SPINE Volume 33, Number 25, pp 2789–2800

Surgical *Versus* Nonoperative Treatment for Lumbar Disc Herniation

Four-Year Results for the Spine Patient Outcomes Research Trial (SPORT)

James N. Weinstein, DO, MS,* Jon D. Lurie, MD, MS,* Tor D. Tosteson, ScD,* Anna N. A. Tosteson, ScD,* Emily A. Blood, MS,* William A. Abdu, MD,* Harry Herkowitz, MD,† Alan Hilibrand, MD,‡ Todd Albert, MD,‡ and Jeffrey Fischgrund, MD†

Study Design. Concurrent, prospective, randomized, and observational cohort study.

Objective. To assess the 4-year outcomes of surgery *versus* nonoperative care.

Summary of Background Data. Although randomized trials have demonstrated small short-term differences in favor of surgery, long-term outcomes comparing surgical to nonoperative treatment remain controversial.

Methods. Surgical candidates with imaging-confirmed lumbar intervertebral disc herniation meeting SPORT eligibility criteria enrolled into prospective, randomized (501 participants), and observational cohorts (743 participants) at 13 spine clinics in 11 US states. Interventions were standard open discectomy *versus* usual nonoperative care. Main outcome measures were changes from baseline in the SF-36 Bodily Pain (BP) and Physical Function (PF) scales and the modified Oswestry Disability Index (ODI - AAOS/Modems version) assessed at 6 weeks, 3 months, 6 months, and annually thereafter.

Results. Nonadherence to treatment assignment caused the intent-to-treat analyses to underestimate the treatment effects. In the 4-year combined as-treated analysis, those receiving surgery demonstrated significantly greater improvement in all the primary outcome measures (mean change surgery *vs.* nonoperative; treatment effect; 95% CI): BP (45.6 *vs.* 30.7; 15.0; 11.8 to 18.1), PF (44.6 *vs.* 29.7; 14.9; 12.0 to 17.8) and ODI (−38.1 *vs.* −24.9; −13.2; −15.6 to −10.9). The percent working was similar between the surgery and nonoperative groups, 84.4% *versus* 78.4% respectively.

Conclusion. In a combined as-treated analysis at 4 years, patients who underwent surgery for a lumbar disc herniation achieved greater improvement than nonoperatively treated patients in all primary and secondary outcomes except work status.

Key words: SPORT, intervertebral disc herniation, surgery, nonoperative care, outcomes. **Spine 2008;33:2789–2800**

Lumbar disc surgery remains one of the most commonly performed operations, with rates exhibiting considerable geographic variation.[1] Two recent randomized trials demonstrated that surgery provides faster pain relief and perceived recovery in patients with herniated disc.[2–4] Outcomes were similar at 1 year for patients assigned to surgery and for those assigned to nonoperative treatment. However, both trials included substantial numbers of surgical patients in the nonoperative comparison arm due to treatment crossover, affecting the interpretation of the intent-to-treat analyses. This paper reports 4-year results for SPORT based on the continued follow-up of the herniated disc randomized and observational cohorts.

Materials and 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 previously.[3–5]

Patient Population

Men and women who had symptoms and confirmatory signs of lumbar radiculopathy that persisted for at least 6 weeks, who had disc herniation at a corresponding level and side on imaging, and who were considered surgical candidates were eligible. The content of pre-enrollment nonoperative care was not prespecified in the protocol.[3–5] Specific enrollment and exclusion criteria are reported elsewhere.[4,5]

A research nurse at each site identified potential participants and verified eligibility. Participants were offered enrollment in

From the *Dartmouth Medical School, Hanover, NH; William Beaumont Hospital, Royal Oak, MI; and Rothman Institute, Philadelphia, PA.

Acknowledgment date: July 24, 2008. First revision date: September 22, 2008. Acceptance date: September 23, 2008.
The manuscript submitted does not contain information about medical device(s)/drug(s).
Federal funds were received in support of this work. No benefits in any form have been or will be received from a commercial party related directly or indirectly to the subject of this manuscript.
This study is under annual review by the Dartmouth Committee for the Protection of Human Subjects (CPHS #17083).
Funding: The National Institute of Arthritis and Musculoskeletal and Skin Diseases (U01-AR45444-01A1) and the 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 NIAMS (P60-AR048094-01A1). Dr. Lurie is supported by a Research Career Award from NIAMS (1 K23 AR 048138-01).
Trial Registration: Spine Patient Outcomes Research Trial (SPORT): Intervertebral Disc Herniation; #NCT00000410; http://www.clinicaltrials.gov/ct/show/NCT00000410?order=2.
Address correspondence to James N. Weinstein, DO, MS, Director, The Dartmouth Institute for Health Policy and Clinical Practice, Dartmouth College, Professor & Chair, Department of Orthopaedics, Dartmouth Medical School, One Medical Center Dr., Lebanon, NH 03756, 603-653-3580, 603-653-3581; E-mail:SPORT@dartmouth.edu

either the randomized trial or the observational cohort. Enrollment began in March of 2000 and ended in November of 2004.

Study Interventions

The surgery was a standard open discectomy with examination of the involved nerve root.[5,6] The nonoperative protocol was "usual care" recommended to include at least: active physical therapy, education/counseling with home exercise instruction, and nonsteroidal anti-inflammatory drugs if tolerated. Nonoperative treatments were individualized for each patient and tracked prospectively.[3–5]

Study Measures

Primary endpoints were the Bodily Pain (BP) and Physical Function (PF) scales of the SF-36 Health Survey[7] and the AAOS/Modems version of the Oswestry Disability Index (ODI)[8] as measured at 6 weeks, 3 months, 6 months, and annually thereafter. If surgery was delayed beyond 6 weeks, additional follow-up data were obtained 6 weeks and 3 months after surgery. Secondary outcomes included patient self-reported improvement, work status, satisfaction with current symptoms and care,[9] and sciatica severity as measured by the sciatica bothersomeness index.[10,11] Treatment effect was defined as the difference in the mean changes from baseline between the surgical and nonoperative groups.

Statistical Considerations

Initial analyses compared means and proportions for baseline patient characteristics between the randomized and observational cohorts and between the initial treatment arms of the individual and combined cohorts. The extent of missing data and the percentage of patients undergoing surgery were calculated by treatment arm for each scheduled follow-up. Baseline predictors of time until surgical treatment (including treatment crossovers) in both cohorts were determined *via* a stepwise proportional hazards regression model with an inclusion criterion of $P < 0.1$ to enter and $P > 0.05$ to exit. Predictors of missing follow-up visits at yearly intervals up to 4 years were separately determined *via* stepwise logistic regression. Baseline characteristics that predicted surgery or a missed visit at any time-point were then entered into longitudinal models of primary outcomes. Those that remained significant in the longitudinal models of outcome were included as adjusting covariates in all subsequent longitudinal regression models to adjust for potential confounding due to treatment selection bias and missing data patterns.[12] In addition, baseline outcome, center, age and gender were included in all longitudinal outcome models.

Primary analyses compared surgical and nonoperative treatments using changes from baseline at each follow-up, with a mixed effects longitudinal regression model including a random individual effect to account for correlation between repeated measurements within individuals. The randomized cohort was initially analyzed on an intent-to-treat basis.[4] Because of crossover, subsequent analyses were based on treatments actually received. In these as-treated analyses, the treatment indicator was a time-varying covariate, allowing for variable times of surgery. Follow-up times were measured from enrollment for the intent-to-treat analyses, whereas for the as-treated analysis the follow-up times were measured from the beginning of treatment (*i.e.*, the time of surgery for the surgical group and the time of enrollment for the nonoperative group), and baseline covariates were updated to the follow-up immediately preceding the time of surgery. This procedure has the effect of including all changes from baseline before surgery in the estimates of the nonoperative treatment effect and all changes after surgery in the estimates of the surgical effect. The 6-point sciatica scales and binary outcomes were analyzed *via* longitudinal models based on generalized estimating equations[13] using the same intent-to-treat and adjusted as-treated analysis definitions as the primary outcomes. The randomized and observational cohorts were each analyzed to produce separate as-treated estimates of treatment effect. These results were compared using a Wald test to simultaneously test all follow-up visit times for differences in estimated treatment effects between the 2 cohorts.[12] Final analyses combined the cohorts.

To evaluate the 2 treatment arms across all time-periods, the time-weighted average of the outcomes (area under the curve) for each treatment group was computed using the estimates at each time period from the longitudinal regression models and compared using a Wald test.[12]

Kaplan-Meier estimates of reoperation rates at 4 years were computed for the randomized and observational cohorts and compared *via* the log-rank test.[14,15] Crossover from assigned surgery to nonoperative treatment and from assigned nonoperative treatment to surgery was compared *via* McNemar's test.[16]

Computations were done using SAS procedures PROC MIXED for continuous data and PROC GENMOD for binary and non-normal secondary outcomes (SAS version 9.1 Windows XP Pro, Cary, NC). Statistical significance was defined as $P < 0.05$ based on a 2-sided hypothesis test with no adjustments made for multiple comparisons. Data for these analyses were collected through April 10, 2008.

Results

Overall, 1244 SPORT participants with lumbar intervertebral disc herniation were enrolled (501 in the randomized cohort, and 743 in the observational cohort) (Figure 1). In the randomized cohort, 245 were assigned to surgical treatment and 256 to nonoperative treatment. Of those randomized to surgery, 57% had surgery by 1 year and 59% by 4 years. In the group randomized to nonoperative care, 41% of patients received surgery by 1 year and 45% received surgery by 4 years. In the observational cohort, 521 patients initially chose surgery and 222 patients initially chose nonoperative care. Of those initially choosing surgery, 95% received surgery by 1 year; at 4 years no further surgeries had been reported. Of those choosing nonoperative treatment, 20% had surgery by 1 year, 24% by 4 years. In both cohorts combined, 805 patients received surgery at some point during the first 4 years; 439 (35%) remained nonoperative. Over the 4 years, 1192 (96%) of the original enrollees completed at least 1 follow-up visit and were included in the analysis (randomized cohort: 94% and observational cohort 97%); between 65% and 87% of enrollees supplied data at each follow-up interval with losses due to dropouts, missed visits, or deaths (Figure 1).

Patient Characteristics

Baseline characteristics are compared in Table 1. Overall, the cohorts were similar. However, patients in the observational cohort had more disability, a strong preference for surgery, more often rated their problem as

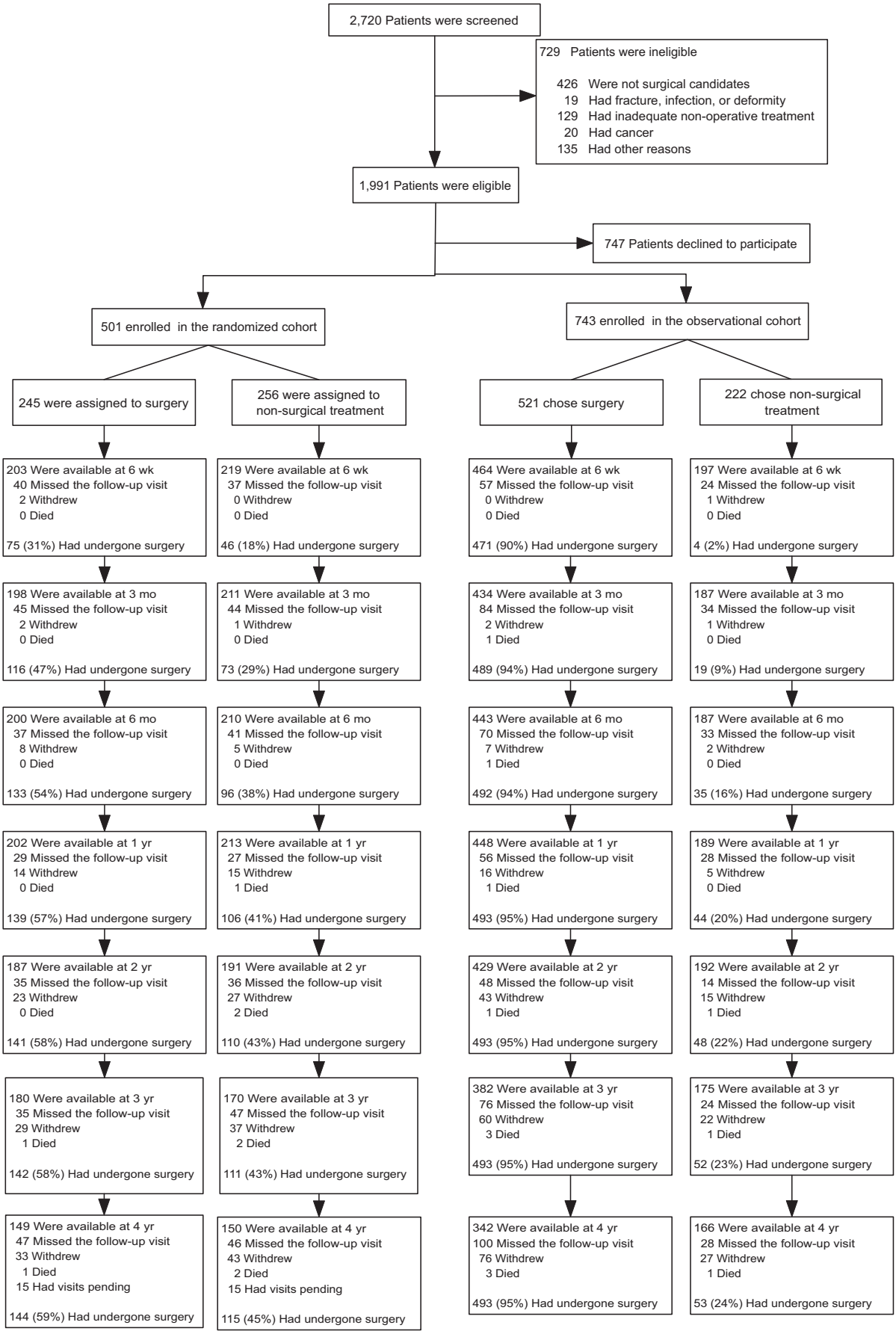

Figure 1. Exclusion, enrollment, randomization, and follow-up of trial participants. The values for surgery, withdrawal, and death are cumulative over 4 years. For example, a total of 1 patient in the group assigned to surgery died during the 4-year follow-up period. [Data set 04/10/2008].

Table 1. Patient Baseline Demographic Characteristics, Comorbidities, and Health Status Measures According to Study Cohort and Treatment Received

	SPORT Study Cohorts			Randomized and Observational Cohorts Combined: Treatment Received*		
	Randomized Cohort (n = 473)	Observational Cohort (n = 719)		Surgery (n = 788)	Non-Operative (n = 404)	
Mean age (stdev)	42.3 (11.6)	41.4 (11.2)	0.16	40.7 (10.8)	43.9 (12.2)	<0.001
Female	194 (41%)	313 (44%)	0.42	340 (43%)	167 (41%)	0.59
Ethnicity: Not Hispanic†	449 (95%)	688 (96%)	0.64	752 (95%)	385 (95%)	0.97
Race-White†	400 (85%)	633 (88%)	0.10	694 (88%)	339 (84%)	0.056
Education-At least some college	356 (75%)	528 (73%)	0.52	571 (72%)	313 (77%)	0.072
Income-Under $50,000	207 (44%)	328 (46%)	0.57	367 (47%)	168 (42%)	0.11
Marital status-Married	333 (70%)	502 (70%)	0.88	549 (70%)	286 (71%)	0.74
Work status			0.70			0.004
Full or part time	291 (62%)	431 (60%)		455 (58%)	267 (66%)	
Disabled	58 (12%)	100 (14%)		121 (15%)	37 (9%)	
Other	124 (26%)	187 (26%)		211 (27%)	100 (25%)	
Compensation-Any‡	76 (16%)	132 (18%)	0.35	161 (20%)	47 (12%)	<0.001
Mean Body Mass Index (BMI), (stdev)§	28 (5.5)	27.9 (5.6)	0.87	28.2 (5.7)	27.5 (5.3)	0.037
Smoker	108 (23%)	174 (24%)	0.64	197 (25%)	85 (21%)	0.15
Comorbidities						
Depression	62 (13%)	79 (11%)	0.31	93 (12%)	48 (12%)	0.96
Joint problem	97 (21%)	124 (17%)	0.18	127 (16%)	94 (23%)	0.003
Other¶	221 (47%)	305 (42%)	0.16	328 (42%)	198 (49%)	0.018
Time since recent episode <6 months	373 (79%)	557 (77%)	0.62	609 (77%)	321 (79%)	0.43
SF-36 Bodily Pain (BP) Score‖	26.9 (17.9)	25.2 (18.3)	0.11	22.3 (16.2)	32.9 (19.7)	<0.001
SF-36 Physical Function (PF) score‖	39.5 (25.3)	36.6 (25.6)	0.056	32.3 (23.4)	48.2 (26.3)	<0.001
Mental Component Summary (MCS) score	45.9 (12)	44.7 (11.2)	0.086	44.6 (11.4)	46.3 (11.8)	0.021
Oswestry Disability Index (ODI)**	46.9 (21)	51.2 (21.4)	<0.001	54.9 (19.6)	38.8 (20.4)	<0.001
Sciatica Frequency Index (0–24)††	15.6 (5.5)	16 (5.3)	0.18	16.7 (5.1)	14.3 (5.6)	<0.001
Sciatica Bothersome Index (0–24)‡‡	15.2 (5.2)	15.8 (5.3)	0.057	16.4 (4.9)	13.9 (5.6)	<0.001
Satisfaction with symptoms-very dissatisfied	370 (78%)	584 (81%)	0.23	696 (88%)	258 (64%)	<0.001
Problem getting better or worse			<0.001			<0.001
Getting better	90 (19%)	89 (12%)		65 (8%)	114 (28%)	
Staying about the same	221 (47%)	313 (44%)		338 (43%)	196 (49%)	
Getting worse	161 (34%)	311 (43%)		379 (48%)	93 (23%)	
Treatment preference			<0.001			<0.001
Preference for non-surg	193 (41%)	201 (28%)		120 (15%)	274 (68%)	
Not sure	154 (33%)	43 (6%)		112 (14%)	85 (21%)	
Preference for surgery	126 (27%)	472 (66%)		553 (70%)	45 (11%)	
Pain radiation	458 (97%)	704 (98%)	0.33	772 (98%)	390 (97%)	0.19
Straight Leg Raise Test-Ipsilateral	291 (62%)	459 (64%)	0.45	515 (65%)	235 (58%)	0.018
Straight Leg Raise Test-Contralateral/Both	67 (14%)	121 (17%)	0.25	149 (19%)	39 (10%)	<0.001
Any Neurological Deficit	351 (74%)	551 (77%)	0.38	617 (78%)	285 (71%)	0.004
Reflexes-Asymmetric Depressed	203 (43%)	278 (39%)	0.16	325 (41%)	156 (39%)	0.42
Sensory-Asymmetric Decrease	222 (47%)	381 (53%)	0.047	429 (54%)	174 (43%)	<0.001
Motor-Asymmetric Weakness	190 (40%)	311 (43%)	0.32	354 (45%)	147 (36%)	0.006
Herniation level			0.087			<0.001
L2–L3/L3–L4	32 (7%)	56 (8%)		42 (5%)	46 (11%)	
L4–L5	165 (35%)	291 (40%)		305 (39%)	151 (37%)	
L5–S1	275 (58%)	372 (52%)		441 (56%)	206 (51%)	
Herniation type			0.85			0.30
Protruding	126 (27%)	196 (27%)		204 (26%)	118 (29%)	
Extruded	314 (66%)	469 (65%)		530 (67%)	253 (63%)	
Sequestered	32 (7%)	54 (8%)		54 (7%)	32 (8%)	
Posterolateral herniation	378 (80%)	541 (75%)	0.071	626 (79%)	293 (73%)	0.009

*Patients in the two cohorts combined were classified according to whether they received surgical treatment or only nonsurgical treatment during the first 2 years of enrollment.
†Race or ethnic group was self-assessed. Whites and blacks could be either Hispanic or non-Hispanic.
‡This category includes patients who were receiving or had applications pending for workers compensation, Social Security compensation, or other compensation.
§The body-mass index is the weight in kilograms divided by the square of the height in meters.
¶Other = problems related to stroke, diabetes, osteoporosis, cancer, fibromyalgia, CFS, PTSD, alcohol, drug dependence, heart, lung, liver, kidney, blood vessel, nervous system, hypertension, migraine, anxiety, stomach or bowel.
‖The SF-36 scores range from 0 to 100, with higher score indicating less severe symptoms.
**The Oswestry Disability Index ranges from 0 to 100, with lower scores indicating less severe symptoms.
††The Sciatica Frequency Index ranges from 0 to 24, with lower scores indicating less severe symptoms.
‡‡The Sciatica Bothersomeness Index ranges from 0 to 24, with lower scores indicating less severe symptoms.

worsening, and were slightly more likely to have a sensory deficit.

Summary statistics for the combined cohorts are also displayed in Table 1 according to treatment received. The study population had an overall mean age of 41.7, with a mean of 40.7 in the surgery group and a mean of 43.9 in the nonoperative group. There were slightly more men than women. Subjects receiving surgery were younger, less likely to be working, more likely to report being disabled and to be receiving compensation, had slightly greater BMI's, fewer joint and other comorbidities, more pain, frequent and bothersome sciatica, depression, dissatisfaction with their symptoms and more often rated them as getting worse, less function, and were more likely to prefer surgery. Subjects receiving surgery also had more ipsilateral and contralateral straight leg tests, and more neurologic, sensory, and motor deficits. Radiographically, their herniations were more likely to be at the L4–L5 and L5–S1 levels and to be posterolateral in location.

Nonoperative Treatments

Nonoperative treatments within 4 years of enrollment were similar between the 2 cohorts. However, more observational patients reported visits to other practitioners (57% observational *vs.* 37% randomized, $P = < 0.001$); and randomized patients had more (randomized *vs.* observational): injections (57% *vs.* 40%, $P = < 0.001$), activity restriction (32% *vs.* 20% $P = 0.004$) and narcotics (50% *vs.* 37% $P = 0.005$).

Surgical Treatment and Complications

Overall surgical treatment and complications were similar between the 2 cohorts (Table 2). The average surgical time was slightly longer in the randomized cohort (80.6 minutes randomized *vs.* 74.9 minutes observational, $P = 0.049$). Median (interquartile range) values for surgical time were 74.5 minutes (57.8, 90.0) for the randomized and 70 minutes (50.0, 90.0) for the observational cohort. The average blood loss was 67.5cc in the randomized cohort *versus* 63.0cc in the observational, $P = 0.56$. Median (25th percentile, 75th percentile) for blood loss was 50cc (25th percentile, 75th percentile) in the randomized cohort and 50cc (25th percentile, 50th percentile) in the observational. Only 6 patients total required intraoperative transfusions. There were no perioperative mortalities. The most common surgical complication was dural tear (3% of cases). Reoperation occurred in a combined 6% of cases by 1 year, 8% at 2 years, 9% at 3 years, and 10% at 4 years post surgery. The rates of reoperation were not significantly different between the randomized and observational cohorts. Seventy-five of the 81 reoperations noted the type of reoperation; approximately 50% of these (48/75) were listed as recurrent herniations at the same level. One death occurred within 90 days postsurgery related to heart surgery at another institution; the death was judged to be unrelated and was reported to the Institutional Review Board and the Data and Safety Monitoring Board.

Table 2. Operative Treatments, Complications and Events

	Randomized Cohort (n = 253)	Observational Cohort (n = 545)	*P*-Value
Discectomy Level			
L2–L3	2 (1%)	12 (2%)	0.27
L3–L4	7 (3%)	20 (4%)	0.69
L4–L5	97 (39%)	215 (40%)	0.98
L5–S1	149 (60%)	305 (56%)	0.32
Median time to surgery in months (95% CI)†	6.8 (4.5, 40.9)	0.5 (0.4, 0.7)	<0.001
Operation time, minutes (SD)	80.6 (41)	74.9 (35.5)	0.049
Blood loss, cc (SD)	67.5 (99.1)	63 (103.2)	0.56
Blood replacement			
Intraoperative replacement	4 (2%)	2 (0%)	0.16
Post-operative transfusion	0 (0%)	0 (0%)	—
Length of stay (SD)	1 (1.1)	0.94 (0.9)	0.26
Post-operative mortality (death within 6 weeks of surgery)	0 (0%)	0 (0%)	
Post-operative mortality (death within 3 months of surgery)	0 (0%)	1‡ (0.2%)	0.70
Intraoperative complications§			
Dural tear/spinal fluid leak	10 (4%)	14 (3%)	0.40
Nerve root injury	1 (0%)	1 (0%)	0.84
Vascular injury	1 (0%)	0 (0%)	0.69
Other	2 (1%)	1 (0%)	0.50
None	239 (94%)	530 (97%)	0.08
Postoperative complications/ events¶			
Nerve root injury	0 (0%)	1 (0%)	0.69
Wound hematoma	0 (0%)	4 (1%)	0.41
Wound infection	4 (2%)	14 (3%)	0.55
Other	9 (4%)	18 (3%)	1
None	237 (95%)	510 (94%)	0.67
Additional surgeries (1-year rate)‖	10 (4%)	36 (7%)	0.13
Additional surgeries (2-year rate)‖	14 (5%)	49 (9%)	0.082
Additional surgeries (3-year rate)‖	18 (7%)	52 (10%)	0.23
Additional surgeries (4-year rate)‖	22 (9%)	59 (11%)	0.32
Recurrent disc herniation	13 (5%)	36 (7%)	
Complication or Other	5 (2%)	16 (3%)	
New condition	2 (1%)	7 (1%)	

*259 RCT and 546 OBS patients had surgery, surgical information was available for 253 RCT patients and 545 observational patients.
‡Patient died after heart surgery at another hospital, the death was judged unrelated to spine surgery.
§None of the following were reported: aspiration, operation at wrong level, vascular injury.
¶Any reported complications up to 8 weeks post operation. None of the following were reported: bone graft complication, CSF leak, paralysis, cauda equina injury, wound dehiscence, pseudarthrosis.
‖One-, two-, three- and four-year post-surgical re-operation rates are Kaplan Meier estimates and p-values are based on the log-rank test. Numbers and percentages are based on the first additional surgery if more than one additional surgery.
†Median and 95% CI based on Kaplan-Meier estimates and p-value based on log-rank test.

Crossover Patients

Nonadherence to treatment assignment affected both treatment arms: patients chose to delay or decline surgery in the surgical arm and crossed over to surgery in the nonoperative arm (Figure 1). Some characteristics of crossover patients were statistically different from patients who did not cross over (Table 3). Patients crossing over to nonoperative care were older, had higher in-

Table 3. Statistically Significant Predictors of Adherence to Treatment Among RCT Patients

	Assigned to Surgery			Assigned to Non-Operative		
	Treatment Received within 4 Years			Treatment Received within 4 Years		
	Surgery (n = 143)	Non-Operative (n = 89)	*P*-Value	Surgery (n = 112)	Non-Operative (n = 129)	*P*-Value
Mean age (stdev)	40.2 (11)	44.1 (12.7)	0.01	41.9 (10)	43.8 (12.3)	0.21
Income-Under $50,000	66 (46%)	27 (30%)	0.02	61 (54%)	53 (41%)	0.052
Straight Leg Raise Test-Ipsilateral	96 (67%)	47 (53%)	0.04	74 (66%)	74 (57%)	0.21
Herniation level			0.008			0.35
L2–L3/L3–L4	4 (3%)	12 (13%)		5 (4%)	11 (9%)	
L4–L5	52 (36%)	28 (31%)		43 (38%)	42 (33%)	
L5-S1	87 (61%)	49 (55%)		64 (57%)	75 (58%)	
Bodily Pain (BP) score	24.3 (16.7)	31.6 (20.4)	0.003	24.6 (16.8)	28.6 (17.7)	0.072
Physical Functioning (PF) score	36.2 (24.1)	45.4 (25.3)	0.006	33.8 (23.2)	43.9 (26.8)	0.002
Oswestry (ODI)	51.2 (21)	41.5 (20.8)	<0.001	51.5 (19.5)	41.7 (20.5)	<0.001
Sciatica Frequency Index (0–24)	16.3 (5.2)	15.1 (6)	0.11	16.4 (5.5)	14.6 (5.4)	0.012
Sciatica Bothersome Index (0–24)	15.9 (4.8)	14.7 (5.6)	0.09	16 (5)	14.1 (5.3)	0.004
Satisfaction with symptoms-very dissatisfied	126 (88%)	58 (65%)	<0.001	96 (86%)	90 (70%)	0.005
Problem getting worse	60 (42%)	22 (25%)	0.01	45 (40%)	34 (26%)	0.032
Treatment preference			<0.001			<0.001
Preference for non-surg	45 (31%)	51 (57%)		29 (26%)	68 (53%)	
Not sure	48 (34%)	31 (35%)		36 (32%)	39 (30%)	
Preference for surgery	50 (35%)	7 (8%)		47 (42%)	22 (17%)	

*The SF-36 scores range from 0 to 100, with higher score indicating less severe symptoms.
†The Oswestry Disability Index ranges from 0 to 100, with lower scores indicating less severe symptoms.
‡The Sciatica Frequency Index ranges from 0 to 24, with lower scores indicating less severe symptoms.
§The Sciatica Bothersomeness Index ranges from 0 to 24, with lower scores indicating less severe symptoms.

comes and less pain and disability, were less likely to have an ipsilateral straight leg raise and to perceive their symptoms as getting worse, more likely to have a high level disc herniation and to express a baseline preference for nonoperative care, and were less dissatisfied with their symptoms. Patients crossing over to surgery within 4 years had lower income, worse physical function and more self-rated disability, were more dissatisfied with their symptoms, perceived they were getting worse and expressed a baseline preference for surgery. While more patients crossed from nonoperative treatment to surgery [112 (24%)] than crossed from surgery to nonoperative treatment [89 (19%)], this difference is not significant based on a McNemar's test ($P = 0.12$).

Main Treatment Effects

Intent-to-Treat Analysis. In the intent-to-treat analysis of the randomized cohort, all measures over 4 years favored surgery, but there were no statistically significant treatment effects in any of the primary outcome measures at any time interval (Table 4 and Figure 2). The secondary outcomes (sciatica bothersomeness index and self-rated improvement) were statistically significant in favor of surgery in the intent-to-treat analysis at 1 year[4]; significance was maintained out to 4 years only for the sciatica bothersomeness index (Table 4 and Figure 3).

As-Treated Analysis. The global hypothesis test (not shown) comparing the as-treated treatment effects between the randomized and observational cohorts over all time periods showed no difference between the randomized and observational cohorts ($P = 0.44$ for SF-36 BP; $P = 0.76$ for SF-36 PF; and $P = 0.90$ for the ODI). Treatment effects for the primary outcomes in the combined as-treated analysis were significant at 2 years and maintained out to 4 years: SF-36 BP 15.0 $P < 0.001$ (95% CI: 11.8 to 18.1); SF-36 PF 14.9 $P < 0.001$ (95% CI: 12.0 to 17.8); ODI −13.2 $P < 0.001$ (95% CI: −15.6 to −10.9) (Table 4). The footnote for Table 4 describes the controlling covariates for the final model.

Results from the intent-to-treat and as-treated analyses of the 2 cohorts are compared in Figure 2. The as-treated treatment effects significantly favored surgery in both cohorts. In the combined analysis, treatment effects were statistically significant in favor of surgery for all primary and secondary outcome measures (with the exception of work status) at each time point (Table 4 and Figure 3).

The treatment effects for the secondary measures of sciatica bothersomeness, satisfaction, and self-rated improvement narrowed between 3 months and 2 years but remained significant at all periods. Work status was significantly worse in the surgery group at 3 months due to surgery patients recovering from surgery; work status thereafter showed a small but nonsignificant benefit for surgery. At 4 years, the adjusted percentage of patients working was 84.4% surgical *versus* 78.4% nonoperative, treatment effect 6.0 (95% CI: −0.9, 12.9) (Table 4 and Figure 3).

Table 4. Primary Analysis Results for Years 3 and 4. Intent-to-treat for the Randomized Cohort and Adjusted* Analyses According to Treatment Received for the Randomized and Observational Cohorts Combined‡‡

		2-Year			3-Year			4-Year		
		Mean Change (SE) or Percent			Mean Change (SE) or Percent			Mean Change (SE) or Percent		
	Baseline Overall Mean	Surgery	Non-Operative	Treatment Effect (95% CI)†	Surgery	Non-Operative	Treatment Effect (95% CI)†	Surgery	Non-Operative	Treatment Effect (95% CI)†
RCT Intent-to-treat										
Primary outcomes		(n = 187)	(n = 191)		(n = 180)	(n = 170)		(n = 149)	(n = 150)	
SF-36 Bodily Pain (BP) (0–100) (SE)‡	26.9 (0.82)	40.5 (1.9)	37.5 (1.9)	3.1 (−2.2, 8.4)	39.6 (2)	36.2 (2)	3.4 (−2.1, 8.9)	41.3 (2.1)	36.8 (2.1)	4.5 (−1.2, 10.3)
SF-36 Physical Function (PF) (0–100) (SE)‡	39.5 (1.2)	36.2 (2)	35.7 (2)	0.5 (−4.9, 6)	37.2 (2)	34.1 (2)	3.1 (−2.5, 8.8)	36.6 (2.1)	34.4 (2.1)	2.2 (−3.7, 8)
Oswestry Disability Index (ODI) (0–100) (SE)§	46.9 (0.97)	−31.5 (1.7)	−28.8 (1.7)	−2.8 (−7.5, 1.9)	−31.5 (1.7)	−27.2 (1.7)	−4.3 (−9.1, 0.6)	−31.2 (1.8)	−27.6 (1.8)	−3.6 (−8.6, 1.4)
Secondary outcomes										
Sciatica Bothersomeness Index (0–24) (SE)¶	15.2 (0.24)	−10.2 (0.47)	−8.6 (0.47)	−1.6 (−2.9, −0.3)	−10.2 (0.48)	−8.4 (0.49)	−1.8 (−3.2, −0.5)	−10.5 (0.51)	−8.8 (0.5)	−1.8 (−3.2, −0.4)
Leg pain (0–6) (SE)‖	4.6 (0.1)	−3.3 (0.2)	−2.8 (0.2)	−0.5 (−0.9, −0.1)	−3.4 (0.2)	−2.8 (0.2)	−0.6 (−1.1, −0.2)	−3.5 (0.2)	−2.9 (0.2)	−0.6 (−1.1, −0.2)
Low back pain bothersomeness (0–6) (SE)**	3.9 (0.1)	−1.9 (0.2)	−1.8 (0.2)	−0.1 (−0.6, 0.3)	−1.9 (0.2)	−1.6 (0.2)	−0.3 (−0.7, 0.2)	−1.8 (0.2)	−1.7 (0.2)	−0.1 (−0.6, 0.3)
Very/somewhat satisfied w/symptoms (%)	3.4 (1.8)	68.3	64.9	3.3 (−6.3, 13)	72.9	62.4	10.5 (0.5, 20.6)	64.7	61.3	3.4 (−7.7, 14.6)
Very/somewhat satisfied w/care (%)		87	84.5	2.4 (−4.8, 9.6)	85.2	85	0.2 (−7.4, 7.8)	89.3	81.2	8.1 (−0.1, 16.2)
Self-rated progress: major improvement (%)		75.7	68.5	7.2 (−2, 16.3)	72.1	66.8	5.3 (−4.6, 15.2)	72.5	65	7.5 (−3.2, 18.1)
Work status: working (%)	64.3 (4.8)	75	75.5	−0.6 (−9.1, 8)	71	74.5	−3.6 (−12.7, 5.6)	71.4	75.1	−3.8 (−13.3, 5.8)
RCT/OC As-treated										
Primary outcomes		(n = 662)	(n = 344)		(n = 581)	(n = 318)		(n = 511)	(n = 283)	
SF-36 Bodily Pain (BP) (0–100) (SE)‡	26.6 (0.5)	43 (0.87)	31.6 (1.2)	11.4 (8.5, 14.2)	43.8 (0.9)	31.8 (1.2)	12 (9, 15)	45.6 (0.95)	30.7 (1.3)	15 (11.8, 18.1)
SF-36 Physical Function (PF) (0–100) (SE)‡	37.9 (0.7)	43.6 (0.82)	30.2 (1.1)	13.4 (10.7, 16)	44.1 (0.85)	30.5 (1.2)	13.6 (10.8, 16.4)	44.6 (0.89)	29.7 (1.2)	14.9 (12, 17.8)
Oswestry Disability Index (ODI) (0–100) (SE)§	49.3 (0.6)	−37.2 (0.68)	−24.6 (0.91)	−12.5 (−14.7, −10.4)	−37.5 (0.71)	−25 (0.96)	−12.5 (−14.8, −10.2)	−38.1 (0.74)	−24.9(1)	−13.2 (−15.6, −10.9)
Secondary outcomes										
Sciatica Bothersomeness Index (0–24) (SE)¶	15.5 (0.1)	−10.9 (0.21)	−8 (0.3)	−2.8 (−3.6, −2.1)	−11.1 (0.22)	−8.3 (0.32)	−2.8 (−3.6, −2)	−11.5 (0.23)	−8.2 (0.33)	−3.3 (−4.1, −2.5)
Leg pain (0–6) (SE)‖	4.7 (0)	−3.5 (0.1)	−2.7 (0.1)	−0.8 (−1, −0.5)	−3.6 (0.1)	−2.9 (0.1)	−0.7 (−1, −0.5)	−3.7 (0.1)	−2.8 (0.1)	−0.9 (−1.2, −0.6)
Low back pain bothersomeness (0–6) (SE)**	3.9 (0)	−2.1 (0.1)	−1.5 (0.1)	−0.7 (−0.9, −0.4)	−2.2 (0.1)	−1.5 (0.1)	−0.7 (−1, −0.5)	−2.2 (0.1)	−1.4 (0.1)	−0.8 (−1, −0.5)
Very/somewhat satisfied w/symptoms (%)	4.1 (2)	75	48.8	26.3 (19.4, 33.1)	74.3	49.9	24.4 (17.1, 31.6)	76.6	46.7	29.9 (22.4, 37.4)
Very/somewhat satisfied w/care (%)		91.7	79.3	12.5 (7.1, 17.8)	90.2	74.6	15.6 (9.4, 21.7)	91.7	77.3	14.4 (8.2, 20.7)
Self-rated progress: major improvement (%)		78.3	58.7	19.7 (12.8, 26.5)	75.9	53.9	22 (14.6, 29.3)	79.2	51.7	27.5 (19.9, 35)
Work status: working (%)	73.6 (4.4)	85.4	85.1	0.4 (−4.8, 5.5)	83.9	79.7	4.2 (−2.1, 10.5)	84.4	78.4	6 (−0.9, 12.9)

*Adjusted for age, gender, marital status, smoking status, race, compensation, herniation location, working status, stomach comorbidity, depression, other *** comorbidity, self-rated health trend, duration of most recent episode, treatment preference and baseline score (for SF-36 and ODI), and center.
†Treatment effect is the difference between the surgical and non-operative mean change from baseline.
‡The SF-36 scores range from 0 to 100, with higher score indicating less severe symptoms.
§The Oswestry Disability Index ranges from 0 to 100, with lower scores indicating less severe symptoms.
¶The Sciatica Bothersomeness Index ranges from 0 to 24, with lower scores indicating less severe symptoms.
‖The Leg Pain Bothersomeness Scale ranges from 0 to 6, with lower scores indicating less severe symptoms.
**The Low Back Pain Bothersomeness Scale ranges from 0 to 6, with lower scores indicating less severe symptoms.
††The sample sizes for the as-treated analyses reflect the number of patients contributing to the estimate in a given time-period using the longitudinal modeling strategy explained in the methods section, and may not correspond to the counts provided for each visit time in Figure 1.
***Other comorbidities include: stroke, diabetes, osteoporosis, cancer, fibromyalgia, cfs, PTSD, alcohol, drug dependency, heart, lung, liver, kidney, blood vessel, nervous system, hypertension, migraine, anxiety, stomach, bowel.
‡‡The estimates for 1Y and 2Y for IDH RCT ITT differ slightly from those presented in JAMA paper (reference) due to modeling differences.

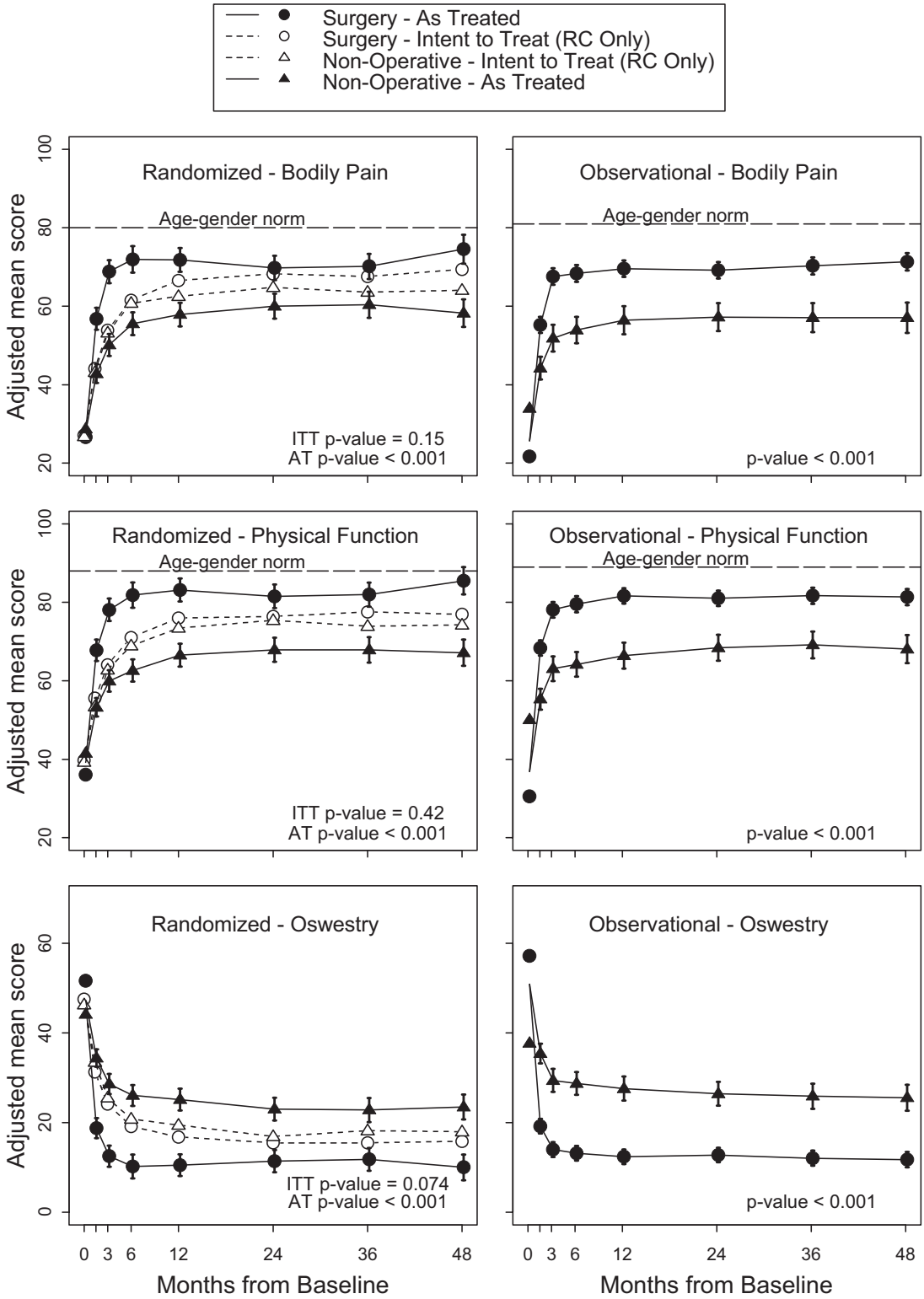

Figure 2. Primary outcomes (SF-36 Bodily Pain and Physical Function, and Oswestry Disability Index) in the randomized and observational cohorts during 2 years of follow-up. The graphs show both the intent-to-treat and the as-treated analyses for the randomized cohort (column on the left) and the as-treated analysis for the observation cohort (column on the right). The horizontal dashed line in each of the 4 SF-36 graphics represents normal values adjusted for age and sex. The I bars represent 95% confidence intervals. At 0 months, the floating data points represent the observed mean scores for each study group, whereas the data points on plot lines represent the overall means used in the adjusted analyses.

■ Discussion

In patients with a herniated disc confirmed by imaging and leg symptoms persisting for at least 6 weeks, surgery was superior to nonoperative treatment in relieving symptoms and improving function. In the as-treated analysis, the treatment effect for surgery was seen as early

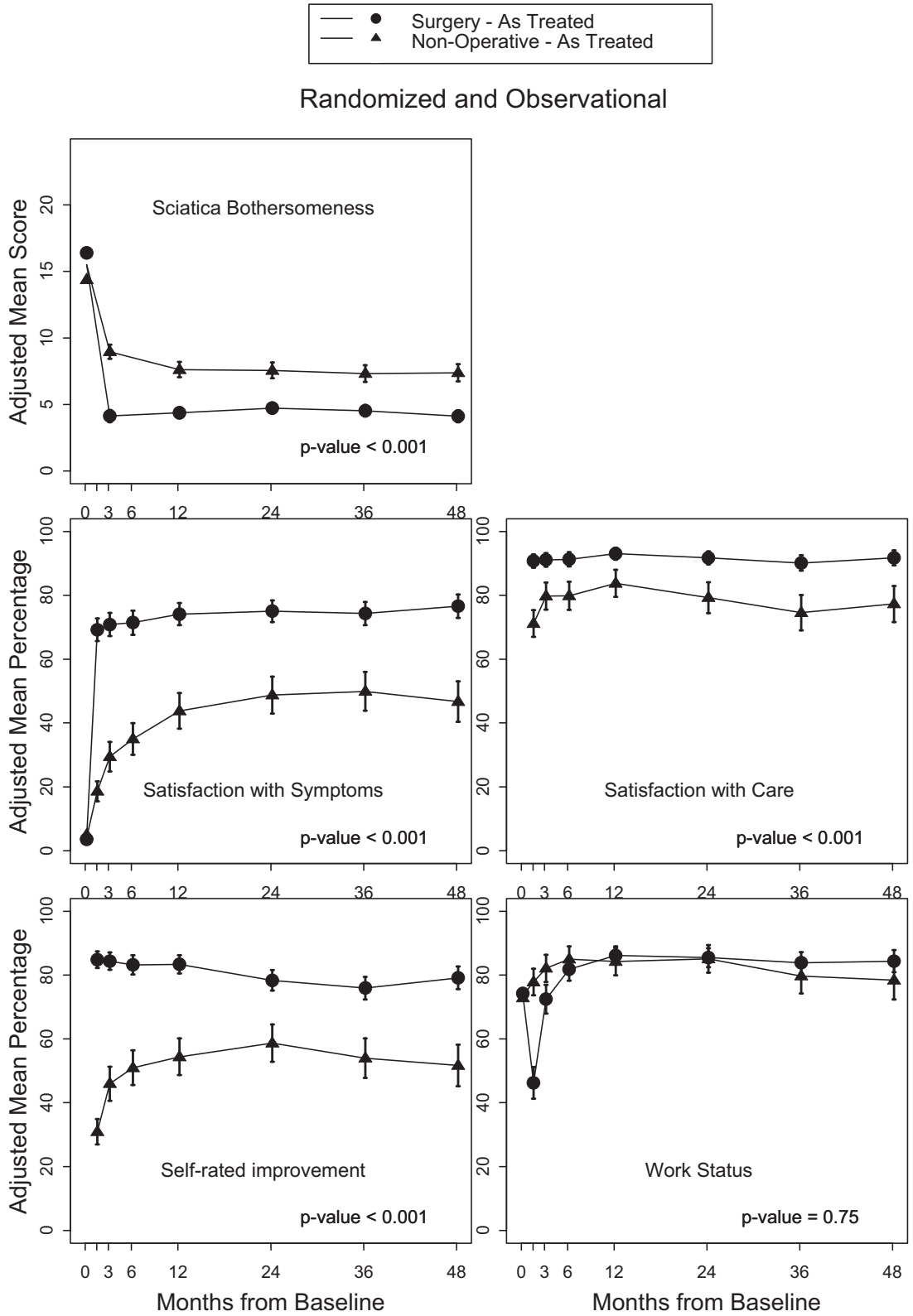

Figure 3. Secondary outcomes in the randomized and observational cohorts during 2 years of follow-up.

as 6 weeks, appeared to reach a maximum by 6 months, and persisted over 4 years. It is notable that the nonoperative group improved significantly, and this improvement persisted throughout the 4-year period. Under reasonable assumptions, this mixing of treatments due to crossover can be expected to create a bias toward the null in the intent-to-treat analyses.[2,17] The large effects seen in the as-treated analysis after adjustments for characteristics of the crossover patients suggest that the intent-to-treat analysis underestimates the true effect of surgery.

The SPORT data are supported by the only other recent randomized trial for disc herniation, Peul *et al*.[2] In that study, 39% of those randomized to nonoperative treatment crossed over to surgery at about 5 months. This is nearly identical to SPORT, in which 38% had crossed into surgery by 6 months. The estimated improvements 1 year after surgery in these 2 studies were similar (Peul *vs*. SPORT): SF-36 BP: 59.3 *versus* 43.7; SF-36 PF: 50.3 *versus* 44.4; and sciatica bothersomeness: −11.5 *versus* −11.2. In addition, the differences at 1 year between randomized groups in the intent-to-treat analyses were also quite similar: SF-36 BP: 2.7 *versus* 3.6; SF-36 PF: 2.2 *versus* 2.0; and sciatica bothersomeness: −0.4 *versus* −1.9. These results further validate the SPORT randomized cohort results but again highlight the need to also consider the as-treated analysis in this study population to estimate the true effect of surgery and to avoid bias towards the null.

SPORT Randomized Versus *Observational Cohorts*

Debate continues in the scientific literature regarding the optimal role of observational studies *versus* randomized trials.[18,19] The design of SPORT and its long-term follow-up provides an opportunity to compare over time the randomized trial results to its simultaneously enrolled observational cohort.[3–5,20,21] These 2 cohorts were remarkably similar at baseline. Patients in the observational cohort were relatively more symptomatic and functionally impaired than those in the randomized cohort; however, the absolute differences were quite small: 4.3 points on the ODI, 2.9 points on SF-36 PF, and 0.6 on the sciatica bothersomeness index. Given these similarities and our formal comparison of treatment effect between cohorts, the combined analyses are well justified.

Comparisons to Other Studies

There are no long-term studies reporting the same primary outcome measures as SPORT. As demonstrated above, the results of SPORT primary outcomes at 1 year and Peul *et al*[2] secondary measures are quite similar, but longer follow-up for the Peul study is necessary for further comparison. Unlike the suggestions made in the Weber study,[23] the differences in the outcomes between treatment groups remained relatively constant after the first year, suggesting the importance of ongoing follow-up of the patients in SPORT (Table 4, Figures 2 and 3).

The results of SPORT are similar to the Maine Lumbar Spine Study (MLSS)[22] and the Weber study.[23] The MLSS reported somewhat larger adjusted treatment effect differences at 5 years between patients who had received surgery within 3 months *versus* those that had not when compared to the SPORT 4-year data: −7.1 *versus* −3.3 (sciatica bothersomeness); −2.0 *versus* −0.9 (leg pain in the past week); and −1.2 *versus* −0.8 (low back pain in the past week). The differences are mainly related to greater improvements in the nonoperative group in SPORT *versus* MLSS; that is, the sciatica bothersomeness nonoperative mean change from baseline for SPORT was −8.2 at 4 years *versus* MLSS −4.6 at 5 years. While there are no validated outcome measures that can be directly compared between SPORT and the Weber study,[23] their 4 year results of 70% more patients in the surgical group and 51% in the conservative treatment group with "good" results is similar to SPORT's 79.2% self-rated major improvement in the surgery group and 51.7% in the nonoperative group at 4 years.

Given the increasing rates of disability and related cost for back conditions worldwide,[24] work status is thought to be an important measure of success in this population. However, return to work appears to be independent of treatment received and does not follow improvement in pain, function or satisfaction with treatment (percent working at 4 years: surgical 84.4%, non operative 78.4%; 6.0 treatment effect (95% CI: −0.9 to 12.9) (Table 4 and Figure 3E). The MLSS also showed no significant effect of surgery on return to work. From varied perspectives—payer, provider, and patient—setting expectations regarding treatment and the issue of work status must be considered in light of these results. A formal cost-effectiveness analysis of SPORT disc herniation patients with workers compensation is underway.[25] We have performed a formal cost effectiveness analysis of the combined SPORT disc herniation cohorts over 2 years.[26] It revealed that the cost per Quality-Adjusted-Life Year (QALY) gained for surgery relative to nonoperative care was $69,403 (95% CI: $49,523; $94,999), concluding that surgery is cost-effective.

Over the 4 years there was little evidence of harm from either treatment. The 4-year rate of reoperation was 10%, which is lower than the 19.4% reported by MLSS at 5 years.[22]

Limitations

Although our results are adjusted for characteristics of crossover patients and control for important baseline covariates, the as-treated analyses presented do not share the strong protection from confounding that exists for an intent-to-treat analysis.[2–4] However, the as-treated analyses yielded results similar to prior studies and to a well-designed, randomized trial by Peul *et al*.[2] Another limitation is the heterogeneity of the nonoperative treatment interventions, discussed in our prior papers.[3,4]

■ Conclusion

In the as-treated analysis combining the randomized and observational cohorts, which carefully controlled for potentially confounding baseline factors, patients treated surgically for intervertebral disc herniation showed significantly greater improvement in pain, function, satisfaction, and self-rated progress over 4 years compared to patients treated nonoperatively. It appears, despite the improvement in symptoms, that except for the first 6

weeks after surgery, work status is not related to treatment.

■ **Key Points**

- At 4-year follow-up, patients who had surgery for intervertebral disc herniation maintained greater improvement in all primary outcomes compared to those who remained nonoperative based on as-treated analyses.
- Except for work status, all secondary measures retained a significant benefit for surgery at 4 years.
- Work status showed a nonsignificant benefit for surgery at 4 years.

Acknowledgments

We thank Tamara S. Morgan, MA, Department of Orthopedic Surgery, Dartmouth Medical School, for graphic design and production, journal interfacing, and shepherding the SPORT study from its original submission and into the foreseeable future. Ms. Morgan is funded through the department, through The Dartmouth Institute, and partially by SPORT.

This study is dedicated to the memory of Brieanna Weinstein.

Contributors: Project coordinator, Judi Lowenburg Forman, MPH. Site Contributors (In order by randomized cohort enrollment): Dartmouth-Hitchcock Medical Center: William A. Abdu, Barbara Butler-Schmidt, J. J. Hebb. William Beaumont Hospital: Harry Herkowitz, Gloria Bradley, Melissa Lurie. Rothman Institute at Thomas Jefferson Hospital: Todd Albert, Allan Hilibrand, Carol Simon. Hospital for Special Surgery: Frank Cammisa, Brenda Green. Nebraska Foundation for Spinal Research: Michael Longley, Nancy Fullmer, Ann Marie Fredericks. Emory University-The Emory Clinic: Scott Boden, Sally Lashley. Washington University: Lawrence Lenke, G. A. Stobbs. University Hospitals of Cleveland/Case-Western-Reserve: Sanford Emery, Chris Furey, Kathy Higgins. Hospital for Joint Diseases: Thomas Errico, Alex Lee. Kaiser-Permanente: Harley Goldberg, Pat Malone. University of California–San Francisco: Serena Hu, Pat Malone. Rush-Presbyterian-St. Luke's: Gunnar Andersson, Margaret Hickey. Maine Spine & Rehabilitation: Robert Keller.

Site Enrollers (In order by randomized enrollment): 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 (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).

References

1. Dartmouth Atlas Working Group. *Dartmouth Atlas of Musculoskeletal Health Care*ed. Chicago, IL: American Hospital Association Press, 2000.
2. Peul WC, van Houwelingen HC, van den Hout WB, et al. Surgery versus prolonged conservative treatment for sciatica. *N Engl J Med* 2007;356: 2245–56.
3. Weinstein JN, Lurie JD, Tosteson TD, et al. Surgical vs nonoperative treatment for lumbar disk herniation: the Spine Patient Outcomes Research Trial (SPORT) observational cohort. *JAMA* 2006;296:2451–9.
4. 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;296:2441–50.
5. Birkmeyer NJ, Weinstein JN, Tosteson AN, et al. Design of the Spine Patient outcomes Research Trial (SPORT). *Spine* 2002;27:1361–72.
6. Delamarter R, McCullough J. Microdiscectomy & Microsurgical Laminotomies. In Frymoyer J ed. *The Adult Spine: Principles and Practice*. 2nd ed. Philadelphia: Lippincott-Raven Publishers, 1996.
7. McHorney CA, Ware JE Jr, Lu JF, et al. 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.
8. Daltroy LH, Cats-Baril WL, Katz JN, et al. The North American Spine Society lumbar spine outcome assessment Instrument: reliability and validity tests. *Spine* 1996;21:741–9.

9. Deyo RA, Diehl AK. Patient satisfaction with medical care for low-back pain. *Spine* 1986;11:28–30.
10. Atlas SJ, Deyo RA, Patrick DL, et al. The Quebec Task Force classification for Spinal Disorders and the severity, treatment, and outcomes of sciatica and lumbar spinal stenosis. *Spine* 1996;21:2885–92.
11. Patrick DL, Deyo RA, Atlas SJ, et al. Assessing health-related quality of life in patients with sciatica. *Spine* 1995;20:1899–908, discussion 909.
12. Fitzmaurice G, Laird N, Ware J. *Applied Longitudinal Analysis*ed. Philadelphia, PA: John Wiley & Sons, 2004.
13. Diggle PJ, Liang K-Y, Zeger SL. *Analysis of Longitudinal Data*ed. Oxford, England, UK: Oxford University Press, 1994.
14. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J of the Am Stat Assoc* 1958;53:457–81.
15. Peto R, Peto J. Asymptotically Efficient Rank Invariant Test Procedures. *Journal of the Royal Statistical Society Series a-General* 1972;135:185.
16. McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* 1949;12:153–7.
17. Meinert CL. *Clinical Trials: Design, Conduct, and Analysis*ed. New York, NY: Oxford University Press, Inc., 1986.
18. Black N. Why we need observational studies to evaluate the effectiveness of health care. *Bmj* 1996;312:1215–8.
19. McKee M, Britton A, Black N, et al. Methods in health services research. Interpreting the evidence: choosing between randomised and non-randomised studies. *Bmj* 1999;319:312–5.
20. Weinstein JN, Lurie JD, Tosteson TD, et al. Surgical versus nonsurgical treatment for lumbar degenerative spondylolisthesis. *N Engl J Med* 2007; 356:2257–70.
21. Weinstein JN, Tosteson TD, Lurie JD, et al. Surgical versus nonsurgical therapy for lumbar spinal stenosis. *N Engl J Med* 2008;358:794–810.
22. Atlas SJ, Keller RB, Chang Y, et al. Surgical and nonsurgical management of sciatica secondary to a lumbar disc herniation: five-year outcomes from the Maine Lumbar Spine Study. *Spine* 2001;26:1179–87.
23. Weber H. Lumbar disc herniation. A controlled, prospective study with ten years of observation. *Spine* 1983;8:131–40.
24. AAOS. *Burden of Musculoskeletal Disease*ed. Rosemont, IL: AAOS, 2008.
25. Atlas SJ, Tosteson ANA, Tosteson TD, et al. PODIUM: Cost-effectiveness of surgery for a lumbar disc herniation in patients with workers' compensation: Results from the Spine Patient Outcomes Research Trial (SPORT). International Society for Study of the Lumbar Spine. Geneva, Switzerland: Lippincott, Williams, and Wilkins, 2008.
26. Tosteson AN, Skinner JS, Tosteson TD, et al. The Cost Effectiveness of Surgical versus Non-Operative Treatment for Lumbar Disc Herniation over Two Years: Evidence from the Spine Patient Outcomes Research Trial (SPORT). *Spine* 2008;In Press.