The Spine Journal 21 (2021) 1489–1496

Clinical Study

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

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

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

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

Abstract

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

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

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

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

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

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

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

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

FDA. Device/drug status: Not applicable

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

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

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

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

1529-9430/© 2021 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Downloaded for Anonymous User (n/a) at Innlandet Hospital Trust from ClinicalKey.com by Elsevier on September 08, 2021. For personal use only. No other uses without permission. Copyright ©2021. Elsevier Inc. All rights reserved.

CONCLUSIONS: In this national register study, ODI derivates were most accurate to identify both failure and worsening after surgery for degenerative lumbar spinal stenosis. We recommend use of ODI percentage change and ODI final score for further studies of failure and worsening in elective spine surgery. © 2021 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

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

Introduction

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

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

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

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

Method

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

Patient population

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

The NORspine registry

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

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

Patient-reported outcome measures (PROMs) and reference

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

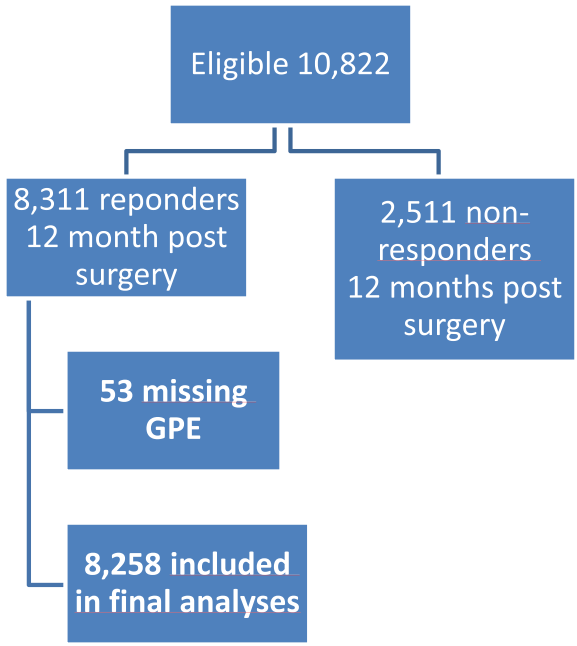

Fig. 1. Flowchart showing eligible patients, responders, non-responders, and those that could not be analyzed due to missing Global Perceived Effect (GPE) score. 8,258 patients were available for final analyses.

Downloaded for Anonymous User (n/a) at Innlandet Hospital Trust from ClinicalKey.com by Elsevier on September 08, 2021. For personal use only. No other uses without permission. Copyright ©2021. Elsevier Inc. All rights reserved.

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

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

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

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

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

Statistical analyses

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

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

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

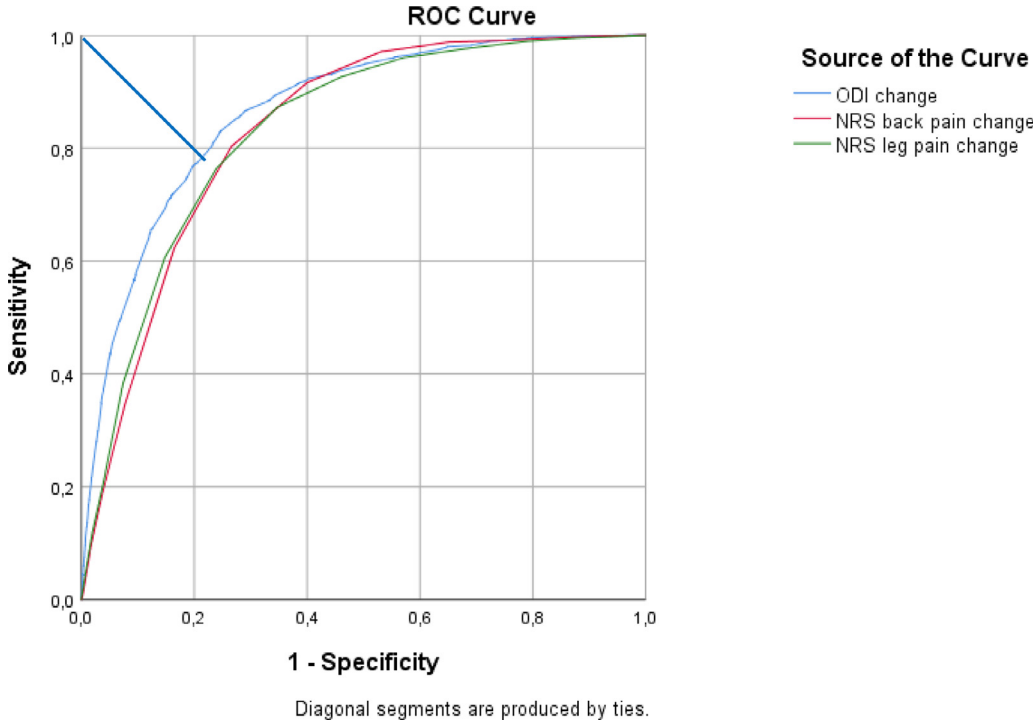

Fig. 2. ROC curves for absolute change in ODI, NRS back pain, and NRS leg pain vs. "failure" (GPE 4−7). The oblique blue line demonstrates "distance to corner" - a method to identify the highest sensitivity and specificity for each curve. The ROC curves also demonstrate "area under the curve" (AUC) -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

Downloaded for Anonymous User (n/a) at Innlandet Hospital Trust from ClinicalKey.com by Elsevier on September 08, 2021. For personal use only. No other uses without permission. Copyright ©2021. Elsevier Inc. All rights reserved.

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

Ethical considerations

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

Results

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

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

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

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

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

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

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

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

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

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

Downloaded for Anonymous User (n/a) at Innlandet Hospital Trust from ClinicalKey.com by Elsevier on September 08, 2021. For personal use only. No other uses without permission. Copyright ©2021. Elsevier Inc. All rights reserved.

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

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

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

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

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

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

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

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

Discussion

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

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

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

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

Downloaded for Anonymous User (n/a) at Innlandet Hospital Trust from ClinicalKey.com by Elsevier on September 08, 2021. For personal use only. No other uses without permission. Copyright ©2021. Elsevier Inc. All rights reserved.

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

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

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

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

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

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

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

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

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

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

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

Conclusions

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

Acknowledgements

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

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

Declarations of Competing Interests

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

Downloaded for Anonymous User (n/a) at Innlandet Hospital Trust from ClinicalKey.com by Elsevier on September 08, 2021. For personal use only. No other uses without permission. Copyright ©2021. Elsevier Inc. All rights reserved.

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

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

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

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

References

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

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

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

[4] Austevoll IM, Gjestad R, Grotle M, Solberg T, Brox JI, Hermansen E, et al. Follow-up score, change score or percentage change score for determining clinical important outcome following surgery? An observational study from the Norwegian registry for Spine surgery evaluating patient reported outcome measures in lumbar spinal stenosis and lumbar degenerative spondylolisthesis. BMC Musculoskelet Disord 2019;20:31. https://doi.org/10.1186/s12891-018-2386-y.

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

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

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

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

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

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

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

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

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

Downloaded for Anonymous User (n/a) at Innlandet Hospital Trust from ClinicalKey.com by Elsevier on September 08, 2021. For personal use only. No other uses without permission. Copyright ©2021. Elsevier Inc. All rights reserved.

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

[15] Kamper SJ, Ostelo RW, Knol DL, Maher CG, de Vet HC, Hancock MJ. Global perceived effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. J Clin Epidemiol 2010;63 (7):760–6 e1. https://doi.org/10.1016/j.jclinepi.2009.09.009.

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

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

[18] Grøvle L, Haugen AJ, Hasvik E, Natvig B, Brox JI, Grotle M, et al. Patients' ratings of global perceived change during 2 years were strongly influenced by the current health status. J Clin Epidemiol 2014;67(5):508–15. https://doi.org/10.1016/j.jclinepi.2013.12.001.

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

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

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

[22] Solberg T, Johnsen LG, Nygaard ØP, Grotle M. Can we define success criteria for lumbar disc surgery? estimates for a substantial amount of improvement in core outcome measures. Acta Orthop 2013;84(2):196–201. https://doi.org/10.3109/17453674.2013.786634.

[23] Ostelo RW, Deyo RA, Stratford P, Waddell G, Croft P, Von Kroff M, et al. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. Spine (Phila Pa 1976) 2008;33(1):90–4. https://doi.org/10.1097/BRS.0b013e31815e3a10.

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

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

[26] van Hooff ML, Mannion AF, Staub LP, Ostelo RW, Fairbank JC. Determination of the Oswestry Disability Index score equivalent to a "satisfactory symptom state" in patients undergoing surgery for degenerative disorders of the lumbar spine-a spine tango registry-based study. *Spine J*: 2016;16(10):1221–30. https://doi.org/10.1016/j.spinee.2016.06.010.

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

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

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

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

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

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

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

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

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

[36] Weinstein JN, Lurie JD, Tosteson TD, Zhao W, Blood EA, Tosteson ANA, et al. Surgical compared with nonoperative treatment for lumbar degenerative spondylolisthesis. four-year results in the spine patient outcomes research trial (SPORT) randomized and observational cohorts. J Bone Joint Surg Am 2009;91(6):1295–304. https://doi.org/10.2106/JBJS.H.00913.

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

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

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

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

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

Downloaded for Anonymous User (n/a) at Innlandet Hospital Trust from ClinicalKey.com by Elsevier on September 08, 2021. For personal use only. No other uses without permission. Copyright ©2021. Elsevier Inc. All rights reserved.