Surgery versus Nonsurgical Treatment for Lumbar Spinal Stenosis: A Comparative Effectiveness Randomized Trial with 2-Year Follow-up

Anthony Delitto, PT, Ph.D, FAPTA [Professor and Chair],
Department of Physical Therapy, School of Health and Rehabilitation Sciences, University of Pittsburgh delitto@pitt.edu

Sara R. Piva, PT, Ph.D [Associate Professor],
Department of Physical Therapy, SHRS, University of Pittsburgh spiva@pitt.edu

Charity G. Moore, Ph.D, MSPH [Professor],
Department of Medicine, University of Pittsburgh. moorecg@upmc.edu

Julie M. Fritz, PT, Ph.D [Professor],
Department of Physical Therapy, School of Health Professions, University of Utah julie.fritz@hsc.utah.edu

Stephen R. Wisniewski, PH.D [Professor],
Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh stevewis@pitt.edu

Deborah A. Josbeno, PT, Ph.D,
Department of Physical Therapy, SHRS, University of Pittsburgh daj11@pitt.edu

Mark Fye, MD, and
University of Pittsburgh Medical Center fyema2@upmc.edu

William C. Welch, MD
University of Pennsylvania william.welch@uphs.upenn.edu

Abstract

Background: Primary care management decisions for patients with symptomatic Lumbar Spinal Stenosis (LSS) remain a challenge and non-surgical guidance is limited by lack of evidence.

Objective: To compare surgical decompression to physical therapy for LSS and to evaluate gender differences.

Design: Multi-Site Randomized Controlled Trial (ClinicalTrials.gov # NCT00022776).

Corresponding Author: Anthony Delitto, Professor and Chair, Department of Physical Therapy, School of Health and Rehabilitation Sciences, University of Pittsburgh, 6035 Forbes Tower, Pittsburgh, PA 16260, Phone: 412-383-6631 (office) 412-445-8384 (cell), Fax: 412-648-5970, delitto@pitt.edu.

Setting: Departments of neurological and orthopaedic surgery and physical therapy clinics.

Participants: Surgical candidates with LSS 50 years or older who consented to surgery.

Intervention: Surgical decompression or a physical therapy regimen.

Measurements: Primary outcome was physical function score on the SF-36 at the 2 years assessed by masked testers.

Results: Study took place from Nov/2000 to Sept/2007. A total of 169 subjects were randomized stratified by surgeon and gender (87 Surgery; 82 Physical Therapy) with 24 month follow-up completed by 74 and 73 subjects in surgery and physical therapy groups. Mean changes in physical function for surgical and physical therapy groups were 22.4 (95% CI 16.9, 27.9) and 19.2 (95% CI 13.6, 24.8), respectively. Intention to treat (ITT) analyses revealed no difference between groups (24 month difference= 0.85, 95% CI (−7.9, 9.6). Sensitivity analyses using causal effects methods to account for the high proportion of crossovers from PT to surgery (57%) showed no significant differences in physical function between the two groups.

Limitations: Without a control group, it is not possible to judge success attributable to either intervention versus alternative explanations (e.g., regression to the mean).

Conclusions: Results endorse shared decision making between patients and health care providers that includes full disclosure of evidence involving surgery for LSS and patient acceptability and access to non-surgical treatments for LSS.

Lumbar spinal stenosis (LSS) is an anatomical impairment characterized by narrowing of the spinal canal or nerve root foramen.(1) When symptomatic, LSS causes pain, weakness in the lower back and buttocks and thighs and claudicating pain.(1) However, anatomic LSS is also commonly present in older patients who are asymptomatic,(2) which underscores the importance of corroborative findings on history and examination.(3) Even after careful examination, management decisions for people with symptomatic LSS remain a challenge that has aptly been described as a "balancing act"(4) that is compounded by a lack of clear, evidence-based non-surgical treatment options.(5)

Surgery for lumbar spinal stenosis remains an option for patients with persistent and severe symptoms that include both back and leg pain.(6–8) In fact, LSS is now the most often cited cause for lumbar surgery in the USA.(9) Studies comparing surgical with non-surgical treatment for LSS have been conducted but reflect the lack of clarity on optimal non-surgical treatment options.(7, 10–12) The surgical approach in these studies has been highly standardized. In contrast, the non-surgical comparator groups commonly lack structure and detail and have inconsistent follow-through to assure that even the most basic evidence-based approaches are included, including activation, exercise and discouragement of passive agents. In the Spine Patient Outcomes Research Trial (SPORT), the largest randomized controlled trial (RCT) comparing surgical and non-surgical treatment for LSS, the surgical group had a standard posterior decompressive laminectomy whereas the non-surgical group received usual care in which surgeons were encouraged to recommend active physical therapy, education or counseling with home exercise instruction, and nonsteroidal anti-inflammatory medication as initial management strategies, but could receive any additional conservative treatments deemed appropriate by the surgeon.(6) The SPORT RCT had cross-

over rates of 67% for the surgical group (e.g., crossed over to non-surgical intervention) and 43% in the non-surgical intervention (crossed over to surgery), both of which were sufficiently high to preclude intention to treat interpretations. Instead, interpretation was based on "as treated" analyses for virtually all of their outcomes.(11, 13–16) Secondary analysis of patients in the non-surgical arm found only 10% were provided education/ counseling and 37% received physical therapy within 6 weeks and that patients receiving physical therapy had higher self-ratings of improvement and were less likely to cross over to surgery then patients in the non-surgical arm who did not receive physical therapy.(17)

Comparisons between surgery and standardized application of a physical therapy program for patients with LSS have not been performed. Therefore, the primary purpose of this RCT was to compare surgical decompression to a specified non-surgical physical therapy regimen in patients considered surgical candidates for symptomatic, degenerative, LSS. The study also evaluated gender differences in outcomes after treatment for LSS. In order to reduce cross over in at least one arm of the study, we randomized patients after consent to surgery.

METHODS

Design and Overview

Patients were randomly assigned to one of two treatment groups; 1) surgical decompression or 2) physical therapy. The primary outcome of physical function was evaluated through patient-reported outcomes. We followed the patients both in the short-term (first ten weeks) and long-term (6, 12 and 24 months).

Targeted Patient Population: We enrolled patients with a diagnosis of LSS identified by either CT scan using criteria of Wiesel et al.(18) or MRI scan using criteria of Boden et al. (2) All patients were considered by a spine surgeon to be candidates for surgical decompression and consented to surgery. Consent to surgery was a key inclusion criterion to ensure all subjects were surgical candidates and minimize crossover from the surgery into the physical therapy group. Patients also had to meet the following additional eligibility criteria: presence of neurogenic claudication (e.g., self-reported inability to walk greater than ¼ mile due to lower extremity pain and/or cramping); agreeing to be randomly assigned to either surgery or to attend a specified physical therapy clinic for twice weekly exercise sessions; and no previous surgery for LSS at the level being considered for decompression. We excluded patients who were less than 50 years of age; had signs of serious dementia; diagnosis of severe vascular disease or recent history of myocardial infarction; concomitant spondylolisthesis requiring fusion (defined as > 5mm of slippage); compression fractures at the level being considered for decompression; or a diagnosis of metastatic cancer.

Recruitment and Consent Procedures: Five Neurosurgeons and one Orthopaedic spine surgeon from two medical centers participated in subject recruitment. All potential subjects were examined initially by spine surgeons and judged to be surgical candidates. After consenting to surgery, the surgeon introduced the study to the patients. If patients indicated an interest in the study, a study coordinator then fully explained the study and obtained consent to participate in the RCT, after which they would be randomly assigned to have surgery or attend physical therapy. The study was approved by a Data and Safety

Monitoring Board chartered by NIH/NIAMS and the IRBs of the University of Pittsburgh and Allegheny General Hospital (now DBA Allegheny Health Network) and conducted in the Western Pennsylvania area.

Study participants completed self-reported demographic and other descriptive information that included sex, age, ethnicity, past and present activity levels, present income level and employment status, and past medical history. Participants also completed self-reported questionnaires of disability and other psychosocial factors.

Randomization: Patients were randomized after the initial assessment procedures. Randomization was conducted using SAS software with permuted blocks of random block sizes and stratification by gender and surgeon. Sequentially numbered and sealed envelopes with assignment allocation were prepared by the data center using computerized number generator. A study coordinator not involved with preparing the allocation sequence opened the envelopes and assigned participants to interventions.

Intervention

Surgical Decompression: The surgical treatment of patients in was performed similar to that described by Rothman and Simeone.(19) The protocol for surgery was identical to that used in other surgical trials for LSS(20) and included decompressive laminectomies, partial facet resection and neuroforaminotomies performed at the levels of radiographic stenosis. No patients underwent fusion. All surgical procedures were performed by either fellowship-trained spine surgeons or surgeons with 20+ years of experience dedicated to spine surgery. Hospital stay following surgery averaged three days. The post-operative course included a graduated ambulation program that began on post-operative day one. Patients were encouraged to increase their level of walking activity as tolerated. No study funds were used to reimburse expenses of surgical and post-operative care.

Physical Therapy: Patients were treated with a physical therapy program emphasizing lumbar flexion exercises, general conditioning exercises, and patient education. Each patient was evaluated by a physical therapist utilizing an examination scheme described in our previous work.(21) This examination scheme was designed to identify impairments in lower extremity strength and flexibility to be addressed during treatment. Treatment fidelity was assessed by the investigators. The following treatments were administered (Appendix 1):

- Instruction in lumbar flexion exercises including posterior pelvic tilts, supine knee-to-chest and quadruped flexion exercises.

- General conditioning exercises including stationary cycling or treadmill walking.

- Lower extremity strengthening exercises deemed appropriate for the patient from the examination (e.g., standing squats, seated knee extension, supine straight leg raises).

- Lower extremity flexibility exercises deemed appropriate for the patient based on individual examination (e.g., hamstring or hip flexor stretching).

- Patient education to avoid postures involving hyperextension of the lumbar spine.

Physical therapy was prescribed for 6 weeks, with a frequency of 2 visits per week, and delivered by licensed Physical Therapists. Patients were instructed that they could cross over to surgery at any point in the trial based on a shared decision-making process with the spine surgeon. No study funds were used to reimburse physical therapy care.

Outcome and Follow up—The primary outcome for the study was the Physical Function score on the Medical Outcomes Survey Short Form 36 (SF-36)(22) at the 2-year endpoint of the study. In addition, we administered the Oswestry Low Back Pain Disability Questionnaire (ODI),(23) and the North American Spine Society (NASS) Pain and Disability, Neurogenic Symptoms, and Expectation for treatment outcome scales.(24) All evaluations were carried out by research assistants blinded to subject's group assignment. Patients were asked to wear t-shirts to cover surgical scars when reporting for follow up visits.

<u>**Sample Size and Statistical Analyses:**</u> We predetermined with a sample size of 96 subjects per group, we would have 93% power to detect an effect size of .5 standard deviations between the surgical and physical therapy treatment arms for primary outcome (two-sided α=0.05). To evaluate gender differences in outcomes, with a 50% split by gender (48 subjects per group), we would have approximately 79% power to detect a difference in treatment effects of 0.8 SD between men and women. No adjustments to sample size were made for attrition or crossovers. Three interim analyses were planned after 48, 96, and 158 subjects completed 12 month follow-up based on Lan and DeMets alpha spending functions with O'Brien-Fleming boundaries.

Primary results of the study are based on the intention to treat approach. The ITT analysis provides an estimated effect of being offered PT compared to being offered surgery.

Function was measured on the SF-36 at baseline, 10 weeks, 6 months, 12 months, and 24 months; therefore we used linear mixed-effects models(25) to assess for differences in the improvement in function over time between the two groups (controlling for gender and surgeon). Baseline and all follow up measurements within a patient were fitted using maximum likelihood methods with unstructured covariance matrix (SAS Proc MIXED)(25). The linear mixed effects model included fixed effects for group, time, and the group by time interaction. The same approach to the analysis was used for secondary outcomes of Oswestry, NASS Pain and Disability, Neurogenic Symptoms, and Expectation for treatment outcomes. Since an 'as treated' analysis can result in biased estimates due to the factors associated with crossing over, we conducted a complier average causal effect (CACE)(26, 27) and an inverse probability weighting (IPW)(28) analysis due to the high rate of crossovers in the PT group.

Last, we compared the rate of successful outcome between the subjects from the surgical group, the patients that crossed over from the physical therapy group to surgery, and the group assigned to physical therapy and who did not cross over. We defined >0.5 standard deviation improvement at 2-year follow-up relative to baseline as a clinically meaningful effect.

Subgroup analyses were only conducted for gender as this was pre-specified as a goal for our study to assess the impact of gender on treatment effects. We only formally tested for a gender interaction for the primary outcome based on the intention to treat approach.

We conducted sensitivity analyses for physical function using multiple imputation and simple imputation based on non-ignorable missingness to evaluate the impact of missing data on the primary ITT results. All tests were two-sided with $\alpha=0.05$ and all analyses were conducted in SAS version 9.3.

Role of Funding Sources—The study was funded through NIH/NIAMS #AR-NS45622. Funding source did not have a role in the design, conduct, and analysis of this study and the decision to submit the manuscript for publication.

RESULTS

Recruitment took place from November 2000 to October 2005 and the last follow-up was done in September 2007. Of the 481 patients who met eligibility criteria, 312 refused to consent to the study, with the vast majority preferring not to risk randomization and instead went straight to surgery (Fig 1). A total of 169 participants were randomized. Study recruitment was halted in October 2005 prior to meeting recruitment goals due to slow recruitment and limited resources to carry out the remainder of the trial. Planned interim analyses were conducted after 48, 96, and 158 subjects completed 12 month follow-up. There were no significant interim analysis which were conducted based on Lan and DeMets spending functions with O'Brien-Fleming boundaries.

Patients:

Patient characteristics were comparable in both arms (Table 1), except that subjects in the surgery group were in average 3 years younger than those in the physical therapy group. Patients entered the study with numeric pain ratings of 7/10 in both groups.

Cross-overs—All but 2 patients assigned to the surgery group received surgery. These crossovers occurred prior to 10 weeks. In contrast, 47 (57%) of the 82 subjects in the physical therapy group crossed-over to surgery over the 2-year period (Fig 1). Of the 47 crossovers to surgery, 79% had physical therapy with an average of 7.8 visits. In addition, 66% (n=31) of crossovers had surgery prior to 10 weeks and among these 74% (n=23) had at least one PT session with an average of 7.5 visits. Most demographic and clinical characteristics of those in the physical therapy group who crossed-over to surgery were similar to the ones who did not cross-over, except that the subjects who crossed-over had higher pain and lower education as compared to those who did not cross-over (Supplemental Table 1).

Treatment Fidelity: Physical Therapy—The average number of physical therapy sessions attended was 8.4 (4.6). Fifty four patients (66%) attended at least 50% of the prescribed 12 sessions. Thirteen patients (16%) failed to attend at least 1 session of whom 77% (n=10) had surgery.

Primary Treatment Effects—Mean changes in physical function for surgical and physical therapy groups were 22.4 (95% CI 16.9, 27.9) and 19.2 (95% CI 13.6, 24.8), respectively. Intention to treat analyses revealed no difference between the surgical and physical therapy arms of the study at all points of follow-up (p>0.50), including the 2-year primary endpoint [adjusted estimate 0.85, 95% CI (−7.9, 9.6)](Table 2 and Figure 2). Sensitivity analyses for missing data did not result in different findings for physical function.

The difference in physical function at 2-years using the CACE approach with predictors of compliance (sex, baseline pain, education) and controlling for design features of the trial (surgeon, sex, baseline physical function, and imbalance in age) was non-significant [estimate 8.5 95% CI (−14.4, 31.4)] with a wider confidence interval than the ITT approach. The IPW estimate with the same predictors for compliance, two predictors for censoring (treatment and depression), and controlling for design was also not significant [estimate 3.2 95% CI (−5.9, 12.1)] but with a confidence interval similar to the ITT analyses. Appendix 2 provides technical description of CACE and IPW analyses and results. Additionally, we also provide the trajectories of physical function according to the different categories of treatment that subjects received during the trial, including subjects randomized to surgery who received surgery, subjects randomized to surgery who had PT, randomized to PT who had surgery, and those randomized to PT who did not have surgery (Supplemental Table 2 and Supplemental Figure 1).

Success/Failure Analyses—In the randomly assigned surgical group, 45 of the 74 (61%) for which there was available data achieved a successful outcome at 2-year follow-up. In the patients that crossed over from the Physical Therapy group to surgery, 24 of the 44 (55%) achieved a successful outcome. In the group that was assigned to Physical Therapy and who did not cross over, 15 of the 29 (52%) had a successful outcome.

Gender-Related Findings—Subgroup analysis from gender revealed no difference in treatment effects over time for men and women (3-way treatment x time x gender interaction p=0.066; within men treatment x time interaction p= 0.10; within women treatment x time interaction p=0.50) (Figure 3 and Supplemental Table 3).

Complications—There were 33 surgery-related complications, 11 of which occurred in the subjects from the physical therapy group who crossed over (Supplemental Table 4).The most common surgery-related complication was reoperation and delay in wound healing/ surgical site infection. There were 9 physical therapy-related complications and they were all reports of worsening symptoms. While 6 subjects died during the study period (4 in the surgery group and 2 in the physical therapy group), the deaths were not related to study participation. The other complications recorded during study implementation were not directly related to study interventions and were similar in both groups. The most common were worsening symptoms, injuries such as broken bones and falls, and neurological and neuromuscular complication such as vestibular disorders and stroke.

Other Therapies—The utilization of therapies outside of study regimens is reported in Supplemental Table 5. Over 50% of subjects in both groups used pain medication throughout the study. The utilization of therapies such as injections, exercises, orthotics and

pain clinic were low and similar in both groups. Both groups also visited health care providers by a similar rate. Subjects continued to seek spine surgeons throughout the study and during year 2 around 10% of subjects sought this service. Visits to general practitioner were around 20% throughout the study. The other health care providers were visited with lower frequency. Additionally, 21 (24%) of subjects in the surgery group had some form of physical therapy after the surgery. After that, physical therapy outside of study regimen was utilized by 18% in the surgery group and 20% in the physical therapy group during year 1 and 22% by both groups during year 2.

DISCUSSION

In patients with lumbar spinal stenosis (LSS) who were considered surgical candidates and who consented to surgery, no differences were detected between a trial of physical therapy and simple decompression surgery with respect to relieving symptoms and improving function. Results of the Intention to Treat (ITT) analyses showed in both groups that subjects began to demonstrate improvement at the 10 week mark, continued to improve through 26 weeks and maintained improvement the remaining points of follow-up, including at the 2-year end-point of the study. When compared to baseline, the magnitude of improvement at the 2-year follow-up for all groups was well beyond minimal clinically important differences (MCID) for measures of function.(29) While the results demonstrated no statistically significant differences in outcome between groups, interpretation of the confidence intervals indicate that the results are likely indeterminate. For example, the confidence bounds for physical function are consistent with as much as 7.9 point benefit with assignment to PT to as much as 9.6 point benefit with assignment to surgery. Both confidence bounds are above the clinically meaningful benefit for SF-36 physical function. (30–32)

Though we were able to minimize crossover in the surgical group, the high percentage of crossover in the physical therapy group presents a challenge in interpretation. However, alternative analytic techniques (e.g., CACE; inverse probability weighting) resulted in similar interpretations of the data, namely, that any differences between the groups were not significant. Similar to the observations associated with the ITT results, the confidence intervals of the IPW and CACE analyses were wide and included the thresholds of meaningful differences and should be interpreted as indeterminate.

By design, subjects were required to consent to surgery prior to consenting to the study, which resulted in almost 75% of eligible patients refusing to consent to the study. The large percentage of patients refusing to consent may include patients with attributes that would limit generalizability of this study, including patients that were inherently less risk averse or those unable to participate in physical therapy. However, given that 43% of the patients assigned to the physical therapy arm (N=35) avoided surgery at the 2-year follow-up mark, it is reasonable to assume that many patients had not exhausted their non-surgical options before consenting to surgery. This could be due to inaccurate self-reported attempts of non-surgical care (e.g., physical therapy) or that the physical therapy care received was inadequate either due to: (1) non-adherence to best care standards in physical therapy environments or (2) barriers to adherence to physical therapy (e.g., burden of co-payments). Overall, these results strongly endorse a shared decision making approach between patients,

surgeons and their primary care providers with full disclosure of the results of this and other studies involving surgery and LSS.

Unlike other LSS RCTs, (6, 7, 11) the patients in this study were considered surgical candidates and consented to surgery prior to randomization. This criterion successfully reduced crossover from surgery to physical therapy to less than 3% while the crossover rate from physical therapy to surgery (57%) was higher than the SPORTS trial (43%), possibly because of difference in patient eligibility. In contrast, Malmivaara et al.(7) excluded subjects whose signs and symptoms suggested "forthcoming treatment" (e.g., surgery), which was just the opposite of this study. In spite of differences in inclusion and exclusion criteria, however, Malmivaara et al.(7) and the SPORT trials all reported similar baseline levels of disability based on the ODI as this study. In the surgical groups (those randomized and those that crossed over), the ODI change over 2-years in this study was also similar to the SPORT trial(6) and slightly higher as compared to Malmivaara.(7) The major difference in this trial compared with previous trials was the greater degree of improvement in the physical therapy arm in this trial as compared to the non-operative arms of SPORT(6) and Malmivaara et al.(7) trials, which might be partially explained by the fact that in this study most of the subjects in the non-operative (e.g., physical therapy) arm received exercise and instruction as opposed to SPORT,(6) which did not control the non-operative interventions.

Though females improved in the surgical and physical therapy arms, the magnitude of change lagged behind men, a gender difference that has been described in the LSS literature (33–35) but not reported in the previous RCTs of LSS (Fig 3). Baseline measures related to demographics and clinical characteristics did not account for this gender difference. The possible gender difference in outcomes for LSS treatments requires further study.

The majority of subjects (84%) in the physical therapy arm attended at least one visit to physical therapy with 66% attending at least 50% of the prescribed visits. We did not see a relationship between attendance in physical therapy and outcome, including the probability of success. Though we did not track the reasons for non-attendance, numerous members of the study team reported that co-payments were a major obstacle to attendance in physical therapy.

Limitations

Sixty-five percent of eligible patients refused admittance to the study primarily because they did not wish to risk the 50% chance of being assigned to the non-surgical group. We recognize there could be key differences between the consented and non-consented cohorts, which should be a strong consideration when generalizing the results of this trial. Though the design of the study successfully controlled for crossover in the surgery arm, 57% of patients assigned to physical therapy cross-over to surgery, which presents a challenge in interpreting intention to treat analyses. Though alternative analytic techniques were used in this study, interpreting ITT analyses remains a challenge. Both groups' functional outcome improved to levels of clinical meaningfulness. However, without a control group it is not possible to judge changes in functional outcome attributable to either intervention versus alternative explanations (e.g., regression to the mean).

CONCLUSION

Patients with LSS who were surgical candidates and consented to surgery achieved similar long-term functional gains with surgical decompression as with an evidence-based physical therapy treatment regimen. Though there were similar proportions of successes with both groups, there were also similar proportions of patients that failed to achieve a level of improvement that would be considered success. Future research should focus on a more expanded approach to defining predictors of success and failures with surgical and non-surgical approaches to symptomatic LSS.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References Cited

1. Katz JN, Harris MB. Clinical practice. Lumbar spinal stenosis. [Review] [44 refs]. New England Journal of Medicine. 2008;358(8):818–25. [PubMed: 18287604]

2. Boden SD. The use of radiographic imaging studies in the evaluation of patients who have degenerative disorders of the lumbar spine. [Review] [128 refs]. Journal of Bone & Joint Surgery - American Volume. 1996;78(1):114–24.

3. Katz JN, Dalgas M, Stucki G, Katz NP, Bayley J, Fossel AH, et al. Degenerative lumbar spinal stenosis. Diagnostic value of the history and physical examination. Arthritis & Rheumatism. 1995;38(9):1236–41. [PubMed: 7575718]

4. Deyo RA. Treatment of lumbar spinal stenosis: a balancing act. Spine Journal: Official Journal of the North American Spine Society. 2010;10(7):625–7.

5. Atlas SJ, Delitto A. Spinal stenosis: surgical versus nonsurgical treatment. [Review] [92 refs]. Clinical Orthopaedics & Related Research. 2006;443:198–207. [PubMed: 16462443]

6. Weinstein JN, Tosteson TD, Lurie JD, Tosteson AN, Blood E, Hanscom B, et al. Surgical versus nonsurgical therapy for lumbar spinal stenosis. New England Journal of Medicine. 2008;358(8): 794–810. [PubMed: 18287602]

7. Malmivaara A, Slatis P, Heliovaara M, Sainio P, Kinnunen H, Kankare J, et al. Surgical or nonoperative treatment for lumbar spinal stenosis? A randomized controlled trial. Spine. 2007;32(1):1–8. [PubMed: 17202885]

8. Atlas SJ, Keller RB, Wu YA, Deyo RA, Singer DE. Long-term outcomes of surgical and nonsurgical management of sciatica secondary to a lumbar disc herniation: 10 year results from the maine lumbar spine study. Spine. 2005;30(8):927–35. [PubMed: 15834338]

9. Weinstein JN, Lurie JD FAU - Olson P, Olson PR FAU - Bronner K, Bronner KK FAU - Fisher E, Fisher ES. United States' trends and regional variations in lumbar spine surgery: 1992–2003. (1528–1159 (Electronic)).

10. Atlas SJ, Keller RB, Wu YA, Deyo RA, Singer DE. Long-term outcomes of surgical and nonsurgical management of lumbar spinal stenosis: 8 to 10 year results from the maine lumbar spine study. Spine. 2005;30(8):936–43. [PubMed: 15834339]

11. Weinstein JN, Tosteson TD, Lurie JD, Tosteson A, Blood E, Herkowitz H, et al. Surgical versus nonoperative treatment for lumbar spinal stenosis four-year results of the Spine Patient Outcomes Research Trial. Spine. 2010;35(14):1329–38. [PubMed: 20453723]

12. Amundsen T, Weber HF, Nordal HJ FAU, Magnaes BF, Abdelnoor MF, Lilleas F. Lumbar spinal stenosis: conservative or surgical management?: A prospective 10-year study. (0362–2436 (Print)).

13. Katz JN, Stucki G, Lipson SJ, Fossel AH, Grobler LJ, Weinstein JN. Predictors of surgical outcome in degenerative lumbar spinal stenosis. Spine. 1999;24(21):2229–33. [PubMed: 10562989]

14. 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(21):1791–802. [PubMed: 23018805]

15. Radcliff KE, Rihn J, Hilibrand A, DiIorio T, Tosteson T, Lurie JD, 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;36(25):2197–210. [PubMed: 21912308]

16. Rihn JA, Radcliff K, Hilibrand AS, Anderson DT, Zhao W, Lurie J, et al. Does obesity affect outcomes of treatment for lumbar stenosis and degenerative spondylolisthesis? Analysis of the Spine Patient Outcomes Research Trial (SPORT). Spine. 2012;37(23):1933–46. [PubMed: 22614793]

17. Fritz JM, Lurie JD, Zhao W, Whitman JM, Delitto A, Brennan GP, et al. Associations between physical therapy and long-term outcomes for individuals with lumbar spinal stenosis in the SPORT study. LID - S1529–9430(13)01586–6 [pii] LID - 10.1016/j.spinee.2013.09.044 [doi]. (1878–1632 (Electronic)).

18. Wiesel SW, Tsourmas N, Feffer HL, Citrin CM, Patronas N. A study of computer-assisted tomography. I. The incidence of positive CAT scans in an asymptomatic group of patients. Spine. 1984;9(6):549–51. [PubMed: 6495024]

19. Rothman S The Spine. 3 ed. Philadelphia: W.B. Saunders; 1992.

20. Birkmeyer NJ, Weinstein JN, Tosteson AN, Tosteson TD, Skinner JS, Lurie JD, et al. Design of the Spine Patient outcomes Research Trial (SPORT). Spine. 2002;27(12):1361–72. [PubMed: 12065987]

21. Fritz JM, Erhard RE, Vignovic M. A nonsurgical treatment approach for patients with lumbar spinal stenosis. Phys Ther. 1997;77(9):962–73. [PubMed: 9291953]

22. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care. 1992;30(6):473–83. [PubMed: 1593914]

23. Fairbank JC, Couper J, Davies JB, O'Brien JP. The Oswestry low back pain disability questionnaire. Physiotherapy. 1980;66(8):271–3. [PubMed: 6450426]

24. 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 (Phila Pa 1976). 1996;21(6):741–9. [PubMed: 8882698]

25. Garrett M. Fitzmaurice NML, James H. Ware. Applied Longitudinal Analysis 2ed. Hoboken: Wiley; 2011.

26. Knox CR, Lall R, Hansen Z, Lamb SE. Treatment compliance and effectiveness of a cognitive behavioural intervention for low back pain: a complier average causal effect approach to the BeST data set. BMC Musculoskelet Disord. 2014;15:17. [PubMed: 24423146]

27. Skrondal R-H. Generalized latent variable modeling: multilevel, longitudinal, and structural equation models. Boca Raton, FL: Chapman & Hall/CRC; 2004.

28. Toh S, Hernan MA. Causal inference from longitudinal studies with baseline randomization. Int J Biostat. 2008;4(1):Article 22.

29. Copay AG, Glassman SD, Subach BR, Berven S, Schuler TC, Carreon LY. Minimum clinically important difference in lumbar spine surgery patients: a choice of methods using the Oswestry Disability Index, Medical Outcomes Study questionnaire Short Form 36, and Pain Scales. The Spine Journal. 2008;8(6):968–74. [PubMed: 18201937]

30. Lauridsen HH, Hartvigsen J, Manniche C, Korsholm L, Grunnet-Nilsson N. Responsiveness and minimal clinically important difference for pain and disability instruments in low back pain patients. BMC Musculoskelet Disord. 2006;7:82. [PubMed: 17064410]

31. Angst F, Aeschlimann A, Stucki G. Smallest detectable and minimal clinically important differences of rehabilitation intervention with their implications for required sample sizes using WOMAC and SF-36 quality of life measurement instruments in patients with osteoarthritis of the lower extremities. Arthritis Rheum. 2001;45(4):384–91. [PubMed: 11501727]

32. Davidson M, Keating JL. A comparison of five low back disability questionnaires: reliability and responsiveness. Phys Ther. 2002;82(1):8–24. [PubMed: 11784274]

33. Johnsson KE, Redlund-Johnell I, Dan A, illner ST. Preoperative and Postoperative Instability in Lumbar Spinal Stenosis. Spine. 1989;14(6).

34. Kim HJ, Suh BG FAU - Lee D-B, Lee DB FAU - Park J-Y, Park JY FAU - Kang K-T, Kang KT FAU - Chang B-S, Chang BS FAU - Lee C-K, et al. Gender difference of symptom severity in lumbar spinal stenosis: role of pain sensitivity. (2150–1149 (Electronic)).

35. Shabat S, Folman Y, Arinzon Z, Adunsky A, Catz A, Gepstein R. Gender differences as an influence on patients' satisfaction rates in spinal surgery of elderly patients. European Spine Journal. 2005;14(10):1027–32. [PubMed: 15912353]
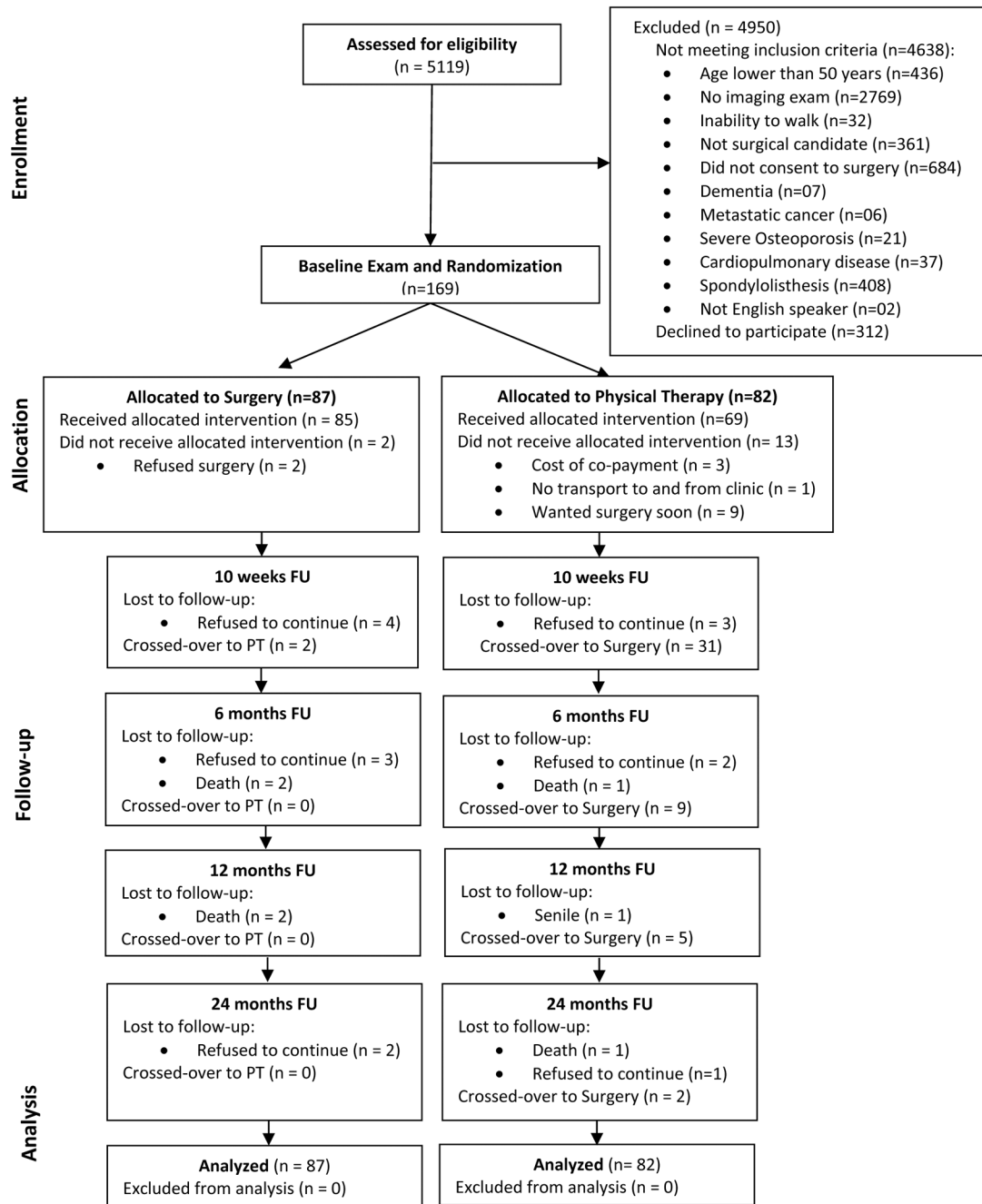
Enrollment

Assessed for eligibility
(n = 5119)

Excluded (n = 4950)
- Not meeting inclusion criteria (n=4638):
  - Age lower than 50 years (n=436)
  - No imaging exam (n=2769)
  - Inability to walk (n=32)
  - Not surgical candidate (n=361)
  - Did not consent to surgery (n=684)
  - Dementia (n=07)
  - Metastatic cancer (n=06)
  - Severe Osteoporosis (n=21)
  - Cardiopulmonary disease (n=37)
  - Spondylolisthesis (n=408)
  - Not English speaker (n=02)
- Declined to participate (n=312)

Baseline Exam and Randomization
(n=169)

Allocation

Allocated to Surgery (n=87)
Received allocated intervention (n = 85)
Did not receive allocated intervention (n = 2)
- Refused surgery (n = 2)

Allocated to Physical Therapy (n=82)
Received allocated intervention (n=69)
Did not receive allocated intervention (n= 13)
- Cost of co-payment (n = 3)
- No transport to and from clinic (n = 1)
- Wanted surgery soon (n = 9)

Follow-up

10 weeks FU
Lost to follow-up:
- Refused to continue (n = 4)
Crossed-over to PT (n = 2)

10 weeks FU
Lost to follow-up:
- Refused to continue (n = 3)
Crossed-over to Surgery (n = 31)

6 months FU
Lost to follow-up:
- Refused to continue (n = 3)
- Death (n = 2)
Crossed-over to PT (n = 0)

6 months FU
Lost to follow-up:
- Refused to continue (n = 2)
- Death (n = 1)
Crossed-over to Surgery (n = 9)

12 months FU
Lost to follow-up:
- Death (n = 2)
Crossed-over to PT (n = 0)

12 months FU
Lost to follow-up:
- Senile (n = 1)
Crossed-over to Surgery (n = 5)

24 months FU
Lost to follow-up:
- Refused to continue (n = 2)
Crossed-over to PT (n = 0)

24 months FU
Lost to follow-up:
- Death (n = 1)
- Refused to continue (n=1)
Crossed-over to Surgery (n = 2)

Analysis

Analyzed (n = 87)
Excluded from analysis (n = 0)

Analyzed (n= 82)
Excluded from analysis (n = 0)

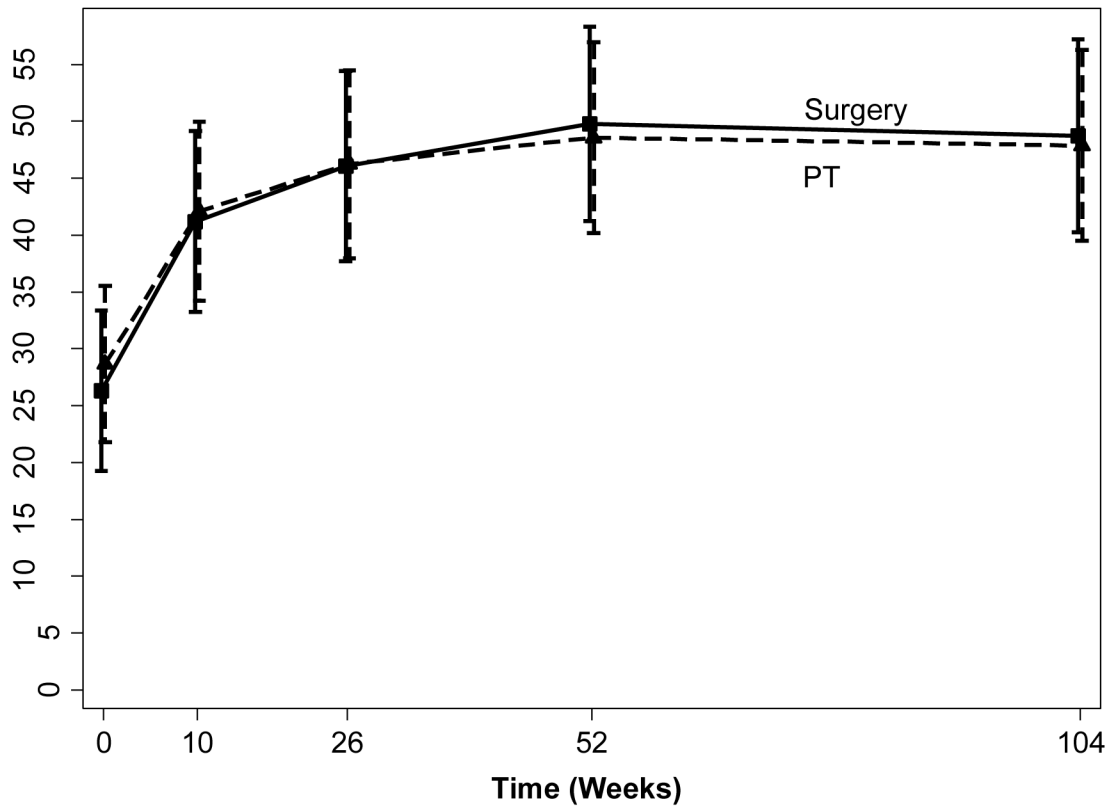**Figure 1.**
Stenosis Trial Flow Diagram.

Figure 2. Adjusted Means for Physical Function over Time in the Surgery and Physical Therapy Groups.

Adjusted means and 95% confidence intervals of physical function scale of the Short Form-36 (SF-36) for the surgery and physical therapy groups over time from linear mixed models (adjusted for gender, surgeon and baseline age). SF-36 scale ranges from 0 to 100, with lower scores indicating more severe symptoms.
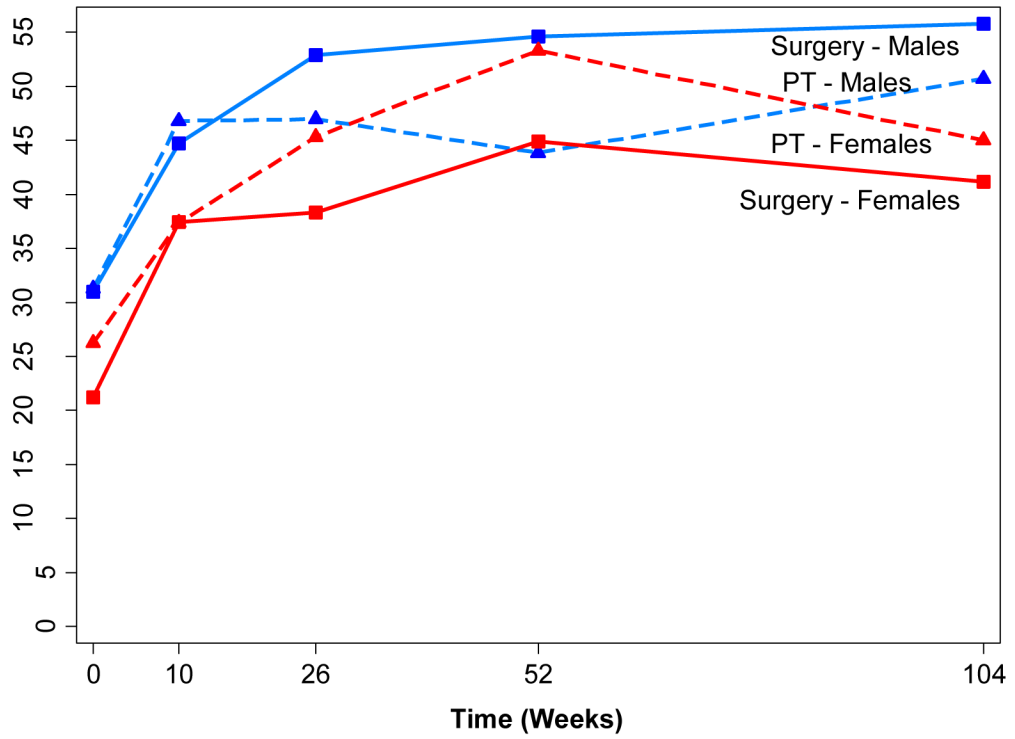
PT stands for Physical Therapy.

Figure 3. Adjusted Means for Physical Function over Time by Treatment and Gender using Intention to Treat Analyses.

Adjusted means of physical function scale of the Short Form-36 (SF-36) for the surgery (solid lines) and physical therapy (dashed lines) groups by gender (male = blue lines; female = red lines) over time from linear mixed models (adjusted for surgeon and baseline age).

Three-way interaction for gender*group*time p=0.066; group*time within men p=0.10; group*time within women p=0.50.

SF-36 scale ranges from 0 to 100, with lower scores indicating more severe symptoms.

PT stands for Physical Therapy.

Table 1.

Baseline Demographics and Clinical Characteristics of Participants.

	Surgery (n = 87)	Physical Therapy (n = 82)
Age at enrollment, years, mean (SD) [*]	66.6 (10.5)	69.8 (9.0)
Male (n, %)	44 (51)	44 (54)
Height, cm, mean (SD)	168.0 (12.5)	170.7 (13.1)
Weight, kg, mean (SD)	89.1 (19.9)	91.0 (19.3)
BMI, mean (SD)	31.5 (6.8)	31.1 (5.4)
Race (n, %)		
Caucasian	83 (95)	77 (94)
African American	4 (5)	5 (6)
Hispanic ethnicity	0 (0)	0 (0)
Education level (n, %)		
High school or less	38 (44)	40 (49)
Some college	23 (26)	21 (26)
College degree	26 (30)	21 (26)
Marital status (n, %)		
Single	1 (1)	3 (4)
Married	63 (72)	59 (72)
Divorced/separated	14 (16)	8 (10)
Widowed/other	9 (10)	12 (15)
Employment situation (n, %)		
Currently working	22 (25)	17 (21)
Unemployed	4 (5)	1 (1)
Retired	50 (57)	52 (63)
Disabled and/or retired due to health problems	11 (13)	12 (15)
Government programs currently enrolled (n, %)		
Social security	54 (62)	63 (77)
Disability	9 (10)	10 (12)
Workers compensation	1 (1)	0 (0)
Present Activity level (n, %)		
Inactive/mildly active	83 (95)	74 (90)
Moderately/very active	4 (5)	8 (10)
Comorbidities (median, Q25 – Q75)	1.0 (0.0 – 2.0)	1.0 (1.0 – 2.0)
Cigarette smoker (n, %)	8 (9)	5 (6)
Current level of pain, mean (SD)	6.8 (2.2)	6.6 (2.2)
Depression score in BAI, mean (SD)	9.3 (6.3)	7.9 (4.5)
Length of back pain episode in days (median, Q25 – Q75)	233 (136 – 602)	282 (111 – 678)
Number of prior back surgeries not for same spinal segment (n, %)		
No prior surgeries	79 (91)	65 (79)
1 prior surgery	7 (8)	12 (15)

	Surgery (n = 87)	Physical Therapy (n = 82)
> 1 prior surgery	1 (1)	5 (6)
Number of prior series of PT for LBP, mean (SD) (n = 77)	4.4 (5.7)	4.4 (4.5)

PT stands for Physical Therapy. LBP stands for Low Back Pain. BDI stands for Beck Depression Inventory. The BDI is a 21-item scale that measures symptoms of depression. Each item is scored from 0-3, with possible range of scores between 0-63. Higher numbers indicating greater severity of depressive symptoms.

*
p<0.05.

Table 2.

Changes in Outcome over Time in the Surgery and Physical Therapy Groups.

Primary Outcome	Group	Baseline		Week 10		Week 26		Week 52		Week 104	
		n	Mean (95% CI)	n	Mean (95% CI)	n	Mean (95% CI)	n	Mean (95% CI)	n	Mean (95% CI)
SF-36 Physical Function [‡]	Surgery	87	26.8 (23.2, 30.4)	80	42.5 (37.1, 47.9)	78	47.2 (41.1, 53.3)	73	51.1 (44.9, 57.3)	74	49.5 (43.1, 55.9)
	PT	82	28.2 (23.9, 32.5)	73	41.0 (35.3, 46.7)	75	45.4 (39.3, 51.5)	75	47.9 (41.4, 54.4)	73	47.6 (40.7, 54.4)
Mean Difference (95% CI)											1.9 (−7.3, 11.2)
Adjusted Mean Difference (95% CI) [†]											0.9 (−7.9, 9.6)
Secondary Outcomes											
Oswestry [§]	Surgery	87	42.6 (39.5, 45.7)	79	33.1 (29.1, 37.1)	75	28.5 (24.5, 32.5)	74	29.2 (25.1, 33.4)	71	25.2 (21.5, 29.0)
	PT	82	40.2 (36.9, 43.5)	74	33.5 (29.6, 37.5)	76	27.2 (23.3, 31.1)	75	29.5 (25.2, 33.8)	73	27.0 (22.8, 31.1)
Mean Difference (95% CI)											−1.7 (−7.3, 3.8)
Adjusted Mean Difference (95% CI) [†]											−1.8 (−7.3, 3.7)
NASS [#] - Pain and Disability	Surgery	87	2.6 (2.5, 2.8)	79	2.0 (1.8, 2.2)	75	1.7 (1.5, 1.9)	74	1.7 (1.5, 2.0)	71	1.6 (1.4, 1.8)
	PT	82	2.5 (2.3, 2.6)	74	2.0 (1.8, 2.2)	75	1.6 (1.4, 1.8)	75	1.8 (1.5, 2.0)	73	1.6 (1.4, 1.9)
Mean Difference (95% CI)											−0.1 (−0.4, 0.2)
Adjusted Mean Difference (95% CI) [†]											−0.1 (−0.3, 0.2)
NASS– Neurogenic Symptoms	Surgery	87	3.7 (3.4, 4.0)	80	2.5 (2.2, 2.8)	77	2.4 (2.2, 2.7)	74	2.6 (2.3, 2.9)	71	2.4 (2.1, 2.7)
	PT	82	3.9 (3.6, 4.1)	74	2.8 (2.4, 3.1)	76	2.6 (2.3, 3.0)	73	2.5 (2.2, 2.9)	71	2.6 (2.2, 2.9)
Mean Difference (95% CI)											−0.2 (−0.6, 0.3)
Adjusted Mean Difference (95% CI) [†]											−0.1 (−0.6, 0.3)
NASS - Treatment Expectation	Surgery	87	3.8 (3.6, 4.0)	75	2.4 (2.1, 2.6)	73	2.5 (2.2, 2.8)	72	2.6 (2.3, 2.9)	69	2.8 (2.5, 3.1)
	PT	82	3.8 (3.5, 4.0)	66	2.7 (2.4, 2.9)	73	2.7 (2.4, 3.0)	68	2.6 (2.3, 2.9)	67	2.9 (2.6, 3.2)
Mean Difference (95% CI)											−0.1 (−0.6, 0.4)
Adjusted Mean Difference (95% CI) [†]											−0.1 (−0.5, 0.4)

PT stands for Physical Therapy. CI stands for Confidence Interval.

Values represent non-adjusted means and 95% confidence intervals unless otherwise noted.

[†] The adjusted difference between the surgery and physical therapy groups over time were compared using linear mixed effects models with adjustments for gender, surgeon and baseline age. P values for group*time interaction effect were above 0.5 for the primary and all secondary outcomes. Adjusted mean differences and 95% confidence intervals are derived from contrasts from the linear mixed models.

†The Short Form-36 (SF-36) physical function is a self-administered questionnaire that assesses the concept of physical functioning and scores range from 0 – 100.

§The Oswestry is a self-administered questionnaire with ten areas of performance including pain intensity, lifting, sitting, standing, walking, traveling, personal hygiene, social activity, sex life and sleeping. Scores range from 0 – 100.

#The North American Spine Society (NASS) is a self-administered questionnaire. The NASS Pain and disability: ranges from 1 – 6 and based on eleven items of performance including back pain, buttock pain, dressing, lifting, sitting, standing, walking, traveling, social activity, sex life and sleeping. The NASS Neurogenic symptoms: ranges from 1 – 6 and based on six items of performance including leg pain, numbness/tingling, weakness in leg/foot. NASS Expectation for Treatment: ranges from 1 – 5 and based on six items of performance including relief from symptoms, do more every day, sleep more comfortably, go back to my usual job, exercise and do recreational activities, and prevent future disability.