the factor with the highest p-value was removed first until a group of significant factors remained.

To evaluate the performance of the disability and arm pain models, we used the following [212]:

1. The explained variance by *Nagelkerke´s R^2* which measures the proportion of the variance in the dependent variable (cutoff estimate) that the independent variables (prognostic factors) explain collectively. In this way, R^2 provides information about whether a prognostic model fits the underlying data. Values range between 0 and 1, where 1 is the optimal fit and signals that the model perfectly predicts the outcome.

2. *The Hosmer–Lemeshow test* which assesses whether an observed event rate match expected event rates in subgroups of the model population. The output is a p-value aiming to be above 0.05 for the prognostic model to be a good fit.

3. Assessment of the discriminative ability of the prognostic model using *AUC*, which is applicable for a binary outcome. The discriminative performance of the model was considered acceptable if the AUC was ≥ 0.7 and good if the AUC was ≥ 0.8 (the c-criterion).

4. Internal validation was conducted by a *bootstrap procedure* (1000 samples) to estimate the amount of optimism in the two final models. Calibration plots were created. Calibration refers to the agreement between

estimated and observed outcomes. The Calibration-In-The-Large (CITL) refers to the difference between the average predicted and observed risk. The slope value indicates whether the model is over- or underestimating the prediction effects between the development and validation sample. The closer the slope value is to 1.0, the less over-optimism of the prognostic model.

3.7 Ethical considerations and juridical aspects

NORspine is administered by the University Hospital of North Norway, and all contained information is approved by the Norwegian Data Inspectorate (Datatilsynet). All registration of data is approved by the patient. The approval contains an acceptance that the data may be used in future research (Appendix A).

The data received by our research team were coded with all the personal identifiers removed and only a participant code present. The key identifier is kept separately by the NORspine administration to avoid identification of patients. However, our data included dates and localization of hospitalizations, as well as other patient information related to the current disease and treatment, so it was common routine to apply the Regional Committee for Medical and Health Research Ethics (REK) for approval and guidelines for handling the data. REK evaluated the registry protocol for this dissertation in 2014 under the identifier 2014/1477 and concluded that the project was to be classified as "a quality control measure". Thus, following the REK approval, the project group was given free access to the NORspine database. Based on the current regulations of 2014, the office for data protection (Personvernombudet) at Oslo University Hospital suggested extraction of the NORspine data on the University of Oslo's Service for Sensitive Data Platform

(TSD). The data where then made available for a selected group of members in our research group. When the General Data Protection Regulation (GDPR) came into effect in 2018, our project obtained a renewed approval by Personvernombudet, so that the data could be kept on the TSD server and handled according to plan.

4. Results

Baseline characteristics of all investigated groups in Paper I-III, including sensitivity analysis groups, are listed in table 7.

Table 7. Baseline characteristics of the different cohorts investigated in Paper I-III

Paper	I					II			III
Patient group	Cervical radiculopathy	Disc herniation	Foraminal stenosis	Ant. surgery	Post. surgery	Cervical myelopathy	Ant. surgery	Post. surgery	Cervical radiculopathy
N	2868	1182	403	2640	228	614	371	243	3142
Age; (SD)	49.4 (9.2)	46.4 (9.0)	53.1 (9.1)	49.0 (9.2)	53.8 (8.9)	57.4 (12.2)	52.3 (10.9)	65.2 (9.9)	49.5 (9.3)
Female; no (%)	1348 (47.0)	595 (50.3)	178 (41.4)	1279 (48.4)	69 (30.3)	233 (37.9)	156 (42.0)	77 (31.7)	1502 (47.8)
ASA level (SD)	1.7 (0.6)	1.6 (0.6)	1.8 (0.6)	1.7 (0.6)	1.7 (0.7)	2.4 (1.6)	2.06 (1.4)	2.84 (1.8)	1.7 (0.6)
Body Mass Index (SD)	26.9 (4.2)	26.7 (4.4)	27.0 (4.3)	26.9 (4.3)	26.7 (3.9)	26.9 (4.2)	27.3 (4.4)	26.9 (5.1)	26.9 (4.3)
Smokers; no (%)	931 (32.5)	385 (33.3)	132 (31.4)	872 (33.7)	59 (26.5)	190 (31.3)	121 (32.8)	69 (28.9)	1043 (33.9)
Comorbidity; no (%)	1115 (39.5)	381 (32.6)	92 (45.8)	1030 (39.7)	85 (37.6)	336 (54.7)	176 (48.5)	160 (66.1)	1272 (40.5)
NDI (SD)	41.2 (15.0)	42.2 (15.2)	40.4 (14.7)	41.7 (15.0)	35.3 (14.2)	34.5 (17.4)	34.7 (16.9)	34.3 (18.2)	41.6 (15.1)
NRS-AP (SD)	6.4 (2.3)	6.5 (2.3)	6.2 (2.3)	6.4 (2.3)	5.8 (2.5)	5.0 (2.9)	5.1 (2.9)	4.9 (3.0)	6.4 (2.4)
NRS-NP (SD)	6.1 (2.5)	6.1 (2.5)	6.1 (2.4)	6.1 (2.4)	5.6 (2.6)	4.8 (3.0)	5.1 (2.9)	4.4 (3.2)	6.2 (2.4)
EQ-5D (SD)	0.43 (0.32)	0.42 (0.33)	0.46 (0.31)	0.4 (0.3)	0.5 (0.3)	0.45 (0.32)	0.47 (0.32)	0.41 (0.33)	0.43 (0.32)
EQ-VAS (SD)	50.3 (20.2)	48.7 (20.7)	51.8 (18.7)	49.8 (20.2)	56.6 (19.0)	48.7 (21.3)	50.0 (20.6)	46.7 (22.2)	49.8 (20.4)
EMS (SD)	---	---	---	---	---	14.5 (2.4)	15.0 (2.1)	13.6 (2.5)	---

SD=Standard Deviation, NDI=Neck Disability Index (0-100), NRS=Numeric Rating Scale for arm pain (0-10), NRS-NP=Numeric Rating Scale for neck pain (0-10), EQ-5D=EuroQol 5-dimension-questionnaire (-0.4-1.0), EQ-VAS= Health-Related Quality-of-Life by EuroQol (0-100), EMS=European Myelopathy Score (5-18), MCID=Minimal Clinically Important Difference.

The myelopathy patients (in Paper II) were significantly older and had higher ASA levels/more comorbidity, but better disability and pain scores compared to the radiculopathy patients in Paper I. A similar pattern was found for posterior surgery patients compared to anterior surgery patients (Table 7).

The clinical course (as measured by NDI) of patients operated for cervical radiculopathy or myelopathy is illustrated in Figures 4 and 5. Both patient groups showed in average a large improvement during the first months after surgery. However, patients undergoing posterior surgery for cervical myelopathy showed a slight deterioration between follow-ups 3 and 12 months after intervention.
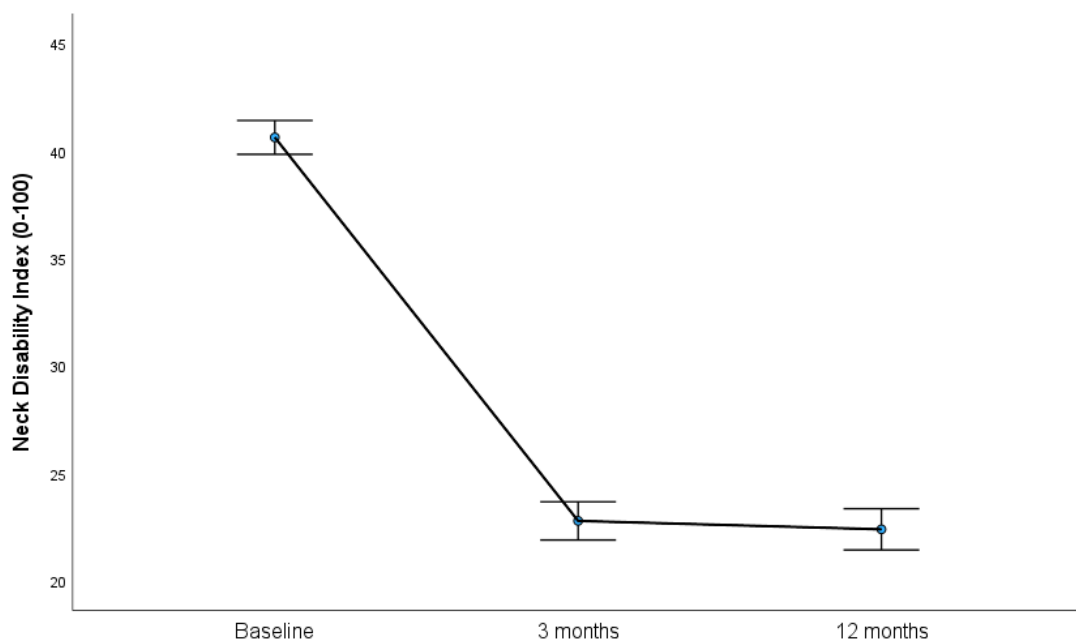
Figure 4. Mean NDI for patients operated for cervical radiculopathy at baseline, 3 and 12 months following surgery.

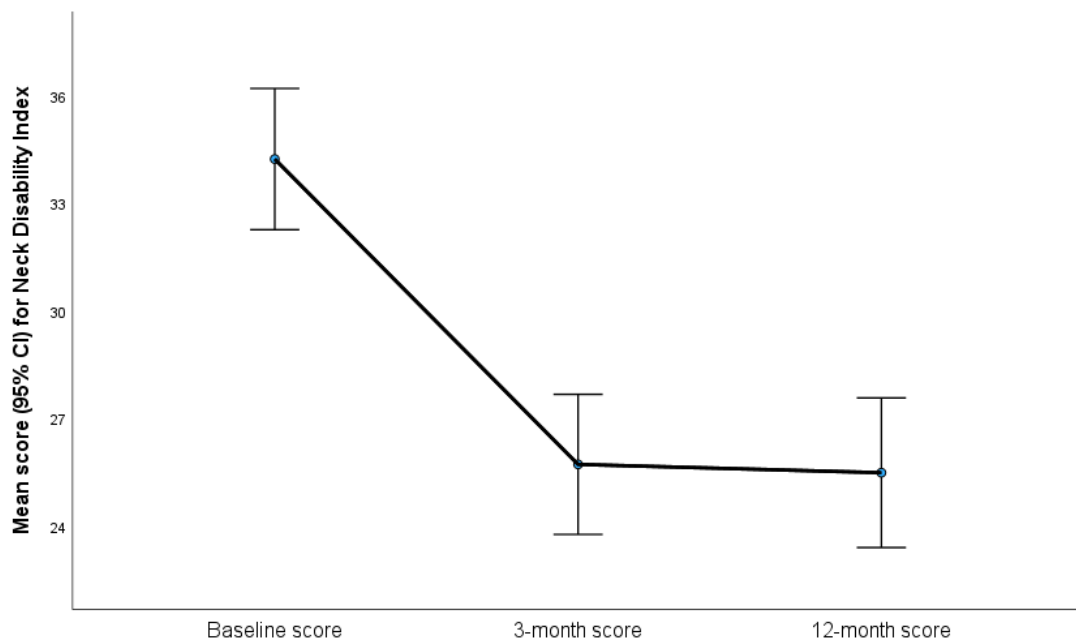**Figure 5a**. Mean NDI for all patients operated for cervical myelopathy at baseline, 3 and 12 months following surgery.
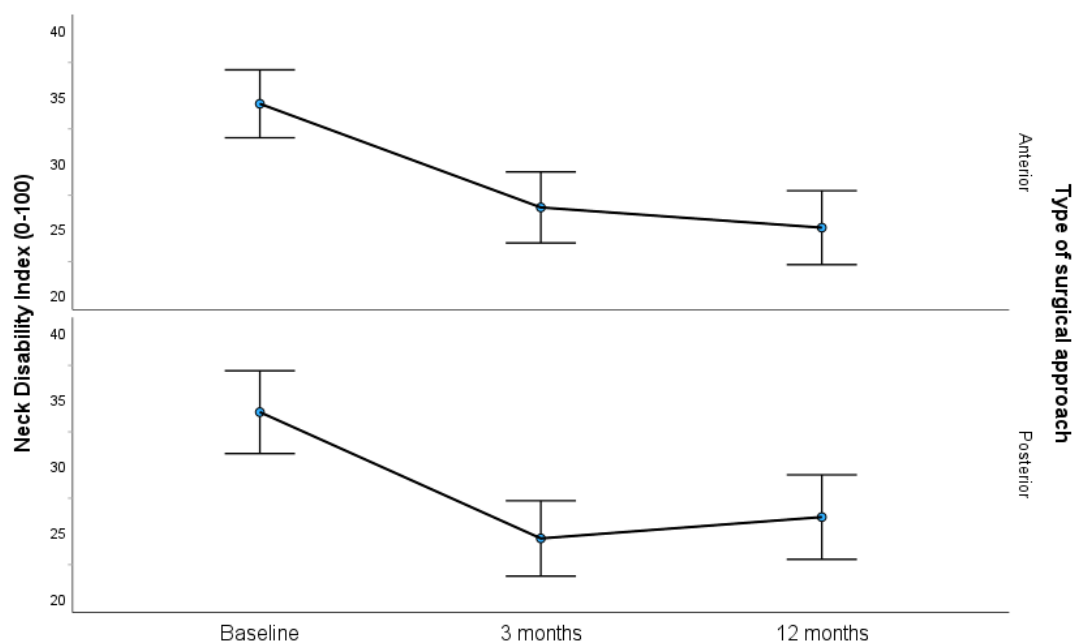

Figure 5b. Mean NDI for patients undergoing anterior surgery and posterior surgery for cervical myelopathy at baseline, 3 and 12 months following surgery.

In terms of correlation between the investigated PROMs and the anchor

(GPE), we found a stepwise decrease in the mean scores for every PROM modality at

3- and 12- month follow-up in Paper I and II (Figures 6a-e). For cervical radiculopathy (Paper I), the Spearman correlation coefficient varied from 0.41 to 0.78. NDI and NRS-AP follow-up scores and percentage change scores showed the best correlation (0.69–0.78). For cervical myelopathy, the correlation ranged between 0.30 and 0.59. The NDI percentage change score showed the best correlation (0.59) together with the NRS-NP percentage change score (0.55).

In Paper I, the NDI showed the highest discriminative ability for patients undergoing degenerative neck surgery for cervical radiculopathy (Table 8). Further, the percentage change and follow-up scores were more accurate than change scores. There were minor differences in results between follow-ups at 3 and 12 months. The results were stable across subgroups of cervical disc herniation and spondylotic root canal stenosis.

In Paper II, there were minor differences in AUC and MCID cutoff estimates in the collected PROMs (NDI, NRS-AP, NRS-NP, EQ-5D and EMS) at 3- and 12- month follow-up for patients undergoing surgery for cervical myelopathy. The percentage change scores of NDI and NRS-NP had the highest sensitivity and specificity (Table 8). There was a tendency to higher discriminative ability for PROMs in the same modality in the anterior surgery group (0.65-0.77) compared to the posterior surgery group (0.62-0.76). However, for EMS, the AUCs were higher in the posterior surgery group (anterior vs. posterior: 0.67 vs 0.72 for change score; and 0.65 vs 0.71 for percentage change score).

6a)
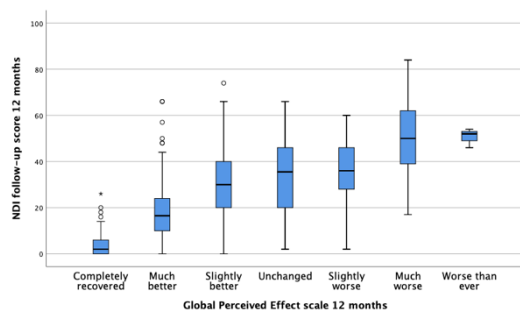
6b)
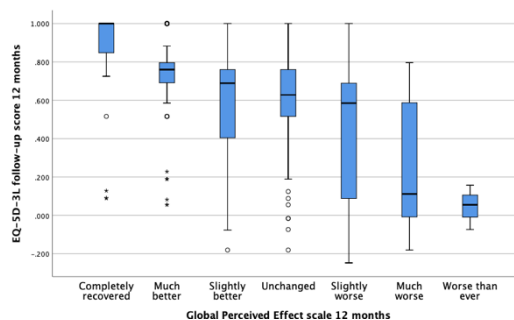
6c)

6d)

6e)

Figures 6a-e. Boxplots illustrating the correlation between Global Perceived Effect scale and 12-month PROM follow-up scores in patients operated for cervical myelopathy.
NDI=Neck Disability Index (0-100), NRS=Numeric Rating Scale for arm pain (0-10), NRS-NP=Numeric Rating Scale for neck pain (0-10), EQ-5D-3L=EuroQol 5-dimension-questionnaire (-0.4-1.0), EMS=European Myelopathy Score (5-18).

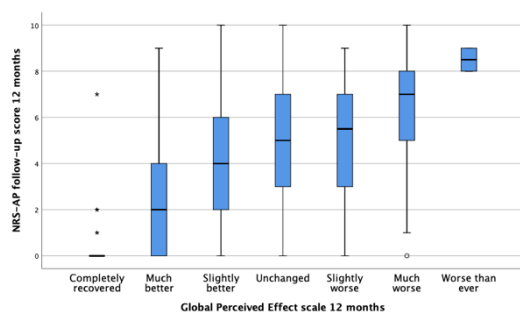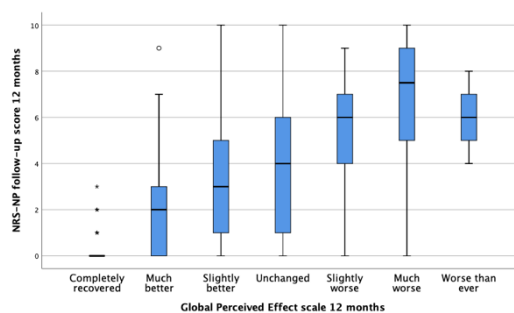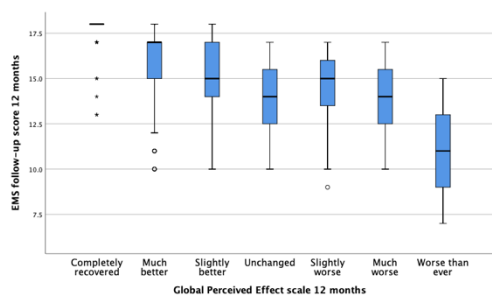**Table 8**. Area under the Receiver Operating Characteristic curve (AUC) for different PROMs and modalities in Papers I and II at 12-month follow-up.

	Follow-up score (points)	Change score (95% CI)	Percentage change score (95% CI)
Paper I: AUC for cutoff estimates for Success			
NDI	0.91 (0.89–0.92)	0.87 (0.85–0.89)	0.91 (0.89–0.93)
NRS-AP	0.86 (0.84–0.88)	0.81 (0.78–0.83)	0.85 (0.82–0.87)
NRS- NP	0.88 (0.86–0.90)	0.79 (0.76–0.81)	0.86 (0.83–0.88)
EQ-5D	0.86 (0.84–0.88)	0.74 (0.71–0.77)	–
EQ-VAS	0.88 (0.86–0.89)	0.78 (0.76–0.81)	0.74 (0.71–0.77)
Paper II: AUC for cutoff estimates for MCID			
NDI	–	0.74 (0.69–0.79)	0.77 (0.72–0.81)
NRS-AP	–	0.64 (0.58–0.70)	0.69 (0.63–0.75)
NRS-NP	–	0.73 (0.67–0.78)	0.76 (0.70–0.81)
EQ-5D	–	0.70 (0.64–0.77)	0.70 (0.64–0.77)
EMS	–	0.69 (0.63–0.75)	0.68 (0.61–0.74)

AUC=Area Under the Receiver Operating Curve, NDI=Neck Disability Index (0-100), NRS=Numeric Rating Scale for arm pain (0-10), NRS-NP=Numeric Rating Scale for neck pain (0-10), EQ-5D= EuroQol 5-dimension-questionnaire (-0.4-1.0), EQ-VAS-3L=Health-Related Quality-of-Life by EuroQol (0-100), EMS=European Myelopathy Score (5-18), MCID=Minimal Clinically Important Difference.

In Paper II, the proportion of patients achieving a MCID after surgery for cervical myelopathy was presented (Figure 7). The rates were slightly higher for the anterior approach group compared to the posterior approach group for both change score and percentage change score.
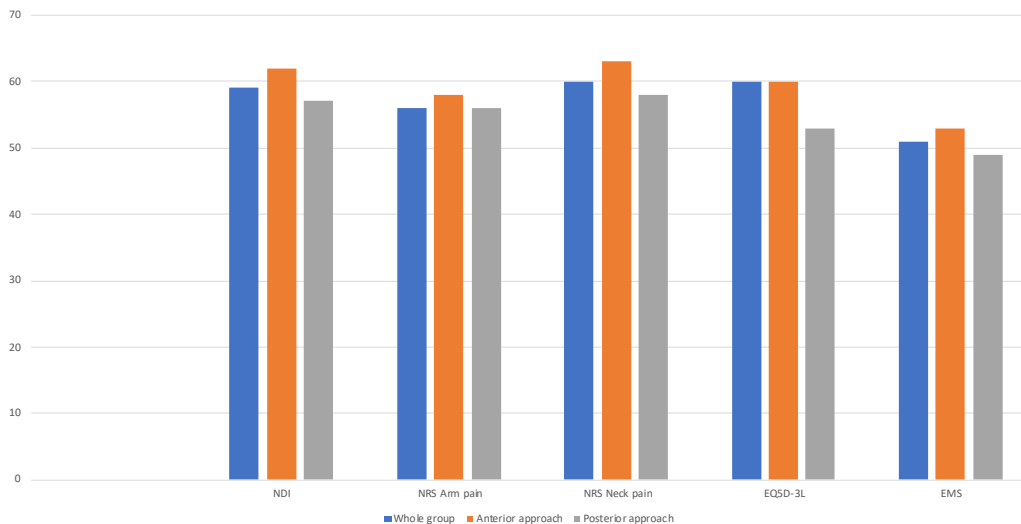
Figure 7. Proportion of patients that achieved a clinical improvement according to MCID estimates for percentage change scores at 12-month follow-up after surgery for cervical myelopathy.
NDI=Neck Disability Index, NRS=Numeric Rating Scale, EQ5D-3L=EuroQol 3L, EMS=European Myelopathy Scale.

In Paper III, we developed a prognostic model for lack of success in patients undergoing ACDF and ACDA for cervical degenerative radiculopathy based on the cutoff estimates generated in Paper I. A total of 2020 patients registered in NORspine were included. Approximately 38% of patients receiving surgery did not achieve an improvement in neck disability that they consider significant or substantial, whereas 35% of the patients did not achieve a similar improvement in arm pain. The final prognostic model for neck disability included seven predictors; physical demanding work, low level of education, pending litigation, previous neck surgery, duration of arm pain>3 months, medium to high levels of baseline disability, as well as anxiety/depression. The prognostic model for arm pain included nine predictors; physical demanding work, low level of education, pending litigation, previous neck surgery, duration of arm pain>3 months, medium to high levels of baseline disability, foreign mother tongue, smoking, and medium to high levels of baseline arm pain. The prognostic model for a nonsuccessful disability outcome showed good performance

with an AUC of 0.78 (95% CI: 0.75, 0.82), whereas the model for arm pain had an

AUC of 0.68 (95% CI: 0.64, 0.72). The calibration plots indicated no overfitting of

the two models (Figures 8a-b).

a)   b)
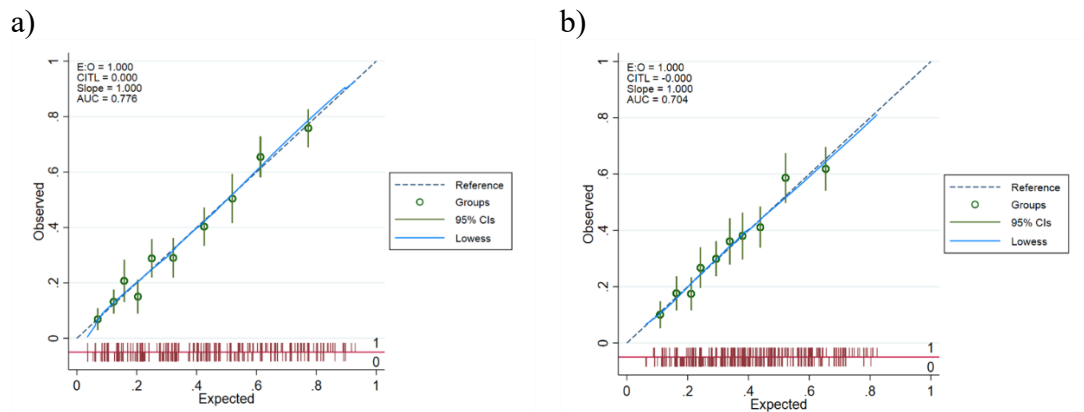
Figures 8a-b. Calibration plots for two models predicting nonsuccess in (a) neck
disability and (b) arm pain at 12-month follow-up.
CITL=Calibration-in-the-large, slope=calibration slope, AUC=Area-under-the-curve, CI=Confidence
Interval

To illustrate the clinical usefulness of the disability model we made two case

examples. One contained only a few of the prognostic factors in the model and

showed a 13% probability of nonsuccess. The second case example contained several

prognostic model factors and showed a 92% probability of nonsuccess.

5. Discussion

This thesis contains three scientific papers assessing benchmarks for clinical improvement and prognostic models for nonsuccess in a cohort of patients undergoing surgery for cervical degenerative disorders and registered in a national database (NORspine). In this work, several methodological frameworks and checklist were used: COSMIN [125, 126] for the two methodological papers, and the PROGRESS [156, 160, 161, 213], TRIPOD [214] and Prediction Model Risk of Bias Assessment tools (PROBAST) [215] for the prognostic model paper. In the following chapter, methodological aspects that could affect internal and external validity of the findings are presented. The validity discussion will be followed by a comparison of the findings in the thesis with results found in current literature, a description of clinical implications of the performed investigations, and, lastly, a recounting of perspectives on further research.

5.1 Internal validity

5.1.1 Study design

All three papers in the present study used a prospective cohort design, based upon the NORspine registry data. Cohort studies are considered the optimal design when conducting research on clinical course or prognosis, prognostic factors, and prognostic models according to the PROGRESS framework [156, 160, 161, 213]. One of the advantages by using a prospective cohort design, and data from the NORspine, is the large sample size and the broad geographical coverage of patients who have undergone spine surgery in Norwegian hospitals. This registry provides a unique opportunity to investigate patient outcomes the year following surgery, which due to

lack of statistical power can be difficult assess in prospective clinical cohorts and trials [216, 217]. Another advantage of clinical registers is that they include patients who often are excluded from trials and, hence, represent the "real life" variation and has a larger generalizability than trial cohorts [104, 218]. However, registry and cohort studies may be challenged by different types of bias, of which the most important will be discussed below.

5.1.2 Study attrition and missing data

Registry studies are particularly prone to *selection bias*, which occurs when the collected study sample, as a result of the process of collecting data, is not representative of the investigated population. The three studies in this thesis are based on data collected through NORspine from the year the registry was established in 2011 and until the fall of 2016. Today, every private and public center performing cervical spine procedures reports regularly to the registry together with interdisciplinary policlinics in rural hospitals. However, during the study period (2011-16), the completeness of the registry – referring to the proportion of individuals with the condition of interest in the target population [219] – was markedly lower than today (2021: 71%) [116-118]. Thus, there is a clear chance that the collected data in this thesis is not representative of the population operated for cervical degenerative disorders in Norway in the same period. We know from the annual reports that some hospitals experienced issues with information security and data handling. Also, there is a clear possibility that some surgical environments have been more conscientious than others. Administration of questionnaires to both patients and surgeons is time-consuming, and it takes time to build a well-working routine which the department staff fully embrace. All these issues may have led to systematic underreporting of

patients from some hospitals or centers treating patients with different characteristics or symptom severity than the remaining population, which, again, may have had an impact on our results.

In NORspine, patient inclusion is based on an opt-in principle. A reason for not wanting to be registered can be lack of trust in the system to keep data confidential or fear of data being used in a liability context. These patients may potentially have traits or features that deviate from the patients opting to be included in NORspine, and, therefore, may represent a non-response bias. However, the Norwegian population has a high trust in the national government [220], and national registries are highly regarded in terms of contributing to improve quality of care in medical practice. Also, the Norwegian system of compensation to patients [221], which offers free compensation claim processing, has been shown to bring down the number of court cases [222]. Thus, medical liability is less of an issue in Norwegian health care than in other comparable Western countries.

On the provider side, a potential bias may be that different hospitals and centers have different routines or procedures related to which patients are selected for surgery and to which type of operation is recommended. Since we have no records of patients treated conservatively, we do not know whether such clinical traditions systematically influence the characteristics of patients included in NORspine. However, the Norwegian neurosurgical environment is small and constituted by only approximately 110 Norwegian specialists [223]. Also, through the Norwegian Medical association, which organize close to 100% of acting professionals in the country [224], there is close collaboration between centers regarding national treatment recommendations and guidelines. Therefore, this type of bias should have limited impact on patient selection.

Another and possibly larger concern is *attrition bias*, which is a type of selection bias caused by systematic differences between the actual study sample and those patients who are lost to follow-up [225]. In our studies, patient groups with certain characteristics or symptoms may have been more prone to loss to follow-up than others. For example, compared to respondents, non-respondents in Paper I and II were more frequently smoking – a characteristic that was shown to predict non-success for arm pain in Paper III. Accordingly, there may be other characteristics not registered in NORspine overrepresented or underrepresented among the non-responding patients that may have contributed to biased results among the investigated outcomes. One such characteristic may be socioeconomical status. Our data were collected prior to the full digitalization of the patient questionnaires (2021), and the magnitude to the necessary efforts for participation may have contributed to a systematic loss to follow-up in groups with less resources. Outcome itself may also have influenced follow-up rates. Patients with more severe symptoms may be less motivated to answer a 12-month questionnaire due to health issues or lack of faith in the health care system. In our studies, poorer health-related quality of life (Paper I and II), as well as poorer disability and pain severity scores (Paper I), were more frequently found among non-respondents than respondents.

Some of the effect of attrition bias can be compensated with the large sample size found in NORspine [226]. Still, the rate of non-respondents in our studies is higher than the suggested rule of thumb saying that more than 20% can lead to biased results, whereas less than 5% will not [227]. However, researchers using registry data often have to deal with a higher number of respondents lost to follow-up than other types of studies. Attempts to trace and retain data from those lost are expensive and resource-demanding in population-based studies according to a report from the

Swedish national registry [113]. Accordingly, a 2015 investigation of international spine registries reported that the respondent rates ranged from 20% to 88%, and the authors concluded that one should strive for a respondent rate of 60-80% at 12-month follow-up to reduce bias in investigations [228]. Our figures fall within the recommended range.

The risk for attrition bias in NORspine have been investigated in two recent studies. Ingebrigtsen et al. interviewed patients lost to follow-up after surgery for cervical degenerative disorders to find that clinical outcomes did not differ between respondents and non-respondents. Forgetfulness (33.3%) and bustle (14.3%) were the two most frequent causes for lack of response. The authors concluded that non-respondents (28.7%) should be regarded as they were missing at random. Another study conducted on NORspine lumbar patients showed that a 22% loss to follow-up did not impact the two-year results after surgery. Again, the authors found no evidence that loss to follow-up was due to more unfavorable outcomes or health problems, but rather due to external factors such as forgetfulness (88%) or questionnaire fatigue (23%). However, the respondent rate was higher in this study (78%) compared to the rates in our studies 64-70% [135]. A study conducted by the Danish national registry yielded similar results, with the 12% loss to follow-up group demonstrating comparable outcomes and patient satisfaction scores to the respondent group [229].

In registry studies, if the issue of missing data is not appropriately addressed, it can lead to a misinterpretation of the findings. In Paper III, an overview of all missing values for the candidate predictors of the prospective models were presented. A missingness rate less than 5% is considered acceptable [230]. In our material, this threshold was exceeded only in a few instances. The variable "physical demands in

work" had the highest rate of missing data (15.5%). There is an option to employ an imputation method in such cases, i.e., impute the mean value of all non-missing observations for the variable in question. However, the imputed variables may not represent the true missing values and lead to biased results. Therefore, we chose not to impute any data before conducting the statistical analyses.

5.1.3 Classification and diagnostic subgroups

In terms of systematic *misclassification* of subjects, there is one concern in our project related to the surgeon´s registration of patient symptom etiology. Under the tagline "indication for operation" in the postoperative NORspine questionnaire, the treating surgeon checkmarks whether the patient has preoperative pain, paresis (including grading 0-5), myelopathy – with the option of choosing sensory or motoric or both, and/or "other" (see Appendix D). In Paper I and II, this categorization was the basis for splitting the registry cohort into two groups: one containing patients with radiculopathy only; and one containing patients with myelopathy with or without radiculopathy. Although the classic symptomatology differs between radiculopathy and myelopathy, both clinical and radiological differentiation can sometimes be challenging and potentially cause misclassification of registry patients. In a study of 127 myelopathy patients, 66 patients (52%) had radiculopathy symptoms, and arm pain presented as a characteristic of combined disease [231]. Today, Magnetic Resonance Imaging (MRI) is the ideal diagnostic tool for investigation of patients with degenerative spine disorders [232]. Preoperative MRI accuracy rates of cervical radiculopathy have been shown to be as high as 92% making it the only necessary imaging technique prior to surgery [233]. In relation to myelopathy, on the other hand, the extent of cord compression does not necessarily correlate well with the

clinical picture [9, 27]. In our study, the interpretation of the term "myelopathy" could be interpreted differently by different medical environments causing a systematic bias in the classification of patients. This, in turn, would then affect the number of patients and characteristics of the cohorts in all three papers. To further make the radiculopathy groups in paper I and III even more homogenous, all patients with a Ranawat score indicating myelopathy were excluded from the study. Due to lack of compliance by operating surgeons, the information about MRI and/or radiological findings were of no additional help in classifying patients correctly.

Another classification issue that might represent a bias in our studies is the grouping of patients undergoing anterior surgery as one population and another for patients undergoing posterior surgery. In Paper III, only 33 of the included 3142 patients received a disc arthroplasty (about 1%), while the rest underwent fusion surgery. Similar rates were found in the "anterior approach" populations in Paper I and II. One may question the relevance of including the small ACDA patient group, and whether the results of the fusion surgeries may be generalized to the disc arthroplasty surgeries, and wise versa. Recent reports have shown conflicting results in terms of outcomes after anterior fusion and arthroplasty. Although several US studies indicate cervical arthroplasty produce favorable outcomes compared to fusion, a randomized controlled study performed Oslo University Hospital found no difference in outcome between these two groups [234]. In our present study, the baseline characteristics, and outcomes between the ACDA and the ACDF group were strikingly similar, with only a few significant differences: higher baseline NDI (p=0.02) and NRS neck pain (p=0.002) for the arthroplasty patients, as well as a lower number of operated levels (p<0.001). Because of the minor differences between groups and since there is no current evidence in terms of differences in outcomes

between ACDA and ACDF, we believe it was justified to include the arthroplasty patients in the anterior approach cohort. Similar considerations were also made for the different techniques included in the posterior surgery groups in Paper I and II.

5.1.4 Cutoff estimates – methodological considerations

In Paper I and II, we present cutoff estimates for radiculopathy and myelopathy patients. Several previous studies have presented cutoff based on procedures rather than etiology, and a similar approach could have been chosen in our studies. ACDF, for instance, is often the preferred procedure for single-level disease regardless of etiology [47, 235]. However, the clinical presentation and symptoms of patients with radiculopathy and myelopathy differs substantially [6]. Also, the indications for surgical treatment, as well as choice of surgical approach, varies as described in the introduction. Therefore, we believe it is justified to investigate each cohort separately.

In both Paper I and II, we produced several sensitivity analyses of subgroups. In Paper I, separate cutoff scores were made for patients with spondylotic foraminal stenosis and patients with a disc herniation. We also made sensitivity analyses for procedural groups (anterior and posterior surgery). The latter analyses were repeated in Paper II.

In Paper I, the cutoff scores for cervical radiculopathy were similar across subgroups indicating that the same cutoff should be applied across different operative techniques and causes of disease. Although there were some differences in characteristics between groups, neither choice of procedure nor etiology seem to be crucial for the outcome. In terms of the differences, our figures suggest that a patient with a disc herniation is more likely to undergo anterior surgery rather than posterior

surgery. Patients with spondylotic foraminal stenosis is more frequently operated with a posterior approach procedure than patients with a disc herniation, but an anterior approach is still the more common of the two surgical modalities in patients with this type of etiology.

In Paper II, lower cutoff estimates for disability and pain were found for the myelopathy group undergoing posterior surgery compared to the group undergoing anterior surgery. The posterior surgery patients were older, had undergone surgery in more levels (1.4 vs 2.7), had a higher ASA-level score and more comorbidity, and were less likely to be working prior to the operation compared to the anterior surgery patients. They also experienced more severe myelopathy symptoms according to EMS. These findings support earlier reports that these procedural groups really are two different patient populations. Anterior approach surgery is typically used for younger myelopathy patients with a herniated disc, while posterior approach procedures are chosen for multi-level disease patients with bony degenerative changes in the cervical spine [47]. In this sense, it is justified to use different cutoff estimates for each population. According to our study, the posterior approach patients seemed to be satisfied with less improvement in pain and disability after surgery. This may be due to a set of different reasons, for instance lowering of expectations by the operating surgeon prior to intervention, or – since the anterior surgery group is younger and more likely to have an active work life – the need to be physically fit is bigger for this patient group.

5.1.5 Information bias

Information bias is a term describing systematic errors in how data are collected, interpreted, or measured. Random or statistical error, which is inherently

plays a role in all types of measurements, is not systematic and, therefore, not a type of information bias [236].

In NORspine, like for most large databases, an important limitation is inaccuracies in relation to data registration [217]. Inaccurate registration becomes an information bias when it is not done at random, i.e., the instrument/questionnaire used for measurement is not valid. For instance, such lack of validity can be caused by illogical questionnaire design creating different interpretations among respondents regarding how to fill out a form. This could be the case among patients answering independent questions about personal characteristics or other variables related to work status. Since Norwegian neurosurgeons are well aware of the NORspine questionnaire and submit forms regularly, this should be a marginal problem in relation to the physician reporting. However, one item in the surgeon´s form seemed to cause some confusion among responders (Appendix D). In relation to registration of operation method, surgeons were asked to checkmark whether the patient were operated with a posterior or anterior procedure. For the anterior alternative, there were two additional checkboxes for either right or left skin incision on the neck. Some surgeons seemed to interpret this as a checkmark for right and/or left nerve root decompression, and, followingly, several responders had check marked both the left and right item box. Although our research group did not use the information about skin incision in our studies, this example is a good illustration of the importance of questionnaire design. It also highlights the upsides of including well-validated PROMs or other validated questionnaires to reduce response variance and differences in question interpretation.

Underreporting is another phenomenon categorized as information bias. In the medical environment, intentional underreporting of adverse events after spine surgery has been documented in the Swedish spine registry [237]. A recent study of lumbar

patient data in NORspine, also showed weak agreement for surgeon-reported complications and perioperative adverse events compared to information in the electronic health record [238]. In our data, complication rates were extremely low although the surgeon's questionnaire contained checkboxes with a list of options to simplify the registration of adverse events. The low rate was most likely caused by underreporting. Due to lack of statistical power, these figures were not included in our studies.

Patients are also prone to underreporting, especially in case of smoking habits and analgesic use [239-241]. Such underreporting may have influenced the statistical power of these variables in the prognostic model in Paper III.

5.1.6 Outcome definitions and measures

In our studies, we applied several commonly used PROMs in relation to cervical degenerative disorders. Among the disease-specific PROMs, NDI is the most widely used and investigated in literature. NDI measures disability across ten different items and has shown to have high discriminative ability for both conservative and operative treatment [187]. EMS [191] is a more seldom used PROM measuring myelopathy severity across several items, and we have found no available cutoff scores in literature. In terms of pain assessment, NRS-AP and NRS-NP are widely used across different treatments and etiologies [189, 242]. They have shown to be accurate for cervical degenerative disorders in several studies [243-245]. Finally, EQ-5D is a generic tool applied across several treatments and diseases [190].

In paper I and II, we created cutoff estimates for the mentioned PROMs by using an anchor question and producing Receiver Operating Characteristic (ROC) curves and cutoff estimates representing yes-or-no-thresholds regarding improvement

[208]. Based on a continuous scale, the cutoff estimates answer the question whether an intervention has been successful or not successful. Each cutoff estimate is the point on the ROC representing the highest sensitivity (highest fraction of true positives) and highest specificity (highest fraction of true negatives). The figures are, therefore, not associated with a p-value.

Since previous methodological studies for PROM cutoff estimates for successful or non-successful outcomes for lower back conditions have shown to be dependent upon the PROM baseline score [132], we chose to provide cutoff estimates in Paper I and II for both absolute scores at follow-up, change scores and percentage change scores from baseline to 3- and 12-month follow-up.

Change scores are widely applied in research of cutoff estimates for beneficial/nonbeneficial outcomes. However, the higher baseline value, the more change or improvement in a condition is needed for a patient to conceive the change as clinically important [246]. Thus, the change score has a clear dependency on the baseline level which makes it less suitable for application across patients with different degrees of condition severity. In Paper I and II, change score estimates generally showed lower discriminatory ability (AUC) than absolute score and percentage change score estimates. Also, the sensitivity and specificity for the change score estimates were for most part lower compared to those for the other two modalities.

There is little tradition for the use of *percentage change* cutoff scores in literature today. Both in Paper I and II we recommend further use of this modality because of high accuracy. This finding is in line with previous NORspine studies conducted on patients undergoing surgery for lumbar disc herniation [136] and lumbar spinal stenosis [132, 133]. The percentage change score adds a different

dimension to the cutoff score, as it tells something about the actual improvement the patient has been through. Also, in a research setting, it is often necessary to adjust for the baseline score when doing statistical analysis in e.g., non-randomized trials. However, percentage change scores are a little more complicated to use compared to the absolute score and the change score since the final score needs to be calculated. For practical purposes, one study suggests a 30% change from baseline may be considered a meaningful improvement [247]. However, this suggestion does not take into consideration cohorts of patients with baseline scores on the upper or lower part of a scale. A patient with a high initial NDI, for example, will expect more improvement than a patient with a medium baseline score.

The percentage change score has been criticized for being highly sensitive to changes in variance. In theory, percentage change score does not correct for baseline imbalance between groups. Also, it can lead to create non-normally distributed outcome data from a normally distributed data sample [248]. This warrants caution when comparing the effect of surgery between different cohorts of patients.

In a clinical setting, the *follow-up/absolute score* has it´s clear advantages, as it does not involve any calculations, it merely states what the patient can expect postoperatively. However, as for the change score, one would expect a dependency on the baseline score and that a patient with high initial arm pain will be less likely to achieve the cutoff for success of less than 3 points, as reported in Paper I. In our analyses of the NRS arm pain results, we found that area under the curve (AUC) for the absolute score was 0.86 and with a 95% confidence interval of 0.84-0.88. An AUC of 0.9 is regarded as "excellent" and suggesting that it is highly accurate to state that patients with less than 3 in postoperative arm pain conceive themselves as being "much improved" or "completely recovered" following intervention. Therefore, based

on the results from our cohort, we believe the use of absolute scores as cutoffs for cervical radiculopathy patients is justified.

A weakness of the methodology applied in Paper I and II was the lack of assessment of the MDC, which quantifies the smallest amount of change possible detected beyond underlying measurement error. A change score should be greater than MDC to be found relevant [206], but these calculations were omitted from the analyses.

In a recent study by Taso et al., cutoff estimates reported in Paper I are compared with outcome expectations reported by patients prior to surgery for cervical radiculopathy. The conclusion is that patients expect "a higher improvement than the proposed criteria of 13.5 NDI points" [249]. This illustrates how the cutoff estimates from our investigations could help set expectations for patients referred to cervical spine surgery and improve the shared decision-making process.

5.1.7 Prognostic measures

After a review of literature, twenty potential prognostic factors associated with lack of success in surgery for cervical degenerative radiculopathy were selected for the univariate analyses in Paper III. However, some of the identified factors from the review were not registered in NORspine and, therefore, could not be included in the analyses. Race, baseline mJOA scores and ambulation assistance prior to surgery were such examples. These factors have shown to have significant impact on surgical outcomes according to several studies (Table 1-3) and including them in our analyses could potentially have improved the performance of our prognostic models.

In retrospect, there are also a few relevant prognostic factors that can be found in NORspine and should have been included in the regression analyses. Quality of life

measures, such as baseline values of EQ-5D or EQ-VAS, are such examples. These factors could have added a different dimension to our prognostic models and potentially improved their performance. Also, radiological evidence of listhesis was registered in the NORspine questionnaire and considered for inclusion in the univariate analyses. However, the surgeon compliance rate related to the documentation of radiological findings was found to be very low, and this underreporting would have influenced the results. The variable was, therefore, omitted from the analyses.

5.1.8 Statistical analyses

In Paper I and II, when investigating the NORspine data, floor and ceiling effects were considered to be present if >15% of the respondents achieved the minimum or maximum possible score [250]. Such effects represent a measurement issue since the clustering of data in either end of a scale illustrates the limited ability of an investigative tool, for instance a PROM, to accurately assess the truth about a respondent and distinguish between different respondents at each end of a scale. Thus, with a floor and ceiling effect present, a change in a patient may not be measurable, and an investigation may fail to capture the true range of PROM values, leading to little variance in the data set. NDI [187] and EQ-5D [251] have been criticized for being subject to floor and ceiling effects. However, we found no such effects among the investigated PROMs.

A limitation in terms of the cutoff score is the use of the GPE scale as an anchor. The GPE scale captures the subjective improvement of a patient by asking how a patient feels compared to the situation at baseline, which in this case is before surgery. One could argue that patients tend to be more focused on their current state

rather than the health change. Therefore, the GPE can be regarded as a very subjective scale. An ideal anchor should objectively measure the patient's status before and after neck surgery. However, there exist no such gold standard today. More objective tools, such as a question about "return to work", come with a lot of limitations as these modalities might not reflect the actual improvement that the patient goes through.

The strengths and weaknesses of the GPE scale have been reviewed in several studies, and the scale has been found to be a reliable assessment of health transition in people with musculoskeletal disorders, as well as for back pain and upper/lower extremity disorders [192, 252, 253]. We argue that, despite its limitations, the GPE is highly applicable as an anchor. According to our investigations, the scale aligns nicely with the PROMs included in the studies of both radiculopathy and myelopathy.

There is an ongoing debate whether variables tested in a univariate or multivariate regression analysis should be categorical and not continuous [254]. On the positive side, dichotomization or categorization makes an interpretation of a variable simpler both in a research and clinical setting. An illustrative example is the use of a normal blood pressure (BP) level of 140/90. Patients with BP above this threshold, are recommended medical treatment and diagnosed with "high BP". This categorization can easily be applied in research as either an endpoint, prognostic factor, or characteristic. The PROM thresholds of success and MCID applied in Paper I-III are of similar stature. However, this simplification causes loss of statistical power; you need more observations of a dichotomized variable than a continuous variable in order to get the same statistically equivalent results [255]. Also, by pooling groups of patients, within-category information is lost. Dichotomized variables do not make full use of the information in the response scale. Due to the introduction of a threshold, the risk of misclassification because of measurement error becomes higher,

as well. Therefore, one should be aware that the alterations could have an impact on the associations between the variable and the outcome in question. In addition, it is important to look to other comparable studies to see how the dichotomizations are done to be able to compare across different models [254, 256].

In Paper III, to illustrate association strength between an unsuccessful outcome and prognostic factors, we used Odds ratios (OR). OR represented the probability of experiencing an unsuccessful outcome for patients with a certain trait or characteristic compared to someone experiencing the outcome without the trait or characteristic. OR should be considered in relation to the frequency of the outcome [257]. Since a fair portion of patients (38% or 35%) can expect to become only slightly better or worse following surgery, achieving an unsuccessful outcome, in this regard, is clinically highly relevant.

In the two case examples of Paper III, we have tried to illustrate the usefulness of the prognostic model for neck disability, which was the model with the best performance presented in the article. We wanted to present cases which were recognizable to surgeons using a limited number of easily assessable and recognizable patient features: physical demands in work, level of education, pending litigation, previous neck surgery, duration of arm pain and baseline NDI. A low-risk patient was exemplified to have university-level education, <3 months of arm pain, medium level NDI score, and none of the remaining features. The high-risk case example, on the other hand, involved a patient who experienced hard physical demands in work, had been through previous neck surgery and was involved in pending litigating activities, had over one year duration of arm pain and high level of baseline NDI. There was a probability of 0.92 or a 92% chance for the high-risk patient to obtain a nonsuccessful

outcome after surgery (i.e., "much better" or "completely recovered" on the GPE scale), while a low-risk patient only had a 13% chance of nonsuccess.

5.2 External validity

After ensuring the internal validity of the study, researchers should evaluate its external validity to determine whether the findings are applicable to similar patients in different settings. This means that the cutoff scores produced in Paper I and II, as well as the prognostic model in Paper III, need to be externally validated for other patient populations. Such a validation process should be performed using independent data from a different location or cohort than the NORspine population [258].

To enhance internal validity, investigators should meticulously plan the study and implement quality control and recruitment strategies, data collection, analysis, and sample size. In contrast, external validity can be improved by using broad inclusion criteria to select a study population that closely resembles real-life patients and feasible interventions easily reproduced in clinical trials [259]. The wide selection criteria of NORspine, therefore, is a strength in terms of external validation. All patients referred to surgery for cervical degenerative disease are included. The criteria for patient selection, however, are not apparent when interpreting the registry data. Thus, these features can come to have an impact on a reproduction study.

In terms of prognostic models that performs less well in a different environment, an adjustment will need to be made. Such adjustment efforts are called *updating*, which can vary from minor recalibration measures to whole model revisions, meaning everything from adjustments of individual prognostic variables to inclusions of new ones [260]. Finally, investigations should be conducted in order to evaluate whether the model has any actual impact on patient care [261].

5.3 Comparison with other studies

5.3.1 Cutoff estimates for cervical radiculopathy and myelopathy

In both Paper I and II, we found that NDI showed superior discriminative ability in defining cutoff estimates for a beneficial outcome following surgical treatment of cervical radiculopathy and myelopathy compared with other PROMs collected in NORspine. NDI is considered one of the more accurate PROMs in relation to cervical degenerative disorders [188], and these findings are supported by studies of lumbar disorders of the Oswestry disability Index (ODI) [134, 136] which NDI originally derived from [262]. NRS-AP was the second-best PROM to distinguish between success and nonsuccess for cervical radiculopathy. Since arm pain is a distinct feature of cervical radiculopathy and measures of arm pain have proven to have acceptable discriminative ability in other studies [243-245], this was an expected finding. For cervical myelopathy patients, neck pain is a main symptom and affecting 40 to 80% of patients [263]. NRS-NP showed to be superior to NRS-AP in assessing cutoff estimates in cervical myelopathy patients. Neck pain AUCs have previously been shown to be acceptable for cervical myelopathy patients [244].

For both the radiculopathy group and the myelopathy group, we found better discriminative ability for the percentage change scores and the absolute scores compared to the change scores. Similar results have been found in NORspine studies for lumbar disc herniation [136] and lumbar spinal stenosis [132]. We found no other studies reporting percentage change scores or absolute scores for patients operated due to cervical radiculopathy and/or myelopathy. For change scores, several studies are published [244, 264-268].

In Paper I, we assessed cutoff estimates for success for patients undergoing one- or two-level surgery due to cervical radiculopathy using NDI, NRS-AP, NRS-

NP, EQ-5D and EQ-VAS. The seven-step GPE scale was used as an anchor, and "success" was defined as one or two steps on the scale. In literature, there are multiple studies that have produced anchor-based cutoffs for cervical radiculopathy patients, especially for NDI and the pain scales. Most of these are MCID cutoff estimates [166, 243, 245, 269], while only a few report cutoff estimates for success, or substantial clinical benefit (SCB) [207, 245, 270]. Among the latter studies, Carreon et al. found slightly higher cutoff estimates compared with our findings for cervical spine fusion despite the inclusion of both radiculopathy and myelopathy patients. In a group of 505 patients followed for a year, the SCB change score was found to be 19 for NDI (vs. 13.5 in our study), while it was 3.5 for pain scales (vs. 2.5 for NRS-AP and 1.5 for NRS-NP) [245]. As in our study, NDI was shown to have the best discriminative ability (AUC=0.823). However, opposite to our findings, neck pain scores (AUC=0.788) were more accurate than arm pain scores (0.716), which may be due to the inclusion of myelopathy patients. There are some differences in the methodology that is notable. Carreon et al. applied only the top level of a five-step anchor as a cutoff (1/5), which represents a higher threshold than the one used in our study (2/7). Also, it was a single study of markedly fewer patients (505) with significantly different baseline characteristics, like higher age (53y vs 49y), lower female ratio (33% vs 47%), fewer smokers (17% vs 33%) and higher baseline NDI scores (53 vs 41).

Two other single-center studies with drastically smaller cohorts applied a five-point anchor and a one-step threshold (1/5) to assess SCB. Donk et al. investigated 80 patients with radiculopathy undergoing single level anterior surgery and found a cutoff of 20 [207]. Patients were significantly younger (45 vs 49), more likely female (59% vs 47%) and had higher mean NDI baseline score (38 vs. 33). AUC (0.7) was

slightly lower compared with our findings. Steinhaus et al. followed a cohort of 48 radiculopathy patients undergoing cervical spine surgery for 6 months [270], and found several SCB cutoff estimates close to our findings (NDI: 11, VAS arm pain: 1.5, VAS neck pain: 3.5). The patients were healthier and experienced lower preoperative pain (4-4.8 vs. 6.1-6.4), although NDI scores were similar (39 vs 41). Again, NDI showed better discriminative ability (0.72), than arm pain (0.68) and neck pain (0.67).

We have found only one study investigating cutoff estimates for EQ-5D. Using a 4-item anchor question, the MCID change score in patients undergoing ACDF (0.05) was found to be lower than our success cutoff estimate (0.11) [243]. As in our study, the AUC of EQ-5D was lower than AUCs found for disease-specific PROMs. Since EQ-5D is a generic PROM, this is as expected. When assessing change in relation to surgery, no single PROM will completely cover the whole clinical picture and using different tools is, therefore, necessary. As in our study (AUC=0.74), EQ-5D was found to be acceptable with an AUC of 0.67.

Cutoff estimates for EQ-VAS have previously been reported only for lumbar surgery [271] with similar results for discriminative ability as in our study. As a PROM, EQ-VAS is easier to use than EQ-5D, and it also allows for calculation of the percentage change score.

In terms of cervical myelopathy, several assessments of MCID have been made although the investigated cohorts are considerably smaller than the one in Paper II. Most of the reported MCID estimates are change scores and slightly higher than our findings (NDI: 4.3, NRS-AP/NP: 0.5, EQ-5D: 0.02, EMS: 0.5). Also, many studies have produced estimates for PROMs which are not collected in NORspine,

such as Patient-Reported Outcomes Measurement Information System (PROMIS) and mJOA.

Chien et al. reported a change score of 6 for NDI using four steps on an eleven-step anchor in patients operated with anterior or posterior techniques by the same surgeon [265]. Although there were only 45 patients in the study, the sensitivity and specificity were both 0.67, which are the same as in our study (0.68/0.68 for MCID change score). Patients were slightly younger (56y vs 59y), more likely female (49% vs 39%), and preoperative NDI was markedly lower (15 vs 34).

Auffinger et al. similarly investigated a smaller sample of myelopathy patients (30) undergoing ACDF and used five different methods to assess MCID for NDI (4.8-13.4) and NRS-NP (0.36-3.11) [244]. The cohort is comparable to our anterior approach cohort of myelopathy patients, for which the change score was 5.9 for NDI and 0.5 for the pain scores. In the study of Auffinger et al., the patients were slightly older (58y vs. 54y), had a higher female ratio (53% vs. 44%), higher percentage of smokers (40% vs. 25.5%), slightly higher Body Mass Index (BMI) (28 vs 27), and baseline NDI was slightly lower (29 vs 34), while preoperative pain was the same (VAS 5).

Kato et al. investigated a MCID change score for NDI and EQ-5D using a 7-point Likert scale in a cohort of 101 myelopathy patients undergoing laminoplasty [266]. The NDI cutoff of 4.2 is slightly higher than the 2.4 cutoff found in our posterior approach group, but similar to the cutoff for the whole myelopathy group (4.3). The EQ-5D cutoff of 0.0485 is also higher than the cutoff estimates for our cohorts (0.02 for all groups). Patients selected for laminoplasty rather than laminectomy often have less neck pain than other patients with cervical myelopathy [82]. The pain level was not reported in the study, but it is likely that the selection of

patients has had an impact on the results and might have affected the NDI cutoff estimate.

Several studies with a mix of cervical radiculopathy and myelopathy patients have assessed MCID estimates for cervical spine surgery [243, 245, 270]. Most of these cutoff estimates are higher than the estimates found in our study. This may be due to the smaller amount of improvement expected for myelopathy patients. For example, Carreon et al. found MCID for NDI to be 7.5 (vs. 4.3), and to be 2.5 for arm pain and neck pain (vs. 0.5) [245].

5.3.2 Prognostic models for outcomes in treatment of cervical radiculopathy

When comparing different prognostic models, both how the models perform, which available variables are tested and what the final output is, need to be considered. In Table 1, the available prognostic models for patients undergoing surgery for cervical radiculopathy found in literature, are listed. Only two of these investigate factors associated with a poor outcome [166, 167], while three focus on factors predicting a beneficial outcome [163, 165, 175]. In addition, there exist some multivariate analyses based on mixed cohorts of both myelopathy and radiculopathy patients [164, 168, 173, 174] (Table 3).

In terms of number of patients, the population in the study of Archer et al. is markedly larger than those in the other available studies in literature. A cohort of 7629 patients undergoing elective cervical spine surgery is investigated with separate cohorts for both cervical radiculopathy (4988) and myelopathy patients (2641) [167]. This study is the only one of the two models for poorer outcome reporting performance measures. Compared to our study, the authors report slightly lower AUCs for a predictive model for NDI (0.64–0.69 vs. 0.78), arm pain (0.63–0.65 vs.

0.68), as well as neck pain (0.63-0.67) one year after intervention. No goodness-of-fit measure is presented by the authors. The three prognostic models for a beneficial outcome after surgery for cervical radiculopathy all have study populations of approximately 100 patients undergoing ACDF [163, 165, 175]. The authors all report goodness-of-fit measures (14-30%) similar to those found in our study (17.3 and 27.3%), but no AUC estimates or calibration plots.

The populations in the four prognostic models investigating mixed cohorts of both radiculopathy and myelopathy patients vary markedly. Investigating only 34 patients, Peolsson et al. report R^2s of 38-78% [173]. Lubelski et al., on the other hand, investigate a cohort of 952 patients to define prognostic models for good quality of life outcomes and find R^2 to vary from 35% to 47% [174], which suggest a high goodness-of-fit for these models. The third multivariate analysis of a mixed cohort of 488 patients reports no performance measures [164] which is a limitation in terms of interpreting the results. The same is the case for the fourth study [168].

In terms of prognostic variables, there is significant overlap between our study and the model presented by Archer et al. for cervical radiculopathy patients [167]. Archer et al. found that longer symptom duration, workers' compensation claims, and higher baseline NDI were significantly associated with worsening of disability and arm pain scores, all of which are included in our equivalent models. Consistent with our findings, Archer et al. also found that depression was only significantly associated with worse NDI scores and not arm pain scores, and that BMI, number of surgical levels and motor deficit was not associated with neither worsening nor improvement in any PROM.

In our study, low level of education was found significant for both disability and arm pain. Accordingly, high level of education was found to be a positive

predictor of improved outcomes by Archer et al. On the other hand, smoking was included as a significant factor in our arm pain model but was not found significant for worse disability or arm pain. A nonsmoking status is found to be significant in several prognostic models for a beneficial outcome [163, 165, 175]. Archer et al. did not investigate the association of physical demands in work or previous neck surgery, which were significant factors in both our models. Foreign mother tongue, found to be significant in our arm pain model, was also not investigated. White race, however, was associated with improved disability and neck pain, while black race was not significantly associated with any outcome.

Gender was not found to be significantly associated with any outcome in our study. However, Archer et al. found that female sex was a predictor for worse neck disability scores but not for worse arm pain scores. This is supported by other studies tying male sex to a beneficial outcome [162, 163]. Archer et al. also found that ambulation assistance prior to surgery, higher baseline neck pain and posterior surgical approach were predictors of a worse outcome. None of these factors was investigated in our study. In terms of improved scores, Archer et al. found greater age, preoperative employment, private insurance, greater arm pain, higher education, and evidence of listhesis to be significant for both disability and arm pain.

Narain et al. produced a prognostic model of risk factors associated with failure to reach MCID for disability, arm pain and neck pain for patients undergoing one- or two-level ACDF for cervical radiculopathy [166]. This study has several limitations. The patient population is constituted by only 84 patients, and the authors do not report any performance measures. For the arm pain models, no significant factors were identified, and for disability, only comorbidity burden, as evidenced by Charlson Comorbidity Index, was found significant. In our study, comorbidity and

ASA level only emerged as significant predictors in our univariate analyses but not in the final multivariate analyses for disability nor arm pain. The results of Narain et al. must be seen in light of the marginal patient population and lack of statistical power.

In terms of mixed cohorts with both cervical radiculopathy and myelopathy patients, Anderson et al. performed a retrospective multivariate analysis of 488 patients undergoing single-level ACDF to identify important predictors of a poor outcome [164]. Again, no performance measures were presented. The outcomes were >15-point NDI improvement and "overall clinical success", which was based on an assessment of the following criteria: >15-point improvement in NDI; maintained or improved neurologic examination; no serious adverse event related to the procedure; and no revision of the plate or graft. For overall success, worker's compensation claims and preoperative sensory loss came out as significant negative prognostic factors. For NDI success, greater age, high baseline NDI, and preoperative working status were positive factors, while litigation was significantly associated with poorer outcome. In conclusion, the results of Anderson et al. further strengthen the evidence that litigation issues and claims are associated with lack of improvement following cervical spine surgery. The findings of greater age and preoperative employment as positive prognostic factors are also supported by Archer et al. Preoperative sensory function, however, was not investigated in our study and not found significant by Archer et al.

Scerrati et al. [168] also investigated a mixed cohort. Again, female sex was found to be a prognostic factor for poor outcome, while the two other included factors (two-level surgery and collar use) has not been reported previously, the latter possibly since it is not a part of routine treatment according to guidelines.

In terms of achieving a good outcome for disability and pain after surgery for cervical radiculopathy, the three models for cervical radiculopathy report varying factors associated with a good outcome [163, 165, 175]. Peolsson et al. [175] highlight a low baseline NDI as an important predictor for long-term NDI success after ACDF. Hermansen et al. [163] report male sex and nonsmoking status as significant prognostic factors for a 10-year improvement in NDI while the neck pain model showed high initial neck pain to be the only significant factor. In a 2003 investigation of 1- to 2-year improvement in NDI and pain (R^2 28% and 30%, respectively), several factors were identified [165]. For NDI, higher educational level, non-smoking, greater kyphosis at the level operated on, a greater flexion mobility, greater right handgrip strength and lower current pain intensity were all significant. For the pain model, male sex, greater kyphosis at the level operated on, nonsmoking, a greater neck mobility, low disability on NDI, and older age were found to be associated with an improvement.

Lubelski et al. is the only study investigating a prognostic model for quality-of-life scores [174]. In the model, the preoperative quality-of-life baseline scores had the largest effect on the outcome. To further increase the goodness-of-fit for future prognostic models, this suggest that generic PROMs, should be included in the multivariate analyses of patients undergoing surgery for cervical degenerative disease. As mentioned earlier, we did not include EQ-5D or EQ-VAS in our analyses in Paper III although they are collected in NORspine and potentially could have strengthened our models' performance.

5.4 Clinical implications and further research

Today, there are major differences between health systems in terms of both choice of payment models and level of spending [272]. The costs are twice as high in the United States compared to other Western countries, but health outcomes on a population level are lower with i.e., lower life-expectancy rates [273].

Although there are international guidelines in spine surgery, the structural differences in each system influence incentives to treat, and operation rates vary greatly across countries [274]. At the same time, spinal surgery rates have increased considerably in many Western countries over the last decades [275-280]. This increase cannot be explained by higher incidence or prevalence rates of spinal disorders, and there are marked variations in spinal surgery rates also within countries or regions [40, 274, 281, 282].

The Nordic countries are single-payer systems with salaried physicians and very similar incentives to treat across systems. Still, a study examining incidents rates for lumbar surgery found major differences between countries [283], and it is likely that the same is the case for cervical degenerative surgery.

Even for Norway, the operation rates for cervical degenerative disorders seem to differ between more rural and urban regions [40]. This variability within the NORspine population may, however, be a benefit in terms of translating the findings to populations elsewhere, as it makes our studies more robust and generalizable. At the same time, spine surgery is constantly evolving. Only the last decades we have seen operating techniques develop from open surgical approaches to minimal invasive surgery. The use of MRI has revolutionized the diagnostic process, and now robotics are slowly being introduced [43]. This development will come to influence the relevance of the cutoff estimates and prognostic models presented in this thesis and

pave the way for constant revalidation processes and adjustments. In this regard, a strength of the NORspine dataset is that registered patients consent to be contacted again for data collection and that there is no time limit for the use of data and merging with other registries (Appendix A).

In terms of healthcare expenditures, the recent report of the national Health Care Commission (Helsekommisjonen) concluded that the sustainability of the current healthcare delivery system will be challenged the coming years by ageing populations, increasing rates of chronic disease and more available treatments [284]. Although numerous efforts to further digitalize workflows for health care personnel, the output has not been more effective work processes. We, therefore, need to start looking at other ways to reform our care delivery, such as redesigning our financing models. Value-based health care, in this context, aims for systems to focus more on value than volume. This also mean having a broader focus on outcome measurement. If we are to do more for less resources, monitoring of care quality is necessary.

In spinal care, validated measures or benchmarks of favorable outcomes are important for patients and providers to set expectations for recovery following surgery prior to treatment and to retrospectively assess whether the wanted result is achieved. On a provider level, the same measures can be used to minimize variation in patient care by comparing data across different treatments or interventions to assess the optimal care. Also, consensus regarding validated benchmarks across institutions can help assess quality of care and contribute to more value for patients.

In terms of contributing to enhance quality in surgery for cervical degenerative disorders, our results can serve at least four purposes:

1) The definitions (cutoff estimates) of a beneficial treatment outcome can be applied to set patient expectations in a clinical setting. By actively using

the results from Paper I-II, for example in relation to pain assessment, the treating surgeon can better communicate the likely outcome of an operation.

2) The cutoff estimates can be applied in monitoring the quality of care on a department or even national level. This is partly done today through the Centre for Patient-Reported Outcomes Data [285]. Nevertheless, it is possible that the Norwegian neurosurgical community could place even greater emphasis on the utilization of collected data. Based on our experience, it seems that surgeons are interested in accessing a greater amount of outcome data than they currently have available today.

3) The cutoff estimates for radiculopathy and myelopathy can be used to create quality benchmarks for bundled payment models in relation to surgery for cervical degenerative disorders. Since the cutoffs are stable across several procedures and for different etiologies, especially for radiculopathy, they should be highly applicable for such initiatives.

4) The application of our prognostic models during patient evaluations of anterior surgery for cervical radiculopathy can improve the shared decision-making process and lead to improved selection of patients for surgery.

In future research, the cutoff values in Paper I and II are representative for the NORspine population and could be used to further investigate the population. A recent study from NORspine did not apply the cutoffs from Paper I, but instead used internationally produced thresholds for substantial clinical benefit [286].

Finally, there is a need for a predictive model for patients operated due to cervical myelopathy. So far, the myelopathy cohort in NORspine has been evaluated not to be large enough in terms of producing results with sufficient statistical power. However, since the registry cohort is constantly growing, such calculations will be possible to conduct in the near future.

Since 2020, the questionnaires of NORspine have become electronically accessible through the national HelseNorge portal [118] offering patients to opt-in online before even seeing a specialist and allowing providers to follow up nonrespondents online. This have already had a major impact on the respondent rate and will further reduce the attrition bias and improve the validity of the database. In this sense, NORspine can come to play a crucial role in the development of international spine surgery in the coming years.

6. Main conclusions

In this thesis, the aim was to assess outcomes and benchmarks for clinical improvement after surgical intervention of cervical radiculopathy and myelopathy in Norwegian public hospitals and private clinics and to provide information about prognosis from anterior surgery for cervical radiculopathy.

In the first paper, we found that NDI had the highest discriminative ability in terms of assessing a benchmark for "success" after surgery for cervical radiculopathy. The NRS-AP cutoff estimate was the second-best benchmark for a successful outcome. All PROMs showed more than "acceptable" accuracy, and percentage change and follow-up/absolute scores were more accurate than change scores.

In the second paper, MCID cutoff estimates for patients undergoing surgery for cervical myelopathy were investigated. The NDI and NRS-NP were the superior PROMs for accurately assessing a MCID, and the percentage change scores were more accurate than the change scores. The cutoff estimates for anterior surgery were slightly higher for NDI, NRS-AP and NRS-NP compared to those for posterior surgery suggesting that patients undergoing anterior surgery should expect more improvement than those undergoing posterior surgery. The latter patient group were older and had higher morbidity, and this may be a contributing factor. The proportion of patients achieving a MCID varied between 51-61% indicating that a majority of patients operated for cervical myelopathy should expect symptom improvement and not only a cease in symptom progression. This study is the first of our knowledge to produce cutoff estimates for EMS, which showed to have acceptable discriminative ability.

In the third paper, prognostic models for lack of surgical success 12 months after surgery for cervical radiculopathy were developed by using absolute scores of

NDI and NRS-AP from Paper I. Approximately 38% (NDI) and 35% (NRS-AP) of patients did not achieve a successful outcome according to the benchmarks. The prognostic model for NDI showed high discriminative performance, and significant characteristics were physical demanding work, low level of education, pending litigation, previous neck surgery, duration of arm pain>3 months, medium to high levels of baseline disability, as well as anxiety/depression. For the arm pain model, the accuracy was slightly lower but still acceptable. Patient characteristics with significant association with a nonsuccessful outcome of arm pain were physical demanding work, low level of education, pending litigation, previous neck surgery, duration of arm pain>3 months, medium to high levels of baseline disability, foreign mother tongue, smoking, and medium to high levels of baseline arm pain. Further validation of the models is warranted.

7. References

1. Dakwar, E., et al., *Chapter 16 - Spinal Pathology, Conditions, and Deformities: Surgical Intervention*, in *Pathology and Intervention in Musculoskeletal Rehabilitation (Second Edition)*, D.J. Magee, et al., Editors. 2016, W.B. Saunders: Maryland Heights, Missouri. p. 584-611.
2. OECD, *Fiscal Sustainability of Health Systems*. 2015.
3. Teisberg, E., S. Wallace, and S. O'Hara, *Defining and Implementing Value-Based Health Care: A Strategic Framework.* Acad Med, 2020. **95**(5): p. 682-685.
4. Fehlings, M.G. and B. Arvin, *Surgical management of cervical degenerative disease: the evidence related to indications, impact, and outcome.* J Neurosurg Spine, 2009. **11**(2): p. 97-100.
5. Woods, B.I. and A.S. Hilibrand, *Cervical radiculopathy: epidemiology, etiology, diagnosis, and treatment.* J Spinal Disord Tech, 2015. **28**(5): p. E251-9.
6. Theodore, N., *Degenerative Cervical Spondylosis.* N Engl J Med, 2020. **383**(2): p. 159-168.
7. McCormack, B.M. and P.R. Weinstein, *Cervical spondylosis. An update.* West J Med, 1996. **165**(1-2): p. 43-51.
8. Bakhsheshian, J., V.A. Mehta, and J.C. Liu, *Current Diagnosis and Management of Cervical Spondylotic Myelopathy.* Global Spine Journal, 2017. **7**(6): p. 572-586.
9. Davies, B.M., et al., *Degenerative cervical myelopathy.* Bmj, 2018. **360**: p. k186.
10. Nagata, K., et al., *Prevalence of cervical cord compression and its association with physical performance in a population-based cohort in Japan: the Wakayama Spine Study.* Spine (Phila Pa 1976), 2012. **37**(22): p. 1892-8.
11. Lee, M.J., E.H. Cassinelli, and K.D. Riew, *Prevalence of cervical spine stenosis. Anatomic study in cadavers.* J Bone Joint Surg Am, 2007. **89**(2): p. 376-80.
12. Kovalova, I., et al., *Prevalence and Imaging Characteristics of Nonmyelopathic and Myelopathic Spondylotic Cervical Cord Compression.* Spine (Phila Pa 1976), 2016. **41**(24): p. 1908-1916.
13. Nagata, K., et al., *Physical performance decreases in the early stage of cervical myelopathy before the myelopathic signs appear: the Wakayama Spine Study.* Eur Spine J, 2019. **28**(5): p. 1217-1224.
14. Donnally, C.J., 3rd, et al., *Current Management of Cervical Spondylotic Myelopathy.* Clin Spine Surg, 2022. **35**(1): p. E68-e76.
15. Abbed, K.M. and J.V. Coumans, *Cervical radiculopathy: pathophysiology, presentation, and clinical evaluation.* Neurosurgery, 2007. **60**(1 Suppl 1): p. S28-34.
16. Galbusera, F., et al., *Ageing and degenerative changes of the intervertebral disc and their impact on spinal flexibility.* Eur Spine J, 2014. **23 Suppl 3**: p. S324-32.
17. Grunhagen, T., et al., *Intervertebral disk nutrition: a review of factors influencing concentrations of nutrients and metabolites.* Orthop Clin North Am, 2011. **42**(4): p. 465-77, vii.

18. Ohara, Y., *Ossification of the Ligaments in the Cervical Spine, Including Ossification of the Anterior Longitudinal Ligament, Ossification of the Posterior Longitudinal Ligament, and Ossification of the Ligamentum Flavum.* Neurosurgery Clinics of North America, 2018. **29**(1): p. 63-68.

19. Boody, B.S., M. Lendner, and A.R. Vaccaro, *Ossification of the posterior longitudinal ligament in the cervical spine: a review.* Int Orthop, 2019. **43**(4): p. 797-805.

20. Howe, J.F., J.D. Loeser, and W.H. Calvin, *Mechanosensitivity of dorsal root ganglia and chronically injured axons: a physiological basis for the radicular pain of nerve root compression.* Pain, 1977. **3**(1): p. 25-41.

21. Rothman, S.M. and B.A. Winkelstein, *Chemical and mechanical nerve root insults induce differential behavioral sensitivity and glial activation that are enhanced in combination.* Brain Res, 2007. **1181**: p. 30-43.

22. Kim, J.H., et al., *Activated macrophage-like THP-1 cells modulate anulus fibrosus cell production of inflammatory mediators in response to cytokines.* Spine (Phila Pa 1976), 2008. **33**(21): p. 2253-9.

23. Kokubo, Y., et al., *Herniated and spondylotic intervertebral discs of the human cervical spine: histological and immunohistological findings in 500 en bloc surgical samples: Laboratory investigation.* Journal of Neurosurgery: Spine SPI, 2008. **9**(3): p. 285-295.

24. Peng, B. and M.J. DePalma, *Cervical disc degeneration and neck pain.* J Pain Res, 2018. **11**: p. 2853-2857.

25. Badhiwala, J.H. and J.R. Wilson, *The Natural History of Degenerative Cervical Myelopathy.* Neurosurg Clin N Am, 2018. **29**(1): p. 21-32.

26. Badhiwala, J.H., et al., *Degenerative cervical myelopathy - update and future directions.* Nat Rev Neurol, 2020. **16**(2): p. 108-124.

27. Nouri, A., et al., *Degenerative Cervical Myelopathy: Epidemiology, Genetics, and Pathogenesis.* Spine (Phila Pa 1976), 2015. **40**(12): p. E675-93.

28. Benoist, M., *Natural history of the aging spine.* Eur Spine J, 2003. **12 Suppl 2**(Suppl 2): p. S86-9.

29. Boden, S.D., et al., *Abnormal magnetic-resonance scans of the cervical spine in asymptomatic subjects. A prospective investigation.* J Bone Joint Surg Am, 1990. **72**(8): p. 1178-84.

30. Kelly, J.C., et al., *The natural history and clinical syndromes of degenerative cervical spondylosis.* Adv Orthop, 2012. **2012**: p. 393642.

31. Lv, Y., et al., *The prevalence and associated factors of symptomatic cervical Spondylosis in Chinese adults: a community-based cross-sectional study.* BMC Musculoskeletal Disorders, 2018. **19**(1): p. 325.

32. Williams, F.M. and P.N. Sambrook, *Neck and back pain and intervertebral disc degeneration: role of occupational factors.* Best Pract Res Clin Rheumatol, 2011. **25**(1): p. 69-79.

33. Hurwitz, E.L., et al., *The Global Spine Care Initiative: a summary of the global burden of low back and neck pain studies.* Eur Spine J, 2018. **27**(Suppl 6): p. 796-801.

34. Mokdad, A.H., et al., *Global burden of diseases, injuries, and risk factors for young people's health during 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013.* Lancet, 2016. **387**(10036): p. 2383-401.

35. Radhakrishnan, K., et al., *Epidemiology of cervical radiculopathy. A population-based study from Rochester, Minnesota, 1976 through 1990.* Brain, 1994. **117 (Pt 2)**: p. 325-35.

36. Mansfield, M., et al., *Cervical spine radiculopathy epidemiology: A systematic review.* Musculoskeletal Care, 2020. **18**(4): p. 555-567.

37. Boogaarts, H.D. and R.H. Bartels, *Prevalence of cervical spondylotic myelopathy.* Eur Spine J, 2015. **24 Suppl 2**: p. 139-41.

38. Wu, J.C., et al., *Epidemiology of cervical spondylotic myelopathy and its risk of causing spinal cord injury: a national cohort study.* Neurosurg Focus, 2013. **35**(1): p. E10.

39. Lad, S.P., et al., *National trends in spinal fusion for cervical spondylotic myelopathy.* Surg Neurol, 2009. **71**(1): p. 66-9; discussion 69.

40. Kristiansen, J.A., et al., *The use of surgery for cervical degenerative disease in Norway in the period 2008-2014 : A population-based study of 6511 procedures.* Acta Neurochir (Wien), 2016. **158**(5): p. 969-74.

41. Ingebrigtsen, T., et al., *Cervical spine surgery in the Northern Norway Regional Health Authority area in 2014-18.* Tidsskr Nor Laegeforen, 2022. **142**(4).

42. Fehlings, M.G., et al., *Future Advances in Spine Surgery: The AOSpine North America Perspective.* Neurosurgery, 2017. **80**(3s): p. S1-s8.

43. Mao, J.Z., et al., *Technologic Evolution of Navigation and Robotics in Spine Surgery: A Historical Perspective.* World Neurosurgery, 2021. **145**: p. 159-167.

44. Rhee, J.M., T. Yoon, and K.D. Riew, *Cervical radiculopathy.* J Am Acad Orthop Surg, 2007. **15**(8): p. 486-94.

45. Wong, J.J., et al., *The course and prognostic factors of symptomatic cervical disc herniation with radiculopathy: a systematic review of the literature.* Spine J, 2014. **14**(8): p. 1781-9.

46. Wolff, M.W. and L.A. Levine, *Cervical radiculopathies: conservative approaches to management.* Phys Med Rehabil Clin N Am, 2002. **13**(3): p. 589-608, vii.

47. Bono, C.M., et al., *An evidence-based clinical guideline for the diagnosis and treatment of cervical radiculopathy from degenerative disorders.* The Spine Journal, 2011. **11**(1): p. 64-72.

48. Liang, L., et al., *The effect of exercise on cervical radiculopathy: A systematic review and meta-analysis.* Medicine (Baltimore), 2019. **98**(45): p. e17733.

49. Lantz, J.M., et al., *The Impact of Physical Therapy Following Cervical Spine Surgery for Degenerative Spine Disorders: A Systematic Review.* Clin Spine Surg, 2021. **34**(8): p. 291-307.

50. Miller, J., et al., *Manual therapy and exercise for neck pain: a systematic review.* Man Ther, 2010. **15**(4): p. 334-54.

51. Young, I.A., et al., *Manual therapy, exercise, and traction for patients with cervical radiculopathy: a randomized clinical trial.* Phys Ther, 2009. **89**(7): p. 632-42.

52. Cicala, R.S., K. Thoni, and J.J. Angel, *Long-term results of cervical epidural steroid injections.* Clin J Pain, 1989. **5**(2): p. 143-5.

53. Abdi, S., et al., *Epidural steroids in the management of chronic spinal pain: a systematic review.* Pain Physician, 2007. **10**(1): p. 185-212.

54. Dreyer, S.J. and S.D. Boden, *Nonoperative treatment of neck and arm pain.* Spine (Phila Pa 1976), 1998. **23**(24): p. 2746-54.

55. Tiso, R.L., et al., *Adverse central nervous system sequelae after selective transforaminal block: the role of corticosteroids.* The Spine Journal, 2004. **4**(4): p. 468-474.

56.   Rosenkranz, M., et al., *Anterior spinal artery syndrome following periradicular cervical nerve root therapy.* Journal of neurology, 2004. **251**: p. 229-231.

57.   Rhee, J.M., et al., *Nonoperative management of cervical myelopathy: a systematic review.* Spine (Phila Pa 1976), 2013. **38**(22 Suppl 1): p. S55-67.

58.   Badran, A., et al., *Is there a role for postoperative physiotherapy in degenerative cervical myelopathy? A systematic review.* Clin Rehabil, 2018. **32**(9): p. 1169-1174.

59.   Shiban, E. and B. Meyer, *Treatment considerations of cervical spondylotic myelopathy.* Neurol Clin Pract, 2014. **4**(4): p. 296-303.

60.   Fehlings, M.G., et al., *Safety and efficacy of riluzole in patients undergoing decompressive surgery for degenerative cervical myelopathy (CSM-Protect): a multicentre, double-blind, placebo-controlled, randomised, phase 3 trial.* Lancet Neurol, 2021. **20**(2): p. 98-106.

61.   Kumarasamy, D., et al., *The Role of Riluzole in Acute Traumatic Cervical Spinal Cord Injury with Incomplete Neurological Deficit: A Prospective, Randomised Controlled Study.* Indian J Orthop, 2022. **56**(12): p. 2160-2168.

62.   Kadanka, Z., et al., *Conservative treatment versus surgery in spondylotic cervical myelopathy: a prospective randomised study.* Eur Spine J, 2000. **9**(6): p. 538-44.

63.   Yoshimatsu, H., et al., *Conservative treatment for cervical spondylotic myelopathy: prediction of treatment effects by multivariate analysis.* The Spine Journal, 2001. **1**(4): p. 269-273.

64.   Gutman, G., D.H. Rosenzweig, and J.D. Golan, *Surgical Treatment of Cervical Radiculopathy: Meta-analysis of Randomized Controlled Trials.* Spine (Phila Pa 1976), 2018. **43**(6): p. E365-e372.

65.   Sherk, H.H. and L. Pratt, *Anterior approaches to the cervical spine.* The Laryngoscope, 1983. **93**(2): p. 168-171.

66.   Mummaneni, P.V., et al., *Clinical and radiographic analysis of cervical disc arthroplasty compared with allograft fusion: a randomized controlled clinical trial.* J Neurosurg Spine, 2007. **6**(3): p. 198-209.

67.   Burkus, J.K., et al., *Long-term clinical and radiographic outcomes of cervical disc replacement with the Prestige disc: results from a prospective randomized controlled clinical trial.* J Neurosurg Spine, 2010. **13**(3): p. 308-18.

68.   Findlay, C., S. Ayis, and A.K. Demetriades, *Total disc replacement versus anterior cervical discectomy and fusion: a systematic review with meta-analysis of data from a total of 3160 patients across 14 randomized controlled trials with both short- and medium- to long-term outcomes.* Bone Joint J, 2018. **100-b**(8): p. 991-1001.

69.   Wang, Q.L., et al., *Long-term Results Comparing Cervical Disc Arthroplasty to Anterior Cervical Discectomy and Fusion: A Systematic Review and Meta-Analysis of Randomized Controlled Trials.* Orthop Surg, 2020. **12**(1): p. 16-30.

70.   Alvin, M.D., et al., *Cervical Degenerative Disease: Systematic Review of Economic Analyses.* Spine, 2014. **39**(22S): p. S53-S64.

71.   Malloy, K.M. and A.S. Hilibrand, *Autograft versus allograft in degenerative cervical disease.* Clin Orthop Relat Res, 2002(394): p. 27-38.

72.   Wen, Z., et al., *Anterior Cervical Corpectomy and Fusion and Anterior Cervical Discectomy and Fusion Using Titanium Mesh Cages for Treatment*

of Degenerative Cervical Pathologies: A Literature Review. Med Sci Monit, 2018. **24**: p. 6398-6404.

73. Weber, M.H., et al., *Graft Subsidence and Revision Rates Following Anterior Cervical Corpectomy: A Clinical Study Comparing Different Interbody Cages.* Clin Spine Surg, 2017. **30**(9): p. E1239-e1245.

74. Kwee, M.M., Y.H. Ho, and W.M. Rozen, *The prone position during surgery and its complications: a systematic review and evidence-based guidelines.* Int Surg, 2015. **100**(2): p. 292-303.

75. Mohamed, E., et al., *Lateral mass fixation in subaxial cervical spine: anatomic review.* Global Spine J, 2012. **2**(1): p. 39-46.

76. Rhee, J.M. and S. Basra, *Posterior surgery for cervical myelopathy: laminectomy, laminectomy with fusion, and laminoplasty.* Asian Spine J, 2008. **2**(2): p. 114-26.

77. Revesz, D.F., A. Charalampidis, and P. Gerdhem, *Effectiveness of laminectomy with fusion and laminectomy alone in degenerative cervical myelopathy.* Eur Spine J, 2022. **31**(5): p. 1300-1308.

78. Bartels, R.H., et al., *Lamifuse: results of a randomized controlled trial comparing laminectomy with and without fusion for cervical spondylotic myelopathy.* J Neurosurg Sci, 2017. **61**(2): p. 134-139.

79. de Dios, E., et al., *Laminectomy alone versus laminectomy with fusion for degenerative cervical myelopathy: a long-term study of a national cohort.* Eur Spine J, 2022. **31**(2): p. 334-345.

80. Zhao, H., et al., *Comparison of Laminoplasty vs. Laminectomy for Cervical Spondylotic Myelopathy: A Systematic Review and Meta-Analysis.* Frontiers in Surgery, 2022. **8**.

81. Manzano, G.R., et al., *A Prospective, Randomized Trial Comparing Expansile Cervical Laminoplasty and Cervical Laminectomy and Fusion for Multilevel Cervical Myelopathy.* Neurosurgery, 2012. **70**: p. 264–277.

82. Weinberg, D.S. and J.M. Rhee, *Cervical laminoplasty: indication, technique, complications.* J Spine Surg, 2020. **6**(1): p. 290-301.

83. Yonenobu, K. and T. Oda, *Posterior approach to the degenerative cervical spine.* Eur Spine J, 2003. **12 Suppl 2**(Suppl 2): p. S195-201.

84. Persson, L.C., C.A. Carlsson, and J.Y. Carlsson, *Long-lasting cervical radicular pain managed with surgery, physiotherapy, or a cervical collar. A prospective, randomized study.* Spine (Phila Pa 1976), 1997. **22**(7): p. 751-8.

85. Engquist, M., et al., *Surgery versus nonsurgical treatment of cervical radiculopathy: a prospective, randomized study comparing surgery plus physiotherapy with physiotherapy alone with a 2-year follow-up.* Spine (Phila Pa 1976), 2013. **38**(20): p. 1715-22.

86. Nygaard Ø, P., et al., *[Neck pain with and without affection of nerve roots].* Tidsskr Nor Laegeforen, 2010. **130**(22): p. 2252-5.

87. Bohlman, H.H., et al., *Robinson anterior cervical discectomy and arthrodesis for cervical radiculopathy. Long-term follow-up of one hundred and twenty-two patients.* J Bone Joint Surg Am, 1993. **75**(9): p. 1298-307.

88. Woertgen, C., et al., *Long term outcome after cervical foraminotomy.* J Clin Neurosci, 2000. **7**(4): p. 312-5.

89. Church, E.W., et al., *Cervical laminoforaminotomy for radiculopathy: Symptomatic and functional outcomes in a large cohort with long-term follow-up.* Surg Neurol Int, 2014. **5**(Suppl 15): p. S536-43.

90.   Liu, W.J., et al., *Comparison of Anterior Cervical Discectomy and Fusion versus Posterior Cervical Foraminotomy in the Treatment of Cervical Radiculopathy: A Systematic Review.* Orthop Surg, 2016. **8**(4): p. 425-431.

91.   Fallah, A., et al., *Anterior cervical discectomy with arthroplasty versus arthrodesis for single-level cervical spondylosis: a systematic review and meta-analysis.* PLoS One, 2012. **7**(8): p. e43407.

92.   Fehlings, M.G., et al., *Efficacy and safety of surgical decompression in patients with cervical spondylotic myelopathy: results of the AOSpine North America prospective multi-center study.* J Bone Joint Surg Am, 2013. **95**(18): p. 1651-8.

93.   Fehlings, M.G., et al., *Anterior versus posterior surgical approaches to treat cervical spondylotic myelopathy: outcomes of the prospective multicenter AOSpine North America CSM study in 264 patients.* Spine (Phila Pa 1976), 2013. **38**(26): p. 2247-52.

94.   Gulati, S., et al., *Surgery for Degenerative Cervical Myelopathy: A Nationwide Registry-Based Observational Study With Patient-Reported Outcomes.* Neurosurgery, 2021. **89**(4): p. 704-711.

95.   Lawrence, B.D., et al., *Anterior versus posterior approach for treatment of cervical spondylotic myelopathy: a systematic review.* Spine (Phila Pa 1976), 2013. **38**(22 Suppl 1): p. S173-82.

96.   Geck, M.J. and F.J. Eismont, *Surgical options for the treatment of cervical spondylotic myelopathy.* Orthop Clin North Am, 2002. **33**(2): p. 329-48.

97.   Herbert, R.D., *Cohort studies of aetiology and prognosis: they're different.* J Physiother, 2014. **60**(4): p. 241-4.

98.   Zabor, E.C., A.M. Kaizer, and B.P. Hobbs, *Randomized Controlled Trials.* Chest, 2020. **158**(1s): p. S79-s87.

99.   Riley, R.D., et al., *Prognosis Research Strategy (PROGRESS) 2: prognostic factor research.* PLoS medicine, 2013. **10**(2): p. e1001380-e1001380.

100.   McLeod, R.S., *Issues in surgical randomized controlled trials.* World J Surg, 1999. **23**(12): p. 1210-4.

101.   Van Spall, H.G., et al., *Eligibility criteria of randomized controlled trials published in high-impact general medical journals: a systematic sampling review.* JAMA, 2007. **297**(11): p. 1233-40.

102.   Wolf, B.R. and J.A. Buckwalter, *Randomized surgical trials and "sham" surgery: relevance to modern orthopaedics and minimally invasive surgery.* Iowa Orthop J, 2006. **26**: p. 107-11.

103.   Nardini, C., *The ethics of clinical trials.* Ecancermedicalscience, 2014. **8**: p. 387.

104.   Keller, R.B. and S.J. Atlas, *Is there a continuing role for prospective observational studies in spine research?* Spine (Phila Pa 1976), 2005. **30**(8): p. 847-9.

105.   Donk, R.D., et al., *What's the best surgical treatment for patients with cervical radiculopathy due to single-level degenerative disease? A randomized controlled trial.* PLoS One, 2017. **12**(8): p. e0183603.

106.   Luo, J., et al., *Comparison of artificial cervical arthroplasty versus anterior cervical discectomy and fusion for one-level cervical degenerative disc disease: a meta-analysis of randomized controlled trials.* Eur J Orthop Surg Traumatol, 2015. **25 Suppl 1**: p. S115-25.

107.   Radcliff, K., D. Coric, and T. Albert, *Five-year clinical results of cervical total disc replacement compared with anterior discectomy and fusion for*

treatment of 2-level symptomatic degenerative disc disease: a prospective, randomized, controlled, multicenter investigational device exemption clinical trial. J Neurosurg Spine, 2016. **25**(2): p. 213-24.

108.	Boselie, T.F., et al., *Arthroplasty versus fusion in single-level cervical degenerative disc disease: a Cochrane review.* Spine (Phila Pa 1976), 2013. **38**(17): p. E1096-107.

109.	*AHRQ Methods for Effective Health Care*, in *Registries for Evaluating Patient Outcomes: A User's Guide*, R.E. Gliklich and N.A. Dreyer, Editors. 2010, Agency for Healthcare Research and Quality (US): Rockville (MD).

110.	Delaunay, C., *Registries in orthopaedics.* Orthopaedics & Traumatology: Surgery & Research, 2015. **101**(1, Supplement): p. S69-S75.

111.	Hannan, E.L., et al., *The New York State cardiac registries: history, contributions, limitations, and lessons for future efforts to assess and publicly report healthcare outcomes.* J Am Coll Cardiol, 2012. **59**(25): p. 2309-16.

112.	Meltzer, S.N. and W.S. Weintraub, *The Role of National Registries in Improving Quality of Care and Outcomes for Cardiovascular Disease.* Methodist Debakey Cardiovasc J, 2020. **16**(3): p. 205-211.

113.	Fritzell, P., B. Strömqvist, and O. Hägg, *A practical approach to spine registers in Europe: the Swedish experience.* Eur Spine J, 2006. **15 Suppl 1**(Suppl 1): p. S57-63.

114.	Strömqvist, B., *Evidence-based lumbar spine surgery. The role of national registration.* Acta Orthop Scand Suppl, 2002. **73**(305): p. 34-9.

115.	Strömqvist, B., et al., *The Swedish Spine Register: development, design and utility.* Eur Spine J, 2009. **18 Suppl 3**(Suppl 3): p. 294-304.

116.	Solberg, T., L. Olsen, and M. Berglund, *Årsrapport for 2018 med plan for forbedringstiltak.* 2019, Norsk Nakke- og Ryggregister: Tromsø, Norway.

117.	Anke, A. and L. Granan, *Årsrapport for 2016 med plan for forbedringstiltak.* 2017, Norsk Nakke- og Ryggregister: Tromsø, Norway.

118.	Wilhelmsen, M. and K. Samuelsen, *Årsrapport 2021 - Resultater og forbedringstiltak.* 2022, Norsk Nakke- og Ryggregister: Tromsø, Norway.

119.	Churruca, K., et al., *Patient-reported outcome measures (PROMs): A review of generic and condition-specific measures and a discussion of trends and issues.* Health Expect, 2021. **24**(4): p. 1015-1024.

120.	Weldring, T. and S.M. Smith, *Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs).* Health Serv Insights, 2013. **6**: p. 61-8.

121.	Black, N., *Patient reported outcome measures could help transform healthcare.* Bmj, 2013. **346**: p. f167.

122.	Finkelstein, J.A. and C.E. Schwartz, *Patient-reported outcomes in spine surgery: past, current, and future directions.* J Neurosurg Spine, 2019. **31**(2): p. 155-164.

123.	McCormick, J.D., B.C. Werner, and A.L. Shimer, *Patient-reported outcome measures in spine surgery.* J Am Acad Orthop Surg, 2013. **21**(2): p. 99-107.

124.	Ghimire, P., et al., *Patient-Reported Outcome Measures in Neurosurgery: A Review of the Current Literature.* Neurosurgery, 2018. **83**: p. 622+.

125.	Mokkink, L.B., et al., *The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content.* BMC Med Res Methodol, 2010. **10**: p. 22.

126. Terwee, C.B., et al., *COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study.* Qual Life Res, 2018. **27**(5): p. 1159-1170.

127. Grimes, D.A. and K.F. Schulz, *Bias and causal associations in observational research.* Lancet, 2002. **359**(9302): p. 248-52.

128. Chung, A.S., et al., *Minimum Clinically Important Difference: Current Trends in the Spine Literature.* Spine (Phila Pa 1976), 2017. **42**(14): p. 1096-1105.

129. de Vet, H.C., et al., *Minimally important change determined by a visual method integrating an anchor-based and a distribution-based approach.* Qual Life Res, 2007. **16**(1): p. 131-42.

130. Froud, R. and G. Abel, *Using ROC curves to choose minimally important change thresholds when sensitivity and specificity are valued equally: the forgotten lesson of pythagoras. theoretical considerations and an example application of change in health status.* PLoS One, 2014. **9**(12): p. e114468.

131. Crosby, R.D., R.L. Kolotkin, and G.R. Williams, *Defining clinically meaningful change in health-related quality of life.* J Clin Epidemiol, 2003. **56**(5): p. 395-407.

132. Austevoll, I.M., et al., *Follow-up score, change score or percentage change score for determining clinical important outcome following surgery? An observational study from the Norwegian registry for Spine surgery evaluating patient reported outcome measures in lumbar spinal stenosis and lumbar degenerative spondylolisthesis.* BMC Musculoskelet Disord, 2019. **20**(1): p. 31.

133. Hermansen, E., et al., *Clinical outcome after surgery for lumbar spinal stenosis in patients with insignificant lower extremity pain. A prospective cohort study from the Norwegian registry for spine surgery.* BMC Musculoskelet Disord, 2019. **20**(1): p. 36.

134. Solberg, T., et al., *Can we define success criteria for lumbar disc surgery? : estimates for a substantial amount of improvement in core outcome measures.* Acta Orthop, 2013. **84**(2): p. 196-201.

135. Solberg, T.K., et al., *Would loss to follow-up bias the outcome evaluation of patients operated for degenerative disorders of the lumbar spine?* Acta Orthop, 2011. **82**(1): p. 56-63.

136. Werner, D.A.T., 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): p. 2650-2659.

137. Muñoz, E., W. Muñoz, 3rd, and L. Wise, *National and surgical health care expenditures, 2005-2025.* Ann Surg, 2010. **251**(2): p. 195-200.

138. Kristiansen, I.S. and K.M. Pedersen, *[Health care systems in the Nordic countries--more similarities than differences?].* Tidsskr Nor Laegeforen, 2000. **120**(17): p. 2023-9.

139. Porter, M.E. and R.S. Kaplan, *How to Pay for Health Care.* Harv Bus Rev, 2016. **94**(7-8): p. 88-98, 100, 134.

140. Porter, M.E. and E.O. Teisberg, *Redefining Health Care: Creating Value-Based Competition on Results.* 2006, Boston: Harvard Business School Press.

141. Kane, R.L., M. Maciejewski, and M. Finch, *The relationship of patient satisfaction with care and clinical outcomes.* Med Care, 1997. **35**(7): p. 714-30.

142. *The NHS Long Term Plan.* 2019, NHS: London, United Kingdom.

143. Jacobs, D., et al., *The Medicare Value-Based Care Strategy: Alignment, Growth, And Equity.* Health Affairs Forefront, 2022.

144. Mou, D., et al., *Impetus of US hospital leaders to invest in patient-reported outcome measures (PROMs): a qualitative study.* BMJ Open, 2022. **12**(7): p. e061761.

145. Joeris, A., et al., *Patient-reported outcome measurements in clinical routine of trauma, spine and craniomaxillofacial surgeons: between expectations and reality: a survey among 1212 surgeons.* BMJ open, 2018. **8**(6): p. e020629-e020629.

146. Mou, D., et al., *The Surgeon's Perceived Value of Patient-reported Outcome Measures (PROMs): An Exploratory Qualitative Study of 5 Different Surgical Subspecialties.* Ann Surg, 2020.

147. Holmes, C. and N. Briffa, *Patient-Reported Outcome Measures (PROMS) in patients undergoing heart valve surgery: why should we measure them and which instruments should we use?* Open Heart, 2016. **3**(1): p. e000315.

148. Meerhoff, G.A., et al., *Exploring the perspective of patients with musculoskeletal health problems in primary care on the use of patient-reported outcome measures to stimulate quality improvement in physiotherapist practice; a qualitative study.* Physiother Theory Pract, 2021. **37**(9): p. 993-1004.

149. Basch, E., et al., *Methods for Developing Patient-Reported Outcome-Based Performance Measures (PRO-PMs).* Value in Health, 2015. **18**(4): p. 493-504.

150. Squitieri, L., K.J. Bozic, and A.L. Pusic, *The Role of Patient-Reported Outcome Measures in Value-Based Payment Reform.* Value Health, 2017. **20**(6): p. 834-836.

151. Mjåset, C., et al., *An introduction to bundled payments.* Tidsskr Nor Laegeforen, 2020. **140**(12).

152. Goude, F., et al., *The effects of competition and bundled payment on patient reported outcome measures after hip replacement surgery.* BMC Health Serv Res, 2021. **21**(1): p. 387.

153. Sedaghat, A.R., *Understanding the Minimal Clinically Important Difference (MCID) of Patient-Reported Outcome Measures.* Otolaryngol Head Neck Surg, 2019. **161**(4): p. 551-560.

154. Mjåset, C., et al., *Value-Based Health Care in Four Different Health Care Systems.* NEJM Catalyst, 2020.

155. Moons, K.G.M., et al., *Prognosis and prognostic research: what, why, and how?* BMJ, 2009. **338**: p. b375.

156. Hemingway, H., et al., *Prognosis research strategy (PROGRESS) 1: A framework for researching clinical outcomes.* BMJ : British Medical Journal, 2013. **346**: p. e5595.

157. Sielatycki, J.A., et al., *Does Obesity Correlate With Worse Patient-Reported Outcomes Following Elective Anterior Cervical Discectomy and Fusion?* Neurosurgery, 2016. **79**(1): p. 69-74.

158. Zhang, G.A., et al., *Impact of Elevated Body Mass Index on Surgical Outcomes for Patients Undergoing Cervical Fusion Procedures: A Systematic Review and Meta-Analysis.* Orthop Surg, 2020. **12**(1): p. 3-15.

159. Narain, A.S., et al., *Impact of body mass index on surgical outcomes, narcotics consumption, and hospital costs following anterior cervical discectomy and fusion.* J Neurosurg Spine, 2018. **28**(2): p. 160-166.

160. Steyerberg, E.W., et al., *Prognosis Research Strategy (PROGRESS) 3: prognostic model research.* PLoS medicine, 2013. **10**(2): p. e1001381-e1001381.

161. Hingorani, A.D., et al., *Prognosis research strategy (PROGRESS) 4: Stratified medicine research.* BMJ : British Medical Journal, 2013. **346**: p. e5793.

162. Peolsson, A. and M. Peolsson, *Predictive factors for long-term outcome of anterior cervical decompression and fusion: a multivariate data analysis.* Eur Spine J, 2008. **17**(3): p. 406-414.

163. Hermansen, A., et al., *Positive predictive factors and subgroup analysis of clinically relevant improvement after anterior cervical decompression and fusion for cervical disc disease: a 10- to 13-year follow-up of a prospective randomized study: clinical article.* J Neurosurg Spine, 2013. **19**(4): p. 403-11.

164. Anderson, P.A., B.R. Subach, and K.D. Riew, *Predictors of outcome after anterior cervical discectomy and fusion: a multivariate analysis.* Spine (Phila Pa 1976), 2009. **34**(2): p. 161-6.

165. Peolsson, A., et al., *Predictive factors for the outcome of anterior cervical decompression and fusion.* Eur Spine J, 2003. **12**(3): p. 274-80.

166. Narain, A.S., et al., *Risk Factors Associated With Failure to Reach Minimal Clinically Important Difference in Patient-Reported Outcomes Following Anterior Cervical Discectomy and Fusion.* Int J Spine Surg, 2019. **13**(3): p. 262-269.

167. Archer, K.R., et al., *Development and Validation of Cervical Prediction Models for Patient-Reported Outcomes at 1 Year After Cervical Spine Surgery for Radiculopathy and Myelopathy.* Spine (Phila Pa 1976), 2020. **45**(22): p. 1541-1552.

168. Scerrati, A., et al., *Factors affecting functional outcome after anterior cervical discectomy and fusion: A multicenter study.* J Craniovertebr Junction Spine, 2021. **12**(2): p. 144-148.

169. Narain, A.S., et al., *Risk Factors for Medical and Surgical Complications after 1-2-Level Anterior Cervical Discectomy and Fusion Procedures.* Int J Spine Surg, 2020. **14**(3): p. 286-293.

170. Tetreault, L.A., et al., *A clinical prediction model to determine outcomes in patients with cervical spondylotic myelopathy undergoing surgical treatment: data from the prospective, multi-center AOSpine North America study.* JBJS, 2013. **95**(18): p. 1659-1666.

171. Tetreault, L., et al., *Predicting the minimum clinically important difference in patients undergoing surgery for the treatment of degenerative cervical myelopathy.* Neurosurg Focus, 2016. **40**(6): p. E14.

172. Tetreault, L.A., et al., *A clinical prediction model to assess surgical outcome in patients with cervical spondylotic myelopathy: internal and external validations using the prospective multicenter AOSpine North American and international datasets of 743 patients.* Spine J, 2015. **15**(3): p. 388-97.

173. Peolsson, A., L. Vavruch, and B. Oberg, *Predictive factors for arm pain, neck pain, neck specific disability and health after anterior cervical decompression and fusion.* Acta Neurochir (Wien), 2006. **148**(2): p. 167-73; discussion 173.

174. Lubelski, D., et al., *Preoperative Nomograms Predict Patient-Specific Cervical Spine Surgery Clinical and Quality of Life Outcomes.* Neurosurgery, 2018. **83**(1): p. 104-113.

175. Peolsson, A. and M. Peolsson, *Predictive factors for long-term outcome of anterior cervical decompression and fusion: a multivariate data analysis.*

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, 2008. **17**(3): p. 406-414.

176. Wilhelmsen, M., K. Samuelsen, and K. Thon, *Årsrapport for 2020 med plan for forbedringstiltak.* 2021, Norsk Nakke- og Ryggregister: Tromsø, Norway.

177. Vandenbroucke, J.P., et al., *Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration.* PLoS Med, 2007. **4**(10): p. e297.

178. Collins, G.S., et al., *Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD).* Ann Intern Med, 2015. **162**(10): p. 735-6.

179. Jaeschke, R., J. Singer, and G.H. Guyatt, *Measurement of health status. Ascertaining the minimal clinically important difference.* Control Clin Trials, 1989. **10**(4): p. 407-15.

180. Oni, P., et al., *Radiological and Clinical Outcome after Multilevel Anterior Cervical Discectomy and/or Corpectomy and Fixation.* J Clin Med, 2018. **7**(12).

181. Shamji, M.F., et al., *Comparison of Anterior Surgical Options for the Treatment of Multilevel Cervical Spondylotic Myelopathy: A Systematic Review.* Spine, 2013. **38**(22S): p. S195-S209.

182. Traynelis, V.C., et al., *Alternative procedures for the treatment of cervical spondylotic myelopathy: arthroplasty, oblique corpectomy, skip laminectomy: evaluation of comparative effectiveness and safety.* Spine (Phila Pa 1976), 2013. **38**(22 Suppl 1): p. S210-31.

183. Chau, A.M. and R.J. Mobbs, *Bone graft substitutes in anterior cervical discectomy and fusion.* Eur Spine J, 2009. **18**(4): p. 449-64.

184. Rao, R.D., K. Gourab, and K.S. David, *Operative treatment of cervical spondylotic myelopathy.* J Bone Joint Surg Am, 2006. **88**(7): p. 1619-40.

185. Gornet, M.F., et al., *Multilevel Cervical Disc Arthroplasty: Long-Term Outcomes at 3 and 4 Levels.* Int J Spine Surg, 2020. **14**(s2): p. S41-s49.

186. Joaquim, A.F. and K.D. Riew, *Multilevel cervical arthroplasty: current evidence. A systematic review.* Neurosurg Focus, 2017. **42**(2): p. E4.

187. MacDermid, J.C., et al., *Measurement properties of the neck disability index: a systematic review.* J Orthop Sports Phys Ther, 2009. **39**(5): p. 400-17.

188. Vernon, H., *The Neck Disability Index: state-of-the-art, 1991-2008.* J Manipulative Physiol Ther, 2008. **31**(7): p. 491-502.

189. Jensen, M.P. and P. Karoly, *Self-report scales and procedures for assessing pain in adults*, in *Handbook of pain assessment, 3rd ed.* 2011, The Guilford Press: New York, NY, US. p. 19-44.

190. Dolan, P., et al., *The time trade-off method: results from a general population study.* Health Econ, 1996. **5**(2): p. 141-54.

191. Herdmann, J., et al. *The European Myelopathy Score.* in *Cerebellar Infarct. Midline Tumors. Minimally Invasive Endoscopic Neurosurgery (MIEN).* 1994. Berlin, Heidelberg: Springer Berlin Heidelberg.

192. Kamper, S.J., et al., *Global Perceived Effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status.* J Clin Epidemiol, 2010. **63**(7): p. 760-766 e1.

193. Yerneni, K., et al., *Preoperative Opioid Use and Clinical Outcomes in Spine Surgery: A Systematic Review.* Neurosurgery, 2020. **86**(6): p. E490-e507.

194.   Faour, M., et al., *Preoperative Opioid Use: A Risk Factor for Poor Return to Work Status After Single-level Cervical Fusion for Radiculopathy in a Workers' Compensation Setting.* Clin Spine Surg, 2018. **31**(1): p. E19-e24.

195.   Tarazona, D., et al., *Longer Preoperative Duration of Symptoms Negatively Affects Health-related Quality of Life After Surgery for Cervical Radiculopathy.* Spine (Phila Pa 1976), 2019. **44**(10): p. 685-690.

196.   Passias, P.G., et al., *Arm Pain Versus Neck Pain: A Novel Ratio as a Predictor of Post-Operative Clinical Outcomes in Cervical Radiculopathy Patients.* Int J Spine Surg, 2018. **12**(5): p. 629-637.

197.   Khan, O., et al., *Machine learning algorithms for prediction of health-related quality-of-life after surgery for mild degenerative cervical myelopathy.* The Spine Journal, 2021. **21**(10): p. 1659-1669.

198.   Tetreault, L., et al., *Significant Predictors of Outcome Following Surgery for the Treatment of Degenerative Cervical Myelopathy: A Systematic Review of the Literature.* Neurosurg Clin N Am, 2018. **29**(1): p. 115-127.e35.

199.   Merali, Z.G., et al., *Using a machine learning approach to predict outcome after surgery for degenerative cervical myelopathy.* PloS one, 2019. **14**(4): p. e0215133-e0215133.

200.   Müller, D., et al., *Development of a machine-learning based model for predicting multidimensional outcome after surgery for degenerative disorders of the spine.* Eur Spine J, 2022. **31**(8): p. 2125-2136.

201.   Cagnie, B., et al., *Individual and work related risk factors for neck pain among office workers: a cross sectional study.* Eur Spine J, 2007. **16**(5): p. 679-86.

202.   Macedo, L.G. and M.C. Battié, *The association between occupational loading and spine degeneration on imaging - a systematic review and meta-analysis.* BMC Musculoskelet Disord, 2019. **20**(1): p. 489.

203.   Palmer, K.T. and J. Smedley, *Work relatedness of chronic neck pain with physical findings--a systematic review.* Scand J Work Environ Health, 2007. **33**(3): p. 165-91.

204.   Kim, E.J., et al., *Factors Associated With Return-to-Work Following Cervical Spine Surgery in Non-Worker's Compensation Setting.* Spine (Phila Pa 1976), 2019. **44**(13): p. 903-907.

205.   Radcliff, K., et al., *Two-Year Results of the Prospective Spine Treatment Outcomes Study: Analysis of Postoperative Clinical Outcomes Between Patients with and without a History of Previous Cervical Spine Surgery.* World Neurosurg, 2018. **109**: p. e144-e149.

206.   Copay, A.G., et al., *Understanding the minimum clinically important difference: a review of concepts and methods.* Spine J, 2007. **7**(5): p. 541-6.

207.   Donk, R.D., et al., *Substantial clinical benefit for Neck Disability Index revisited: establishing the goal for treatment?* Eur Spine J, 2018. **27**(6): p. 1262-1265.

208.   Zweig, M.H. and G. Campbell, *Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine.* Clin Chem, 1993. **39**(4): p. 561-77.

209.   Unal, I., *Defining an Optimal Cut-Point Value in ROC Analysis: An Alternative Approach.* Comput Math Methods Med, 2017. **2017**: p. 3762651.

210.   Mandrekar, J.N., *Receiver Operating Characteristic Curve in Diagnostic Test Assessment.* Journal of Thoracic Oncology, 2010. **5**(9): p. 1315-1316.

211. Schober, P., C. Boer, and L.A. Schwarte, *Correlation Coefficients: Appropriate Use and Interpretation.* Anesth Analg, 2018. **126**(5): p. 1763-1768.

212. Steyerberg, E.W., et al., *Assessing the performance of prediction models: a framework for traditional and novel measures.* Epidemiology, 2010. **21**(1): p. 128-38.

213. Riley, R.D., et al., *Prognosis Research Strategy (PROGRESS) 2: prognostic factor research.* PLoS Med, 2013. **10**(2): p. e1001380.

214. Moons, K.G., et al., *Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration.* Ann Intern Med, 2015. **162**(1): p. W1-73.

215. Moons, K.G.M., et al., *PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration.* Ann Intern Med, 2019. **170**(1): p. W1-w33.

216. Taipale, H. and J. Tiihonen, *Registry-based studies: What they can tell us, and what they cannot.* Eur Neuropsychopharmacol, 2021. **45**: p. 35-37.

217. Deyo, R.A., et al., *Analysis of Automated Administrative and Survey Databases to Study Patterns and Outcomes of Care.* Spine, 1994. **19**(18): p. 2083S-2091S.

218. McLeod, R., *Randomized, controlled trials: is there a role for them in surgery?* Ann Surg, 2006. **244**(5): p. 684-5.

219. Thygesen, L.C. and A.K. Ersbøll, *When the entire population is the sample: strengths and limitations in register-based epidemiology.* Eur J Epidemiol, 2014. **29**(8): p. 551-8.

220. *"Trust in government, policy effectiveness and the governance agenda", in Government at a Glance 2013.* 2013, OECD: Paris, France.

221. Bratlid, D., *Is the Norwegian System of Patient Injury Compensation in the best interests of patients?* Tidsskr Nor Laegeforen, 2017. **137**(23-24).

222. Jørstad, R.G., *The Norwegian system of compensation to patients.* Med Law, 2002. **21**(4): p. 681-6.

223. *Antall godkjente spesialister per 14.04.2021 samt historikk 2008-2021 med fordeling på kjønn.* 2022 Jan 24, 2022 [cited 2023 Feb 27].

224. *Legestatistikk.* 2023 [cited 2023 May 23, 2023]; Available from: https://www.legeforeningen.no/om-oss/legestatistikk/#53298.

225. Kristman, V., M. Manno, and P. Côté, *Loss to follow-up in cohort studies: how much is too much?* Eur J Epidemiol, 2004. **19**(8): p. 751-60.

226. Nunan, D., J. Aronson, and C. Bankhead, *Catalogue of bias: attrition bias.* BMJ Evid Based Med, 2018. **23**(1): p. 21-22.

227. Schulz, K.F. and D.A. Grimes, *Sample size slippages in randomised trials: exclusions and the lost and wayward.* Lancet, 2002. **359**(9308): p. 781-5.

228. van Hooff, M.L., et al., *Evidence and practice in spine registries.* Acta Orthop, 2015. **86**(5): p. 534-44.

229. Hojmark, K., et al., *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): p. 282-6.

230. Schafer, J.L., *Multiple imputation: a primer.* Statistical Methods in Medical Research, 1999. **8**(1): p. 3-15.

231. Choi, B.-W., et al., *Cervical radiculopathy combined with cervical myelopathy: prevalence and characteristics.* European Journal of Orthopaedic Surgery & Traumatology, 2017. **27**(7): p. 889-893.

232. Modic, M.T., J.S. Ross, and T.J. Masaryk, *Imaging of degenerative disease of the cervical spine.* Clin Orthop Relat Res, 1989(239): p. 109-20.

233. Wilson, D.W., R.T. Pezzuti, and J.N. Place, *Magnetic resonance imaging in the preoperative evaluation of cervical radiculopathy.* Neurosurgery, 1991. **28**(2): p. 175-9.

234. Sundseth, J., et al., *The Norwegian Cervical Arthroplasty Trial (NORCAT): 2-year clinical outcome after single-level cervical arthroplasty versus fusion-a prospective, single-blinded, randomized, controlled multicenter study.* Eur Spine J, 2017. **26**(4): p. 1225-1235.

235. Kavanagh, R., et al., *Operative techniques for cervical radiculopathy and myelopathy.* Advances in orthopedics, 2012. **2012**.

236. Rothman, K.J., S. Greenland, and T.L. Lash, *Modern epidemiology.* 2008, Wolters Kluwer Health/Lippincott Williams & Wilkins: Philadelphia.

237. Ohrn, A., et al., *Adverse events in spine surgery in Sweden: a comparison of patient claims data and national quality register (Swespine) data.* Acta Orthop, 2011. **82**(6): p. 727-31.

238. Alhaug, O.K., et al., *Accuracy and agreement of national spine register data for 474 patients compared to corresponding electronic patient records.* Eur Spine J, 2022. **31**(3): p. 801-811.

239. Loo, R.L., et al., *A comparison of self-reported analgesic use and detection of urinary ibuprofen and acetaminophen metabolites by means of metabonomics: the INTERMAP Study.* Am J Epidemiol, 2012. **175**(4): p. 348-58.

240. Alberg, A.J., et al., *The Validity of Self-reported Recent Smoking in Head and Neck Cancer Surgical Patients.* Otolaryngol Head Neck Surg, 2015. **153**(6): p. 990-995.

241. Wolvers, P.J.D., et al., *Self-Reported Smoking Compared to Serum Cotinine in Bariatric Surgery Patients: Smoking Is Underreported Before the Operation.* Obes Surg, 2020. **30**(1): p. 23-37.

242. Turk, D.C. and R. Melzack, *Handbook of pain assessment.* 3rd ed. 2011, New York: Guilford Press. xvii, 542 p.

243. Parker, S.L., et al., *Assessment of the minimum clinically important difference in pain, disability, and quality of life after anterior cervical discectomy and fusion: clinical article.* J Neurosurg Spine, 2013. **18**(2): p. 154-60.

244. Auffinger, B.M., et al., *Measuring surgical outcomes in cervical spondylotic myelopathy patients undergoing anterior cervical discectomy and fusion: assessment of minimum clinically important difference.* PLoS One, 2013. **8**(6): p. e67408.

245. Carreon, L.Y., et al., *Neck Disability Index, short form-36 physical component summary, and pain scales for neck and arm pain: the minimum clinically important difference and substantial clinical benefit after cervical spine fusion.* Spine J, 2010. **10**(6): p. 469-74.

246. Kovacs, F.M., et al., *Minimum detectable and minimal clinically important changes for pain in patients with nonspecific neck pain.* BMC Musculoskelet Disord, 2008. **9**: p. 43.

247. Ostelo, R.W., et al., *Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change.* Spine (Phila Pa 1976), 2008. **33**(1): p. 90-4.

248. Vickers, A.J., *The use of percentage change from baseline as an outcome in a controlled trial is statistically inefficient: a simulation study.* BMC Medical Research Methodology, 2001. **1**(1): p. 6.

249. Taso, M., 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, 2022. **31**(10): p. 2761-2768.

250. McHorney, C.A. and A.R. Tarlov, *Individual-patient monitoring in clinical practice: are available health status surveys adequate?* Qual Life Res, 1995. **4**(4): p. 293-307.

251. Garratt, A.M., et al., *Evaluation of the EQ-5D-3L and 5L versions in low back pain patients.* Health and Quality of Life Outcomes, 2021. **19**(1): p. 155.

252. Kamper, S.J., C.G. Maher, and G. Mackay, *Global rating of change scales: a review of strengths and weaknesses and considerations for design.* J Man Manip Ther, 2009. **17**(3): p. 163-70.

253. Bobos, P., et al., *Psychometric properties of the global rating of change scales in patients with low back pain, upper and lower extremity disorders. A systematic review with meta-analysis.* J Orthop, 2020. **21**: p. 40-48.

254. Naggara, O., et al., *Analysis by categorizing or dichotomizing continuous variables is inadvisable: an example from the natural history of unruptured aneurysms.* AJNR Am J Neuroradiol, 2011. **32**(3): p. 437-40.

255. Fedorov, V., F. Mannino, and R. Zhang, *Consequences of dichotomization.* Pharm Stat, 2009. **8**(1): p. 50-61.

256. Royston, P., D.G. Altman, and W. Sauerbrei, *Dichotomizing continuous predictors in multiple regression: a bad idea.* Stat Med, 2006. **25**(1): p. 127-41.

257. Norton, E.C., B.E. Dowd, and M.L. Maciejewski, *Odds Ratios-Current Best Practice and Use.* Jama, 2018. **320**(1): p. 84-85.

258. Altman, D.G., et al., *Prognosis and prognostic research: validating a prognostic model.* BMJ, 2009. **338**: p. b605.

259. Akobeng, A.K., *Assessing the validity of clinical trials.* J Pediatr Gastroenterol Nutr, 2008. **47**(3): p. 277-82.

260. Steyerberg, E.W., et al., *Validation and updating of predictive logistic regression models: a study on sample size and shrinkage.* Stat Med, 2004. **23**(16): p. 2567-86.

261. Moons, K.G., et al., *Prognosis and prognostic research: application and impact of prognostic models in clinical practice.* Bmj, 2009. **338**: p. b606.

262. Vernon, H. and S. Mior, *The Neck Disability Index: a study of reliability and validity.* J Manipulative Physiol Ther, 1991. **14**(7): p. 409-15.

263. Chan, A.K., et al., *Cervical spondylotic myelopathy with severe axial neck pain: is anterior or posterior approach better?* J Neurosurg Spine, 2023. **38**(1): p. 42-55.

264. Badhiwala, J.H., et al., *Minimum Clinically Important Difference in SF-36 Scores for Use in Degenerative Cervical Myelopathy.* Spine (Phila Pa 1976), 2018. **43**(21): p. E1260-e1266.

265. Chien, A., et al., *Responsiveness of the Chinese versions of the Japanese Orthopaedic Association Cervical Myelopathy Evaluation Questionnaire and Neck Disability Index in postoperative patients with cervical spondylotic myelopathy.* Spine (Phila Pa 1976), 2015. **40**(17): p. 1315-21.

266. Kato, S., et al., *Minimum clinically important difference in outcome scores among patients undergoing cervical laminoplasty.* Eur Spine J, 2019. **28**(5): p. 1234-1241.

267. Kato, S., et al., *Minimum Clinically Important Difference and Patient Acceptable Symptom State of Japanese Orthopaedic Association Score in*

Degenerative Cervical Myelopathy Patients. Spine (Phila Pa 1976), 2019. **44**(10): p. 691-697.

268. Tetreault, L., et al., *The Minimum Clinically Important Difference of the Modified Japanese Orthopaedic Association Scale in Patients with Degenerative Cervical Myelopathy.* Spine (Phila Pa 1976), 2015. **40**(21): p. 1653-9.

269. Hung, M., et al., *What Are the MCIDs for PROMIS, NDI, and ODI Instruments Among Patients With Spinal Conditions?* Clin Orthop Relat Res, 2018. **476**(10): p. 2027-2036.

270. Steinhaus, M.E., et al., *Minimal Clinically Important Difference and Substantial Clinical Benefit Using PROMIS CAT in Cervical Spine Surgery.* Clin Spine Surg, 2019. **32**(9): p. 392-397.

271. Joelson, A., et al., *The minimal important change for the EQ VAS based on the SF-36 health transition item: observations from 25772 spine surgery procedures.* Quality of Life Research, 2022. **31**(12): p. 3459-3466.

272. Petrou, P., G. Samoutis, and C. Lionis, *Single-payer or a multipayer health system: a systematic literature review.* Public Health, 2018. **163**: p. 141-152.

273. Papanicolas, I., L.R. Woskie, and A.K. Jha, *Health Care Spending in the United States and Other High-Income Countries.* JAMA, 2018. **319**(10): p. 1024-1039.

274. Cherkin, D.C., et al., *An international comparison of back surgery rates.* Spine (Phila Pa 1976), 1994. **19**(11): p. 1201-6.

275. Steiger, H.J., M. Krämer, and H.J. Reulen, *Development of neurosurgery in Germany: comparison of data collected by polls for 1997, 2003, and 2008 among providers of neurosurgical care.* World Neurosurg, 2012. **77**(1): p. 18-27.

276. Jansson, K.A., et al., *Surgery for herniation of a lumbar disc in Sweden between 1987 and 1999. An analysis of 27,576 operations.* J Bone Joint Surg Br, 2004. **86**(6): p. 841-7.

277. Martin, B.I., et al., *Trends in Lumbar Fusion Procedure Rates and Associated Hospital Costs for Degenerative Spinal Diseases in the United States, 2004 to 2015.* Spine (Phila Pa 1976), 2019. **44**(5): p. 369-376.

278. Sivasubramaniam, V., et al., *Trends in hospital admissions and surgical procedures for degenerative lumbar spine disease in England: a 15-year time-series study.* BMJ Open, 2015. **5**(12): p. e009011.

279. Wang, M.C., et al., *Trends and variations in cervical spine surgery in the United States: Medicare beneficiaries, 1992 to 2005.* Spine (Phila Pa 1976), 2009. **34**(9): p. 955-61; discussion 962-3.

280. Grøvle, L., et al., *The Rates of LSS Surgery in Norwegian Public Hospitals: A Threefold Increase From 1999 to 2013.* Spine (Phila Pa 1976), 2019. **44**(6): p. E372-e378.

281. Weinstein, J.N., et al., *United States' trends and regional variations in lumbar spine surgery: 1992-2003.* Spine (Phila Pa 1976), 2006. **31**(23): p. 2707-14.

282. Einstadter, D., et al., *Variation in the Rate of Cervical Spine Surgery in Washington State.* Medical Care, 1993. **31**(8): p. 711-718.

283. Lagerbäck, T., et al., *Effectiveness of surgery for sciatica with disc herniation is not substantially affected by differences in surgical incidences among three countries: results from the Danish, Swedish and Norwegian spine registries.* Eur Spine J, 2019. **28**(11): p. 2562-2571.

284.	Bovim, G., et al., *NOU 2023:4. Tid for handling. Personellet i en bærekraftig helse- og omsorgstjeneste*. Helse- og omsorgsdepartementet: Oslo.

285.	*About Centre on patient-reported outcomes data*. 2023 [cited 2023 May 3, 2023]; Available from: https://www.kvalitetsregistre.no/artikkel/information-english.

286.	Hara, S., et al., *Return to Work After Surgery for Cervical Radiculopathy: A Nationwide Registry-based Observational Study*. Spine (Phila Pa 1976), 2023. **48**(4): p. 253-260.

Appendices A-D and Papers I-III

Appendix A.
Patient declaration of consent form, NORspine

NORSK NAKKE- OG RYGGREGISTER

E-post: nakkerygg@unn.no
Hjemmeside: www.nakkeryggreg.no

Versjon 2.0

Samtykkeerklæring

Til deg som skal vurderes for nakke- og ryggplager

Norsk Nakke og Ryggregister er et nasjonalt medisinsk kvalitetsregister. Hensikten med registeret er å forbedre kvaliteten på tilbudet til personer med nakke og ryggplager på sykehusene i Norge. Universitetssykehuset Nord-Norge HF (UNN) er dataansvarlig. Rettslig grunnlag for registeret er personvernforordningen artikkel 6 nr. 1 bokstav e (allmenn interesse) og forskrift om medisinske kvalitetsregistre.

Hva skal registreres?
Ditt personnummer og navn, opplysninger om diagnose, samt opplysninger som beskriver plagene dine, grad av funksjonsnedsettelse og yrkesstatus. I tillegg registreres vanlige journalopplysninger som sykehistorie, radiologiske funn og opplysninger om hvilken type behandling du får.

Hvordan samles opplysningene inn?
Opplysninger fra deg samles inn til registeret i forkant av første konsultasjon ved poliklinikken. Ved den første konsultasjonen vil også relevante opplysninger fra helsepersonell som undersøker deg registreres. Norsk Nakke- og Ryggregister vil i tillegg sende deg et spørreskjema 6 og 12 måneder etter undersøkelsen på poliklinikk.

Hvem kan få tilgang til opplysningene?
Opplysninger som samles inn på spørreskjemaene gjøres tilgjengelig for den sykehusavdelingen eller institusjon som undersøker deg, og det er kun de som får tilgang til dine direkte personidentifiserbare opplysninger. I registeret behandles opplysningene videre uten navn og fødselsnummer eller andre direkte personidentifiserende kjennetegn. Opplysninger fra registeret kan utleveres etter søknad til bruk for kvalitetsforbedring og forskning, gitt at det er innenfor registerets formål og gjeldende lovkrav.

Kvalitetsforbedring og forskning
Når du samtykker til registeret kan opplysninger brukes til kvalitetsforbedring i sykehusene lokalt og nasjonalt, samt i forskningsprosjekter. Forskere vil kunne bruke registeret til å evaluere hva som har betydning for gode eller dårlige resultat for pasienter med nakke- og ryggproblemer. For spesielle forskningsprosjekter kan det være aktuelt å sammenstille informasjon fra registeret med andre offentlige registre (se vedlagte liste). Dersom du samtykker, samtykker du også til at du kan kontaktes på nytt utenom 6 og 12 måneders oppfølgings-skjemaet.

Dine rettigheter
Spørreskjemaene utfylles vanligvis elektronisk, men kan også fylles ut på papir. Data vil lagres elektronisk. Opplysninger i databasen lagres på en trygg måte som ivaretar personvernet. De vil bli lagret uten tidsbegrensning.

Å bidra med opplysninger til registeret er frivillig. Du kan velge å ikke samtykke. Du kan når som helst, og uten å oppgi noen grunn, trekke ditt samtykke. Det vil ikke ha noen negative konsekvenser for deg eller din behandling dersom du ikke vil delta eller senere velger å trekke deg. Du har rett til å få vite hva som står om deg i registeret, og hvem som har hatt tilgang til eller fått utlevert opplysninger som er knyttet til deg. Du kan kreve at opplysninger blir rettet eller slettet fra registeret. Å trekke samtykke tilbake vil gjelde for fremtidige behandlinger av opplysninger, og gjelder ikke for allerede utførte analyser eller anonymt materiale. Dersom du senere ønsker å trekke deg eller har spørsmål, kan du kontakte registeret på mail nakkerygg@unn.no.

Dersom du mener at informasjon om deg ikke blir brukt i samsvar med relevant regelverk kan du henvende deg til Datatilsynet eller Statens Helsetilsyn. Spørsmål vedrørende personvern kan rettes til Personvernombudet@unn.no.

Du finner videre informasjon om registeret her: www.nakkeryggreg.no

Med vennlig hilsen
Norsk nakke- og ryggregister (NNRR)

Maja Wilhelmsen
Registerleder

Koblinger til andre registre
Informasjon fra Norsk Nakke- og Ryggregister kan i tråd med helseregisterloven sammenstilles med andre offentlige registre og befolkningsundersøkelser. Eksempler på registre som særlig kan være aktuelle:
Nasjonalt Kvalitetsregister for Ryggkirurgi
NAV
Dødsårsaksregisteret
Medisinsk Fødselsregister
Norsk Pasientregister
Kreftregisteret
Reseptregisteret
Kommunalt pasient- og brukerregister
Kontroll og utbetaling av helserefusjoner (KUHR)
Statistisk sentralbyrå
Nasjonalt register for leddproteser
Befolkningsundersøkelsene som inngår i Conor (Cohort of Norway)
Befolkningsundersøkelsene som inngikk i Statens Helseundersøkelser (SHuS)
Helseundersøkelsen i Trøndelag (HUNT)
Tromsøundersøkelsen
Skattedirektoratets databaser

_____

Jeg har lest informasjonen ovenfor og samtykker i at de nevnte opplysningene registreres og gjøres tilgjengelig for   kvalitetssikring og forskning.

Sted: ....   Dato:…………………………………….

Underskrift: ....

Appendix B.
Preoperative patient questionnaire, NORspine

Spørreskjema for pasienter som skal opereres for degenerative tilstander i nakken

Nasjonalt Kvalitetsregister for Ryggkirurgi
Degenerativ nakke

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

0510 – Versjon 2

Formålet med dette spørreskjemaet er å gi leger, sykepleiere og fysioterapeuter bedre forståelse av hvilke plager pasienter med degenerative tilstander i nakke har og hvor effektiv og trygg behandlingen er.
Slik kunnskap kan brukes til å gi nakkepasienter et bedre behandlingstilbud i fremtiden

Pasientdata (Barkode)

Navn

Fødselsnr. (11 siffer)

Adresse

E-post
(For bruk ved etterkontroll)

Mobil
(For bruk ved etterkontroll)

Dato for utfylling   Dag   Måned   År

Familie og barn

1. Sivilstatus (sett kun ett kryss)

☐ Gift   ☐ Samboende   ☐ Enslig

2. Hvor mange barn har du?

Morsmål

☐ Norsk   ☐ Samisk   ☐ Annet, angi hvilket

Røyker du?   ☐ Ja   ☐ Nei
Bruker du snus?   ☐ Ja   ☐ Nei

Høyde og vekt

Høyde ⬚,⬚⬚ (m)   Vekt ⬚⬚⬚ (kg)

Utdanning

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)

Yrke (Sett kun ett kryss)

☐ Har en jobb der jeg arbeider mye med armene over skuldernivå
☐ Har en jobb der jeg i stor grad bruker datamaskin
☐ Har en jobb som medfører tungt fysisk arbeid
☐ Har en jobb som medfører lett fysisk arbeid med variert arbeidsstilling
☐ Har en stillesittende jobb

Arbeidsstatus (Sett kun ett kryss)

☐ I arbeid   ☐ Sykemeldt
☐ Hjemmeværende   ☐ Aktivt sykemeldt
☐ Student/skoleelev   ☐ Arbeidsavklaringspenger
☐ Alderspensjonist   ☐ Uførepensjon
☐ Arbeidsledig   ☐ Uførepensjon + sykemeldt

Hvis du er delvis sykemeldt eller ikke har full uførepensjon, angi prosent

_____ % Sykemeldt   _____% Ufør

Har du søkt om uførepensjon pga din sykdom/tilstand som du opereres for nå? (Sett kun ett kryss)

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

Har du søkt om erstatning fra forsikringsselskap eller folketrygden (eventuelt yrkesskadeerstatning) pga din tilstand/sykdom som du opereres for nå? (Sett kun ett kryss)

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

Symptomvarighet

Varighet av nåværende hode-/nakkesmerter (Sett kun ett kryss):

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

Varighet av nåværende utstrålende armsmerter :

- [ ] 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)

Hvor sterke smerter har du hatt siste uke?

Hvordan vil du gradere smertene du har hatt i **hodet** 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 **nakken** 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 **armen** (en 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

Hvor stråler smertene? (Sett kun ett kryss)

- [ ] I begge skuldre/armer
- [ ] Kun i en skulder/arm
- [ ] Ingen strålesmerter

Hvor langt ut går armsmertene dine? (Sett kun ett kryss)

- [ ] Til skulder
- [ ] Til overarm/albue
- [ ] Til underarm/håndledd
- [ ] Til finger/fingre
- [ ] Ingen arm/skuldersmerte

Har du vært undersøkt eller behandlet for skulderplager tidligere?

- [ ] Ja   [ ] Nei

Smertestillende medisiner

Bruker du smertestillende medisiner på grunn av dine nakke- og/eller armsmerter?

- [ ] Ja   [ ] Nei

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

- [ ] Sjeldnere enn hver uke
- [ ] Hver uke
- [ ] Daglig
- [ ] Flere ganger daglig

Redusert styrke (kraftsvikt)

Har du redusert styrke (kraftsvikt) i den aktuelle skulder, arm eller hånd?

- [ ] Ja   [ ] Nei

Hvis ja: Angi varighet så nøyaktig som mulig, bruk kun en av følgende valgmuligheter

- [ ] Mindre enn 24 timer, evt. antall timer ....
- [ ] Mindre enn en uke, evt. antall døgn ....
- [ ] 1 uke til 3 mnd, evt. antall uker ....
- [ ] 3 mnd til 12 mnd
- [ ] Mer enn 12 mnd

Funksjonsskår (NDI)

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

1. Smerteintensitet

- [ ] Jeg har ingen smerter akkurat nå
- [ ] Smertene er svært svake akkurat nå
- [ ] Smertene er moderate akkurat nå
- [ ] Smertene er nokså sterke akkurat nå
- [ ] Smertene er er meget sterke akkurat nå
- [ ] Smertene er de verst tenkelige akkurat nå

2. Personlig stell

- [ ] Jeg kan stelle meg selv som normalt, uten at det gir ekstra smerter
- [ ] Jeg kan stelle meg selv som normalt, men det gir ekstra smerter
- [ ] Det er smertefullt å stelle seg, og jeg er langsom og forsiktig
- [ ] Jeg trenger noe hjelp, men klarer mesteparten av mitt personlige stell
- [ ] Jeg trenger hjelp hver dag med mesteparten av mitt personlige stell
- [ ] Jeg kler ikke på meg, har vansker med å vaske meg og holder meg i sengen

3. Løfting

- [ ] Jeg kan løfte noe tungt uten at det gir ekstra smerter
- [ ] Jeg kan løfte noe tungt, men det gir ekstra smerter
- [ ] Smertene hindrer meg i å løfte noe tungt opp fra gulvet, men jeg kan klare det hvis det er gunstig plassert, for eksempel på et bord
- [ ] Smertene hindrer meg i å løfte noe tungt, men jeg klarer noe lett eller middels tungt hvis det er gunstig plassert
- [ ] Jeg kan bare løfte noe som er meget lett
- [ ] Jeg kan ikke løfte eller bære noe i det hele tatt

4. Lesing

- [ ] Jeg kan lese så mye som jeg ønsker, uten at det gir smerter i nakken
- [ ] Jeg kan lese så mye som jeg ønsker, men med svake smerter i nakken
- [ ] Jeg kan lese så mye som jeg ønsker, men med moderate smerter i nakken
- [ ] Jeg kan ikke lese så mye som jeg ønsker, på grunn av nokså sterke smerter i nakken
- [ ] Jeg kan omtrent ikke lese i det hele tatt, på grunn av meget sterke smerter i nakken
- [ ] Jeg kan ikke lese i det hele tatt på grunn av smerter i nakken

5. Hodepine

- [ ] Jeg har ikke hodepine i det hele tatt.
- [ ] Jeg har svak hodepine som kommer nå og da
- [ ] Jeg har moderat hodepine som kommer nå og da
- [ ] Jeg har moderat hodepine som kommer jevnlig
- [ ] Jeg har sterk hodepine som kommer jevnlig
- [ ] Jeg har hodepine nesten hele tiden

6. Konsentrasjon

- [ ] Jeg kan konsentrere meg uten vansker
- [ ] Jeg kan konsentrere meg med små vansker
- [ ] Jeg har nokså store vansker med å konsentrere meg
- [ ] Jeg har store vansker med å konsentrere meg
- [ ] Jeg har svært store vansker med å konsentrere meg
- [ ] Jeg kan ikke konsentrere meg i det hele tatt

7. Arbeid (eller daglige gjøremål)

- [ ] Jeg kan gjøre så mye arbeid jeg ønsker
- [ ] Jeg kan gjøre mitt vanlige arbeid, men ikke mer
- [ ] Jeg kan gjøre mesteparten av mitt vanlige arbeid, men ikke mer
- [ ] Jeg kan ikke gjøre mitt vanlige arbeid
- [ ] Jeg kan omtrent ikke gjøre noe arbeid i det hele tatt
- [ ] Jeg kan ikke gjøre noe arbeid i det hele tatt

8. Bilkjøring

- [ ] Jeg kan kjøre bil uten at det gir smerter i nakken
- [ ] Jeg kan kjøre bil så lenge som jeg ønsker, men med svake smerter i nakken
- [ ] Jeg kan kjøre bil så lenge som jeg ønsker, men med moderate smerter i nakken
- [ ] Jeg kan ikke kjøre bil så lenge som jeg ønsker, på grunn av nokså sterke smerter i nakken
- [ ] Jeg kan omtrent ikke kjøre bil i det hele tatt, på grunn av meget sterke smerter i nakken
- [ ] Jeg kan ikke kjøre bil i det hele tatt, på grunn av smerter i nakken

9. Søvn

- [ ] Jeg har ikke problemer med å sove
- [ ] Søvnen min er litt forstyrret (mindre enn 1 times søvnløshet)
- [ ] Søvnen min er noe forstyrret (1–2 timers søvnløshet)
- [ ] Søvnen min er moderat forstyrret (2–3 timers søvnløshet)
- [ ] Søvnen min er sterkt forstyrret (3–5 timers søvnløshet)
- [ ] Søvnen min er fullstendig forstyrret (5–7 timers søvnløshet)

10. Fritid

- [ ] Jeg er i stand til å drive med alle mine fritidsaktiviteter uten at det gir smerter i nakken overhodet
- [ ] Jeg er i stand til å drive med alle mine fritidsaktiviteter, men med noe smerter i nakken
- [ ] Jeg er i stand til å drive med de fleste av, men ikke alle, mine fritidsaktiviteter på grunn av smerter i nakken
- [ ] Jeg er bare i stand til å drive med noen få av mine vanlige fritidsaktiviteter på grunn av smerter i nakken
- [ ] Jeg kan omtrent ikke drive med fritidsaktiviteter på grunn av smerter i nakken
- [ ] Jeg kan ikke drive med fritidsaktiviteter i det hele tatt

Europeisk myelopati skår (EMS)

Skalaen består av fem spørsmål som belyser ulike aspekter på ryggmargsfunksjon. Vær snill å besvare spørsmålene ved å sette kryss (kun ett kryss for hvert avsnitt) i de rutene som beskriver din situasjon best. w

1. Gangfunksjon

- [ ] Jeg kan ikke gå, trenger rullestol
- [ ] Jeg kan gå på flatt underlag med stokk eller annet hjelpemiddel
- [ ] Jeg trenger stokk eller annet hjelpemiddel i trapper, men jeg kan gå uten støtte på flatt underlag
- [ ] Jeg går klossete, men trenger ikke hjelpemidler
- [ ] Jeg går normalt, selv i trapper

2. Håndfunksjon

- [ ] Jeg kan ikke skrive for hånd eller spise med kniv og gaffel
- [ ] Jeg har problemer med å skrive for hånd eller spise med kniv og gaffel
- [ ] Jeg kan skrive for hånd og knytte slips og skolisser, men jeg gjør det klossete
- [ ] Jeg har ingen vansker med å skrive

3. Koordinasjon

- [ ] Jeg trenger hjelp med påkledning
- [ ] Jeg kan kle på meg selv, men er klossete og det tar tid
- [ ] Jeg har ingen vansker med å kle på meg

4. Blære og tarmkontroll

- [ ] Jeg har ingen kontroll over blære- og/eller tarmfunksjon
- [ ] Jeg har dårlig kontroll over blære- og/eller tarmfunksjon
- [ ] Jeg har normal blære- og tarmfunksjon

5. Nummenhet/smerte

- [ ] Jeg har store invalidiserende smerter
- [ ] Jeg opplever nummenhet og smerte, men kan leve med det
- [ ] Jeg har ingen nummenhet eller smerte

Beskrivelse av helsetilstand (EQ-5D)

Under hver overskrift ber vi deg krysse av den ENE boksen som best beskriver helsen din I DAG.

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, famile- 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 vanlige 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

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

100
90
80
70
60
50
40
30
20
10
0
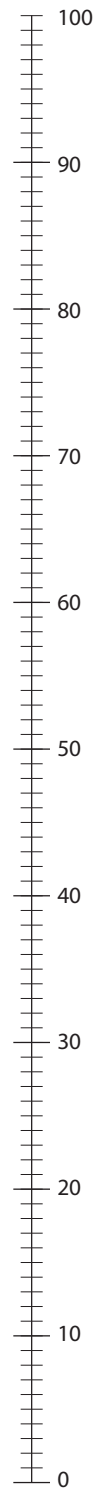
Den dårligste
helsen du kan
tenke deg

HELSEN DIN I DAG = ☐

Appendix C.

3- and 12-month postoperative patient questionnaire, NORspine

Spørreskjema for pasienter
3 måneder etter nakkeoperasjon

Nasjonalt Kvalitetsregister for Ryggkirurgi
Degenerativ nakke

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

0510 – Versjon 2

Formålet med dette spørreskjemaet er å gi leger, sykepleiere og fysioterapeuter bedre forståelse av hvilke plager pasienter med degenerative tilstander i nakke har og hvor effektiv og trygg behandlingen er. Slik kunnskap kan brukes til å gi nakkepasienter et bedre behandlingstilbud i fremtiden

Dato for utfylling

Dag   Måned   År

Friskmeldt? (tilbake i arbeid, helt eller delvis)

Hvis ja, angi dato

Dag   Måned   År

Varighet av sykemelding etter operasjon   (uker)

Arbeidsstatus (Sett kun ett kryss)

☐ I arbeid
☐ Hjemmeværende
☐ Student/skoleelev
☐ Alderspensjonist
☐ Arbeidsledig

☐ Sykemeldt
☐ Delvis sykemeldt
☐ Arbeidsavklaringspenger
☐ Uførepensjon
☐ Uførepensjon + sykemeldt

Hvis du er delvis sykemeldt eller ikke har full uførepensjon, angi prosent

_____% Sykemeldt   _____ % Ufør

Har du søkt om uførepensjon pga din sykdom/tilstand som du er operert for? *(Sett kun ett kryss)*

☐ Ja
☐ Planlegger å søke

☐ Nei
☐ Er allerede innvilget

Har du søkt om erstatning fra forsikringsselskap eller folketrygden (eventuelt yrkesskadeerstatning) pga din tilstand/sykdom som du er operert for? *(Sett kun ett kryss)*

☐ Ja
☐ Planlegger å søke

☐ Nei
☐ Er allerede innvilget

Hvilken nytte mener du at du har hatt av operasjonen?

☐ 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

Redusert styrke (kraftsvikt)

Dersom du hadde redusert styrke (kraftsvikt) i skuler/arm/hånd før operasjonen; - har dette endret seg?

☐ Har blitt helt bra
☐ Har blitt bedre
☐ Uendret
☐ Har blitt dårligere

Komplikasjoner til inngrepet? (Sett evt. flere kryss)

☐ 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 diagnosen «dyp venetrombose» (blodpropp i benet) og blitt 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 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 kraftsvikt i arm eller ben etter operasjonen?

☐ Har du etter operasjonen vedvarende ubehag ved svelging av mat/drikke?

☐ Har du etter operasjonen vedvarende problemer med stemmen din (f.eks. heshet/svak stemme)?

Hvor sterke smerter har du hatt siste uke?

Hvordan vil du gradere smertene du har hatt i **hodet** 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 **nakken** 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 **armen** (en 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

Hvor stråler smertene? (Sett kun ett kryss)

- [ ] I begge skuldre/armer
- [ ] Kun i en skulder/arm
- [ ] Ingen strålesmerter

Hvor langt ut går armsmertene dine? (Sett kun ett kryss)

- [ ] Til skulder
- [ ] Til overarm/albue
- [ ] Til underarm/håndledd
- [ ] Til finger/fingre

Smertestillende medisiner

Bruker du smertestillende medisiner på grunn av dine nakke- og/eller armsmerter?

- [ ] Ja   - [ ] Nei

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

- [ ] Sjeldnere enn hver uke
- [ ] Hver uke
- [ ] Daglig
- [ ] Flere ganger daglig

Europeisk myelopati skår (EMS)

Skalaen består av fem spørsmål som belyser ulike aspekter på ryggmargsfunksjon. Vær snill å besvare spørsmålene ved å sette kryss (kun ett kryss for hvert avsnitt) i de rutene som beskriver din situasjon best.

1.   Gangfunksjon

- [ ] Jeg kan ikke gå, trenger rullestol
- [ ] Jeg kan gå på flatt underlag med stokk eller annet hjelpemiddel
- [ ] Jeg trenger stokk eller annet hjelpemiddel i trapper, men jeg kan gå uten støtte på flatt underlag
- [ ] Jeg går klossete, men trenger ikke hjelpemidler
- [ ] Jeg går normalt, selv i trapper

2.   Håndfunksjon

- [ ] Jeg kan ikke skrive for hånd eller spise med kniv og gaffel
- [ ] Jeg har problemer med å skrive for hånd eller spise med kniv og gaffel
- [ ] Jeg kan skrive for hånd og knytte slips og skolisser, men jeg gjør det klossete
- [ ] Jeg har ingen vansker med å skrive

3.   Koordinasjon

- [ ] Jeg trenger hjelp med påkledning
- [ ] Jeg kan kle på meg selv, men er klossete og det tar tid
- [ ] Jeg har ingen vansker med å kle på meg

4.   Blære og tarmkontroll

- [ ] Jeg har ingen kontroll over blære- og/eller tarmfunksjon
- [ ] Jeg har dårlig kontroll over blære- og/eller tarmfunksjon
- [ ] Jeg har normal blære- og tarmfunksjon

5.   Nummenhet/smerte

- [ ] Jeg har store invalidiserende smerter
- [ ] Jeg opplever nummenhet og smerte, men kan leve med det
- [ ] Jeg har ingen nummenhet eller smerte

Funksjonsskår (NDI)

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

1.   Smerteintensitet

- [ ] Jeg har ingen smerter akkurat nå
- [ ] Smertene er svært svake akkurat nå
- [ ] Smertene er moderate akkurat nå
- [ ] Smertene er nokså sterke akkurat nå
- [ ] Smertene er er meget sterke akkurat nå
- [ ] Smertene er de verst tenkelige akkurat nå

2. Personlig stell

- [ ] Jeg kan stelle meg selv som normalt, uten at det gir ekstra smerter
- [ ] Jeg kan stelle meg selv som normalt, men det gir ekstra smerter
- [ ] Det er smertefullt å stelle seg selv, og jeg er langsom og forsiktig
- [ ] Jeg trenger noe hjelp, men klarer mesteparten av mitt personlige stell
- [ ] Jeg trenger hjelp hver dag med mesteparten av mitt personlige stell
- [ ] Jeg kler ikke på meg, har vansker med å vaske meg og holder meg i sengen

3. Løfting

- [ ] Jeg kan løfte noe tungt uten at det gir ekstra smerter
- [ ] Jeg kan løfte noe tungt, men det gir ekstra smerter
- [ ] Smertene hindrer meg i å løfte noe tungt opp fra gulvet, men jeg kan klare det hvis det er gunstig plassert, for eksempel på et bord
- [ ] Smertene hindrer meg i å løfte noe tungt, men jeg klare noe lett eller middels tungt hvis det er gunstig plassert
- [ ] Jeg kan bare løfte noe meget lett
- [ ] Jeg kan ikke løfte eller bære noe i det hele tatt

4. Lesing

- [ ] Jeg kan lese så mye som jeg ønsker, uten at det gir smerter i nakken
- [ ] Jeg kan lese så mye som jeg ønsker, men med svake smerter i nakken
- [ ] Jeg kan lese så mye som jeg ønsker, men med moderate smerter i nakken
- [ ] Jeg kan ikke lese så mye som jeg ønsker, på grunn av nokså sterke smerter i nakken
- [ ] Jeg kan omtrent ikke lese i det hele tatt, på grunn av meget sterke smerter i nakken
- [ ] Jeg kan ikke lese i det hele tatt på grunn av smerter i nakken

5. Hodepine

- [ ] Jeg har ikke hodepine i det hele tatt.
- [ ] Jeg har svak hodepine som kommer nå og da
- [ ] Jeg har moderat hodepine som kommer nå og da
- [ ] Jeg har moderat hodepine som kommer jevnlig
- [ ] Jeg har sterk hodepine som kommer jevnlig
- [ ] Jeg har hodepine nesten hele tiden

6. Konsentrasjon

- [ ] Jeg kan konsentrere meg uten vansker
- [ ] Jeg kan konsentrere meg små vansker
- [ ] Jeg har nokså store vansker med å konsentrere meg
- [ ] Jeg har store vansker med å konsentrere meg
- [ ] Jeg har svært store vansker med å konsentrere meg
- [ ] Jeg kan ikke konsentere meg i det hele tatt

7. Arbeid (eller daglige gjøremål)

- [ ] Jeg kan gjøre så mye arbeid jeg ønsker
- [ ] Jeg kan gjøre mitt vanlige arbeid, men ikke mer
- [ ] Jeg kan gjøre mesteparten av mitt vanlige arbeid, men ikke mer
- [ ] Jeg kan ikke gjøre mitt vanlige arbeid
- [ ] Jeg kan omtrent ikke gjøre noe arbeid i det hele tatt
- [ ] Jeg kan ikke gjøre noe arbeid i det hele tatt

8. Bilkjøring

- [ ] Jeg kan kjøre bil uten at det gir smerter i nakken
- [ ] Jeg kan kjøre bil så lenge som jeg ønsker, men med svake smerter i nakken
- [ ] Jeg kan kjøre bil så lenge som jeg ønsker, men med moderate smerter i nakken
- [ ] Jeg kan ikke kjøre bil så lenge som jeg ønsker, på grunn av nokså sterke smerter i nakken
- [ ] Jeg kan omtrent ikke kjøre bil i det hele tatt, på grunn av meget sterke smerter i nakken
- [ ] Jeg kan ikke kjøre bil i det hele tatt, på grunn av smerter i nakken

9. Søvn

- [ ] Jeg har ikke problemer med å sove
- [ ] Søvnen min er litt forstyrret (mindre enn 1 times søvnløshet)
- [ ] Søvnen min er noe forstyrret (1–2 timers søvnløshet)
- [ ] Søvnen min er moderat forstyrret (2–3 timers søvnløshet)
- [ ] Søvnen min er sterkt forstyrret (3–5 timers søvnløshet)
- [ ] Søvnen min er fullstendig forstyrret (5–7 timers søvnløshet)

10. Fritid

- [ ] Jeg er i stand til å drive med alle mine fritidsaktiviteter uten at det gir smerter i nakken overhodet
- [ ] Jeg er i stand til å drive med alle mine fritidsaktiviteter, men med noe smerter i nakken
- [ ] Jeg er i stand til å drive med de fleste av, men ikke alle, mine fritidsaktiviteter på grunn av smerter i nakken
- [ ] Jeg er bare i stand til å drive med noen få av mine vanlige fritidsaktiviteter på grunn av smerter i nakken
- [ ] Jeg kan omtrent ikke drive med fritidsaktiviteter på grunn av smerter i nakken
- [ ] Jeg kan ikke drive med fritidsktiviteter i det hele tatt

Beskrivelse av helsetilstand (EQ-5D)

Under hver overskrift ber vi deg krysse av den ENE boksen som best beskriver helsen din I DAG.

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, famile- 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 vanlige 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

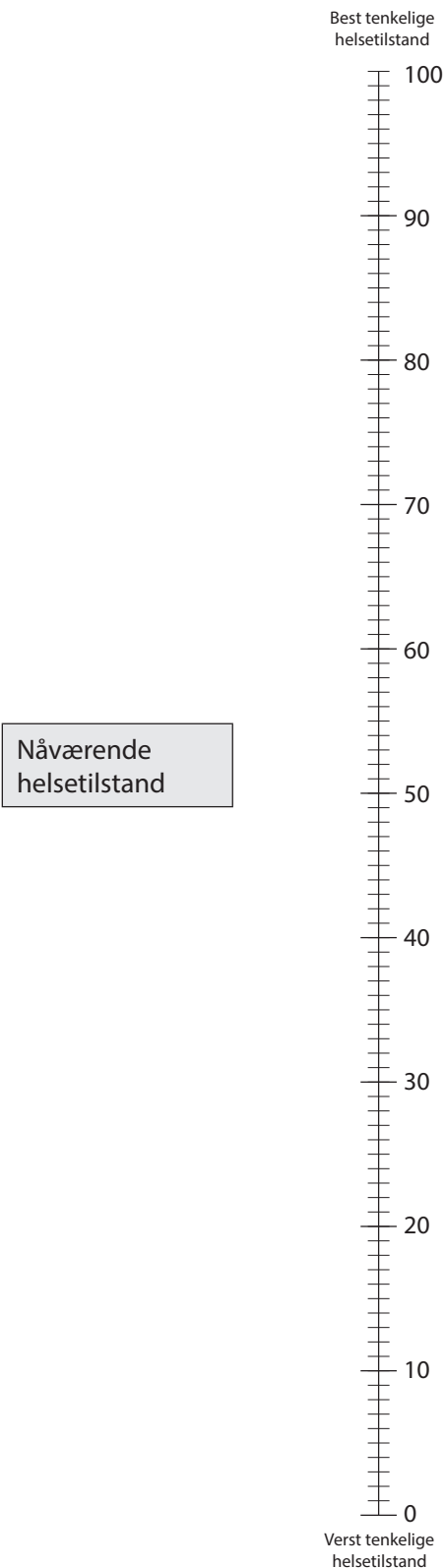☐ Jeg har svært sterke smerter eller ubehag

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

Helsetilstand

For at du skal kunne vise oss hvor god eller dårlig din helse-tilstand 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 «Nåværende helsetilstand» nedenfor til det punkt på skalaen som passer best med din helsetilstand.

Best tenkelige helsetilstand

100

90

80

70

60

Nåværende helsetilstand

50

40

30

20

10

0

Verst tenkelige helsetilstand

Appendix D.
Surgeon questionnaire, NORspine

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

Registreringsskjema for pasienter som opereres for degenerative tilstander i nakken

Nasjonalt Kvalitetsregister for Ryggkirurgi
Degenerativ nakke

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

0510 – Versjon 1

Operasjonsdato

(Må fylles ut)   Dag   Måned   År

Dato for utfylling

Dag   Måned   År

Pasientdata (Barkode)

Navn

Fødselsnr. (11 siffer)

Sykehistorie (Sett evt. flere kryss)

Tidligere nakkeoperert?

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

- Pasienten har vært operert ☐☐ ganger tidligere i Cervical columna

Andre relevante sykdommer, skader eller plager

☐ Nei

Ja, spesifiser:

☐ RA   ☐ Hodepine
☐ Bechterew   ☐ Cerebrovaskulær sykdom
☐ Annen revmatisk sykdom   ☐ Kronisk nevrologisk sykdom
☐ Under steroid/immuno-modulerende behandling   ☐ Hypertensjon
☐ Kroniske smerter i muskel skjelettsystemet   ☐ Hjertekar sykdom
☐ Carpal tunnel syndrom   ☐ Vaskulær claudikatio
☐ Skulderartrose/impingment   ☐ Kreftsykdom
☐ Whiplash/nakkeskade   ☐ Astma/kronisk lunge sykdom
☐ Osteoporose   ☐ Diabetes mellitus
☐ Depresjon/angst   ☐ Annen endokrinologisk sykdom

Annet, spesifiser _____

Radiologisk vurdering (Sett evt. flere kryss)

1. Undersøkelse

☐ CT   ☐ Rotblokkade
☐ MR   ☐ Rtg cervical columna
☐ Myelografi   ☐ Med fleksjon/ekstensjon
☐ EMG/Nevrografi

2. Funn

☐ Normal   ☐ Rotkanalstenose
☐ Skiveprolaps   ☐ Spondylolistese
☐ Cervical spinalstenose   ☐ Intramedullære signalforandringer ved MR
☐ Degenerative forandringer på flere nivå enn opererte
☐ Annet, spesifiser _____

Operasjonsindikasjon (Sett evt. flere kryss)

☐ Smerter: ☐ Nakke   ☐ Arm
☐ Parese, Grad (0-5): _____
☐ Myelopati: ☐ Sensorisk   ☐ Motorisk
☐ Annet, spesifiser _____

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

☐ Liquorlekkasje   ☐ Løsning/feilplassering av osteosyntesemateriale
☐ Dyp infeksjon   ☐ Feilplassering av implantat
☐ Overfladisk infeksjon   ☐ Operert i feil nivå
☐ Hematom   ☐ Ufullstendig dekompresjon
☐ Postoperativ spondylolistese
☐ Annet, spesifiser _____

Operasjonskategori (Kun ett kryss)

☐ Elektiv   ☐ Øyeblikkelig hjelp   ☐ ½ øyeblikkelig hjelp

Dagkirurgi (ingen døgnopphold på avdelingen)

☐ Ja   ☐ Nei

ASA-klassifisering fra anestesiskjema (Kun ett kryss)

☐ 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

RANAWAT klassifikasjon for medullopati (Kun ett kryss)

- [ ] I   Ingen nevrologiske utfall
- [ ] II   Subjektiv svakhet, hyperreflexi og dysestesi
- [ ] III   Objektiv svakhet og langbane symptomer
  - [ ] A   Oppegående
  - [ ] B   Quadriparese og ikke oppegående

Operasjonsmetode (Sett evt. flere kryss)

Tilgang (Sett evt flere kryss)

- [ ] Bakre
- Fremre:
  - [ ] Hø
  - [ ] Ve

Har operatøren brukt synsfremmende midler?

- [ ] Mikroskop
- [ ] Lupe-briller
- [ ] Endo-skopi
- [ ] Ingen

Fremre cervical diskektomi og fusjon eller skiveprotese

- [ ] Diskektomi
- [ ] Benblokk
- [ ] Plate
- [ ] Cage/bur
- [ ] Skiveprotese

Kirurgisk dekompresjon

- [ ] Bakre foramenotomi
- [ ] Unilateral
- [ ] Bilateral

Annen bakre dekompresjon

- [ ] Laminektomi
- [ ] Laminoplastikk
- [ ] Skip laminektomi
- [ ] Hemilaminektomi

- [ ] Korpektomi
- [ ] Plate
- [ ] Cage/Bur
- [ ] Beinblokk

Andre operasjonsmetoder (Sett kun ett kryss)

- [ ] Revisjon av osteosyntesemateriale
- [ ] Revisjon av cage
- [ ] Fjerning av osteosyntesemateriale
- [ ] Revisjon av skiveprotese

Annet, spesifiser _____

Bakre fusjon (Sett evt. flere kryss)

- [ ] Cervical
- [ ] Occipitocervical

Instrumentering   [ ] Wire   [ ] Skruer   [ ] Stag

Proximale nivå, f.eks C0 [ ][ ]   Distale nivå, f.eks. TH1 [ ][ ][ ]

Type bengraft (Sett evt. flere kryss)

- [ ] Autograft
- [ ] Bensubstitutt
- [ ] Bank-ben

Dekomprimert nivå og side (Sett evt. flere kryss)

- [ ] C0/C1   [ ] Hø   [ ] Ve
- [ ] C1/C2   [ ] Hø   [ ] Ve
- [ ] C2/C3   [ ] Hø   [ ] Ve
- [ ] C3/C4   [ ] Hø   [ ] Ve
- [ ] C4/C5   [ ] Hø   [ ] Ve
- [ ] C5/C6   [ ] Hø   [ ] Ve
- [ ] C6/C7   [ ] Hø   [ ] Ve
- [ ] C7/TH1   [ ] Hø   [ ] Ve

Annet, spesifiser _____

Antibiotikaprofylakse

- [ ] Nei   [ ] Ja, spesifiser;

     Medikament:.................... Dose:.......... Antall:..........
      Eks: Keflin   2g   x1

- [ ] Kun operasjonsdagen
- [ ] Evt. antall døgn..............

Sårdren

- [ ] Ja   [ ] Nei

Knivtid (hud til hud) angi klokkeslett

Opr. start [ ][ ] [ ][ ] (timer/min)   Opr. slutt [ ][ ] [ ][ ] (timer/min)

Perioperative komplikasjoner (Sett evt. flere kryss):

- [ ] Durarift
- [ ] Anafylaktisk reaksjon
- [ ] Nerverotskade
- [ ] Medullaskade
- [ ] Operert på feil nivå/side
- [ ] Øsofagusskade
- [ ] Feil plassering av implantat
- [ ] Skade på større blodkar
- [ ] Transfusjonskrevende peroperativ blødning
- [ ] Kardiovaskulære komplikasjoner
- [ ] Respiratoriske komplikasjoner
- [ ] Annen nerveskade
- [ ] Annet, spesifiser _____

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

[ ][ ][ ][ ][ ]   [ ][ ][ ][ ][ ]

Fylles ut ved endt opphold/utskrivelse

Utskrivelsesdato samt totalt antall liggedøgn

Utskrivelsesdato [ ][ ] [ ][ ] [ ][ ]   [ ][ ][ ] (dager)
                 dag   mnd   år

Ved dødsfall under oppholdet, oppgi årsak

Spesifiser _____

Paper I

Clinical Study

Criteria for success after surgery for cervical radiculopathy—estimates for a substantial amount of improvement in core outcome measures

Christer Mjåset, MD[a,b,]*, John-Anker Zwart, MD, PhD[g,h], Caroline M.W. Goedmakers, BSc[c,d], Timothy R. Smith[d], Tore K. Solberg, MD, PhD[e,f], Margreth Grotle, PT, PhD[i,j]

[a] *Research and Communication Unit for Musculoskeletal Health (FORMI), Oslo University Hospital, Oslo, Norway*
[b] *Department of Neurosurgery, Oslo University Hospital, Oslo, Norway*
[c] *Department of Neurosurgery, Leiden University Medical Center, Leiden, The Netherlands*
[d] *Computational Neuroscience Outcomes Center (CNOC), Department of Neurosurgery, Brigham and Women's Hospital, School of Medicine, Harvard University, Boston, MA, USA*
[e] *Faculty of Health Sciences, Oslo Metropolitan University, Oslo, Norway*
[f] *Institute of Clinical Medicine, The Arctic University of Norway (UiT), Tromsø, Norway*
[g] *Faculty of Medicine, University of Oslo, Oslo, Norway*
[h] *Research and Communication Unit for Musculoskeletal Health (FORMI), Oslo University Hospital, Oslo, Norway*
[i] *Department of Physiotherapy, Oslo Metropolitan University, Oslo, Norway*
[j] *Research and Communication Unit for Musculoskeletal Health (FORMI), Oslo University Hospital, Oslo, Norway*

Received 24 January 2020; revised 24 May 2020; accepted 25 May 2020

ABSTRACT

BACKGROUND CONTEXT: Defining clinically meaningful success criteria from patient-reported outcome measures (PROMs) is crucial for clinical audits, research and decision-making.

PURPOSE: We aimed to define criteria for a successful outcome 3 and 12 months after surgery for cervical degenerative radiculopathy on recommended PROMs.

STUDY DESIGN: Prospective cohort study with 12 months follow-up.

PATIENT SAMPLE: Patients operated at one or two levels for cervical radiculopathy included in the Norwegian Registry for Spine Surgery (NORspine) from 2011 to 2016.

OUTCOME MEASURES: Neck disability index (NDI), Numeric Rating Scale for neck pain (NRS-NP) and arm pain (NRS-AP), health-related quality-of-life EuroQol 3L (EQ-5D), general health status (EQ-VAS).

METHODS: We included 2,868 consecutive cervical degenerative radiculopathy patients operated for cervical radiculopathy in one or two levels and included in the Norwegian Registry for Spine Surgery (NORspine). External criterion to determine accuracy and optimal cut-off values for success in the PROMs was the global perceived effect scale. Success was defined as "much better" or "completely recovered." Cut-off values were assessed by analyzing the area under the receiver operating curves for follow-up scores, mean change scores, and percentage change scores.

RESULTS: All PROMs showed high accuracy in defining success and nonsuccess and only minor differences were found between 3- and 12-month scores. At 12 months, the area under the receiver operating curves for follow-up scores were 0.86 to 0.91, change scores were 0.74 to 0.87, and percentage change scores were 0.74 to 0.91. Percentage scores of NDI and NRS-AP showed the best

accuracy. The optimal cut-off values for each PROM showed considerable overlap across those operated due to disc herniation and spondylotic foraminal stenosis.

CONCLUSIONS: All PROMs, especially NDI and NRS-AP, showed good to excellent discriminative ability in distinguishing between a successful and nonsuccessful outcome after surgery due to cervical radiculopathy. Percentage change scores are recommended for use in research and clinical practice. © 2020 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license. (http://creativecommons.org/licenses/by/4.0/)

Introduction

The last decade's advances in surgical technique and equipment have increased the effectiveness and safety of surgical intervention for cervical degenerative radiculopathy (CDR) making operations for disc herniation and spondylotic foraminal stenosis high volume procedures [1,2]. Since surgery is a costly treatment with potential risks, there has been a need to define criteria for substantial benefit to facilitate doctor-patient communication and assess quality of surgical care [3,4]. In this way, the introduction of patient-reported outcome measures (PROMs) [5] and the concept of minimal important change (MIC) have been important to establish evidence-based practice. The MIC represents the smallest difference in PROM score that is clinically beneficial within a patient group, as recommended by consensus-based standards for the selection of health status measurement instruments [5,6]. Other similar concepts are currently being used, like minimal clinically important difference (MCID) [7].

The concept of success, representing a more optimal treatment goal than the MIC, can be used both in communication with patients in clinical practice and in research but is often poorly defined or surgeon-reported. One way to assess it more accurately is to align it with the concept of substantial improvement which was first described for patients undergoing lumbar surgery [8] and later assessed for heterogeneous patient populations undergoing surgery for degenerative spine conditions [9,10]. For CDR patients, however, PROM-based definitions of substantial change after surgery have not been well defined.

The aim of this study was to define success criteria after surgery for cervical radiculopathy performed in daily clinical practice based on frequently used PROMs; the neck disability index (NDI), the Euro-Qol (EQ-5D-3L) with visual analogue scale (EQ-VAS), and numeric rating scale for arm pain (NRS-AP) and neck pain (NRS-NP).

Materials and methods

Data source

All data were collected through the Norwegian Registry for Spine Surgery (NORspine). NORspine is a government funded comprehensive clinical registry receiving no industry funding and used for quality assessment and research. Informed consent is obtained from all patients before

they enter the registry. Currently, all centers performing cervical spine surgery in Norway report data to NORspine (coverage=100%) and the operation recording rate is 78% (completeness) [11].

The board of NORspine allowed us to access the data after the Norwegian Committee for Medical and Health Research Ethics Midt approved our research protocol (2014/344).

Design

This is a prospective cohort study with follow-up at 3 and 12 months. This report is consistent with the strengthening the reporting of observational studies in epidemiology statement [12] and the methods used are in accordance with the consensus-based standards for the selection of health measurement instruments recommendations [6].

Eligibility criteria

Of 4,229 consecutive patients operated for degenerative disorders in the cervical spine between January 2011 and August 2016 in ten private or public clinics, 2,868 were included for the main investigation. Eligible patients were those who had undergone surgery with either anterior cervical discectomy and fusion (ACDF) or arthroplasty (ACDA) (n=2,640) or posterior cervical foraminotomy or hemilaminectomy (n=228) at one or two levels due to CDR, excluding patients with more complex pathology, verified or possible myelopathy, and former operation(s) at the index level (Fig. 1).

Two diagnostic subgroups were investigated separately: patients with disc herniation (n=1,182) and patients with spondylotic foraminal stenosis (n=430). Since these degenerative changes often coexist, we excluded patients operated for both diagnoses. Also, patients operated at more than one level, indicating more widespread cervical spondylosis, were excluded in these subgroup analyses. We chose this strategy because it may be difficult to decide the clinical relevance of multiple nerve root compressions found on MRI. Therefore, the total number of patients in the two diagnostic subgroups (n=1,612) do not add up to the number of patients for the whole material (n=2,868) in Fig. 1.

Measurements

The comprehensive NORspine self-administered questionnaire consists of information about sociodemographic
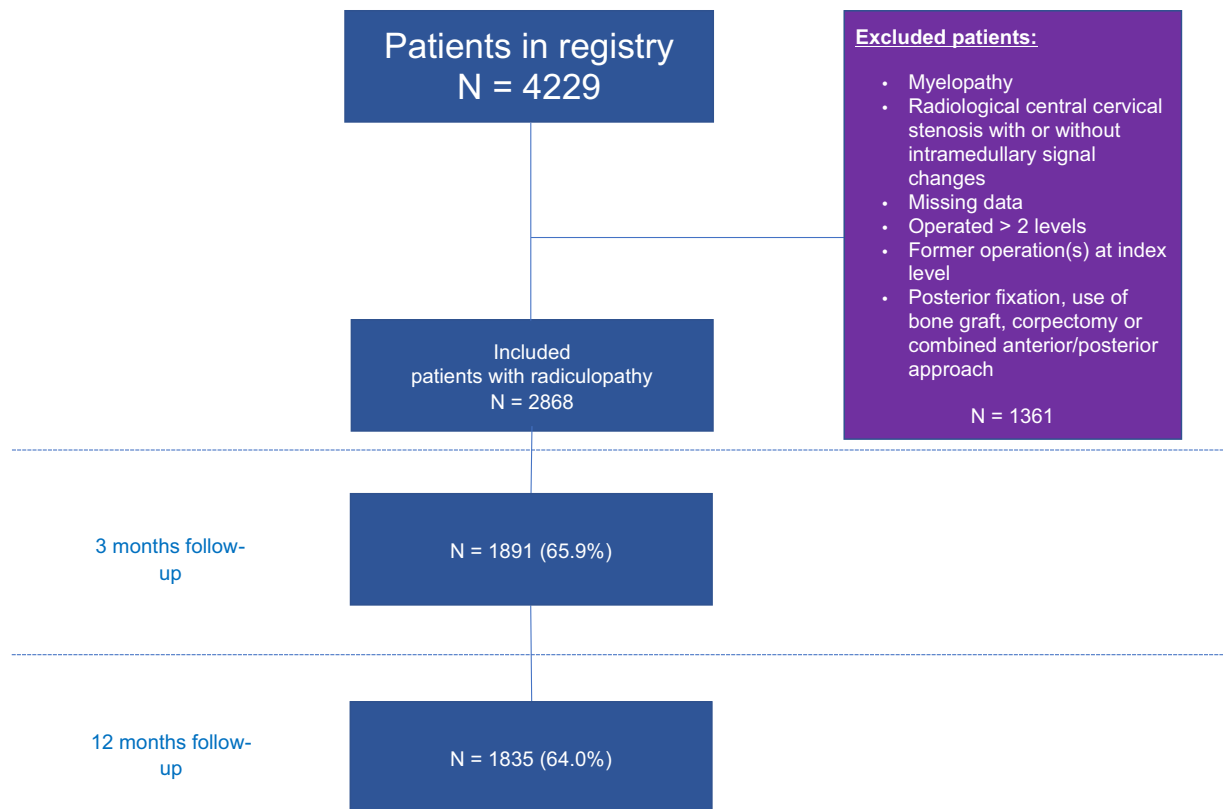
Fig. 1. Exclusion criteria for patients with follow-up rates.

factors, lifestyle, work, pain location and duration of symptoms in addition to PROMs. Patients complete it at admission for surgery (baseline) and at home 3 and 12 months after surgery after receiving it by postal mail. To avoid selective reporting, the NORspine central unit collects follow-up data without involvement of the treating hospitals. The patient receives a reminder with a new questionnaire if he or she does not respond.

After the operation, the surgeon completes a separate form with information about diagnosis, treatment, comorbidity (including the American Society of Anesthesiologists physical status (ASA), surgical indication (radiculopathy, myelopathy, pain paresis and others) and type of operation.

The following PROMs were included at all time points:

Neck disability index (NDI) [13] is a measure of neck pain related disability, containing 10 items (pain, personal care, lifting, reading, headaches, concentration, work, driving, sleeping and recreation), all scored on a 6-point ordinal scale (0−5). The 10 items are summarized and recalculated to a percentage score ranging from 0 to 100 (no to maximum disability).

EuroQoL (EQ-5D-3L) [14] is a generic measurement and preference-weighted measure of health-related quality-of-life based on five dimensions: mobility, self-care, usual activity, pain/discomfort and anxiety/discomfort. For each dimension the patient assesses three possible levels (3L) of problems; "none," "mild to moderate," and "severe." The score ranges from −0.59 to 1, where 1 corresponds to

perfect health and 0 to death and negative values worse than death. In the second part, called the EQ-VAS, the patient is asked to indicate overall health on a vertical analogue scale, ranging from 0 to 100 ("worst to "best imaginable health").

Numeric rating scale for arm (NRS-AP) and neck pain (NRS-NP) [15,16] assesses pain severity ranging from 0 to 10 ("no" to "worst conceivable pain") on two separate scales. Information about joint pain is not collected.

Included in the two follow-up questionnaires is also *The Global Perceived Effect scale (GPE)* [17] which measures the patient perceived benefit of an operation by asking how the situation is for the patient after the procedure. There are seven response categories; (1) "completely recovered," (2) "much improved," (3) "slightly improved," (4) "unchanged," (5) "slightly worse," (6) "much worse", and (7) "worse than ever." In this study, the GPE scale was applied as an external criterion to define cut-offs for success on the PROM scales. Patients reporting to be "completely recovered" or "much improved" (1−2) were classified as having a "successful outcome," while those who considered themselves to be "slightly improved," "unchanged" or worse (3−7) were classified as having a "nonsuccessful" outcome. The same method has previously been applied on several datasets from NORspine [18−21].

Statistical analyses

All statistical analyses were performed with the Statistical Package for the Social Sciences (SPSS, version 25).

Baseline characteristics and preoperative PROMs were reported as means and standard deviations of continuous variables and as percentages of categorical variables. The patient cohort was analyzed as a whole, then separately for 3- and 12-month follow-ups, procedural groups (the posterior approach group and the anterior approach group) and diagnostic groups (the disc herniation group and the spondylotic foraminal stenosis group).

We calculated the change score as the absolute difference between the pre- and postoperative scores. The percentage change score equals the absolute difference divided by the baseline score, multiplied by 100.

The distribution of 3- and 12-month scores, that is the follow-up, mean change and percentage change scores according to each of the response alternatives of the GPE scale, were analyzed by ANOVA analysis. Because the EQ-5D-3L questionnaire values range from −0.6 to 1.0, it is not mathematically possible to evaluate the percent change. However, percentage change score was measured for EQ-VAS (0−100). The correlations between the ordinal GPE scale and the PROMs were analyzed by the Spearman rank coefficient, *rho*.

Receiver operating curves (ROC) were used to identify discriminative ability of the PROMs and to define the optimal cut-off with the highest sensitivity and specificity. ROC-curves were made by plotting the sensitivity against (1—specificity) for each possible cut-off value for success. The sensitivity refers to the probability of correctly classifying an individual replying "completely recovered" or "much improved" into the group with a successful outcome (1−2) based on the simultaneously reported PROM score. Correspondingly, the specificity refers to the probability of

correctly classifying a patient reporting anything less than "much improved" into the "nonsuccessful" group (3−7). The area under the ROC-curves (AUC) with 95 % confidence interval was used for discriminative ability as it describes the test's accuracy in correctly classifying a case according to the anchor. The larger the area under the curve, the greater is the accuracy of the test. The AUC is classified as "excellent" from 1.0 to 0.90, "good" from 0.90 to 0.80, "fair" from 0.80 to 0.70, "poor" from 0.70 to 0.60, and "failed" from 0.60 to 0.50 [22].

Results

Out of the 4,229 patients operated for CDR in the NORspine registry, 2,868 patients met the inclusion criteria. Of these patients, 2,640 patients had undergone either anterior cervical discectomy and fusion (n=2,609) or anterior cervical discectomy and arthroplasty (n=31). Another 228 patients were operated with posterior approach procedures, meaning either unilateral or bilateral posterior cervical foraminotomy (n=227) or hemilaminectomy (n=1).

A total of 66% and 64% of the patients responded to the 3- and 12-months follow-up, respectively (Fig. 1). The nonresponding patients were slightly older, were more likely to be men, to smoke, to have less comorbidity and low ASA level, and to score slightly poorer on levels of pain severity, disability, and health-related quality-of-life (Table 1). Baseline characteristics of the whole radiculopathy group and of the two diagnostic subgroups operated on one-level (disc herniation and foraminal stenosis group) are presented in Table 2. The spondylotic foraminal stenosis group had a

Table 1

Baseline characteristics of respondents and nonrespondents to follow-up at 12 months

	Respondents N=1,843		Nonrespondents N=1,025		
	N		N		Sig. (2-tailed)/ chi-square
Age (years); Mean (SD*)	1,843	50.9 (9.2)	1,023	46.6 (8.7)	0.001
Female, no (%)	1,843	910 (49.4)	1,025	438 (42.7)	<0.001
ASA level (1−4); Mean (SD)	1,770	1.7 (0.6)	1,006	1.6 (0.6)	0.076
Body mass index; Mean (SD)	1,803	26.8 (4.2)	996	26.9 (4.4)	0.443
Smokers, no (%)	1,807	521 (28.8)	1,001	410 (41.0)	<0.001
University/College education	1,799	684 (38.0)	994	334 (33.6)	0.02
Degenerative neck changes, no (%)	1,843	538 (29.2)	1,025	265 (25.9%)	0.056
Comorbidity, no (%)	1,816	745 (41.0)	1,004	370 (36.9)	0.03
Preoperative paresis	1,798	1,411 (78.5)	1,002	799 (79.7)	0.432
Emergency surgery	1,833	120 (6.6)	1,023	59 (5.7)	0.412
NDI[†]; Mean (SD)	1,837	40.6 (15.1)	1,022	42.1 (14.9)	0.011
NRS-AP[‡]; Mean (SD)	1,810	6.4 (2.3)	1,002	6.3 (2.4)	0.226
NRS-NP[§]; Mean (SD)	1,801	6.0 (2.5)	999	6.2 (2.4)	0.011
EQ-5D-3L[‖]; Mean (SD)	1,763	0.44 (0.32)	973	0.41 (0.33)	0.029
EQ-VAS[¶]; Mean (SD)	1,753	51.0 (20.2)	947	48.9 (20.1)	0.011

* Standard deviation.
† Neck disability index (0−100).
‡ Numeric rating scale for arm pain (0−10).
§ Numeric rating scale for neck pain (0−10).
‖ Health-related quality-of-life by EuroQol (−0.4−1.0).
¶ General health status by EuroQol (0−100).

Table 2
Baseline characteristics. Characteristics of the whole radiculopathy group and of the two diagnostic groups operated on one-level and with either disc herniation or spondylotic foraminal stenosis

	Whole radiculopathy group N=2,868		Disc herniation N=1,182		Spondylotic foraminal stenosis N=430		
	N		N		N		Sig. (2-tailed)/ chi-square
Age (years); Mean (SD*)	2,866	49.4 (9.2)	1,181	46.4 (9.0)	430	53.1 (9.1)	<0.001
Female, no (%)	2,868	1,348 (47.0)	1,182	595 (50.3)	430	178 (41.4)	0.002
ASA level (1−4); Mean (SD)	2,776	1.7 (0.6)	1,147	1.6 (0.6)	415	1.8 (0.6)	<0.001
Body mass index; mean (SD)	2,799	26.86 (4.2)	1,148	26.7 (4.4)	418	27.0 (4.3)	0.326
Smokers, no (%)	2,864	931 (32.5)	1,155	385 (33.3)	421	132 (31.4)	0.497
Comorbidity, no (%)	2,820	1,115 (39.5)	1,167	381 (32.6)	419	192 (45.8)	<0.001
Anterior surgical approach, no (%)	2,868	2,640 (92.1)	1,182	1,169 (98.9)	430	315 (73.3)	<0.001
NDI[†] (SD) (0−100)	2,859	41.2 (15.0)	1,179	42.2 (15.2)	428	40.4 (14.7)	<0.001
NRS-AP[‡] (SD) (0−10)	2,812	6.4 (2.3)	1,168	6.5 (2.3)	417	6.2 (2.3)	<0.001
NRS-NP[§] (0−10) (SD)	2,800	6.1 (2.5)	1,164	6.1 (2.5)	416	6.1 (2.4)	<0.001
EQ-5D-3L[‖] (SD) (-0.6−1)	2,736	0.43 (0.32)	1,134	0.42 (0.33)	405	0.46 (0.31)	0.005
EQ-VAS[¶] (SD) (0−100)	2,700	50.3 (20.2)	1,120	48.7 (20.7)	405	51.8 (18.7)	<0.001

* Standard deviation.
† Neck disability index (0−100).
‡ Numeric rating scale for arm pain (0−10).
§ Numeric rating scale for neck pain (0−10).
‖ Health-related quality-of-life by EuroQol (−0.4−1.0).
¶ General health status by EuroQol (0−100).

higher proportion of men, higher age, ASA level, degenerative changes in the neck and comorbidity as compared to the disc herniation group. Patients with disc herniation had more severe symptoms at baseline than patients with spondylotic foraminal stenosis, as well as lower health condition scores. There were minor differences in the baseline PROM scores between the two diagnostic subgroups. For the procedural groups, patients operated with posterior approach procedures had significantly better PROM scores than the anterior approach group: NDI 35.3 versus 41.7, p<.001; NRS-AP 5.5 versus 6.4, p<.001, NRS-NP 5.8 versus 6.1, p<.001; EQ-5D-3L 0.4 versus 0.5, p=.005; EQ-VAS 56.6 versus 49.8, p<.001.

The mean follow-up scores of PROMs at 12 months according to each GPE category are presented in Fig. 2A−E. For all PROMs, there was a stepwise decrease in follow-up scores for patients who reported themselves to be completely recovered and much better compared to those reporting no change or worsening. The results of the mean change scores and the mean percentage change scores at 12 months showed a similar pattern (Appendix A), as well as the follow-up score, change score and percentage change score at 3 months (obtained on request). The correlations between the PROMs and the GPE were moderate to strong, especially for NDI and NRS-AP follow-up scores and percentage change scores (0.7−0.8) but weaker for mean change scores (0.5−0.7). The correlations were generally weaker for the NRS-NP, EQ-5D-3L and EQ-VAS (0.4−0.7) scores.

We found minor differences in AUC and cut-off values between 3- and 12-month scores. Therefore, further analysis of the data is presented only for PROMs at 12-month

follow-up. 3-month scores can be found in Appendix B. AUC for NDI and NRS-AP follow-up scores and percentage change scores showed from "good" to "excellent" test accuracy (Table 3). NRS-NP, EQ-5D-3L and EQ-VAS showed either "good" or "fair" test accuracy. In general, AUC was slightly lower for the change scores than for the follow-up scores and the percentage change scores.

In Table 3, we present the cut-off values for follow-up scores, change scores and percentage change scores with highest sensitivity and specificity for the PROMs at 12 months. The cut-off values for the NDI and NRS-AP had highest sensitivity and specificity, showing that at follow-up for example a NDI percentage change score of 35% or more provided a sensitivity and specificity of 84% in distinguishing between a successful outcome or not. The NRS-AP had a larger percentage change score of 47%, whereas the NRS-NP score was 39%. Both these PROMs had slightly lower accuracy estimates. The EQ-5D-3L and EQ-VAS showed the poorest discriminative ability of success versus nonsuccess. For the subgroup analyses there were only minor variations across the two diagnoses. Finally, we also found minor differences between anterior approach and posterior approach procedural groups regarding cut-off scores (Table 4) and AUC (Appendix C).

Discussion

We found very good to excellent discriminative ability in distinguishing between success and nonsuccess following neck surgery due to radiculopathy for the most commonly used PROMs. The NDI and the NRS-AP had the highest
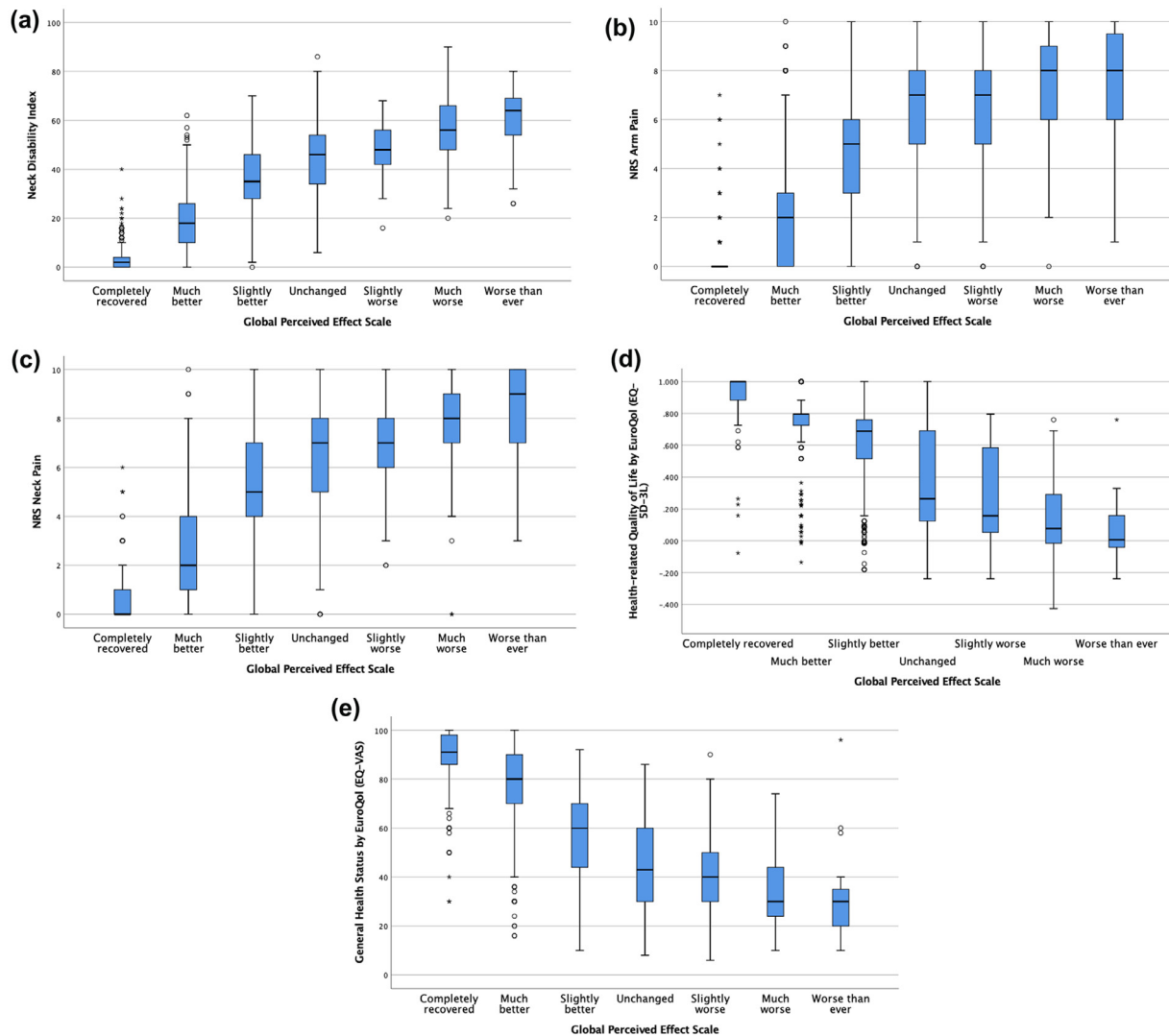
Fig. 2. (A−E). Boxplots of global perceived effect scale (GPE) and follow-up scores of patient-reported outcome measures (PROMs) at 12 months. Values which are more than three box lengths from either end of the box are denoted by asterisks ("*"). Values which are between one and a half and three box lengths from either end of the box are denoted by "o" (outliers). (A): Boxplot of neck disability index (NDI) and GPE at 12 months. (B): Boxplot of numeric rating scale for arm pain (NRS-AP) and GPE at 12 months. (C): Boxplot of numeric rating scale for neck pain (NRS-NP) and GPE at 12 months. (D): Boxplot of health-related quality-of-life by EuroQol (EQ-5D-3L) and GPE at 12 months. (E): Boxplot of general health status by EuroQol (EQ-VAS) and GPE at 12 months.

discriminative ability at 3 and 12 months. The NRS-NP, EQ-5D-3L and EQ-VAS showed markedly lower accuracy.

We found a better discriminative ability for the percentage change scores and the follow-up scores compared to the change scores. This finding is in line with previous studies conducted on surgery for lumbar disc herniation [18] and lumbar spinal stenosis [19,20]. Furthermore, the use of change scores for benchmarking has been criticized for not taking into account the patient's baseline score [23−25]. The percentage change score, on the other hand, tells something about the actual improvement the patient has been through. Also, our impression is that patients seem to put more emphasis on the follow-up score rather than the change score in clinical practice. We therefore recommend using the cut offs for success on follow-up and percentage change scores in clinical practice and future studies.

We found only minor differences in cut-off values across the two diagnostic groups and between 3 and 12 months after surgery. This means that the same cut-off scores can be applied on different time intervals and across subgroups of patients operated for CDR. One exception was the cut-off value for the NRS-NP percentage change score. Patients with spondylotic foraminal stenosis had to undergo a considerably greater change for the procedure to be considered a success (43.7%) than patients with disc herniation (35.4%). Since this is the only major difference between the two diagnostic groups, the result should be interpreted carefully.

For the two procedural groups, one cut-off score can be used. This is supported by findings in recent studies [26,27]. However, the posterior approach group was small in comparison to the anterior approach group (n=228 vs.

Table 3
Area under the curve and cut-off values for "success" for all patient-reported outcome measures at 12 months

		Follow-up score(points)	Change score(points)	Percentage change score (%)
NDI*	AUC[†] (95% CI)	0.91 (0.89−0.92)	0.87 (0.85−0.89)	0.91 (0.89−0.93)
	Cut-off (% sensitivity, % specificity)	24.2 (83.1, 83.1)	13.5 (79.4, 76.1)	35.1 (83.7, 83.6)
NRS-AP[‡]	AUC (95% CI)	0.86 (0.84−0.88)	0.81 (0.78−0.83)	0.85 (0.82−0.87)
	Cut-off (% sensitivity, % specificity)	2.50 (83.0, 75.5)	2.50 (80.0, 66.6)	47.2 (82.1, 74.2)
NRS- NP[§]	AUC (95% CI)	0.88 (0.86−0.90)	0.79 (0.76−0.81)	0.86 (0.83−0.88)
	Cut-off (% sensitivity, % specificity)	3.50 (80.1, 81.9)	1.50 (78.5, 61.9)	38.8 (79.6, 78.8)
EQ-5D-3L[‖]	AUC (95% CI)	0.86 (0.84−0.88)	0.74 (0.71−0.77)	Not possible to calculate
	Cut-off (% sensitivity, % specificity)	0.75 (79.5, 72.0)	0.11 (70.3, 68.7)	Not possible to calculate
EQ-VAS[¶]	AUC (95% CI)	0.88 (0.86−0.89)	0.78 (0.76−0.81)	0.74 (0.71−0.77)
	Cut-off (% sensitivity, % specificity)	69.0 (83.6, 24.1)	10.5 (76.7, 66.1)	24.2 (72.0, 63.3)

* Neck disability index (0−100).
[†] Area under the curve.
[‡] Numeric rating scale for arm pain (0−10).
[§] Numeric rating scale for neck pain (0−10).
[‖] Health-related quality-of-life by EuroQol (−0.4−1.0).
[¶] General health status by EuroQol (0−100).

n=2,540) and one should be careful to conclude on the basis of our results alone.

Conceptually, "success," implying a substantial improvement, is different from the MIC. Therefore, we chose to use "much better" or "completely recovered" as success criteria on the GPE (1−2) and defined "slightly better" and the other categories (GPE 3−7) as a "nonsuccess." Substantial improvement has previously been assessed for populations constituted by both radiculopathy and myelopathy patients [9,10] and on lumbar spine surgery cohorts [8,19,21], but not for radiculopathy patients alone. Fig. 2 illustrates that our definitions were reasonable.

Often in studies of MIC/MCID, the category "slightly better" is placed in the "improved" class [28]. This distinction is important to consider when interpreting our results. For instance, the cut-off values for NDI change score was 13.5 points, which is in line with previous definitions of MIC for neck patients [10,29−31]. Similar concordance with MIC was also found for the other PROMs. Also, in previous NORspine studies on lumbar surgery patients, cut-off values for a successful outcome assessed by the Oswestry Disability Index, NRS leg pain and NRS back pain were found to be at the same or slightly higher level as compared to NDI, NRS-AP and NRS-NP in this study [19,21].

Table 4
Cut-off values with sensitivity and specificity for all patient-reported outcome measures in the two diagnostic subgroups and the two procedural groups. Estimates for the 12-months follow-up score, and the change score and percentage change score from baseline to 12-months follow-up

		Disc herniation (% sensitivity, % specificity)	Spondylotic foraminal stenosis (% sensitivity, % specificity)	Anterior approach procedures (% sensitivity, % specificity)	Posterior approach procedures (% sensitivity, % specificity)
NDI*	Follow-up score (points)	25.9 (84.5, 83.8)	23.3 (82.7, 78.7)	24.2 (83.7, 82.0)	21.0 (83.6, 80.2)
	Change score (points)	13.5 (80.8, 76.1)	13.5 (81.7, 72.0)	13.5 (79.6, 76.5)	12.5 (78.6, 72.3)
	Percentage change score (%)	36.2 (84.6, 84.1)	36.3 (86.2, 84.5)	36.3 (84.2, 84.3)	38.0 (81.8, 80.8)
NRS-AP[†]	Follow-up score (points)	2.50 (81.6, 78.8)	2.50 (83.7, 72.6)	2.50 (84.2, 74.6)	1.50 (90.0, 70.2)
	Change score (points)	2.50 (81.6, 66.5)	2.50 (76.7, 72.7)	2.50 (80.3, 65.8)	2.50 (75.7, 74.5)
	Percentage change score (%)	47.2 (83.2, 73.9)	47.2 (79.8, 76.1)	47.2 (82.4, 74.5)	46.6 (86.1, 72.8)
NRS-NP[‡]	Follow-up score (points)	3.50 (83.1, 81.6)	2.50 (85.6, 70.7)	3.50 (80.9, 81.0)	2.50 (81.7, 73.8)
	Change score (points)	1.50 (77.4, 65.8)	2.50 (71.7, 74.7)	1.50 (79.2, 62.4)	1.50 (74.3, 66.0)
	Percentage change score (%)	35.4 (79.6, 79.7)	43.7 (83.0, 81.7)	35.4 (78.4, 78.5)	36.7 (87.0, 78.6)
EQ-5D-3L[§]	Follow-up score (points)	0.75 (81.3, 75.1)	0.74 (78.1, 65.4)	0.75 (79.4, 73.1)	0.74 (80.7, 61.4)
	Change score (points)	0.11 (71.3, 70.3)	0.09 (70.8, 66.6)	0.11 (70.0, 68.8)	0.12 (74.3, 70.2)
EQ-VAS[‖]	Follow-up score (points)	69 (85.1, 77.0)	68 (84.6, 72.9)	69.0 (83.0, 77.8)	73.0 (78.3, 75.4)
	Change score (points)	15.5 (71.7, 71.0)	12.5 (75.0, 74.7)	13.5 (75.0, 69.8)	13.5 (72.3, 67.9)
	Percentage change score (%)	25.5 (70.2, 63.0)	24.5 (70.2, 64.8)	27.6 (66.4, 67.0)	24.5 (65.1, 66.7)

* Neck disability index (0−100).
[†] Numeric rating scale for arm pain (0−10).
[‡] Numeric rating scale for neck pain (0−10).
[§] Health-related quality-of-Life by EuroQol (−0.4−1.0).
[‖] General health status by EuroQol (0−100).

Limitations and strengths of study

The main limitation of this study is using the GPE scale as an anchor, since it is a self-reported scale, influenced by the current health status of the patient [17]. Using a more objective anchor could be advisable [32,33]. However, no objective golden standard currently exists. The psychometric properties of the GPE seems to be good [17,34−36]. It has therefore been recommended, despite its limitations [23,37].

Another limitation is the nonrespondent rate of approximately 35%. Although it may be regarded as acceptable for a spine registry [38], it might represent a selection bias. Some of the baseline characteristics of the nonrespondents (Table 1) have been associated with poorer outcomes [39], though others have not. Also, two previous studies found no differences in outcome when comparing respondents and nonrespondents at follow-up [40,41].

A major strength of this study is the large sample size of patients operated in daily clinical practice [11] indicating a high external validity of our results.

Conclusion

In conclusion, this study showed the best ability in distinguishing between a successful and nonsuccessful outcome 12 months after surgery for a NDI follow-up score lower than 24 or a percentage change score of larger than 35% and for a NRS-AP follow-up score lower than 2.5 or a percentage change score larger than 47%. In this cohort, these criteria were stable at both 3 and 12 months of follow-up, and across subgroups of patients operated for CDR. Further research is needed to see if these scores are similar for other cohorts.

Acknowledgment

Supplementary materials

Supplementary material associated with this article can be found in the online version at https://doi.org/10.1016/j.spinee.2020.05.549.

References

[1] Rihn JA. Chapter 33 - anterior cervical discectomy and fusion. In: Frank RM, Forsythe B, Provencher MT, editors. Case competencies in orthopaedic surgery. Philadelphia, PA: Elsevier; 2017. p. 318–25.

[2] Dakwar E, Deukmedjian A, Ritter Y, Dain Allred C, Rechtine GR. Chapter 16 - spinal pathology, conditions, and deformities: surgical intervention. In: Magee DJ, Zachazewski JE, Quillen WS, Manske RC, editors. Pathology and intervention in musculoskeletal rehabilitation. Second ed. Maryland Heights, Missouri: W.B. Saunders; 2016. p. 584–611.

[3] Kepler CK, Wilkinson SM, Radcliff KE, Vaccaro AR, Anderson DG, Hilibrand AS, et al. Cost-utility analysis in spine care: a systematic review. Spine J 2012;12(8):676–90.

[4] Alvin MD, Lubelski D, Alam R, Williams SK, Obuchowski NA, Steinmetz MP, et al. Spine surgeon treatment variability: the impact on costs. Global Spine J 2018;8(5):498–506.

[5] Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. Qual Life Res 2018;27(5):1159–70.

[6] Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res 2010;19(4):539–49.

[7] Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials 1989;10(4):407–15.

[8] 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.

[9] Carreon LY, Glassman SD, Campbell MJ, Anderson PA. Neck Disability Index, short form-36 physical component summary, and pain scales for neck and arm pain: the minimum clinically important difference and substantial clinical benefit after cervical spine fusion. Spine J 2010;10(6):469–74.

[10] Steinhaus ME, Iyer S, Lovecchio F, Khechen B, Stein D, Ross T, et al. Minimal clinically important difference and substantial clinical benefit using PROMIS CAT in cervical spine surgery. Clin Spine Surg 2019;32(9):392–7.

[11] Solberg T, Olsen LR, Berglund ML. Nasjonalt kvalitetsregister for ryggkirugi (NKR) Årsrapport for 2018 medplan for forbedringstiltak. Available at: https://www.kvalitetsregistre.no/sites/default/files/28_arsrapport_2018_nkr_0.pdf. Accessed January 15, 2020.

[12] von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP, et al. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet 2007;370(9596):1453–7.

[13] MacDermid JC, Walton DM, Avery S, Blanchard A, Etruw E, McAlpine C, et al. Measurement properties of the neck disability index: a systematic review. J Orthop Sports Phys Ther 2009;39(5):400–17.

[14] Dolan P, Gudex C, Kind P, Williams A. The time trade-off method: results from a general population study. Health Economics 1996;5(2):141–54.

[15] Jensen MP, Karoly P. Self-report scales and procedures for assessing pain in adults. Handbook of pain assessment. 3rd ed. New York, NY, US: The Guilford Press; 2011. p. 19–44.

[16] Turk DC, Melzack R. Handbook of pain assessment. 3rd ed. New York: Guilford Press; 2011.

[17] 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.

[18] Werner DAT, Grotle M, Gulati S, Austevoll IM, Lonne G, Nygaard OP, 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.

[19] 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(1):31.

[20] Hermansen E, Myklebust TA, Austevoll IM, Rekeland F, Solberg T, Storheim K, et al. Clinical outcome after surgery for lumbar spinal stenosis in patients with insignificant lower extremity pain. A

prospective cohort study from the Norwegian registry for spine surgery. BMC Musculoskelet Disord 2019;20(1):36.

[21] Solberg T, Johnsen LG, Nygaard OP, Grotle M. Can we define success criteria for lumbar disc surgery?: Estimates for a substantial amount of improvement in core outcome measures. Acta Orthop 2013;84(2):196–201.

[22] Tape TG. Interpreting diagnostic tests: Available at: http://gim.unmc.edu/dxtests/Default.htm. Accessed Janyary 15, 2020.

[23] Dworkin RH, Turk DC, Wyrwich KW, Beaton D, Cleeland CS, Farrar JT, et al. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. J Pain 2008;9(2):105–21.

[24] Fekete TF, Haschtmann D, Kleinstuck FS, Porchet F, Jeszenszky D, Mannion AF. What level of pain are patients happy to live with after surgery for lumbar degenerative disorders? Spine J. 2016;16(4 Suppl.):S12–8.

[25] de Vet HC, Ostelo RW, Terwee CB, van der Roer N, Knol DL, Beckerman H, et al. Minimally important change determined by a visual method integrating an anchor-based and a distribution-based approach. Qual Life Res 2007;16(1):131–42.

[26] Foster MT, Carleton-Bland NP, Lee MK, Jackson R, Clark SR, Wilby MJ. Comparison of clinical outcomes in anterior cervical discectomy versus foraminotomy for brachialgia. Br J Neurosurg 2019;33:3–7.

[27] Liu WJ, Hu L, Chou PH, Wang JW, Kan WS. Comparison of anterior cervical discectomy and fusion versus posterior cervical foraminotomy in the treatment of cervical radiculopathy: a systematic review. Orthop Surg 2016;8(4):425–31.

[28] Sedaghat AR. Understanding the minimal clinically important difference (MCID) of patient-reported outcome measures. Otolaryngol Head Neck Surg 2019;161(4):551–60.

[29] Hung M, Saltzman CL, Kendall R, Bounsanga J, Voss MW, Lawrence B, et al. What are the MCIDs for PROMIS, NDI, and ODI instruments among patients with spinal conditions? Clin Orthop Relat Res 2018;476(10):2027–36.

[30] Parker SL, Godil SS, Shau DN, Mendenhall SK, McGirt MJ. Assessment of the minimum clinically important difference in pain, disability, and quality of life after anterior cervical discectomy and fusion: clinical article. J Neurosurg Spine 2013;18(2):154–60.

[31] Narain AS, Hijji FY, Khechen B, Haws BE, Patel DV, Bohl DD, et al. Risk factors associated with failure to reach minimal clinically important difference in patient-reported outcomes following anterior cervical discectomy and fusion. Int J Spine Surg 2019;13(3):262–9.

[32] Gatchel RJ, Mayer TG. Evidence-informed management of chronic low back pain with functional restoration. Spine J 2008;8(1):65–9.

[33] Mayer TG, Gatchel RJ, Mayer H, Kishino ND, Keeley J, Mooney V. A prospective two-year study of functional restoration in industrial low back injury. An objective assessment procedure. Jama 1987;258 (13):1763–7.

[34] Bobos P, MacDermid J, Nazari G, Furtado R. Psychometric properties of the global rating of change scales in patients with neck disorders: a systematic review with meta-analysis and meta-regression. BMJ Open 2019;9(11):e033909.

[35] Bobos P, Ziebart C, Furtado R, Lu Z, MacDermid JC. Psychometric properties of the global rating of change scales in patients with low back pain, upper and lower extremity disorders. A systematic review with meta-analysis. J Orthop 2020;21:40–8.

[36] Kamper SJ, Maher CG, Mackay G. Global rating of change scales: a review of strengths and weaknesses and considerations for design. J Man Manip Ther 2009;17(3):163–70.

[37] McLeod LD, Coon CD, Martin SA, Fehnel SE, Hays RD. Interpreting patient-reported outcome results: US FDA guidance and emerging methods. Expert Rev Pharmacoecon Outcomes Res 2011;11(2):163–9.

[38] van Hooff ML, Jacobs WC, Willems PC, Wouters MW, de Kleuver M, Peul WC, et al. Evidence and practice in spine registries. Acta Orthop 2015;86(5):534–44.

[39] Mannion AF, Elfering A. Predictors of surgical outcome and their assessment. Eur Spine J 2006;15(Suppl. 1):S93–108.

[40] Solberg TK, Sorlie A, Sjaavik K, Nygaard OP, Ingebrigtsen T. Would loss to follow-up bias the outcome evaluation of patients operated for degenerative disorders of the lumbar spine? Acta Orthop 2011;82 (1):56–63.

[41] Hojmark K, Stottrup 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.
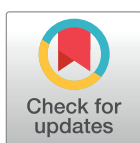
Paper II

PLOS ONE

Data Availability Statement: The data cannot be shared publicly since it contains potentially identifiable personal information about the

Clinical improvement after surgery for degenerative cervical myelopathy; A comparison of Patient-Reported Outcome Measures during 12-month follow-up

Christer Mjåset[1,2,3]*, John-Anker Zwart[1,3], Frode Kolstad[2], Tore Solberg[4,5,6], Margreth Grotle[3,7]

1 Faculty of Medicine, University of Oslo, Oslo, Norway, 2 Department of Neurosurgery, Oslo University Hospital, Oslo, Norway, 3 Department of Research and Innovation, Division of Clinical Neuroscience, Oslo University Hospital, Oslo, Norway, 4 Institute of Clinical Medicine, The Arctic University of Norway, Tromsø, Norway, 5 Department of Neurosurgery, The University Hospital of North Norway, Tromsø, Norway, 6 The Norwegian Registry for Spine Surgery, The University Hospital of North Norway, Tromsø, Norway, 7 Department of Physiotherapy, Faculty of Health Sciences, Oslo Metropolitan University, Oslo, Norway

* chrmja@gmail.com

Abstract

Object

Although many patients report clinical improvement after surgery due to degenerative cervical myelopathy, the aim of intervention is to stop progression of spinal cord dysfunction. We wanted to provide estimates and assess achievement rates of Minimal Clinically Important Difference (MCID) at 3- and 12-month follow-up for Neck Disability Index (NDI), Numeric Rating Scale for arm pain (NRS-AP) and neck pain (NRS-NP), Euro-Qol (EQ-5D-3L), and European Myelopathy Score (EMS).

Methods

614 degenerative cervical myelopathy patients undergoing surgery responded to Patient-Reported Outcome Measures (PROMs) prior to, 3 and 12 months after surgery. External criterion was the Global Perceived Effect Scale (1–7), defining MCID as "slightly better", "much better" and "completely recovered". MCID estimates with highest sensitivity and specificity were calculated by Receiver Operating Curves for change and percentage change scores in the whole sample and in anterior and posterior procedural groups.

Results

The NDI and NRS-NP percentage change scores were the most accurate PROMs with a MCID of 16%. The change score for NDI and percentage change scores for NDI, NRS-AP and NRS-NP were slightly higher in the anterior procedure group compared to the posterior procedure group, while remaining PROM estimates were similar across procedure type. The MCID achievement rates at 12-month follow-up ranged from 51% in EMS to 62% in NRS-NP.

Conclusion

The NDI and NRS-NP percentage change scores were the most accurate PROMs to measure clinical improvement after surgery for degenerative cervical myelopathy. We recommend using different cut-off estimates for anterior and posterior approach procedures. A MCID achievement rate of 60% or less must be interpreted in the perspective that the main goal of surgery for degenerative cervical myelopathy is to prevent worsening of the condition.

Introduction

Degenerative cervical myelopathy (DCM) describes a range of conditions in the cervical spine causing cord compression and neurological dysfunction [1]. There is current lack of evidence for nonoperative management in terms of preventing neurological deterioration, although physical rehabilitation and close observation can be considered in mild to asymptomatic cases. For moderate to severe cases, individualized surgical treatment is recommended [2–4]. Anterior Cervical Discectomy and Fusion (ACDF) and Anterior Cervical Disc Arthroplasty (ACDA) are frequently used in patients with disc herniation, while posterior approach procedures are well-established treatments options for patients with posterior and/or multi-level spinal cord compression [5]. In cases where symptoms are caused by spinal cord compression due to cervical ossification of the posterior longitudinal ligament, no treatment consensus is obtained and various anterior and posterior approach procedures are currently applied [6, 7].

The aim of surgery has traditionally been to stop progression of spinal cord dysfunction symptoms. However, recent studies have shown that many patients report improvement post intervention both regarding functionality and disability, as well as quality-of-life outcomes [2, 8]. Depending on PROMs used, severity of preoperative disease and length of follow-up, improvement rates range from around 20 to 80% [9, 10].

Patient-Reported Outcome Measures (PROMs) are commonly used to measure clinical improvement or worsening in spine literature. In combination with the concept of Minimal Clinically Important Difference (MCID), defined as the smallest change in an outcome score that is clinically beneficial within a patient group [11], optimal cut-off estimates for an individual PROM can be assessed [12, 13]. The traditional method is to assess the MCID change score, or the delta value. However, since the interpretation of a change score is dependent on the baseline score, the percentage change score can provide a more representative result at group level [14]. To date, MCID estimates for PROM percentage change scores have not been reported for DCM patients undergoing surgery. Further, there is current lack of evidence in terms of which PROMs are the more accurate in capturing changes in health status among these patients and whether results differ across surgical approach.

The purpose of this study was to estimate MCID for frequently used PROMs 3 and 12 months after surgery for DCM; NDI, Numeric Rating Scale for arm pain (NRS-AP) and neck pain (NRS-NP), Euro-Qol (EQ-5D-3L), and European Myelopathy Score (EMS). A secondary aim was to report achievement rates of MCID through 12 months of follow-up. The MCID estimates are reported for change scores and percentage changes scores for the whole sample, as well as for anterior and posterior approach procedural groups.

Materials and methods

Data collection

All data were collected through the Norwegian Registry for Spine Surgery (NORspine) which is a government funded comprehensive clinical registry. Participation in NORspine is not required for a patient to gain access to the health care, or for payment/reimbursement to a provider. All Norwegian health care providers offering cervical spine surgery (six public hospitals and three private clinics) report to NORspine. The proportion of operated patients reported to the registry was 75–78% over the study period [15].

Our research protocol was approved by the Norwegian Committee for Medical and Health Research Ethics Midt (2014/344). Informed consent was obtained from all patients before entering the registry.

Design

This is a multicenter observational study with follow-up at 3 and 12 months. Results are reported consistent with the Strengthening The Reporting of Observational Studies in Epidemiology (STROBE) statement [16], and methods are in accordance with the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) recommendations [12].

Eligibility criteria

A cohort of 614 patients undergoing surgery for DCM between January 2011 and August 2016 were found to be eligible (Fig 1). Exclusion criteria were: 1) prior surgery the index level; and 2) patients undergoing combined anterior and posterior approach, since these patients commonly are selected on a case-by-case basis [17]. Of the 614 patients, 371 underwent either ACDF (363, 98%) or ACDA (8, 2%), and 243 patients underwent posterior approach procedures, such as laminectomy with or without fusion, hemilaminectomy or laminoplasty.

Measurements

At admission for surgery (baseline), patients complete the NORspine questionnaire which cover demographics, location and extent of pain and PROMs. During the hospital stay, the surgeon records data concerning diagnosis, American Society of Anesthesiologists physical status (ASA), surgical treatment and comorbidity on a separate form. Under 'indication for operation' the surgeon can checkmark if he/she considers the patient to have myelopathy based on clinical assessment and radiological findings. To avoid selective reporting, the 3- and 12-month follow-up is conducted by the NORspine central registry unit without involvement from treating hospitals. After surgery, a questionnaire identical to that used at baseline is distributed by mail to every registered patient. One reminder questionnaire is sent to those who do not respond. The following PROMs are collected:

1. Neck Disability Index (NDI): a patient-completed questionnaire focusing on the patient's functional status and scores ranging from 0 (no disability) to 100 (greatest disability) [18].

2. Numeric Rating Scale for arm (NRS-AP) and neck pain (NRS-NP): a scale that assesses pain level ranging from 0 (no pain) to 10 (worst conceivable pain) [19].

3. EuroQoL (EQ-5D-3L): a generic measure assessing health-related quality of life with scores ranging from -0.59 (worse than death) to 1 (perfect health) [20].
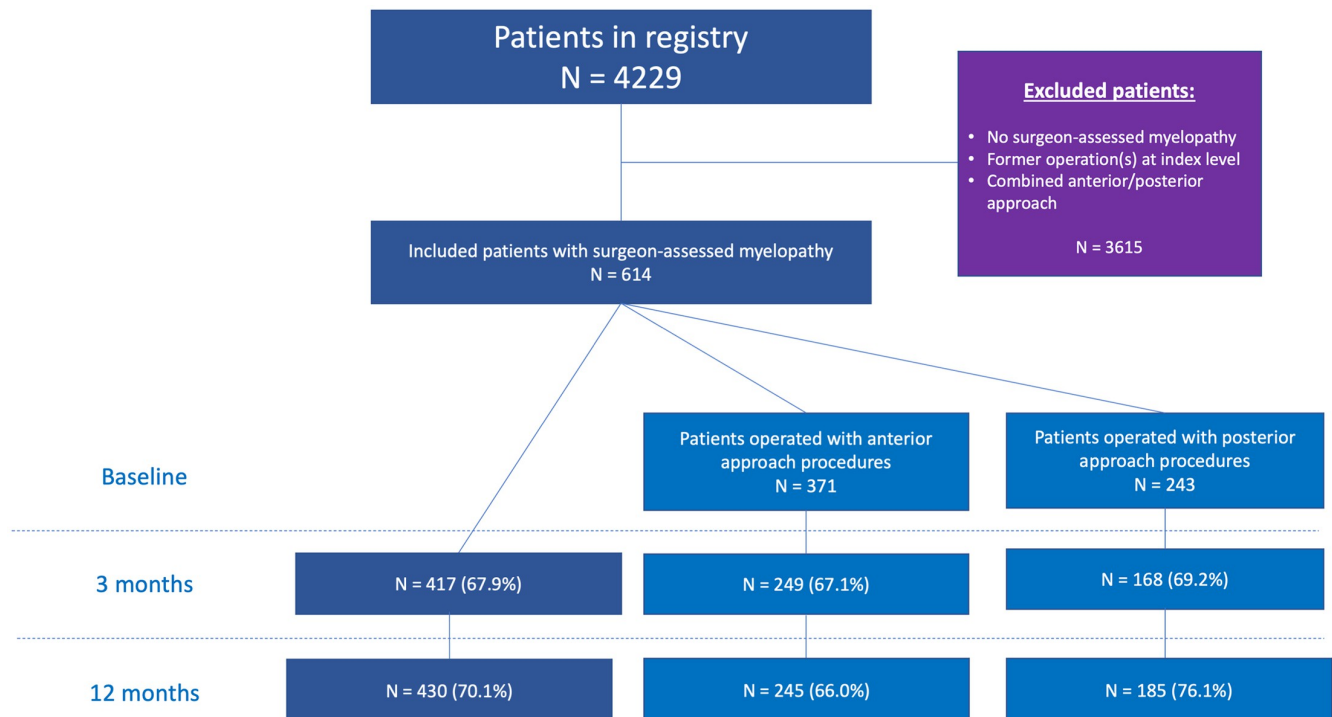
Fig 1. Project flow chart. Exclusion criteria for patients included in the study with follow-up rates.

https://doi.org/10.1371/journal.pone.0264954.g001

4. European Myelopathy Score (EMS): a patient-based questionnaire derived for assessing spinal cord function. Scoring is between 5 (severe deficit) and 18 (no symptoms) [21].

The Global Perceived Effect scale (GPE) was in the present study used as an external criterion for defining MCID. The GPE measures patient-reported treatment outcome through one single question and seven answer choices; "completely recovered", "much improved", "slightly improved", "unchanged, "slightly worse", "much worse" and "worse than ever" [22]. Patients reporting to be "completely recovered", "much improved" or "slightly improved" (1–3) were classified as having achieved a MCID. Those who considered themselves to be "unchanged" or worse (4–7) were classified as not improved.

Statistics

All statistical analyses were performed with the Statistical Package for the Social Sciences (SPSS, version 26). Continuous variables were reported as means and standard deviations and categorical variables as numbers and percentages. Differences were evaluated by Chi-square test for categorical variables and by t-tests for continuous variables. PROM change scores were obtained by subtracting the follow-up score from the baseline score. The percentage change score was calculated by dividing the change score with the baseline score and multiplying by 100. To be able to calculate the EQ-5D-3L percentage change score we converted the value range from -0.6 to 1.0 into a relative score from 0 to 100.

The correlations between the GPE scale and the different PROMs were analyzed using the Spearman correlation coefficient. Receiver Operating Curves (ROCs) were used to assess

discriminative ability of each PROM and to define the optimal cut-off with the highest sensitivity and specificity. ROCs were made by plotting the sensitivity against (1 –specificity) for each possible MCID cut-off estimate. The sensitivity refers to the probability of correctly classifying an individual replying "slightly improved" or better (1–3) according to the PROM score. Correspondingly, the specificity refers to the probability of correctly classifying a patient reporting to be "unchanged" or worse as having "not improved" after surgery (4–7). The area under the ROC (AUC) with 95% confidence interval (CI) describes the test's accuracy of correctly classifying a case according to the anchor. The AUC is classified as "acceptable" from 0.7 to 0.8, "excellent" from 0.8 to 0.9 and "outstanding" from 0.9 to 1.0 [23]. To determine MCID cut-off estimates with highest sensitivity and specificity, the closest point to the upper left corner of the ROC-curve was calculated from the coordinates of the curve. Cut-off estimates were assessed for the whole DCM group and for both procedural groups. Lastly, proportions of patients achieving MCID for the whole group and both procedural groups were calculated using the cut-off estimates for each PROM.

Results

Respondents and baseline characteristics

Of 4229 consecutive patients undergoing surgery for degenerative disorders in the cervical spine between January 2011 and August 2016, 614 patients were included. Of these patients, 371 underwent an anterior approach procedure, while 243 underwent a posterior approach procedure. A total of 67.9% and 70.1% of patients responded to the 3- and 12-month follow-up questionnaire, respectively (Fig 1). The non-responding patients were slightly younger (p<0.001), less likely to be retired (p<0.001), and more likely to smoke (p<0.001) (Table 1). There were no statistically significant differences in PROM scores, except for the EQ-5D-3L

Table 1. Baseline characteristics of respondents and non-respondents to 12-month follow-up.

	Respondents N = 430		Non-respondents N = 184		Sig. (2-tailed)/ chi-square
	N		N		
Age (years); Mean (SD)	430	59.1 (11.9)	184	53.5 (12.2)	<0.001
Female, no (%)	430	167 (38.8)	184	66 (35.9)	0.488
ASA level (1–4); Mean (SD)	430	2.4 (1.7)	184	2.3 (1.5)	0.414
Body Mass Index; Mean (SD)	417	27.0 (4.5)	179	27.5 (5.2)	0.220
Smokers, no (%)	428	106 (24.8)	184	84 (45.7)	<0.001
University/College education	402	137 (31.6)	173	56 (32.4)	0.823
Retired, no (%)	430	121 (28.1)	183	23 (12.6)	<0.001
Comorbidity, no (%)	422	227 (53.8)	183	109 (59.6)	0.189
Levels operated, Mean (SD)	418	1.9 (1.1)	184	1.85 (1.1)	0.376
NDI; Mean (SD)	428	33.7 (17.3)	178	36.6 (17.4)	0.060
NRS-AP; Mean (SD)	399	5.0 (2.9)	164	5.1 (3.0)	0.794
NRS-NP; Mean (SD)	401	4.7 (3.0)	162	5.1 (2.9)	0.134
EQ-5D-3L; Mean (SD)	392	0.47 (0.32)	171	0.39 (0.34)	0.008
EMS; Mean (SD)	384	14.5 (2.3)	165	14.4 (2.5)	0.750

SD, Standard Deviation; NDI, Neck Disability Index (0–100); NRS-AP Numeric Rating Scale for arm pain (0–10); NRS-NP, Numeric Rating Scale for neck pain (0–10); EQ-5D-3L, Health-Related Quality-of-Life by EuroQol (-0.4– 1.0); EMS, European Myelopathy Score (5–18).

https://doi.org/10.1371/journal.pone.0264954.t001

Table 2. Baseline characteristics of the whole myelopathy group and of the two procedural groups.

	Whole myelopathy group N = 430		Anterior approach group N = 245		Posterior approach group N = 185		Sig. (2-tailed)/ chi-square
	N		N		N		
Age (years); Mean (SD)	430	59.1 (11.9)	245	53.7 (11.0)	185	66.1 (8.9)	<0.001
Female; no (%)	430	167 (38.8)	245	108 (44.1)	185	59 (31.9)	0.01
ASA level (1–4); Mean (SD)	430	2.4 (1.7)	245	2.0 (1.4)	185	2.9 (1.9)	<0.001
Body Mass Index; Mean (SD)	417	27.0 (5.0)	363	27.3 (4.4)	178	26.8 (5.1)	0.260
Smokers; no (%)	425	106 (24.9)	243	62 (25.5)	182	44 (24.2)	0.752
No of levels operated; Mean (SD)	418	1.9 (1.1)	241	1.4 (0.6)	177	2.7 (1.2)	<0.001
Comorbidity; no (%)	422	227 (53.8)	238	110 (46.2)	184	117 (63.6)	<0.001
Currently working; no (%)	430	110 (25.9)	240	85 (35.4)	184	25 (13.6)	<0.001
Retired; no (%)	430	121 (28.1)	245	34 (13.9)	185	87 (47.0)	<0.001
NDI; Mean (SD)	428	33.7 (17.3)	244	33.9 (16.9)	184	33.4 (18.0)	0.753
NRS-AP; Mean (SD)	399	5.0 (2.9)	232	5.1 (3.0)	167	4.9 (2.9)	0.442
NRS-NP; Mean (SD)	401	4.7 (3.0)	234	4.9 (2.9)	167	4.4 (3.1)	0.062
EQ-5D-3L; Mean (SD)	392	0.47 (0.32)	225	0.49 (0.30)	167	0.44 (0.33)	0.084
EMS; Mean (SD)	427	14.5 (2.4)	243	14.9 (2.2)	184	13.9 (2.5)	<0.001

SD, Standard Deviation; NDI, Neck Disability Index (0–100); NRS-AP Numeric Rating Scale for arm pain (0–10); NRS-NP, Numeric Rating Scale for neck pain (0–10); EQ-5D-3L, Health-Related Quality-of-Life by EuroQol (-0.4–1.0); EMS, European Myelopathy Score (5–18).

https://doi.org/10.1371/journal.pone.0264954.t002

mean, which was lower (poorer health-related quality-of-life) among non-responders (p<0.008) (Table 1).

Baseline characteristics of the whole myelopathy group and the two procedural groups are presented in Table 2. Compared to the anterior approach procedure group, patients in the posterior approach group were more likely to be male, not working, and to be operated at a higher number of levels. Also, they had significantly higher mean age, higher mean ASA level, more comorbidity, and more severe myelopathy symptoms according to EMS.

Correlation between the PROMs and the external criterion

For all PROMs, there was a stepwise decrease in mean change scores and mean percentage change scores at 12 months for patients who reported themselves to be completely recovered, much better and slightly better compared to those reporting no change or some degree of worsening (S1 Table). A similar pattern was found for results at 3 months (obtained on request). For the whole group, the Spearman correlation coefficients ranged from 0.30 to 0.59. The NDI showed the strongest correlation with the external anchor.

AUC and MCID

We found minor differences in AUC and MCID cut-off estimates at 3 and 12 months. Therefore, further analysis of the data is presented only for the PROMs at 12-month follow-up. 3-month scores are presented in S2 Table.

The change scores of NDI, NRS-NP and the EQ-5D-3L showed acceptable AUC values (>0.70), whereas AUC values of the NRS-AP change score and EMS percentage change score were slightly lower than acceptable (0.69 and 0.68, respectively) (Table 3). Most of the AUC change score values (0.64–0.74) were similar to or lower than the corresponding AUC percentage change score value (0.68–0.77). Only for EMS, the change score AUC (0.69) was higher

Table 3. Area under the curve and cut-off estimates for Minimal Clinically Important Difference for all Patient-Reported Outcome Measures at 12 months.

		Change score (points)	Percentage change score (%)
NDI	**AUC (95% CI)**	0.74 (0.69, 0.79)	0.77 (0.72, 0.81)
	Cut-off (% sensitivity, % specificity)	4.3 (0.68, 0.68)	15.7 (0.71, 0.71)
NRS-AP	**AUC (95% CI)**	0.64 (0.58, 0.70)	0.69 (0.63, 0.75)
	Cut-off (% sensitivity, % specificity)	0.5 (0.66, 0.53)	23.6 (0.63, 0.61)
NRS-NP	**AUC (95% CI)**	0.73 (0.67, 0.78)	0.76 (0.70, 0.81)
	Cut-off (% sensitivity, % specificity)	0.5 (0.71, 0.64)	15.5 (0.72, 0.71)
EQ-5D-3L	**AUC (95% CI)**	0.70 (0.64, 0.77)	0.70 (0.64, 0.77)
	Cut-off (% sensitivity, % specificity)	0.02 (0.70, 0.66)	2.2 (0.68, 0.66)
EMS	**AUC (95% CI)**	0.69 (0.63, 0.75)	0.68 (0.61, 0.74)
	Cut-off (% sensitivity, % specificity)	0.5 (0.58, 0.69)	4.2 (0.58, 0.69)

NDI, Neck Disability Index (0–100); AUC, Area Under the Curve, NRS-AP, Numeric Rating Scale for arm pain (0–10); NRS-NP, Numeric Rating Scale for neck pain (0–10); EQ-5D-3L, Health-Related Quality-of-Life by EuroQol (-0.4–1.0); EMS, European Myelopathy Score (5–18).

than the percentage change score AUC (0.68) (Table 3). The percentage change scores of the NDI and NRS-NP had the highest sensitivity and specificity.

Similar results were found for AUCs analyzed for the anterior and posterior approach groups. However, there was a tendency to lower discriminative ability for all PROMs in the posterior approach group except for EMS in which case the AUCs were higher in this group (Table 4).

Proportions of patients with clinical improvement at 12-month follow-up

In Fig 2, we present the proportions of patients that achieved a clinical improvement according to MCID estimates for percentage change scores at 12-month follow-up. Overall, NDI (59%),

Table 4. Minimal Clinically Important Difference cut-off estimates for all Patient-Reported Outcome Measures in the two procedural subgroups at 12 months.

		Anterior approach (% sensitivity, % specificity)	AUC (95% Confidence Interval)	Posterior approach (% sensitivity, % specificity)	AUC (95% Confidence Interval)
NDI	Change score (points)	5.9 (0.70, 0.70)	0.74 (0.67, 0.81)	2.4 (0.68, 0.68)	0.73 (0.66, 0.81)
	Percentage change score (%)	16.2 (0.72, 0.71)	0.77 (0.71, 0.84)	14.4 (0.71, 0.71)	0.76 (0.68, 0.83)
NRS-AP	Change score (points)	0.5 (0.66, 0.52)	0.66 (0.58, 0.74)	0.5 (0.65, 0.54)	0.62 (0.52, 0.72)
	Percentage change score (%)	23.6 (0.64, 0.59)	0.69 (0.62, 0.77)	21.1 (0.62, 0.61)	0.69 (0.60, 0.77)
NRS-NP	Change score (points)	0.5 (0.76, 0.62)	0.77 (0.69, 0.84)	0.5 (0.63, 0.66)	0.66 (0.58, 0.75)
	Percentage change score (%)	18.3 (0.73, 0.73)	0.77 (0.69, 0.85)	11.8 (0.69, 0.69)	0.73 (0.65, 0.81)
EQ-5D-3L	Change score (points)	0.02 (0.72, 0.71)	0.74 (0.66, 0.82)	0.02 (0.67, 0.61)	0.66 (0.57, 0.76)
	Percentage change score (%)	2.2 (0.70, 0.71)	0.74 (0.66, 0.82)	2.3 (0.63, 0.61)	0.66 (0.57, 0.76)
EMS	Change score (points)	0.5 (0.58, 0.66)	0.67 (0.58, 0.76)	0.5 (0.59, 0.72)	0.72 (0.63, 0.80)
	Percentage change score (%)	4.2 (0.58, 0.66)	0.65 (0.55, 0.74)	4.2 (0.59, 0.72)	0.71 (0.62, 0.81)

AUC, Area Under the Curve; NDI, Neck Disability Index (0–100); NRS-AP, Numeric Rating Scale for arm pain (0–10); NRS-NP, Numeric Rating Scale for neck pain (0–10); EQ-5D-3L, Health-Related Quality-of-Life by EuroQol (-0.4–1.0); EMS, European Myelopathy Score (5–18).
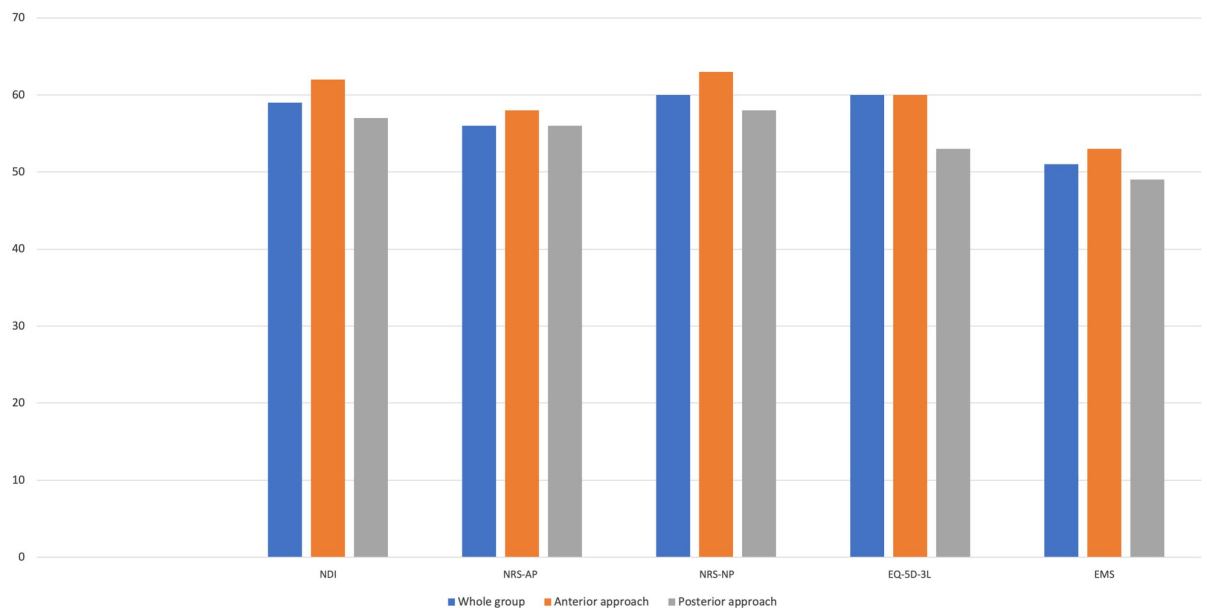
Fig 2. Clinical improvement rates. Percentage of patients achieving improvement larger than the Minimal Clinically Important Difference (MCID) according to the Neck Disability Index (NDI), Numeric Rating Scale for arm pain (NRS-AP) and neck pain (NRS-NP), Euro-Quol-5D-3L (EQ-5D-3L) and European Myelopathy Score (EMS). Results are provided by the percentage change score from baseline to 12-month follow-up.

NRS-NP (61%) and EQ-5D-3L (59%) showed similar proportions of achieving a MCID, whereas NRS-NP (56%) and, in particular, EMS (51%) showed lower proportions of improvement. The rates were slightly higher for the anterior approach group compared to the posterior approach group for both change score and percentage change score (S3 Table).

Discussion

This study showed that NDI and NRS-NP were the most accurate PROMs to measure MCID among patients undergoing surgery due to DCM. EQ-5D-3L also showed acceptable accuracy for both change and percentage change score. Further, achievement of clinical improvement according to the optimal MCID estimates of the investigated PROMs ranged from 51% to 62%, depending on type of PROM, type of MCID and surgical approach.

Although there are several studies investigating MCID for DCM patients undergoing surgery [24–28], there are no reports of percentage change scores for this patient group. In our study, the majority of the percentage change scores were more accurate than the change scores. As shown in Table 3, percentage change scores for NDI, NRS-AP and NRS-NP showed higher AUC, including higher sensitivity and specificity, compared to the change scores. For EQ-5D, the AUCs were identical, while the EMS AUC was slightly higher for the change score than for the percentage change score (0.69 vs. 0.68). Since the use of change scores for benchmarking has been criticized for not taking into account the baseline score [29–31], we recommend using percentage change scores in future research.

The observed MCID estimate of 4.3 points for the NDI 12-month change score is similar to a previous study of Kato et al., who found a cut-off estimate of 4.2 in 101 myelopathy patients undergoing cervical laminoplasty [32]. Chien et al. report a slightly higher cut-off of 6 for NDI which might be due to a very small patient sample (n = 45) [26]. Similarly, in a study of 30

DCM patients by Auffinger et al., five statistical methods used for calculation of cut-off estimates showed similar or substantially higher findings for both NDI (4.8–13.4) and NRS-NP (0.36–3.11) [25].

The accuracy of EQ-5D-3L has also been assessed in a previous study. Kato et al. reported a MCID estimate of 0.05 for EQ-5D-3L with an AUC of 0.704 [32], which is in accordance with the results in the present study. Since the accuracy of EQ-5D-3L has been found to be acceptable (>0.70) in both these studies, we recommend further use of this PROM for DCM patients.

Several studies have reported MCID estimates for degenerative neck surgery patients. However, in many of the investigated cohorts there have been a mix of radiculopathy and myelopathy patients [33–35]. We argue that it is necessary to distinguish between myelopathy and radiculopathy patient cohorts considering the smaller amount of expected improvement among DCM patients. For example, Carreon et al., who analyzed a mixed sample of 505 patients, reported higher MCID estimates than our study for both NDI (7.5 vs. 4.3), NRS-AP (2.5 vs. 0.5) and NRS-NP (2.5 vs. 0.5) [34].

As far as we know, no previous study has presented MCID estimates for EMS and NRS-AP in a DCM cohort.

Surgical approach

We found minor differences in accuracy of NDI and NRS-NP across patients undergoing anterior versus posterior surgical procedures. However, there was a tendency to lower discriminative ability for NDI, the two NRS scores and EQ-5D-3L in the posterior approach group (Table 4). In each group, NDI and NRS-NP showed the best discriminative ability.

The MCID estimates for NDI, NRS-AP and NRS-NP were lower in the posterior approach group compared to the anterior approach group. This may indicate that posterior approach patients, which were older and had multilevel degenerative disease, were satisfied with less improvement compared to the younger and healthier patients in the anterior approach group. These results confirm that it is reasonable to analyze these two surgical groups separately. They also suggest that the interpretation of change and percentage change scores of PROMs should be different across anterior and posterior procedures.

Proportion of patients achieving MCID

The proportion of DCM patients that achieved MCID varied between 51% and 61% for the percentage change score. This is in line with a previous study by Stull et al. which reported that 40 to 61% achieve MCID in a sample of 53 DCM patients [9]. Although Stull et al. found little or no difference in achievement rates between radiculopathy and myelopathy patients, others have shown that the proportion of patients achieving a MCID is substantially higher among radiculopathy patients. Applying a cut-off estimate of 15 for NDI, two recent studies found NDI success rates of 80–92% for patients undergoing ACDF or ACDA [36, 37].

Limitations and strengths

GPE is a self-reported scale and not an objective anchor. This represents the main limitation of our study as global scale ratings tend to be influenced by the current health status of the patient [22]. However, no alternative gold standard currently exists, and the GPE is still the most frequently used anchor in scientific literature [38–42].

The main inclusion criterion for all patients was that the operating surgeon had made a checkmark for myelopathy (yes/no) in the post-operative questionnaire under "indication for operation". This response represents a subjective judgement based on patient history, clinical

features, and radiological findings. Since we have no objective reference for evaluating the accuracy of the surgeons' judgment, misclassifications could exist.

The non-respondent rate of approximately 30% is usually regarded as acceptable for a spine registry [43]. As some of the baseline characteristics of the non-respondents have been associated with poorer outcomes [44], this might still be considered a selection bias especially since we are estimating the proportion of patients achieving MCID. However, this should be of less importance when assessing actual cut-off estimates across a wide range of outcomes. Two previous studies found no differences in outcome when comparing respondents and non-respondents at follow-up, though both had slightly lower non-respondent rates [45, 46].

A major strength of this study is the large sample size of surgical patients in daily clinical practice and the high coverage rate [15] indicating a high external validity of our results.

Conclusion

NDI and NRS-NP were the most accurate PROMs to measure a clinical improvement according to MCID estimates 12 months after surgery for DCM. Also, EQ-5D-3L showed acceptable discriminative ability.

Percentage change scores were more accurate than change scores, hence, we recommend using percentage change cut-off estimates in future studies. The cut-off estimates and MCID achievement rates were also slightly higher for the anterior approach group compared to the posterior approach group indicating that separate cut-off estimates should be used for each surgical approach.

An achievement of a MCID of 60% or less among DCM patients must be interpreted in the perspective that the main goal of surgery is to prevent worsening of the condition.

Supporting information

S1 Table. Mean scores with standard deviation of the Patient-Reported Outcome Measures at 12 months for the whole myelopathy group according to the Global Perceived Effect Scale. Spearman, Spearman's rank correlation coefficient; Neck Disability Index (0–100); SD, Standard Deviation; NRS-AP, Numeric Rating Scale for arm pain (0–10); NRS-NP, Numeric Rating Scale for neck pain (0–10), EQ-5D-3L, Health-Related Quality-of-Life by EuroQol (-0.4–1.0), EMS, European Myelopathy Score (5–18).
(DOCX)

S2 Table. Area under the curve and cut-off estimates for "Minimal Clinically Important Difference" for all Patient-Reported Outcome Measures at 3 months. NDI, Neck Disability Index (0–100); AUC, Area Under the Curve, NRS-AP, Numeric Rating Scale for arm pain (0–10), NRS-NP, Numeric Rating Scale for neck pain (0–10), EQ-5D-3L, Health-Related Quality-of-Life by EuroQol (-0.4–1.0); EMS, European Myelopathy Score (5–18).
(DOCX)

S3 Table. Proportion of patients with an improvement larger than "Minimal Clinically Important Difference" at 12-months follow-up according to Patient-Reported Outcome Measures. NDI, Neck Disability Index (0–100), NRS-AP, Numeric Rating Scale for arm pain (0–10), NRS-NP, Numeric Rating Scale for neck pain (0–10), EQ-5D-3L, Health-Related Quality-of-Life by EuroQol (-0.4–1.0), EMS, European Myelopathy Score (5–18).
(DOCX)

Acknowledgments

Author Contributions

Conceptualization: Christer Mjåset, John-Anker Zwart, Frode Kolstad, Tore Solberg, Margreth Grotle.

Data curation: Christer Mjåset.

Formal analysis: Margreth Grotle.

Methodology: Tore Solberg.

Project administration: Christer Mjåset.

Supervision: John-Anker Zwart, Frode Kolstad, Margreth Grotle.

Validation: John-Anker Zwart, Frode Kolstad, Tore Solberg.

Visualization: Christer Mjåset.

Writing – original draft: Christer Mjåset.

Writing – review & editing: Christer Mjåset, John-Anker Zwart, Frode Kolstad, Tore Solberg, Margreth Grotle.

References

1. Nouri A, Tetreault L, Singh A, Karadimas SK, Fehlings MG. Degenerative Cervical Myelopathy: Epidemiology, Genetics, and Pathogenesis. Spine (Phila Pa 1976). 2015; 40(12):E675–93.

2. Fehlings MG, Wilson JR, Kopjar B, Yoon ST, Arnold PM, Massicotte EM, et al. Efficacy and safety of surgical decompression in patients with cervical spondylotic myelopathy: results of the AOSpine North America prospective multi-center study. J Bone Joint Surg Am. 2013; 95(18):1651–8. https://doi.org/10.2106/JBJS.L.00589 PMID: 24048552

3. Rhee JM, Shamji MF, Erwin WM, Bransford RJ, Yoon ST, Smith JS, et al. Nonoperative management of cervical myelopathy: a systematic review. Spine (Phila Pa 1976). 2013; 38(22 Suppl 1):S55–67. https://doi.org/10.1097/BRS.0b013e3182a7f41d PMID: 23963006

4. Fehlings MG, Barry S, Kopjar B, Yoon ST, Arnold P, Massicotte EM, et al. Anterior versus posterior surgical approaches to treat cervical spondylotic myelopathy: outcomes of the prospective multicenter AOSpine North America CSM study in 264 patients. Spine (Phila Pa 1976). 2013; 38(26):2247–52. https://doi.org/10.1097/BRS.0000000000000047 PMID: 24108289

5. Lawrence BD, Jacobs WB, Norvell DC, Hermsmeyer JT, Chapman JR, Brodke DS. Anterior versus posterior approach for treatment of cervical spondylotic myelopathy: a systematic review. Spine (Phila Pa 1976). 2013; 38(22 Suppl 1):S173–82. https://doi.org/10.1097/BRS.0b013e3182a7eaaf PMID: 23962995

6. Yoshii T, Egawa S, Hirai T, Kaito T, Mori K, Koda M, et al. A systematic review and meta-analysis comparing anterior decompression with fusion and posterior laminoplasty for cervical ossification of the posterior longitudinal ligament. J Orthop Sci. 2020; 25(1):58–65. https://doi.org/10.1016/j.jos.2019.03.004 PMID: 30905611

7. Nakashima H, Tetreault L, Nagoshi N, Nouri A, Arnold P, Yukawa Y, et al. Comparison of Outcomes of Surgical Treatment for Ossification of the Posterior Longitudinal Ligament Versus Other Forms of Degenerative Cervical Myelopathy: Results from the Prospective, Multicenter AOSpine CSM-International Study of 479 Patients. J Bone Joint Surg Am. 2016; 98(5):370–8. https://doi.org/10.2106/JBJS.O.00397 PMID: 26935459

8. Fehlings MG, Ibrahim A, Tetreault L, Albanese V, Alvarado M, Arnold P, et al. A global perspective on the outcomes of surgical decompression in patients with cervical spondylotic myelopathy: results from the prospective multicenter AOSpine international study on 479 patients. Spine (Phila Pa 1976). 2015; 40(17):1322–8.

9. Stull JD, Goyal DKC, Mangan JJ, Divi SN, McKenzie JC, Casper DS, et al. The Outcomes of Patients With Neck Pain Following ACDF: A Comparison of Patients With Radiculopathy, Myelopathy, or Mixed Symptomatology. Spine (Phila Pa 1976). 2020; 45(21):1485–90. https://doi.org/10.1097/BRS.0000000000003613 PMID: 32796460

10. Goh GS, Liow MHL, Ling ZM, Soh RCC, Guo CM, Yue WM, et al. Severity of Preoperative Myelopathy Symptoms Affects Patient-reported Outcomes, Satisfaction, and Return to Work After Anterior Cervical Discectomy and Fusion for Degenerative Cervical Myelopathy. Spine (Phila Pa 1976). 2020; 45 (10):649–56.

11. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: Ascertaining the minimal clinically important difference. Controlled Clinical Trials. 1989; 10(4):407–15. https://doi.org/10.1016/0197-2456(89)90005-6 PMID: 2691207

12. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res. 2010; 19(4):539–49. https://doi.org/10.1007/s11136-010-9606-8 PMID: 20169472

13. Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. Qual Life Res. 2018; 27(5):1159–70. https://doi.org/10.1007/s11136-018-1829-0 PMID: 29550964

14. Kaiser L. Adjusting for baseline: change or percentage change? Stat Med. 1989; 8(10):1183–90. https://doi.org/10.1002/sim.4780081002 PMID: 2682909

15. Solberg T, Olsen L, Berglund M. Årsrapport for 2018 med plan for forbedringstiltak. Tromsø, Norway: Nasjonalt kvalitetsregister for ryggkirugi (NKR); 2019.

16. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet. 2007; 370(9596):1453–7. https://doi.org/10.1016/S0140-6736(07)61602-X PMID: 18064739

17. Kim PK, Alexander JT. Indications for circumferential surgery for cervical spondylotic myelopathy. Spine J. 2006; 6(6 Suppl):299s–307s. https://doi.org/10.1016/j.spinee.2006.04.025 PMID: 17097550

18. MacDermid JC, Walton DM, Avery S, Blanchard A, Etruw E, McAlpine C, et al. Measurement properties of the neck disability index: a systematic review. J Orthop Sports Phys Ther. 2009; 39(5):400–17. https://doi.org/10.2519/jospt.2009.2930 PMID: 19521015

19. Jensen MP, Karoly P. Self-report scales and procedures for assessing pain in adults. Handbook of pain assessment, 3rd ed. New York, NY, US: The Guilford Press; 2011. pp. 19–44.

20. 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. PMID: 8733106

21. Herdmann J, Linzbach M, Krzan M, Dvořák J, Bock W. The European Myelopathy Score. In: Bauer B, Brock M, Klinger M, editors. Advances in Neurosurgery. 22. Berlin, Heidelberg: Springer; 1994. pp. 266–8.

22. 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 PMID: 20056385

23. Hosmer DW, Lemeshow S. Applied Logistic Regression. 2nd ed. NY: John Wiley and Sons; 2000.

24. Badhiwala JH, Witiw CD, Nassiri F, Akbar MA, Jaja B, Wilson JR, et al. Minimum Clinically Important Difference in SF-36 Scores for Use in Degenerative Cervical Myelopathy. Spine (Phila Pa 1976). 2018; 43(21):E1260–E6.

25. Auffinger BM, Lall RR, Dahdaleh NS, Wong AP, Lam SK, Koski T, et al. Measuring surgical outcomes in cervical spondylotic myelopathy patients undergoing anterior cervical discectomy and fusion: assessment of minimum clinically important difference. PLoS One. 2013; 8(6):e67408. https://doi.org/10.1371/journal.pone.0067408 PMID: 23826290

26. Chien A, Lai DM, Cheng CH, Wang SF, Hsu WL, Wang JL. Responsiveness of the Chinese versions of the Japanese Orthopaedic Association Cervical Myelopathy Evaluation Questionnaire and Neck Disability Index in postoperative patients with cervical spondylotic myelopathy. Spine (Phila Pa 1976). 2015; 40(17):1315–21. https://doi.org/10.1097/BRS.0000000000001005 PMID: 26020849

27. Tetreault L, Nouri A, Kopjar B, Côté P, Fehlings MG. The Minimum Clinically Important Difference of the Modified Japanese Orthopaedic Association Scale in Patients with Degenerative Cervical Myelopathy. Spine (Phila Pa 1976). 2015; 40(21):1653–9. https://doi.org/10.1097/BRS.0000000000001127 PMID: 26502097

28. Kato S, Oshima Y, Matsubayashi Y, Taniguchi Y, Tanaka S, Takeshita K. Minimum Clinically Important Difference and Patient Acceptable Symptom State of Japanese Orthopaedic Association Score in Degenerative Cervical Myelopathy Patients. Spine (Phila Pa 1976). 2019; 44(10):691–7. https://doi.org/10.1097/BRS.0000000000002928 PMID: 30395093

29. Dworkin RH, Turk DC, Wyrwich KW, Beaton D, Cleeland CS, Farrar JT, et al. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. J Pain. 2008; 9(2):105–21. https://doi.org/10.1016/j.jpain.2007.09.005 PMID: 18055266

30. Fekete TF, Haschtmann D, Kleinstuck FS, Porchet F, Jeszenszky D, Mannion AF. What level of pain are patients happy to live with after surgery for lumbar degenerative disorders? Spine J. 2016; 16(4 Suppl):S12–8. https://doi.org/10.1016/j.spinee.2016.01.180 PMID: 26850172

31. de Vet HC, Ostelo RW, Terwee CB, van der Roer N, Knol DL, Beckerman H, et al. Minimally important change determined by a visual method integrating an anchor-based and a distribution-based approach. Qual Life Res. 2007; 16(1):131–42. https://doi.org/10.1007/s11136-006-9109-9 PMID: 17033901

32. Kato S, Oshima Y, Matsubayashi Y, Taniguchi Y, Tanaka S, Takeshita K. Minimum clinically important difference in outcome scores among patients undergoing cervical laminoplasty. Eur Spine J. 2019; 28 (5):1234–41. https://doi.org/10.1007/s00586-019-05945-y PMID: 30877386

33. Steinhaus ME, Iyer S, Lovecchio F, Khechen B, Stein D, Ross T, et al. Minimal Clinically Important Difference and Substantial Clinical Benefit Using PROMIS CAT in Cervical Spine Surgery. Clin Spine Surg. 2019; 32(9):392–7. https://doi.org/10.1097/BSD.0000000000000895 PMID: 31569175

34. Carreon LY, Glassman SD, Campbell MJ, Anderson PA. Neck Disability Index, short form-36 physical component summary, and pain scales for neck and arm pain: the minimum clinically important difference and substantial clinical benefit after cervical spine fusion. Spine J. 2010; 10(6):469–74. https://doi.org/10.1016/j.spinee.2010.02.007 PMID: 20359958

35. Parker SL, Godil SS, Shau DN, Mendenhall SK, McGirt MJ. Assessment of the minimum clinically important difference in pain, disability, and quality of life after anterior cervical discectomy and fusion: clinical article. J Neurosurg Spine. 2013; 18(2):154–60. https://doi.org/10.3171/2012.10.SPINE12312 PMID: 23176164

36. Matthew FG, Burkus JK, Mark ES, Francine WS, Anne GC. Cervical disc arthroplasty: 10-year outcomes of the Prestige LP cervical disc at a single level. Journal of Neurosurgery: Spine SPI. 2019; 31 (3):317–25.

37. Gornet MF, Lanman TH, Burkus JK, Hodges SD, McConnell JR, Dryer RF, et al. Cervical disc arthroplasty with the Prestige LP disc versus anterior cervical discectomy and fusion, at 2 levels: results of a prospective, multicenter randomized controlled clinical trial at 24 months. Journal of Neurosurgery: Spine SPI. 2017; 26(6):653–67. https://doi.org/10.3171/2016.10.SPINE16264 PMID: 28304237

38. Werner DAT, Grotle M, Gulati S, Austevoll IM, Lonne G, Nygaard OP, 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 PMID: 28616747

39. 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(1):31. https://doi.org/10.1186/s12891-018-2386-y PMID: 30658613

40. Solberg T, Johnsen LG, Nygaard OP, Grotle M. Can we define success criteria for lumbar disc surgery?: estimates for a substantial amount of improvement in core outcome measures. Acta Orthop. 2013; 84(2):196–201. https://doi.org/10.3109/17453674.2013.786634 PMID: 23506164

41. Hermansen E, Myklebust TA, Austevoll IM, Rekeland F, Solberg T, Storheim K, et al. Clinical outcome after surgery for lumbar spinal stenosis in patients with insignificant lower extremity pain. A prospective cohort study from the Norwegian registry for spine surgery. BMC Musculoskelet Disord. 2019; 20(1):36. https://doi.org/10.1186/s12891-019-2407-5 PMID: 30669998

42. Mjåset C, Zwart JA, Goedmakers CMW, Smith TR, Solberg TK, Grotle M. Criteria for success after surgery for cervical radiculopathy-estimates for a substantial amount of improvement in core outcome measures. Spine J. 2020; 20(9):1413–21. https://doi.org/10.1016/j.spinee.2020.05.549 PMID: 32502657

43. van Hooff ML, Jacobs WC, Willems PC, Wouters MW, 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 PMID: 25909475

44. Mannion AF, Elfering A. Predictors of surgical outcome and their assessment. Eur Spine J. 2006; 15 Suppl 1:S93–108. https://doi.org/10.1007/s00586-005-1045-9 PMID: 16320033

45. Solberg TK, Sorlie A, Sjaavik K, Nygaard OP, Ingebrigtsen T. Would loss to follow-up bias the outcome evaluation of patients operated for degenerative disorders of the lumbar spine? Acta Orthop. 2011; 82 (1):56–63. https://doi.org/10.3109/17453674.2010.548024 PMID: 21189113

46. Hojmark K, Stottrup 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 PMID: 26208938

Paper III

ORIGINAL ARTICLE - SPINE DEGENERATIVE

Anterior surgical treatment for cervical degenerative radiculopathy: a prediction model for non-success

Christer Mjåset[1,2,3] · Tore K. Solberg[4,5] · John-Anker Zwart[1,3] · Milada C. Småstuen[6] · Frode Kolstad[2] · Margreth Grotle[3,6]

Abstract

Purpose By using data from the Norwegian Registry for Spine Surgery, we wanted to develop and validate prediction models for non-success in patients operated with anterior surgical techniques for cervical degenerative radiculopathy (CDR).

Methods This is a multicentre longitudinal study of 2022 patients undergoing CDR surgery and followed for 12 months to find prognostic models for non-success in neck disability and arm pain using multivariable logistic regression analysis. Model performance was evaluated by area under the receiver operating characteristic curve (AUC) and a calibration test. Internal validation by bootstrapping re-sampling with 1000 repetitions was applied to correct for over-optimism. The clinical usefulness of the neck disability model was explored by developing a risk matrix for individual case examples.

Results Thirty-eight percent of patients experienced non-success in neck disability and 35% in arm pain. Loss to follow-up was 35% for both groups. Predictors for non-success in neck disability were high physical demands in work, low level of education, pending litigation, previous neck surgery, long duration of arm pain, medium-to-high baseline disability score and presence of anxiety/depression. AUC was 0.78 (95% CI, 0.75, 0.82). For the arm pain model, all predictors for non-success in neck disability, except for anxiety/depression, were found to be significant in addition to foreign mother tongue, smoking and medium-to-high baseline arm pain. AUC was 0.68 (95% CI, 0.64, 0.72).

Conclusion The neck disability model showed high discriminative performance, whereas the arm pain model was shown to be acceptable. Based upon the models, individualized risk estimates can be made and applied in shared decision-making with patients referred for surgical assessment.

Keywords Degenerative neck surgery · Predictors · Prognostic model · Outcome · Neck disability · Arm pain

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

✉ Christer Mjåset
  chrmja@gmail.com

1 Faculty of Medicine, University of Oslo, Oslo, Norway

2 Department of Neurosurgery, Oslo University Hospital, Oslo, Norway

3 Division of Clinical Neuroscience, Department of Research and Innovation, Oslo University Hospital, P.O. Box 4956, 0424 Oslo, Nydalen, Norway

4 Institute of Clinical Medicine, The Arctic University of Norway, Tromsø, Norway

5 Department of Neurosurgery and The Norwegian Registry for Spine Surgery (NORspine), The University Hospital of North Norway, Tromsø, Norway

6 Department of Rehabilitation and Technology, Faculty of Health Science, Oslo Metropolitan University, St. Olavs Plass, P.O. Box 4, 0130 Oslo, Norway

Introduction

Cervical degenerative radiculopathy (CDR) is caused by nerve root compression by a herniated or bulging disc and/ or ligament hypertrophy and bony spurs. The incidence rate is reported to be approximately 80 per 100,000 people [32], and surgical treatment is usually offered to patients with persistent arm pain and/or paresis [10]. With the introduction of modern operative techniques like anterior cervical discectomy and fusion or disc arthroplasty, treatment safety and effectiveness have increased dramatically [10]. Currently, day surgery is practiced in many clinics worldwide [14, 20]. Still, far from all patients improve after surgery [5–7, 12]. Many studies have investigated what predicts a beneficial outcome [21, 31], but there is current lack of evidence concerning factors associated with unfavourable or non-successful outcomes. A high body mass index [47],

mental health problems [1, 19] and lower social class [15] are individual patient characteristics that have been linked to poor treatment outcomes after cervical degenerative surgery. Predictive models can aid in calibrating surgeons' and patients' expectations prior to intervention, thus enhancing clinical decision-making and patient selection for surgical intervention.

The primary objective of this study was to develop and validate a prediction model for non-success in neck disability 12 months after surgery for CDR. Secondary objectives were to provide the same analysis for arm pain and to develop a risk matrix for the primary outcome to exemplify the use of the model in a clinical setting.

Methods

Design and ethics

This is a multicentre longitudinal study following the recommendations for reporting in observational studies, STROBE criteria [44] and the methodological framework proposed by the PROGRESS framework [33, 40]. The manuscript is reported according to the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) guidelines [9]. We used the Prediction model Risk Of Bias Assessment Tool (PROBAST) to minimize the risk of bias [27, 45]

Our research protocol was approved by the Norwegian Committee for Medical and Health Research Ethics Midt (2014/344). Written informed consent was obtained from all patients.

Patients and surgical treatment

Data from the Norwegian Registry for Spine Surgery (NORspine) from 2011 to 2016 was used. NORspine is a government-funded comprehensive clinical registry receiving no industry funding and used for quality assessment and research. Informed consent is obtained from all patients before they enter the registry. Currently, all centres performing cervical spine surgery in Norway report data to NORspine (coverage = 100%), and the operation recording rate is 78% (completeness) [3]. Patients who had undergone anterior cervical discectomy and fusion or arthroplasty surgery due to cervical degenerative radiculopathy in the period were included. For both groups, baseline characteristics and 12-month outcome data were similar, except from baseline Neck Disability Index and neck pain scores, which were slightly higher in the arthroplasty group ($p = 0.02$ and $p = 0.002$, respectively). Also, arthroplasty patients were operated in significantly lower number of levels ($p < 0.001$). Patients undergoing posterior cervical procedures due to

CDR, as well as all patients operated for myelopathy symptoms, were excluded. Patients operated for tumours, fractures and primary infections are not included in NORspine.

Patients completed data at admission for surgery (baseline), after 3 and 12 months. Surgeon's forms containing information about diagnosis, treatment and comorbidity were completed during the hospital stay. Only cohort participants responding to the 12-month questionnaire were included in the present study. Follow-up was conducted by the central registry unit without involvement of the treating hospital. The patients responded by questionnaires sent and returned by mail. One reminder with a new form was sent to non-respondents within 2 weeks.

Outcome definitions

For the primary outcome, the Neck Disability Index (NDI) (0–100), non-success was defined as an absolute score of > 26 at 12-month follow-up. For the secondary outcome, arm pain intensity assessed by numerical rating scale (NRS-AP) (0–10), non-success was defined as a score ≥ 3 at 12-month follow-up. These estimates are based on a previous study for patients undergoing CDR surgery in Norway [25].

Candidate predictors

Candidate predictors for non-success were selected from the comprehensive NORspine questionnaire administered before surgery, which consists of information about sociodemographic factors, lifestyle, work and clinical variables in addition to patient-reported outcome measures (PROMs). Data from the surgeons' forms were used for information about diagnosis, treatment, comorbidity, the American Society of Anaesthesiologists physical status (ASA), surgical indication and type of operation. The selection of the final set of predictors was made after a thorough literature review where we identified the factors that have been found to be significantly and consistently associated with outcomes after CDR surgery.

The following predictors were selected for the model: gender, age groups (below 40 years, between 40 and 60 years or above 60 years), work status prior to operation (on sick leave, retired or disabled; on rehabilitation, out of work or on work return training; student, fully working or housewife/househusband), physical demands in work (working with computers, sitting, light physical work or hard physical work), educational level (high school or less, less than 4 years university or 4 or more years of university), mother tongue (Norwegian or non-native speaker), pending litigation (yes/no) (pending litigation defined as unresolved claims or litigation issues against the Norwegian Public Welfare Agency Fund concerning permanent disability pension or compensation claims against private insurance

companies or the public Norwegian System of Compensation to Patients), duration of arm pain (less than 3 months, 3 to 12 months or more than 12 months), duration of pre-operative paresis (no paresis, less than 3 months or more than 3 months), body mass index (BMI) (equal to or below 30 or above 30), smoking (yes/no), comorbidities (yes/no), previous neck surgery (yes/no), number of surgical levels (one or more than one), daily use of analgetic drugs (yes/no), ASA level (level 1–2 or level 3–4), arm pain neck pain ratio (above 1 or below or equal to 1) [30] and anxiety/depression by the item on the EuroQol-5D-3L (EQ-5D) questionnaire ("moderate" to "extremely" anxious or depressed or "not anxious or depressed"). In addition, baseline outcome scores were included as potential predictors; the baseline scores were categorized into low, medium and high by percentile distribution.

Sample size considerations

Since no consensus on sample size in prognostic modelling exists, we chose to follow the recommendations by Steyerberg [38]: (a) aiming for at least 100 events as a minimum for reliable estimation of the average risk and (b) aiming for at least 10 events per variable (EPV) and preferably 20, for reliable prediction modelling if the event rate is < 20% and higher EPV values if the event rate is between 20 and 80%. In the present material, approximately 700 cases had non-success at 12-month follow-up, and with this large number of EPV, we had nearly 40 cases per event and good statistical power for the prediction model analyses. The large EPV will reduce the potential for overfitting and optimism of the final models. Overfitting is defined as fitting a statistical model with too many effective degrees of freedom in the modelling process. Estimation bias is defined as the overestimation of effects of predictors because selection of the effects withstood a statistical test, whereas optimism is defined as the difference between true performance (performance in the underlying population, e.g. external validation sample) and apparent performance (development sample) [38].

Statistical analysis

Statistical analyses were performed using IBM SPSS Statistics version 26 for Windows and the STATA version 16 for Windows. Missing data was checked for all variables and are reported together with descriptive data. Frequencies were used for categorical data and mean and standard deviation (SD) for continuous data. Continuous variables, such as baseline disability, were categorized to be adapted into a risk matrix. The distribution of baseline and 12-month scores of the two outcome measures are presented by mean scores and SD.

First, a univariable analysis of the candidate predictors was performed to assess the crude association between each candidate predictors and the two outcomes. Associations between outcomes and predictors are expressed as odd ratios (OR) with a 95% confidence interval (CI). Predictors reaching $p < 0.1$ in these analyses were entered into two multivariable logistic regression models (for primary and secondary outcome), where a stepwise backward elimination method was used. Variables that were not statistically significant ($p > 0.05$) in the multivariable models were removed from the model. The performance of the two final models was evaluated with (1) the explained variance by Nagelkerke's R^2, (2) the Hosmer–Lemeshow test $p > 0.05$) and (3) the discriminative ability of the model (the likelihood that the model allocates higher predicted risks to patients who achieve non-substantial improvement and lower predicted risks to those who do not) assessed by calculating the area under the receiver operating curves (AUC), also often referred to as the c-index [41]. The larger the AUC, the greater is the discriminative ability of the model. The discriminative performance of the models was considered acceptable if the AUC was ≥ 0.7 and good if the AUC was ≥ 0.8 (the c-criterion).

Internal validation was conducted by a bootstrap procedure (1000 samples) to estimate the amount of optimism in the two final models [26, 39]. A slope value was calculated (the closer to 1.0, the less over-optimism) and used to correct and shrink the regression coefficients, the R^2 and the c-index.

Clinical usefulness (risk matrix)

We assessed the potential clinical utility of the final prediction model for non-success in neck disability by developing a risk matrix for two hypothetical patient case profiles with few and many predictors present. Regression coefficients from the final disability model were converted into probabilities, and a risk score for each of the two individual case profiles was calculated by the sum of the products of individual values of each predictor variable and its regression coefficient. Depending on the presence or absence of the risk factors, the matrix was then calculated as probability for a non-substantial improvement after 12 months for each of the patients.

Results

There were 3142 patients who had undergone either anterior discectomy and fusion (3109) or arthroplasty (33) due to CDR during the study period. Out of these, 2022 (64.4%) completed 12-month follow-up and were included in the analyses (2020 for the NDI analysis and 1980 for the NRS-AP analysis). Compared to responders at 12-month

follow-up, non-responders were less likely to be female, significantly younger, had higher neck disability, more neck pain, lower quality of life, were less educated and more likely to be smoking.

Demographic and clinical characteristics at baseline for the included participants are summarized in Table 1, including the missing values for each variable. Gender distribution was equal (50%), and the average age at baseline was 51 years. Most patients were on sick leave before surgery. Approximately 40% of the patients reported that their job involved hard physical work, and only 17% had high level of education. Nearly half of the patients had experienced neck pain for more than 1 year. Only 5% of the patients had an ASA level of 3 or more prior to surgery. There were few missing values for the candidate predictor variables, except for physical demands in work, pending litigation, duration of pre-operative paresis, previous neck surgery and arm/neck pain ratio (Table 1).

The mean scores of the NDI and NRS arm pain at 12-month follow-up was 23.4 (SD 18.8) and 2.9 (SD 2.8), respectively. A total of 38.0% had non-successful outcomes in neck disability and 35.3% in arm pain.

Table 2 presents the univariable analysis of all candidate predictors. Most candidate predictors showed a statistical univariate relationship to the two outcomes: female gender, being retired or receiving disability or rehabilitation pension, high physical demands in work, low education level, being a non-native speaker, having a pending litigation, smoking, presence of comorbidity, having undergone previous neck surgery, having long duration of arm pain or long duration of paresis prior to surgery, high ASA level, daily use of analgetic drugs, arm pain worse than neck pain, presence of anxiety/depression or high baseline scores of NDI or arm pain. Age, obesity and number of surgical levels were not significantly associated to any of the two outcomes.

Table 3 shows the results from the multivariable analyses. Seven predictors (hard physical demands in work, low level of education, pending litigation, previous neck surgery, duration of arm pain > 3 months, medium or high levels of baseline disability and anxiety/depression) showed statistically significant association with non-success in neck disability. The model displayed good overall performance with Nagelkerke R^2 of 28.3%, non-significant Hosmer–Lemeshow test and AUC 0.78 (95% CI 0.75, 0.82). The prediction model for non-success in arm pain included six of the same predictors (hard physical demands in work, low level of education, pending litigation, previous neck surgery, duration of arm pain > 3 months, medium or high levels of baseline disability) in addition to foreign mother tongue, smoking and medium or high levels of arm pain. This model showed acceptable performance with Nagelkerke R^2 of 15.5% and AUC of 0.68 (95% CI 0.64, 0.72). The calibration plots for the two models are displayed in Figs. 1 and 2. Both models

had high calibration slopes of 1.0, indicating no overfitting of the models.

Two risk matrices were developed for cases with low and high risk for non-success in neck disability, respectively. "Low-risk" was defined as having three out of the eleven risk factors in the prognostic model, while "high-risk" was defined as having six out of the eleven factors. The matrices are displayed in Table 4 and show that a low-risk individual the risk for non-success was 13%, whereas for a high-risk individual, the risk for non-success was 92%.

Discussion

In this study, we found that more than one third of the patients reported non-successful outcome in neck disability or arm pain at 12-month follow-up after surgery for cervical degenerative radiculopathy. Patients with high risk for non-success in neck disability were characterized by physical demanding work, low level of education, pending litigation, previous neck surgery, duration of arm pain > 3 months and medium-to-high levels of baseline disability as well as anxiety/depression. The predictors for non-success in arm pain were foreign mother tongue, smoking, medium-to-high levels of baseline arm pain and all neck disability model predictors except for anxiety/depression.

The discriminative performance of the neck disability model was found to be good with an AUC of 0.78, whereas the arm pain model was slightly less accurate but still acceptable (0.68). A recent study on patients undergoing elective cervical spine surgery by Archer et al. reports slightly lower AUCs for a predictive model of worse NDI scores (0.64–0.69) and of worse arm pain scores (0.63–0.65) 1 year after intervention [2]. There is a large overlap of significant predictors between our two studies. For example, Archer et al. found that worsening of NDI and arm pain scores were significantly associated with longer symptom duration, workers' compensation claims and higher baseline NDI — all of which are included in our two present models. In accordance with our results, Archer et al. found depression only to be significantly associated with worse NDI scores. Several other studies have shown a negative impact of mental health on outcomes after surgery for CDR [1, 11, 19, 23]. Further, Archer et al. found no association between worsening of scores and smoking or pre-operative pain level. In the present study, both factors were significantly associated with non-success in arm pain.

There exists conflicting evidence regarding gender and its impact on PROMs and other outcomes, such as length of hospital stay and complication rates after degenerative neck surgery [4, 18, 34]. Archer et al. found that female sex was among the predictors for worse neck disability scores but not for worse arm pain scores. In another multivariate

Table 1 Characteristics of participants at baseline (*n*=3142) and 12-months follow-up (2022), including number of missing values in each of the variables

Characteristics and domain	Baseline (*n*=3142)	Complete sample (*n*=2022)
Sociodemographic		
Female gender, *n* (%)	1502 (47.8)	1005 (49.7)
Age, mean years (SD)	49.5 (9.3)	51.0 (9.2)
Age <40	404 (12.9)	189 (9.4)
Age 40-60	2355 (75.0)	1534 (75.9)
Age >60	380 (12.1)	298 (14.7)
Missing	3 (0)	1 (0)
Work status prior to operation, *n* (%)		
Student, in work or at home	974 (31.4)	187 (9.3)
Retired or disability pension	460 (14.8)	334 (16.6)
Rehabilitation pension	278 (9.0)	157 (7.8)
Sick leave	1390 (44.8)	1333 (66.3)
Missing	40 (1.3)	11 (0.5)
Physical demands in work, *n* (%)		
Working in front of a computer/sitting still	1039 (38.8)	712 (41.6)
Light physical work	486 (18.1)	319 (18.6)
Hard physical work	1156 (43.1)	681 (39.8)
Missing	461 (14.7)	310 (15.5)
Educational level, *n* (%)		
High school or less	1963 (64.1)	1231 (62.3)
Less than 4 years of university	626 (20.5)	405 (20.5)
4 or more years of university	472 (15.4)	339 (17.2)
Missing	81 (2.5)	47 (2.3)
Non-native speaker, *n* (%)	229 (7.3)	134 (6.6)
Missing	2 (0)	2 (0)
Pending litigation[1], *n* (%)		
No	2764 (89.0)	1680 (83.1)
Yes	342 (11.0)	342 (16.9)
Missing	36 (1.1)	0 (0)
Physical/somatic		
Obesity (Body Mass Index ≥ 30), *n* (%)	657 (21.4)	419 (21.2)
Missing	78 (2.5)	48 (2.4)
Smoking	1043 (33.9)	582 (29.4)
Missing	63 (2.0)	40 (2.0)
Comorbidity	1272 (40.5)	842 (41.6)
Missing	0 (0)	0 (0)
Previous neck surgery (same level)	265 (8.6)	160 (8.1)
Missing	52 (1.7)	40 (2.0)
Number of surgical levels		
One level	2347 (75.3)	1488 (74.4)
Two or more levels	768 (24.7)	512 (25.6)
Missing	27 (0.9)	22 (1.0)
Type of surgery		
Anterior discectomy and fusion	3109 (98.9)	2003 (99.0)
Anterior discectomy and arthroplasty	33 (1.1)	19 (1.0)
Missing	0 (0)	0 (0)
Clinical self-report		
Duration of arm pain		
< 3 months	440 (14.4)	284 (14.4)
3-12 months	1120 (36.7)	714 (36.2)

Table 1 (continued)

Characteristics and domain	Baseline (n=3142)	Complete sample (n=2022)
> 1 year	1494 (48.9)	975 (49.4)
Missing	88 (2.8)	49 (2.4)
Duration of pre-operative paresis		
No paresis	641 (22.6)	425 (22.4)
< 3 months	450 (15.8)	357 (18.8)
3 months or more	1750 (61.6)	1116 (58.8)
Missing	301 (9.6)	124 (6.1)
Daily use of analgetic drugs (vs < daily use)	1634 (52.8)	1042 (52.4)
Missing	48 (1.5)	25 (1.2)
ASA level of 3 or more[2]	151 (5.0)	101 (5.2)
Missing	99 (3.2)	76 (3.8)
Arm pain worse than neck pain	945 (32.0)	591 (31.2)
Missing	187 (6.0)	125 (6.1)
Baseline neck pain (NRS[3]), mean (SD)	6.1 (2.5)	6.1 (2.5)
Baseline arm pain (NRS[3]), mean (SD)	6.4 (2.4)	6.4 (2.4)
Baseline disability (NDI[4]), mean (SD)	41.6 (15.1)	41.0 (15.2)
Psychological		
Anxiety or depression[5]	1355 (44.0)	832 (41.8)
Missing	62 (2.0)	31 (1.0)

[1]Pending medical claim/litigation against the Norwegian public welfare agency fund concerning disability pension or pending medical compensation claim/litigation against private insurance companies or the public Norwegian System of Compensation to Patients. [2]American Society of Anesthesiologists grade. [3]Numeric rating scale (0-10). [4]Neck Disability Index, 0-100 (no-maximal disability). [5]Based on scoring "moderate" or "extremely" anxious or depressed in the item in EQ-5D-3L questionnaire

analysis, Scerrati et al. found that female sex and two-level surgery (vs. one-level surgery) correlated with worse outcomes in NDI, as well as the use of postoperative collars, while BMI only was shown to be significant in an univariate analysis [34]. In the present model, neither gender, number of surgical levels nor obesity did show significant association with non-success in neck disability or arm pain. There are also conflicting results in literature regarding the impact of obesity on neck disability. For example, similar to our results, Sielatycki et al. found no correlation between a high BMI and cut-offs for several PROMs, including NDI [35], whereas Zhang et al. found that high BMI was associated with longer hospital stay, duration of surgery and higher postoperative complication rates [47].

The present study could not find that high age was a predictor of non-success in neither neck disability nor arm pain. This is supported by other multivariate studies [2, 29, 34]. Further, both comorbidity and ASA level only came out as significant predictors for non-success in the present univariate analyses but not in the final multivariate analysis. In a study of risk factors for failure to achieve a minimal clinically important difference (MCID) in NDI 12 months after surgery for cervical radiculopathy, a higher burden of comorbidity was found to be the most significant predictor [29]. Other studies have emphasized the significance of age

and pre-operative functional status as a predictor of complications and mortality after cervical degenerative surgery [24, 28]. Since changing demographics are likely to significantly increase the age and frailty of those who seek operative care for cervical degenerative disease in the coming years, further research is warranted in relation to these aspects.

Impact of findings

In the current healthcare environment, value-based thinking has brought more focus on quality and appropriateness of care. Also, as degenerative neck surgery is becoming increasingly safe and efficient, there is a need for more knowledge about which patients are not improving from surgery. The two present models can be used in a clinical setting to predict which patients will benefit from a surgical intervention and who will be better off being treated conservatively. To exemplify how these models can be used in a surgical practice, we produced a risk matrix constituted of two hypothetical patient scenarios for disability; one where the patient had several of the risk factors and another where the patient had only a few risk factors (Table 4). The patients with few predictors had low probability for non-success (0.13), while several predictors involved a high risk for non-success (0.92). According to our model, a patient with similar characteristics

Table 2 Univariate associations at 12-month follow-up between candidate predictors and the two outcomes; non-substantial improvement in disability and arm pain. Regression coefficient and odds ratio (95% confidence intervals) (n=2022)

	Total number of cases n=2020	Non-success in neck disability (12-mo NDI ≥26) * n=768 (38%)	Odds Ratio (95% CI)	Total number of cases n=1980	Non-success in arm pain (12-mo arm pain ≥3) ** n=698 (35.3%)	Odds Ratio (95% CI)
Socio-demographic						
Female gender	2020	408	1.25 (1.04, 1.49)	1980	366	1.18 (0.98, 1.42)
Age, years	2019			1979		
Age <40		69	Ref		59	Ref
Age 40-60		593	1.09 (0.79, 1.49)		549	1.23 (0.89, 1.70)
Age >60		105	0.94 (0.65, 1.38)		89	0.99 (0.67, 1.49)
Work status	2009			1969		
Student, in work or stay-at-home		61	Ref		62	Ref
Retired or disability pension		173	2.25 (1.55, 3.27)		134	1.41 (0.96, 2.05)
Rehabilitation pension		113	5.31 (3.31, 8.43)		95	3.20 (2.04, 5.00)
Sick leave		415	0.93 (0.67,1.30)		403	0.86 (0.62, 1.20)
Physical demands in work	1712			1683		
Computers/ sitting		187	Ref		174	Ref
Light physical work		103	1.34 (1.00, 1.79)		99	1.41 (1.05, 1.89)
Hard physical work		296	2.16 (1.72, 2.70)		286	2.28 (1.81, 2.87)
Educational level	1974			1935		
High school or less		528	Ref		486	Ref
Less than 4 years of university		143	0.73 (0.58, 0.92)		129	0.7 (0.55, 0.89)
4 or more years of university		78	0.40 (0.30, 0.52)		61	0.33 (0.25, 0.45)
Non-native speaker	2018	70	1.08 (1.03, 1.13)	1978	72	1.11 (1.06, 1.16)
Pending litigation[1]	1996			1957		
None		565	Ref		537	Ref
Yes		119	3.06 (2.26, 4.15)		95	2.01 (1.49, 2.72)
Already approved		71	3.13 (2.12, 4.60)		59	2.27 (1.55, 3.33)
Physical/ somatic						
Obesity (Body Mass Index ≥ 30)	1972	169	1.14 (0.92, 1.42)	1932	141	0.98 (0.78, 1.24)
Smoking	1980	273	1.71 (1.40, 2.08)	1941	263	1.93 (1.58, 2.36)
Comorbidities	2020	378	1.66 (1.38, 1.99)	1980	323	1.55 (1.12, 1.63)
Previous neck surgery	1980	89	1.09 (1.05, 1.15)	1940	84	1.10 (1.06, 1.14)
Number of surgical levels	1998			1958		
One level		550	Ref		512	Ref
Two or more levels		209	1.18 (0.96, 1.45)		177	0.91 (0.82, 1.25)
Pain and symptoms						

Table 2 (continued)

	Total number of cases	Non-success in neck disability (12-mo NDI ≥26) *	Odds Ratio (95% CI)	Total number of cases	Non-success in arm pain (12-mo arm pain ≥3) **	Odds Ratio (95% CI)
	n=2020	n=768 (38%)		n=1980	n=698 (35.3%)	
Duration of arm pain	1971			1933		
< 3 months		65	Ref		57	Ref
3-12 months		237	1.66 (1.21, 2.29)		214	1.67 (1.20, 2.33)
> 1 year		444	2.81 (2.07, 3.81)		406	2.83 (2.06, 3.90)
Duration of pre-operative paresis	1828			1860		
No paresis		124	Ref		107	Ref
< 3 months		110	1.08 (0.80, 1.47)		96	1.08 (0.78, 1.49)
3 months or more		467	1.75 (1.38, 2.23)		441	1.91 (1.49, 2.46)
Daily use of Analgetic drugs	1985	268	0.45 (0.37, 0.54)	1946	269	0.58 (0.48, 0.71)
ASA level 3 or more[2]	1944	47	1.50 (1.00, 2.24)	44	44	1.57 (1.05, 2.37)
Neck pain worse than arm pain	1895	270	1.13 (1.08, 1.17)	1951	195	0.88 (0.72, 1.08)
Baseline NDI score (0-100)[3]	2013			1973		
Low (0-40)		173	Ref		225	Ref
Medium (41-60)		448	4.84 (3.92, 5.99)		358	2.29 (1.87, 2.28)
High (> 60)		144	10.74 (7.62, 15.14)		113	4.21 (3.06, 5.80)
Baseline NRS arm pain (0-10)[4]	1982			1944		
Low (0-5)		117	Ref		86	Ref
Medium (6-7)		300	1.45 (1.12, 1.87)		273	1.87 (1.42, 2.48)
High (> 8)		334	2.06 (1.59, 2.67)		330	3.01 (2.28, 3.40)
Psychological						
Anxiety or depression[5]	1989	422	2.62 (2.18, 3.16)	1950		1.75 (1.45, 2.11)

*38% did not achieve a substantial improvement in disability. **35.3% did not achieve a substantial improvement in arm pain. [1]Pending medical claim/litigation against the Norwegian public welfare agency fund concerning disability pension or pending medical compensation claim/litigation against private insurance companies or the public Norwegian System of Compensation to Patients. [2]American Society of Anesthesiologists grade. [3]Neck Disability Index, 0-100 (no-maximal disability). [4]Numeric Rating Scale (0-10). [5]Based on scoring "moderate" to "extremely" anxious or depressed in the item in EQ-5D-3L questionnaire

and symptomatology as described in the case study with few predictors should be reassured that surgery is a safe option in terms of improving from baseline arm pain and disability. Patients with a similar clinical picture as patient 2 with several positive predictors, on the other hand, should be counselled about alternative treatment strategies.

The present models can be further developed into a risk calculator to assess the probability of success or failure to achieve substantial change for every patient in a surgical practice. However, the model will first need to be further validated in other study populations. The feasibility of a risk calculator should also be evaluated.

Table 3 Predictors for non-success in neck disability or arm pain at 12-months after surgery. Results are presented by Odds Ratio (OR) and bootstrapped 95% Confidence Intervals (CI) for the significant variables

	Non-success in neck disability (12-mo NDI[1]≥26*) n=1593	Non-success in arm pain (12-mo NRS-AP[2]≥3**) n=1546
Hard physical demands in work (vs computers/sitting still or light physical work)	1.56 (1.22, 2.00)	1.46 (1.15, 1.85)
High educational level (4 or more years of university) (vs high school or less than 4 years of university)	0.57 (0.41, 0.78)	0.51 (0.36, 0.71)
Pending litigation[3] (vs none)	2.38 (1.70, 3.34)	1.68 (1.21, 2.33)
Previous neck surgery (vs not)	2.52 (1.61, 3.96)	2.01 (1.33, 3.03)
Duration of arm pain		
< 3 months	Ref	Ref
3-12 months	1.81 (1.24, 2.65)	1.48 (1.01, 2.18)
> 1 year	2.51 (1.72, 3.66)	2.42 (1.66, 3.52)
Anxiety or depression	1.74 (1.38, 2.19)	-
Baseline NDI[1] score		
Low (0-40)	Ref	Ref
Medium (41-60)	4.20 (3.22, 5.48)	1.79 (1.40, 2.29)
High (> 60)	7.79 (5.07, 11.98)	2.53 (1.67, 3.85)
Baseline Arm Pain score		
Low (0-4)	-	Ref
Medium (5-7)	-	1.66 (1.18, 2.32)
High (8-10)	-	2.04 (1.43, 2.90)
Foreign mother tongue (vs Norwegian)	-	1.71 (1.12, 2.61)
Smoking (vs no smoking)	-	1.44 (1.11, 1.85)
Nagelkerke R square	28.3%	17.3%
Discrimination, AUC[4]	0.78 (0.75, 0.82)	0.68 (0.64, 0.72)
Hosmer-Lemeshow test	p=0.455	p=0.753

* Number of participants with poor primary outcome n=768 (38%), ** Number of participants with poor secondary outcome n=698 (35,3%).
[1]Neck Disability Index, 0-100 (no-maximal disability).[2]A Numeric Rating Scale for arm pain (0-10). [3]Pending medical claim/litigation against the Norwegian public welfare agency fund concerning disability pension or pending medical compensation claim/litigation against private insurance companies or the public Norwegian System of Compensation to Patients.[4]Area Under the operating Curve

Strengths and limitations

An advantage of the present study is the large sample size of data captured in a national registry. NORspine was designed to prospectively capture important candidate predictors and PROMs prior to and during the year following surgery. The registry covers all the hospitals and private clinics conducting surgery on spinal disorders in Norway. A total of 78% of the operations are recorded in the registry [36]. Furthermore, our two models were well balanced with respect to the risk of overfitting, in particular the disability model which showed high accuracy with only seven included predictors.

In our study, we chose to include patients operated with both arthroplasty and fusion. The baseline characteristics and 12-month outcome data were similar between the groups, except for slightly higher NDI and NRS-NP scores for the arthroplasty patients at baseline, as well as a lower number of operated levels. There is no current consensus about the use of arthroplasty vs fusion in patients with

degenerative cervical disease [8, 13, 16, 17, 42, 46]. One may question whether the results of the fusion group in our study can be generalized to the arthroplasty group since there are only 1% of arthroplasty patients in our cohort. Further studies are warranted to elucidate this issue.

Loss to follow-up was 35.6% at 12-month follow-up and could represent a selection bias. However, two recent Scandinavian spine registry studies based on similar cohorts have found that a loss to follow-up did not bias conclusions about treatment effects [22, 37].

Another potential limitation is related to the cut-off estimates of the applied PROMs. In the present study, we decided to use estimates of non-success instead of the concept of MCID. The main reason is that MCID often show to be less than measurement errors or estimates for smallest detectable change [43], making it difficult for a patient and/or a clinician to judge the clinical meaningfulness of these estimates. By using stricter estimates reflecting a substantial rather than minimal change, we argue that these cut-offs are

Fig. 1 Calibration plot for the final model predicting no substantial improvement in neck disability at 12-month follow-up (E:O, expected/observed; CITL, calibration-in-the-large; slope, calibration slope; AUC, area under the curve; CIs, confidence intervals)
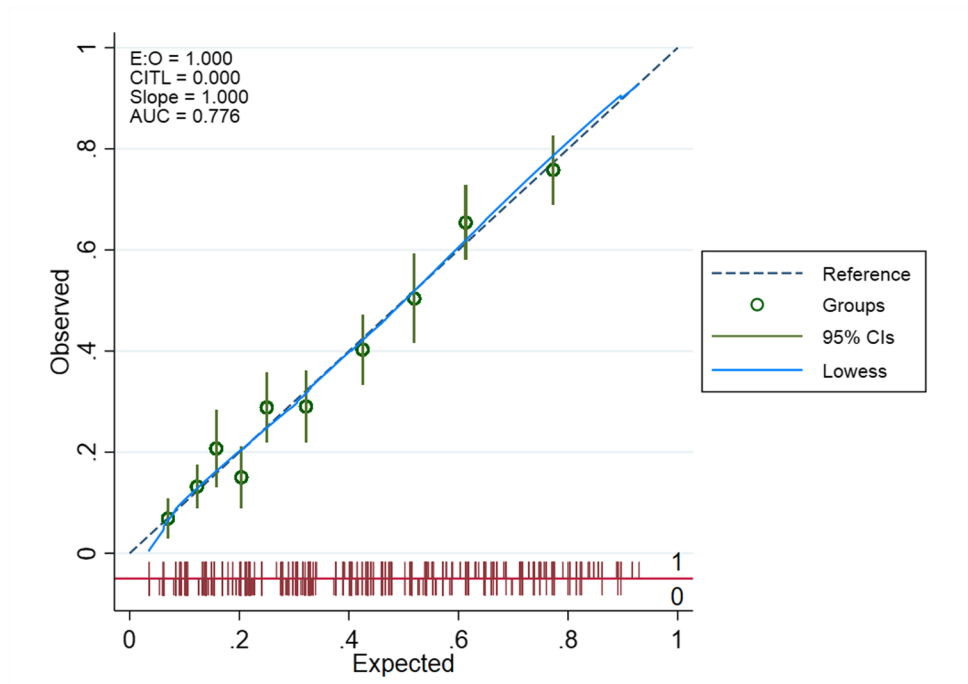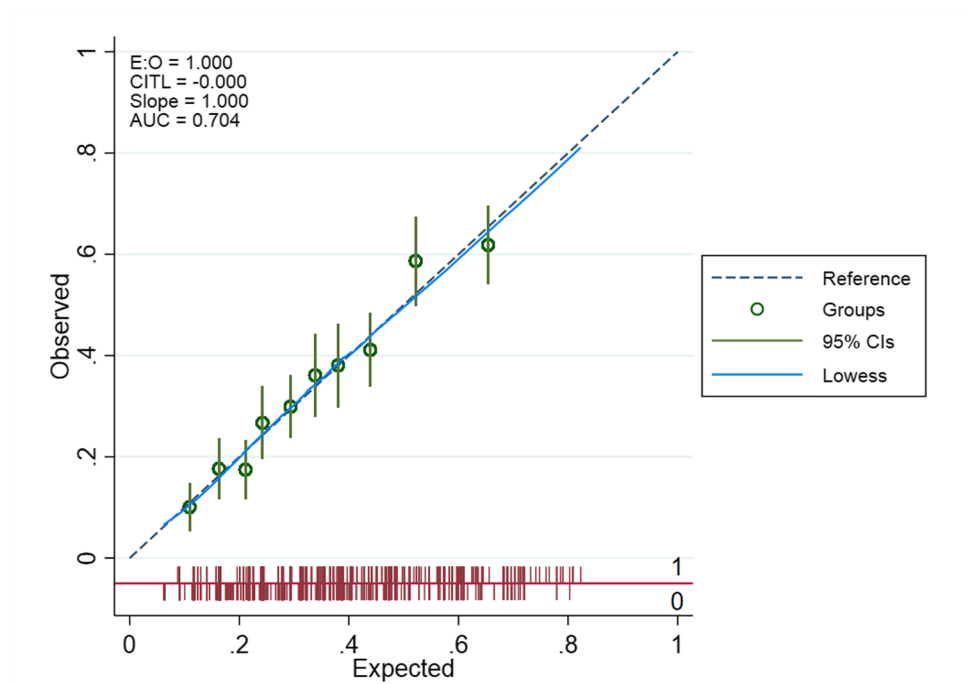

Fig. 2 Calibration plot for the final model predicting no substantial improvement in arm pain at 12-month follow-up (E:O, expected/observed; CITL, calibration-in-the-large; slope, calibration slope; AUC, area under the curve; CIs, confidence intervals)

better suited for use in the development of prediction models for non-success (or success).

The major limitation of the present study is that we did not externally validate the final models. External validation is necessary before these models can be further developed into a risk calculator used in clinical settings. Risk calculators may help inform discussions of surgical treatment

options between surgeons and patients and lead to more accurate judgement of operative risk. In a clinical decision-making process, the probability of successful or non-successful outcomes of conservative treatment strategies also needs to be taken into consideration. The present study only investigated outcomes after surgical treatment and cannot be generalized to outcomes after non-surgical treatment

Table 4 Example of two cases with few or several positive predictors from the final prediction model for non-success in neck disability at 12 months. For each of the cases, the predicted probability has been calculated based upon the presence (yes) or absence (no) of each predictor

	Patient 1, few positive predictors	Patient 2, several positive predictors
Hard physical demands in work (vs computers/sitting still or light physical work)	No	Yes
High educational level (4 or more years of university) (vs high school or less than 4 years of university)	Yes	No
Pending litigation[1] (vs none)	No	Yes
Previous neck surgery (vs not)	No	Yes
Duration of arm pain		
< 3 months	Yes	No
3–12 months	No	No
> 1 year	No	Yes
Baseline NDI[2] score		
Low (0–40)	No	No
Medium (41–60)	Yes	No
High (> 60)	No	Yes
Anxiety or depression[3]	No	Yes
Probability	0.13 [0.12 to 0.15]	0.92 [0.91 to 0.94]

[1]Pending medical claim/litigation against the Norwegian public welfare agency fund concerning disability pension or pending medical compensation claim/litigation against private insurance companies or the public Norwegian System of Compensation to Patients. [2]Neck Disability Index, 0–100 (no-maximal disability). [3]Based on scoring "moderate" to "extremely" anxious or depressed in the item in EQ-5D-3L questionnaire

options. Thus, there is a large need for exploring prediction models for both surgical and non-surgical treatment trajectories and outcomes.

Conclusions

The final prediction model for non-successful outcome in neck disability 12 months after CDR surgery showed high discriminative performance, whereas the prediction model for arm pain was slightly less predictive. Based upon the two prediction models, individualized risk estimates can be made and used in shared decision-making with patients referred for surgical assessment. The models need to be externally validated and further tested in a clinical setting.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Alvin MD, Miller JA, Lubelski D, Nowacki AS, Scheman J, Mathews M, McGirt MJ, Benzel EC, Mroz TE (2016) The impact of preoperative depression and health state on quality-of-life outcomes after anterior cervical diskectomy and fusion. Global Spine J 6:306–313. https://doi.org/10.1055/s-0035-1562932

2. Archer KR, Bydon M, Khan I, Nian H, Pennings JS, Harrell FE Jr, Sivaganesan A, Chotai S, McGirt MJ, Foley KT, Glassman SD, Mummaneni PV, Bisson EF, Knightly JJ, Shaffrey CI, Asher AL, DEVIN CJ (2020) Development and validation of cervical prediction models for patient-reported outcomes at 1 year after cervical spine surgery for radiculopathy and myelopathy. Spine (Phila Pa 1976) 45:1541–1552. https://doi.org/10.1097/brs.0000000000003610

3. Austevoll IM, Gjestad R, Brox JI, Solberg TK, Storheim K, Rekeland F, Hermansen E, Indrekvam K, Hellum C (2017) 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 26:404–413. https://doi.org/10.1007/s00586-016-4683-1

4. Basques BA, Hijji FY, Khechen B, Haws BE, Mayo BC, Massel DH, Louie PK, Cardinal KL, Guntin JA, Singh K (2018) Sex differences for anterior cervical fusion: complications and length of stay. Spine (Phila Pa 1976) 43:1025–1030. https://doi.org/10.1097/brs.0000000000002512

5. Bertalanffy H, Eggert HR (1988) Clinical long-term results of anterior discectomy without fusion for treatment of cervical

radiculopathy and myelopathy. A follow-up of 164 cases. Acta Neurochir (Wien) 90:127–135. https://doi.org/10.1007/bf015 60567

6. Burkus JK, Traynelis VC, Haid RW Jr, Mummaneni PV (2014) Clinical and radiographic analysis of an artificial cervical disc: 7-year follow-up from the Prestige prospective randomized controlled clinical trial: Clinical article. J Neurosurg Spine 21:516–528. https://doi.org/10.3171/2014.6.Spine13996

7. Buttermann GR (2018) Anterior cervical discectomy and fusion outcomes over 10 years: a prospective study. Spine (Phila Pa 1976) 43:207–214. https://doi.org/10.1097/brs.0000000000 002273

8. Byvaltsev VA, Stepanov IA, Riew DK (2020) Mid-term to long-term outcomes after total cervical disk arthroplasty compared with anterior diskectomy and fusion: a systematic review and meta-analysis of randomized controlled trials. Clin Spine Surg 33:192–200. https://doi.org/10.1097/bsd.0000000000000929

9. Collins GS, Reitsma JB, Altman DG, Moons KG (2015) Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD). Ann Intern Med 162:735–736. https://doi.org/10.7326/l15-5093-2

10. Dakwar E, Deukmedjian A, Ritter Y, Dain Allred C, Rechtine GR (2016) Chapter 16 - Spinal pathology, conditions, and deformities: surgical intervention. In: Magee DJ, Zachazewski JE, Quillen WS, Manske RC (eds) Pathology and intervention in musculoskeletal rehabilitation (second edition). Elsevier, Maryland Heights, MO, pp 584–611. https://doi.org/10.1016/B978-0-323-31072-7. 00016-6

11. Diebo BG, Lavian JD, Liu S, Shah NV, Murray DP, Beyer GA, Segreto FA, Maffucci F, Poorman GW, Cherkalin D, Torre B, Vasquez-Montes D, Yoshihara H, Cukor D, Naziri Q, Passias PG, Paulino CB (2018) The impact of comorbid mental health disorders on complications following cervical spine surgery with minimum 2-year surveillance. Spine (Phila Pa 1976) 43:1455–1462. https://doi.org/10.1097/brs.0000000000002651

12. Engquist M, Löfgren H, Öberg B, Holtz A, Peolsson A, Söderlund A, Vavruch L, Lind B (2013) Surgery versus nonsurgical treatment of cervical radiculopathy: a prospective, randomized study comparing surgery plus physiotherapy with physiotherapy alone with a 2-year follow-up. Spine (Phila Pa 1976) 38:1715–1722. https://doi.org/10.1097/BRS.0b013e31829ff095

13. Gao X, Yang Y, Liu H, Meng Y, Zeng J, Wu T, Hong Y (2018) A comparison of cervical disc arthroplasty and anterior cervical discectomy and fusion in patients with two-level cervical degenerative disc disease: 5-year follow-up results. World Neurosurg. https://doi.org/10.1016/j.wneu.2018.10.231

14. Gennari A, Mazas S, Coudert P, Gille O, Vital JM (2018) Outpatient anterior cervical discectomy: a French study and literature review. Orthop Traumatol Surg Res 104:581–584. https://doi.org/ 10.1016/j.otsr.2018.04.014

15. Godlewski B, Stachura MK, Twardowska-Staszek E, Czepko RA, Czepko R (2018) Effect of social factors on surgical outcomes in cervical disc disease. Anesth Pain Med 8:e84140. https://doi.org/ 10.5812/aapm.84140

16. Goldstein ZH, Boody B, Sasso R (2020) Two-level anterior cervical discectomy and fusion versus cervical disc arthroplasty-long-term evidence update. Int J Spine Surg 14:36–40. https://doi.org/ 10.14444/7089

17. Gornet MF, Lanman TH, Burkus JK, Dryer RF, McConnell JR, Hodges SD, Schranck FW (2019) Two-level cervical disc arthroplasty versus anterior cervical discectomy and fusion: 10-year outcomes of a prospective, randomized investigational device exemption clinical trial. J Neurosurg Spine 21:1–11. https://doi. org/10.3171/2019.4.SPINE19157

18. Gruskay JA, Fu M, Basques BA, Bohl DD, Buerba RA, Webb ML, Grauer JN (2016) Factors affecting length of stay and complications after elective anterior cervical discectomy and fusion: a study of 2164 patients from the American College of Surgeons National Surgical Quality Improvement Project Database (ACS NSQIP). Clin Spine Surg 29:E34-42. https://doi.org/ 10.1097/bsd.0000000000000080

19. He J, Xiong W, Li F, Luo W, Gao SC (2017) Depression influences pain and function after cervical disc arthroplasty. J Neurosurg Sci 61:39–45. https://doi.org/10.23736/s0390-5616.16. 03032-0

20. Helseth Ø, Lied B, Heskestad B, Ekseth K, Helseth E (2019) Retrospective single-centre series of 1300 consecutive cases of outpatient cervical spine surgery: complications, hospital readmissions, and reoperations. Br J Neurosurg 33:613–619. https://doi. org/10.1080/02688697.2019.1675587

21. Hermansen A, Hedlund R, Vavruch L, Peolsson A (2013) Positive predictive factors and subgroup analysis of clinically relevant improvement after anterior cervical decompression and fusion for cervical disc disease: a 10- to 13-year follow-up of a prospective randomized study: clinical article. J Neurosurg Spine 19:403–411. https://doi.org/10.3171/2013.7.Spine12843

22. Hojmark K, Stottrup 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 25:282–286. https://doi.org/10.1007/ s00586-015-4127-3

23. McNeely EL, Sachdev R, Rahman R, Zhang B, Skolasky RL (2021) Associations of depression and sociodemographic characteristics with patient activation among those presenting for spine surgery. J Orthop 26:8–13. https://doi.org/10.1016/j.jor.2021.06. 001

24. Minhas SV, Mazmudar AS, Patel AA (2017) Pre-operative functional status as a predictor of morbidity and mortality after elective cervical spine surgery. Bone Joint J 99-b:824–828. https://doi. org/10.1302/0301-620x.99b6.Bjj-2016-1149.R1

25. Mjåset C, Zwart J-A, Goedmakers CMW, Smith TR, Solberg TK, Grotle M (2020) Criteria for success after surgery for cervical radiculopathy; estimates for a substantial amount of improvement in core outcome measures. Spine J 20:1413–1421. https://doi.org/ 10.1016/j.spinee.2020.05.549

26. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF, Collins GS (2015) Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 162:1–73. https://doi.org/10.7326/ m14-0698

27. Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, Reitsma JB, Kleijnen J, Mallett S (2019) PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. Ann Intern Med 170:1–33. https://doi.org/10.7326/m18-1377

28. Narain AS, Hijji FY, Haws BE, Khechen B, Kudaravalli KT, Yom KH, Singh K (2020) Risk factors for medical and surgical complications after 1-2- level anterior cervical discectomy and fusion procedures. Int J Spine Surg 14:286–293. https://doi.org/ 10.14444/7038

29. Narain AS, Hijji FY, Khechen B, Haws BE, Patel DV, Bohl DD, Yom KH, Kudaravalli KT, Singh K (2019) Risk factors associated with failure to reach minimal clinically important difference in patient-reported outcomes following anterior cervical discectomy and fusion. Int J Spine Surg 13:262–269. https://doi.org/10.14444/ 6035

30. Passias PG, Hasan S, Radcliff K, Isaacs R, Bianco K, Jalai CM, Poorman GW, Worley NJ, Horn SR, Boniello A, Zhou PL, Arnold PM, Hsieh P, Vaccaro AR, Gerling MC (2018) Arm pain versus neck pain: a novel ratio as a predictor of post-operative clinical

outcomes in cervical radiculopathy patients. Int J Spine Surg 12:629–637. https://doi.org/10.14444/5078

31. Peolsson A, Vavruch L, Oberg B (2006) Predictive factors for arm pain, neck pain, neck specific disability and health after anterior cervical decompression and fusion. Acta Neurochir 148:167–173. https://doi.org/10.1007/s00701-005-0660-x

32. Radhakrishnan K, Litchy WJ, O'Fallon WM, Kurland LT (1994) Epidemiology of cervical radiculopathy. A population-based study from Rochester, Minnesota, 1976 through 1990. Brain 117:325–335. https://doi.org/10.1093/brain/117.2.325

33. Riley RD, Hayden JA, Steyerberg EW, Moons KGM, Abrams K, Kyzas PA, Malats N, Briggs A, Schroter S, Altman DG, Hemingway H, Group P (2013) Prognosis research strategy (PROGRESS) 2: prognostic factor research. PLoS Med 10:e1001380–e1001380. https://doi.org/10.1371/journal.pmed.1001380

34. Scerrati A, Germano A, Montano N, Visani J, Cacciola F, Raffa G, Ghetti I, Pignotti F, Cavallo MA, Olivi A, de Bonis P (2021) Factors affecting functional outcome after anterior cervical discectomy and fusion: a multicenter study. J Craniovertebr Junction Spine 12:144–148. https://doi.org/10.4103/jcvjs.jcvjs_1_21

35. Sielatycki JA, Chotai S, Kay H, Stonko D, McGirt M, Devin CJ (2016) Does obesity correlate with worse patient-reported outcomes following elective anterior cervical discectomy and fusion? Neurosurgery 79:69–74. https://doi.org/10.1227/neu.0000000000001252

36. Solberg T, Olsen L, Berglund M (2019) Nasjonalt kvalitetsregister for ryggkirugi (NKR). Årsrapport for 2018 med plan for forbedringstiltak https://unn.no/Documents/Kvalitetsregistre/Nasjonalt%20kvalitetsregister%20for%20ryggkirurgi/Årsrapporter/Årsrapport_NKR_2018%20.pdf. Accessed 5 May 2021

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

38. Steyerberg E (2009) Clinical prediction models: a practical approach to development, validation, and updating. Spinger-Verlag New York, New York, NY, Statistics for Biology and Health. https://doi.org/10.1007/978-0-387-77244-8

39. Steyerberg EW, Harrell FE Jr (2016) Prediction models need appropriate internal, internal-external, and external validation. J Clin Epidemiol 69:245–247. https://doi.org/10.1016/j.jclinepi.2015.04.005

40. Steyerberg EW, Moons KGM, van der Windt DA, Hayden JA, Perel P, Schroter S, Riley RD, Hemingway H, Altman DG, Group P (2013) Prognosis research strategy (PROGRESS) 3: prognostic model research. PLoS Med 10:e1001381–e1001381. https://doi.org/10.1371/journal.pmed.1001381

41. Steyerberg EW, Vergouwe Y (2014) Towards better clinical prediction models: seven steps for development and an ABCD for validation. Eur Heart J 35:1925–1931. https://doi.org/10.1093/eurheartj/ehu207

42. Sundseth J, Fredriksli OA, Kolstad F, Johnsen LG, Pripp AH, Andresen H, Myrseth E, Muller K, Nygaard OP, Zwart JA (2017) The Norwegian Cervical Arthroplasty Trial (NORCAT): 2-year clinical outcome after single-level cervical arthroplasty versus fusion-a prospective, single-blinded, randomized, controlled multicenter study. Eur Spine J 26:1225–1235. https://doi.org/10.1007/s00586-016-4922-5

43. Terwee CB, Roorda LD, Knol DL, De Boer MR, De Vet HC (2009) Linking measurement error to minimal important change of patient-reported outcomes. J Clin Epidemiol 62:1062–1067. https://doi.org/10.1016/j.jclinepi.2008.10.011

44. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, Poole C, Schlesselman JJ, Egger M (2007) Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration. PLoS Med 4:e297. https://doi.org/10.1371/journal.pmed.0040297

45. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, Reitsma JB, Kleijnen J, Mallett S (2019) PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med 170:51–58. https://doi.org/10.7326/m18-1376

46. Xie L, Liu M, Ding F, Li P, Ma D (2016) Cervical disc arthroplasty (CDA) versus anterior cervical discectomy and fusion (ACDF) in symptomatic cervical degenerative disc diseases (CDDDs): an updated meta-analysis of prospective randomized controlled trials (RCTs). Springerplus 5:1188. https://doi.org/10.1186/s40064-016-2851-8

47. Zhang GA, Zhang WP, Chen YC, Hou Y, Qu W, Ding LX (2020) Impact of elevated body mass index on surgical outcomes for patients undergoing cervical fusion procedures: a systematic review and meta-analysis. Orthop Surg 12:3–15. https://doi.org/10.1111/os.12572

Enhancing Quality in Surgery for Cervical Degenerative Disorders: Benchmarks for Clinical Improvement and Prognostic Models for Nonsuccess

Doctoral thesis by
Christer Mjåset

Institute of Clinical Medicine, Faculty of Medicine, University of Oslo

Department of Research and Innovation, Division of Clinical Neuroscience, Oslo University Hospital

Department of Neurosurgery, Oslo University Hospital

June 2023