

Preoperative Evaluation of Oswestry Disability Index in Lumbar Spinal Stenosis: New Evidence of Time Independence of Variation Up to 1 Year

Juho Hatakka, Katri Pernaa, Joel Kostensalo, Keijo Mäkelä and Inari Laaksonen

Int J Spine Surg published online 29 January 2025
<https://www.ijssurgery.com/content/early/2025/01/22/8699>

This information is current as of January 30, 2025.

Email Alerts Receive free email-alerts when new articles cite this article. Sign up at:
<http://ijssurgery.com/alerts>

Preoperative Evaluation of Oswestry Disability Index in Lumbar Spinal Stenosis: New Evidence of Time Independence of Variation Up to 1 Year

JUHO HATAKKA, MD; KATRI PERNAÄ, MD, PhD¹; JOEL KOSTENSALO, PhD²; KEIJO MÄKELÄ, MD, PhD; AND INARI LAAKSONEN, MD, PhD

¹Department of Orthopedics and Traumatology, Turku University Hospital, and University of Turku, Turku, Finland; ²Natural Resources Institute Finland, Natural Resources, Joensuu, Finland

ABSTRACT

Background: The Oswestry Disability Index (ODI) is a well-validated and widely used patient-reported outcome instrument to evaluate lumbar spinal stenosis (LSS) patients' treatment outcomes. The objective of the present study was to determine long the average interval between 2 preoperative measurements can be before a clinically significant difference of 10 points or more might appear.

Methods: This was a retrospective observational study utilizing prospectively collected data from a single university hospital database, which was compatible with the national registry. One hundred and four surgically treated LSS patients were included in this observational study using systematic sampling. The preoperative ODI score was obtained at 2 timepoints. The 2-month mark as a potential turning point was of special interest, as the registry in question excludes preoperative data as outdated if the data are older than 2 months. Possible time dependence of the change in ODI scores was explored using a linear mixed-effects model with ODI as the dependent variable and interval length, sex, age, body mass index (BMI), and the presence of a concomitant disease as fixed effects.

Results: The mean ODI score was 41.7 points (SD = 16.0) at the first and 41.1 points (SD = 15.5) at the second measurement. Mean time between the ODI scores was 74 days (range 8–361). On average, ODI changed by 9.17 points (SD = 7.16) between the 2 measurements, increasing for 48 patients, remaining unchanged for 9 patients, and decreasing for 47 patients. The arithmetic mean of the changes was -0.60 points and the median was 0.00 points. The estimated change in the population mean was -0.0005 points/day (95% CI $[-0.022, 0.022]$, $P = 0.97$), meaning that we have strong evidence that the change in the mean is not clinically significant for up to 15 months (95% CI between ± 10 points). Furthermore, no evidence was found that age, sex, BMI, or concomitant diseases were associated with the change of ODI score over time. Furthermore, the probability to observe a clinically significant change in a patient did not depend on the number of days between the 2 measurements (OR 1.003, 95% CI $[0.997, 1.010]$, $P = 0.30$). Variance in ODI change did not grow over time.

Conclusions: The probability of observing a clinically significant differences does not depend on the length of the observation interval, and ODI scores can be considered equally reliable for a significantly longer time than 2 months, even up to 1 year.

Clinical Relevance: Preoperative ODI scores do not lose reliability up to 1 year in patients undergoing operatively treatment for LSS.

Level of Evidence: 3.

Lumbar Spine

Keywords: spine, lumbar spinal stenosis, patient reported outcome, oswestry disability index, ODI score, registry study

INTRODUCTION

Degenerative lumbar spinal stenosis (LSS)¹ is the most common cause of spinal disability.² LSS patients typically experience pain, numbness, or discomfort in the lower back, buttocks, or lower extremities, distinct or all together, while standing or walking. Decompressive surgery with or without fusion has shown a positive effect on patients' symptoms compared with conservative treatment, especially leg pain, claudication, and

overall disability.^{3–5} As with all spine surgery, the incidence of LSS surgery has increased over past few decades.^{6,7}

The first nationwide spine registry (the Swedish Spine Registry, or SWESPINE) was established in 1993.⁸ SWESPINE has since provided several peer-reviewed publications on the results of spine surgery.⁹ The Finnish Spine Registry (FinSpine) development started in 2015. Besides operative data, both registries collect patient-reported outcome (PRO) data evaluating

back-related symptoms such as the Oswestry Disability Index (ODI) both preoperatively and postoperatively. ODI was initially developed in 1976¹⁰ and first published in 1980.¹¹ It is 1 of the most commonly used patient-reported outcome measures in spine surgery.¹² The Finnish version of ODI is currently in use in Finland.¹³

The adequacy of a PRO can be evaluated with different methods, such as validity and responsiveness. Validity reflects the ability of a certain instrument to measure what it is supposed to measure. The validity of the ODI has been established previously.^{10,14} ODI has acceptable internal consistency and reliability. Especially good test-retest reliability, that is, the stability of an instrument over a specific duration, most often 1 to 6 weeks, has been reported.^{15,16} However, to our knowledge, there are no data relating to longer time intervals beyond 6 weeks for LSS patients. This information is necessary for evaluating the reliability of registry data as time intervals between outpatient clinic visit and operative treatment tend to be longer (up to several months) in a real-life setting.

The objective of the present study was to find out how long the interval between 2 measurements can be before a clinically significant difference (10 points or more) might appear. The 2-month mark as a potential turning point was of special interest, as the registry in question excludes preoperative data more than 2 months old as outdated.

METHODS

This study was an observational investigation of consecutive patients with operatively treated LSS from a single university hospital database that collects registry data compatible with the Finnish Spine registry (FinSpine). All patients who had LSS diagnosed at an outpatient clinic and were scheduled for operative treatment between January 2019 and December 2019 were screened. All patients completed the first ODI before their outpatient clinic visit and, due to FinSpine requirements, the second ODI preoperatively no more than 2 months before the operation. Due to this limit of 2 months between the ODI score and operation set by the registry, we decided to study patients with time intervals ≤ 2 months and > 2 months between the ODI measurements as separate groups in addition to the full sample analysis. Based on a power analysis, described in more detail later, data on 104 patients were gathered using systematic quota sampling from the registry: the first 52 patients with an ODI time interval ≤ 2 months and the first 52 patients with ODI interval > 2 months,

Table 1. Inclusion and exclusion criteria.

Variable	Inclusion Criteria	Exclusion Criteria
Diagnosis	Lumbar spinal stenosis with or without spondylolisthesis	Any concomitant spinal disorders (scoliosis, vertebral fracture, isthmic spondylolysis, tumor, and metastases)
ODI scores	Measured at 2 timepoints prior to the operation	Incomplete or missing scores, or the interval was < 1 week ¹⁰

Abbreviation: ODI, Oswestry Disability Index.

fulfilling inclusion criteria, were included in the sample (Table 1).

All of the study patients underwent upright lumbar radiography or a full-body scan (EOS imaging) and lumbar spine MRI, and they had symptoms related to LSS such as buttock pain, neural claudication, or lower limb radicular pain. Collected data included patient demographics, 2 preoperative ODI scores, VAS scores for back and leg pain, employment (employed, unemployed, retired, or unable to work), smoking status (smoker or non-smoker), duration of symptoms (< 6 weeks, 6–12 weeks, 3–12 months, or > 12 months), usage of pain medication (none, occasionally, or regularly), and concomitant diseases.

Sex, body mass index (BMI), age, and concomitant diseases (diabetes mellitus, lung disease, rheumatoid arthritis, and heart disease) were systematically investigated as potential confounding factors.

Power Analysis

Because the registry considers 2 months as a cut-off for reliability, the sample size was determined so that clinically significant changes in either of these 2 subgroups (≤ 2 months and > 2 months) could be reliably observed. The mean (SD) minimal clinically significant difference for ODI has been reported to be 10 (20).¹⁷ Assuming the ODI scores are normally distributed, it is straightforward to carry out a power calculation for a paired (ie, 1 sample) *t* test.¹⁸ Based on these, the total number of patients needed to achieve 95% power was 104, with 52 patients in each subgroup. This sample size should also be sufficient for the linear mixed model used to quantify the expected day-to-day changes, as the use of exact interval length and additional covariates provide additional precision.

Statistical Methods

The time dependence of ODI score change was investigated from 3 different points of view:

1. Does the mean change depend on the interval length?
2. Does the probability of an individual patient experiencing a clinically significant change depend on the interval length?
3. Does the variation in ODI change depending on the interval length?

Possible time dependence of the mean change in ODI scores was explored using a linear mixed-effects model with ODI as the dependent variable while interval length, sex, age, BMI, and the presence of a concomitant disease were fixed effects. A random intercept term was included at the patient level. The need to include interaction terms between the interval length and the other covariates was systematically tested using the Akaike Information Criterion. Initial ODI score, smoking status, duration of symptoms, use of pain medication, and employment status were explored in post-hoc analyses. An analogous generalized linear model where the response variable was whether a clinically significant difference was observed (0 = no and 1 = yes) was fitted to test whether the probability of an individual patient experiencing a clinically significant change depends on time. The time dependence of variation in the change of ODI scores was tested using Levene's test for the equality of variances. For this test, the data set was divided into groups with 15, 30, 60, and 90 days.

The change in ODI scores was studied separately for patients with a measurement interval of less than 61 days, those with a measurement interval of 61 days or more, and for the whole sample. The normality of the ODI scores in the full sample and the 2 subsamples was investigated using Shapiro-Wilk test for normality,¹⁹ and no evidence for the non-normality of the distribution was found. The statistical significance of the difference between the 2 measurement points was tested using a paired *t* test with a 2-tailed alternative hypothesis for the complete sample and the subsamples separately. The equality of variances was also tested between these 2 groups using Levene's test.

All statistical analyses were carried out using the statistical software R.²⁰ Graphical investigations with some figures were produced using the package ggplot2.²¹

Ethics

The present study was based on registry data, and the patients were not directly contacted. Therefore, this study was exempt from local ethical committee review.

Table 2. Results of the mixed linear model fit.

Variable	Estimate ^a	SE ^a	P ^a
Fixed effects			
Intercept ^b	44.0	6.5	<0.0001
Interval (day)	-0.0005	0.011	0.97
Sex (male)	-8.1	2.9	0.007
Age	0.18	0.45	0.32
Concomitant diseases (yes or NA)	0.35	0.34	0.96
BMI	0.68	0.34	0.05
Random effects			
Patient	$\sigma = 12.8$		
Residual	$\sigma = 8.3$		
Fit quality			
$R^2_{\text{mar}} / R^2_{\text{con}}$ ^c	0.09/0.73		

Abbreviations: BMI, body mass index; NA, not applicable; ODI, Oswestry Disability Index.

^aFor the fixed effects.

^bThe intercept reflects the expected ODI score for a 70-year-old woman with no concomitant diseases and a BMI of 29.

^cMarginal R^2 value/conditional R^2 value.

RESULTS

The mean age was 71 years, and 64 patients (62%) were women. The mean (SD) ODI score at the first measurement was 41.7 (16.0) points and 41.1 (15.5) points at the second measurement. Mean time between the ODI scores was 74 days (range, 8–361). On average, ODI changed by 9.17 points (SD = 7.16) between the 2 measurements, with the ODI score increasing for 48 patients, remaining unchanged for 9 patients, and decreasing for 47 patients. The arithmetic mean of the changes was -0.60 points and the median 0.00 points.

For the linear mixed model, no interaction terms were found to improve the model fit, and the final model fit can be found in Table 2. The population-level estimates for ODI score changes were found to be -0.0005 points/day (95% CI [-0.022, 0.022], $P = 0.97$). The 95% CI is contained within the clinically significant limits of ± 10 points for the first 446 days, that is, for about 15 months. Thus, the population-level mean is unlikely to change in a clinically significant way over this period. Women and patients with higher BMIs had higher ODI scores on average, while age and concomitant diseases had no statistically significant association with the ODI score. For patients with ≤ 2 months between the ODI scores, the mean (SD) ODI score at the first time point was 43.7 (17.3) points and 41.3 (17.1) points at the second time point. For patients with > 2 months between the ODI scores, the mean (SD) ODI score at the first time point was 39.6 (14.3) points and 40.8 (13.9) points at the second time point (Figures 1 and 2). Also, when patients with a time interval ≤ 2 months and > 2 months between the ODI scores were studied separately with a *t* test, no statistically or clinically significant changes were observed (2.4 points, 95% CI [-5.20, 0.37], P

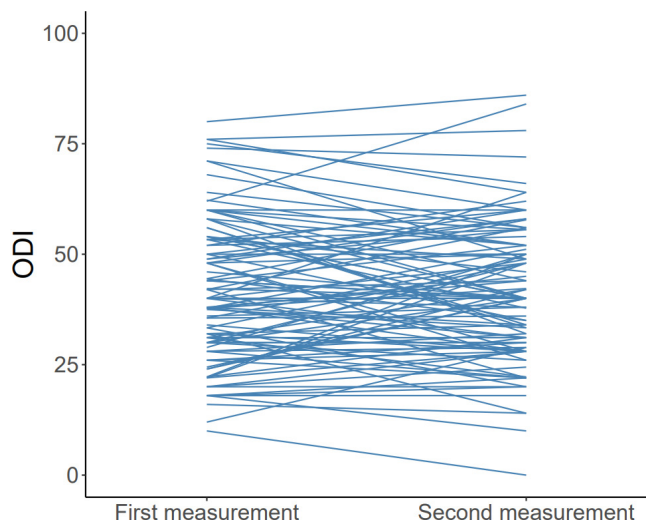


Figure 1. Time evolution of Oswestry Disability Index (ODI) results for the individual patients between the 2 measurement times.

= 0.09; 1.2 points, 95% CI [-2.39, 4.82], $P = 0.50$, respectively).

For 62 patients, there was no clinically significant change in the ODI score between the measurement points; for 20 patients, the ODI score decreased clinically significantly (≥ 10 points; 7 patients in the ≤ 2 months and 13 patients in the >2 months group); for 22 patients, the score increased clinically significantly (13 patients in the ≤ 2 months and 9 patients in the >2 months group). The shortest interval associated with a clinically significant change (-26 points) was 8 days. Based on the generalized linear model fit, none of the covariates nor the length of the time interval was associated with an increased or decreased risk of having a clinically significant difference occur. The OR for

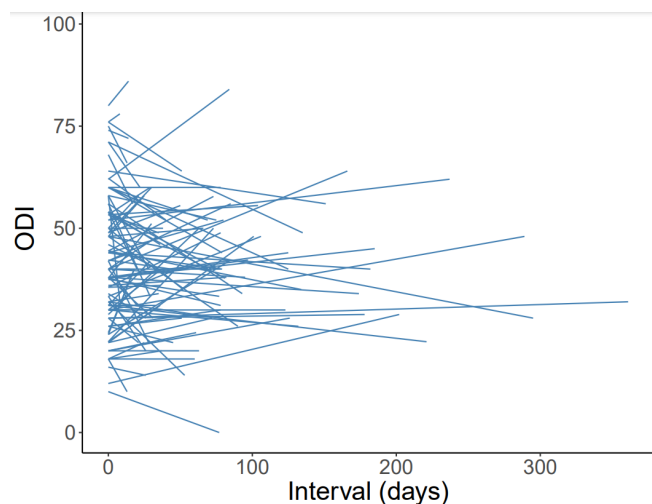


Figure 2. Time evolution between the 2 Oswestry Disability Index measurement points with the time interval between the 2 points on the horizontal axis.

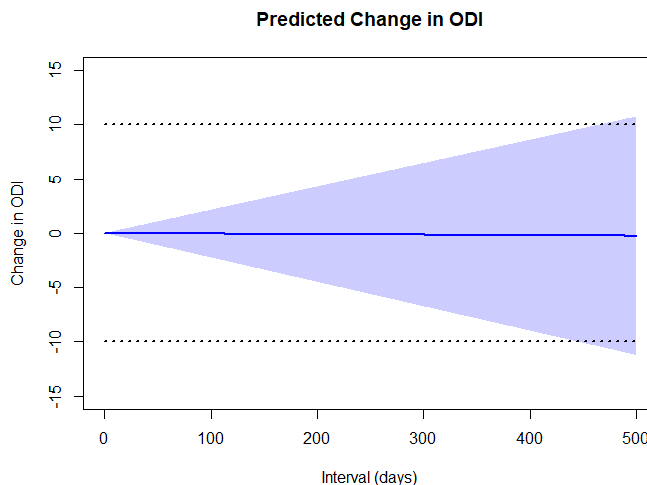


Figure 3. Predicted change in the population mean due to interval length between 2 Oswestry Disability Index (ODI) measurements. The shaded area corresponds to the 95% CI. The dotted lines denote the clinically significant difference of 10 ODI points.

observing a clinically significant difference for an individual 1 day longer interval was 1.003 (95% CI [0.997, 1.010] $P = 0.30$). Thus, there is no indication that a longer interval between the measurements is connected to a higher probability for an individual to experience clinically significant changes in ODI (Figure 3).

The variance of ODI score change did not depend on the interval length either with 15-day binning of observations ($P = 0.22$), 30-day binning ($P = 0.36$), 60-day binning ($P = 0.21$), or 90-day binning ($P = 0.12$). No difference in the variance of ODI score change was found between the ≤ 2 months and >2 months groups either ($P = 0.20$).

The preoperative ODI score did not have any correlation with the delay for surgery ($P = 0.13$). Smoking or employment status, use of pain medication, or duration of symptoms were not found to improve the model fits.

DISCUSSION

The objective of the present study was to assess preoperative changes in ODI score in LSS patients waiting for operative treatment. There was no statistically or clinically significant difference in the means of the 2 preoperative ODI scores measured at different occasions with waiting time of ≤ 2 months or >2 months. Based on our results, it seems that ODI scores even for patients with severe LSS do not progress within a few months, and the decision of operative treatment does not affect the ODI score. Furthermore, we did not find any potential factors contributing to the change in the ODI scores. Based on our registry data, preoperative ODI score at outpatient clinic seems to present patient's

preoperative symptom state reliably even though there would be a delay between the operative treatment decision and surgical treatment.

Medical registries have been shown to be valuable tools for improving patient care. Proper utilization of registry data relies on the accuracy of the data contained in the database. The assessment of data quality is of utmost importance in improving the reliability of registry-based studies in the future.

ODI is a well validated and widely used PRO to assess patients with spine-related conditions with several adaptations in different languages.^{10,11,13,22} Test-retest reliability is a feature of PRO instrument quality that indicates how well an instrument produces similar results on repeated measures when no change is expected. ODI has good test-retest reliability with an interval between measurements less than 6 weeks used in most studies.^{15,22} Given that it is evident that the instrument has good measurement properties, we were able to assess possible changes with longer intervals in this study. This is relevant in clinical settings, where a number of factors affect the interval between the operative treatment decision and the surgery itself, as well as for the adequacy of registry data. In our data, longer waiting time was not associated with a clinically relevant increase in the average ODI score when the interval between scores was less than 15 months. For intervals longer than 15 months, the number of observations was too low to assess the change; however, there was no evidence that the change would be clinically significant after this time point. It should also be noted that the longest observed interval in our data was 361 days, so making claims regarding intervals longer than 1 year is not possible.

The ability of an instrument to identify possible changes in the condition to be measured is called responsiveness. The responsiveness for the ODI has been confirmed in a number of clinical conditions, such as back pain and LSS.^{23,24} While test-retest validity ensures that there are no instrument-related errors expected in repeated measures, with good responsiveness the change, if there is such, can be expected to manifest between repeated measures. Even though the prevalence of symptomatic LSS is higher in the elderly population, in our sample, age was not associated with the change between 2 preoperative ODI score time-points.²

Preoperative ODI scores in our population were comparable to earlier registry data assessing PRO results of patients waiting for surgical treatment for LSS.²⁵ Weinstein et al compared results of LSS treatment in

both randomized and observational groups, and in both groups, the ODI remained stable.⁵ However, there was a significant crossover between the study groups in the randomized cohort: at 1 year only 63% from the surgery group had undergone operative treatment and 42% from the non-surgical group had had an operation. The mean change for ODI score for the non-surgical treatment group was -7.4 points at 6 weeks, -8.1 points at 3 months, and -12.7 points at 1 year. A randomized controlled trial comparing long-term effects of operative vs nonoperative treatment of LSS by Slätis et al showed improvement in the ODI on both groups favoring the operative group, and there was no remarkable crossover from the conservative group to the operative group.⁴ In their conservative treatment group, the mean ODI change was -7.4 points at 6 weeks, -5.2 points at 3 months, and -7.2 points at 1 year. In our sample, all patients were scheduled to undergo operative treatment with no additional treatment provided after the decision, and there was no clinically significant change in the ODI. In our material, the severe symptomatic LSS symptom state seems to remain stable within our time interval. For 20 patients, the ODI score decreased clinically significantly (≥ 10 points), and for 22 patients, the score increased clinically significantly (≥ 10 points). As the number of changes to both directions was comparable and there was no significant change in the mean ODI scores, it is likely that the change in these patients is explained by daily variation. Also, based on a recent study, repeated preoperative MRI scans do not provide benefit, which is in line with these findings of changes in preoperative patient-reported outcome measure results.²⁶

In reliable assessment of registry data, as well as individual patient care, it is important to consider potential contributing factors to each condition. Knutsson et al observed that smoking and the risk of having LSS surgery are correlated in the Swedish working population.²⁷ The risk was dose correlated, and heavy smokers were more likely to undergo LSS surgery. A registry-based study, also of a Swedish population, noted that nonsmokers were more satisfied with the treatment outcome and used less analgesics than smokers after LSS surgery.²⁸ Sekiguchi et al found an association between lack of regular exercise, strenuous use of low back and legs, and lower job satisfaction with LSS.²⁹ Hypertension and diabetes mellitus were shown to be associated in a study by Uesugi et al.³⁰ In our study, we found no confounding covariates related to change of the ODI. The tested covariates included sex, BMI, smoking status, preoperative occupational status, preoperative

use of pain medication, and concomitant diseases. One potential factor affecting PRO scores is the operative treatment decision and outpatient clinic visit. In our data, the ODI scores did not change significantly, thus suggesting that these factors did not affect the scores. Furthermore, the preoperative ODI score had no effect on the delay between the operative treatment decision and surgical treatment. Based on this finding, it seems that institutional factors affected the surgery delay more than patients' preoperative symptom state.

We acknowledge that our study has several limitations. First, the sampling was done with the main research question in mind, that is, the influence of the interval between 2 measurements to the potential change in the ODI. Therefore, in the subgroup analysis, there are a limited number of patients, such as patients with lung disease, and the results of this analysis must be interpreted with care. Second, as always is the case with a retrospective study setting, there might be some selection bias. In surgical studies, patient selection to nonoperative and operative treatment is a major bias, and as all the patients included in this study were waiting for operative treatment, we think that the risk was low. Third, the analysis was carried out only with subjects diagnosed with LSS. Thus, extrapolation of these results to other spine conditions or less severe forms of LSS must be conducted with concern.

CONCLUSION

There was no statistically nor clinically significant change in the population mean of the ODI score between 2 preoperative measurements when the interval between ODI measurements was less than 1 year. Furthermore, there was no evidence that the probability of an individual patient experiencing a clinically significant change would be associated with the length of the interval between consecutive ODI measurements. Finally, the variation in ODI change was not associated with the length of the time interval. Therefore, the preoperative ODI score gathered at the outpatient clinic before the surgical treatment decision can be considered to equally reliably present patient's preoperative symptoms as long as the score is no older than 1 year. However, a clinically significant change could be experienced in as little as 8 days based on the data. There was no evidence of the treatment effect of outpatient clinic visit or treatment decision for surgery on the change in preoperative ODI scores. Furthermore, we did not find any other factors contributing to the change in the ODI scores. However, until we get more data with longer intervals, a new ODI score is needed when the

time interval between preoperative ODI measurements exceeds 12 months to reliably assess patient's preoperative symptom state.

REFERENCES

- Spengler DM. Degenerative stenosis of the lumbar spine. *J Bone Joint Surg Am.* 1987;69(2):305–308.
- Kalichman L, Cole R, Kim DH, et al. Spinal stenosis prevalence and association with symptoms: the framingham study. *Spine J.* 2009;9(7):545–550. doi:10.1016/j.spinee.2009.03.005
- Malmivaara A, Slätis P, Heliövaara M, et al. Surgical or nonoperative treatment for lumbar spinal stenosis? *Spine (Phila Pa 1986).* 2007;32(1):1–8. doi:10.1097/01.brs.0000251014.81875.6d
- Slätis P, Malmivaara A, Heliövaara M, et al. Long-term results of surgery for lumbar spinal stenosis: a randomised controlled trial. *Eur Spine J.* 2011;20(7):1174–1181. doi:10.1007/s00586-010-1652-y
- Weinstein JN, Tosteson TD, Lurie JD, et al. Surgical versus nonsurgical therapy for lumbar spinal stenosis. *N Engl J Med.* 2008;358(8):794–810. doi:10.1056/NEJMoa0707136
- Deyo RA, Mirza SK, Martin BI, Kreuter W, Goodman DC, Jarvik JG. Trends, major medical complications, and charges associated with surgery for lumbar spinal stenosis in older adults. *JAMA.* 2010;303(13):1259–1265. doi:10.1001/jama.2010.338
- Gray DT, Deyo RA, Kreuter W, et al. Population-based trends in volumes and rates of ambulatory lumbar spine surgery. *Spine (Phila Pa 1986).* 2006;31(17):1957–1963. doi:10.1097/01.brs.0000229148.63418.c1
- Strömquist B, Jönsson B. Computerized follow-up after surgery for degenerative lumbar spine diseases. *Acta Orthop Scand.* 1993;64(sup251):138–142. doi:10.3109/17453679309160145
- Parai C, Hägg O, Lind B, Brisby H. The value of patient global assessment in lumbar spine surgery: an evaluation based on more than 90,000 patients. *Eur Spine J.* 2018;27(3):554–563. doi:10.1007/s00586-017-5331-0
- Fairbank JCT, Pynsent PB. The oswestry disability index. *Spine (Phila Pa 1986).* 2000;25(22):2940–2953. doi:10.1097/00007632-200011150-00017
- Fairbank JCT, Davies JB, Couper J, O'Brien JP. The oswestry low back pain disability questionnaire. *Physiotherapy.* 1980;8(66):271–273.
- Boden SH, Farley KX, Campbell C, Boden SD, Gottschalk MB. Rational selection of patient-reported outcomes measures in lumbar spine surgery patients. *Int J Spine Surg.* 2020;14(3):347–354. doi:10.14444/7046
- Pekkanen L, Kautiainen H, Ylinen J, Salo P, Häkkinen A. Reliability and validity study of the finnish version 2.0 of the oswestry disability index. *Spine (Phila Pa 1986).* 2011;36(4):332–338. doi:10.1097/BRS.0b013e3181cdd702
- Deyo RA, Battie M, Beurskens AJHM, et al. Outcome measures for low back pain research. *Spine (Phila Pa 1986).* 1998;23(18):2003–2013. doi:10.1097/00007632-199809150-00018
- Davidson M, Keating JL. A comparison of five low back disability questionnaires: reliability and responsiveness. *Phys Ther.* 2002;82(1):8–24. doi:10.1093/ptj/82.1.8
- Fritz JM, Irrgang JJ. A comparison of a modified oswestry low back pain disability questionnaire and the quebec back pain disability scale. *Phys Ther.* 2001;81(2):776–788. doi:10.1093/ptj/81.2.776

17. Hägg O, Fritzell P, Nordwall A, Swedish Lumbar Spine Study Group. The clinical importance of changes in outcome scores after treatment for chronic low back pain. *Eur Spine J*. 2003;12(1):12–20. doi:10.1007/s00586-002-0464-0
18. Chow SC, Shao J, Wang H, Lokhnygina Y. *Sample Size Calculations in Clinical Research*. Third Edition. New York: Imprint Chapman and Hall/CRC; 2017. doi:10.1201/9781315183084
19. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). *Biometrika*. 1965;52(3/4):591. doi:10.2307/2333709
20. R Core Team. *R Foundation for Statistical Computing V, Austria. R: A Language and Environment for Statistical Computing*. 2021. 2020.
21. Wickham H. *Ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag; 2016.
22. Mannion AF, Junge A, Fairbank JCT, Dvorak J, Grob D. Development of a german version of the Oswestry disability index. Part 1: cross-cultural adaptation, reliability, and validity. *Eur Spine J*. 2006;15(1):55–65. doi:10.1007/s00586-004-0815-0
23. Frost H, Lamb SE, Stewart-Brown S. Responsiveness of a patient specific outcome measure compared with the Oswestry disability index v2.1 and Roland and Morris disability questionnaire for patients with subacute and chronic low back pain. *Spine (Phila Pa 1986)*. 2008;33(22):2450–2457. doi:10.1097/BRS.0b013e31818916fd
24. Vishwanathan K, Braithwaite I. Construct validity and responsiveness of commonly used patient reported outcome instruments in decompression for lumbar spinal stenosis. *J Clin Orthop Trauma*. 2021;16:125–131. doi:10.1016/j.jcot.2021.01.002
25. 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):2187–2197. doi:10.1007/s00586-019-05989-0
26. Dybvik V, Hermansen E, Banitalebi H, Myklebust TÅ, Indrekvam K. Is repeated preoperative magnetic resonance imaging necessary before planned decompressive surgery for lumbar spinal stenosis? *Int J Spine Surg*. 2023;17(3):449–453. doi:10.14444/8469
27. Knutsson B, Mukka S, Wahlström J, Järvholm B, Sayed-Noor AS. The association between tobacco smoking and surgical intervention for lumbar spinal stenosis: cohort study of 331,941 workers. *Spine J*. 2018;18(8):1313–1317. doi:10.1016/j.spinee.2017.11.018
28. Sandén B, Försth P, Michaëlsson K. Smokers show less improvement than nonsmokers two years after surgery for lumbar spinal stenosis. *Spine (Phila Pa 1986)*. 2011;36(13):1059–1064. doi:10.1097/BRS.0b013e3181e92b36
29. Sekiguchi M, Yonemoto K, Kakuma T, et al. Relationship between lumbar spinal stenosis and psychosocial factors: a multicenter cross-sectional study (DISTO project). *Eur Spine J*. 2015;24(10):2288–2294. doi:10.1007/s00586-015-4002-2
30. Uesugi K, Sekiguchi M, Kikuchi S, Konno S. Relationship between lumbar spinal stenosis and lifestyle-related disorders. *Spine (Phila Pa 1986)*. 2013;38(9):E540–E545. doi:10.1097/BRS.0b013e31828a2517

Funding: This study was funded by State research funding of the Hospital District of Southwest Finland.

Disclosures: Inari Laaksonen reports payment for expert testimony from the supreme court of Finland and support for attending meetings and/or travel from Arthrex and Stryker. The remaining authors have no disclosures.

Corresponding Author: Inari Laaksonen, Department of Orthopaedics and Traumatology, Turku University Hospital, PO Box 52, FI-20521, Turku, Finland; inari.laaksonen@varha.fi

This manuscript is generously published free of charge by ISASS, the International Society for the Advancement of Spine Surgery. Copyright © 2025 ISASS. To see more or order reprints or permissions, see <http://ijssurgery.com>.