Surgical vs Nonoperative Treatment for Lumbar Disk Herniation

The Spine Patient Outcomes Research Trial (SPORT) Observational Cohort

James N. Weinstein, DO, MSc
Jon D. Lurie, MD, MS
Tor D. Tosteson, ScD
Jonathan S. Skinner, PhD
Brett Hanscom, MS
Anna N. A. Tosteson, ScD
Harry Herkowitz, MD
Jeffrey Fischgrund, MD
Frank P. Cammisa, MD
Todd Albert, MD
Richard A. Deyo, MD, MPH

Context For patients with lumbar disk herniation, the Spine Patient Outcomes Research Trial (SPORT) randomized trial intent-to-treat analysis showed small but not statistically significant differences in favor of diskectomy compared with usual care. However, the large numbers of patients who crossed over between assigned groups precluded any conclusions about the comparative effectiveness of operative therapy vs usual care.

Objective To compare the treatment effects of diskectomy and usual care.

Design, Setting, and Patients Prospective observational cohort of surgical candidates with imaging-confirmed lumbar intervertebral disk herniation who were treated at 13 spine clinics in 11 US states and who met the SPORT eligibility criteria but declined randomization between March 2000 and March 2003.

Interventions Standard open diskectomy vs usual nonoperative care.

Main Outcome Measures Changes from baseline in the Medical Outcomes Study Short-Form Health Survey (SF-36) bodily pain and physical function scales and the modified Oswestry Disability Index (American Academy of Orthopaedic Surgeons/ MODEMS version).

Results Of the 743 patients enrolled in the observational cohort, 528 patients received surgery and 191 received usual nonoperative care. At 3 months, patients who chose surgery had greater improvement in the primary outcome measures of bodily pain (mean change: surgery, 40.9 vs nonoperative care, 26.0; treatment effect, 14.8; 95% confidence interval, 10.8-18.9), physical function (mean change: surgery, 40.7 vs nonoperative care, 25.3; treatment effect, 15.4; 95% CI, 11.6-19.2), and Oswestry Disability Index (mean change: surgery, –36.1 vs nonoperative care, –20.9; treatment effect, –15.2; 95% CI, –18.5. to –11.8). These differences narrowed somewhat at 2 years: bodily pain (mean change: surgery, 42.6 vs nonoperative care, 32.4; treatment effect, 10.2; 95% CI, 5.9-14.5), physical function (mean change: surgery, 43.9 vs nonoperavtive care 31.9; treatment effect, 12.0; 95% CI; 7.9-16.1), and Oswestry Disability Index (mean change: surgery –37.6 vs nonoperative care –24.2; treatment effect, –13.4; 95% CI, –17.0 to –9.7).

Conclusions Patients with persistent sciatica from lumbar disk herniation improved in both operated and usual care groups. Those who chose operative intervention reported greater improvements than patients who elected nonoperative care. However, nonrandomized comparisons of self-reported outcomes are subject to potential confounding and must be interpreted cautiously.

Trial Registration clinicaltrials.gov Identifier: NCT00000410

JAMA. 2006;296:2451-2459 www.jama.com

SEVERAL STUDIES HAVE COMPARED surgical and nonoperative treatment of patients with herniated disk, but baseline differences between treatment groups, small sample sizes with limited geographic participation, or lack of validated outcome measures in these studies limit evidence-based conclusions regarding optimal treatment.[1-3] Results for the Spine Patient Outcomes Research Trial (SPORT) randomized trial for intervertebral disk herniation are reported in a companion article. In that study, both surgical and nonoperative patients experienced significant improvement over time and intent-to-treat analyses showed no significant differences between the randomized groups for the primary outcome measures.

Randomized trials of surgery face a number of challenges[4] and their gen-

See also pp 2441, 2483, and 2485.

Author Affiliations: Dartmouth Medical School, Hanover, NH (Drs Weinstein, Lurie, T. Tosteson, Skinner, and A. Tosteson, and Mr Hanscom); William Beaumont Hospital, Royal Oak, Mich (Drs Herkowitz and Fischgrund); Hospital for Special Surgery, New York, NY (Dr Cammisa); Rothman Institute, Philadelphia, Pa (Dr Albert); and Center for Cost and Outcomes Research, University of Washington, Seattle (Dr Deyo).
Corresponding Author: James N. Weinstein, DO, MSc, Department of Orthopaedics, Dartmouth Medical School, One Medical Center Drive, Lebanon, NH 03756 (james.n.weinstein@dartmouth.edu).

eralizability has been questioned.[5-9] Are patients willing to be randomized between surgery and nonoperative treatment representative of those seen in clinical practice? In addition, when the surgical procedure is elective (as in the SPORT trial), treatment crossover is more common, complicating the interpretation of intent-to-treat effects.

In anticipation of these concerns, SPORT was designed to include a concurrent observational cohort study in which identical selection and outcomes assessment occurred, but participants declined randomization. This article reports the 2-year follow-up results for the SPORT intervertebral disk herniation observational cohort.

METHODS

Study Design

SPORT was conducted in 11 US states at 13 medical centers with multidisciplinary spine practices. The human subjects committees at each participating institution approved a standardized protocol for both the observational and the randomized cohorts. Patient inclusion and exclusion criteria, study interventions, outcome measures, and follow-up procedures have been reported.[5]

Patient Population

All men and women who had symptoms and confirmatory signs of lumbar radiculopathy that persisted for at least 6 weeks, who had disk herniation at a corresponding level and side on imaging, who were considered surgical candidates, and who met inclusion criteria were eligible. The content of preenrollment nonoperative care was not prespecified in the protocol but included the following: physical therapy (73%); epidural injections (50%); chiropractic (38%); anti-inflammatories (58%); and opioid analgesics (49%).

A research nurse at each site identified potential participants and verified eligibility. Participants were offered enrollment in either the randomized trial or the observational cohort; participants in the observational cohort chose their treatment (surgery vs nonoperative treatment) at enrollment after consultation with their physician. Enrollment began in March of 2000 and ended March 2003.

Study Interventions

The surgery was a standard open diskectomy with examination of the involved nerve root.[5,10] The nonoperative protocol was "usual care" recommended to include at least active physical therapy, education and counseling with home exercise instruction, and nonsteroidal anti-inflammatory drugs if tolerated. Nonoperative treatments were individualized for each patient and tracked prospectively.

Study Measures

Primary end points were 2 scales of the Medical Outcomes Study Short-Form Health Survey (SF-36)—bodily pain scale and physical function scale[11]—and the American Academy of Orthopaedic Surgeons MODEMS version of the Oswestry Disability Index (ODI)[12] as measured at 6 weeks, 3 months, 6 months, and 1 and 2 years. Secondary outcomes included patient self-reported improvement, work status, satisfaction with current symptoms and care,[13] and sciatica severity as measured by the Sciatica Bothersomeness Index.[2,14]

Statistical Considerations

Primary analyses compared changes from baseline and percentages of patients showing improvement at each follow-up time based on treatments received. In these analyses, the treatment indicator (ie, surgery vs nonoperative) was a time-varying covariate, allowing for variable times of surgery. Prior to the time of surgery, all changes from baseline were included in the estimates of the effect of nonoperative treatment. Following surgery, subsequent changes in outcomes were assigned to the surgical group with follow-up times measured from the date of surgery. Due to the allowable windows for scheduled visits, the actual time of outcome assessment varied (eg, a 6-week follow-up might occur at 5 weeks or 7 weeks). To adjust for this variation, individual visit times were used to fit a linear trend for each planned visit, and the linearly interpolated mean value was used to compute the treatment effect at that follow-up.

To adjust for potential confounding, baseline variables associated with missing data or treatment received were included as adjusting covariates in longitudinal regression models.[15] A random effect was specified to account for the correlation between the repeated measurements on individuals. Computations were done using SAS procedures PROC MIXED for continuous data with normal random effects, and PROC GENMOD for binary and non-normal secondary outcomes, software version 9.1 (SAS Institute Inc, Cary, NC). Statistical significance was defined as $P<.05$ based on a 2-sided hypothesis test.

RESULTS

Overall, 1244 SPORT participants with lumbar intervertebral disk herniation were enrolled out of 1991 eligible for enrollment (FIGURE 1). Five hundred one patients agreed to participate in the randomized controlled trial and are reported in another article in this issue of *JAMA*.[16] The 743 patients who declined to enroll in the randomized controlled trial comprised the observational cohort. Seven hundred nineteen patients (97%) completed at least 1 follow-up visit and were included in the analysis; between 82% and 89% of enrollees supplied data at each follow-up interval.

Five hundred twenty-one patients initially choosing surgery and 222 patients initially choosing nonoperative care were enrolled. For the group initially choosing surgery, 91% received surgery within 6 weeks of enrollment, with an additional 4% receiving surgery by 6 months; at 2 years 4% remained nonoperative. In the group initially choosing nonoperative treatment, 2% underwent surgery in the first 6 weeks; while 16% had surgery by 6 months, and 22% had surgery by 2 years. Overall, 528 patients received surgery during the first 2 years and 191 remained nonoperative (TABLE 1).

Patient Characteristics

The baseline characteristics of participants are shown in Table 1, according to whether they actually received surgery during the 2 years of follow-up. A comparison between the SPORT observational and randomized cohorts is also provided.

The study population was a mean age of 41.4 years with a majority being men, of white race, completing some college, and working full-time or part-time; 18% were receiving disability compensation. Ninety-eight percent had classic dermatomal pain radiation. Most of the herniations were at L5-S1, were posterolateral, and were extrusions by imaging criteria.[17]

At baseline, the surgery group was younger, heavier, less likely to be working, more likely to be receiving disability compensation, and reported fewer comorbid joint problems than those in the nonoperative group. They had more disk extrusions, positive contralateral straight leg raise, and neurological deficits; more severe bodily pain and back pain–related disability; lower levels of physical function; worse sciatica; and more often rated symptoms as getting worse at enrollment than those in the nonoperative group. The final model controlled for age, sex, race, marital status, work status, compensation, body mass index, smoking status, joint problems, migraines, neurological deficit, herniation (type, level, location), baseline score (for SF-36 and ODI), baseline sciatica bothersomeness, baseline satisfaction with symptoms, self-rated health trend, center, and health insurance status.

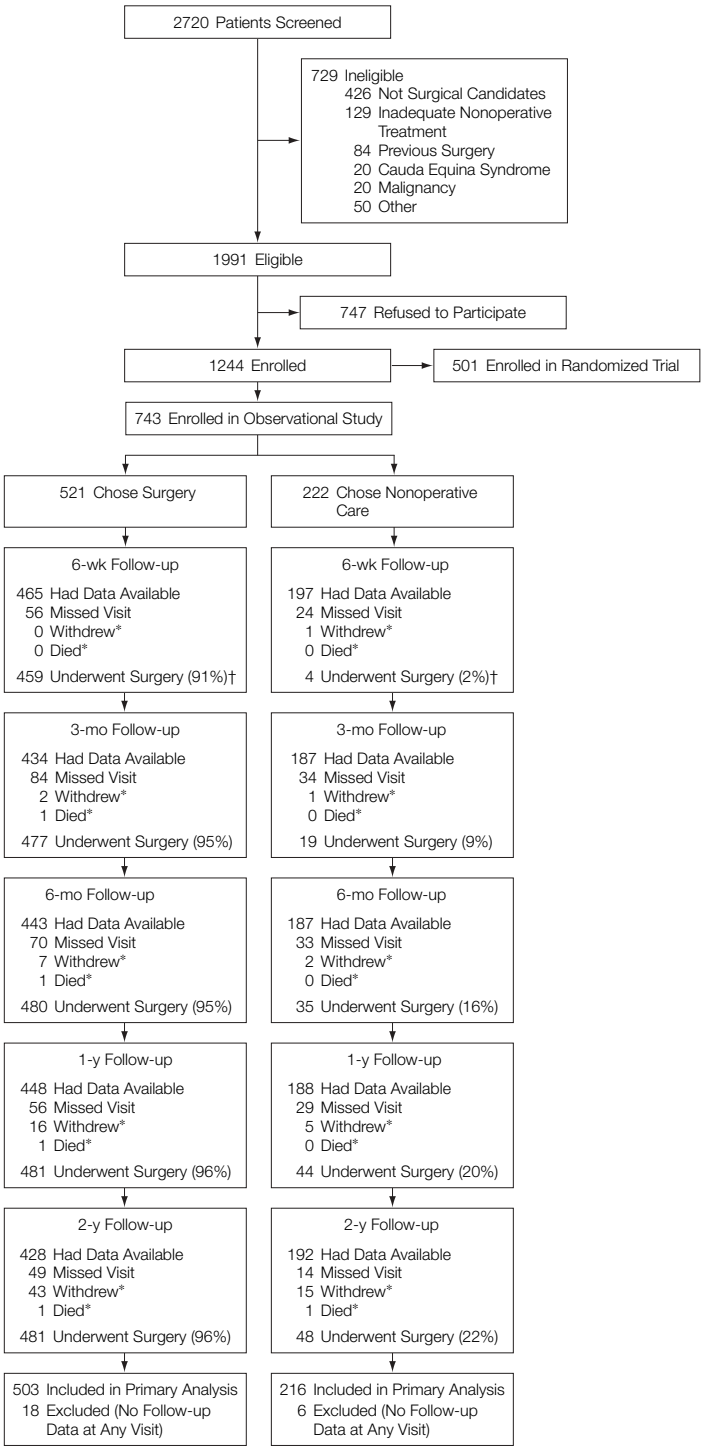
Nonoperative Treatments

A variety of nonoperative treatments were used during SPORT. In the observational cohort, 92% received education and counseling, 58% received nonsteroidal anti-inflammatory drugs, 35% received narcotic analgesic agents, 43% underwent physical therapy, and 38% underwent epidural injections.

Surgical Treatment and Complications

The median surgical time was 70 minutes (interquartile range, 15-333 minutes) with an median blood loss of 50

Figure 1. Flow Diagram of SPORT Observational Cohort for Herniated Disk: Exclusion, Enrollment, and Follow-up



SPORT indicates Spine Patient Outcomes Research Trial.
*Cumulative.
†Percentages of patients undergoing surgery were calculated using the number included in the primary analysis as the denominator (n=503 for surgery; n=216 for nonoperative care).

Table 1. Patient Baseline Demographic Characteristic, Comorbidities, Clinical Findings, and Health Status Measures by Treatment Received, and Also Compared With the Patients in the SPORT Randomized Controlled Trial*

	Treatment Received			All Observational Patients (n = 719)	SPORT Randomized Controlled Trial Patients (n = 472)	
	Surgery (n = 528)	Nonoperative (n = 191)	*P* Value			*P* Value
Demographics						
Age, mean (SD), y	40.5 (10.9)	43.7 (11.9)	<.001	41.4 (11.2)	42.3 (11.6)	.15
Women	227 (43)	86 (45)	.69	313 (44)	194 (41)	.44
Non-Hispanic ethnicity	506 (96)	183 (96)	.84	689 (96)	448 (95)	.55
White race	472 (89)	162 (85)	.12	634 (88)	399 (85)	.08
Education: at least some college	379 (72)	149 (78)	.12	528 (73)	355 (75)	.54
Annual income <$50 000	237 (45)	91 (48)	.57	328 (46)	207 (44)	.59
Married	367 (70)	135 (71)	.83	502 (70)	332 (70)	.90
Work status						
Full-time or part-time	302 (57)	129 (68)	.04	431 (60)	290 (61)	.71
Disabled	80 (15)	20 (10)		100 (14)	58 (12)	
Other	145 (27)	42 (22)		187 (26)	124 (26)	
Compensation†	110 (21)	22 (12)	.006	132 (18)	76 (16)	.35
Clinical						
Body mass index, mean (SD)	28.3 (5.8)	26.9 (5)	.004	27.9 (5.6)	28 (5.5)	.86
Current smoker	136 (26)	38 (20)	.13	174 (24)	108 (23)	.65
Comorbidities						
Depression	53 (10)	26 (14)	.22	79 (11)	62 (13)	.30
Joint problem	76 (14)	48 (25)	.001	124 (17)	97 (21)	.17
Other‡	217 (41)	88 (46)	.27	305 (42)	221 (47)	.15
Time since recent episode <6 mo	406 (77)	151 (79)	.61	557 (77)	372 (79)	.63
Dermatomal pain radiation	519 (98)	185 (97)	.37	704 (98)	457 (97)	.32
Straight leg raise (ipsilateral)	344 (65)	115 (60)	.26	459 (64)	290 (61)	.44
Straight leg raise (contralateral or both)	108 (20)	13 (7)	<.001	121 (17)	67 (14)	.26
Any neurological deficit	418 (79)	133 (70)	.01	551 (77)	350 (74)	.36
Reflexes-asymmetrical depressed	213 (40)	65 (34)	.15	278 (39)	202 (43)	.17
Sensory-asymmetrical decrease	295 (56)	86 (45)	.01	381 (53)	222 (47)	.05
Motor-asymmetrical weakness	246 (47)	65 (34)	.004	311 (43)	190 (40)	.33
Herniation level§						
L2-L3 or L3-L4	32 (6)	24 (13)	.005	56 (8)	32 (7)	.09
L4-L5	209 (40)	82 (43)		291 (40)	165 (35)	
L5-S1	287 (54)	85 (45)		372 (52)	274 (58)	
Herniation type						
Protruding	134 (25)	62 (32)	.05	196 (27)	126 (27)	.86
Extruded	358 (68)	111 (58)		469 (65)	313 (66)	
Sequestered	36 (7)	18 (9)		54 (8)	32 (7)	
Posterolateral herniation	408 (77)	133 (70)	.05	541 (75)	377 (80)	.07
SF-36 scale, mean (SD)						
Bodily pain‖	21.2 (15.8)	36.2 (20.3)	<.001	25.2 (18.3)	26.9 (17.9)	.11
Physical function ‖	30.8 (23.0)	52.5 (25.9)	<.001	36.6 (25.6)	39.4 (25.3)	.06
Mental component summary‖	44.2 (11.1)	46.1 (11.6)	.05	44.7 (11.2)	45.9 (12)	.09
Oswestry Disability Index, mean (SD)¶	56.7 (18.9)	35.9 (20.1)	<.001	51.2 (21.4)	46.9 (21)	<.001
Sciatica Frequency Index, mean (SD)¶	16.9 (4.9)	13.6 (5.6)	<.001	16.0 (5.3)	15.6 (5.5)	.18
Sciatica Bothersomeness Index, mean (SD)¶	16.7 (4.9)	13.4 (5.8)	<.001	15.8 (5.3)	15.2 (5.2)	.05
Satisfaction with symptoms: very dissatisfied	471 (89)	113 (59)	<.001	584 (81)	369 (78)	.23
Patient self-assessed health trend						
Problem getting better	31 (6)	58 (30)	<.001	89 (12)	90 (19)	<.001
Problem staying about the same	221 (42)	92 (48)		313 (44)	220 (47)	
Problem getting worse	272 (52)	39 (20)		311 (43)	161 (34)	

Abbreviation: SF-36, Medical Outcomes Study Short-Form Health Survey.
*Data are presented as number (percentage) unless otherwise indicated.
†Receiving workers' compensation, Social Security compensation, or other compensation or have a pending application.
‡Indicates problems related to stroke, diabetes, osteoporosis, cancer, fibromyalgia, chronic fatigue syndrome, posttraumatic stress disorder, alcohol or drug dependency, heart, lung, liver, kidney, blood vessel, nervous system, migraine, anxiety, stomach, bowel.
§The diagnosis for approximately 97% of patients evaluated with magnetic resonance imaging and 3% with computed tomography.
‖For SF-36 scales, a hight score indicates less severe symptoms.
¶For the Oswestry Disability Index and Sciatica Frequency and Bothersomeness Indices, a lower score indicates less severe symptoms.

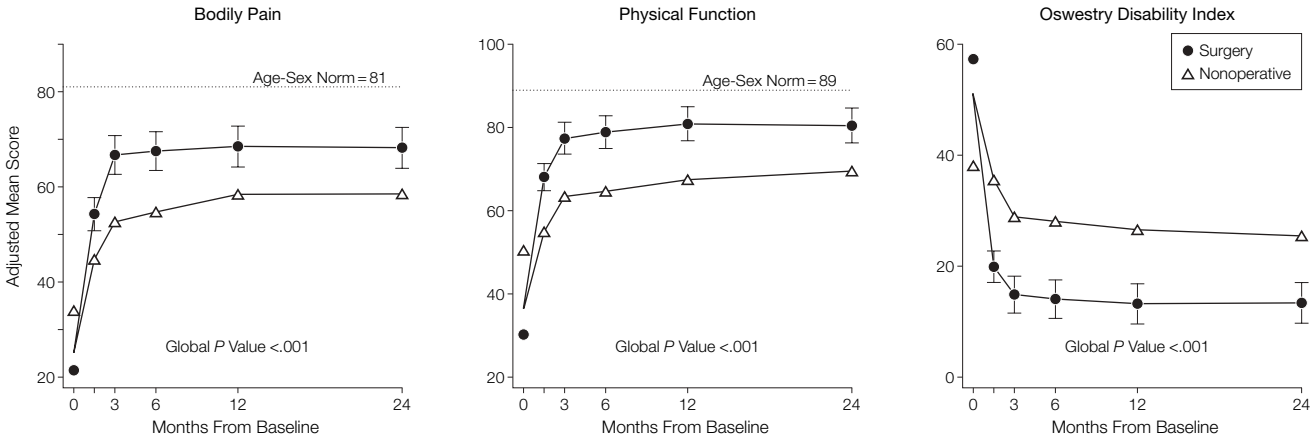
mL (interquartile range, 0-1500 mL). Only 2 patients required transfusions. There were no perioperative mortalities. The most common surgical complication was dural tear in 2% of cases. Reoperation occurred in 7% of cases by 1 year and in 9% of cases at 2 years; more than half were recurrent herniations at the same level.

Table 2. Adjusted Primary and Secondary Outcomes Change Scores, Percent and Treatment Effects for the Intervertebral Disk Herniation Observational Cohort According to Treatment Received*

	3 Months			1 Year			2 Years		
	Surgery (n = 466)	Nonoperative (n = 190)	Treatment Effect (95% CI)†	Surgery (n = 460)	Nonoperative (n = 171)	Treatment Effect (95% CI)†	Surgery (n = 456)	Nonoperative (n = 165)	Treatment Effect (95% CI)†
Primary outcomes									
SF-36 scale, mean (SE)‡									
Bodily pain	40.9 (1.1)	26.0 (1.8)	14.8 (10.8 to 18.9)	42.8 (1.1)	32.0 (1.9)	10.8 (6.5 to 15.0)	42.6 (1.1)	32.4 (1.9)	10.2 (5.9 to 14.5)
Physical function	40.7 (1.0)	25.3 (1.7)	15.4 (11.6 to 19.2)	44.3 (0.99)	29.2 (1.9)	15.0 (10.9 to 19.2)	43.9 (0.99)	31.9 (1.9)	12.0 (7.9 to 16.1)
Oswestry Disability Index, mean (SE)§	–36.1 (0.87)	–20.9 (1.5)	–15.2 (–18.5 to –11.8)	–37.7 (0.85)	–22.4 (1.7)	–15.2 (–18.9 to –11.6)	–37.6 (0.85)	–24.2 (1.7)	–13.4 (–17.0 to –9.7)
Secondary outcomes									
Sciatica Bothersomeness Index, mean (SE)‖	–11.4 (0.27)	–7.5 (0.45)	–3.8 (–4.9 to –2.8)	–11.2 (0.26)	–8.6 (0.48)	–2.6 (–3.6 to –1.5)	–10.8 (0.26)	–8.7 (0.48)	–2.1 (–3.2 to –1.0)
Working full or part time, % (SE)	77.0 (2.5)	81.8 (3.6)	–4.9 (–13.5 to 3.7)	89.3 (1.5)	80.0 (4.0)	9.3 (1.0 to 17.7)	89.1 (1.5)	86.5 (3.0)	2.5 (–4.1 to 9.1)
Posttreatment satisfaction, % (SE)									
Very or somewhat satisfied with symptoms	68.1 (2.3)	29.4 (3.7)	38.7 (30.0 to 47.4)	71.1 (2.2)	44.7 (4.3)	26.4 (16.8 to 36.1)	71.5 (2.2)	49.1 (4.3)	22.4 (12.8 to 32.0)
Very or somewhat satisfied with care	91.3 (1.3)	77.3 (3.6)	13.9 (6.4 to 21.5)	92.4 (1.2)	82.3 (3.4)	10.1 (3.0 to 17.2)	92.5 (1.2)	80.9 (3.4)	11.7 (4.5 to 18.9)
Self-rated progress since enrollment: major improvement, % (SE)	82.6 (1.8)	48.2 (4.2)	34.4 (25.4 to 43.4)	80.4 (1.9)	60.1 (4.3)	20.2 (10.9 to 29.5)	75.8 (2.1)	58.0 (4.2)	17.9 (8.4 to 27.3)

Abbreviations: SF-36, Medical Outcomes Study Short-Form Health Survey.
*Adjusted for age, sex, race, marital status, work status, compensation, body mass index, smoking status, joint problems, migraines, any neurological deficit, herniation (type, level, location), baseline evaluation scores (SF36, ODI, and sciatica scales), baseline sciatica bothersomeness, baseline satisfaction with symptoms, self-rated health trend, center, insurance. Note, for sciatica bothersomeness and satisfaction with symptoms the "baseline score" is equivalent to "baseline sciatica bothersomeness" and "baseline satisfaction with symptoms," respectively.
†The global *P* value assessing all time points simultaneously is less than .001 for all measures.
‡SF-36 scale scores range from 0 to 100, with a higher score indicating less severe symptoms.
§Scores for the Oswestry Disability Index range from 0 to 100 with a low score indicating less severe symptoms.
‖Scores from the Sciatica Bothersomeness Index range from 0 to 24 with a low score indicating less severe symptoms. (n=455 for surgery group at 3 months; data not collected at 3 months for late surgeries.)

Figure 2. Main Outcomes at Baseline and Each Follow-up Visit Through 2 Years



The data markers at time 0 indicate actual mean baseline scores. The curves begin at the overall baseline mean, and the subsequent data markers indicate means adjusted for baseline variables. The adjusting baseline variables are named in the footnotes of Table 2 and include the score plotted. The length of the error bars indicates the 95% confidence interval for the treatment difference between the study groups at each time point. The error bars are centered on the values of the surgery group. If the 95% confidence interval crosses the value in the nonoperative group, the *P* value for the difference between the groups is greater than .05.

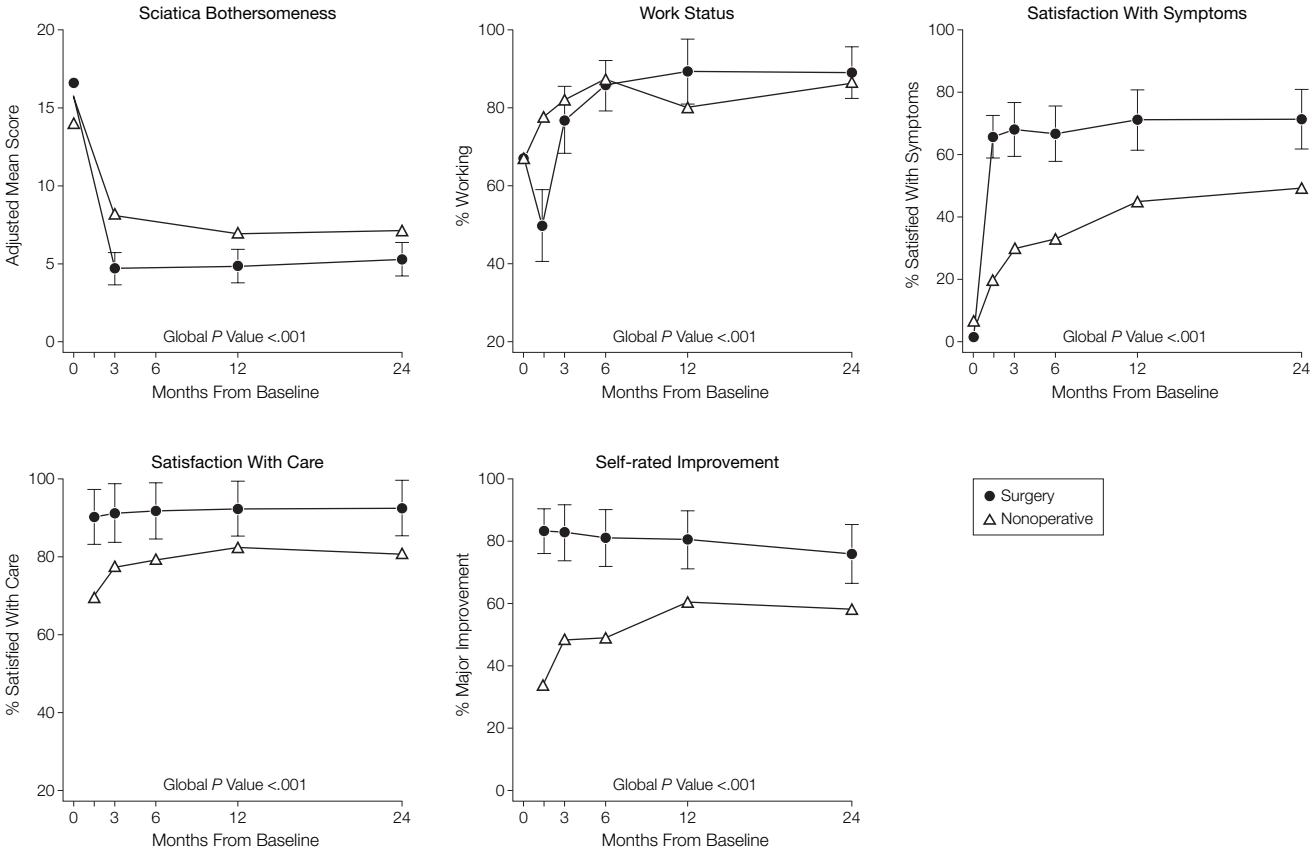
Main Treatment Effects

Treatment outcomes for the observational cohort are summarized in TABLE 2, FIGURE 2, and FIGURE 3. Treatment effects were statistically significant in favor of surgery for the primary outcome measures at 3 months: bodily pain (mean change: surgery, 40.9 vs nonoperative, 26.0; treatment effect, 14.9; 95% confidence interval [CI], 10.8-18.9), physical function (mean change: surgery, 40.7 vs nonoperative, 25.3; treatment effect, 15.4; 95% CI, 11.6-19.2), and ODI (mean change: surgery, –36.1 vs nonoperative –20.9; treatment effect, –15.2; 95% CI, –18.5 to –11.8); at 1 year: bodily pain (mean change: surgery, 42.8 vs nonoperative, 32.0; treatment effect, 10.8; 95% CI, 6.5-15.0), physical function (mean change: surgery, 44.3 vs nonoperative, 29.2; treatment effect, 15.0; 95% CI, 10.9-19.2), and ODI (mean change: surgery, –37.7 vs nonoperative, –22.4; treatment effect, –15.2; 95% CI, –18.9 to –11.6), and 2 years: bodily pain (mean change: surgery, 42.6 vs nonoperative, 32.4; treatment effect, 10.2; 95% CI, 5.9-14.5), physical function (mean change: surgery, 43.9 vs nonoperative, 31.9; treatment effect, 12.0; 95% CI, 7.9-16.1), and ODI (mean change: surgery, –37.6 vs nonoperative, –24.2; treatment effect, –13.4; 95% CI, –17.0 to –9.7). The secondary measures of sciatica bothersomeness, satisfaction, and self-rated improvement also demonstrated significant treatment effects. The treatment effects narrowed between 3 months and 2 years but remained significant at all periods. Work status was worse in the surgery group at 6 weeks but this had equalized at 3 and 6 months; work status then showed a small benefit for surgery at 1 year but not at 2 years.

Missing Data and Shifting Baselines

The percentages of participants with missing data were equivalent between

Figure 3. Work Status, Satisfaction With Symptoms, Satisfaction With Care, and Self-rated Health Trend at Baseline and Each Follow-up Visit Through 2 Years



The data markers at time 0 indicate actual mean baseline scores or proportions, except for satisfaction with care and self-rated improvement, which were not measured at baseline. The curves begin at the overall baseline mean or proportion, and the subsequent data markers indicate means or proportions adjusted for baseline variables. The adjusting baseline variables are named in the footnotes of Table 2 and include the score or factor plotted, except for satisfaction with care and self-rated improvement. The length of the error bars indicates the 95% confidence interval for the treatment difference between the study groups at each time point. The error bars are centered on the values of the surgery group. If the 95% confidence interval crosses the value in the nonoperative group, the *P* value for the difference between the groups is greater than .05.

the groups at each time point with no evidence of differential dropout (Figure 1). At year 2 the missing data percentages were 17% for the surgery group and 14% for the nonoperative group. Sensitivity analysis was completed comparing our primary analysis using longitudinal models including covariates associated with missed visits with alternative analytic methods using single-imputation of missing data—baseline value carried forward and last value carried forward.[15] Treatment effect estimates at 1 year ranged from 9.0 to 11.3 for bodily pain, 14.3 to 15.0 for physical function, –13.9 to –15.2 for ODI, and –2.1 to –2.6 for sciatica. Given these ranges, there appear to be no substantial differences among these methods.

Several alternative approaches for other features of the primary treatment effect analyses were also evaluated. Models using the enrollment values as baseline for the surgically treated group, rather than the visit prior to surgery, and which evaluated outcomes from the time of enrollment rather than the time from surgery, produced similar estimates for the 1-year outcomes. Strategies excluding the nonoperative experience of patients ultimately undergoing surgery or ignoring the correlation between patients contributing both nonoperative and surgical visits showed smaller but still statistically significant treatment effects in favor of surgery. Models without adjustment for baseline differences between the groups showed much larger treatment effects in favor of surgery as would be expected from regression to the mean since the surgery group started out with worse health status scores. Controlling for this regression to the mean in the adjusted models is important for estimating the true treatment effect.

COMMENT

Patients presenting with signs and symptoms of radiculopathy for at least 6 weeks secondary to an image-confirmed lumbar disk herniation experienced substantial improvement over time in both treatment groups, but improvement was significantly greater for those patients who underwent surgery. The benefit of surgery was seen as early as 6 weeks and was maintained for at least 2 years.

Interpretation of the clinical significance of changes seen in quality-of-life scales is important. Despite interest in knowing the minimal clinically important difference for various scales, no consensus exists with regards to methods for providing such benchmarks.[18,19] However, based on published work, reasonable estimates for the minimal clinically important difference for the scales used in SPORT were 10 points for the SF-36 subscales,[2] and 8 to 12 points for the ODI.[20,21] The SPORT results based on the observational cohort exceed this threshold for at least 2 years, arguing that the results seen are indeed of clinical importance.

Debate continues in the scientific literature regarding the optimal role of observational studies vs randomized trials. The design of SPORT provided an opportunity to compare randomized trial results with results for a simultaneously enrolled observational cohort. These 2 groups were similar at baseline. Patients in the observational cohort were relatively more symptomatic and functionally impaired than those in the randomized controlled trial; however, the absolute differences were small: 4 points on the ODI, <3 on the SF-36 PF, and 0.6 on the Sciatica Bothersomeness Index.

Patient perception that the problem was getting worse at enrollment was a more striking factor predicting participation in the observational cohort as well as in initially choosing surgery. This preference for surgery seemed to be an important factor for those declining randomization. Arega et al[22] reported those preferring surgery were only one fourth as likely as those preferring nonoperative care to randomize; alternatively, those who were unsure about their treatment preference at baseline were 3.6 times more likely to participate in the randomized trial.

The results of SPORT are similar to the Maine Lumbar Spine Study[1] and the classic Weber study.[3] The former reported unadjusted treatment effect differences at 1 year of 24 (bodily pain) and 22 (physical function), similar to SPORT's 15.3 and 25.1, respectively (unadjusted data not shown). However, these unadjusted results overestimate the true effect of surgery because of baseline differences between groups. While there are no validated outcome measures that can be directly compared between SPORT and the Weber study, its 1-year results of 33% more patients with "good" results in the surgical group is similar to SPORT's 21% more patients with major improvement and 26% having more satisfaction with symptoms 1 year after surgery than those who received nonoperative care. In these prior studies, the differences in the outcomes between treatment groups continued to narrow over time, suggesting the importance of ongoing follow-up of the patients in SPORT.

Limitations

The strict eligibility criteria may limit the generalizability of the SPORT results, eg, patients unable to tolerate symptoms for 6 weeks or who prefer early surgical intervention were not included and we can draw no conclusions regarding the effectiveness of surgery in that group. However, SPORT entry criteria followed published guidelines for patient selection for elective diskectomy and therefore these results should apply to the majority of patients with a herniated disk facing a surgical decision.[23]

The protocol for nonoperative treatment was usual care individualized to each patient and in keeping with published guidelines. The same basic approach was used in the Maine Lumbar Spine Study.[23] This flexible nonoperative protocol reflects current practice among multidisciplinary spine practices but precludes evaluation of the results of surgery compared with specific nonoperative

treatments. To the degree that some of the nonoperative treatments used were ineffective or inappropriate, the benefits of surgery may be overestimated. However, the 1-year improvements in the usual care group (bodily pain, 32.0; physical function, 29.2; Sciatica, –8.6) were excellent and were greater than the 20-, 18-, and –3.0-point improvements, respectively reported in the Maine Lumbar Spine Study. Usual care appeared to have been generally effective, although we cannot say which components were or were not effective. Nor can we say what the nonoperative outcomes would have been with a hypothetical optimal nonoperative regimen.

Missing data was an important limitation in interpreting study results. Although it did not appear that data were missing differentially between treatment and usual care groups, the effects of missing data in 14% to 18% of follow-up surveys cannot be certain. Multiple sensitivity analyses were used to determine the impact of missing data, and all suggest that the observed differences persist even if missing data were accounted for in the most conservative fashion.

An important limitation in this study design and in all nonmasked treatment intervention studies is that, when measuring subjective outcomes, the differences in motivation for recovery, expectation of treatment success, and perception of changes in health status may affect the results. Patients who elected to have an operation were different in some ways that suggested that they had a greater burden of disease, but they may have been different in other unmeasured ways. Furthermore, in any unmasked study, differences in perceptions of care may also affect subjective outcomes.

The results in this observational cohort were similar to the as-treated results from the randomized cohort reported in another article in this issue of *JAMA*.[16] The greater proportion of patients who elected to have surgery in the observational cohort did not substantially alter the treatment outcomes. However, observational comparisons cannot account for all patient- and surgeon-level factors that differ between the groups and it remains unclear if some of these account for part or all of the differential effect observed between treatment groups.

CONCLUSION

In this nonrandomized evaluation of patients with persistent sciatica from lumbar disk herniation who had operative or usual care, both treatment groups improved considerably over 2 years. Nonrandomized comparisons of self-reported outcomes are subject to potential confounding and must be interpreted cautiously. Nevertheless, patients who underwent diskectomy had significantly better self-reported outcomes than those who had usual care.

Author Contributions: Dr Weinstein had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Study concept and design: Weinstein, Lurie, T. Tosteson, A. Tosteson
Acquisition of data: Weinstein, Lurie, T. Tosteson, Herkowitz, Fischgrund, Cammisa, Albert.
Analysis and interpretation of data: Weinstein, Lurie, T. Tosteson, Skinner, Hanscom, A. Tosteson, Deyo.
Drafting of the manuscript: Weinstein, Lurie, T. Tosteson.
Critical revision of the manuscript for important intellectual content: Weinstein, Lurie, T. Tosteson, Skinner, Hanscom, A. Tosteson, Herkowitz, Fischgrund, Cammisa, Albert, Deyo.
Statistical analysis: Weinstein, T. Tosteson, Skinner, Hanscom, A. Tosteson.
Obtained funding: Weinstein, T. Tosteson, A. Tosteson.
Administrative, technical, or material support: Weinstein, Lurie, T. Tosteson.
Study supervision: Weinstein, Lurie.
Financial Disclosures: Dr Weinstein reports that he is Editor-in-Chief of *Spine,* has been a consultant to United Health Care (proceeds are donated to the Brie Fund, a fund for children with disabilities in the name of his daughter who passed away from leukemia), and has been a consultant to the Foundation for Informed Medical Decision-Making, proceeds to the Department of Orthopaedics, Dartmouth. Dr Lurie reports that he receives grant support from St Francis Medical Technologies and American Board of Orthopaedic Surgery; has served on advisory boards for Ortho-MacNeil Pharmaceuticals, the Robert Graham Center of the American Academy of Family Practice, Pfizer, and Centocor; and as a consultant for Myexpertdoctor.com, Pacific Business Group on Health, and the Foundation for Informed Medical Decision-Making. Mr Hanscom reports working for the National Spine Network and receiving funding from Medtronic. Dr A. Tosteson reports receiving funding from St Francis Medical Technologies. Dr Herkowitz, reports that he is the inventor of the Premier Cervical Spine Plate, serves as a consultant for Medtronics and is a board of director for Stryker Corp and Globus Medical. Dr Cammisa reports consulting for Spinal Kinetics, Healthpoint Capital, K2 Medical LLC, Mazor Surgical Technologies, Nuvasive, Orthovita Inc, Raymedica Inc, and Scient'x USA; has investments in Alphatec Holdings Inc, Healthpoint Capital Partners, Ivy Health Care, K2 Medical LLC, Nuvasive, Orthopedic Investment Partners Fund LP, Orthovita Inc, Paradigm Spine LLC, Pioneer Labs, Raymedica Inc, Small Bone Innovations Inc, and Spinal Kinetics; Dr Albert reports that he served as a consultant for DePuy Spine, serves on the advisory board of and owns stock options in Axiamed-Medical, K2 Medical, and Gentis. Dr Deyo reports that his research program benefits from a gift to the university from Symthes, a manufacturer of surgical devices, for which he receives no personal benefits. No other disclosures were reported.
Funding/Support: This study was supported by grant U01-AR45444-01A1 from NIAMS and by the NIH Office of Research on Women's Health, the National Institutes of Health, and the National Institute of Occupational Safety and Health, the Centers for Disease Control and Prevention. The Multidisciplinary Clinical Research Center in Musculoskeletal Diseases is funded by grant P60-AR048094-01A1 from NIAMS. Dr Lurie is supported by a Research Career Award from NIAMS (1 K23 AR 048138-01).
Role of the Sponsors: These sources had no role in design or conduct of the study; collection, management, analysis, or interpretation of the data; or in preparation, review, or approval of the manuscript.
Principal Investigators, Participating Institutions, Project Coordinator, and Nurse Coordinators (in order by enrollment): Jeffrey Fischgrund and Sig Berven, MD (University of California, San Francisco); Judi L. Forman, MPH (Project Coordinator), William A. Abdu, MD, MS, Barbara Butler-Schmidt, RN, MSN, J. J. Hebb, RN, BSN (Dartmouth-Hitchcock Medical Center); Harry Herkowitz, MD, Gloria Bradley, BSN, Melissa Lurie, RN (William Beaumont Hospital); Todd Albert, MD, Alan Hilibrand, MD, Carole Simon, RN, MS (Rothman Institute at Thomas Jefferson Hospital); Frank Cammisa, MD, Brenda Green, RN, BSN (Hospital for Special Surgery); Michael Longley, MD, Nancy Fullmer, RN, Ann Marie Fredericks, RN, MSN, CPNP (Nebraska Foundation for Spinal Research); Scott Boden, MD, Sally Lashley, BSN, MSA (Emory University–The Emory Clinic); Lawrence Lenke, MD, Georgia Stobbs, RN, BA (Washington University); Sanford Emery, MD, Chris Furey, MD, Kathy Higgins, RN, PhD, CNS (University Hospitals of Cleveland/Case Western Reserve); Thomas Errico, MD, Alex Lee, RN, BSN (Hospital for Joint Diseases); Harley Goldberg, DO, Pat Malone, RN, MSN, ANP (Kaiser Permanente); Serena Hu, MD, Pat Malone, RN, MSN, ANP (University of California–San Francisco); Gunnar Andersson, MD, PhD, Howard An, MD, Margaret Hickey, RN, MS (Rush-Presbyterian-St Luke's Medical Center); Robert Keller, MD (Maine Spine and Rehabilitation).
Enrolling Physicians: William Abdu (Dartmouth-Hitchcock Medical Center); David Montgomery, Harry Herkowitz (William Beaumont Hospital); Ted Conliffe, Alan Hilibrand (Rothman Institute at Thomas Jefferson Hospital); Perry Ball (Dartmouth); Frank Cammisa (Hospital for Special Surgery); S. Tim Yoon (Emory University–The Emory Clinic); Randall Woodward (Nebraska Foundation for Spinal Research); Brett Taylor (Washington University); Todd Albert (Rothman); Richard Schoenfeldt (Hospital for Joint Diseases); Jonathan Fuller (Nebraska Foundation for Spinal Research); Harvinder Sandhu (Hospital for Special Surgery); Scott Boden (Emory); Carolyn Murray (Dartmouth); Michael Longley (Nebraska Foundation for Spinal Research); Ronald Moskovich (Hospital for Joint Diseases); Keith Bridwell (Washington University); John McClellan (Nebraska Foundation for Spinal Research);

Lawrence Lenke (Washington University); Ferdy Massimino (Kaiser Permanente); Lawrence Kurz (Beaumont); Joseph Dryer (Hospital for Joint Diseases); Sanford Emery (University Hospitals of Cleveland/Case Western Reserve); Susan Dreyer, Howard Levy (Emory); Patrick Bowman (Nebraska Foundation for Spinal Research); Thomas Errico (Hospital for Joint Diseases); Lee Thibodeau, MD (Maine Spine and Rehabilitation); Jeffrey Fischgrund (Beaumont); Mark Splaine (Dartmouth); John Bendo (Hospital for Joint Diseases); Taylor Smith (University of California–San Francisco); Eric Phillips (Nebraska Foundation for Spinal Research); Dilip Sengupta (Dartmouth); David Hubbell (Emory); Henry Schmidek (Dartmouth); Harley Goldberg (Kaiser); Robert Rose (Dartmouth); Sig Berven (University of California–San Francisco); Frank Phillips, Howard An (Rush-Presbyterian-St Luke's Medical Center); Colleen Olson (Dartmouth); Anthony Margherita, John Metzler (Washington University); Jeffrey Goldstein (Hospital for Joint Diseases); Phaedra Mcdonough (Dartmouth); James Farmer (Hospital for Special Surgery); Marsolais (Case Western); Gunnar Andersson (Rush-Presbyterian-St Luke's); Hilda Magnadottir, Jim Weinstein, Jon Lurie (Dartmouth); J. X. Yoo (Case Western); John Heller (Emory); Jeffrey Spivak (Hospital for Joint Diseases); Roland Hazard (Dartmouth); Michael Schaufele (Emory); Jeffrey Florman (Maine Spine and Rehabilitation); Philip Bernini (Dartmouth); Eeric Truumees (Beaumont); K. Daniel Riew (Washington University); Timothy Burd (Nebraska Foundation for Spinal Research); John Rhee (Emory); Henry Bohlman (Case Western); Richard Perry (Hospital for Joint Diseases); Edward Goldberg (Rush-Presbyterian-St Luke's); Christopher Furey (Case Western).

Acknowledgment: Thanks to Tamara S. Morgan, Department of Orthopaedic Surgery, Dartmouth Medical School, for graphic design and production, journal interfacing and for shepherding the SPORT study from its original submission and into the foreseeable future. She is funded through the department and partially by SPORT. This study is dedicated to the memory of Brieanna Weinstein.

REFERENCES

1. Atlas SJ, Deyo RA, Keller RB, et al. The Maine Lumbar Spine Study, II: 1-year outcomes of surgical and nonsurgical management of sciatica. *Spine*. 1996;21:1777-1786.

2. 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.

3. Weber H. Lumbar disc herniation: a controlled, prospective study with ten years of observation. *Spine*. 1983;8:131-140.

4. McCulloch P, Taylor I, Sasako M, Lovett B, Griffin D. Randomised trials in surgery: problems and possible solutions. *BMJ*. 2002;324:1448-1451.

5. Birkmeyer NJ, Weinstein JN, Tosteson AN, et al. Design of the Spine Patient Outcomes Research Trial (SPORT). *Spine*. 2002;27:1361-1372.

6. Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ*. 1996;312:1215-1218.

7. McKee M, Britton A, Black N, McPherson K, Sanderson C, Bain C. Interpreting the evidence: choosing between randomised and non-randomised studies. *BMJ*. 1999;319:312-315.

8. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med*. 2000;342:1887-1892.

9. Pocock SJ, Elbourne DR. Randomized trials or observational tribulations? *N Engl J Med*. 2000;342:1907-1909.

10. Delamarter R, McCullough J. Microdiscectomy & microsurgical laminotomies. In: Frymoyer J, ed. *The Adult Spine: Principles and Practice*. 2nd ed. Philadelphia, Pa: Lippincott-Raven Publishers; 1996.

11. McHorney CA, Ware JE Jr, Lu JF, Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36), III: tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Med Care*. 1994;32:40-66.

12. 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.

13. Deyo RA, Diehl AK. Patient satisfaction with medical care for low-back pain. *Spine*. 1986;11:28-30.

14. 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.

15. Fitzmaurice G, Laird N, Ware J. *Applied Longitudinal Analysis*. Philadelphia, Pa: John Wiley & Sons; 2004.

16. Weinstein JN, Tosteson TD, Lurie JD, et al. Surgical vs nonoperative treatment for lumbar disk herniation: the Spine Patient Outcomes Research Trial (SPORT): a randomized trial. *JAMA*. 2006;296:2441-2450.

17. Fardon DF, Milette PC. Nomenclature and 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.

18. Beaton DE. Understanding the relevance of measured change through studies of responsiveness. *Spine*. 2000;25:3192-3199.

19. Beaton DE, Tarasuk V, Katz JN, Wright JG, Bombardier C. "Are you better?" a qualitative study of the meaning of recovery. *Arthritis Rheum*. 2001;45:270-279.

20. Hagg O, Fritzell P, Nordwall A. The clinical importance of changes in outcome scores after treatment for chronic low back pain. *Eur Spine J*. 2003;12:12-20.

21. Mannion AF, Junge A, Grob D, Dvorak J, Fairbank JC. Development of a German version of the Oswestry Disability Index, II: sensitivity to change after spinal surgery. *Eur Spine J*. 2006;15:66-73.

22. Arega A, Birkmeyer NJ, Lurie JD, et al. Racial variation in treatment preferences and willingness to randomize in the Spine Patient Outcomes Research Trial (SPORT). *Spine*. 2006;31:2263-2269.

23. AHCPR. *Acute Low Back Problems in Adults*. Vol 14. Bethesda, Md: US Dept of Health and Human Services; 1994.