ORIGINAL ARTICLE

Case-matched radiological and clinical outcome evaluation of interlaminar versus microsurgical decompression of lumbar spinal stenosis

Babak Saravi[1,2] · Sara Ülkümen[1,2] · Gernot Lang[1,2] · Sébastien Couillard-Després[3,4] · Frank Hassel[2]

Abstract

Purpose Endoscopic spine surgery is a globally expanding technique advocated as less invasive for spinal stenosis treatment compared to the microsurgical approach. However, evidence on the efficiency of interlaminar full-endoscopic decompression (FED) vs. conventional microsurgical decompression (MSD) in patients with lumbar spinal stenosis is still scarce. We conducted a case-matched comparison for treatment success with consideration of clinical, laboratory, and radiologic predictors.
Methods We included 88 consecutive patients (FED: 36/88, 40.9%; MSD: 52/88, 59.1%) presenting with lumbar central spinal stenosis. Surgery-related (operation time, complications, length of stay (LOS), American Society of Anesthesiologists physical status (ASA) score, C-reactive protein (CRP), white blood cell count, side of approach (unilateral/bilateral), patient-related outcome measures (PROMs) (Oswestry disability index (ODI), numeric rating scale of pain (NRS; leg-, back pain), EuroQol questionnaire (eQ-5D), core outcome measures index (COMI)), and radiological (dural sack cross-sectional area, Schizas score (SC), left and right lateral recess heights, and facet angles, respectively) parameters were extracted at different time points up to 1-year follow-up. The relationship of PROMs was analyzed using Spearman's rank correlation. Surgery-related outcome parameters were correlated with patient-centered and radiological outcomes utilizing a regression model to determine predictors for propensity score matching.
Results Complication (most often residual sensorimotor deficits and restenosis due to hematoma) rates were higher in the FED (33.3%) than MSD (13.5%) group ($p < 0.05$), while all complications in the FED group were observed within the first 20 FED patients. Operation time was higher in the FED, whereas LOS was higher in the MSD group. Age, SC, CRP revealed significant associations with PROMs. We did not observe significant differences in the endoscopic vs. microsurgical group in PROMs. The correlation between ODI and COMI was significantly high, and both were inversely correlated with eQ-5D, whereas the correlations of these PROMs with NRS findings were less pronounced.
Conclusions Endoscopic treatment of lumbar spinal stenosis was similarly successful as the conventional microsurgical approach. Although FED was associated with higher complication rates in our single-center study experience, the distribution of complications indicated surgical learning curves to be the main factor of these findings. Future long-term prospective studies considering the surgical learning curve are warranted for reliable comparisons of these techniques.

Keywords Minimal-invasive · Spine · Endoscopic · Learning curve · Complications · Spinal stenosis

---

Babak Saravi and Sara Ülkümen have contributed equaly to this work.

✉ Babak Saravi
babak.saravi@jupiter.uni-freiburg.de

1 Department of Orthopedics and Trauma Surgery, Faculty of Medicine, Medical Centre - Albert-Ludwigs-University of Freiburg, Albert-Ludwigs-University of Freiburg, Hugstetterstrasse 55, 79106 Freiburg, Germany

2 Department of Spine Surgery, Loretto Hospital, Freiburg, Germany

3 Spinal Cord Injury and Tissue Regeneration Center Salzburg (SCI-TReCS), Institute of Experimental Neuroregeneration, Paracelsus Medical University, 5020 Salzburg, Austria

4 Austrian Cluster for Tissue Regeneration, Vienna, Austria

Introduction

Lumbar spinal stenosis (LSS) is one of the most prevalent clinical conditions, with a prevalence ranging from 11 to 25% in the general population [1]. As the incidence of LSS is known to increase with age, it is also the most common reason for spinal surgery in 65-years-old patients [2]. LSS is characterized by spinal canal narrowing due to flavum ligament hypertrophy, disk herniation, facet joint hypertrophy, trauma, tumors, or other congenital and acquired bone diseases [3]. Considering the biomechanical and neurological role of the lumbar area, symptomatic patients are highly restricted in their daily life, further perpetuating the medical condition and leading to a huge socioeconomic burden [4].

For a long time, conservative treatments, including physical therapy, analgesic medications, and epidural injections, remained the mainstay of therapy. Today as well, initial treatment of patients presenting with LSS typically includes several weeks of physical therapy, leading to reduced pain, disability, and the amount of pain medication intake [5, 6]. Surgical treatments are warranted when the aforementioned conservative therapies fail. Minimal-invasive therapeutic approaches seek to decrease tissue damage and the risk for iatrogenic instability while aiming for a faster rehabilitation of patients. Open micro-decompressive laminectomy is still considered the gold standard in patients with LSS. Although this technique allows good visualization of the operative field due to skin incision and preparation along with the paraspinal muscle while affording comfort for the performing surgeon, this might come at the cost of collateral tissue injury, in particular the multifidus muscle, which functions as a stabilizer of the spine [7, 8]. Further development in minimal-invasive spine surgery was made by the introduction of full-endoscopic decompression of spinal stenosis in the late 1990s [9]. Although it has been globally expanding since then, most surgeries utilizing this technique are still performed in specialized centers. High-resolution optics with a broad field of view, small-caliber working cannulas, spared paraspinal muscle, a multichannel irrigation system, and the possibility to introduce multiple small instruments enable a minimal-invasive manipulation of the surgical situs [10].

A reliable outcome evaluation will eventually require either a randomized controlled trial or a case-matched comparison when comparing two surgical techniques. Both types of evidence focusing on a broad range of clinically relevant, radiological, and laboratory outcome parameters are currently scarce. A systematic review and meta-analysis conducted by Phan et al. in 2017 included 23 studies comparing the full-endoscopic and the microsurgical technique [11]. Although the results of Phan et al.

indicate similar outcomes between both techniques, none of the assessed studies considered the surgical learning curve and a combination of clinical, radiological, and laboratory factors. Phan et al. concluded that further validation is required. An updated systematic review on this topic was recently published by Perez-Roman et al. in 2021 [12]. Like Phan et al. they concluded that both techniques were similarly successful. However, the endoscopic patients were reported to have reduced hospital stay and a trend to less perioperative blood loss. Nevertheless, most of the included studies focused on specific outcome parameters or considered only small numbers of confounding variables. A broader view on this topic focusing on several different datasets of patients and the learning curve could allow for more precise comparison. To include all patients already treated at our institution since the first introduction of the full-endoscopic interlaminar technique, we first sought to evaluate predictive factors affecting patient-related outcome measures (PROMs) and subsequently apply these findings to case-match compare patients from both groups. The general aim was to assess whether or not the full-endoscopic technique was superior to the microsurgical technique when patients were adjusted for relevant clinical and radiological baseline characteristics.

Methods

Study design

We performed a retrospective cohort study including consecutive patients treated with microsurgical decompression or full-endoscopic interlaminar decompression of lumbar spinal stenosis between 2018 and 2020.

The main inclusion criterion involved patients with lumbar spinal stenosis treated with either microsurgical or full-endoscopic decompression in the aforementioned period. The iLESSYS® system (Joimax GmbH, Karlsruhe, Germany) was utilized for the endoscopic group. After collecting all data from patients fulfilling our inclusion criteria, we applied our exclusion criteria for filtering the initial dataset. Exclusion criteria included: (1) < 18-years-old patients, patients with tumors of the spine, patients having spinal fusion, patients having less than one year of follow-up data, and patients who have declined the usage of their data for research purposes.

Data handling

Patients were collected from the in-house patient information system and extracted into a predefined datasheet. Data were pseudonymized utilizing a code generated with the

"encode" command in Stata statistical software release 15 (StataCorp. 2011, College Station, TX, USA).

Four main groups of variables were included in the data extraction form. The surgery-related and clinical factor variable group included operation time (OT), length of stay (LOS), the American Society of Anesthesiologists (ASA) physical status classification. This group also contained the demographic data for descriptive statistics (e.g., sex, age). The laboratory variable group included C-reactive protein (CRP) and white blood cell count (WBS). The radiological variable group included preoperative dural sack cross-sectional area (DSCA), Schizas score (SC), left (LRH) and right (RRH) lateral recess heights, and left (LFA) and right (RFA) facet angles. Two examiners conducted radiological measurements using the approach shown by Schizas et al. [13] for SC, Iwahashi et al. [14] for DSCA, and Wu et al. [15] for LRH, RRH, LFA, RFA. The mean of both examiner values was used for statistics. The patient-related outcome measures (PROMs) group were targeted as the dependent variables for outcome evaluations and included the German version of the Oswestry disability index (ODI) [16], core outcome measures index (COMI) [17], the numeric rating scale of leg and back pain [18], and the eQ-5D health questionnaire [19]. Patients had follow-ups at regular time points to evaluate treatment success, and PROMs were assessed preoperatively, 3 weeks, and 1 year postoperatively. Patient data were screened for the following complications: residual sensorimotor deficits or new-onset sensorimotor deficits, hematomas requiring revision, persisting stenosis requiring revision, postoperative instability, and fracture.

Statistical analysis

Data were first evaluated with descriptive statistics. Comparability of baseline characteristics was assessed with Chi^2-tests (categorical variables) and t-tests or its nonparametric alternative (continuous variables) where applicable. PROMs for the three time points (preoperatively, 3 weeks, and 1 year postoperatively) were then compared utilizing a two-way ANOVA with repeated measures design and a multiple comparison post hoc test. Bar charts with

mean ± standard error of the mean were created in GraphPad Prism Software version 8.2.1 (GraphPad Software, Inc., San Diego, CA). Hereafter, we checked for variables showing an association with the PROMs utilizing a multivariate linear mixed effect model that includes both fixed and random effects to account for the within-participant repeated measures outcome evaluation. Variables showing significant associations with PROMs were then used to match patients from the two treatment groups utilizing propensity score matching analysis using at least 8 matches to estimate the coefficient in the logit treatment model [20]. We further compared the included PROMs to evaluate how they correlate with each other using Spearman's rank correlation test with a Bonferroni adjusted significance level, considering that the data did not pass the Shapiro–Wilk test for normality. Spearman's rank correlation test was also utilized to examine the relationship between the number of surgeries and the operation time as indicators for the surgical learning curve. Finally, we applied a simple multilayer perceptron (MLP) neural network model with three layers (input-, hidden-, and output-layer; maximal 50 units per layer) and a hyperbolic tangent activation function to predict the PROMs from the input data (predictors). Standardized rescaling was applied to all variables. We used 70% of the data for training and 30% for testing. Subsequently, a feature importance analysis was performed, and the relative error was assessed to evaluate the model. A p-value < 0.05 was deemed to be significant. Statistical analyses were conducted in Stata Statistical Software Release 15 (StataCorp. 2011, College Station, TX, USA) and SPSS v26 (IBM, Armonk, NY, USA).

Results

Baseline characteristics

The cohort consisted of $n = 88$ patients. A total of $n = 36$ (male: 20, female: 16) patients underwent full-endoscopic decompression (FED), whereas $n = 52$ (male: 31, female: 21) patients received microsurgical decompression (MSD). The most frequent operated surgical levels

Table 1 Comparison of baseline characteristics between the full-endoscopic decompression (FED) and the microsurgical decompression (MSD) group

Study groups				
Variable	All ($n=88$) n	FED ($n=36$) n (%)	MSD ($n=52$) n (%)	p-value
Gender				
Female	37	16 (4444)	21 (40.38)	
Male	51	20 (55.56)	31 (59.62)	ns
Age (years)	70.89 ± 1.06	69.25 ± 1.64	72.04 ± 1.39	ns
Body mass index (kg/m²)	30.12 ± 16.99	27.43 ± 4.73	31.98 ± 21.66	ns

Ns—not significant

were L4/L5 (37/88, 42.05%), followed by L3-L5 (12/88, 13.64%) and L3/L4 (12/88, 13.64%). Table 1 summarizes and compares the baseline characteristics of both groups. The mean age of patients from the FED group (69.25 ± 9.84 years) did not significantly differ from the MSD group (72.03 ± 10.16 years; $p = 0.1994$). This was also found for the BMI and the distribution of the gender between the groups.

Table 2 summarizes and compares the intraoperative, perioperative and postoperative characteristics of both groups. The distribution of Schizas Scores did not reveal a significant difference between the groups as assessed with the Chi²-test ($p = 0.163$). The FED group had score C as the most frequent Schizas score ($n = 12$), followed by D ($n = 4$), B ($n = 4$, and A4. Similarly, the MSD group had score C as the most frequent Schizas score ($n = 14$), followed by A2 ($n = 12$), A3 ($n = 9$), and D ($n = 3$). In addition, median ASA scores for the FED (median: 2; range: 1–3) and MSD (median: 3; range: 1–4) were homogeneously distributed and did not show significant differences ($p = 0.465$). Also, the mean number of treated levels did not significantly differ between the FED (2.19 ± 0.53) and the MSD (2.48 ± 0.69). In contrast, we found a significant difference in OT (FED: 78.67 ± 3.89 min versus MSD: 68.54 ± 5.04 min; $p = 0.0138$) and LOS (FED: 7.14 ± 8.09 days versus 11.19 ± 6.12 days; $p < 0.0001$). Furthermore, the preoperative DSCA was significantly lower in FED (81.39 ± 4.79 mm²) than MSD (95.67 ± 4.89 mm²) ($p = 0.0243$).

Complications

We observed a total of 19 complications (Fig. 1). The complication rate was higher in the FED (33.3%) than in the MSD group (13.46%). In the FED group, complications were residual sensorimotor deficits ($n = 9$) and restenosis due to hematoma ($n = 2$). In the MSD group, complications were postoperative fractures requiring revision ($n = 2$), hip flexor paresis ($n = 1$), restenosis due to hematoma ($n = 2$), and revision due to persisting stenosis ($n = 1$). Assessment of the distribution of complications over the consecutive patients included since the first FED surgery revealed that all complications in the FED group were observed in the first 20 patients after introducing the full-endoscopic technique in our hospital and were then absent. To consider the learning curve evaluation findings in the statistical comparisons, we also compared the first 20 full-endoscopic patients versus the last 16 full-endoscopic patients and the MSD group regarding the operation time. We found that the MSD group has a significantly lower operation time compared to the first full-endoscopic patients ($p < 0.05$). However, we found no significant differences between the MSD group and the last $n = 16$ full-endoscopic patients treated at our institution ($p = 0.322$). The first $n = 16$ full-endoscopic patients showed an operation time of 88.7 ± 24.6 min, whereas the last $n = 20$ endoscopic

Table 2 Comparison of intraoperative, perioperative, and postoperative characteristics between the full-endoscopic decompression (FED) and the microsurgical decompression (MSD) group

Variable	All ($n=88$) n	FED ($n=36$) n (%)	MSD ($n=52$) n (%)	p-value
Number of levels	2.36 ± 0.65	2.19 ± 0.53	2.48 ± 0.69	ns
Side of approach				
Unilateral	53	22 (61.11)	31 (59.62)	
Bilateral	35	14 (38.89)	21 (40.38)	ns
Length of stay (days)	9.53 ± 7.23	7.14 ± 8.09	11.19 ± 6.12	< 0.0001****
Complications				
No	69	24 (66.67)	45 (86.54)	
Yes	19	12 (33.33)	7 (13.46)	0.026*
Operation time (minutes)	72.68 ± 3.41	78.67 ± 3.89	68.54 ± 5.04	0.0138*
White blood cell count (preoperative) (per nL)	8.65 ± 1.08	9.64 ± 2.59	7.96 ± 0.36	ns
C-reactive protein (preoperative) (mg/L)	3.80 ± 0.65	2.7 ± 0.56	4.57 ± 1.03	ns
Hemoglobin (preoperative) (g/dL)	13.9 ± 1.38	13.79 ± 1.45	14.09 ± 1.31	ns
Dural sack cross-sectional area (preoperative) (mm²)	89.83 ± 3.57	81.39 ± 4.79	95.67 ± 4.89	0.0243*
Left recess height (preoperative) (mm)	2.04 ± 0.11	2.1 ± 0.19	2.0 ± 0.14	ns
Right recess height (preoperative) (mm)	2.2 ± 0.12	2.37 ± 0.22	2.16 ± 0.13	ns
Left recess angle (preoperative)	6.18 ± 0.51	6.73 ± 0.86	5.79 ± 0.63	ns
Right recess angle (preoperative)	4.51 ± 0.46	4.58 ± 0.57	4.47 ± 0.67	ns

Ns—not significant

* indicates statistical significance

Fig. 1 Distribution of complications in the full-endoscopic (FED; **A**), microsurgical (MSD; **B**) decompression group and total (**C**)
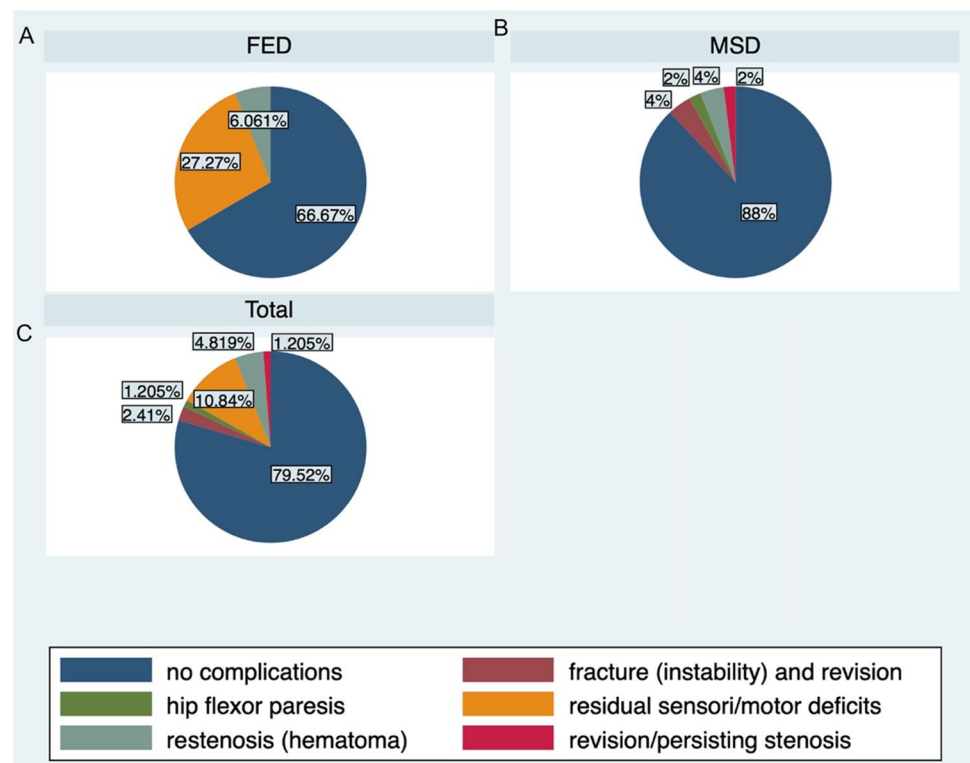
patients showed an operation time of 70.65 ± 19.42 min ($p = 0.006$). We further examined whether this finding was due to a lower number of segments treated or the side of surgical approach. The number of segments treated in the first $n = 16$ full-endoscopic patients was 2.06 ± 0.25 whereas the number of segments treated in the last $n = 20$ full-endoscopic patients was 3.10 ± 0.31. Moreover, the number of unilateral and bilateral surgical approach in the first $n = 16$ full-endoscopic patients was $n = 8$ for each approach. For the last $n = 20$ full-endoscopic patients the number of unilateral and bilateral surgical approach was $n = 13$ and $n = 7$, respectively. Figure 2 illustrates the development of the study's learning curve. It is evident that a low number of case surgeries performed leads to highly complication rates and longer operation times for the FED group, whereas for the MSD group this finding was not clearly evident, as the surgeons had already experience with the MSD technique. These findings indicate that the lower operation time in the last $= 20$ patients was probably not due to a higher number of segments or a higher number of bilateral surgical approaches.

PROMs

All PROMs showed marked improvements for both groups over time (Fig. 3). In the FED group, the ODI decreased from 39.94 ± 11.87 preoperatively to 19.89 ± 14.67 and 7.67 ± 7.58 after 3 weeks and 1 year postoperatively

($p < 0.0001$). Similarly, ODI decreased from 39.39 ± 14.29 preoperatively to 23.08 ± 14.47 and 13.22 ± 17.19 after 3 weeks and 1 year postoperatively. Comparable to the ODI evaluations, COMI showed a significant decrease from the preoperative situation (FED: 5.97 ± 1.01; MSD: 5.62 ± 1.51) to the 3 weeks (FED: 2.51 ± 2.82; MSD: 2.23 ± 1.87) and 1-year (FED: 1.39 ± 1.66; MSD: 1.84 ± 2.02) values ($p < 0.0001$). In addition, there was an increase in eQ-5D from preoperative assessments (FED: 0.79 ± 0.05; MSD: 0.79 ± 0.05) to the 3 weeks (FED: 0.89 ± 0.08; MSD: 0.89 ± 0.09) and 1 year (FED: 0.95 ± 0.08; MSD: 0.93 ± 0.09) data ($p < 0.0001$). NRS for leg and back pain also showed marked improvements up to the 1-year data. However, most improvements occurred within the 3 weeks postoperatively, and there was no more significant improvement between the 3 weeks and the 1-year data for both groups.

Mixed-effects linear regression model

Additionally, we constructed a mixed-effects linear regression model for each PROM to find variables and factors significantly associated with PROMs (Fig. 4). Results for the dependent variable ODI showed that age ($p = 0.010$) DIH ($p < 0.0001$), OT ($p = 0.031$), DSCA ($p < 0.0001$), Schizas score ("D" vs. "A2" $p = 0.003$), ASA (2 vs 1: $p = 0.036$, 3 vs. 1: $p = 0.024$), sex (male vs. female: $p = 0.005$), and preoperative CRP ($p < 0.0001$) showed significant associations.
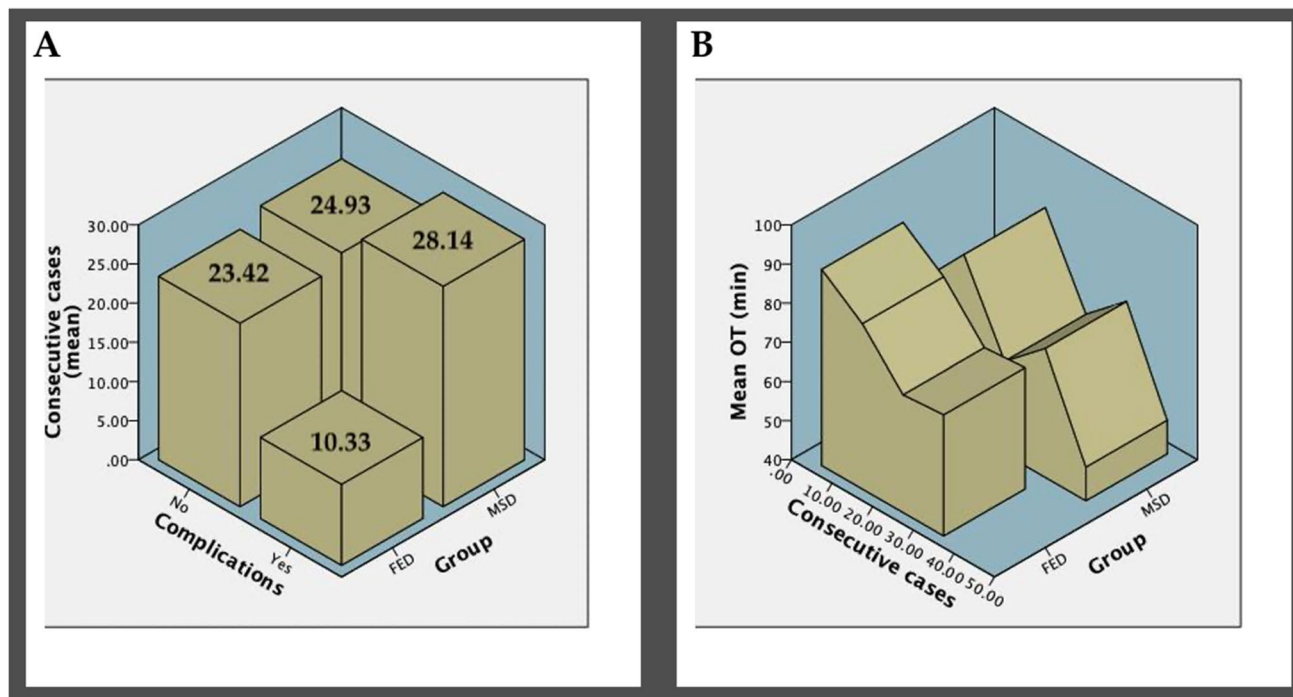
Fig. 2 Illustration of the study's learning curve. Subfigure **A** shows the mean number of consecutive cases in each group stratified by whether a complication was observed or not and the surgical technique (FED versus MSD). As can be seen from the figure, complications were present in the FED group when a small number of cases were performed, whereas for the MSD group, this phenomenon was not present, as the surgeons already had experience with the MSD technique. Subfigure **B** illustrates the nearly linear decrease in operation time (OT) in the FED group until the 20th consecutive case surgery performed. In the case of MSD, this phenomenon was not clearly evident

Furthermore, measurements after 3-weeks (reg. coeff.: −3.33, 95% CI −3.8 to −2.77, $p < 0.0001$) and 1-year (reg. coeff.: −4.76, 95% CI −4.77 to −3.65, $p < 0.0001$) revealed significant improvement of ODI. COMI, age, and Schizas score revealed significant associations, whereas, for the eQ-5D, only preoperative CRP showed significance. NRS-leg, age, surgical side, sex, and CRP levels showed significant associations, whereas in NRS-back only surgical side, Schizas Score and preoperative CRP levels showed significance. All PROMs showed significant improvements over time regardless of these associations between the assessed independent variables and the PROMs. We decided to use age, Schizas score, and preoperative CRP levels to match patients based on these findings.

Propensity score matching based on the independent variables age, Schizas score, and preoperative CRP revealed no significant associations between the treatment dependent (FED vs MSD) and the ODI (reg. coeff.: 1.65, 95% CI −4.34 to 7.64, $p = 0.589$), COMI (reg. coeff.: −0.33, 95% CI −1.18 to 0.52, $p = 0.449$), eQ-5D (reg. coeff.: 0.0004, 95% CI −0.33 to 0.034, $p = 0.0983$), NRS-Leg (reg. coeff.: 0.32, 95% CI −0.76 to 1.40, $p = 0.560$), and NRS- Leg (reg. coeff.: −0.901, 95% CI −2.05 to 0.245, $p = 0.123$). Overall, results indicate no difference between both techniques in improving

the PROMs. The correlation between ODI and COMI was significantly high, and both were inversely correlated with eQ-5D, whereas the correlations of these PROMs with NRS findings were less pronounced. Spearman's rho revealed a significant inverse relationship between the number of full-endoscopic surgeries and the operation time (rho = −0.4219; $p = 0.0104$), whereas there was no significant correlation for the microsurgical group.

Multilayer perceptron (MLP) neural network model to predict PROMs

The feature importance analysis results of the multilayer perceptron neural network model were similar to the findings of the linear mixed-effects model, revealing that time, preoperative CRP, length of stay, age, and the preoperative dural sack cross-sectional area are at least partly effective in predicting PROMs in our cohort (Fig. 5). Whereas the NRS for back, leg, eQ-5D were not adequately predicted with relative errors of more than 40%, the analysis for COMI reached a relative error of 29.1%. Adjustment of the model, e.g., dropping variables and changing the train-test-split fraction, did not improve the model further.

Fig. 3 Bar charts show the ODI (**A**), COMI (**B**), NRS-leg (**C**), NRS-back (**D**), and eQ-5D (**E**) difference between time points for the full-endoscopic (FED) and the microsurgical (MSD) group. Pre-OP: preoperatively; Post-OP (3 weeks): 3 weeks postoperatively; Post-OP (1 year): 1 year postoperatively. Bar charts show the mean $\pm$ sd. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, $****p < 0.0001$; FED: full-endoscopic group MSD: microsurgical decompression group
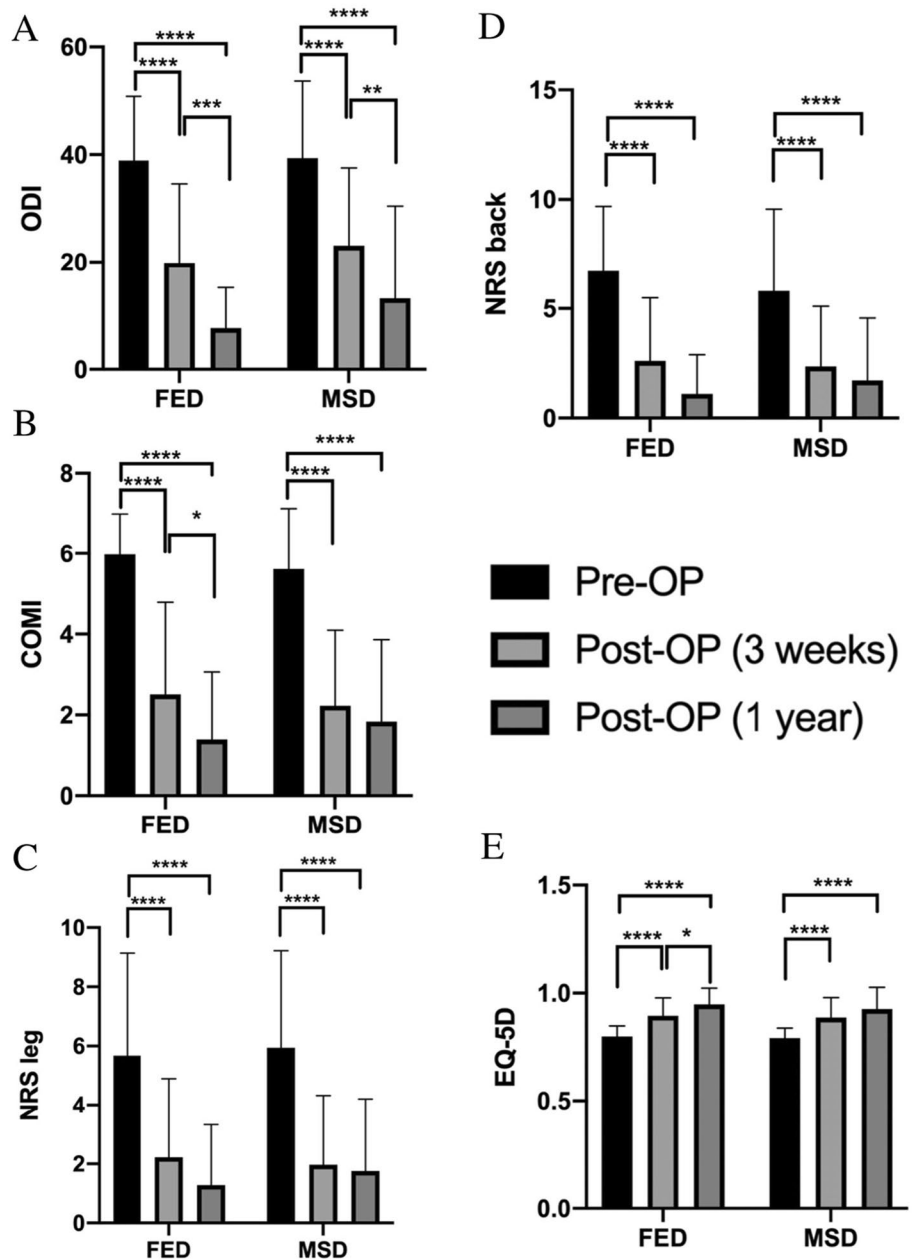
Discussion

Our study aimed first to investigate a range of variables relevant for a reliable comparison of the full-endoscopic versus microsurgical decompression in our institution; secondly, we sought to assess predictive factors associated with PROMs and third use these findings to compare the techniques' ability to improve PROMs reliably. Our approach allowed us to overcome the lack of a randomized controlled trial design by using adequate case-matching of patients assessed retrospectively. Endoscopic treatment of lumbar spinal stenosis was similarly successful as the conventional microsurgical approach, although it was associated with higher complication rates in our single-center study experience. The distribution of complications indicated surgical learning curves to be the main factor of these findings.

FED and MSD provide equivalent PROM-improvement, but FED comprises higher complication rates

The results show that both techniques are comparable in improving PROMs without one showing signs of superiority with regards to PROMs as the outcome of interest. However, we observed more complications in the FED group. Notably, these complications occurred within the
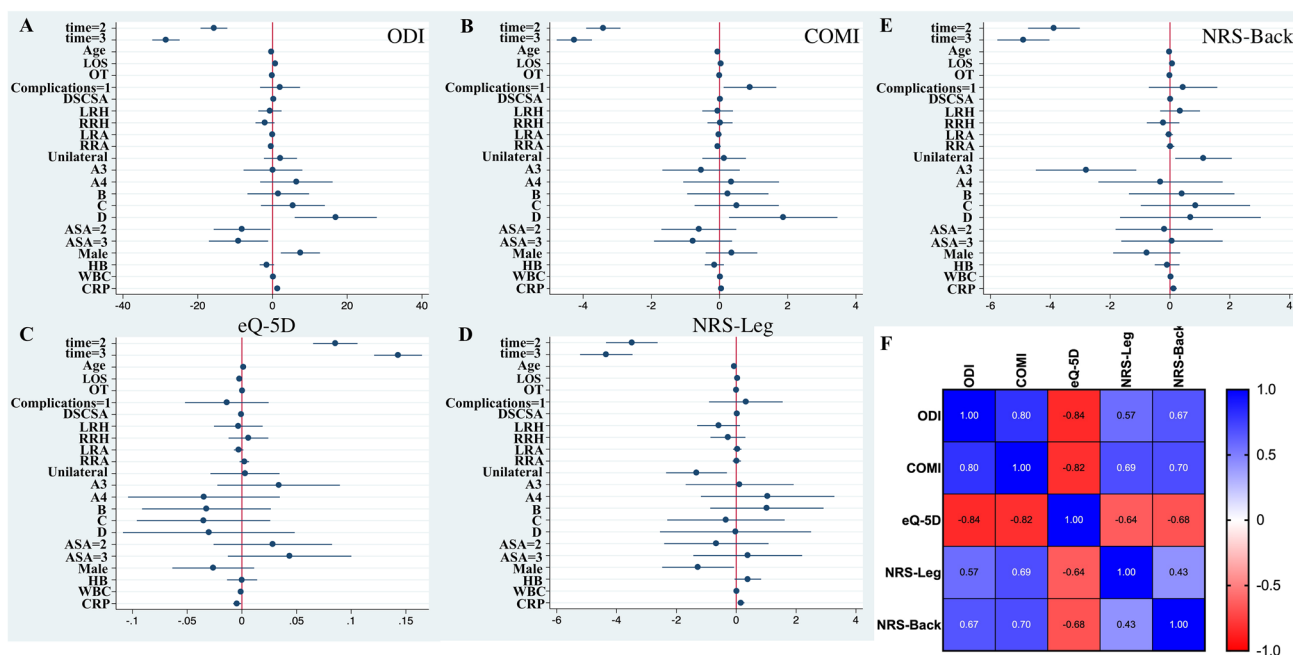
Fig. 4 Illustration of the regression analysis via forest Plots (**A**–**E**) and correlation analysis via a heatmap (**F**). The forest plots show the regression coefficients and the 95% CI of the mixed-effects linear regression models for the dependent variables ODI (**A**), COMI (**B**), eQ-5D (**C**), NRS-Leg (**D**), and NRS-back(**E**). Nominal variables are compared by comparing the shown factors against a dummy variable. The heatmap (**F**) shows the Spearman rank correlation coefficients for the comparison of the PROMs. All shown comparisons in the heatmap were significant ($p < 0.0001$). DSCA—dural sack cross-sectional area; LRH—left recess height; RRH—right recess height; LRA—left recess angle; RRA—right recess angle; HB—preoperative hemoglobin; WBC—preoperative white blood cell count; LOS—Length of Stay; OT—operation time; CRP—preoperative c-reactive protein; ASA—physical status score (dummy factor was ASA score 1); A3/A4/B/C/D: Schizas Score against the dummy factor A2; time=2/time=3: timepoints 3 weeks and 1-year postoperatively against the dummy factor "preoperatively"; Complications=1 (binary): complication occurred against not occurred (0). Male: coefficients are shown against the dummy factor "female"

first $n = 20$ patients treated with the full-endoscopic technique in our hospital, and thus this high rate was most likely based on the surgeon's learning curve. Further, our analysis revealed a significant inverse relationship between the number of full-endoscopic surgeries and the operation time (rho $= -0.4219$; $p = 0.0104$), whereas there was no significant correlation for the microsurgical group. Further, we did not observe significant differences in operation time between the last 16 FED-surgeries and the MSD group. These findings indicate that the number of surgeries performed significantly lowered the complication rates and reduced the operation time. Our results are consistent with the learning curve assessment of Zelenkov et al. who reported that the plateau of the learning curve of full-endoscopic interlaminar and transforaminal surgery would be achieved within the first 20 patients [21]. As our surgeon had extensive practice in the microsurgical technique, we cannot currently define whether the full-endoscopic procedure can be generally classified as having a "steep learning curve" or a "shallow learning curve" according to the definition provided by Benzel et al. [22].

In the most extensive analysis of the learning curve in endoscopic decompression of lumbar spinal stenosis, Lee et al. showed that the complication rates were higher and operation times were longer in the first cohort of patients treated with FED [23]. After the 100th case, the plateau of the operation time was reached, translating to a rather steep learning curve [23]. In addition, the complication rates in the first cohort of patients were twice as high compared to the more experienced phase of the learning curve [23]. This might also explain the higher operation time in FED compared to MSD in our cohort. The first 20 FED patients were included in the statistics and generally showed higher OT than those treated hereafter. Other authors generally reported lower complication rates for FED compared to MSD [24–26]. However, no information was provided regarding the learning curves of the surgeons. To summarize, literature evidence and our findings indicate that complication rates might be higher for the FED group in the first cohort of patients, but are likely to become lower versus MSD after the plateau of the learning curve has been reached for FED.
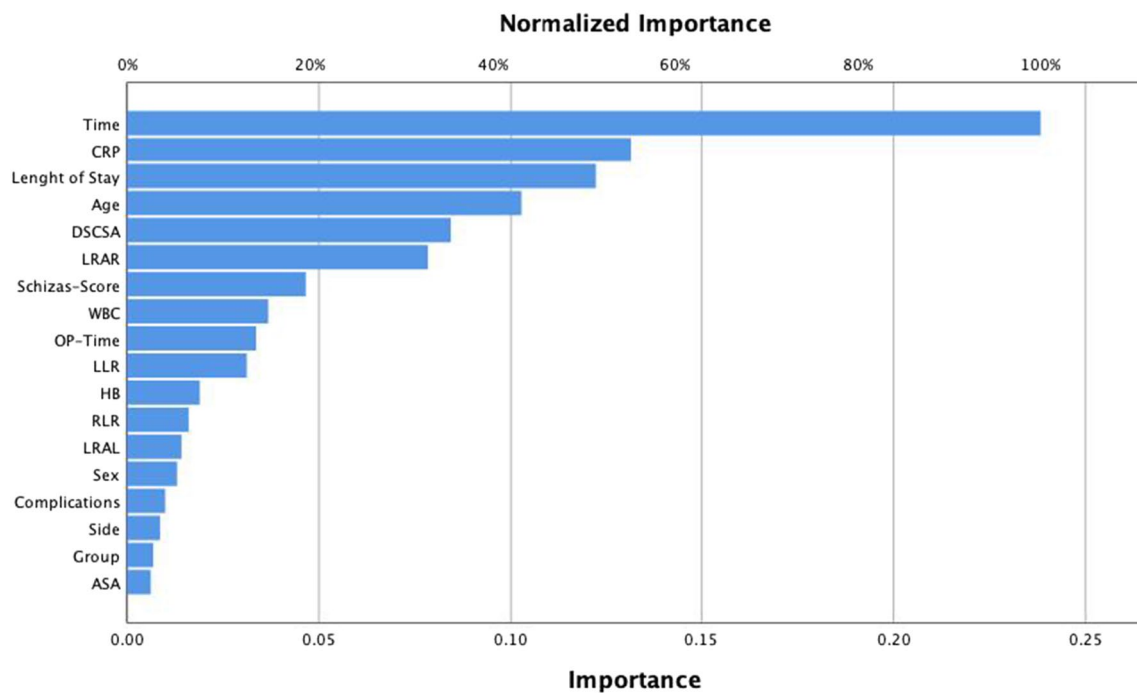
Normalized Importance

Fig. 5 Feature importance analysis utilizing the multilayer perceptron neural network model. DSCA—dural sack cross-sectional area; LRH—left recess height; RRH—right recess height; LRA—left recess angle; RRA—right recess angle; HB—preoperative hemoglobin; WBC—preoperative white blood cell count; CRP: preoperative c-reactive protein; ASA: physical status score

Further considerations and perspectives

In contrast to Marković et al. we did not find that the full-endoscopic technique has better outcomes in pain and disability scales [27]. Nevertheless, their data were obtained over a 3-years period, and we cannot rule out the possibility that PROMs will become different after 1-year examinations. Thus, a long-term evaluation is warranted using either a case-matched design or a randomized controlled trial. Consequently, we are currently conducting a comprehensive cohort trial utilizing a broad range of relevant outcomes to overcome this lack of evidence [German clinical trials register (DRKS): DRKS-ID: DRKS00025786]. In addition, we did not focus on other relevant factors which might influence the implementation of the technique in hospitals, such as cost analysis comparisons. In a previous report, cost analysis comparisons between MSD and FED revealed that both procedures had similar costs in hospitalization, radiology, and follow-up visits. Although costs for FED were 5.7% higher for the unit to run the operations, MSD was 28.1% more expensive than FED when comparing complication rates, which were 3.8% for FED and 7.5% for MSD [24].

The full-endoscopic technique to treat lumbar spinal stenosis is in advance. Scarring of the epidural space, the route of access potentially leading to instability of the coordination system, and the generally larger amount of soft-tissue resection might justify the shift toward a more tissue-sparing technique [25]. Constant technical advantages regarding the visualization of the operation situs utilizing modern optics might allow better progress for the FED than the MSD, probably affecting future outcome evaluations. Furthermore, the broad application of cadaver courses might improve and enhance the complicated learning curve [28]. Especially the fact that surgeons in Asia report higher self-reported skill levels and that endoscopic spine surgery training in Asia is reported to be better implemented in the daily practice of spine surgeons might be the reason for the tendency of better outcomes results for the FED compared to MSD in Asian publications [28]. Interestingly, reports from non-Asian countries generally include more comparable results for the FED verses MSD [25, 29–41], compared to Asian country publications [42, 43], which seem to favor FED. However, a future meta-analysis using the publication region as the confounding variable in the meta-regression and subgroup analysis model is warranted to provide an in-depth analysis of this phenomenon. In accordance with Chen et al. we did not find a general tendency of the assessed variables to affect all PROMs similarly [44]. However, they determined alcohol use to be associated with higher re-operation rates. Unfortunately, we could not include this variable as there was no sufficient data available. In contrast, we additionally assessed laboratory markers and radiological markers compared to Chen

et al. We found that preoperative CRP levels and a high Schizas score were associated with worse PROMs, particularly for the ODI. Notably, the fact that preoperative CRP levels influenced the PROMs might be of relevance and requires future exploration in prospective studies. The neural network model applied by us did confirm the relevance of several predictors for PROMs. However, the relative errors were not satisfying, probably due to the limited sample size. Furthermore, there might be an information loss when radiological images are measured, and the data are fed into a machine learning model compared to an approach where the radiological images are directly combined with the clinical data (multi-input, mixed-data model). We are currently collecting prospective data to feed and train a multi-input, mixed-data neural network model and predict spine surgery patients' PROMs based on a combination of radiological, clinical, and laboratory predictors. This will allow evaluating whether the patient could benefit from surgery and which surgical approach could be better suited for each patient based on the patient's individual data.

The present study showed that the early learning curve for the FED technique is associated with complications and action must be taken to improve the outcomes in the implementation phase. In order to decrease the surgical learning curve, it could be necessary for surgeons to perform more supervised surgeries before they are able to perform a particular surgery on their own. As part of the requirement, there may be several "hands-off" observations, followed by assistance and then supervision. It is also possible to incorporate simulations, cadavers, and animal surgeries into the course. There are many factors that have to be considered in defining the parameters of each phase of the learning curve. As shown, the appropriate number of surgeries before a surgeon can apply the technique with confidence will depend on the type of surgery and how many experience the surgeon has gained with other similar techniques. Rather than assuming that things will go according to plan, comprehensive training should include simulations of responses to known contingencies. A regimen such as that would be far more extensive than the brief training periods currently common. The surgeon must accept that a high level of supervised training will be required. There are logistical, cost, and personal considerations. Furthermore, hospitals will have to take a greater role in determining what type of surgery individual surgeons are permitted to perform and what training and experience are required. A hospital's credentialing process must take this into account before implementing new techniques. According to the present findings, fast implementations could negatively impact patient outcomes and result in higher costs in the long run than rigorous training at the early phase of the learning curve.

Strengths and limitations

Our study is associated with certain strengths and limitations. One of the main strengths is the extraction of several confounding variables which potentially affect PROMs. Considering these variables in our regression model allowed a more precise estimation of regression coefficients than existent in studies to date on this topic. Furthermore, we applied a propensity score matching based on this finding, one of the state-of-the-art techniques to maintain comparability between groups in non-randomized study designs [20]. Another advantage is the consideration of the surgeon's learning curve in our results and interpretations as we included all FED patients since inception. Therefore, the results are especially interesting for clinicians who want to apply this technique in their institutions and are interested in relevant outcomes in the "implementation phase." Study limitations include the retrospective design, which has several disadvantages, such as the necessity to apply multiple statistical models to allow comparability, which themselves can introduce some susceptibility to error. Furthermore, this study type is prone to selection bias and misclassification bias, as we had to use the data as provided in our patient information system without further validation, and additional consultations to extract missing data in some variables are often not possible. Alterations in CRP levels are known to be associated with surgical trauma, and peaks are usually observed after 48 h postoperatively [45, 46]. Nevertheless, the CRP alterations as a response to the iatrogenic traumatic injury are highly variable and dependent on numerous patient characteristics [45] and may even be absent in some patients [47]. Due to the retrospective design, we cannot validate whether other patient characteristics might have affected the CRP findings as no randomization was performed. Thus, comparisons of CRP levels between studies might be limited and outcome interpretations and comparisons of the techniques should rely more on the PROMs than on surrogate markers such as CRP. Nevertheless, they can help to identify adverse outcomes, and CRP levels can be used as one quantifying parameter for the degree of surgical trauma [48], although there are also controversial statements in this regard [49].

Furthermore, we could not include other variables which could be relevant such as interleukin-6 as a surrogate marker for tissue damage, as these are not regularly measured and thus not available in the patient information system. Therefore, a large-scale prospective cohort focusing on a broad range of relevant outcomes is warranted to improve the current knowledge.

Conclusions

Endoscopic treatment of lumbar spinal stenosis was similarly successful as the conventional microsurgical approach, although it was associated with higher complication rates in our single-center study experience. The distribution of complications indicated different phases on the learning curve to be the main factor of these findings. Operation time was higher in the FED group, whereas LOS was higher in the MSD group. Future long-term prospective studies considering the learning curve are warranted for reliable comparisons of these techniques.

Declarations

Conflict of interest The authors of the present manuscript declare no conflict of interest.

Ethical approval This retrospective observational study was approved by the local Ethics Committee Freiburg, Germany [Number: 116/200]. Written informed consent to participate in observational studies was obtained from each patient.

References

1. Jensen RK, Jensen TS, Koes B, Hartvigsen J (2020) Prevalence of lumbar spinal stenosis in general and clinical populations: a systematic review and meta-analysis. Eur Spine J 29:2143–2163. https://doi.org/10.1007/s00586-020-06339-1

2. Deyo RA, Gray DT, Kreuter W et al (2005) United States trends in lumbar fusion surgery for degenerative conditions. Spine 30:1441–1445. https://doi.org/10.1097/01.brs.0000166503.37969.8a

3. Levy RM, Deer TR (2012) Systematic safety review and meta-analysis of procedural experience using percutaneous access to treat symptomatic lumbar spinal stenosis. Pain Med 13:1554–1561. https://doi.org/10.1111/j.1526-4637.2012.01504.x

4. Bagley C, Macallister M, Dosselman L et al (2019) Current concepts and recent advances in understanding and managing lumbar spine stenosis. F1000 Res 8:137. https://doi.org/10.12688/f1000research.16082.1

5. Diwan S, Sayed D, Deer TR et al (2019) An algorithmic approach to treating lumbar spinal stenosis: an evidenced-based approach. Pain Med 20:S23–S31. https://doi.org/10.1093/pm/pnz133

6. Slater J, Kolber MJ, Schellhase KC et al (2016) The influence of exercise on perceived pain and disability in patients with lumbar spinal stenosis: a systematic review of randomized controlled trials. Am J Lifestyle Med 10:136–147. https://doi.org/10.1177/1559827615571510

7. Hasan S, McGrath LB, Sen RD et al (2019) Comparison of full-endoscopic and minimally invasive decompression for lumbar spinal stenosis in the setting of degenerative scoliosis and spondylolisthesis. Neurosurg Focus 46:E16. https://doi.org/10.3171/2019.2.FOCUS195

8. Lee C-H, Choi M, Ryu DS et al (2018) Efficacy and Safety of Full-endoscopic Decompression via Interlaminar Approach for Central or Lateral Recess Spinal Stenosis of the Lumbar Spine: A Meta-analysis. Spine 43:1756–1764. https://doi.org/10.1097/BRS.0000000000002708

9. Kim M, Kim H-S, Oh SW et al (2019) Evolution of spinal endoscopic surgery. Neurospine 16:6–14. https://doi.org/10.14245/ns.1836322.161

10. Liu Y, Kim J-S, Chen C-M et al (2021) A review of full-endoscopic interlaminar discectomy for lumbar disc disease: a historical and technical overview. J Minim Invasive Spine Surg Tech 6:S109–S116. https://doi.org/10.21182/jmisst.2021.00059

11. Phan K, Xu J, Schultz K et al (2017) Full-endoscopic versus micro-endoscopic and open discectomy: a systematic review and meta-analysis of outcomes and complications. Clin Neurol Neurosurg 154:1–12. https://doi.org/10.1016/j.clineuro.2017.01.003

12. Perez-Roman RJ, Gaztanaga W, Lu VM, Wang MY (2021) Endoscopic decompression for the treatment of lumbar spinal stenosis: an updated systematic review and meta-analysis. J Neurosurg Spine. https://doi.org/10.3171/2021.8.SPINE21890

13. Schizas C, Theumann N, Burn A et al (2010) Qualitative grading of severity of lumbar spinal stenosis based on the morphology of the dural sac on magnetic resonance images. Spine 35:1919–1924. https://doi.org/10.1097/BRS.0b013e3181d359bd

14. Iwahashi H, Yoshimura N, Hashizume H et al (2016) The association between the cross-sectional area of the dural sac and low back pain in a large population: the Wakayama spine study. PLoS ONE 11:e0160002. https://doi.org/10.1371/journal.pone.0160002

15. Wu A-M, Zou F, Cao Y et al (2017) Lumbar spinal stenosis: an update on the epidemiology, diagnosis and treatment. AME Med J 2:63

16. Fairbank JC, Pynsent PB (2000) The Oswestry disability index. Spine 25:2940–2952. https://doi.org/10.1097/00007632-20001150-00017

17. Mannion AF, Vila-Casademunt A, Domingo-Sàbat M et al (2016) The core outcome measures index (COMI) is a responsive instrument for assessing the outcome of treatment for adult spinal deformity. Eur Spine J 25:2638–2648. https://doi.org/10.1007/s00586-015-4292-4

18. Haefeli M, Elfering A (2006) Pain assessment. Eur Spine J 15(Suppl 1):S17-24. https://doi.org/10.1007/s00586-005-1044-x

19. Garratt AM, Furunes H, Hellum C et al (2021) Evaluation of the EQ-5D-3L and 5L versions in low back pain patients. Health Qual Life Outcomes 19:155. https://doi.org/10.1186/s12955-021-01792-y

20. Kuss O, Blettner M, Börgermann J (2016) Propensity score: an alternative method of analyzing treatment effects. Deutsches Aerzteblatt Online. https://doi.org/10.3238/arztebl.2016.0597

21. Zelenkov P, Nazarov VV, Kisaryev S et al (2020) Learning curve and early results of interlaminar and transforaminal full-endoscopic resection of lumbar disc herniations. Cureus 12:e7157. https://doi.org/10.7759/cureus.7157

22. Benzel EC, Orr RD (2011) A steep learning curve is a good thing! Spine J 11:131–132. https://doi.org/10.1016/j.spinee.2010.12.012

23. Lee C-W, Yoon K-J, Kim S-W (2019) percutaneous endoscopic decompression in lumbar canal and lateral recess stenosis—the surgical learning curve. Neurospine 16:63–71. https://doi.org/10.14245/ns.1938048.024

24. Cheung PWH, Wong CKH, Lau ST, Cheung JPY (2020) Cost analysis comparison between conventional microsurgical decompression and full-endoscopic interlaminar decompression for lumbar spinal stenosis surgery. J Spine Surg 6:721–728. https://doi.org/10.21037/jss-20-552

25. Komp M, Hahn P, Oezdemir S et al (2015) Bilateral spinal decompression of lumbar central stenosis with the full-endoscopic interlaminar versus microsurgical laminotomy technique: a prospective, randomized, controlled study. Pain Physician 18:61–70

26. Ruetten S, Komp M, Merk H, Godolias G (2009) Surgical treatment for lumbar lateral recess stenosis with the full-endoscopic interlaminar approach versus conventional microsurgical technique: a prospective, randomized, controlled study. J Neurosurg Spine 10:476–485. https://doi.org/10.3171/2008.7.17634

27. Marković M, Živković N, Aleksić V (2015) Full-endoscopic interlaminar versus microsurgical operations in lumbar compressive lesions surgery. J Neurol Surg A Cent Eur Neurosurg 76:s-0035-1566385. https://doi.org/10.1055/s-0035-1566385

28. Lin G-X, Kotheeranurak V, Mahatthanatrakul A et al (2020) Worldwide research productivity in the field of full-endoscopic spine surgery: a bibliometric study. Eur Spine J 29:153–160. https://doi.org/10.1007/s00586-019-06171-2

29. Caputy AJ, Luessenhop AJ (1992) Long-term evaluation of decompressive surgery for degenerative lumbar stenosis. J Neurosurg 77:669–676. https://doi.org/10.3171/jns.1992.77.5.0669

30. Fox MW, Onofrio BM, Onofrio BM, Hanssen AD (1996) Clinical outcomes and radiological instability following decompressive lumbar laminectomy for degenerative spinal stenosis: a comparison of patients undergoing concomitant arthrodesis versus decompression alone. J Neurosurg 85:793–802. https://doi.org/10.3171/jns.1996.85.5.0793

31. Getty CJ, Johnson JR, Kirwan EO, Sullivan MF (1981) Partial undercutting facetectomy for bony entrapment of the lumbar nerve root. J Bone Joint Surg Br 63-B:330–335. https://doi.org/10.1302/0301-620X.63B3.7263743

32. Mayer HM, List J, Korge A, Wiechert K (2003) Microsurgery of acquired degenerative lumbar spinal stenosis. Bilateral over-the-top decompression through unilateral approach. Orthopade 32:889–895. https://doi.org/10.1007/s00132-003-0536-9

33. Ragab AA, Fye MA, Bohlman HH (2003) Surgery of the lumbar spine for spinal stenosis in 118 patients 70 years of age or older. Spine 28:348–353. https://doi.org/10.1097/01.BRS.0000048494.66599.DF

34. Sanderson PL, Getty CJ (1996) Long-term results of partial undercutting facetectomy for lumbar lateral recess stenosis. Spine 21:1352–1356. https://doi.org/10.1097/00007632-199606010-00015

35. Johnsson KE, Redlund-Johnell I, Udén A, Willner S (1989) Preoperative and postoperative instability in lumbar spinal stenosis. Spine 14:591–593. https://doi.org/10.1097/00007632-198906000-00008

36. LaRocca H, Macnab I (1974) The laminectomy membrane. Studies in its evolution, characteristics, effects and prophylaxis in dogs. J Bone Joint Surg Br 56B:545–550

37. Frank EH, Hsu FPK (2002) An endoscopic dural retractor for spinal stenosis surgery. Minim Invasive Neurosurg 45:136–138. https://doi.org/10.1055/s-2002-34340

38. Guiot BH, Khoo LT, Fessler RG (2002) A minimally invasive technique for decompression of the lumbar spine. Spine 27:432–438. https://doi.org/10.1097/00007632-200202150-00021

39. Khoo LT, Fessler RG (2002) Microendoscopic decompressive laminotomy for the treatment of lumbar stenosis. Neurosurgery 51:S146-154

40. Young S, Veerapen R, O'Laoire SA (1988) Relief of lumbar canal stenosis using multilevel subarticular fenestrations as an alternative to wide laminectomy: preliminary report. Neurosurgery 23:628–633. https://doi.org/10.1227/00006123-198811000-00014

41. Katz JN, Lipson SJ, Chang LC et al (1996) Seven- to 10-year outcome of decompressive surgery for degenerative lumbar spinal stenosis. Spine 21:92–98. https://doi.org/10.1097/00007632-199601010-00022

42. Lee C-W, Yoon K-J, Ha S-S (2019) Comparative analysis between three different lumbar decompression techniques (microscopic, tubular, and endoscopic) in lumbar canal and lateral recess stenosis: preliminary report. Biomed Res Int 2019:1–11. https://doi.org/10.1155/2019/6078469

43. Yang F, Chen R, Gu D et al (2020) Clinical comparison of full-endoscopic and microscopic unilateral laminotomy for bilateral decompression in the treatment of elderly lumbar spinal stenosis: a retrospective study with 12-month follow-up. J Pain Res 13:1377–1384. https://doi.org/10.2147/JPR.S254275

44. Chen C-M, Sun L-W, Tseng C et al (2020) Surgical outcomes of full endoscopic spinal surgery for lumbar disc herniation over a 10-year period: a retrospective study. PLoS ONE 15:e0241494. https://doi.org/10.1371/journal.pone.0241494

45. Colley CM, Fleck A, Goode AW et al (1983) Early time course of the acute phase protein response in man. J Clin Pathol 36:203–207. https://doi.org/10.1136/jcp.36.2.203

46. White J, Kelly M, Dunsmuir R (1998) C-reactive protein level after total hip and total knee replacement. J Bone Joint Surg Br 80:909–911. https://doi.org/10.1302/0301-620x.80b5.8708

47. Bourguignat A, Férard G, Jenny JY, Gaudias J (1997) Incomplete or absent acute phase response in some postoperative patients. Clin Chim Acta 264:27–35. https://doi.org/10.1016/s0009-8981(97)00071-5

48. Brewster N, Guthrie C, McBirnie J (1994) CRP levels as a measure of surgical trauma: a comparison of different general surgical procedures. J R Coll Surg Edinb 39:86–88

49. Cole DS, Watts A, Scott-Coombes D, Avades T (2008) Clinical Utility of peri-operative C-reactive protein testing in general surgery. Ann R Coll Surg Engl 90:317–321. https://doi.org/10.1308/003588408X285865