



Responsiveness of the Oswestry Disability Index and Zurich Claudication Questionnaire in patients with lumbar spinal stenosis: evaluation of surgically treated patients from the NORDSTEN study

Kari Indrekvam^{1,2} · Tor Åge Myklebust^{3,4} · Ivar Magne Austevoll^{1,2} · Erland Hermansen^{1,3,5} · Hasan Banitalebi^{6,7} · Ingrid Fjeldheim Bånerud⁸ · Clemens Weber^{9,10} · Helena Brisby^{11,12} · Jens Ivar Brox^{13,14} · Christian Hellum¹⁵ · Kjersti Storheim^{8,16}

Received: 24 April 2024 / Revised: 24 April 2024 / Accepted: 30 July 2024 / Published online: 13 August 2024
© The Author(s) 2024

Abstract

Purpose To evaluate the responsiveness of the original low back pain specific Oswestry Disability Index (ODI) and the spinal stenosis specific Zürich Claudication Questionnaire (ZCQ), and to investigate cut-off values for clinical “success” for ODI and ZCQ in surgically treated patients with lumbar spinal stenosis (LSS).

Methods We included 601 LSS patients (218 with, 383 without degenerative spondylolisthesis) from the NORDSTEN trials. Outcome measures included ODI and ZCQ (symptom severity and physical function scales) with three alternative response parameters: scores at follow-up, absolute and relative changes from baseline to two-year follow-up. Effect size and standardised response mean evaluated internal responsiveness. External responsiveness was assessed by the Spearman rank correlation between patient-reported global perceived effect scale (GPE) and ODI and ZCQ, and receiver operating characteristics (ROC). We evaluated which cut-off values could maximise the percentage of correctly classified patients according to the GPE-anchor “completely recovered” / “much improved” for each parameter.

Results Internal and external responsiveness were high for all three indices with effect sizes, standardized response means, ROC and corresponding area under the curve > 0.8. Correlations with GPE responses were moderate (> 0.50) for absolute change and strong (> 0.67) for relative change and follow-up scores. The 30% ODI relative change cut-off correctly classified 81% of patients to “success”, within a range of accurate cut-offs according to the GPE-anchor.

Conclusion ODI and ZCQ demonstrate comparable responsiveness in evaluating outcomes for surgically treated LSS patients. The 30% ODI threshold was consistent with treatment “success” in NORDSTEN trials.

Trial registration ClinicalTrials.gov; NCT02007083 10/12/2013, NCT02051374 31/01/2014 and NCT03562936 20/06/2018.

Keywords Lumbar spinal stenosis · Degenerative spondylolisthesis · NORDSTEN-study · Responsiveness ODI and ZCQ · Cut-off values

Abbreviations

ASA	American society of Anesthesiologists classification	IQR	Interquartile range
AUC	Area under the curve	LSS	Lumbar spinal stenosis
DS	Degenerative spondylolisthesis trial	NORDSTEN	NORwegian degenerative spondylolisthesis and spinal STENosis study
FORMI	Research and communication unit for musculoskeletal health	ODI	Oswestry Disability Index
GPE	Global perceived effect	PROM	Patient reported outcome measures
		ROC	The receiver operating characteristics
		SD	Standard deviation

Communicated by NORWAY.

Extended author information available on the last page of the article

Spearman Rho	Spearman's rank correlation
SRM	standardized response mean
SST	Spinal Stenosis Trial
ZCQ	Zurich claudication questionnaire

Introduction

Lumbar spinal stenosis (LSS) is a common disease in the adult spine caused by degenerative changes narrowing the lumbar spinal canal, causing compression of neurovascular structures. It can occur with or without degenerative spondylolisthesis. LSS is associated with pain and reduced physical function, including walking problems, and is the most frequent reason for lumbar surgery increasing in numbers in the ageing spine [1, 2]. The optimal way of treating these patients non-surgically or/and surgically is still debated [3–6].

Reliable, valid, and responsive outcome measurements are cornerstones in evaluating the disease progress in patients regardless of treatment. There are two major aspects of responsiveness: “internal responsiveness”, characterizes the ability of a measure to change over a prespecified time frame, and “external responsiveness”, reflects the extent to which change in a measure relates to a corresponding change in a reference measure of clinical or health status [7]. Patient reported outcome measures (PROMs) use various questions to reflect the severity of the condition or disability level. In patients with LSS, the most used PROM is the Oswestry Disability Index (ODI) [8, 9], designed to explain how back or leg pain affects the ability to function in everyday life. ODI has been widely used in studies of patients with lumbar spinal disorders treated both surgically and non-surgically. The Zurich Claudication Questionnaire (ZCQ), also named the Swiss Spinal Stenosis Scale [10], is developed specifically for patients with lumbar spinal stenosis. However, ZCQ is less investigated, and few studies in patients with LSS have applied both PROMs [9, 11].

This study aimed to evaluate the responsiveness of the ODI and ZCQ, and investigate the cut-off values in ODI and ZCQ used for defining clinical “success” for surgically treated patients with LSS.

Materials and methods

Study design

This is a secondary analysis of prospective data derived from the Norwegian Degenerative spondylolisthesis and spinal STENosis (NORDSTEN) multicentre study (18 public hospitals located in all regions of Norway). This current

study was based on baseline and two-year follow-up data from the surgically treated patients in the NORDSTEN-SST (patients were randomized to three different decompression techniques) and the NORDSTEN-DS (patients were randomized to surgical decompression alone or decompression with instrumental fusion). A detailed description of these trials results after two-year follow-up have been published previously [4, 5].

Participants

Inclusion criteria were clinical symptoms of LSS defined as neurogenic claudication or radiating pain into the lower limbs, not responding to at least three months of non-surgical treatment; radiological findings corresponding to the clinical findings; men and women age > 18 , ≤ 80 years. A single-level spondylolisthesis ≥ 3 mm gave participation in the NORDSTEN-DS. Exclusion criteria were former surgery at the level of stenosis; former fracture/fusion of the thoracolumbar spine; cauda equina syndrome or fixed complete motor deficit; ASA grade 4 or 5; lumbar scoliosis > 20 grades; stenosis in more than three lumbar levels; distinct symptoms in lower limbs due to other diseases. Inclusion and exclusion criteria for the NORDSTEN trials are previously described in detail [4, 5, 12]. Furthermore, to be included in the present study, patients must have responded to the ODI and ZCQ at inclusion and ODI, ZCQ and GPE scale at two-year follow-up (Fig. 1).

Patient-reported outcome measures

The primary outcome measure for the NORDSTEN trials is the Norwegian-validated version of ODI version 2.0 [13]. The Norwegian-validated version of ZCQ was a secondary outcome measure [14].

The ODI was proposed in 1980 [15] and was originally developed to assess chronic low back pain. The ODI version 2.0 [8] also includes leg symptoms. There are ten questions addressing different aspects of function. Each question is scored on a 6-point Likert response scale ranging from 0 to 5. An index (0-100) is calculated, where a higher score represents higher levels of disability.

The ZCQ was developed in 1996 [10] to measure symptoms associated with LSS specifically. The tool assigns a higher priority to leg symptoms than to low back pain. The ZCQ consists of three scales addressing symptom severity, physical function, and patient satisfaction, the latter only relevant at follow-up. On a Likert response scale, each question scores from 1 to 4 or 5. Each scale score is calculated as a mean of all answered questions and a higher score represents a higher level of disability.

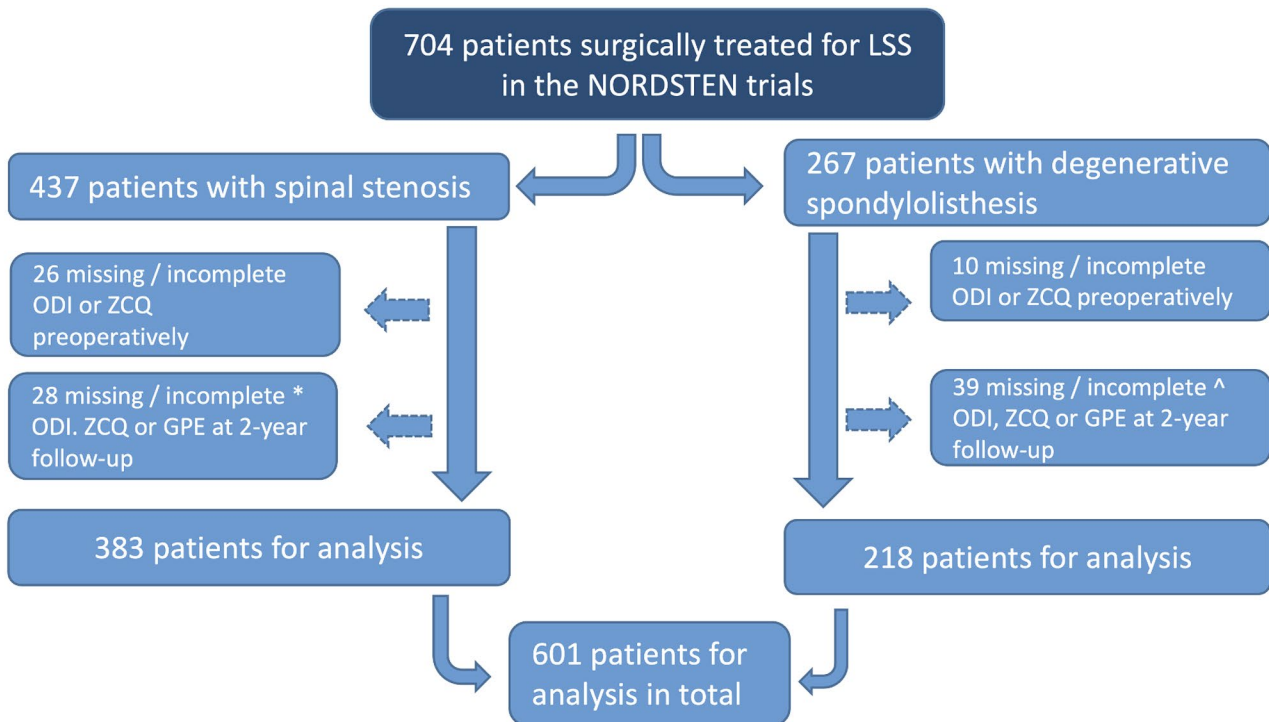


Fig. 1 Flowchart of the included surgically treated lumbar spinal stenosis patients from the NORDSTEN-trials. *LSS*: Lumbar spinal stenosis, *NORDSTEN*: NORwegian degenerative spondylolisthesis and spinal STENosis, *ODI*: Oswestry Disability Index, *ZCQ*: Zürich claudica-

tion questionnaire, *GPE*: Global perceived effect scale. * = included two dead and two withdrawn consents, ^ = included three dead and one withdrawn consent

Two years after surgery, the patients also responded to the Norwegian-validated version of the GPE scale [16] to estimate patient-reported overall judgement of the outcome of the surgery. There were seven response choices: ‘completely recovered’, ‘much improved’, ‘slightly improved’, ‘unchanged’, ‘slightly worse’, ‘much worse’, or ‘worse than ever’.

Statistical analysis

Standard descriptive statistics were presented using mean and standard deviation (SD) or median and interquartile range (IQR) for continuous variables and absolute frequencies and percentage distribution for categorical variables. In addition, histograms and box plots were used to illustrate the distribution of outcome measures, also in relation to the GPE scale. Internal responsiveness was evaluated using effect size and standardized response mean (SRM). The effect size was calculated as the mean difference between pre- and postoperative outcome measures divided by the standard deviation of the measure preoperatively. SRM was defined as the mean change divided by the standard deviation of the change and was calculated for both absolute and relative change in outcome measures. Effect sizes and SRMs greater than 0.80 were considered large. External responsiveness

was evaluated by estimating Spearman’s rank correlation between outcome measures and the GPE scale. Correlation coefficients in the range of 0.5–0.67 were considered moderate, and larger correlation coefficients were considered strong. External responsiveness was also evaluated using receiver operating characteristics (ROC) curves and corresponding area under the curve (AUC) using the GPE scale dichotomized into “success” vs “non-success” (‘completely recovered’/‘much improved’ vs. remaining categories) as an external anchor. The higher the AUC, the higher the ability of a continuous measurement to discriminate patients into a successful or non-successful outcome after treatment. AUCs were considered moderate (0.70–0.79), high (0.80–0.89) or excellent (≥ 0.90). We evaluated the sensitivity, specificity and correct classification rates of all possible cut-offs to evaluate clinically appropriate cut-off values. All analyses were done using Stata version 17.1.

Results

In the NORDSTEN study, 704 LSS patients were included and treated surgically between February 2014 and September 2018: 437 in the NORDSTEN-SST and 267 in the NORDSTEN-DS trial [4, 5, 12]. The present analysis

Table 1 The demographic data, general health condition, and clinical scores at baseline in the NORDSTEN-study for the surgically treated lumbar spinal stenosis patients

	Total (n = 601)	SST (n = 383)	DS (n = 218)
Demographics			
Age - median (IQR)	68 (61–73)	68 (62–73)	67 (61–72)
Female gender - N (%)	328 (54.6)	181 (47.3)	147 (67.4)
Higher education - N (%)	167 (27.8)	112 (29.2)	55 (25.2)
Married/partner - N (%)	455 (75.7)	291 (76.0)	164 (75.2)
Smoker (yes) - N (%)	114 (19.0)	76 (19.8)	38 (17.4)
BMI - mean (SD)	27.7 (4.2)	27.7 (4.1)	27.8 (4.4)
General health condition			
ASA 1 - N (%)	70 (11.7)	44 (11.5)	26 (11.9)
ASA 2 - N (%)	409 (68.1)	256 (66.8)	153 (70.1)
ASA 3 - N (%)	109 (18.1)	77 (20.1)	32 (14.7)
Clinical scores			
ODI mean (SD)	38.5 (14.1)	38.1 (14.4)	39.3 (13.5)
ZCQ-symptom mean (SD)	3.36 (0.56)	3.36 (0.56)	3.34 (0.58)
ZCQ-function mean (SD)	2.50 (0.53)	2.51 (0.54)	2.50 (0.52)

SST: NORDSTEN spinal stenosis trial, DS: NORDSTEN degenerative spondylolisthesis trial, N: Numbers, IQR: Interquartile range, SD: Standard deviation, BMI: Body mass index, ASA: American Society of Anaesthesiologists' score: 1 indicates no disease, 2 mild systemic disease, and 3 severe systemic disease that is not life-threatening, ODI: Oswestry disability index (0–100) a higher score represents higher levels of disability, ZCQ-symptom: Zürich claudication questionnaire symptom severity (1–5) a higher score represents a higher level of disability, ZCQ-function: Zürich claudication questionnaire physical function (1–4) a higher score represents a higher level of disability

excluded 103 patients due to missing or incomplete ODI, ZCQ, or GPE scale preoperatively and/or at two-year follow-up. Consequently, 601 (85%) patients were included in the present study (Fig. 1).

The demographic data and patients' characteristic at baseline are shown in Table 1.

The outcome two years after surgery is shown in Table 2. Patients rated outcome as 'completely recovered'/'much improved' in 392 (65.3%) cases.

Figure 2 shows the distribution of preoperative scores and scores at two-year follow-up for ODI and ZCQ. It also shows the absolute and relative PROM changes from surgery until two-year follow-up. Data for the NORDSTEN-SST and NORDSTEN-DS separately are found in supplementary information Fig. 1.

ODI and ZCQ scores after two-year follow-up, as well as absolute and relative changes from baseline, are summarized in Table 2 and graphically described within each patient group based on response to the GPE scale in Fig. 3. Data for the NORDSTEN-SST and NORDSTEN-DS separately are found in supplementary information Fig. 2.

Table 2 Outcome after surgery for lumbar spinal stenosis assessed by patients at two-year follow-up in the NORDSTEN trials: global perceived effect scale (GPE scale), Oswestry disability index (ODI), and Zurich claudication questionnaire (ZCQ)

	Total (n = 601)	SST (n = 383)	DS (n = 218)
Outcome measures after 2years - mean (SD)			
ODI	18.9 (16.3)	19.0 (16.5)	18.6 (15.9)
ZCQ-symptom severity	2.35 (0.86)	2.34 (0.88)	2.36 (0.84)
ZCQ-physical function	1.67 (0.66)	1.66 (0.65)	1.69 (0.68)
Absolute change in outcome measures - mean (SD)			
ODI	-19.7 (16.5)	-19.1 (17.0)	-20.7 (15.6)
ZCQ-symptom severity	-1.01 (0.81)	-1.02 (0.81)	-0.99 (0.81)
ZCQ-physical function	-0.83 (0.67)	-0.85 (0.67)	-0.80 (0.67)
Relative change (%) in outcome measures - mean (SD)			
ODI	-51.2 (39.6)	-50.1 (41.3)	-53.3 (36.6)
ZCQ-symptom severity	-42.8 (37.4)	-44.0 (34.8)	-40.6 (41.6)
ZCQ-physical function	-56.4 (42.3)	-57.3 (42.0)	-54.9 (43.1)
GPE scale - N (%)			
Completely recovered	135 (22.5)	89 (23.2)	46 (21.1)
Much improved	257 (42.8)	157 (41.0)	100 (45.9)
Slightly improved	106 (17.6)	67 (17.5)	39 (17.9)
Unchanged	50 (8.3)	35 (9.1)	15 (6.9)
Slightly worse	20 (3.3)	14 (3.7)	6 (2.8)
Much worse	26 (4.3)	14 (3.7)	12 (5.5)
Worse than ever	7 (1.2)	7 (1.8)	0 (0.0)

SST: NORDSTEN spinal stenosis trial, DS: NORDSTEN degenerative spondylolisthesis trial, N: Numbers, SD Standard deviations, ODI: Oswestry disability index (0–100) a higher score represents higher levels of disability, ZCQ-symptom severity: Zürich claudication questionnaire symptom severity (1–5) a higher score represents a higher level of disability, ZCQ-physical function: Zürich claudication questionnaire physical function (1–4) a higher score represents a higher level of disability, and GPE: Global perceived effect scale

Internal responsiveness

Effect sizes and SRMs were large (>0.8) and similar across all outcome measures. Effect sizes ranged from -1.39 for ODI and -1.79 for ZCQ symptom severity score. In comparison, SRM ranged from -1.19 for absolute change in ODI and -1.25 for absolute change in ZCQ symptom severity score (Table 3).

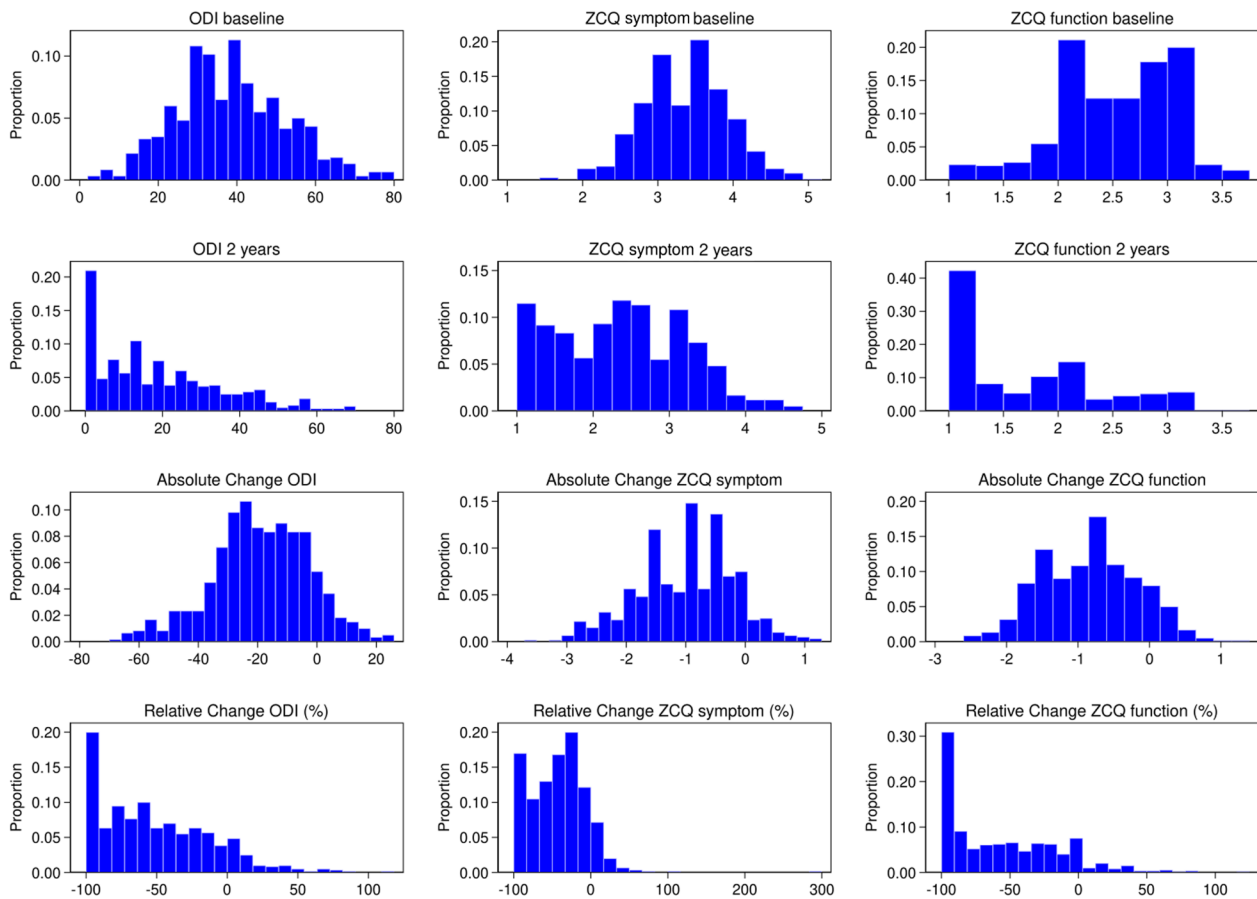


Fig. 2 Histograms showing the distribution of preoperative scores, follow-up scores and changes (absolute and relative) in scores for Oswestry disability index (ODI) and Zurich claudication questionnaire (ZCQ) symptom severity and physical function in patients surgically treated for lumbar spinal stenosis in the NORDSTEN trials. *ODI*: Oswestry

disability index (0–100) a higher score represents higher levels of disability, *ZCQ-symptom*: Zürich claudication questionnaire symptom severity (1–5) a higher score represents a higher level of disability, *ZCQ-function*: Zürich claudication questionnaire physical function (1–4) a higher score represents a higher level of disability

External responsiveness

All outcome measurements correlated significantly to the GPE scale. ODI, ZCQ symptoms severity, and -physical function all had strong or close to strong Spearman's correlation coefficients (≥ 0.67) for follow-up scores and relative change score (Table 3), whereas it was moderate (≥ 0.50) for absolute change score.

ODI, ZCQ symptom severity, and -physical function all had high test accuracy (AUC from 0.8 to 0.9) in ROC curves (Table 3; Fig. 4), but somewhat higher for the follow-up and relative change scores than for the absolute change scores. Data for the NORDSTEN-SST and NORDSTEN-DS separately are found in supplementary information Fig. 3.

Cut-off values for defining clinical “success”

The correct classification rate for all possible cut-off values and the corresponding sensitivity and specificity for

defining “success”/“non-success” for follow-up scores, absolute change scores, and relative change scores are illustrated in Fig. 5. For relative change scores with varying sensitivity and specificity, all ODI cut-offs between 27.5% and 49% did correctly classify 81 to 82 per cent of the patients. For ZCQ symptom severity, the correct classification score was 78–79 per cent in the interval between 30 and 38%, and for ZCQ physical function, 80–81 per cent in the interval between 35 and 50% relative change. The figure illustrates that regardless of chosen combination of sensitivity and specificity, the maximum accuracy was lower for the absolute change score than for the follow-up score and the relative change score.

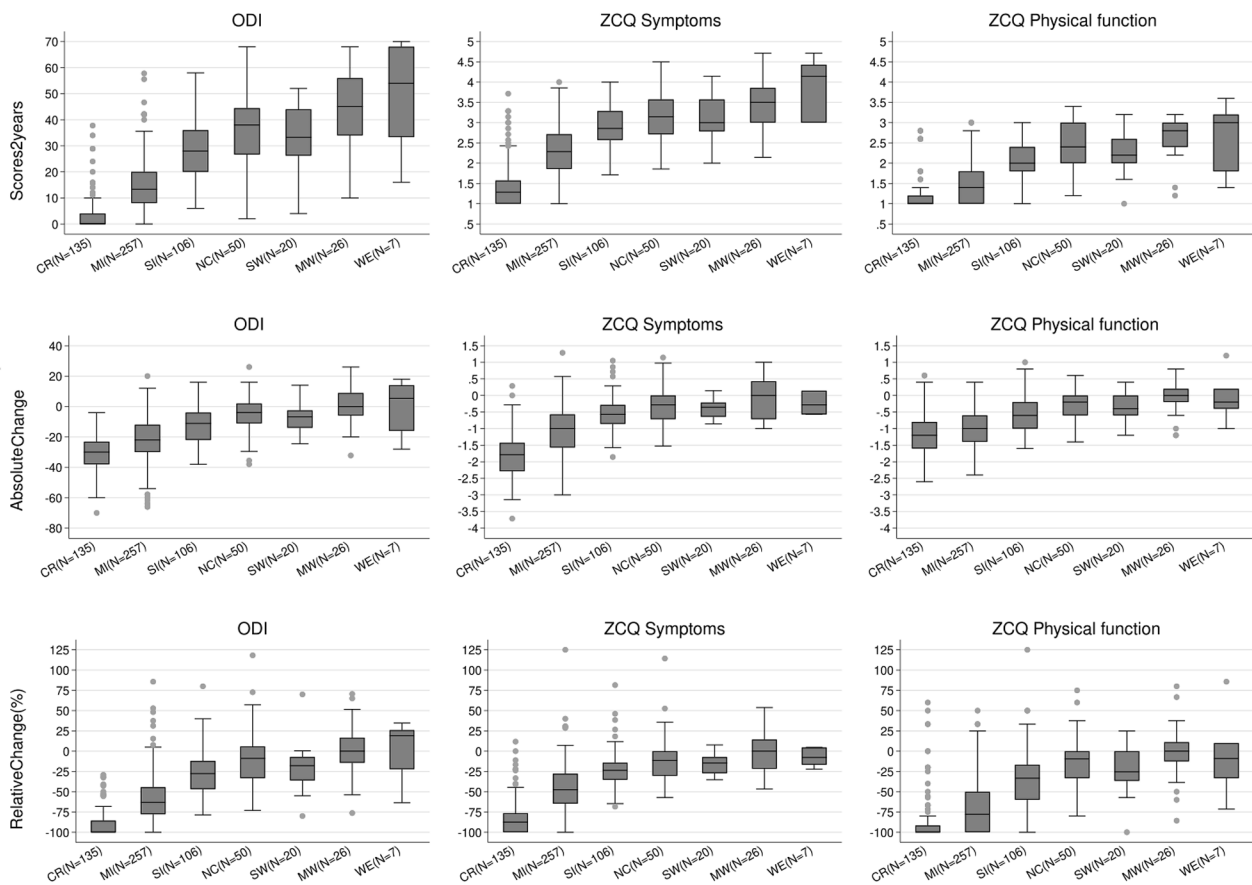


Fig. 3 The follow-up scores, absolute change scores, and relative change scores at two-year-follow-up for Oswestry disability index (ODI) and Zurich claudication questionnaire (ZCQ) symptom severity and physical function in each Global Perceived effect scale (GPE) group for patients surgically treated for lumbar spinal stenosis in the NORDSTEN trials. *GPE* scale groups: *CR* (Completely recovered), *MI* (Much improved), *SI* (Slightly improved), *NC* (Not changed),

SW (Slightly worse), *MW* (Much worse), *WE* (Worse than ever), *ODI*: Oswestry disability index (0-100) a higher score represents higher levels of disability, *ZCQ-symptoms*: Zürich claudication questionnaire symptom severity (1–5) a higher score represents a higher level of disability, *ZCQ-physical function*: Zürich claudication questionnaire physical function (1–4) a higher score represents a higher level of disability

Table 3 Responsiveness for Oswestry disability index (ODI) and Zurich claudication questionnaire (ZCQ) for surgically treated lumbar stenosis patients. Internal responsiveness was assessed by effect size and standard reponse mean (SRM), and external responsiveness by Spearman rank correlation coefficient, and the receiver operating characteristics area under curve (ROC AUC) using the global perceived effect scale (GPE) as the external criterion

	ODI	ZCQ symptom severity	ZCQ physical function
Effect size	-1.39	-1.79	-1.58
SRM (absolute change)	-1.19	-1.25	-1.24
SRM (relative change)	-1.29	-1.14	-1.33
Spearman Rho (follow-up score)	0.77	0.72	0.69
Spearman Rho (absolute change)	0.58	0.64	0.51
Spearman Rho (relative change)	0.76	0.71	0.67
ROC AUC (follow-up score)	0.90	0.87	0.87
ROC AUC (absolute change)	0.81	0.82	0.80
ROC AUC (relative change)	0.89	0.86	0.87

Discussion

In the present study, ODI and ZCQ showed good responsiveness for assessing clinical outcomes in patients treated surgically for LSS with or without degenerative spondylolisthesis. The internal responsiveness was good for both tested PROMs. The external responsiveness and the ability to discriminate between “success” and “non-success” were strong for follow-up and relative change scores, whereas they were moderate for absolute change scores. The 30% threshold for the ODI relative change used in the NORDSTEN trials was within the range of cut-off values with accuracy.

The two PROMs, ODI and ZCQ, are specific for spinal conditions. Since ODI was originally developed for patients with low back pain, and ZCQ was developed for LSS patients, the finding of similar and good responsiveness for both in surgically treated LSS patients is interesting.

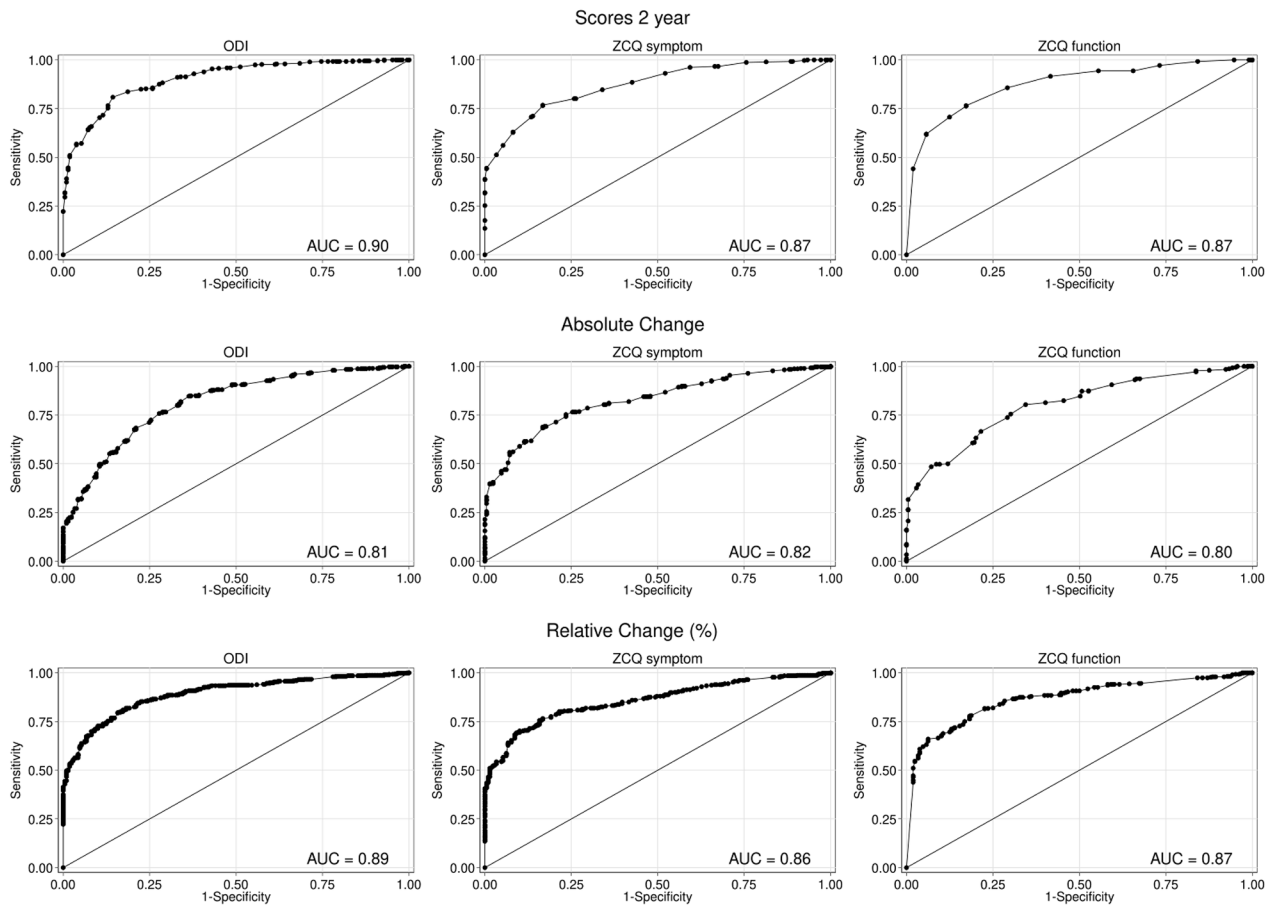


Fig. 4 The receiver operating characteristics (ROC) and area under the curve (AUC) using the Global Perceived effect scale (GPE) as the external criterion. The follow-up scores, absolute change scores, and

relative change scores for Oswestry disability index (ODI) and Zurich claudication questionnaire (ZCQ) for patients surgically treated for lumbar spinal stenosis in the NORDSTEN-trials

Follow-up score, absolute change score, and relative change score for the PROMs are alternative response parameters used in evaluations. The relative change (percentage) score has been recommended to account for the influence of the baseline score on the outcome score [17–20]. In the present study, the follow-up score and the relative change score performed better than the absolute change score, which is in accordance with a previous study from The Norwegian Spine registry [20].

Internal responsiveness was good. We found a large effect size and SRM for both ODI and ZCQ, of the same magnitude as Fujimori et al.'s investigation of patients operated for LSS at one-year follow-up [11]. In the present study, external responsiveness evaluated by Spearman correlation coefficients was higher than that reported by Fujimori et al. (coefficients around 0.50), whereas the ROC test accuracy was similar.

Contrary to a previously published article on this topic [11], we do not provide a definitive ranking of the instruments compared. Such rankings might be sensitive to random variation and misleading because one puts too much

confidence in one instrument being better than others. We intentionally focus on the similarities between ODI and ZCQ, both in numerical results and concerning clinical relevance and usefulness.

Formerly published cut-off values for ODI and ZCQ defining a clinical, minimal, or substantial important difference have been calculated with various methods and based on follow-up, absolute, or relative change scores. Different external anchors have been used; however, the patient's perceived global assessment of outcome or satisfaction (GPE-scales) is the most used. Patient response has been given on a five- or seven-point Likert scale, and the anchor has been the two or three best answer options. The calculations have been based on different study designs, such as clinical or register studies, and the time for follow-up has varied. In addition, most studies have used heterogeneous cohorts of various spine conditions. Therefore, a wide range of cut-off values have been proposed, and comparison is difficult. The present paper's results demonstrate that a range of cut-off values gave similar results for the proportion of correctly

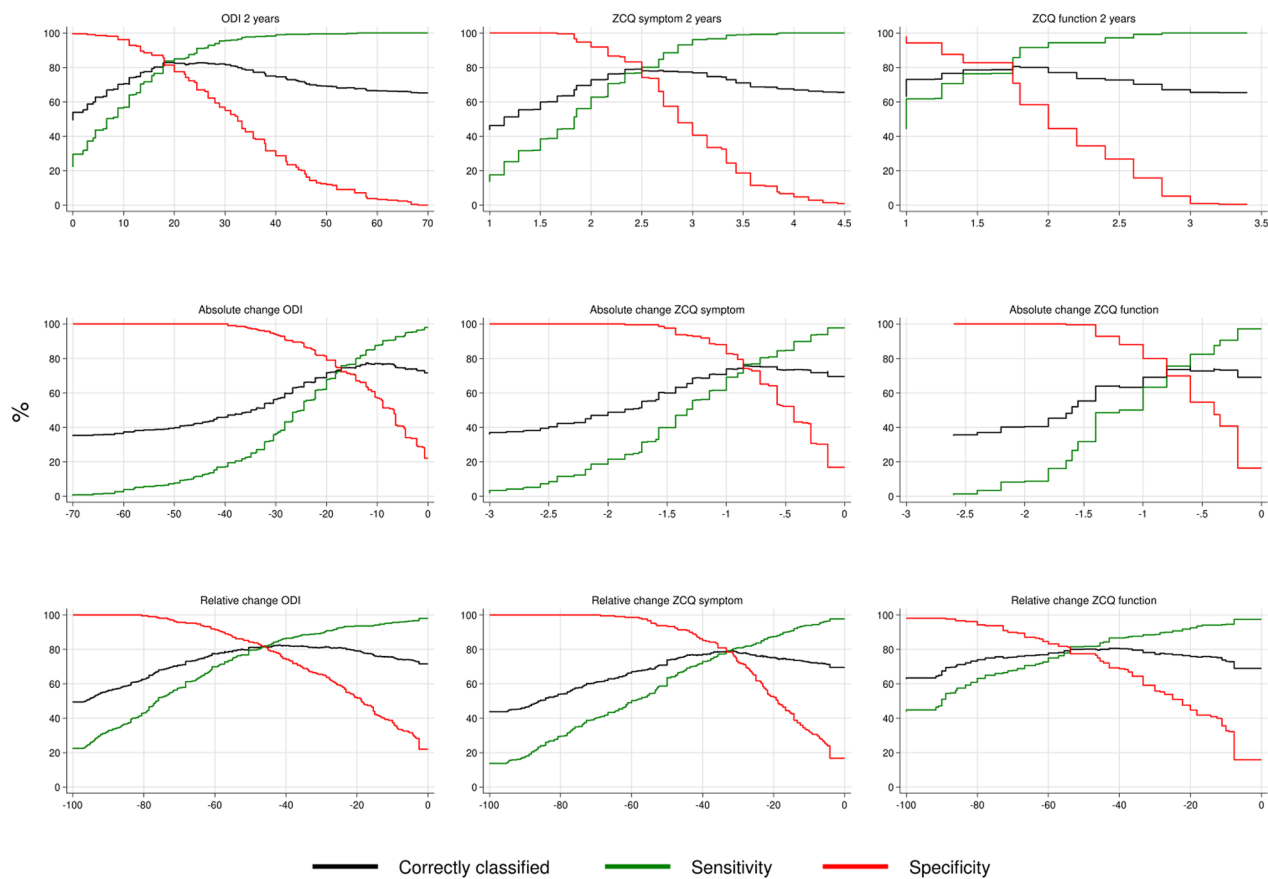


Fig. 5 Illustrating possible cut-off values for Oswestry disability index (ODI) and Zurich claudication questionnaire (ZCQ) symptom severity and - physical function defining “success” after surgical treatment of lumbar stenosis patients (“success”= Global Perceived effect scale

(GPE) “completely recovered” and “much improved”). The proportion correctly classified patients balanced against sensitivity and specificity. The curves for follow-up score, absolute change score and relative change score are shown

classified patients, and these must be balanced against sensitivity and specificity.

Previous reports of the clinically important cut-off values for ODI in surgically treated LSS patients were in the range of possible cut-off values reported in the present study both for follow-up scores [20, 21] and for absolute change scores [11, 17, 20]. A report about the cut-off value for absolute change score for ZCQ symptom severity and physical function [11] seemed to be lower than in the present study. The explanation might be that they used a five-point, not seven-point, Likert scale. In reports from others, the relative change for ODI has been suggested between 10 and 40% [19, 20]. The 30% threshold for defining treatment as a clinical “success” recommended and predefined in the NORSTEN trials, is based on a registry study with one-year follow-up [20] and a “gathering to consensus” paper [17], is within accuracy. In the registry study [20], the anchor was completely recovered/much improved. Because of the strict anchor (not including slightly improved), high sensitivity was favoured to ensure the detection of true positive “successes”. The present study showed that the 30% ODI cut-off

was within the interval giving a high correct classification rate. However, the percentage of correctly classified patients would also be high using higher cut-offs than 30%. When planning future comparative studies, one should consider not only using one cut-off but also performing sensitivity analyses using different cut-offs with high accuracy. It may also be reasonable that patients’ expectations are higher during a clinical study with follow-up than when they are part of a registry (more disappointed and answering “slightly improved”). In the registry, there will also be a more heterogeneous patient population.

Strength and limitations

The present analyses were based on a large cohort of surgical patients with a high follow-up rate. International guidelines for outcome measures were followed, along with translated and validated PROMs.

There is limited consensus about the best anchor for measuring changes in disease severity by PROMs. We selected the GPE scale since it has been commonly used

and recommended [18, 22]. There is advice to use a seven-point rating scale of change and setting the cut-off for clinically relevant improvements between “much improved” and “slightly improved” [19, 22]. The present paper’s GPE scale was about outcomes, but some studies have also used satisfaction [21]. Despite being commonly used, the GPE scale also has some weaknesses [16, 19, 20]; there is a possible recall bias in responding two years after the surgical intervention, the scale is domain unspecific, one does not know what kind of deterioration patients had in mind, or if other diseases were interfering, or if patients were more satisfied with care than treatment, and in addition, variation in mood may influence the patient’s response. Another critical concern is that in evaluating the PROMs (ODI, ZCQ), another subjective measurement (GPE scale) was used. In the present study, both tested PROMs correlated moderately to strongly with the GPE scale, as should be expected. Fujimori et al. found a discrepancy between the questionnaires’ improvement and the GPE scale [11]. Since they used a five-point Likert scale for GPE, it might be harder to reveal improvement. Recall bias and individual expectations of outcome may also play a role in a cultural frame.

Strictly, we did not ask the patients if the change was clinically important or a “success”. Still, we considered that the anchor answers, “completely recovered” and “much improved” were indicators of a significant improvement at follow-up. These concepts were discussed in some recent studies [23, 24].

Patients lost to follow-up represent a potential source of bias. In the present study, there was < 10% lost to follow-up, which made a high risk of bias unlikely. Also, a recent study based on data from the Norwegian Spine Registry showed that non-respondents had similar clinical outcomes [25].

These present analyses provided corroboration for the responsiveness of two commonly used outcome measurements in a large sample of patients with LSS treated surgically. Even though the responsiveness was comparable, when choosing an instrument for a study, one should remember that these PROMs were developed for different purposes. Our results for surgically treated LSS patients may not be reproduced if the patients, for instance, had some conservative treatment. ODI having as good responsiveness as ZCQ in the present study might be related to surgery preceding the observed change in scores for the included patients. Furthermore, ZCQ focuses on all symptoms in lower limbs and walking trouble, whereas ODI measures the influence of back and leg pain on daily life function.

Conclusion

ODI and ZCQ demonstrate comparable responsiveness in evaluating clinical outcomes for surgically treated LSS patients. To reflect “success” the follow-up and relative change score seems more accurate than the absolute change score. The 30% ODI threshold defining treatment success in NORDSTEN trials favoured sensitivity over specificity and was within accuracy.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00586-024-08440-1>.

Acknowledgements We would like to thank the patients who volunteered for this study, to Inger Ljostad, our patient representative, for her involvement in implementing the study, and to Katarina Mølsæter (Data and Safety Monitoring Board, Møre and Romsdal Health Trust) for her dedication and her diligent review of the data. We gratefully acknowledge the contributions of Anniken Remseth in managing the trial finances. From the NORDSTEN scientific board we thank Tore Solberg, Frode Rekeland, Eric Franssen, Oliver Grundnes, Jørn Aaen and Masoud Anwar. Thank you to the following members of the NORDSTEN working group: Turid Rognsvåg, Janne Haugland, Anita Karin Vassbakk, Vidar Punsvik, Kirstine Eikenes, Therese Gundersen, Trine Myrvold, Sara Sørøng, Silje Nilsen, Arild Hjulstad, Grete Ward, Turid Fjesme, Nikolaos Ikonou, Ove Bjørnstad, Yngve Sporstøl, Ted P. Lundgren, Anne-Charlotte Fosse Haug, Elisabeth Lilleholdt Muller, Kjartan Krogedal, Bettina Timenes, Vidar Opland, Merete Finjarn, Håvard Furunes, Hege Bergum Nilsen, Samer Habiba, Kristine Helland, Maria Rieber-Mohn, Andreas Seip, Greger Lønne, Gisle Szacinski, Espen Sjøberg, Ellen Aksnes, Øystein Nygaard, Hege Andresen, Andreas Sørli, Terje Fallås, Paal Arnesen, Kristine Evanger, Knut Jørgen Haug, Ingrid Edborg, Arnfinn Pedersen, Hege Westgard, Kamaran Raza, Odd Arild Ågedal and all other local study coordinators and surgeons not named here who contributed to the realization of this trial. A special thanks to the staff at The Research and Communication Unit for Musculoskeletal Health (FORMI), Oslo University Hospital, for meticulous work with data collection (Eira Kathleen Ebbs, Marie Skovli Pettersen, Ørjan Nesse Vigdal, Fiona Aanesen).

Author contributions KI, EH, IMA, JIB, CH, TÅM, KS contributed to the planning of the study; IFB contributed to collect the data; KI, TÅM and KS analysed the results, interpreted the results and wrote the draft of the manuscript. All authors contributed to revision of the draft and approved the final version to be published.

Funding Open access funding provided by University of Bergen (incl Haukeland University Hospital). This work has been supported by grants from The Western Norway Regional Health Authority, The Central Norway Regional Health Authority and the Møre and Romsdal Hospital Trust. The funding authorities were not involved in the study design, collection or analysis of data, interpretation of the results or drafting of the manuscript.

Open access funding provided by University of Bergen (incl Haukeland University Hospital)

Data availability The datasets generated during the current study are available from the corresponding author on reasonable request. Medical researchers’s request has to be in accordance with local registration and ethical approval, and datasets will be provided following review and approval of a research proposal and Statistical Analysis Plan (SAP). Data requests can be submitted at any time and the data will be accessible for

12 months, with possible extensions considered. All proposals requesting data access will need approval of the scientific board before any data can be released.

Declarations

Ethics approval and consent to participate The Norwegian Committees for Medical and Health Research Ethics approved the trials (NOR-DTEN-SST trial and NORDSTEN-OC; 2011/2034, NORDSTEN-DS trial; 2013/366). Data protection was approved by the Norwegian Data Inspectorate and data are stored at “Services for sensitive data” (TSD) hosted by the University of Oslo, to allow secure data sharing between researchers. Informed consent was obtained from patients after oral and written information. The study was performed according to the Helsinki Declaration and registered at ClinicalTrials.gov under the following identifiers; SST: NCT02007083 10/12/2013 / DS: NCT02051374 31/01/2014 / OC: NCT03562936 20/06/2018. The trial is monitored according to requirements of the Good Clinical Practice (GCP) guidelines.

Competing interests The authors declare that they have no financial or non-financial that are directly or indirectly related to the work submitted for publication.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Machado GC et al (2017) Trends, complications, and costs for hospital admission and surgery for lumbar spinal stenosis. *Spine (Phila Pa 1976)* 42(22):1737–1743
- Grøvle L et al (2019) The Rates of LSS Surgery in Norwegian public hospitals: a threefold increase from 1999 to 2013. *Spine (Phila Pa 1976)* 44(6):E372–e378
- Minamide A, Yoshida M, Maio K (2013) The natural clinical course of lumbar spinal stenosis: a longitudinal cohort study over a minimum of 10 years. *J Orthop Sci* 18(5):693–698
- Austevoll IM et al (2021) Decompression with or without Fusion in degenerative lumbar spondylolisthesis. *N Engl J Med* 385(6):526–538
- Hermansen E et al (2022) Comparison of 3 different minimally Invasive Surgical techniques for lumbar spinal stenosis: a Randomized Clinical Trial. *JAMA Netw Open* 5(3):e224291
- Zaina F et al Surgical versus non-surgical treatment for lumbar spinal stenosis. *Cochrane Database Syst Rev*, 2016(1): p. Cd010264.
- Husted JA et al (2000) Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol* 53(5):459–468
- Fairbank JC, Pynsent PB (2000) *The Oswestry disability index* *Spine* 2000.Nov 15;25(22.):2940.-53., 25(22): pp. 2940–2953
- Winebrake JP et al (2020) Wide variability in patient-reported outcomes measures after Fusion for lumbar spinal stenosis: a systematic review. *Global Spine J* 10(2):209–215
- Stucki G et al (1996) Measurement properties of a self-administered outcome measure in lumbar spinal stenosis. *Spine* 21(7):796–803
- Fujimori T et al (2022) Responsiveness of the Zurich Claudication Questionnaire, the Oswestry Disability Index, the Japanese Orthopaedic Association Back Pain evaluation questionnaire, the 8-Item short Form Health Survey, and the Euroqol 5 dimensions 5 level in the assessment of patients with lumbar spinal stenosis. *Eur Spine J* 31(6):1399–1412
- Indrekvam K et al (2023) The Norwegian degenerative spondylolisthesis and spinal stenosis (NORDSTEN) study: study overview, organization structure and study population. *Eur Spine J*
- Grotle M, Brox JI, Vollestad NK (2003) Cross-cultural adaptation of the Norwegian versions of the Roland-Morris disability questionnaire and the Oswestry Disability Index. *J Rehabil Med* 35(5):241–247
- Thornes E, Grotle M (2008) Cross-cultural adaptation of the Norwegian version of the spinal stenosis measure. *Eur Spine J* 17(3):456–462
- Fairbank JC et al (1980) The Oswestry low back pain disability questionnaire. *Physiotherapy* 66(8):271–273
- Grøvle L et al (2014) Patients’ ratings of global perceived change during 2 years were strongly influenced by the current health status. *J Clin Epidemiol* 67(5):508–515
- Ostelo RW et al (2008) Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. *Spine (Phila Pa 1976)* 33(1):90–94
- Dworkin RH et al (2008) Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. *J Pain* 9(2):105–121
- Bråten LCH et al (2022) Minimal important change was on the lower spectrum of previous estimates and responsiveness was sufficient for core outcomes in chronic low back pain. *J Clin Epidemiol* 151:75–87
- Austevoll IM et al (2019) Follow-up score, change score or percentage change score for determining clinical important outcome following surgery? An observational study from the Norwegian registry for spine surgery evaluating patient reported outcome measures in lumbar spinal stenosis and lumbar degenerative spondylolisthesis. *BMC Musculoskelet Disord* 20(1):31
- van Hooff ML et al (2016) Determination of the Oswestry Disability Index score equivalent to a satisfactory symptom state in patients undergoing surgery for degenerative disorders of the lumbar spine—a spine Tango registry-based study. *Spine J* 16(10):1221–1230
- Ostelo RW, de Vet HC (2005) Clinically important outcomes in low back pain. *Best Pract Res Clin Rheumatol* 19(4):593–607
- Taso M et al (2022) What is success of treatment? Expected outcome scores in cervical radiculopathy patients were much higher than the previously reported cut-off values for success. *Eur Spine J* 31(10):2761–2768
- Parai C et al (2020) ISSLS prize in clinical science 2020: the reliability and interpretability of score change in lumbar spine research. *Eur Spine J* 29(4):663–669
- Kaur S et al (2023) Characteristics and outcomes of patients who did not respond to a national spine surgery registry. *BMC Musculoskelet Disord* 24(1):164

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Kari Indrekvam^{1,2}  · Tor Åge Myklebust^{3,4} · Ivar Magne Austevoll^{1,2} · Erland Hermansen^{1,3,5} · Hasan Banitalebi^{6,7} · Ingrid Fjeldheim Bånerud⁸ · Clemens Weber^{9,10} · Helena Brisby^{11,12} · Jens Ivar Brox^{13,14} · Christian Hellum¹⁵ · Kjersti Storheim^{8,16}

✉ Kari Indrekvam
kari.indrekvam@helse-bergen.no

¹ Kysthospitalet in Hagevik, Orthopedic Clinic, Haukeland University Hospital, Bergen, Norway

² Department of Clinical Medicine, University of Bergen, Bergen, Norway

³ Department of Research and Innovation, Møre and Romsdal Hospital Trust, Ålesund, Norway

⁴ Department of Registration, Cancer Registry Norway, Oslo, Norway

⁵ Institute of Health Sciences, Norwegian University of Technology and Science, Ålesund, Norway

⁶ Department of Diagnostic Imaging, Akershus University Hospital, Nordbyhagen, Norway

⁷ Department of Clinical Medicine, University of Oslo, Oslo, Norway

⁸ Norway Research and Communication Unit for Musculoskeletal Health (FORMI), Oslo University Hospital, Oslo, Norway

⁹ Department of Neurosurgery, Stavanger University Hospital, Stavanger, Norway

¹⁰ Department of Quality and Health Technology, University of Stavanger, Stavanger, Norway

¹¹ Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

¹² Department of Orthopedics, Sahlgrenska University Hospital, Gothenburg, Sweden

¹³ Department of Physical Medicine and Rehabilitation, Oslo University Hospital, Oslo, Norway

¹⁴ Faculty of Medicine, University of Oslo, Oslo, Norway

¹⁵ Division of Orthopedic Surgery, Oslo University Hospital Ullevål, Oslo, Norway

¹⁶ Department of Physiotherapy, Oslo Metropolitan University, Oslo, Norway