Spine

SPINE Volume 39, Number 1, pp 3-16

RANDOMIZED TRIAL

Surgical *Versus* Nonoperative Treatment for Lumbar Disc Herniation

Eight-Year Results for the Spine Patient Outcomes Research Trial

Jon D. Lurie, MD, MS,* Tor D. Tosteson, ScD,* Anna N. A. Tosteson, ScD,* Wenyan Zhao, PhD,* Tamara S. Morgan, MA,* William A. Abdu, MD, MS,* Harry Herkowitz, MD,† and James N. Weinstein, DO, MS*

Study Design. Concurrent prospective randomized and observational cohort studies.

Objective. To assess the 8-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 with nonoperative treatment remain controversial.

Methods. Surgical candidates with imaging-confirmed lumbar intervertebral disc herniation meeting Spine Patient Outcomes Research Trial 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 and Physical Function scales and the modified Oswestry Disability Index-AAOS/Modems version assessed at 6 weeks, 3 months, and 6 months, and annually thereafter.

Results. Advantages were seen for surgery in intent-to-treat analyses for the randomized cohort for all primary and secondary outcomes other than work status; however, with extensive nonadherence to treatment assignment (49% patients assigned to nonoperative therapy receiving surgery *versus* 60% of patients assigned to surgery) these observed effects were relatively small and not statistically significant for primary outcomes (bodily pain, physical function, Oswestry Disability Index). Importantly, the overall comparison of secondary outcomes was significantly greater with surgery in the intent-to-treat analysis (sciatica bothersomeness [$P > 0.005$], satisfaction with symptoms [$P > 0.013$], and self-rated improvement [$P > 0.013$]) in long-term follow-up. An as-treated analysis showed significant surgical treatment effects for primary outcome measures (mean change, surgery *vs*. nonoperative care; treatment effect; 95% confidence interval): bodily pain (45.3 *vs*. 34.4; 10.9; 7.7 to 14); PF (42.2 *vs*. 31.5; 10.6; 7.7 to 13.5); and Oswestry Disability Index (−36.2 *vs*. −24.8; −11.3; −13.6 to −9.1).

Conclusion. Carefully selected patients who underwent surgery for a lumbar disc herniation achieved greater improvement than nonoperatively treated patients; there was little to no degradation of outcomes in either group (operative and nonoperative) from 4 to 8 years.

Key words: SPORT, intervertebral disc herniation, surgery, nonoperative care, outcomes.

Level of Evidence: 2

Spine 2014;39:3–16

Lumbar discectomy for relief of sciatica in patients with intervertebral disc herniation (IDH) is a well-researched and common indication for spine surgery, yet rates of this surgery exhibit considerable geographic variation.[1] Several randomized trials and large prospective cohorts have demonstrated that surgery provides faster pain relief and perceived recovery in patients with herniated disc.[2–6] The effect of surgery on longer-term outcomes remains less clear.

In a classic RCT evaluating surgery *versus* nonoperative treatment for lumbar IDH, Weber[2] showed a greater improvement in the surgery group at 1 year that was statistically significant; there was also greater improvement for surgery at 4 years, although not statistically significant, but no apparent difference in outcomes at 10 years. However, a number

From the *Geisel School of Medicine at Dartmouth, Hanover, NH, and Dartmouth-Hitchcock Medical Center, Lebanon, NH; and †William Beaumont Hospital, Royal Oak, MI.

Acknowledgment date: September 4, 2013. Revision date: October 14, 2013. Acceptance date: October 15, 2013.

†Died June 7, 2013.

The manuscript submitted does not contain information about medical device(s)/drug(s).

The National Institute of Arthritis and Musculoskeletal and Skin Diseases (U01-AR45444; P60-AR062799) 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 grant funds were received in support of this work.

Relevant financial activities outside the submitted work: consultancy, grants, stocks.

Address correspondence and reprint requests to Jon D. Lurie, MD, MS, Associate Professor, Department of Medicine, of Orthopaedics, and of the Dartmouth Institute for Health Policy and Clinical Practice, Geisel School of Medicine at Dartmouth and Dartmouth-Hitchcock, One Medical Center Dr., Lebanon, NH 03756; E-mail: tamara.s.morgan@dartmouth.edu

DOI: 10.1097/BRS.0000000000000088

of patients in the nonoperative group eventually underwent surgery during that time, complicating the interpretation of the long-term results. The Maine Lumbar Spine Study, a prospective observational cohort, found greater improvement at 1 year in the surgery group that narrowed over time, but remained significantly greater in the surgical group for sciatica bothersomeness, PF, and satisfaction, but no different for work or disability outcomes.[3] This article reports 8-year results from the Spine Patient Outcomes Research Trial (SPORT) based on the continued follow-up of the herniated disc randomized and observational cohorts.

MATERIALS AND METHODS

Study Design

SPORT is a randomized trial with a concurrent observation cohort 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.[5–8]

Patient Population

Males and females were eligible if they had symptoms and confirmatory signs of lumbar radiculopathy persisting for at least 6 weeks, disc herniation at a corresponding level and side on imaging, and were considered surgical candidates. The content of pre-enrollment nonoperative care was not prespecified in the protocol.[5–7] Specific enrollment and exclusion criteria are reported elsewhere.[6,7]

A research nurse at each site identified potential participants, verified eligibility, and used a shared decision making video for uniformity of enrollment. Participants were offered enrollment in 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.[7,9] 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.[5–8]

Study Measures

Primary endpoints were the Bodily Pain and PF scales of the SF-36 Health Survey[10] and the AAOS/Modems version of the Oswestry Disability Index (ODI)[11] 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 postoperatively. Secondary outcomes included patient self-reported improvement; work status; satisfaction with current symptoms and care;[12] and sciatica severity as measured by the sciatica bothersomeness index.[13,14] 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 8 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 because of treatment selection bias and missing data patterns.[15] In addition, baseline outcome, center, age, and sex 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.[6] Because of crossover, additional analyses were performed 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 prior to 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[16] with linear and logit link functions respectively, 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 test all follow-up visit times simultaneously for differences in

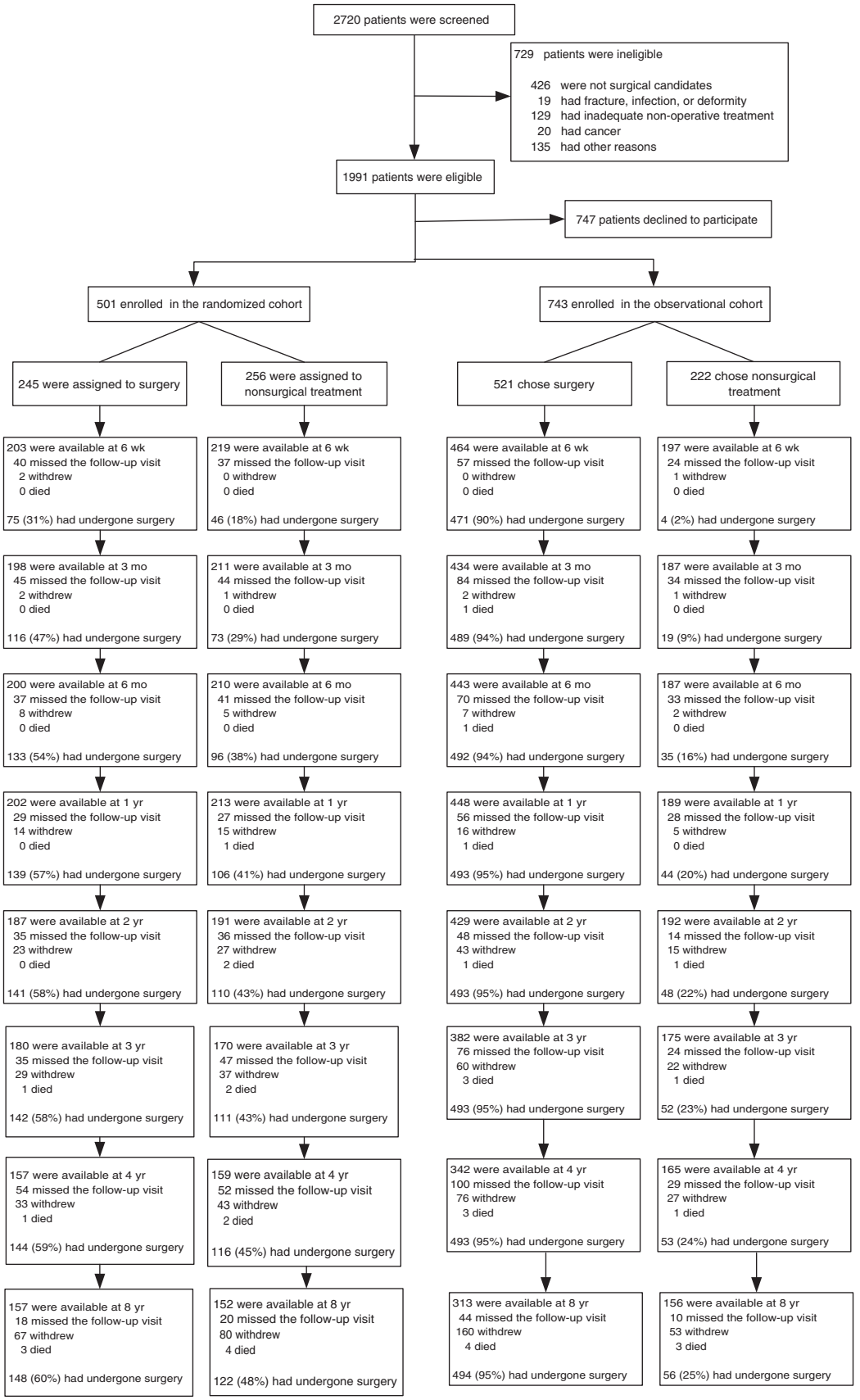

Figure 1. Exclusion, enrollment, randomization and follow-up of trial participants. The values for surgery, withdrawal, and death are cumulative during the 8 years. For example, a total of 3 patients in the group assigned to surgery died during the 8-year follow-up period. (Data set April 10, 2008).

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*		
IDH	Randomized Cohort (n = 474)	Observational Cohort (n = 721)	*P*	Surgery (n = 803)	Nonoperative (n = 392)	*P*
Mean age (SD)	42.3 (11.6)	41.4 (11.2)	0.18	40.7 (10.8)	43.8 (12.3)	<0.001
Female	194 (41%)	313 (43%)	0.43	346 (43%)	161 (41%)	0.55
Ethnicity, not Hispanic†	450 (95%)	690 (96%)	0.63	766 (95%)	374 (95%)	0.89
Race, Caucasian†	401 (85%)	635 (88%)	0.10	707 (88%)	329 (84%)	0.061
Education, at least some college	357 (75%)	530 (74%)	0.53	583 (73%)	304 (78%)	0.077
Income, under $50,000	208 (44%)	329 (46%)	0.59	373 (46%)	164 (42%)	0.15
Marital status, married	333 (70%)	504 (70%)	0.95	562 (70%)	275 (70%)	0.99
Work status‡			0.71			0.007
Full or part time	292 (62%)	433 (60%)		467 (58%)	258 (66%)	
Disabled	58 (12%)	100 (14%)		122 (15%)	36 (9%)	
Other	124 (26%)	187 (26%)		213 (27%)	98 (25%)	
Compensation, any	76 (16%)	132 (18%)	0.35	162 (20%)	46 (12%)	<0.001
Mean BMI, (SD)§	28 (5.5)	28 (5.6)	0.88	28.2 (5.7)	27.5 (5.3)	0.064
Smoker	108 (23%)	175 (24%)	0.60	201 (25%)	82 (21%)	0.13
Comorbidities						
Depression	62 (13%)	79 (11%)	0.31	94 (12%)	47 (12%)	0.96
Joint problem	98 (21%)	124 (17%)	0.15	130 (16%)	92 (23%)	0.003
Other¶	221 (47%)	305 (42%)	0.16	334 (42%)	192 (49%)	0.019
Time since recent episode <6 mo	374 (79%)	559 (78%)	0.62	619 (77%)	314 (80%)	0.27
BP score‖	28.3 (19.9)	26.4 (20.3)	0.13	23.4 (18)	34.8 (22.1)	<0.001
PF score‖	39.5 (25.3)	36.7 (25.7)	0.066	32.6 (23.5)	48.4 (26.4)	<0.001
MCS score	45.9 (12)	44.7 (11.2)	0.081	44.7 (11.4)	46.2 (11.8)	0.035
ODI**	46.9 (20.9)	51.1 (21.4)	<0.001	54.7 (19.6)	38.6 (20.5)	<0.001
Sciatica Frequency Index (0–24)††	15.6 (5.5)	16.1 (5.3)	0.19	16.7 (5.1)	14.2 (5.6)	<0.001
Sciatica Bothersomeness Index (0–24)‡‡	15.2 (5.2)	15.8 (5.3)	0.057	16.4 (4.9)	13.8 (5.6)	<0.001
Satisfaction with symptoms, very dissatisfied	371 (78%)	585 (81%)	0.25	705 (88%)	251 (64%)	<0.001
Problem getting better or worse			<0.001			<0.001
Getting better	90 (19%)	89 (12%)		66 (8%)	113 (29%)	
Staying about the same	221 (47%)	315 (44%)		348 (43%)	188 (48%)	
Getting worse	162 (34%)	311 (43%)		383 (48%)	90 (23%)	
Treatment preference			<0.001			<0.001

(*Continued*)

TABLE 1. (*Continued*)

IDH	SPORT Study Cohorts		*P*	Randomized and Observational Cohorts Combined: Treatment Received*		*P*
	Randomized Cohort (n = 474)	Observational Cohort (n = 721)		Surgery (n = 803)	Nonoperative (n = 392)	
Preference for nonoperative care	193 (41%)	202 (28%)		130 (16%)	265 (68%)	
Not sure	154 (32%)	43 (6%)		114 (14%)	83 (21%)	
Preference for surgery	127 (27%)	473 (66%)		556 (69%)	44 (11%)	
Pain radiation	459 (97%)	706 (98%)	0.33	787 (98%)	378 (96%)	0.15
Straight leg raise test, ipsilateral	291 (61%)	460 (64%)	0.43	520 (65%)	231 (59%)	0.058
Straight leg raise test, contralateral/both	68 (14%)	121 (17%)	0.29	153 (19%)	36 (9%)	<0.001
Any neurological deficit	352 (74%)	552 (77%)	0.40	625 (78%)	279 (71%)	0.014
Reflexes, asymmetric depressed	203 (43%)	279 (39%)	0.17	330 (41%)	152 (39%)	0.48
Sensory, asymmetric decrease	223 (47%)	382 (53%)	0.051	433 (54%)	172 (44%)	0.001
Motor, asymmetric weakness	191 (40%)	311 (43%)	0.36	359 (45%)	143 (36%)	0.008
Herniation level			0.10			<0.001
L2–L3/L3–L4	32 (7%)	56 (8%)		42 (5%)	46 (12%)	
L4–L5	166 (35%)	291 (40%)		314 (39%)	143 (36%)	
L5–S1	275 (58%)	374 (52%)		446 (56%)	203 (52%)	
Herniation type			0.86			0.49
Protruding	126 (27%)	196 (27%)		210 (26%)	112 (29%)	
Extruded	315 (66%)	471 (65%)		537 (67%)	249 (64%)	
Sequestered	32 (7%)	54 (7%)		55 (7%)	31 (8%)	
Posterolateral herniation	379 (80%)	542 (75%)	0.064	636 (79%)	285 (73%)	0.015

**Patients in the 2 cohorts combined were classified according to whether they received surgical treatment or only nonsurgical treatment during the first 8 years of enrollment.*

†Race or ethnic group was self-assessed. Caucasians and African Americans 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 denotes 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.

BMI indicates body mass index; SD, standard deviation; BP, bodily pain; PF, physical function; MCS, mental component summary; ODI, Oswestry Disability Index; CFS, chronic fatigue syndrome; PTSD, post-traumatic stress disorder.

estimated treatment effects between the 2 cohorts.[15] 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.[15]

Kaplan-Meier estimates of reoperation rates at 8 years were computed for the randomized and observational cohorts and compared *via* the log-rank test.[17,18]

TABLE 2. Operative Treatments, Complications and Events

IDH	Randomized Cohort* (n = 262)	Observational Cohort* (n = 548)	*P*
Discectomy level			
L2–L3	3 (1%)	12 (2%)	0.47
L3–L4	8 (3%)	20 (4%)	0.85
L4–L5	102 (40%)	217 (40%)	0.94
L5–S1	152 (59%)	306 (56%)	0.43
Median time to surgery (95% CI)†, mo	7.4 (4.7–42.3)	0.5 (0.4–0.7)	<0.001
Operation time (SD), min	80.5 (40.9)	74.9 (35.4)	0.049
Blood loss (SD), mL	75.3 (110.9)	63.2 (102.8)	0.13
Blood replacement			
Intraoperative replacement	4 (2%)	2 (0%)	0.16
Postoperative transfusion	0 (0%)	0 (0%)	
Length of stay (SD)	1 (1.1)	0.94 (0.9)	0.20
Postoperative mortality (death within 6 wk of surgery)	0 (0%)	0 (0%)	
Postoperative mortality (death within 3 mo of surgery)‡	0 (0%)	1 (0.2%)	0.72
Intraoperative complications§			
Dural tear/spinal fluid leak	12 (5%)	14 (3%)	0.19
Nerve root injury	1 (0%)	1 (0%)	0.82
Other	2 (1%)	1 (0%)	0.51
None	247 (94%)	533 (97%)	0.056
Postoperative complications/events¶			
Nerve root injury	0 (0%)	1 (0%)	0.70
Wound hematoma	0 (0%)	4 (1%)	0.40
Wound infection	4 (2%)	14 (3%)	0.52
Other	9 (4%)	18 (3%)	0.96
None	244 (95%)	513 (94%)	0.62
Additional surgical procedures (1-year rate)‖	11 (4%)	37 (7%)	0.13
Additional surgical procedures (2-year rate)‖	16 (6%)	50 (9%)	0.12
Additional surgical procedures (3-year rate)‖	20 (7%)	53 (10%)	0.29
Additional surgical procedures (4-year rate)‖	24 (9%)	61 (11%)	0.32
Additional surgical procedures (5-year rate)‖	25 (9%)	65 (12%)	0.27
Additional surgical procedures (6-year rate)‖	29 (11%)	73 (13%)	0.31
Additional surgical procedures (7-year rate)‖	33 (12%)	79 (14%)	0.40
Additional surgical procedures (8-year rate)‖	35 (13%)	84 (15%)	0.38
Recurrent disc herniation	17 (7%)	57 (11%)	
Complication or other	9 (3%)	21 (4%)	
New condition	3 (1%)	10 (2%)	

**A total of 270 RCT and 550 OBS patients had surgery. Surgical information was available for 262 RCT patients and 548 observational patients.*

†Median and 95% Confidence Interval based on Kaplan-Meier estimates and p-value based on log-rank test.

‡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, and vascular injury.

¶Any reported complications up to 8-week postoperation. None of the following were reported: bone graft complication, CSF leak, paralysis, cauda equina injury, wound dehiscence, and pseudarthrosis.

‖One-, 2-, 3-, 4-, 5-, 6-, 7-, and 8-year postsurgical reoperation 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.

SD indicates standard deviation; CI, confidence interval; CSF, cerebrospinal fluid.

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

	Assigned to Surgery			Assigned to Nonoperative		
	Treatment Received Within 8 yr			Treatment Received Within 8 yr		
	Surgery	Nonoperative	*P*	Surgery	Nonoperative	*P*
IDH	n = 146	n = 87		n = 119	n = 122	
Mean age (SD)	40.2 (10.9)	43.9 (13)	0.023	42.4 (10.2)	43.5 (12.4)	0.45
Income, under $50,000	67 (46%)	27 (31%)	0.036	63 (53%)	51 (42%)	0.11
Satisfaction with symptoms, very dissatisfied	128 (88%)	57 (66%)	<0.001	101 (85%)	85 (70%)	0.008
Herniation level			0.005			0.17
L2–L3/L3–L4	4 (3%)	12 (14%)		5 (4%)	11 (9%)	
L4–L5	54 (37%)	27 (31%)		47 (39%)	38 (31%)	
L5–S1	88 (60%)	48 (55%)		66 (55%)	73 (60%)	
Problem getting worse	60 (41%)	23 (26%)	0.03	47 (39%)	32 (26%)	0.04
Treatment preference			<0.001			<0.001
Preference for nonoperative care	47 (32%)	49 (56%)		34 (29%)	63 (52%)	
Not sure	49 (34%)	30 (34%)		37 (31%)	38 (31%)	
Preference for surgery	50 (34%)	8 (9%)		48 (40%)	21 (17%)	
BP score*	25.9 (18.9)	32.6 (22.5)	0.015	26.9 (19.1)	29.3 (19.7)	0.33
PF score*	36.6 (24.1)	45 (25.4)	0.012	34 (23.7)	44.2 (26.6)	0.002
ODI†	50.8 (21.1)	42 (20.8)	0.002	51 (19.3)	41.6 (20.8)	<0.001
Sciatica Frequency Index (0–24)‡	16.2 (5.2)	15.1 (6.1)	0.15	16.4 (5.5)	14.5 (5.4)	0.009
Sciatica Bothersomeness Index (0–24)§	15.9 (4.8)	14.7 (5.5)	0.11	16 (5.1)	14 (5.3)	0.003

**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.
BP indicates bodily pain; SD, standard deviation; PF, physical function; ODI, Oswestry Disability Index; IDH, intervertebral disc herniation.

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 February 4, 2013.

RESULTS

Overall, 1244 SPORT participants with lumbar IDH 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 60% by 8 years. In the group randomized to nonoperative care, 41% of patients had surgery by 1 year and 48% by 8 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 8 years 12 additional patients had undergone primary surgery. Of those choosing nonoperative treatment, 20% had surgery by 1 year and 25% by 8 years. In both cohorts combined, 820 patients received surgery at some point during the first 8 years; 424 (34%) remained nonoperative. During the 8 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%); 63% of initial enrollees supplied data at 8 years with losses due to dropouts, missed visits, or deaths (Figure 1).

Patient Characteristics

Baseline characteristics have been previously reported and are summarized in Table 1.[5,6,8] The combined cohorts had an overall mean age of 41.7 with slightly more males than females. Overall, the randomized and observational cohorts were similar. However, patients in the observational cohort had more baseline disability (higher ODI scores), were more likely to prefer surgery, more often rated their problem as

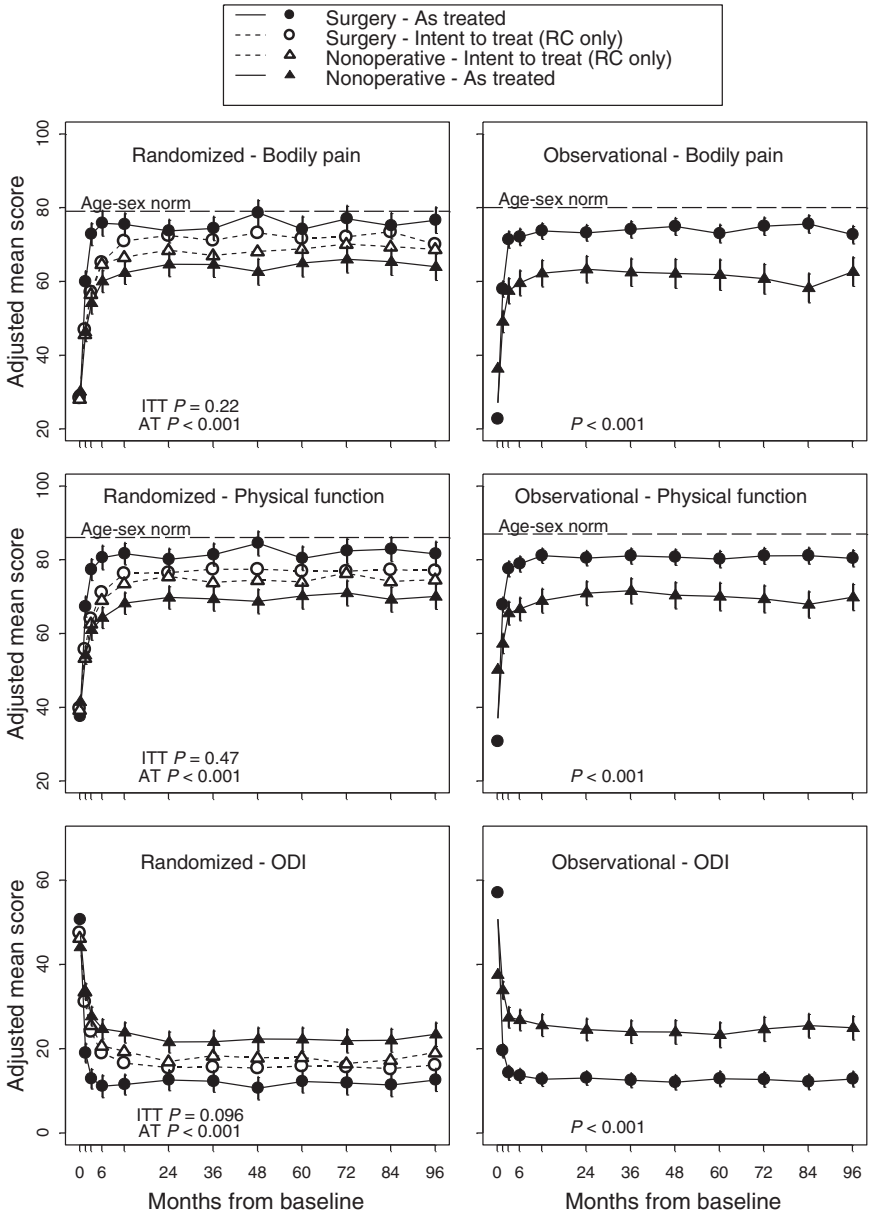

Figure 2. Primary outcomes (SF-36 bodily pain and physical function, and Oswestry Disability Index) in the randomized and observational cohorts during 8 years of follow-up. The graphs show both the intent-to-treat and the as-treated analyses for the randomized cohort (column on the left) and the as-treated analysis for the observational 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 vertical bars represent 95% confidence intervals. At 0 months, the floating data points represent the observed baseline mean scores for each study group, whereas the data points on plot lines represent the estimated means from the adjusted analyses. ODI indicates Oswestry Disability Index.

worsening, and were slightly more likely to have a sensory deficit. Subjects receiving surgery during the course of the study were younger, less likely to be working, more likely to report being on worker's compensation, had more severe baseline pain and functional limitations, fewer joint and other comorbidities, greater dissatisfaction with their symptoms; more often rated their condition as getting worse at enrollment, and were more likely to prefer surgery. Subjects receiving surgery were also more likely to have a positive straight leg test, as well as more frequent neurological, sensory, and motor deficits. Radiographically, their herniations were more likely to be at the L4–5 and L5–S1 levels and to be posterolateral in location.

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 (80.5 randomized *vs.* 74.9 min observational, $P = 0.049$). The average blood loss was 75.3 mL in the randomized cohort *versus* 63.2 mL in the observational, $P = 0.13$. Only 6 patients total required intraoperative transfusions. There were no perioperative mortalities. The most common surgical complication was dural tear (combined 3% of cases). Reoperation occurred in a combined 11% of cases by 5 years, 12% by 6 years, 14% by 7 years, and 15% by 8 years postsurgery. The rates of reoperation were not significantly different between the randomized and observational cohorts. Eighty-seven of the 119 reoperations noted the type of reoperation; approximately 85% of these (74/87) 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.

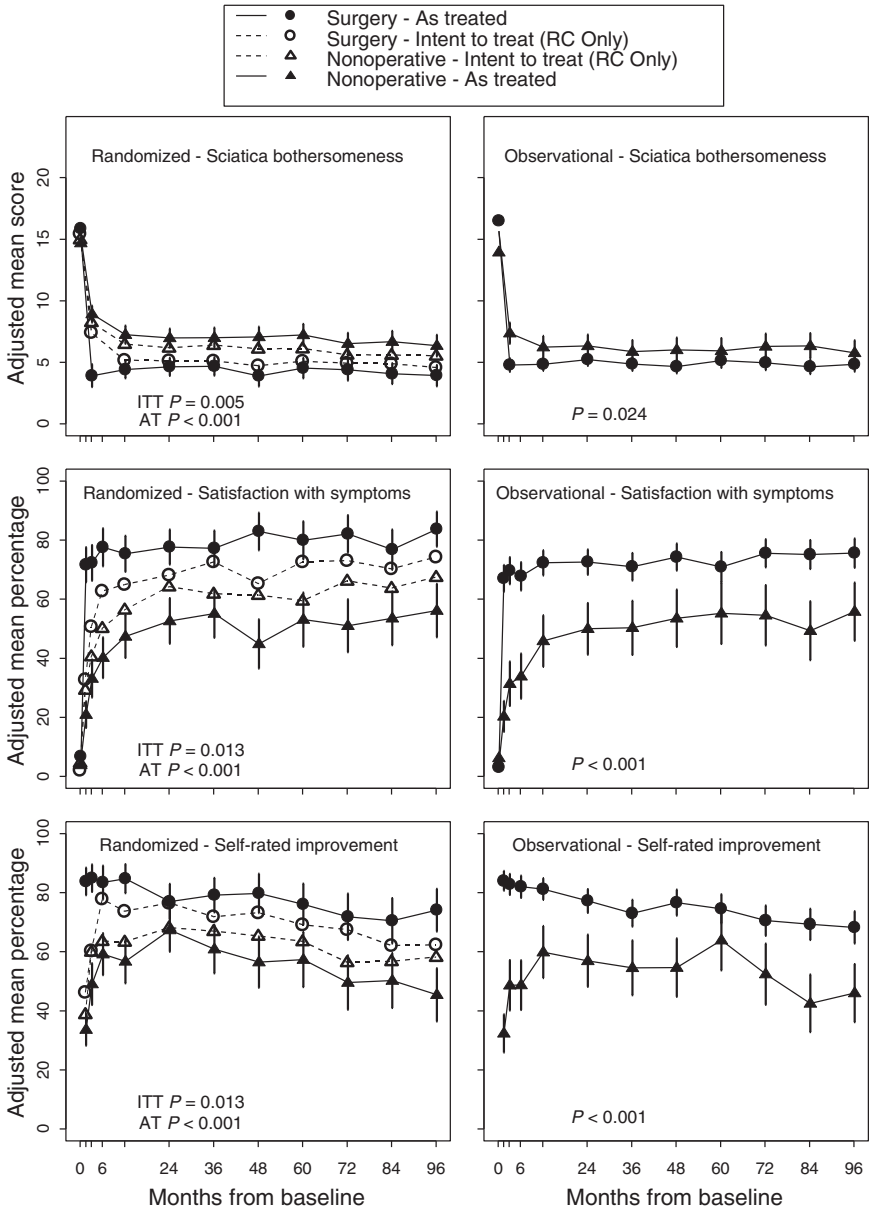

Figure 3. Secondary outcomes (sciatica bothersomeness, satisfaction with symptoms, and selfrated global improvement) in the randomized and observational cohorts during 8 years of follow-up. The graphs show both the intent-to-treat and the as-treated analyses for the randomized cohort (column on the left) and the as-treated analysis for the observational cohort (column on the right). The vertical bars represent 95% confidence intervals. At 0 months, the floating data points represent the observed baseline mean scores for each study group, whereas the data points on plot lines represent the estimated means from the adjusted analyses.

Crossover

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). Statistically significant differences of patients crossing over to nonoperative care within 8 years of enrollment were that they were older, had higher incomes, less dissatisfaction with their symptoms, more likely to have a disc herniation at an upper lumbar level, more likely to express a baseline preference for nonoperative care, less likely to perceive their symptoms as getting worse at baseline, and had less baseline pain and disability (Table 3). Patients crossing over to surgery within 8 years were more dissatisfied with their symptoms at baseline, were more likely to perceive they were getting worse at baseline, more likely to express a baseline preference for surgery, and had worse baseline PF and more self-rated disability.

Main Treatment Effects

Intent-to-Treat Analysis

In the intent-to-treat analysis of the randomized cohort, all measures during the 8 years favored surgery, but there were no statistically significant treatment effects in the primary outcome measures (Supplemental Digital Content Table 4, available at http://links.lww.com/BRS/A841, and Figure 2). In the overall intent-to-treat comparison between the 2 treatment groups over time (area under the curve), secondary outcomes were significantly greater with surgery in the intent-to-treat analysis (sciatica bothersomeness [$P = 0.005$], satisfaction with symptoms [$P = 0.013$], and self-rated improvement [$P = 0.013$]) (Figure 3). Improvement in sciatica bothersomeness index was also statistically significant in favor of surgery at most individual time point comparisons (although

TABLE 5. Patient Baseline Demographic Characteristics, Comorbidities, and Health Status Measures According to Patient Follow-up Status as of February 1, 2013, When the IDH 8-Year Data Were Pulled

IDH	Patients Currently in Study (n = 816)	Patients Lost to Follow-up (n = 379)	*P*
Mean age (SD)	42.2 (11.2)	40.7 (11.7)	0.039
Female	369 (45%)	138 (36%)	0.005
Ethnicity, not Hispanic	782 (96%)	358 (94%)	0.36
Race, Caucasian†	725 (89%)	311 (82%)	0.002
Education, at least some college	625 (77%)	262 (69%)	0.007
Income, under $50,000	367 (45%)	170 (45%)	0.98
Marital status, married	595 (73%)	242 (64%)	0.002
Work status			<0.001
Full or part time	536 (66%)	189 (50%)	
Disabled	73 (9%)	85 (22%)	
Other	207 (25%)	104 (27%)	
Compensation, any‡	115 (14%)	93 (25%)	<0.001
Mean BMI, (SD)§	27.8 (5.6)	28.3 (5.5)	0.16
Smoker	163 (20%)	120 (32%)	<0.001
Comorbidities			
Depression	89 (11%)	52 (14%)	0.19
Joint problem	150 (18%)	72 (19%)	0.86
Other¶	351 (43%)	175 (46%)	0.34
Time since recent episode <6 mo	645 (79%)	288 (76%)	0.27
BP score‖	28.1 (20.6)	25.1 (19)	0.015
PF score‖	38.8 (25.5)	35.7 (25.5)	0.052
MCS score‖	46 (11.5)	43.4 (11.4)	<0.001
ODI**	48.4 (21)	51.7 (21.9)	0.011
Sciatica Frequency Index (0–24)††	15.7 (5.4)	16.3 (5.5)	0.089
Sciatica Bothersomeness Index (0–24)‡‡	15.3 (5.2)	16.1 (5.3)	0.022
Satisfaction with symptoms, very dissatisfied	658 (81%)	298 (79%)	0.47
Problem getting better or worse			0.092
Getting better	133 (16%)	46 (12%)	
Staying about the same	370 (45%)	166 (44%)	
Getting worse	310 (38%)	163 (43%)	
Treatment preference			0.57
Preference for nonoperative care	277 (34%)	118 (31%)	
Not sure	136 (17%)	61 (16%)	
Preference for surgery	402 (49%)	198 (52%)	
Pain radiation	798 (98%)	367 (97%)	0.43
Straight leg raise test, ipsilateral	505 (62%)	246 (65%)	0.35

(*Continued*)

TABLE 5. (*Continued*)

IDH	Patients Currently in Study (n = 816)	Patients Lost to Follow-up (n = 379)	*P*
Straight leg raise test, contralateral/both	136 (17%)	53 (14%)	0.27
Any neurological deficit	630 (77%)	274 (72%)	0.077
Reflexes, asymmetric depressed	342 (42%)	140 (37%)	0.12
Sensory, asymmetric decrease	425 (52%)	180 (47%)	0.16
Motor, asymmetric weakness	347 (43%)	155 (41%)	0.64
Herniation level			0.43
L2–L3/L3–L4	65 (8%)	23 (6%)	
L4–L5	314 (38%)	143 (38%)	
L5–S1	436 (53%)	213 (56%)	
Herniation type			0.61
Protruding	223 (27%)	99 (26%)	
Extruded	530 (65%)	256 (68%)	
Sequestered	62 (8%)	24 (6%)	
Posterolateral herniation	631 (77%)	290 (77%)	0.81

†Race or ethnic group was self-assessed. Caucasians and African Americans 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 indicates 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.

BP indicates bodily pain; SD, standard deviation; PF, physical function; MCS, mental component summary; ODI, Oswestry Disability Index; BMI, body mass index; CFS, chronic fatigue syndrome; PTSD, post-traumatic stress disorder.

nonsignificant in years 6 and 7) (Supplemental Digital Content Table 4, available at http://links.lww.com/BRS/A841).

As-Treated Analysis

The adjusted as-treated effects seen in the randomized and observational were similar. Accordingly, the cohorts were combined for the final analyses. Treatment effects for the primary outcomes in the combined as-treated analysis were clinically meaningful and statistically significant out to 8 years: SF-36 bodily pain 10.9 $P < 0.001$ (95% confidence interval [CI], 7.7 to 14); SF-36 PF 10.6 $P < 0.001$ (95% CI, 7.7 to 13.5); ODI −11.3 $P < 0.001$ (95% CI, −13.6 to −9.1) (Supplemental Digital Content Table 4, available at http://links.lww.com/BRS/A841). The footnote for Supplemental Digital Content Table 4, available at http://links.lww.com/BRS/A841, describes the adjusting covariates selected for the final model.

Results from the intent-to-treat and as-treated analyses of the 2 cohorts are compared in Figure 2. 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 that did not differ between treatment groups) at each time point (Supplemental Digital Content Table 4, available at http://links.lww.com/BRS/A841, and Figure 3).

Loss to Follow-up

At the 8-year follow-up, 63% of initial enrollees supplied data, with losses due to dropouts, missed visits, or deaths. Table 5 summarized the baseline characteristics of those lost to follow-up compared with those retained in the study at 8 years. Those who remained in the study at 8 years were somewhat older; more likely to be female, Caucasian, college educated, and working at baseline; less likely to be disabled, receiving compensation, or a smoker; less symptomatic at baseline with somewhat less bodily pain, better PF, less disability on the ODI, better mental health, and less sciatica bothersomeness. These differences were small but statistically significant. Table 6 summarizes the short-term outcomes during the first 2 years for those retained in the study at 8 years compared with those lost to follow-up. Those lost to follow-up had worse outcomes on average; however, this was true in both the surgical and nonoperative groups with nonsignificant

TABLE 6. Time-Weighted Average of Treatment Effects at 2 Years (AUC) From Adjusted* as-Treated Randomized and Observational Cohorts Combined Primary Outcome Analysis, According to Treatment Received and Patient Follow-up Status

IDH	Patient Follow-up Status	Surgical	Nonoperative	Treatment Effect† (95% CI)
SF-36 BP (SE)‡	Currently in study	43.8 (0.7)	31.2 (0.9)	12.6 (10.4–14.8)
	Lost to follow-up	38.8 (1.2)	27.4 (1.6)	11.4 (7.7–15.1)
	P	<0.001	0.036	0.55
SF-36 PF (SE)‡	Currently in study	40.9 (0.7)	28.1 (0.9)	12.8 (10.8–14.9)
	Lost to follow-up	37 (1.1)	25 (1.5)	12 (8.6–15.4)
	P	0.003	0.071	0.65
ODI (SE)§	Currently in study	−35.7 (0.6)	−23.1 (0.7)	−12.7 (−14.3 to −11)
	Lost to follow-up	−31 (1)	−20.9 (1.3)	−10.1 (−13 to −7.3)
	P	<0.001	0.13	0.11
Sciatica Bothersomeness Index (SE)¶	Currently in study	−8.8 (0.2)	−6.6 (0.2)	−2.2 (−2.7 to −1.6)
	Lost to follow-up	−8.4 (0.3)	−5.5 (0.4)	−3 (−3.8 to −2.1)
	P	0.24	0.004	0.10

Adjusted for age, sex, race, marital status, compensation, smoking status, herniation location, working status, stomach comorbidity, depression, diabetes, other comorbidity, self-rated health trend, duration of most recent episode, treatment preference, baseline score (for SF-36, ODI, and Sciatica Bothersomeness Index), and center.*

†Treatment effect is the difference between the surgical and nonoperative 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 range from 0 to 24, with lower scores indicating less severe symptoms.

BP indicates bodily pain; PF, physical function; ODI, Oswestry Disability Index; CI, confidence interval; CFS, chronic fatigue syndrome; PTSD, post-traumatic stress disorder.

***Other comorbidities are: stroke, diabetes, osteoporosis, cancer, fibromyalgia, chronic fatigue syndrome, PTSD, alcohol, drug dependence, heart, lung, liver, kidney, blood vessel, nervous system, hypertension, migraine, anxiety, stomach or bowel.*

differences in treatment effects. The long-term outcomes are therefore likely to be somewhat overoptimistic on average in both groups, but the comparison between surgical and nonoperative outcomes seem likely to be unbiased despite the long-term loss to follow-up.

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 as 6 weeks, seemed to reach a maximum by 6 months and persisted during the 8 years; it is notable that the nonoperative group also improved significantly and this improvement persisted with little to no degradation of outcomes in either group (operative and nonoperative) between 4 and 8 years. In the longitudinal intent-to-treat analysis, all the outcomes showed small advantages for surgery, but only the secondary outcomes of sciatica bothersomeness, satisfaction with symptoms, and self-rated improvement were statistically significant. The persistent small benefit in the surgery group over time has made the overall intent-to-treat comparison more statistically significant over time despite high levels of crossover. The large effects seen in the as-treated analysis after adjustments for characteristics of the crossover patients suggest that the intent-to-treat analysis may underestimate the true effect of surgery since the mixing of treatments due to crossover can be expected to create a bias toward the null in the intent-to-treat analyses.[4,19] Loss to follow-up among patients who were somewhat worse at baseline and with worse short-term outcomes probably leads to overly optimistic estimated long-term outcomes in both surgery and nonoperative groups but unbiased estimates of surgical treatment effects.

Comparisons With Other Studies

There are no other long-term randomized studies reporting the same primary outcome measures as SPORT. The results of SPORT primary outcomes at 2 years were quite similar to those of Peul *et al* but longer follow-up for the Peul study is necessary for further comparison.[4,20] In contrast to the Weber study,[2] the differences in the outcomes in SPORT between treatment groups remained relatively constant between 1 and 8 years of follow-up. One of the factors in this difference may be the sensitivity of the outcome measures, for example, sciatica bothersomeness, which was significantly different out to 8 years in the intent-to-treat, may be a more sensitive marker of treatment success than the general outcome measure used by Weber.[2]

The long-term results of SPORT are similar to the Maine Lumbar Spine Study (MLSS).[21–22] The MLSS reported statistically significantly greater improvements at 10 years in sciatica bothersomeness for the surgery group (−11.9) than the nonsurgical groups (−5.8) with a treatment effect of −6.1 $P = 0.004$; in SPORT the improvement in sciatica bothersomeness in the surgical group at 8 years was similar to the 10-year result in MLSS (−11) though the nonoperative cohort in SPORT did better than their MLSS counterparts (−9.1); however, the treatment effect in SPORT, whereas smaller, remained statistically significant (−1.5; $P < 0.001$) because of the much larger sample size. Greater improvements in the nonoperative cohorts between SPORT and MLSS may be related to differences in nonoperative treatments over time, differences between the 2 cohorts since the MLSS and did not require imaging confirmation of IDH.

During the 8 years, there was little evidence of harm from either treatment. The 8-year rate of reoperation was 14.7%, which is lower than the 25% reported by MLSS at 10 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.[4–6] However, intent-to-treat analyses are known to be biased in the presence of noncompliance at the level observed in SPORT, and our adjusted as-treated analyses have been shown to produce accurate results under reasonable assumptions about the dependence of compliance on longitudinal outcomes.[23] Another potential limitation is the heterogeneity of the nonoperative treatment interventions, as discussed in our prior articles.[5,6,8] Finally, attrition in this long-term follow-up study meant that only 63% of initial enrollees supplied data at 8 years with losses due to dropouts, missed visits, or deaths; based on analyses at baseline and at short-term follow-up, this likely leads to somewhat overly optimistic estimated long-term outcomes in both treatment groups but an unbiased estimation of surgical treatment effect.

CONCLUSION

In the intent-to-treat analysis, small, statistically insignificant surgical treatment effects were seen for the primary outcomes but statistically significant advantages for sciatica bothersomeness, satisfaction with symptoms, and self-rated improvement were seen out to 8 years despite high levels of treatment crossover. The as-treated analysis combining the randomized and observational cohorts, which carefully controlled for potentially confounding baseline factors, showed significantly greater improvement in pain, function, satisfaction, and self-rated progress during the 8 years than patients treated nonoperatively. The nonoperative group, however, also showed substantial improvements over time, with 54% reporting being satisfied with their symptoms and 73% satisfied with their care after 8 years. Even among patients with strong surgical indications such as those in SPORT, fully a third chose to remain nonoperative during the long term.

➢ Key Points

- ❑ In the randomized intent-to-treat analysis at the 8-year follow-up, patients with IDH randomized to surgery showed small, statistically insignificant treatment effects compared with nonoperative care in primary outcome measures but statistically significantly greater improvements in secondary outcome measures of sciatica bothersomeness, satisfaction with symptoms, and self-rated improvement, despite high levels of "crossover" in both groups.
- ❑ At the 8-year follow-up, patients who had surgery for IDH maintained clinically significant greater improvement in all primary outcomes than those who remained nonoperative based on adjusted analyses according to treatment received.
- ❑ There was little to no degradation of outcomes in either group (operative and nonoperative) between 4 and 8 years.
- ❑ Even among patients with strong surgical indications, many (34%) remained in the nonoperative group out to 8 years.

Acknowledgment

This study is dedicated to the memories of Brieanna Weinstein and Harry Herkowitz, leaders in their own rights, who simply made the world a better place.

Supplemental digital content is available for this article. Direct URL citations appearing in the printed text are provided in the HTML and PDF version of this article on the journal's web site (www.spinejournal.com).

References

1. Dartmouth Atlas Working Group. *Dartmouth Atlas of Musculoskeletal Health Care*. Chicago, IL: American Hospital Association Press; 2000.
2. Weber H. Lumbar disc herniation. A controlled, prospective study with ten years of observation. *Spine (Phila Pa 1976)* 1983;8:131–40.
3. Atlas SJ, Deyo RA, Keller RB, et al. The Maine Lumbar Spine Study, part II. 1-year outcomes of surgical and nonsurgical management of sciatica. *Spine (Phila Pa 1976)* 1996;21:1777–86.
4. 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.
5. 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.
6. 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.
7. Birkmeyer NJ, Weinstein JN, Tosteson AN, et al. Design of the Spine Patient Outcomes Research Trial (SPORT). *Spine (Phila Pa 1976)* 2002;27:1361–72.
8. Weinstein JN, Lurie JD, Tosteson TD, et al. Surgical versus nonoperative treatment for lumbar disc herniation: four-year results for the Spine Patient Outcomes Research Trial (SPORT). *Spine (Phila Pa 1976)* 2008;33:2789–800.

9. Delamarter R, McCullough J. Microdiscectomy & microsurgical laminotomies. In: Frymoyer J, ed. *The Adult Spine: Principles and Practice*. 2nd ed. Philadelphia: Lippincott-Raven Publishers; 1996.
10. 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.
11. Daltroy LH, Cats-Baril WL, Katz JN, et al. The North American Spine Society Lumbar Spine Outcome Assessment Instrument: reliability and validity tests. *Spine (Phila Pa 1976)* 1996;21:741–9.
12. Deyo RA, Diehl AK. Patient satisfaction with medical care for low-back pain. *Spine (Phila Pa 1976)* 1986;11:28–30.
13. 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 (Phila Pa 1976)* 1996;21:2885–92.
14. Patrick DL, Deyo RA, Atlas SJ, et al. Assessing health-related quality of life in patients with sciatica. *Spine (Phila Pa 1976)* 1995;20:1899–908; discussion 1909.
15. Fitzmaurice G, Laird N, Ware J. *Applied Longitudinal Analysis*. Philadelphia, PA: John Wiley & Sons; 2004.
16. Diggle PJ, Liang K-Y, Zeger SL. *Analysis of Longitudinal Data*. Oxford, UK: Oxford University Press; 1994.
17. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Statist Assoc* 1958;53:457–81.
18. Peto R, Peto J. Asymptotically efficient rank invariant test procedures. *J R Statist Soc A* 1972;135:185.
19. Meinert CL. *Clinical Trials: Design, Conduct, and Analysis*. New York, NY: Oxford University Press, Inc.; 1986.
20. Peul WC, van den Hout WB, Brand R, et al. Prolonged conservative care versus early surgery in patients with sciatica caused by lumbar disc herniation: two year results of a randomised controlled trial. *BMJ* 2008;336:1355–8.
21. 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 (Phila Pa 1976)* 2001;26:1179–87.
22. Atlas SJ, Keller RB, Wu YA, et al. Long-term outcomes of surgical and nonsurgical management of sciatica secondary to a lumbar disc herniation: 10 year results from the Maine Lumbar Spine Study. *Spine (Phila Pa 1976)* 2005;30:927–35.
23. Sitlani CM, Heagerty PJ, Blood EA, et al. Longitudinal structural mixed models for the analysis of surgical trials with noncompliance. *Stat Med* 2012;31:1738–60.