Published in final edited form as:

Spine (Phila Pa 1976). 2013 October 1; 38(21): . doi:10.1097/BRS.0b013e3182a314d0.

Who should have surgery for degenerative spondylolisthesis?:

Treatment Effect Predictors in SPORT

Adam M. Pearson, MD, MS^{1,2}, Jon D. Lurie, MD, MS^{1,2,3}, Tor D. Tosteson, ScD^{1,3}, Wenyan Zhao, MS^{1,3}, William A. Abdu, MD, MS^{1,2,3}, and James N. Weinstein, DO, MS^{1,2,3}

¹Geisel School of Medicine at Dartmouth, Hanover NH 03755

²Dartmouth-Hitchcock Medical Center, Lebanon, NH 03756

³The Dartmouth Institute for Health Policy and Clinical Practice, Lebanon, NH 03756

Abstract

Study Design—Combined prospective randomized controlled trial and observational cohort study of degenerative spondylolisthesis (DS) with an as-treated analysis.

Objective—To determine modifiers of the treatment effect (TE) of surgery (the difference between surgical and nonoperative outcomes) for DS using subgroup analysis.

Summary of Background Data—SPORT demonstrated a positive surgical TE for DS at the group level. However, individual characteristics may affect TE.

Methods—DS patients were treated with either surgery (n=395) or nonoperative care (n=206) and were analyzed according to treatment received. Fifty-five baseline variables were used to define subgroups for calculating the time-weighted average TE for the Oswestry Disability Index (ODI) over 4 years ($TE = ODI_{surgery} - ODI_{nonoperative}$). Variables with significant subgroup-by-treatment interactions ($p < 0.05$) were simultaneously entered into a multivariate model to select independent TE predictors.

Results—All analyzed subgroups that included at least 50 patients improved significantly more with surgery than with nonoperative treatment ($p < 0.05$). Multivariate analyses demonstrated that age ≥ 67 (TE -15.7 vs. -11.8 for age < 67 , $p = 0.014$); female gender (TE -15.6 vs. -11.2 for males, $p = 0.01$); the absence of stomach problems (TE -15.2 vs. -11.3 for those with stomach problems, $p = 0.035$); neurogenic claudication (TE -15.3 vs. -9.0 for those without claudication, $p = 0.004$); reflex asymmetry (TE -17.3 vs. -13.0 for those without asymmetry, $p = 0.016$); opioid use (TE -18.4 vs. -11.7 for those not using opioids, $p < 0.001$); not taking antidepressants (TE -14.5 vs. -5.4 for those on antidepressants, $p = 0.014$); dissatisfaction with symptoms (TE -14.5 vs. -8.3 for those satisfied or neutral, $p = 0.039$); and anticipating a high likelihood of improvement with surgery (TE -14.8 vs. -5.1 for anticipating a low likelihood of improvement with surgery, $p = 0.019$) were independently associated with greater TE.

Conclusions—Patients who met strict inclusion criteria improved more with surgery than with nonoperative treatment, regardless of other specific characteristics. However, TE varied significantly across certain subgroups.

Corresponding Author: Adam M. Pearson, MD, MS, One Medical Center Dr., Dartmouth-Hitchcock Medical Center, Lebanon, NH 03756, 603-653-3559, FAX: 603-653-3554, tamara.s.morgan@dartmouth.edu; adam.m.pearson@hitchcock.org.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

degenerative spondylolisthesis; surgery; treatment effect; prediction

INTRODUCTION

The Spine Patient Outcomes Research Trial (SPORT) demonstrated better outcomes for degenerative spondylolisthesis patients (DS) patients treated surgically than those treated nonoperatively out to four years.^{1,2} Other prospective studies of spinal stenosis (SpS) that did not stratify patients based on the presence of listhesis found similar results.^{3–5} However, these findings were all at the group level, and it has been established that individual patient characteristics are strongly related to surgical and nonoperative outcomes.^{6–13} In the SPORT DS cohort, duration of symptoms, diabetes, the number of stenotic levels, predominant pain location (leg vs. back), and the presence of radiographic “instability” have been specifically evaluated as outcome predictors.^{14–18} Both SPORT and prior studies have empirically selected variables for evaluation as potential outcome predictors based on clinical judgment, yet demographic and psychosocial characteristics may be more strongly associated with outcomes than clinical or disease characteristics.^{19–21} As such, the current study aims to evaluate a large pool of demographic, psychosocial, and clinical variables to determine which are predictors of surgical and nonoperative outcomes in DS.

The vast majority of literature focusing on predicting outcomes in patients with degenerative lumbar conditions has focused on exclusively surgical outcomes, usually due to the lack of a nonoperative cohort.^{6,10,13} While it is useful to counsel patients about their likely surgical results, the factor that should drive decision making is their likely benefit from surgery *relative* to nonoperative outcomes, taking into account their specific demographic, psychosocial, and clinical characteristics. This is the predicted treatment effect (TE) of surgery. If only surgical outcomes are considered, surgeons could erroneously suggest surgery to patients who may do well with nonoperative treatment and nonoperative treatment to patients who are likely to fail without surgery. Prior analyses of the SPORT SpS and disk herniation cohorts surprisingly revealed that some characteristics that predicted worse surgical outcomes (i.e. obesity, low educational attainment) were actually associated with a greater TE due to the very poor nonoperative outcomes in comparable subgroups.^{22,23} Thus, taking only the surgical outcome predictors into account could lead to an inappropriate treatment decision for an individual patient. As such, the goal of this study was to determine which variables were significant, independent TE modifiers in DS.

MATERIALS AND METHODS

Study Design

The SPORT DS investigation consisted of a randomized controlled trial with a concurrent observational cohort study conducted in 11 states at 13 institutions with multidisciplinary spine practices.²⁴ In the first year of follow-up in the randomized trial, 43% of patients assigned to surgery did not have surgery, and 44% of patients assigned to nonoperative treatment did have surgery.²⁵ By 4 years, 34% of the patients randomized to surgery had still not undergone surgery, and 54% of the patients assigned to nonoperative treatment had received surgery.²⁶ Given this rate of protocol nonadherence and the consistency of the findings between the randomized and the observational cohorts, the data were combined in an as-treated analysis in this study.^{1,26} The rationale for this decision has been discussed previously.²⁷

Patient Population

Patients were considered for inclusion in the study if they had neurogenic claudication or radicular pain for at least twelve weeks, degenerative spondylolisthesis on a standing lateral radiograph, and confirmatory cross-sectional imaging study demonstrating stenosis at one or more levels. Exclusion criteria included spondylolysis or isthmic spondylolisthesis, cauda equina syndrome, malignancy, significant deformity, prior back surgery, and other established contra-indications to elective surgery.²⁴

Study Interventions

Surgery consisted of a standard open decompressive laminectomy with or without a bilateral single-level fusion using iliac crest autograft with or without instrumentation.²⁴ The nonoperative treatment group received “usual care”—defined as including at least physical therapy, education and counseling with home exercise instruction, and non-steroidal anti-inflammatory drugs if tolerated. Physicians were instructed to individualize nonoperative treatment and explore a wide range of nonoperative options.²⁴

Imaging Studies

All patients underwent standing x-rays and cross-sectional imaging. The cross-sectional imaging was evaluated to determine which levels were stenotic and the location (central, lateral recess, and/or neuroforamen) and severity of the stenosis (mild, moderate or severe).^{28,29} The kappa scores for intra-rater reliability of the location and severity classification have been reported to range from 0.75 to 0.82, while inter-rater reliability ranged from 0.49 to 0.73.²⁹

Baseline Characteristics and Primary Outcome Measure

At baseline, patients and clinicians answered questionnaires evaluating demographic, socioeconomic, clinical, and radiographic characteristics.^{24,25} Baseline scores on the Short-Form 36 (SF-36),³⁰ AAOS/Modems version of the Oswestry Disability Index (ODI),³¹ Stenosis Bothersomeness Index,^{32,33} and Leg and Back Pain Bothersomeness Scale (0–6 point Likert-type scale) were also recorded. The ODI was also recorded at 6 weeks, 3 months, 6 months, 1 year and yearly thereafter out to 4 years. Fifty-five variables were selected as potential TE modifiers. Continuous variables were converted to categorical variables as follows: ODI defined by quartiles; Stenosis Bothersomeness and Age by the median; Duration of Symptoms as less than or greater than 6 months; Income as greater than \$50,000, less than \$50,000 or not working; Body Mass Index (BMI) as less than 30 or greater than or equal to 30; and the SF-36 Mental Component Score (MCS) as less than or equal to 35 or greater than 35.³⁴ Stenosis level was classified as either involving a given level or not. Stenosis location was classified as involving or not involving the central canal, lateral recess or neuroforamen.²⁸ Predominant pain location was classified as predominant leg pain (baseline leg pain bothersomeness score greater than or equal to back pain pain bothersomeness score) or predominant back pain (back pain bothersomeness score greater than leg pain bothersomeness score).

Statistical Considerations

The primary aim of the analysis was to identify baseline variables that were significant indicators of differential treatment effects of surgery (i.e. variables with significant subgroup by treatment interactions). The treatment effect (TE) of surgery was defined as:

$$\text{Treatment Effect} = \text{Change in ODI}_{\text{surgery}} - \text{Change in ODI}_{\text{nonoperative}}$$

In all analyses, TE was calculated as the time-weighted average (area under the curve) over 4 years. Note that a negative TE indicated that surgery was more effective than nonoperative

treatment. To identify candidate variables, a “minimally-adjusted” analysis (controlling only for age, gender, center, and baseline ODI score) was performed for each of the fifty-five variables to identify those that had significant TE differences across subgroups. Those that were identified as potential TE modifiers ($p < 0.05$) were then entered along with their treatment interaction terms into a longitudinal regression model that controlled for factors predicting treatment received or missing data in order to control for selection and attrition bias (age, gender, work status, BMI, neuroforaminal involvement, depression, osteoporosis, joint problems, duration of symptoms, reflex deficit, number of moderately or severely stenotic levels, hypertension, treatment preference, other comorbidities, baseline SF-36 score, baseline ODI score, baseline Stenosis Bothersomeness score, and center).²⁶ Variables with significant treatment interaction terms ($p < 0.05$) were identified as independent TE modifiers and included in the final multivariate mixed effects longitudinal regression model including a random individual effect to account for correlation between repeated measurements within individuals. The outcomes were stratified by modifiers and overall comparisons of area-under-the-curve across the four-year follow-up were made using a Wald test.³⁵ This analytical approach, starting with “minimally adjusted” analyses, was performed in order to ensure that no potential independent modifiers were missed in the first stage. In the second stage, the complete multivariate model was used in order to definitively identify independent TE modifiers. Given the large number of comparisons being made, the possibility for a Type I error (false positive) exists. We elected not to perform a conservative Bonferroni correction in order to avoid increasing the Type II error (false negative) rate. As such, the results should be viewed as hypothesis-generating rather than definitive conclusions.

RESULTS

Patient Distribution and Follow-Up

There were 607 patients enrolled overall, with 304 in the randomized trial and 303 in the observational cohort. By 4 years, 395 (65%) had undergone surgery. Completeness of follow-up ranged from 70% to 92% at the different follow-up intervals. Details can be found in the primary analysis.²⁶

Minimally Adjusted Analyses

All examined subgroups including at least fifty patients had a significant treatment effect of surgery (Table 1). Patients taking antidepressants ($n=31$, TE = -5.7, 95% CI -12.8–1.4) and those who anticipated a small chance of being pain free with surgery ($n=46$, TE = -5.8, 95% CI -12.3–0.7) did not have significant TEs, but this was in part due to the small number of patients in these groups and the large confidence intervals around their estimated TEs. The majority of variables were significantly associated with surgical and/or nonoperative outcomes. However, the direction of association was usually the same for both treatments, so most variables were not significantly associated with TE (e.g. patients with predominant back pain improved less than those with predominant leg pain with both surgical and nonoperative treatment but had similar TE). Twelve variables predicted a greater TE: age 67 (TE -16.0 vs. -12.5 for age > 67, $p = 0.028$), female gender (TE -16.1 vs. -11.6 for males, $p = 0.009$), duration of symptoms greater than 6 months (TE -16.5 vs. -12.1 for those with symptoms less than 6 months, $p = 0.006$), not reporting a stomach problem (TE -15.6 vs. -11.5 for those with stomach problems, $p = 0.029$), not reporting a joint problem (TE -16.4 vs. -13.1 for those with joint problems, $p = 0.041$), having neurogenic claudication (TE -15.9 vs. -8.3 for those without claudication, $p < 0.001$), reflex asymmetry (TE -17.6 vs. -13.5 for no reflex asymmetry, $p = 0.021$), taking opioids (TE -19.0 vs. -12.3 for those not taking opioids, $p < 0.001$), not taking antidepressants (TE -15.0 vs. -5.7 for those taking antidepressants, $p = 0.013$), dissatisfaction with symptoms (TE -15.1 vs. -8.7

for those neutral or satisfied with symptoms, $p = 0.035$), worsening symptoms (TE -23.4 vs. -16.0 for those getting better vs. -11.1 for those staying the same, $p < 0.001$), and anticipating a high likelihood of being pain free with surgery (TE -15.0 vs. -5.8 for those anticipating a low likelihood vs. -14.6 for those anticipating a moderate likelihood, $p = 0.029$).

Multivariate Analyses

The twelve significant TE predictors identified in the minimally adjusted analyses and their interaction terms were added to a previously published model evaluating the time-weighted average TE over four years of follow-up.²⁶ Age ≥ 67 , female gender, not reporting stomach problems, having neurogenic claudication, reflex asymmetry, opioid use, not taking antidepressants, dissatisfaction with symptoms, and anticipating a high likelihood of being pain free with surgery remained significant independent predictors of greater TE ($p < 0.05$) (Table 2).

To illustrate the range of TEs across subgroups defined by multiple variables, we compared outcomes between patients with some characteristics associated with greater TE (neurogenic claudication, taking opioids, not taking antidepressants, and anticipating a moderate or high likelihood of being pain free with surgery, $n=134$) and patients with some characteristics predicting a lower TE (no neurogenic claudication, not taking opioids, $n=49$; Table 3). Patients with the characteristics associated with greater TE had a predicted TE of -22.6 compared to -3.5 for the patients without neurogenic claudication who were not taking opioids ($p < 0.001$). We were unable to compare outcomes between the groups with the combinations of variables predicting the most extreme TE differences since no patients had these constellations of characteristics.

DISCUSSION

In the primary analysis, SPORT demonstrated a significant treatment effect of surgery for the “average” degenerative spondylolisthesis patient.^{1,2} However, clinical decision making takes place at the individual patient level, not the group level. As such, we sought to evaluate a wide range of patient and disease characteristics to determine which were independent TE predictors. Unlike most prior studies of degenerative lumbar conditions that have included only surgical patients, SPORT had the advantage of including both a surgical and nonoperative arm. This allowed for the calculation of TE--the difference between surgical and nonoperative outcomes between patients with similar baseline characteristics—rather than reporting only the absolute change from baseline associated with surgery. We believe that TE is the more important parameter for guiding treatment decisions as it defines the likely differences in outcomes between the patient’s treatment options. Another unique aspect of SPORT is that DS patients and SpS patients without listhesis were studied separately, and we have demonstrated that these two patient populations are markedly different in their baseline characteristics, how they were treated, and in their surgical outcomes.³⁶ Somewhat surprisingly, the current analysis found that essentially all of the DS subgroups defined by the fifty-five variables evaluated in this study had a significant benefit from surgery compared to nonoperative treatment. The lack of a significant TE in patients taking antidepressants and in those anticipating a low chance of being pain free with surgery was likely due in part to the small size of these groups and an underpowered comparison between surgical and nonoperative outcomes.

Many prior studies have documented the strong effect of patient and disease characteristics on surgical outcomes in SpS and DS.^{6,9,11,12,20,21,37} The current study demonstrated that the majority of variables we evaluated were associated with surgical and/or nonoperative outcomes. Some of the strongest predictors of worse surgical and nonoperative outcomes

were increasing comorbidity burden, being involved in litigation, and receiving worker's compensation. Slover et al. convincingly demonstrated an inverse relationship between increasing comorbidity burden and surgical outcomes, and this study confirms that finding.³⁷ While there is little data reporting outcomes in SpS or DS worker's compensation patients, Atlas et al. demonstrated worse surgical and nonoperative outcomes in lumbar disk herniation patients receiving worker's compensation.³⁸ Not surprisingly, the current study found similar results and adds to the extensive literature demonstrating worse outcomes in this population across the entire orthopaedic literature. However, the current analysis differs in that TE was evaluated for these patients who are known to have worse outcomes, and TE was not significantly associated with comorbidity burden, worker's compensation or litigation. In other words, these patients tended to have worse outcomes with either surgery or nonoperative treatment compared to patients without these characteristics, but the relative difference between surgical and nonoperative outcomes was similar across the different subgroups. As such, surgical treatment should be considered as an option for patients with characteristics associated with worse surgical outcomes (i.e. high comorbidity burden, worker's compensation, litigation, etc.) as they are likely to have minimal improvement with nonoperative treatment.

Two of the strongest TE predictors in the DS population were patient expectations and baseline satisfaction with symptoms. The greatest TE differences were observed between patients who anticipated a high and low likelihood of success with surgical treatment. The forty-six patients who anticipated a low likelihood of success with surgery had a TE of only -5.1, with the difference in surgical and nonoperative outcomes failing to reach significance due to the relatively poor surgical outcomes in this group. Not surprisingly, patient expectations can be a self-fulfilling prophecy. A similar finding was reported by Cobo Soraiano et al. who found better outcomes in patients optimistic about their surgical outcomes.²⁰ Patients who were dissatisfied with their symptoms at baseline had markedly worse nonoperative outcomes, and this resulted in a significantly larger TE in this group. Again, it is not surprising that patients who were dissatisfied with their progress with nonoperative treatment failed to improve with further nonoperative treatment.

An unanticipated finding of the current study was that patients taking antidepressants had a significantly lower TE due to very good non-operative outcomes in this group. The role of antidepressants in SpS and DS has not been well-studied, though there is a case series reporting improvement with tricyclic antidepressants in this population.³⁹ On the other hand, meta-analysis has suggested that antidepressants do not improve outcomes in a chronic low back pain population.⁴⁰ With only thirty one patients taking antidepressants, it is hard to know how to interpret this finding, and future work should be done to evaluate the role of these medications in SpS and DS. Also somewhat contrary to popular belief was the finding that patients taking opioids had a significantly greater TE. These patients had similar surgical outcomes to those not taking opioids, however, they had poor nonoperative outcomes. Opioid use in this population was likely a marker for failed nonoperative treatment yet was not associated with worse surgical outcomes.

Compared to patient perceptions of their disease process and medication use, neurological symptoms and demographic characteristics were weaker TE predictors. Patients with neurogenic claudication had slightly better surgical outcomes and slightly worse nonoperative outcomes than those with isolated radicular symptoms, resulting in a higher TE for those with claudication. This result is consistent with a prior study reporting a greater ODI improvement in surgical patients with lower walking capacity at baseline.²⁰ Younger patients and women did slightly better with surgery and somewhat worse with nonoperative treatment, resulting in a higher TE in these groups, though the differences were small.

A similar analysis was conducted for the SPORT SpS cohort, and there was no overlap between the variables predicting TE in the SpS and DS populations.²³ While this seems counterintuitive, comparison of the SPORT SpS and DS cohorts showed marked differences between the groups.³⁶ The most striking difference was that smokers in the SpS cohort were the only subgroup that did not improve more with surgery compared to nonoperative treatment, yet smoking was not associated with outcomes or TE in DS. Conventional wisdom would suggest that smoking would have a greater negative effect on patients undergoing fusion, which was much more common in DS than in SpS.³⁶ While we do not have a definitive explanation for the differential effect of smoking in the two populations, SPORT did not capture data on smoking cessation prior to surgery. It is possible that DS patients were more likely to be counseled to stop smoking due to the known negative effect on fusion rates, while surgeons may have been less likely to be concerned about smoking in patients undergoing decompression alone. Patient expectations, satisfaction with symptoms, and opioid use had similar but weaker, non-significant effects in the SpS analysis.

There are a number of limitations of this study. While SPORT was designed with a randomized cohort, the high rate of protocol non-adherence precluded meaningful interpretation of those data on an intent-to-treat basis alone.²⁵ Additionally, the subgroup analyses used to evaluate potential TE modifiers were possibly underpowered for some variables (as indicated by broad confidence intervals around some TE estimates), and this may have resulted in the failure to detect some meaningful modifiers (Type II error). On the other hand, fifty-five potential TE modifiers were evaluated, which put the study at risk for finding spurious associations due to chance alone (Type I error). Additionally, the subgroup analyses used to evaluate potential TE modifiers were possibly underpowered for some variables (as indicated by broad confidence intervals around some TE estimates), and this may have resulted in the failure to detect some meaningful modifiers (Type II error). On the other hand, fifty-five potential TE modifiers were evaluated, which put the study at risk for finding spurious associations due to chance alone (Type I error).

The current study demonstrates the importance of considering specific patient characteristics when making individual treatment decisions in DS. While single variables tended to predict relatively modest differences in TE, certain combinations of variables—as could exist in real patients—resulted in marked TE differences (Table 4). Large clinical trials provide important information that can be used to formulate treatment guidelines and inform policy decisions, though their findings are less helpful when advising an individual patient about their treatment options and likely outcomes. The next step in applying evidence-based medicine at the individual patient level will be the creation of prediction models that can inform patients about their likely outcomes given their specific baseline characteristics.

Acknowledgments

The manuscript submitted does not contain information about medical device(s)/drug(s). NIH/NIAMS grant funds were received in support of this work. Relevant financial activities outside the submitted work: consultancy, grants.

REFERENCES

1. Weinstein JN, Lurie JD, Tosteson TD, et al. Surgical versus nonsurgical treatment for lumbar degenerative spondylolisthesis. *The New England journal of medicine*. 2007; 356:2257–2270. [PubMed: 17538085]
2. Weinstein JN, Tosteson TD, Lurie JD, et al. Surgical versus nonoperative treatment for lumbar spinal stenosis four-year results of the Spine Patient Outcomes Research Trial. *Spine (Phila Pa 1976)*. 2010; 35:1329–1338. [PubMed: 20453723]

3. Atlas SJ, Deyo RA, Keller RB, et al. The Maine Lumbar Spine Study, Part III 1-year outcomes of surgical and nonsurgical management of lumbar spinal stenosis. *Spine*. 1996; 21:1787–1794. discussion 94–5. [PubMed: 8855463]
4. Atlas SJ, Keller RB, Robson D, Deyo RA, Singer DE. Surgical and nonsurgical management of lumbar spinal stenosis: four-year outcomes from the maine lumbar spine study. *Spine*. 2000; 25:556–562. [PubMed: 10749631]
5. Malmivaara A, Slati P, Heliovaara M, et al. Surgical or nonoperative treatment for lumbar spinal stenosis? A randomized controlled trial. *Spine*. 2007; 32:1–8. [PubMed: 17202885]
6. Airaksinen O, Herno A, Turunen V, Saari T, Suomlainen O. Surgical outcome of 438 patients treated surgically for lumbar spinal stenosis. *Spine (Phila Pa 1976)*. 1997; 22:2278–2282. [PubMed: 9346149]
7. Hurri H, Slati P, Soini J, et al. Lumbar spinal stenosis: assessment of long-term outcome 12 years after operative and conservative treatment. *Journal of spinal disorders*. 1998; 11:110–115. [PubMed: 9588466]
8. Iguchi T, Kurihara A, Nakayama J, Sato K, Kurosaka M, Yamasaki K. Minimum 10-year outcome of decompressive laminectomy for degenerative lumbar spinal stenosis. *Spine*. 2000; 25:1754–1759. [PubMed: 10888941]
9. Jonsson B, Annertz M, Sjöberg C, Stromqvist B. A prospective consecutive study of surgically treated lumbar spinal stenosis Part II: Five-year follow-up by an independent observer. *Spine*. 1997; 22:2938–2944. [PubMed: 9431630]
10. Katz JN, Stucki G, Lipson SJ, Fossel AH, Grobler LJ, Weinstein JN. Predictors of surgical outcome in degenerative lumbar spinal stenosis. *Spine*. 1999; 24:2229–2233. [PubMed: 10562989]
11. Kleinstuck FS, Grob D, Lattig F, et al. The influence of preoperative back pain on the outcome of lumbar decompression surgery. *Spine (Phila Pa 1976)*. 2009; 34:1198–1203. [PubMed: 19407677]
12. Mariconda M, Zanforlino G, Celestino GA, Brancalone S, Fava R, Milano C. Factors influencing the outcome of degenerative lumbar spinal stenosis. *Journal of spinal disorders*. 2000; 13:131–137. [PubMed: 10780688]
13. Spratt KF, Keller TS, Szpalski M, Vandeputte K, Gunzburg R. A predictive model for outcome after conservative decompression surgery for lumbar spinal stenosis. *Eur Spine J*. 2004; 13:14–21. [PubMed: 14658061]
14. Freedman MK, Hilibrand AS, Blood EA, et al. The impact of diabetes on the outcomes of surgical and nonsurgical treatment of patients in the spine patient outcomes research trial. *Spine (Phila Pa 1976)*. 2011; 36:290–307. [PubMed: 21270715]
15. Park DK, An HS, Lurie JD, et al. Does multilevel lumbar stenosis lead to poorer outcomes?: a subanalysis of the Spine Patient Outcomes Research Trial (SPORT) lumbar stenosis study. *Spine (Phila Pa 1976)*. 2010; 35:439–446. [PubMed: 20081560]
16. Pearson A, Blood E, Lurie J, et al. Predominant leg pain is associated with better surgical outcomes in degenerative spondylolisthesis and spinal stenosis: results from the Spine Patient Outcomes Research Trial (SPORT). *Spine (Phila Pa 1976)*. 2011; 36:219–229. [PubMed: 21124260]
17. Pearson AM, Lurie JD, Blood EA, et al. Spine patient outcomes research trial: radiographic predictors of clinical outcomes after operative or nonoperative treatment of degenerative spondylolisthesis. *Spine (Phila Pa 1976)*. 2008; 33:2759–2766. [PubMed: 19050582]
18. Radcliff K, Rihn J, Hilibrand A, et al. Does the Duration of Symptoms in Patients with Spinal Stenosis and Degenerative Spondylolisthesis Affect Outcomes? Analysis of the Spine Outcomes Research Trial. *Spine*. 2011
19. Atlas SJ, Tosteson TD, Blood EA, Skinner JS, Pransky GS, Weinstein JN. The impact of workers' compensation on outcomes of surgical and nonoperative therapy for patients with a lumbar disc herniation: SPORT. *Spine (Phila Pa 1976)*. 2010; 35:89–97. [PubMed: 20023603]
20. Cobo Soriano J, Sendino Revuelta M, Fabregate Fuente M, Cimarra Diaz I, Martinez Urena P, Deglane Meneses R. Predictors of outcome after decompressive lumbar surgery and instrumented posterolateral fusion. *Eur Spine J*. 2010; 19:1841–1848. [PubMed: 20135333]
21. Sinikallio S, Aalto T, Airaksinen O, Herno A, Kroger H, Viinamäki H. Depressive burden in the preoperative and early recovery phase predicts poorer surgery outcome among lumbar spinal

- stenosis patients: a one-year prospective follow-up study. *Spine (Phila Pa 1976)*. 2009; 34:2573–2578. [PubMed: 19927107]
22. Pearson A, Lurie J, Tosteson T, et al. Who should have surgery for an intervertebral disc herniation? Comparative effectiveness evidence from the spine patient outcomes research trial. *Spine*. 2012; 37:140–149. [PubMed: 21681140]
 23. Pearson A, Lurie J, Tosteson T, Zhao W, Abdu W, Weinstein JN. Who Should Have Surgery for Spinal Stenosis? Treatment Effect Predictors in SPORT. *Spine*. 2012; 37:1791–1802. [PubMed: 23018805]
 24. Birkmeyer NJ, Weinstein JN, Tosteson AN, et al. Design of the Spine Patient outcomes Research Trial (SPORT). *Spine*. 2002; 27:1361–1372. [PubMed: 12065987]
 25. Weinstein JN, Tosteson TD, Lurie JD, et al. Surgical versus nonsurgical therapy for lumbar spinal stenosis. *The New England journal of medicine*. 2008; 358:794–810. [PubMed: 18287602]
 26. Weinstein JN, Lurie JD, Tosteson TD, et al. Surgical compared with nonoperative treatment for lumbar degenerative spondylolisthesis: four-year results in the Spine Patient Outcomes Research Trial (SPORT) randomized and observational cohorts. *The Journal of bone and joint surgery American volume*. 2009; 91:1295–1304. [PubMed: 19487505]
 27. Tosteson, TD.; Hanscom, B.; Blood, EA.; Lurie, JD.; Tosteson, ANA.; Weinstein, JN. Statistical methods for cross-over in the SPORT lumbar disc herniation trial. *International Society for the Study of the Lumbar Spine Annual Meeting*; Hong Kong. 2007.
 28. Fardon DF, Milette PC. Nomenclature classification of lumbar disc pathology. Recommendations of the Combined task Forces of the North American Spine Society, American Society of Spine Radiology, and American Society of Neuroradiology. *Spine*. 2001; 26:E93–E113. [PubMed: 11242399]
 29. Lurie JD, Tosteson AN, Tosteson TD, et al. Reliability of readings of magnetic resonance imaging features of lumbar spinal stenosis. *Spine*. 2008; 33:1605–1610. [PubMed: 18552677]
 30. Ware JE Jr, Sherbourne CD. The MOS36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical care*. 1992; 30:473–483. [PubMed: 1593914]
 31. Daltroy LH, Cats-Baril WL, Katz JN, Fossel AH, Liang MH. The North American spine society lumbar spine outcome assessment Instrument: reliability and validity tests. *Spine*. 1996; 21:741–749. [PubMed: 8882698]
 32. Atlas SJ, Deyo RA, Patrick DL, Convery K, Keller RB, Singer DE. The Quebec Task Force classification for Spinal Disorders and the severity, treatment, and outcomes of sciatica and lumbar spinal stenosis. *Spine*. 1996; 21:2885–2892. [PubMed: 9112713]
 33. Patrick DL, Deyo RA, Atlas SJ, Singer DE, Chapin A, Keller RB. Assessing health-related quality of life in patients with sciatica. *Spine*. 1995; 20:1899–1908. discussion 909. [PubMed: 8560339]
 34. Walsh TL, Homa K, Hanscom B, Lurie J, Sepulveda MG, Abdu W. Screening for depressive symptoms in patients with chronic spinal pain using the SF-36 Health Survey. *Spine J*. 2006; 6:316–320. [PubMed: 16651227]
 35. Fitzmaurice, G.; Laird, N.; Ware, J. *Applied Longitudinal Analysis*. Philadelphia, PA: John Wiley & Sons; 2004.
 36. Pearson A, Blood E, Lurie J, et al. Degenerative spondylolisthesis versus spinal stenosis: does a slip matter? Comparison of baseline characteristics and outcomes (SPORT). *Spine (Phila Pa 1976)*. 2010; 35:298–305. [PubMed: 20075768]
 37. Slover J, Abdu WA, Hanscom B, Weinstein JN. The impact of comorbidities on the change in short-form 36 and oswestry scores following lumbar spine surgery. *Spine*. 2006; 31:1974–1980. [PubMed: 16924216]
 38. Atlas SJ, Chang Y, Keller RB, Singer DE, Wu YA, Deyo RA. The impact of disability compensation on long-term treatment outcomes of patients with sciatica due to a lumbar disc herniation. *Spine*. 2006; 31:3061–3069. [PubMed: 17173004]
 39. Orbai AM, Meyerhoff JO. The effectiveness of tricyclic antidepressants on lumbar spinal stenosis. *Bull NYU Hosp Jt Dis*. 2010; 68:22–24. [PubMed: 20345358]
 40. Kuijpers T, van Middelkoop M, Rubinstein SM, et al. A systematic review on the effectiveness of pharmacological interventions for chronic non-specific low-back pain. *European spine journal* :

official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society. 2011; 20:40–50.

KEY POINTS

- All patient subgroups improved more with surgery than with nonoperative treatment.
- Age less than 67, female gender, the absence of stomach problems, neurogenic claudication, reflex asymmetry, opioid use, not taking antidepressants, dissatisfaction with symptoms, and anticipating a high likelihood of being pain free with surgery predicted a greater treatment effect of surgery.
- These data can be used to help individualize shared decision making discussions about likely outcomes following surgical or nonoperative treatment for DS.

Table 1

Area under curve subgroup results for Oswestry Disability Index (ODI) change scores for variables that are statistically significantly associated with treatment effect, adjusted for age, gender, center, and baseline ODI score (*Time weighted average 4 years treatment effects*).

Characteristics	N	Baseline ODI Mean (SE)	Mean Change from Baseline		Treatment Effect (95% CI) *
			Surgical (n = 391)	Non-operative (n = 210)	
Age 67	312	45.3 (1.1)	-23.4 (0.9)	-7.4 (1.1)	-16 (-18.2, -13.8)
Age > 67	289	39.5 (1)	-21 (1)	-8.5 (1.1)	-12.5 (-14.8, -10.2)
p-value			0.064	0.46	0.028
Female	412	45.3 (0.9)	-22.2 (0.8)	-6.1 (0.9)	-16.1 (-18, -14.1)
Male	189	36.5 (1.3)	-23.2 (1.1)	-11.6 (1.4)	-11.6 (-14.4, -8.8)
p-value			0.47	<0.001	0.009
Symptoms Less Than 6 Months	240	42.3 (1.2)	-22.1 (1)	-10 (1.2)	-12.1 (-14.6, -9.7)
Symptoms More Than 6 Months	361	42.7 (1)	-22.8 (0.8)	-6.3 (1)	-16.5 (-18.6, -14.4)
p-value			0.62	0.016	0.006
No Stomach Problem	461	42.1 (0.9)	-24.2 (0.7)	-8.5 (0.9)	-15.6 (-17.5, -13.8)
Stomach Problem	133	43.8 (1.6)	-17 (1.4)	-5.5 (1.6)	-11.5 (-14.8, -8.3)
p-value			<0.001	0.089	0.029
No Joint Problem	250	40.3 (1.2)	-25.6 (1)	-9.2 (1.2)	-16.4 (-18.9, -13.9)
Joint Problem	344	44.1 (1)	-20.1 (0.9)	-7 (1)	-13.1 (-15.2, -11)
p-value			<0.001	0.18	0.041
Neurogenic Claudication	511	42.9 (0.8)	-22.8 (0.7)	-6.9 (0.8)	-15.9 (-17.6, -14.1)
No Neurogenic Claudication	90	40.4 (2.1)	-20.8 (1.8)	-12.4 (1.9)	-8.3 (-12.3, -4.4)
p-value			0.28	0.009	<0.001
Reflex Asymmetry	150	41 (1.4)	-26.6 (1.3)	-9 (1.6)	-17.6 (-20.7, -14.6)
No Reflex Asymmetry	451	43 (0.9)	-21 (0.8)	-7.5 (0.9)	-13.5 (-15.3, -11.6)

Characteristics	N	Baseline ODI Mean (SE)	Mean Change from Baseline		Treatment Effect (95% CI) *
			Surgical (n = 391)	Non-operative (n = 210)	
p-value			<0.001	0.43	0.021
No Opioid Use	343	39.6 (0.9)	-22.5 (0.9)	-10.2 (1)	-12.3 (-14.4, -10.2)
Opioid Use	180	51.5 (1.3)	-22.2 (1.2)	-3.2 (1.7)	-19 (-22.3, -15.7)
p-value			0.84	<0.001	<0.001
No Antidepressant Use	492	43.3 (0.8)	-22.9 (0.7)	-8 (0.9)	-15 (-16.8, -13.2)
Taking Antidepressants	31	49.9 (3.1)	-18.4 (3)	-12.8 (3.7)	-5.7 (-12.8, 1.4)
p-value			0.14	0.21	0.013
Dissatisfied with Symptoms	557	43.6 (0.8)	-22.4 (0.7)	-7.3 (0.8)	-15.1 (-16.7, -13.4)
Satisfied with Symptoms or Neutral	44	29 (2.7)	-21.7 (2.9)	-13.1 (2.3)	-8.7 (-14.4, -2.9)
p-value			0.83	0.02	0.035
Problem Getting Better	38	35.4 (3.3)	-31.5 (2.8)	-8.1 (2.6)	-23.4 (-29.4, -17.4)
Problem Getting Worse	361	45.8 (0.9)	-21.8 (0.8)	-5.9 (1.1)	-16 (-18.1, -13.8)
Problem Staying about the Same	194	37.8 (1.4)	-21.9 (1.2)	-10.8 (1.2)	-11.1 (-13.8, -8.4)
p-value			0.004	0.01	<0.001
Small Chance of Being Pain Free with Surgery	46	33.8 (3.2)	-13.8 (3.3)	-8 (2.3)	-5.8 (-12.3, 0.7)
Moderate Chance of Being Pain Free with Surgery	207	42.9 (1.3)	-22.5 (1.2)	-7.9 (1.2)	-14.6 (-17, -12.1)
Big Chance of Being Pain Free with Surgery	334	43.6 (1)	-22.9 (0.8)	-7.9 (1.2)	-15 (-17.5, -12.6)
p-value			0.027	1	0.029

* Treatment effect is the difference between the surgical and non-operative mean change from baseline. Analysis is done using a mixed effects longitudinal regression model including a random individual effect to account for correlation between repeated measurements within individuals. Treatment is a time-varying covariate where a patients' experience prior to surgery is attributed to the non-operative arm and time is measured from enrollment and his/her post-surgery outcomes are attributed to the surgical arm and time is measured from time of surgery.

Table 2

Area under curve subgroup results for Oswestry Disability Index (ODI) change scores for variables that are not statistically significantly associated with treatment effect, adjusted for age, gender, center, and baseline ODI score* (*Time weighted average 4 years treatment effects*).

Characteristics	N	Baseline ODI Mean (SE)	Mean Change from Baseline		Treatment Effect (95% CI)**
			Surgical (n = 391)	Non-operative (n = 210)	
Baseline ODI 29	165	22.5 (0.8)	-17 (1.4)	0.3 (1.3)	-17.3 (-20.2, -14.4)
29 < Baseline ODI 42	161	38 (0.8)	-21.4 (1.2)	-4.2 (1.5)	-17.2 (-20.3, -14)
42 < Baseline ODI 54	125	49.4 (0.7)	-23.3 (1.4)	-5.9 (1.8)	-17.4 (-21, -13.8)
Baseline ODI > 54	148	63.9 (0.9)	-33.2 (1.3)	-19.7 (1.9)	-13.4 (-17.2, -9.6)
p-value			<0.001	<0.001	0.35
White	506	41.5 (0.8)	-23 (0.7)	-8.2 (0.8)	-14.9 (-16.6, -13.2)
Black	61	49.7 (2.6)	-18.4 (2.2)	-6.1 (2.3)	-12.3 (-17.3, -7.3)
Others	34	45 (3.3)	-21.2 (2.9)	-5.8 (3.5)	-15.4 (-23.2, -7.6)
p-value			0.12	0.59	0.61
High School or Less	198	43.6 (1.2)	-19.9 (1.1)	-6.1 (1.3)	-13.8 (-16.7, -11)
At Least Some College or More	400	42 (0.9)	-24 (0.8)	-8.8 (0.9)	-15.2 (-17.1, -13.3)
p-value			0.003	0.10	0.42
Divorced/Widowed	173	43.7 (1.3)	-19.7 (1.3)	-7.3 (1.4)	-12.4 (-15.3, -9.6)
Married	396	41.9 (0.9)	-23.9 (0.8)	-8.5 (1)	-15.4 (-17.4, -13.4)
Single	26	47.1 (4.2)	-19.1 (3.2)	-1.9 (3.6)	-17.3 (-25, -9.5)
p-value			0.01	0.19	0.18
Not Working	370	43.3 (1)	-21.2 (0.9)	-6.8 (1)	-14.4 (-16.5, -12.3)
Income Over \$50,000	93	39.1 (1.8)	-24.6 (1.7)	-10.5 (2)	-14.1 (-18.1, -10.2)
Income Under \$50,000	137	43 (1.6)	-24.6 (1.4)	-8.8 (1.6)	-15.8 (-19, -12.6)
p-value			0.06	0.24	0.73
Body Mass Index < 30	376	40.3 (0.9)	-23.8 (0.8)	-10.2 (1)	-13.6 (-15.6, -11.5)

Characteristics	N	Baseline ODI Mean (SE)	Mean Change from Baseline		Treatment Effect ** (95% CI)
			Surgical (n = 391)	Non-operative (n = 210)	
Body Mass Index 30	225	46.2 (1.2)	-20.4 (1)	-3.7 (1.3)	-16.7 (-19.2, -14.1)
p-value			0.012	<0.001	0.059
Stenosis Bothersomeness 15	326	37.1 (1)	-23.3 (0.9)	-9.2 (1)	-14.2 (-16.3, -12.1)
Stenosis Bothersomeness > 15	267	49.1 (1.1)	-21.4 (1)	-6.5 (1.2)	-14.9 (-17.3, -12.4)
p-value			0.14	0.11	0.68
Baseline MCS 35	73	55.2 (2)	-15.7 (1.9)	-2.2 (2.3)	-13.5 (-18.2, -8.9)
Baseline MCS > 35	528	40.8 (0.8)	-23.4 (0.7)	-8.7 (0.8)	-14.7 (-16.4, -12.9)
p-value			<0.001	0.008	0.66
Smoker	51	44 (2.7)	-20.3 (2.2)	-8.1 (2.7)	-12.2 (-17.6, -6.8)
Non-Smoker	543	42.4 (0.8)	-22.8 (0.7)	-7.9 (0.8)	-14.9 (-16.6, -13.2)
p-value			0.27	0.94	0.35
No Hypertension	319	41.2 (1)	-23.5 (0.9)	-9.3 (1)	-14.2 (-16.3, -12.1)
Hypertension	275	44 (1.1)	-21.4 (1)	-6.1 (1.2)	-15.3 (-17.8, -12.9)
p-value			0.10	0.039	0.49
No Diabetes	514	42 (0.8)	-23.2 (0.7)	-8.1 (0.8)	-15 (-16.7, -13.3)
Diabetes	80	45.8 (2.3)	-18.8 (1.8)	-6.5 (2.2)	-12.3 (-17.1, -7.6)
p-value			0.025	0.48	0.29
No osteoporosis	525	42.5 (0.8)	-23.4 (0.7)	-8.6 (0.8)	-14.8 (-16.5, -13)
Osteoporosis	69	42.7 (2.1)	-16.1 (2)	-3.1 (2.2)	-13 (-17.7, -8.3)
p-value			<0.001	0.019	0.49
No Heart Problem	472	42.6 (0.8)	-23 (0.7)	-8.5 (0.9)	-14.5 (-16.3, -12.7)
Heart Problem	122	42.1 (1.7)	-20.1 (1.5)	-6.2 (1.6)	-14 (-17.5, -10.4)

Characteristics	N	Baseline ODI Mean (SE)	Mean Change from Baseline		Treatment Effect ** (95% CI)
			Surgical (n = 391)	Non-operative (n = 210)	
p-value			0.085	0.20	0.80
No Bowel or Intestinal Problem	551	42.3 (0.8)	-23.2 (0.7)	-8.4 (0.8)	-14.8 (-16.5, -13.2)
Bowel or Intestinal Problem	43	45.7 (2.8)	-14.6 (2.3)	0.1 (3.3)	-14.6 (-21.3, -8)
p-value			<0.001	0.011	0.95
No Depression	496	40.8 (0.8)	-23.5 (0.7)	-8.6 (0.8)	-14.9 (-16.6, -13.1)
Depression	98	51 (1.8)	-17.8 (1.6)	-4.6 (2)	-13.2 (-17.1, -9.3)
p-value			0.001	0.064	0.44
No Other Comorbidities	367	40.6 (0.9)	-24.2 (0.8)	-10.5 (1)	-13.7 (-15.8, -11.6)
Other Comorbidities [†]	234	45.5 (1.2)	-19.7 (1)	-3.9 (1.2)	-15.8 (-18.3, -13.4)
p-value			<0.001	<0.001	0.18
Zero or One Comorbidity	187	38.4 (1.3)	-26.8 (1.1)	-12.9 (1.3)	-13.9 (-16.7, -11.2)
Two or Three Comorbidities	257	42.3 (1.1)	-23.4 (1)	-7.6 (1.2)	-15.8 (-18.2, -13.3)
Four or More Comorbidities	150	47.9 (1.5)	-14.8 (1.3)	-2.8 (1.5)	-12.1 (-15.3, -8.9)
p-value			<0.001	<0.001	0.18
Positive Tension Sign	85	42.8 (2.1)	-22.7 (1.7)	-10.9 (1.9)	-11.8 (-15.7, -7.8)
Negative Tension Sign	516	42.5 (0.8)	-22.5 (0.7)	-7.2 (0.8)	-15.2 (-17, -13.5)
p-value			0.92	0.08	0.11
No Dermatomal Pain Radiation	133	40.7 (1.5)	-23.2 (1.4)	-6.9 (1.6)	-16.3 (-19.5, -13.1)
Some Dermatomal Pain Radiation	468	43 (0.9)	-22.3 (0.8)	-8.1 (0.9)	-14.3 (-16.1, -12.4)
p-value			0.58	0.53	0.27
Any Neurological Deficit	327	42 (1)	-23.2 (0.9)	-7.7 (1)	-15.6 (-17.7, -13.4)
No Neurological Deficit	274	43.1 (1.1)	-21.6 (1)	-8.1 (1.2)	-13.6 (-15.9, -11.2)

Characteristics	N	Baseline ODI Mean (SE)	Mean Change from Baseline		Treatment Effect ** (95% CI)
			Surgical (n = 391)	Non-operative (n = 210)	
p-value			0.24	0.80	0.21
Sensory Deficit	169	42.6 (1.5)	-21.9 (1.2)	-7 (1.5)	-14.9 (-17.9, -11.9)
Sensory Intact	432	42.5 (0.9)	-22.8 (0.8)	-8.2 (0.9)	-14.6 (-16.5, -12.7)
p-value			0.53	0.50	0.88
Motor Deficit	146	41.1 (1.7)	-22.4 (1.4)	-8.6 (1.6)	-13.8 (-17.1, -10.5)
Motor Intact	455	43 (0.8)	-22.5 (0.8)	-7.6 (0.9)	-14.9 (-16.8, -13.1)
p-value			0.92	0.58	0.54
No L2-L3 Stenosis	548	42.8 (0.8)	-22.7 (0.7)	-8.1 (0.8)	-14.6 (-16.3, -12.9)
L2-L3 Stenosis	53	39.3 (2.5)	-20.6 (2.2)	-4.2 (2.9)	-16.4 (-22.1, -10.6)
p-value			0.37	0.19	0.56
No L3-L4 Stenosis	365	42.7 (1)	-23.3 (0.8)	-8.8 (1)	-14.5 (-16.6, -12.5)
L3-L4 Stenosis	236	42.2 (1.2)	-21.2 (1.1)	-6 (1.3)	-15.2 (-17.7, -12.7)
p-value			0.13	0.091	0.68
No L4-L5 Stenosis	21	37.3 (3.9)	-27 (3.7)	-9.8 (4)	-17.2 (-25.7, -8.7)
L4-L5 Stenosis	580	42.7 (0.8)	-22.4 (0.7)	-7.8 (0.8)	-14.6 (-16.3, -13)
p-value			0.22	0.62	0.56
No L5-S1 Stenosis	544	42.3 (0.8)	-22.7 (0.7)	-7.7 (0.8)	-15 (-16.7, -13.2)
L5-S1 Stenosis	57	45.1 (2.3)	-20.6 (2.2)	-8 (2.3)	-12.6 (-17.3, -7.9)
p-value			0.35	0.93	0.36
No Moderate/Severe Stenotic Levels	23	48.9 (4.2)	-22.7 (3.7)	-4.4 (3.2)	-18.3 (-25.8, -10.8)
One Moderate/Severe Stenotic Level	370	42 (1)	-23.8 (0.8)	-9.5 (1)	-14.3 (-16.3, -12.3)

Characteristics	N	Baseline ODI Mean (SE)	Mean Change from Baseline		Treatment Effect ** (95% CI)
			Surgical (n = 391)	Non-operative (n = 210)	
Two Moderate/Severe Stenotic Levels	172	43.3 (1.4)	-20.5 (1.2)	-4.9 (1.5)	-15.6 (-18.6, -12.6)
Three+ Moderate/Severe Stenotic Levels	36	40.4 (3)	-18 (2.6)	-9.6 (3.5)	-8.4 (-15.3, -1.6)
p-value			0.048	0.035	0.21
No Central Stenosis	52	45.8 (2.6)	-20.3 (2.3)	-7 (2.4)	-13.3 (-18.4, -8.2)
Central Stenosis	549	42.2 (0.8)	-22.7 (0.7)	-7.9 (0.8)	-14.8 (-16.5, -13.1)
p-value			0.31	0.73	0.57
No Lateral Recess Stenosis	55	46.1 (3)	-19.5 (2.2)	-1.5 (2.4)	-18 (-22.8, -13.2)
Lateral Recess Stenosis	546	42.2 (0.8)	-22.8 (0.7)	-8.6 (0.8)	-14.2 (-15.9, -12.5)
p-value			0.15	0.005	0.15
No Neuroforaminal Stenosis	358	42.2 (1)	-21 (0.9)	-6.6 (1)	-14.4 (-16.5, -12.3)
Neuroforaminal Stenosis	243	43 (1.1)	-24.6 (1)	-9.8 (1.2)	-14.9 (-17.3, -12.4)
p-value			0.006	0.04	0.78
Mild/Moderate Stenosis Severity	238	42.4 (1.2)	-21.8 (1.1)	-8.8 (1.2)	-13.1 (-15.5, -10.6)
Severe Stenosis Severity	363	42.6 (1)	-22.9 (0.8)	-7 (1)	-15.9 (-18, -13.8)
p-value			0.43	0.27	0.078
L3-L4 Listhesis Level	57	42 (2.4)	-26.8 (2.1)	-8.8 (2.5)	-18 (-23, -13)
L4-L5 Listhesis Level	544	42.6 (0.8)	-22.1 (0.7)	-7.7 (0.8)	-14.4 (-16.1, -12.7)
p-value			0.034	0.68	0.18
No Prior Injections	270	39.4 (1.1)	-23.7 (1)	-9.6 (1.1)	-14.2 (-16.4, -11.9)
Prior Injections	331	45.1 (1)	-21.6 (0.9)	-6.2 (1.1)	-15.4 (-17.6, -13.1)
p-value			0.10	0.028	0.44

Characteristics	N	Baseline ODI Mean (SE)	Mean Change from Baseline		Treatment Effect ** (95% CI)
			Surgical (n = 391)	Non-operative (n = 210)	
No NSAID Use	316	45.2 (1)	-21.3 (0.9)	-7.9 (1.1)	-13.4 (-15.7, -11.1)
Taking NSAID	207	41.3 (1.3)	-24.6 (1.1)	-8.9 (1.3)	-15.7 (-18.3, -13)
p-value			0.024	0.55	0.21
No Physical Therapy	179	39.9 (1.4)	-21.2 (1.2)	-8.4 (1.4)	-12.8 (-15.8, -9.8)
Had Physical Therapy	422	43.6 (0.9)	-22.9 (0.8)	-7.6 (0.9)	-15.3 (-17.2, -13.4)
p-value			0.25	0.64	0.16
Compensation ^{††}	41	56.7 (2.6)	-12.1 (2.4)	-2.3 (3.7)	-9.8 (-17.1, -2.4)
No Compensation	559	41.5 (0.8)	-23.3 (0.7)	-8.4 (0.8)	-14.9 (-16.6, -13.3)
p-value			<0.001	0.11	0.17
No Litigation	585	42.1 (0.8)	-22.9 (0.7)	-8.3 (0.8)	-14.6 (-16.3, -13)
Litigation Pending/Resolved	15	58.5 (3.6)	-9.8 (3.8)	5.6 (5)	-15.3 (-24.6, -6.1)
p-value			<0.001	0.006	0.88
Predominant Back Pain	154	43.4 (1.5)	-19.4 (1.3)	-5 (1.5)	-14.3 (-17.3, -11.4)
Predominant Leg Pain	437	42.5 (0.9)	-23.7 (0.8)	-9.1 (0.9)	-14.7 (-16.6, -12.7)
p-value			0.003	0.019	0.86
Treatment Preference Not Sure	137	42 (1.4)	-21.7 (1.3)	-7.3 (1.5)	-14.4 (-17.5, -11.3)
Prefer Nonoperative Treatment	235	38 (1.2)	-21.7 (1.2)	-8.9 (1)	-12.7 (-15.1, -10.3)
Prefer Surgery	228	47.6 (1.2)	-22.7 (1)	-6 (2.6)	-16.7 (-21.6, -11.7)
p-value			0.74	0.46	0.33
No Missed Work	177	37.1 (1.2)	-24.1 (1.1)	-8.8 (1.3)	-15.4 (-18, -12.7)
Missed Work	98	48 (1.9)	-22.9 (1.5)	-9.5 (1.9)	-13.4 (-17.2, -9.7)
p-value			0.50	0.76	0.40

Characteristics	N	Baseline ODI Mean (SE)	Mean Change from Baseline		Treatment Effect (95% CI) ^{**}
			Surgical (n = 391)	Non-operative (n = 210)	
No Lifting at Work	323	40.4 (1)	-22.7 (0.9)	-8.9 (1)	-13.9 (-16, -11.7)
Lifting at Work	277	45 (1.1)	-22.3 (1)	-6.5 (1.1)	-15.8 (-18.1, -13.4)
p-value			0.72	0.13	0.24
Small Chance of Being Pain Free with Nonoperative Treatment	309	45.2 (1)	-21.5 (0.9)	-6.6 (1.3)	-14.9 (-17.5, -12.3)
Moderate Chance of Being Pain Free with Nonoperative Treatment	195	38.6 (1.3)	-22.9 (1.2)	-7.6 (1.2)	-15.4 (-17.9, -12.8)
Big Chance of Being Pain Free with Nonoperative Treatment	82	41.5 (2.1)	-24.9 (1.9)	-13.5 (1.9)	-11.4 (-15.4, -7.4)
p-value			0.22	0.008	0.24

^{*} For the ODI subgroup analysis, the baseline ODI quartile variable was in the model instead of the continuous ODI baseline variable.

^{**} Treatment effect is the difference between the surgical and non-operative mean change from baseline. Analysis is done using a mixed effects longitudinal regression model including a random individual effect to account for correlation between repeated measurements within individuals. Treatment is a time-varying covariate where a patients' experience prior to surgery is attributed to the non-operative arm and time is measured from enrollment and his/her post-surgery outcomes are attributed to the surgical arm and time is measured from time of surgery.

[†] Other comorbidities include: stroke, cancer, fibromyalgia, chronic fatigue syndrome (CFS), post-traumatic stress disorder (PTSD), alcohol, drug dependency, lung, liver, kidney, blood vessel, nervous system, migraine or anxiety.

^{††} This category includes patients who were receiving or had applications pending for workers compensation, social security compensation, or other compensation.

Table 3

Area under curve subgroup results from the adjusted* as-treated Oswestry Disability Index (ODI) outcome analysis for the nine significant treatment effect predictors (time weighted average 4 year treatment effects)

Characteristics	N	Baseline ODI Mean (SE)	Mean Change from Baseline		Treatment Effect (95% CI) [†]
			Surgical (n = 391)	Non-operative (n = 210)	
Age 67	312	45.3 (1.1)	-22.9 (0.9)	-7.2 (1.1)	-15.7 (-17.9, -13.4)
Age > 67	289	39.5 (1)	-21.4 (1)	-9.7 (1.1)	-11.8 (-14.1, -9.4)
p-value			0.31	0.15	0.014
Female	412	45.3 (0.9)	-22.7 (0.8)	-7.2 (0.9)	-15.6 (-17.6, -13.6)
Male	189	36.5 (1.3)	-22 (1.2)	-10.7 (1.4)	-11.2 (-14.1, -8.4)
p-value			0.58	0.029	0.01
No Stomach Problem	461	42.1 (0.9)	-23.7 (0.7)	-8.5 (0.9)	-15.2 (-17.1, -13.3)
Stomach Problem	133	43.8 (1.6)	-18.5 (1.4)	-7.2 (1.5)	-11.3 (-14.5, -8.1)
p-value			<0.001	0.48	0.035
Neurogenic Claudication	511	42.9 (0.8)	-22.8 (0.7)	-7.6 (0.8)	-15.3 (-17.1, -13.5)
No Neurogenic Claudication	90	40.4 (2.1)	-20.5 (1.7)	-11.5 (1.9)	-9 (-13, -5)
p-value			0.21	0.056	0.004
Reflex Asymmetry	150	41 (1.4)	-26.7 (1.3)	-9.3 (1.5)	-17.3 (-20.4, -14.2)
No Reflex Asymmetry	451	43 (0.9)	-21 (0.7)	-8 (0.9)	-13 (-14.9, -11.1)
p-value			<0.001	0.45	0.016
No Opioid Use	343	39.6 (0.9)	-22 (0.9)	-10.3 (1)	-11.7 (-13.9, -9.6)
Opioid Use	180	51.5 (1.3)	-23.2 (1.2)	-4.8 (1.6)	-18.4 (-21.7, -15.2)
p-value			0.41	0.004	<0.001
No Antidepressant Use	492	43.3 (0.8)	-22.8 (0.7)	-8.3 (0.9)	-14.5 (-16.4, -12.6)
Taking Antidepressants	31	49.9 (3.1)	-21.1 (3)	-15.8 (3.7)	-5.4 (-12.5, 1.7)

Characteristics	N	Baseline ODI Mean (SE)	Mean Change from Baseline		Treatment Effect (95% CI) [†]
			Surgical (n = 391)	Non-operative (n = 210)	
p-value			0.58	0.05	0.014
Dissatisfied with Symptoms	557	43.6 (0.8)	-22.4 (0.7)	-7.8 (0.8)	-14.5 (-16.3, -12.8)
Satisfied with Symptoms or Neutral	44	29 (2.7)	-21.8 (2.9)	-13.5 (2.3)	-8.3 (-14, -2.6)
p-value			0.85	0.02	0.039
Small Chance of Being Pain Free with Surgery	46	33.8 (3.2)	-13.8 (3.3)	-8.7 (2.3)	-5.1 (-11.5, 1.4)
Moderate Chance of Being Pain Free with Surgery	207	42.9 (1.3)	-22.8 (1.2)	-8.5 (1.2)	-14.2 (-16.7, -11.8)
Big Chance of Being Pain Free with Surgery	334	43.6 (1)	-22.7 (0.8)	-7.9 (1.2)	-14.8 (-17.2, -12.3)
p-value			0.027	0.93	0.019

* Adjusted for center, age, gender, baseline ODI score, work status, BMI, neuroforaminal stenosis, duration of symptoms, reflex deficit, number of moderate/severe stenotic levels, treatment preference, baseline stenosis bothersomeness, other^{**} comorbidity, hypertension, depression, osteoporosis, and joint problems.

** Other comorbidities include: stroke, cancer, fibromyalgia, chronic fatigue syndrome (CFS), post-traumatic stress disorder (PTSD), alcohol, drug dependency, lung, liver, kidney, blood vessel, nervous system, migraine, anxiety.

[†] Treatment effect is the difference between the surgical and non-operative mean change from baseline. Analysis is done using a mixed model with a random subject intercept term. Treatment is a time-varying covariate where a patients' experience prior to surgery is attributed to the non-operative arm and time is measured from enrollment and his/her post-surgery outcomes are attributed to the surgical arm and time is measured from time of surgery.

Table 4

Area under curve subgroup results from the adjusted* as-treated Oswestry Disability Index (ODI) outcome analysis for combinations of significant treatment effect predictors (time weighted average 4 year treatment effects).

Characteristics	N	Baseline ODI Mean (SE)	Surgical	Non-operative	Treatment Effect† (95% CI)
Opioid users with neurogenic claudication, moderate/big chance of being pain free with surgery, and no anti-depressants use	134	51.6 (1.4)	-24.4 (1.3)	-1.8 (2)	-22.6 (-26.5, -18.7)
Non-Opioid users without neurogenic claudication	49	37.4 (2.8)	-18.8 (2.5)	-15.3 (2.5)	-3.5 (-9.1, 2.2)
Others	336	41.3 (1)	-21.8 (0.9)	-9.4 (1)	-12.5 (-14.6, -10.3)
p-value			0.093	<0.001	<0.001

* Adjusted for center, age, gender, baseline ODI score, work status, BMI, neuroforaminal stenosis, duration of symptoms, reflex deficit, number of moderate/severe stenotic levels, treatment preference, baseline stenosis bothersomeness, other** comorbidity, hypertension, depression, osteoporosis, and joint problems.

** Other comorbidities include: stroke, cancer, fibromyalgia, chronic fatigue syndrome (CFS), post-traumatic stress disorder (PTSD), alcohol, drug dependency, lung, liver, kidney, blood vessel, nervous system, migraine, anxiety.

† Treatment effect is the difference between the surgical and non-operative mean change from baseline. Analysis is done using a mixed model with a random subject intercept term. Treatment is a time-varying covariate where a patients' experience prior to surgery is attributed to the non-operative arm and time is measured from enrollment and his/her post-surgery outcomes are attributed to the surgical arm and time is measured from time of surgery.