



Universiteit
Leiden
The Netherlands

Outcome after anterior cervical discectomy: from inferential statistics to Machine Learning

Goedmakers, C.M.W.

Citation

Goedmakers, C. M. W. (2023, December 20). *Outcome after anterior cervical discectomy: from inferential statistics to Machine Learning*. Retrieved from <https://hdl.handle.net/1887/3674247>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3674247>

Note: To cite this publication please use the final published version (if applicable).



Chapter 3

Criteria for success after surgery for cervical radiculopathy

Estimates for a substantial amount of improvement in core outcome measures

Mjåset C^{1,2,*}, Zwart JA^{7,8*}, Goedmakers CMW^{3,4}, Smith TR⁴,
Solberg TK^{5,6}, Grotle M^{9,10}

Spine J. 2020 Sep;20(9):1413-1421.

¹ Research and Communication Unit for Musculoskeletal Health (FORMI), Oslo University Hospital, Oslo, Norway, ² Department of Neurosurgery, Oslo University Hospital, Oslo, Norway, ³ Department of Neurosurgery, Leiden University Medical Center, Leiden, The Netherlands, ⁴ Computational Neuroscience Outcomes Center (CNOC), Department of Neurosurgery, Brigham and Women's Hospital, School of Medicine, Harvard University, Boston, MA, USA, ⁵ Faculty of Health Sciences, Oslo Metropolitan University, Oslo, Norway, ⁶ Institute of Clinical Medicine, The Arctic University of Norway (UiT), Tromsø, Norway, ⁷ Faculty of Medicine, University of Oslo, Oslo, Norway, ⁸ Research and Communication Unit for Musculoskeletal Health (FORMI), Oslo University Hospital, Oslo, Norway, ⁹ Department of Physiotherapy, Oslo Metropolitan University, Oslo, Norway, ¹⁰ Research and Communication Unit for Musculoskeletal Health (FORMI), Oslo University Hospital, Oslo, Norway. * Co-first author

Abstract

Background and 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.

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 (Figure 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 on 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) does not add up to the total number of patients (n=2,868) in Figure 1.

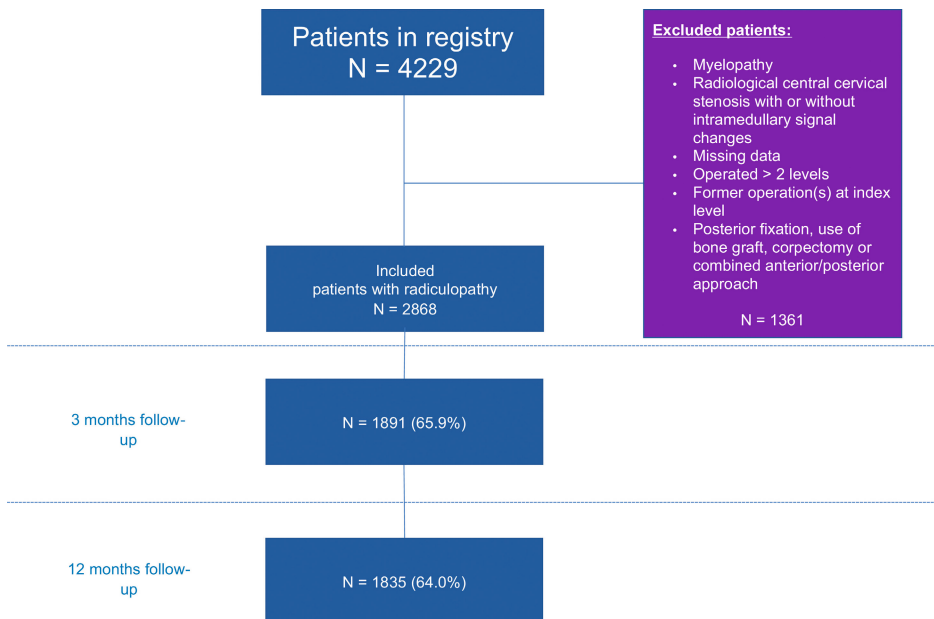


Figure 1. Exclusion criteria for patients with follow-up rates.

Measurements

The comprehensive NORspine self-administered questionnaire consists of information about sociodemographic 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 (Figure 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 radic-

ulopathy 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 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, $p < .001$.

Table 1. Baseline characteristics of respondents and nonrespondents to follow-up at 12 months

	Respondents N=1,843		Nonrespondents N=1,025		Sig. (2-tailed)/ chi-square
	N		N		
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

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), EQ-VAS = General health status by EuroQol (0–100).

Table 2. Baseline characteristics

	Whole radiculopathy group N=2,868		Disc herniation N=1,182		Spondylotic foraminal stenosis N=430		Sig. (2-tailed)/chi-square
	N	N	N	N	N	N	
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

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
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), EQ-VAS = General health status by EuroQol (0-100).

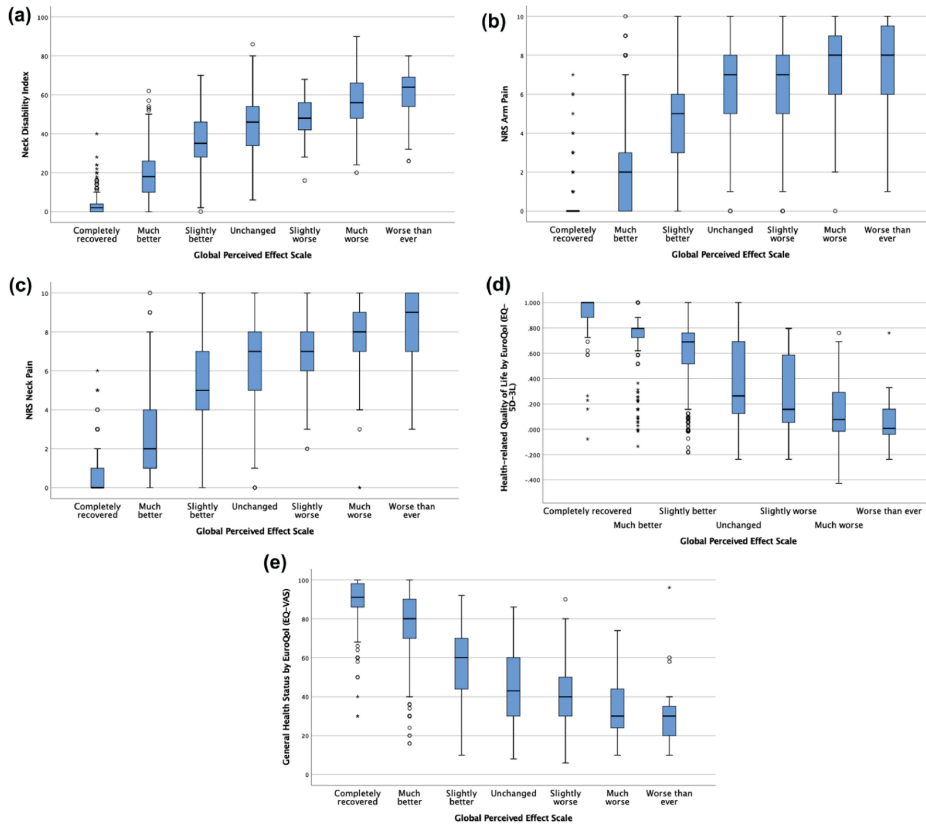


Figure 2. Mean follow-up scores of PROMs at 12 months according to each GPE category

(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.

The mean follow-up scores of PROMs at 12 months according to each GPE category are presented in Figure 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.

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)

AUC = Area Under the Curve, 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), EQ-VAS = General health status by EuroQol (0–100).

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).

Table 4. Cut-off values

	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)

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

Discussion

We found very good to excellent discriminative ability in distinguishing between success and non-success following neck surgery due to radiculopathy for the most commonly used PROMs. The NDI and the NRS-AP had the highest 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. 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. Figure 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].

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.

In conclusion, this study showed the best ability in distinguishing between a successful and non-successful 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.

References

1. Rihn JA. Anterior Cervical Discectomy and Fusion. Case Competencies in Orthopaedic Surgery: Elsevier; 2017. p. 318-25.
2. Dakwar E, Deukmedjian A, Ritter Y. Spinal pathology, conditions, and deformities: Surgical intervention. Pathology and intervention in musculoskeletal rehabilitation. 2015;2.
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 JOURNAL. 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 JOURNAL. 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. QUALITY OF LIFE RESEARCH. 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. QUALITY OF LIFE RESEARCH. 2010;19(4):539-49.
7. Jaeschke R, Singer J, Guyatt GH. MEASUREMENT OF HEALTH-STATUS - ASCERTAINING THE MINIMAL CLINICALLY IMPORTANT DIFFERENCE. CONTROLLED CLINICAL 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. JBJS. 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. The Spine Journal. 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. CLINICAL SPINE SURGERY. 2019;32(9):392-7.
11. Solberg TK, Olsen LR, Berglund ML. Nasjonalt kvalitetsregister for ryggkirurgi (NKR) Årsrapport for 2018 med plan for forbedringstiltak. 2019.
12. 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. Annals of internal medicine. 2007;147(8):573-7.
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. JOURNAL OF ORTHOPAEDIC & SPORTS PHYSICAL THERAPY. 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. 2011.
16. Turk DC, Melzack R. Handbook of pain assessment: Guilford Press; 2011.

17. Kamper SJ, Ostelo R, Knol DL, Maher CG, de Vet HCW, Hancock MJ. Global Perceived Effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. *JOURNAL OF CLINICAL EPIDEMIOLOGY*. 2010;63(7):760-6.
18. Werner DA, Grotle M, Gulati S, Austevoll IM, Lønne G, 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. *European Spine Journal*. 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 musculoskeletal disorders*. 2019;20(1):1-15.
20. Hermansen E, Myklebust TÅ, 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 Musculoskeletal Disorders*. 2019;20(1):1-8.
21. 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 orthopaedica*. 2013;84(2):196-201.
22. Tape TG. Interpreting diagnostic tests: The area under an ROC curve. Retrieved December. 2002;4:2002.
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. *The journal of pain*. 2008;9(2):105-21.
24. Fekete TF, Haschtmann D, Kleinstück FS, Porchet F, Jeszenszky D, Mannion AF. What level of pain are patients happy to live with after surgery for lumbar degenerative disorders? *The Spine Journal*. 2016;16(4):S12-S8.
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. *Quality of life research*. 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. *British Journal of Neurosurgery*. 2019;33(1):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. *Orthopaedic surgery*. 2016;8(4):425-31.
28. Sedaghat AR. Understanding the minimal clinically important difference (MCID) of patient-reported outcome measures. *Otolaryngology–Head and Neck Surgery*. 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? *Clinical orthopaedics and related research*. 2018;476(10):2027.
30. Parker SL, Godil SS, Shah 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. JOURNAL OF NEUROSURGERY-SPINE*. 2013;18(2):154-60.
31. Narain AS, Hijji FY, Khechen B, Haws BE, Bohl DD, Yom KH, et al. Risk factors associated with failure to reach minimal clinically important difference in patient-reported outcomes following anterior cervical discectomy and fusion. *International Journal of Spine Surgery*. 2019;13(3):262-9.

32. Gatchel RJ, Mayer TG. Evidence-informed management of chronic low back pain with functional restoration. *The Spine Journal*. 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. *Journal of Orthopaedics*. 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. *Journal of Manual & Manipulative Therapy*. 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 review of pharmacoeconomics & outcomes research*. 2011;11(2):163-9.
38. Van Hooff ML, Jacobs WCH, Willems PC, Wouters M, de Kleuver M, Peul WC, et al. Evidence and practice in spine registries A systematic review, and recommendations for future design of registries. *ACTA ORTHOPAEDICA*. 2015;86(5):534-44.
39. Mannion AF, Elfering A. Predictors of surgical outcome and their assessment. *European Spine Journal*. 2006;15(1):S93-S108.
40. Solberg TK, Sørle 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? A study of responding and non-responding cohort participants from a clinical spine surgery registry. *Acta orthopaedica*. 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. *EUROPEAN SPINE JOURNAL*. 2016;25(1):282-6.

Appendix (A-C)

Appendix A: Mean Patient-Reported Outcome Measures scores at 12 months stratified according to the Global Perceived Effect (GPE) scale.

Global Perceived Effect Scale									
	Completely recovered	Much better	Slightly better	Un-changed	Slightly worse	Much worse	Worse than ever	Spearman's rank correlation	
N (%)	358 (19.6)	856 (46.8%)	366 (20.0%)	114 (6.2%)	73 (4.0%)	41 (2.2%)	23 (1.3%)		
NDI*									
Follow-up score	3.9	17.7	35.5	43.1	48.9	56.9	58.7	0.78	
Mean (SD*****)	(6.2)	(11.3)	(14.3)	(16.5)	(11.4)	(15.3)	(14.9)		
Change score	31.3	22.2	8.1	0.2	-2.0	-2.9	-4.2	0.64	
Mean (SD*****)	(13.6)	(13.8)	(10.8)	(11.0)	(12.0)	(14.9)	(14.5)		
Percentage Change score	88.9	53.5	17.2	-5.6	-10.0	-10.7	-11.9	0.78	
Mean (SD*****)	(16.0)	(30.3)	(26.4)	(38.3)	(29.0)	(39.0)	(30.4)		
NRS-AP**									
Follow-up score	0.4	2.1	4.4	6.3	6.3	7.3	7.6	0.69	
Mean (SD*****)	(1.0)	(2.0)	(2.4)	(2.5)	(2.5)	(2.3)	(2.4)		
Change score	6.1	4.2	2.1	0.0	0.7	0.1	-0.4	0.58	
Mean (SD*****)	(2.4)	(2.8)	(2.6)	(2.3)	(2.8)	(2.3)	(2.4)		
Percentage Change score	93.8	63.4	28.0	-11.2	4.8	-7.3	-13.3	0.68	
Mean (SD*****)	(16.6)	(47.2)	(45.2)	(75.9)	(43.1)	(64.1)	(56.7)		
NRS-NP***									
Follow-up score	0.5	2.5	5.1	6.2	6.7	7.5	7.9	0.74	
Mean (SD*****)	(1.0)	(1.8)	(2.2)	(2.6)	(1.8)	(2.4)	(2.3)		
Change score	4.8	3.4	1.3	0.2	0.4	0.0	-0.5	0.51	
Mean (SD*****)	(2.7)	(2.7)	(2.2)	(2.3)	(2.4)	(2.9)	(2.6)		
Percentage Change score	90.5	52.4	12.5	-0.5	-6.5	-11.3	-12.7	0.68	
Mean (SD*****)	(20.8)	(43.5)	(60.7)	(71.2)	(53.3)	(63.5)	(46.3)		
EQ-5D-3L****									
Follow-up score	0.94	0.77	0.57	0.39	0.25	0.17	0.07	0.72	
Mean (SD*****)	(0.13)	(0.17)	(0.28)	(0.33)	(0.29)	(0.28)	(0.22)		
Change score Mean (SD*****)	0.44	0.31	0.16	0.00	-0.11	-0.07	-0.08	0.46	
	(0.31)	(0.33)	(0.32)	(0.30)	(0.36)	(0.30)	(0.17)		

Appendix A: Mean Patient-Reported Outcome Measures scores at 12 months stratified according to the Global Perceived Effect (GPE) scale. (continued)

Global Perceived Effect Scale									
	Completely recovered	Much better	Slightly better	Un-changed	Slightly worse	Much worse	Worse than ever	Spearman's rank correlation	
EQ-VAS*****	89.4 (11.1)	76.6 (15.4)	56.8 (18.5)	45.2 (19.6)	40.3 (17.7)	35.9 (16.0)	30.6 (19.5)	0.70	
Follow-up score Mean (SD*****)									
Change score Mean (SD*****)	34.7 (22.5)	24.7 (22.3)	8.5 (20.6)	-6.5 (21.2)	-5.9 (22.4)	-7.6 (24.7)	-4.9 (13.8)	0.50	
Percentage Change score Mean (SD*****)	105.7 (235.7)	85.3 (199.0)	37.9 (84.7)	-4.8 (45.3)	-1.2 (63.1)	9.5 (103.6)	-11.7 (35.0)	0.41	

* 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), ***** Standard Deviation

Appendix B: The area under the curve (AUC) with 95% Confidence Interval (CI) and cut-off values for success (Cut-off) for all Patient-Reported Outcome Measures (PROMs) at 3 months.

		Follow-up score	Change score	Percentage change
NDI*	AUC**	0.88	0.85	0.89
	(95% CI)	(0.86-0.90)	(0.83-0.87)	(0.87-0.91)
	Cut-off	25.0	13.5	33.0%
NRS-AP ***	AUC**	0.84	0.82	0.83
	(95% CI)	(0.82-0.86)	(0.79-0.84)	(0.81-0.86)
	Cut-off	2.5	3.5	56.3%
NRS-NP ****	AUC**	0.85	0.77	0.82
	(95% CI)	(0.83-0.87)	(0.74-0.79)	(0.79-0.84)
	Cut-off	3.5	1.5	35.4%
EQ-5D-3L*****	AUC**	0.85	0.75	Not possible to calculate
	(95% CI)	(0.84-0.87)	(0.72-0.78)	Not possible to calculate
	Cut-off	0.73	0.11	Not possible to calculate
EQ-VAS *****	AUC**	0.87	0.79	0.74
	(95% CI)	(0.85-0.88)	(0.77-0.81)	(0.71-0.77)
	Cut-off	69.0	13.5	24.6%

* 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), ***** Standard Deviation

Appendix C: Area under the Receiver Operating Curve (AUC) with 95% Confidence Intervals (CI) 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.

	NDI*	NRS-AP***	NRS-NP****	EQ-5D-3L*****	EQ-VAS*****
	AUC** (95% CI)	AUC** (95% CI)	AUC** (95% CI)	AUC** (95% CI)	AUC** (95% CI)
Disc herniation					
Follow-up score	0.92 (0.90-0.94)	0.86 (0.83-0.90)	0.89 (0.86-0.92)	0.88 (0.85-0.91)	0.89 (0.86-0.91)
Change score	0.87 (0.84-0.90)	0.81 (0.76-0.85)	0.78 (0.74-0.83)	0.79 (0.75-0.83)	0.76 (0.71-0.81)
Percentage change score	0.92 (0.89-0.94)	0.86 (0.82-0.89)	0.84 (0.81-0.87)	Not possible to calculate	0.73 (0.67-0.78)
Spondylotic foraminal stenosis					
Follow-up score	0.91 (0.88-0.94)	0.85 (0.81-0.90)	0.91 (0.87-0.94)	0.83 (0.79-0.88)	0.87 (0.83-0.91)
Change score	0.85 (0.79-0.90)	0.79 (0.72-0.85)	0.79 (0.72-0.85)	0.73 (0.65-0.80)	0.81 (0.75-0.87)
Percentage change score	0.91 (0.86-0.95)	0.85 (0.80-0.91)	0.87 (0.83-0.91)	Not possible to calculate	0.75 (0.67-0.82)
Anterior approach					
Follow-up score	0.92 (0.90-0.93)	0.86 (0.84-0.88)	0.89 (0.87-0.91)	0.87 (0.85-0.88)	0.88 (0.87-0.90)
Change score	0.87 (0.85-0.89)	0.81 (0.78-0.83)	0.80 (0.77-0.82)	0.74 (0.71-0.77)	0.79 (0.76-0.82)
Percentage change score	0.91 (0.90-0.93)	0.85 (0.83-0.87)	0.84 (0.82-0.86)	Not possible to calculate	0.74 (0.72-0.77)

Appendix C: Area under the Receiver Operating Curve (AUC) with 95% Confidence Intervals (CI) 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. (continued)

	NDI*		NRS-AP***		NRS-NP****		EQ-5D-3L*****		EQ-VAS*****	
	AUC** (95% CI)		AUC** (95% CI)		AUC** (95% CI)		AUC** (95% CI)		AUC** (95% CI)	
Posterior approach										
Follow-up score	0.90 (0.84-0.95)		0.89 (0.83-0.95)		0.88 (0.82-0.95)		0.84 (0.78-0.91)		0.87 (0.80-0.93)	
Change score	0.88 (0.81-0.94)		0.85 (0.78-0.93)		0.79 (0.70-0.88)		0.77 (0.67-0.86)		0.77 (0.68-0.87)	
Percentage change score	0.91 (0.85-0.97)		0.89 (0.82-0.95)		0.88 (0.83-0.94)		Not possible to calculate		0.72 (0.62-0.82)	

* 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)