

Spine Patient Outcomes Research Trial

Radiographic Predictors of Clinical Outcomes After Operative or Nonoperative Treatment of Degenerative Spondylolisthesis

Adam M. Pearson, MD, MS,* Jon D. Lurie, MD, MS,* Emily A. Blood, MS,*
John W. Frymoyer, MD,* Heike Braeutigam, MD,† Howard An, MD,‡
Federico P. Girardi, MD,§ and James N. Weinstein, DO, MS*

Study Design. Subgroup analyses according to treatment received.

Objective. To evaluate whether baseline radiographic findings predicted outcomes in patients with degenerative spondylolisthesis.

Summary of Background Data. The spine patient outcomes research trial combined randomized and observational DS cohorts.

Methods. The Meyerding listhesis grade was determined on the neutral radiograph ($n = 222$). Patients were classified as having low disc height if disc height was less than 5 mm. Flexion-extension radiographs ($n = 185$) were evaluated for mobility. Those with greater than 10° rotation or 4 mm translation were considered hypermobile. Changes in outcome measures were compared between listhesis (grade 1 vs. grade 2), disc height (low vs. normal), and mobility (stable vs. hypermobile) groups using longitudinal regression models adjusted for potential confounders. Outcome measures included SF-36 bodily pain and physical function scales, Oswestry disability index (ODI), stenosis bothersomeness index, and low back pain bothersomeness scale.

Results. Overall, 86% had a grade 1 listhesis, 78% had normal disc height, and 73% were stable. Baseline symptom severity was similar between groups. Overall, surgery patients improved more than patients treated nonoperatively. At 1 year, outcomes were similar in surgery patients across listhesis, disc height, and mobility groups (ODI: grade 1 -23.7 vs. grade 2 -23.3 , $P = 0.90$; normal disc height -23.5 vs. low disc height -21.9 , $P = 0.66$; stable -21.6 vs. hypermobile -25.2 , $P = 0.30$). Among

those treated nonoperatively, grade 1 patients improved more than grade 2 patients (bodily pain $+13.1$ vs. -4.9 , $P = 0.019$; ODI -8.0 vs. $+4.8$, $P = 0.010$ at 1 year), and hypermobile patients improved more than stable patients (ODI -15.2 vs. -6.6 , $P = 0.041$; stenosis bothersomeness index -7.8 vs. -2.7 , $P = 0.002$ at 1 year).

Discussion. Regardless of listhesis grade, disc height or mobility, patients who had surgery improved more than those treated nonoperatively. These differences were due, in part, to differences in nonoperative outcomes, which were better in patients classified as grade 1 or hypermobile.

Key words: radiograph, surgery, degenerative spondylolisthesis, outcomes, nonoperative. **Spine 2008;33:2759–2766**

Since early clinical descriptions of degenerative spondylolisthesis (DS),^{1–3} it has been suggested that certain radiographic features are related to surgical outcomes, including the magnitude of the slip,^{4–6} the degree of disc space narrowing,^{6–8} and angular and translational hypermobility identified on functional radiographs.^{6,9–14} Many of these clinical studies on degenerative “instability” were not specific for DS and included patients with a variety of degenerative conditions.^{13–18} Because most of these studies have focused on surgical outcomes,^{5,9,13–16,19–21} the role of radiographic findings in predicting the natural history and nonoperative outcomes in patients with DS remains unknown.^{8,22,23} As a result, it is uncertain to what extent treatment decisions for DS should be influenced by radiographic findings.

Recently, we reported the results of the Spine Patient Outcomes Research Trial (SPORT) for DS, which demonstrated that patients treated surgically had quantitatively better outcomes than patients managed nonoperatively over 2 years of observation.²³ The specific goals of the current study were to: (1) describe the baseline characteristics of DS patients stratified by listhesis grade, disc height, and hypermobility; and (2) determine if surgical and nonoperative outcomes were associated with these baseline radiographic findings.

Materials and Methods

Study Design

The initial design of SPORT consisted of a randomized controlled trial with a concurrent observational cohort study conducted in 11 states at 13 institutions with multidisciplinary spine practices.²⁴ The human subject committees at each par-

From the *Dartmouth Medical School, Lebanon, NH; †Helios-Rosmann-Hospital, Breisach, Germany; ‡Rush University Medical Center, Chicago, IL; and §Hospital for Special Surgery, New York, NY.

Acknowledgment date: October 17, 2007. Acceptance date: January 21, 2008.

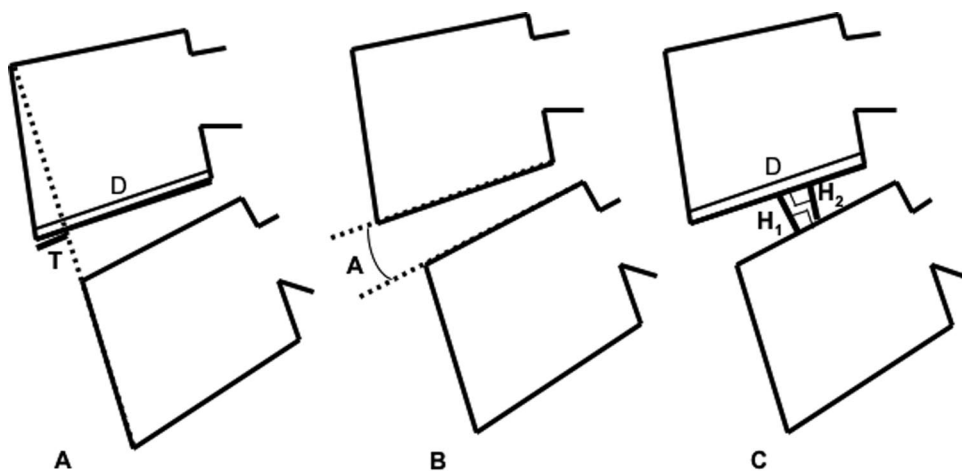
The manuscript submitted does not contain information about medical device(s)/drug(s).

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.

The authors received funding from the following sources: 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. Pearson was funded by NIAMS (T32-AR-049710). Dr. Lurie received support from a Research Career Award from NIAMS (1 K23 AR 048138-01).

Address correspondence and reprint requests to Adam M. Pearson, MD, MS, Department of Orthopaedics, Dartmouth-Hitchcock Medical Center, One Medical Center Dr., Lebanon, NH 03756; E-mail: adam.m.pearson@dartmouth.edu

Figure 1. Measurement techniques to determine translation (A), intervertebral angle (B), and disc height (C). Anterior-posterior translation (T) was measured from the antero-inferior corner of the upper vertebra relative to the line of the anterior cortex of the lower vertebra and normalized to vertebral depth (D). The intervertebral angle (A) was measured between the inferior endplate of the upper vertebra and the superior endplate of the lower vertebra. Disc height was calculated as the average of the distance from the midpoint of the inferior endplate of the upper vertebra to its perpendicular intersection with the superior endplate of the lower vertebra (H_1) and of the distance from the midpoint of the superior endplate of the lower vertebra to its perpendicular intersection with the inferior endplate of the upper vertebra (H_2).



icipating institution approved a standardized protocol for the study.

Patient Population

Patients were considered for inclusion in the DS cohort of SPORT if they: were over 18 years old; had neurogenic claudication or radicular pain with associated neurologic signs for at least 12 weeks; spinal stenosis on cross-sectional imaging; DS identified on the standing lateral radiograph; and were considered surgical candidates by their treating physicians.^{23,24} Exclusion criteria included cauda equina syndrome, malignancy, other significant deformities, prior back surgery, and other established contraindications to elective surgery.²⁴

Imaging Studies

All patients enrolled in the study ($n = 607$) had standing lateral neutral and flexion-extension radiographs obtained at baseline evaluation. Because of the technical complexities and expense involved in digitizing radiographs, the neutral lateral radiographs of 222 patients were available for independent review. Of this group, 185 patients also had digitized flexion-extension radiographs available. Thus, the 222 patients included in this study represent a “convenience sample” of the entire DS cohort, including patients from 11 of the 13 sites.

Listhesis Grade

The listhesis grade was quantified on neutral lateral radiographs using Meyerding’s classification.²⁵ Anterior translation of the listhetic vertebra less than 25% of the anterior-posterior (AP) vertebral depth was classified as grade I, and translation of 25% to 50% was classified as grade II. There were no translations greater than 50%.

Disc Height

Disc height was measured on the lateral radiograph using Quint’s method specific for spondylolisthesis (Figure 1).²⁶ The calculated disc height was normalized to the AP vertebral depth to account for differences in magnification. A patient was classified as having low disc height if the normalized disc height was less than or equal to 0.139. This stratification was based on Wilke *et al*²⁷ who classified a decrease in disc height by two-thirds as severe, and Frobin *et al*²⁸ who reported the average disc height at L4–L5 as 0.41 (normalized to the AP depth of L5).

Mobility

The AP translation of the listhetic vertebra relative to the lower vertebra from the extension to flexion radiograph was determined using Quint’s method of digitizing the corners of the vertebral bodies,²⁶ and then calculating intervertebral rotation and translation using the method of Morgan and King (Figure 1).²⁹ Intervertebral rotation was calculated as the change in intervertebral angle from extension to flexion.²⁶ Based on Hanley’s definition of instability, patients were classified as hypermobile if anterior translation of the affected motion segment exceeded 10% of the AP vertebral depth, or if intervertebral rotation exceeded 10°. ³⁰ Patients not meeting these criteria were classified as stable.

Reliability

We previously have reported the intrarater reliability of the radiographic measurements employed in this study and found intraclass correlation coefficients of 0.90, 0.89, and 0.93 for translation, intervertebral angle, and disc height, respectively.³¹

Study Interventions

All patients treated surgically had a decompressive laminectomy. If fusion was performed, it consisted of iliac crest bone grafting with or without instrumentation based on the surgeon’s preferences.²⁴ The nonoperative treatment group received “usual care,” recommended to include at least physical therapy, education, and counseling with home exercise instruction, and nonsteroidal anti-inflammatory drugs if tolerated. Details are reported elsewhere.²³

Study Measures

Data used in this study were obtained from patient questionnaires completed at baseline, 1 and 2 years after enrollment or surgery that included the SF-36,³² Oswestry disability index (ODI),³³ low back pain bothersomeness scale,³⁴ and the stenosis bothersomeness index (SBI).^{35,36} The SF-36 scales and the ODI range from 0 to 100, the SBI from 0 to 24, and the low back pain bothersomeness scale from 0 to 6. Higher scores indicated more severe symptoms on the ODI, SBI, and low back pain bothersomeness scale, whereas higher scores indicated less severe symptoms on the SF-36.

Statistical Considerations

The initial design of SPORT included both a randomized and an observational cohort. In the first 2 years of surveillance of the DS randomized trial, 36% of patients assigned to surgery did not have that intervention, and 49% of patients assigned to nonoperative treatment did have surgery. We previously reported a comparison of the baseline characteristics between the randomized and observational cohorts.²³ The only significant differences between the cohorts were a lower frequency of L3–L4 involvement and lateral recess stenosis in the observational group. Furthermore, there were no significant differences in the treatment effects of surgery between the 2 groups. Given the high rate of treatment cross-over and the consistency of the baseline characteristics between the randomized and the observational cohorts, the data from both cohorts were combined in an as-treated analysis. The detailed statistical rationale for this strategy has been published elsewhere.³⁷

Differences in baseline radiographic characteristics were compared for the listhesis (grade I *vs.* grade II), disc height (normal *vs.* low), and mobility (stable *vs.* hypermobile) groups, using χ^2 tests for categorical data and Student *t* test for continuous data. Comparisons of clinical baseline characteristics and outcomes were also made between the patients included in the convenience sample and those not included to evaluate if the convenience sample was representative of the overall cohort.

The primary analyses compared changes in the clinical outcome measures from baseline as a function of the degree of listhesis, disc height, and mobility within each treatment arm (*i.e.*, surgery or nonoperative). In addition, the treatment effects of surgery were also compared between the listhesis, mobility, and disc height groups. The treatment effect of surgery was defined as:

Treatment Effect

$$= \text{Change in Outcome Measure}_{\text{surgery}} - \text{Change in Outcome Measure}_{\text{nonoperative}}$$

Positive treatment effects for SF-36 scores and negative treatment effects for ODI, SBI, and low back pain bothersomeness score indicated that surgery was more effective than nonoperative treatment. In these analyses, the treatment indicator (surgery or nonoperative) was assigned according to the actual treatment received at each time point. For surgery patients, all changes from baseline before surgery were included in the estimates of the effect of nonoperative treatment. After surgery, follow-up times were measured from the date of surgery.

To adjust for potential confounding, baseline variables associated with missing data or treatment received (age, sex, work status, depression, osteoporosis, joint problems, duration of current symptoms, reflex deficit, number of moderate or severe stenotic levels, medical center, and baseline SF-36, ODI, and SBI) were included as adjusting covariates in longitudinal regression models.³⁸ A random effect was specified to account for the repeated measurements of individual patients. Statistical analysis was performed on SAS Software (SAS Institute Inc., Cary, NC) using PROC MIXED for continuous data with normal random effects (bodily pain (BP), physical function scales, ODI, sciatica bothersomeness) and PROC GENMOD for non-normal outcomes (low back pain bothersomeness). At each time point, adjusted mean scores were estimated, and differences between the listhesis, mobil-

ity, and disc height groups were compared using a Wald test. Statistical significance was defined as $P < 0.05$ on the basis of a 2-sided hypothesis test.

Results

Of the 892 DS patients eligible for the study, 607 (68%) were enrolled, while 285 (32%) declined to participate. The convenience sample selected from the 607 enrolled patients included 222 (37%) patients with neutral lateral digitized radiographs and 185 (30%) who also had lateral flexion-extension radiographs available for analysis. The baseline characteristics of the convenience sample patients were not significantly different from the 385 patients not included except that a lower proportion of the convenience sample had motor weakness (18% *vs.* 28%, $P = 0.008$) and L3–L4 listhesis (5% *vs.* 12%, $P = 0.014$). About 60% of available images were from the observational cohort and 40% from the randomized cohort. Outcome data were available for 98% of patients at 1 year and 94% of patients at 2 years.

The study population had a mean age of 66 years, and 70% were women (Table 1). Nine percent had applied for or were receiving disability compensation, and 23% were working full time. Of the 222 patients with lateral radiographs, surgery was performed in 139 patients (63%) within the first 2 years, and the remaining 83 (37%) were treated nonoperatively.

Among surgical patients, 71% underwent decompression with instrumented fusion, 24% decompression with uninstrumented fusion, and 5% decompression alone. Detailed information about surgical complications has been published elsewhere.²³ In the 83 patients treated nonoperatively, treatment included education and counseling (88%), nonsteroidal anti-inflammatory drugs (52%), narcotic pain medication (36%), physical therapy (41%), and epidural injections (36%).

Listhesis Grade

Of the 222 patients, 192 (86%) had grade I listhesis, whereas the remaining 30 (14%) demonstrated grade II slips (Table 1). The grade II group included a higher proportion of women (90% in grade II *vs.* 67% in grade I, $P = 0.02$) and patients with low disc height (60% in grade II *vs.* 16% in grade I, $P < 0.001$). There were no other significant differences in baseline characteristics between the grade I and grade II groups, nor did the type of surgical procedure performed vary by listhesis grade.

The treatment effect of surgery for the grade II group at 1 year was greater than the grade I group for BP (41.2 *vs.* 14.5, $P = 0.003$) and ODI (−28.0 *vs.* −15.7, $P = 0.04$) (Table 2). These differences resulted from the poorer outcomes in the nonoperatively treated grade II group (grade II *vs.* grade I BP −4.9 *vs.* +13.1, $P = 0.02$; ODI +4.8 *vs.* −8.0, $P = 0.01$). At 2 years, these differences in treatment effect were no longer significant, with

Table 1. Baseline Patient Characteristics

	Listhesis Grade		<i>P</i>	Normal (n = 173)	Disc Height Low (n = 49)	<i>P</i>	Stable (n = 135)	Mobility Hypermobile (n = 50)	<i>P</i>	Combined (n = 222)
	Grade 1 (n = 192)	Grade 2 (n = 30)								
Mean age (stdev)	65.7 (10.4)	66 (10.9)	0.88	64.7 (10.7)	69.3 (8.6)	0.007	66.2 (10.2)	65.5 (10.4)	0.69	65.7 (10.4)
Gender-female	129 (67%)	27 (90%)	0.02	117 (68%)	39 (80%)	0.15	101 (75%)	29 (58%)	0.041	156 (70%)
Race-white	154 (80%)	28 (93%)	0.14	137 (79%)	45 (92%)	0.068	109 (81%)	43 (86%)	0.54	182 (82%)
Work status-full time	44 (23%)	6 (20%)	0.90	41 (24%)	9 (18%)	0.55	25 (19%)	15 (30%)	0.14	50 (23%)
Compensation status			0.96			0.32			0.28	
None	175 (91%)	27 (90%)		155 (90%)	47 (96%)		120 (89%)	48 (96%)		202 (91%)
Receiving/pending	16 (8%)	3 (10%)		17 (10%)	2 (4%)		14 (10%)	2 (4%)		19 (9%)
Mean body mass index (BMI), (stdev)	29.8 (6)	29 (5.6)	0.48	29.9 (5.9)	28.8 (6.2)	0.27	30 (5.9)	28.9 (6.6)	0.25	29.7 (6)
Bodily pain (BP) score	32 (16.5)	28.8 (15.1)	0.33	32 (16.9)	29.9 (14.4)	0.43	30.8 (16.4)	32.4 (16.7)	0.56	31.6 (16.3)
Physical functioning (PF) score	34.8 (22.9)	32.2 (22.7)	0.56	36.2 (23)	28.4 (21.5)	0.03	32.3 (22.7)	37.8 (23.8)	0.15	34.5 (22.9)
Mental component summary (MCS) score	50.8 (11.5)	51.2 (9.4)	0.85	50.1 (11.6)	53.4 (9.2)	0.06	50.6 (11.1)	49.9 (11.6)	0.72	50.8 (11.2)
Oswestry (ODI)	41.6 (17.5)	41.2 (17.2)	0.89	41.7 (17.8)	41.2 (16.3)	0.85	41.8 (17.3)	41.5 (19.1)	0.90	41.6 (17.5)
Stenosis bothersome index (0–24)	14.6 (5.8)	14.7 (4.2)	0.88	14.5 (5.8)	14.9 (5.2)	0.64	14.5 (5.8)	14.5 (5.8)	0.99	14.6 (5.6)
Back pain bothersomeness score	4.2 (1.9)	4.7 (1.5)	0.19	4.3 (1.9)	4.2 (1.9)	0.93	4.3 (1.9)	4.3 (1.8)	0.97	4.3 (1.9)
Straight leg raise test-present	24 (12%)	3 (10%)	0.93	21 (12%)	6 (12%)	0.82	16 (12%)	6 (12%)	0.82	27 (12%)
Femoral tension sign	7 (4%)	0 (0%)	0.62	5 (3%)	2 (4%)	0.97	5 (4%)	2 (4%)	0.73	7 (3%)
Depressed reflexes	42 (22%)	11 (37%)	0.12	36 (21%)	17 (35%)	0.068	32 (24%)	12 (24%)	0.88	53 (24%)
Depressed sensation	58 (30%)	10 (33%)	0.89	54 (31%)	14 (29%)	0.86	44 (33%)	12 (24%)	0.34	68 (31%)
Motor weakness	36 (19%)	4 (13%)	0.64	30 (17%)	10 (20%)	0.78	29 (21%)	7 (14%)	0.35	40 (18%)
Pseudoclaudication	167 (87%)	27 (90%)	0.87	149 (86%)	45 (92%)	0.41	123 (91%)	41 (82%)	0.14	194 (87%)
Treatment received within 2 yr			0.056			0.20			0.41	
Surgery	115 (60%)	24 (80%)		104 (60%)	35 (71%)		81 (60%)	34 (68%)		139 (63%)
Nonoperative	77 (40%)	6 (20%)		69 (40%)	14 (29%)		54 (40%)	16 (32%)		83 (37%)
Procedure (among those receiving surgery)			0.40			0.13			0.034	
Decompression only	7 (4%)	0 (0%)		3 (2%)	4 (8%)		6 (4%)	1 (2%)		7 (5%)
Noninstrumented fusion	28 (15%)	5 (17%)		26 (15%)	7 (14%)		23 (17%)	3 (6%)		33 (24%)
Instrumented fusion	80 (42%)	19 (63%)		75 (43%)	24 (49%)		52 (39%)	30 (60%)		99 (71%)
RCT or OBS			0.96			0.86			0.72	
RCT	74 (39%)	12 (40%)		66 (38%)	20 (41%)		51 (38%)	21 (42%)		86 (39%)
OBS	118 (61%)	18 (60%)		107 (62%)	29 (59%)		84 (62%)	29 (58%)		136 (61%)
DS level			0.92			0.54			0.96	
L3–L4	11 (6%)	1 (3%)		8 (5%)	4 (8%)		7 (5%)	2 (4%)		12 (5%)
L4–L5	181 (94%)	29 (97%)		165 (95%)	45 (92%)		128 (95%)	48 (96%)		210 (95%)
Diabetes	25 (13%)	1 (3%)	0.22	22 (13%)	4 (8%)	0.53	13 (10%)	9 (18%)	0.19	26 (12%)
Listhesis grade			n/a			<0.001			0.32	
Grade 1	192 (100%)	0 (0%)		161 (93%)	31 (63%)		120 (89%)	41 (82%)		192 (86%)
Grade 2	0 (0%)	30 (100%)		12 (7%)	18 (37%)		15 (11%)	9 (18%)		30 (14%)
Disc height (>0.139)			<0.001			n/a			0.53	
Normal	161 (84%)	12 (40%)		173 (100%)	0 (0%)		103 (76%)	41 (82%)		173 (78%)
Low	31 (16%)	18 (60%)		0 (0%)	49 (100%)		32 (24%)	9 (18%)		49 (22%)
Mobility			0.32			0.53			n/a	
Stable	120 (62%)	15 (50%)		103 (60%)	32 (65%)		135 (100%)	0 (0%)		135 (61%)
Hypermobile	41 (21%)	9 (30%)		41 (24%)	9 (18%)		0 (0%)	50 (100%)		50 (23%)

a delayed improvement in BP between 1 and 2 years in the nonoperatively treated group with grade II slips.

Disc Height

Seventy-eight percent of the 222 patients had normal disc height at the affected level, whereas the remaining 22% were classified as low disc height (Table 1). The low disc height group was older (69.3 years *vs.* 64.7 years, $P = 0.007$) and had worse physical function scores (28.4 *vs.* 36.2, $P = 0.03$), but the type of surgical procedures performed were similar for the 2 groups.

Among patients treated surgically, the normal disc height group improved more than the low disc height group on SBI at 1 year (-9.8 *vs.* -6.3 , $P = 0.01$), though this difference was no longer significant at 2 years (-8.5 *vs.* -6.1 , $P = 0.10$) (Table 2). Comparison of the low disc height and normal disc height patients treated nonoperatively revealed no significant differences in outcomes, and there were no significant differences in treatment effect sizes between the disc height groups.

Table 2. Outcomes by Listhesis and Mobility Groups

	N	1 yr					2 yr						
		Surgery (n = 111)	P†	Nonoperative (n = 97)	P†	Treatment Effect	P†	Surgery (n = 129)	P†	Nonoperative (n = 79)	P†	Treatment Effect	P†
Listhesis													
SF-36 bodily pain													
Grade 1	192	27.6 (2)	0.097	13.1 (2.1)	0.019	14.5 (9, 19.9)	0.003	28.3 (2.1)	0.13	7.6 (2.3)	0.57	20.8 (14.9, 26.6)	0.78
Grade 2	30	36.3 (4.8)		-4.9 (7.4)		41.2 (24.3, 58.2)		36.4 (5)		12.7 (8.7)		23.7 (4.2, 43.2)	
SF-36 physical function													
Grade 1	192	29 (2.2)	0.32	8.6 (2.2)	0.54	20.5 (14.8, 26.1)	0.88	26.5 (2.2)	0.92	7 (2.4)	0.89	19.5 (13.5, 25.6)	0.84
Grade 2	30	22.1 (6.7)		3.1 (8.8)		19 (1.3, 36.8)		25.8 (6.8)		8.4 (10)		17.4 (-3.2, 37.9)	
ODI													
Grade 1	192	-23.7 (1.7)	0.90	-8 (1.7)	0.01	-15.7 (-20.1, -11.3)	0.04	-22.9 (1.7)	0.54	-6.5 (1.9)	0.11	-16.4 (-21.1, -11.7)	0.095
Grade 2	30	-23.3 (2.9)		4.8 (4.6)		-28 (-38.9, -17.2)		-25 (3)		2.8 (5.5)		-27.8 (-40.4, -15.2)	
Stenosis bothersomeness													
Grade 1	192	-9 (0.61)	0.93	-4 (0.64)	0.19	-5 (-6.7, -3.3)	0.26	-7.9 (0.62)	0.34	-3.5 (0.72)	0.88	-4.4 (-6.3, -2.6)	0.76
Grade 2	30	-9.1 (1.3)		-1 (2.2)		-8.1 (-13.2, -3)		-9.3 (1.4)		-3.9 (2.7)		-5.4 (-11.5, 0.6)	
Back pain bothersomeness													
Grade 1	192	-2.3 (0.2)	0.41	-1.1 (0.2)	0.42	-1.2 (-1.7, -0.7)	0.29	-2 (0.2)	0.021	-1 (0.2)	0.69	-1 (-1.5, -0.4)	0.46
Grade 2	30	-2.7 (0.4)		-0.6 (0.6)		-2 (-3.5, -0.6)		-3 (0.4)		-1.3 (0.7)		-1.6 (-3.3, 0.1)	
Disc height													
SF-36 bodily pain													
Normal	173	30.3 (2.2)	0.21	13.1 (2.2)	0.068	17.2 (11.4, 22.9)	0.47	30.1 (2.2)	0.42	8.3 (2.4)	0.50	21.8 (15.7, 28)	0.31
Low	49	24.3 (4.3)		1.9 (5.8)		22.4 (9.4, 35.5)		26 (4.6)		14.3 (8.7)		11.7 (-6.9, 30.3)	
SF-36 physical function													
Normal	173	29.5 (2.3)	0.21	8.4 (2.4)	0.68	21.2 (15.1, 27.2)	0.55	27.7 (2.3)	0.23	6.6 (2.6)	0.29	21.1 (14.6, 27.6)	0.085
Low	49	23 (4.6)		5.8 (5.7)		17.2 (5.7, 28.7)		21.3 (4.8)		15.6 (8.1)		5.7 (-10.6, 22)	
ODI													
Normal	173	-23.5 (1.8)	0.66	-7.8 (1.8)	0.33	-15.7 (-20.2, -11.2)	0.60	-23.4 (1.7)	0.42	-6.4 (2)	0.38	-17 (-21.8, -12.2)	0.23
Low	49	-21.9 (3.2)		-3.3 (4.3)		-18.6 (-28.7, -8.6)		-20.3 (3.4)		-12.4 (6.6)		-7.8 (-22.1, 6.4)	
Stenosis bothersomeness													
Normal	173	9.8 (0.63)	0.011	-3.9 (0.66)	0.20	-5.9 (-7.7, -4.2)	0.62	-8.5 (0.62)	0.10	-3.4 (0.74)	0.66	-5.1 (-7, -3.3)	0.23
Low	49	-6.3 (1.2)		-1.5 (1.8)		-4.8 (-8.9, -0.7)		-6.1 (1.3)		-4.6 (2.6)		-1.5 (-7.1, 4.1)	
Back pain bothersomeness													
Normal	173	-2.5 (0.2)	0.31	-1.1 (0.2)	0.61	-1.4 (-1.9, -0.9)	0.87	-2.2 (0.2)	0.65	-1.1 (0.2)	0.36	-1.1 (-1.6, -0.5)	0.32
Low	49	-2.1 (0.3)		-0.9 (0.5)		-1.3 (-2.4, -0.1)		-2 (0.4)		-1.8 (0.7)		-0.2 (-1.8, 1.4)	
Mobility													
SF-36 bodily pain													
Stable	135	28.4 (2.5)	0.95	13.1 (2.6)	0.58	15.3 (8.6, 22)	0.64	28.2 (2.5)	0.46	8.7 (2.8)	0.97	19.6 (12.4, 26.7)	0.68
Hypermobility	50	28.1 (3.9)		16.2 (5)		11.9 (-0.4, 24.2)		31.6 (3.8)		8.9 (5.6)		22.7 (9.5, 36)	
SF-36 physical function													
Stable	135	25.4 (2.6)	0.57	8.8 (2.7)	0.26	16.6 (9.7, 23.4)	0.60	24.7 (2.8)	0.19	7.6 (3)	0.15	17.1 (9.7, 24.5)	0.72
Hypermobility	50	28.1 (3.9)		15.1 (4.8)		13 (1.4, 24.5)		30.9 (3.8)		16.5 (5.4)		14.4 (1.9, 26.9)	
ODI													
Stable	135	-21.6 (2)	0.30	-6.6 (2)	0.041	-15 (-20.2, -9.7)	0.36	-22.8 (2)	0.89	-4.9 (2.2)	0.12	-18 (-23.5, -12.4)	0.24
Hypermobility	50	-25.2 (2.9)		-15.2 (3.7)		-10 (-19.2, -0.8)		-23.3 (2.8)		-12.1 (4.1)		-11.2 (-21.1, -1.2)	
Stenosis bothersomeness													
Stable	135	-9.2 (0.73)	0.57	-2.7 (0.78)	0.002	-6.5 (-8.6, -4.5)	0.007	-8.1 (0.74)	0.89	-2.3 (0.88)	0.016	-5.8 (-8.1, -3.6)	0.062
Hypermobility	50	-8.5 (1)		-7.8 (1.4)		-0.7 (-4.4, 2.9)		-8.3 (1)		-6.8 (1.6)		-1.5 (-5.4, 2.4)	
Back pain bothersomeness													
Stable	135	-2.4 (0.2)	0.64	-1 (0.2)	0.32	-1.4 (-2, -0.8)	0.65	-2.2 (0.2)	0.059	-0.9 (0.2)	0.52	-1.3 (-1.9, -0.6)	0.57
Hypermobility	50	-2.6 (0.3)		-1.4 (0.4)		-1.1 (-2, -0.2)		-2.8 (0.3)		-1.2 (0.4)		-1.6 (-2.6, -0.6)	

*Adjusted for age, gender, work status, depression, osteoporosis, joint problems, current symptom duration, reflex deficit, no. of moderate/severe stenotic levels, baseline score (for SF-36, ODI, and Stenosis Scales), baseline stenosis bothersomeness, center. Adjusted as treated analysis of change from baseline.

†P is based on a Wald test comparing treatment effects between listhesis, disc height, or mobility groups at each timepoint.

‡P is based on a Wald test comparing group effects between listhesis, disc height, or mobility groups within treatment arm at each timepoint.

Mobility

Of the 185 patients with flexion-extension radiographs, 131 (73%) were classified as stable, whereas the remaining 44 (27%) demonstrated hypermobility. The hypermobile patients included fewer women than the stable group (58% vs. 75%, $P = 0.04$), but the 2 groups were otherwise similar at baseline (Table 1). The hypermobile group was more likely to undergo instrumented fusion than the stable group (88% vs. 64%, $P = 0.03$).

The hypermobile group tended to improve more with nonoperative treatment than did the stable group. As a result, hypermobile patients had a smaller treatment effect of surgery. Among nonoperatively treated patients, the hypermobile group improved more than the stable group on ODI (-15.2 vs. -6.6, $P = 0.04$ at 1 year) and

SBI (-7.8 vs. -2.7, $P = 0.002$ at 1 year and -6.8 vs. -2.3, $P = 0.02$ at 2 years). Consequently, the treatment effect of surgery on SBI was smaller for the hypermobile group at 1 year (-0.7 vs. -6.5, $P = 0.007$), and the difference was on the threshold of significance at 2 years (-1.5 vs. -5.8, $P = 0.06$).

Discussion

This analysis of a subset of the SPORT DS patients demonstrated only modest associations between baseline radiographic parameters and surgical and nonoperative treatment outcomes. In this study, we have focused on 3 radiographic measurements: the degree of listhesis, disc space narrowing, and hypermobility at the affected motion segment. We did not evaluate other radiographic

factors that may predispose patients to DS, such as facet orientation and segmentation abnormalities,^{39,40} and this study did not allow for measurement of deformity progression over time.

The effect of listhesis grade on clinical outcomes has not been carefully analyzed in DS.^{4,5,9} Sengupta and Herkowitz⁶ recommended instrumented fusion for the uncommon case of DS where the slip exceeds 50%, but they did not make specific recommendations for grade I and II listhesis. In accordance with prior studies, we found the grade II listhesis group included a higher proportion of women and low disc height patients than did the grade I group.^{6,8} We also found the treatment effect of surgery was significantly greater for grade II than grade I patients at 1 year. These data suggest a high grade slip portends a poorer outcome with nonoperative treatment.

Disc height was generally not associated with outcomes in either the surgically or nonoperatively treated patients. Patients with low disc height were older, had baseline physical function scores indicative of more severe symptoms, and, as noted, were more likely to have grade II slips. Some investigators have suggested that loss of disc height, although indicative of greater degeneration, could lead to “restabilization” and diminution of back pain over time.^{8,28,41} However, that same process could be associated with worsening stenotic symptoms due to advancing facet degeneration accompanying the disc space narrowing. This study does not shed light on that question.

Of the 3 radiographic parameters evaluated, hypermobility (referred to by some as “segmental instability”) has received the most attention in the literature. Many investigators have suggested instrumented fusion is most appropriate for hypermobile DS patients despite a lack of rigorous outcome data to support this.^{6,9–11,13,14} Surgeons participating in the current study seem to have been influenced by this recommendation, insofar as instrumented fusion was more commonly performed in patients with increased angular and translational movements.

Our mobility analyses did yield some unexpected results. First, the hypermobile group had a significantly lower proportion of women compared with the stable group, a finding contrary to the traditional belief that women are more likely to have segmental hypermobility than men. However, our findings are consistent with those of McGregor *et al*⁴² who found that women had a smaller total lumbar flexion-extension range of motion compared with men. Second, radiographically hypermobile patients had better nonoperative treatment results than stable patients. Comparison of this finding to prior results is limited because no prior DS study has compared surgical and nonoperative outcomes stratified by baseline radiographic mobility. In a small study that combined spinal stenosis and DS patients, Yone *et al*^{13,14} demonstrated that unstable patients undergoing decompression with instrumented fusion had better results than

unstable patients undergoing decompression alone. In the current study, we were unable to compare surgical outcomes among hypermobile patients who underwent decompression alone, decompression with uninstrumented fusion or decompression with instrumented fusion because of the small sample sizes.

This unexpected finding does bring into sharper focus the long-unanswered question, “What constitutes instability in patients with degenerative diseases?” Traditionally, DS has been viewed as the prototype for degenerative instabilities because it was the condition most associated with the 4 criteria for instability at a motion segment: pain, translational and rotational hypermobility, risk for progressive deformity, and the potential for neurologic injury.^{21,43,44} Because we have not obtained follow-up radiographs, we cannot be certain if the deformity increased or if progressive listhesis was associated with increased pain. However, we do know that symptoms decreased over 2 years with nonoperative treatment, particularly in those patients with stringently defined radiographic hypermobility. Although surgery resulted in better outcomes for both hypermobile and stable patients, hypermobility should not be considered a contraindication to nonoperative treatment. Given that hypermobile patients improved with nonoperative treatment, the need for fusion in these patients, particularly those with significant medical comorbidities, should be further explored.

The current study has some limitations. Although all patients enrolled in the study had spinal radiographs, we were able, for logistic and economic reasons, to digitize and analyze images from only about one-third of the patients. Comparison of baseline characteristics of the patients with available images to those not included suggested that the 2 groups were very similar, so we felt that the convenience sample was fairly representative. Another issue was the relatively small proportion of patients that were classified as having grade II listhesis, hypermobility or low disc height. These relatively small subgroups may have limited our power to detect significant differences.

A statistical limitation related to the overall SPORT study has been the substantial cross-over between treatment arms.²³ In our prior publications, we have addressed this concern and the rationale behind the decision to perform an as-treated analysis using multiple regression models to control for baseline differences.^{23,37} Despite this approach, it is possible the current analysis could be vulnerable to residual confounding by unmeasured variables.

Similar to prior spinal radiographic studies, the criteria for classifying listhesis, disc height, and mobility were somewhat arbitrary, although they were based on literature standards.^{25,27,30} The greatest continuing controversy is the radiographic definition of instability.^{21,43,45–49} Some authors have suggested that there is substantial overlap in radiographic findings between “unstable” and “normal” subjects.^{50,51} For this reason, we chose a strin-

gent radiographic definition of hypermobility. We also performed analyses in which the listhesis, disc height, and mobility groups were defined by the median value for each characteristic and found similar results (data not shown). The more perplexing problem is that anterolisthesis can be found in 10% of women over age 65, yet the majority of patients are asymptomatic.^{52,53} In addition, radiographic signs of “instability” do not correlate with symptoms.^{45,54} SPORT used strict clinical and radiographic inclusion criteria, and we wish to emphasize that our findings do not apply to patients without both clinical signs and symptoms of spinal stenosis and associated radiographic abnormalities.

Finally, obtaining reliable radiographic measurements can be difficult.⁵⁵ Given that the intraclass correlation coefficients for the measurements used in this study ranged from 0.89 to 0.93, we feel that the measurement technique and classification system were reasonably reliable.³¹

Despite its limitations, the current study has identified associations between baseline radiographic findings and outcomes in DS patients which should be useful to clinicians caring for these patients. Patients with grade II listhesis had a greater treatment effect of surgery compared with grade I patients at 1 year, suggesting that surgery is more strongly favored in patients with higher grade slips. The other significant and unexpected finding was that hypermobile patients had better nonoperative outcomes than did stable patients. This indicates that hypermobility should not be considered a contraindication to nonoperative treatment, and leaves open the questions “Who benefits most from fusion?” and “Which patients require instrumented fusion?”

■ Key Points

- The study evaluated whether baseline radiographic findings predicted outcomes in patients with degenerative spondylolisthesis.
- Regardless of listhesis grade, disc height or mobility, patients who had surgery improved more than those treated nonoperatively.
- These differences were due, in part, to differences in nonoperative outcomes, which were better in patients classified as Grade 1 or Hypermobility.

References

1. Junghanns H. Spondylolisthesen ohne Spalt in Zwischengelenkstueck. *Arch Orthop Unfallchir* 1930;29:118–27.
2. Macnab I. Spondylolisthesis with an intact neural arch—the so-called pseudospondylolisthesis. *J Bone Joint Surg [Am]* 1950;32:325–33.
3. Newman PH, Stone KH. The etiology of spondylolisthesis. *J Bone Joint Surg [Br]* 1963;45:39–59.
4. Fischgrund JS, Mackay M, Herkowitz HN, et al. 1997 Volvo Award winner in clinical studies. Degenerative lumbar spondylolisthesis with spinal stenosis: a prospective, randomized study comparing decompressive laminectomy and arthrodesis with and without spinal instrumentation. *Spine* 1997;22:2807–12.
5. Herkowitz HN, Kurz LT. Degenerative lumbar spondylolisthesis with spinal stenosis. A prospective study comparing decompression with decompression and intertransverse process arthrodesis. *J Bone Joint Surg Am* 1991;73:802–8.
6. Sengupta DK, Herkowitz HN. Degenerative spondylolisthesis: review of current trends and controversies. *Spine* 2005;30:S71–81.
7. Lombardi JS, Wiltse LL, Reynolds J, et al. Treatment of degenerative spondylolisthesis. *Spine* 1985;10:821–7.
8. Matsunaga S, Ijiri K, Hayashi K. Nonsurgically managed patients with degenerative spondylolisthesis: a 10- to 18-year follow-up study. *J Neurosurg* 2000;93:194–8.
9. Bridwell KH, Sedgewick TA, O'Brien MF, et al. The role of fusion and instrumentation in the treatment of degenerative spondylolisthesis with spinal stenosis. *J Spinal Disord* 1993;6:461–72.
10. Frymoyer JW, Selby DK. Segmental instability. Rationale for treatment. *Spine* 1985;10:280–6.
11. Herkowitz HN. Spine update. Degenerative lumbar spondylolisthesis. *Spine* 1995;20:1084–90.
12. Kanayama M, Hashimoto T, Shigenobu K, et al. Intraoperative biomechanical assessment of lumbar spinal instability: validation of radiographic parameters indicating anterior column support in lumbar spinal fusion. *Spine* 2003;28:2368–72.
13. Yone K, Sakou T. Usefulness of Posner's definition of spinal instability for selection of surgical treatment for lumbar spinal stenosis. *J Spinal Disord* 1999;12:40–4.
14. Yone K, Sakou T, Kawauchi Y, et al. Indication of fusion for lumbar spinal stenosis in elderly patients and its significance. *Spine* 1996;21:242–8.
15. Friberg O. Lumbar instability: a dynamic approach by traction-compression radiography. *Spine* 1987;12:119–29.
16. Johnsson KE, Redlund-Johnell I, Uden A, et al. Preoperative and postoperative instability in lumbar spinal stenosis. *Spine* 1989;14:591–3.
17. McGregor AH, McCarthy ID, Dore CJ, et al. Quantitative assessment of the motion of the lumbar spine in the low back pain population and the effect of different spinal pathologies of this motion. *Eur Spine J* 1997;6:308–15.
18. Wood KB, Popp CA, Transfeldt EE, et al. Radiographic evaluation of instability in spondylolisthesis. *Spine* 1994;19:1697–703.
19. Cauchoix J, Benoist M, Chassaing V. Degenerative spondylolisthesis. *Clin Orthop Relat Res* 1976:122–9.
20. Herron LD, Trippi AC. L4–5 degenerative spondylolisthesis. The results of treatment by decompressive laminectomy without fusion. *Spine* 1989;14:534–8.
21. Nizard RS, Wybier M, Laredo JD. Radiologic assessment of lumbar intervertebral instability and degenerative spondylolisthesis. *Radiol Clin North Am* 2001;39:55–71, v–vi.
22. Matsunaga S, Sakou T, Morizono Y, et al. Natural history of degenerative spondylolisthesis. Pathogenesis and natural course of the slippage. *Spine* 1990;15:1204–10.
23. Weinstein JN, Lurie JD, Tosteson TD, et al. Surgical versus nonsurgical treatment for lumbar degenerative spondylolisthesis. *N Engl J Med* 2007;356:2257–70.
24. Birkmeyer NJ, Weinstein JN, Tosteson AN, et al. Design of the Spine Patient Outcomes Research Trial (SPORT). *Spine* 2002;27:1361–72.
25. Meyerding H. Spondylolisthesis: surgical treatment and results. *SurgGynecol Obstet* 1932;54:371–7.
26. Quint DJ, Tuite GF, Stern JD, et al. Computer-assisted measurement of lumbar spine radiographs. *Acad Radiol* 1997;4:742–52.
27. Wilke HJ, Rohlmann F, Neidlinger-Wilke C, et al. Validity and interobserver agreement of a new radiographic grading system for intervertebral disc degeneration: Part I. Lumbar spine. *Eur Spine J* 2006;15:720–30.
28. Frobin W, Brinckmann P, Kramer M, et al. Height of lumbar discs measured from radiographs compared with degeneration and height classified from MR images. *Eur Radiol* 2001;11:263–9.
29. Morgan F, King T. Primary instability of lumbar vertebrae as a common cause of low back pain. *J Bone Joint Surg Br* 1957;39B:6–22.
30. Hanley EN Jr. The indications for lumbar spinal fusion with and without instrumentation. *Spine* 1995;20:143S–53S.
31. Lurie J, Braeutigam H, Spratt K, et al. Reliability of computer-assisted radiographic measurements in patients with degenerative spondylolisthesis in SPORT. In: *Annual Meeting of the International Society for the Study of the Lumbar Spine*. Bergen, Norway; 2006.
32. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30:473–83.
33. 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.
34. Atlas SJ, Deyo RA, Keller RB, et al. The Maine Lumbar Spine Study, Part III. 1-year outcomes of surgical and nonsurgical management of lumbar spinal stenosis. *Spine* 1996;21:1787–94; discussion 94–5.

35. 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.
36. 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.
37. Tosteson TD, Hanscom B, Blood EA, et al. Statistical methods for cross-over in the SPORT lumbar disc herniation trial. In: *International Society for the Study of the Lumbar Spine Annual Meeting*. Hong Kong; 2007.
38. Fitzmaurice G, Laird N, Ware J. *Applied Longitudinal Analysis*. Philadelphia, PA: John Wiley & Sons; 2004.
39. Grobler LJ, Robertson PA, Novotny JE, et al. Etiology of spondylolisthesis. Assessment of the role played by lumbar facet joint morphology. *Spine* 1993;18:80–91.
40. Rosenberg NJ. Degenerative spondylolisthesis. Predisposing factors. *J Bone Joint Surg Am* 1975;57:467–74.
41. Kirkaldy-Willis WH, Wedge JH, Yong-Hing K, et al. Pathology and pathogenesis of lumbar spondylosis and stenosis. *Spine* 1978;3:319–28.
42. McGregor AH, Cattermole HR, Hughes SP. Global spinal motion in subjects with lumbar spondylolysis and spondylolisthesis: does the grade or type of slip affect global spinal motion? *Spine* 2001;26:282–6.
43. White A, Panjabi M. *Clinical Biomechanics of the Spine*. 2nd ed. Philadelphia, PA: JB Lippincott; 1990.
44. Nachemson A. Lumbar spine instability. A critical update and symposium summary. *Spine* 1985;10:290–1.
45. Dupuis PR, Yong-Hing K, Cassidy JD, et al. Radiologic diagnosis of degenerative lumbar spinal instability. *Spine* 1985;10:262–76.
46. Muggleton JM, Kondracki M, Allen R. Spinal fusion for lumbar instability: does it have a scientific basis? *J Spinal Disord* 2000;13:200–4.
47. Murata M, Morio Y, Kuranobu K. Lumbar disc degeneration and segmental instability: a comparison of magnetic resonance images and plain radiographs of patients with low back pain. *Arch Orthop Trauma Surg* 1994;113:297–301.
48. Pope MH, Frymoyer JW, Krag MH. Diagnosing instability. *Clin Orthop Relat Res* 1992:60–7.
49. Soini J, Antti-Poika I, Tallroth K, et al. Disc degeneration and angular movement of the lumbar spine: comparative study using plain and flexion-extension radiography and discography. *J Spinal Disord* 1991;4:183–7.
50. Boden SD, Wiesel SW. Lumbosacral segmental motion in normal individuals. Have we been measuring instability properly? *Spine* 1990;15:571–6.
51. Hayes MA, Howard TC, Gruel CR, et al. Roentgenographic evaluation of lumbar spine flexion-extension in asymptomatic individuals. *Spine* 1989;14:327–31.
52. Epstein NE, Epstein JA, Carras R, et al. Degenerative spondylolisthesis with an intact neural arch: a review of 60 cases with an analysis of clinical findings and the development of surgical management. *Neurosurgery* 1983;13:555–61.
53. Valkenburg HA, Haanen HCM. The epidemiology of low back pain. In: White AA, Gordon SL, eds. *Proc Am Assoc Orthop Surg Symposium on Low Back Pain*. 1982:9–22.
54. Sengupta DK. Dynamic stabilization devices in the treatment of low back pain. *Orthop Clin North Am* 2004;35:43–56.
55. Shaffer WO, Spratt KF, Weinstein J, et al. 1990 Volvo Award in clinical sciences. The consistency and accuracy of roentgenograms for measuring sagittal translation in the lumbar vertebral motion segment. An experimental model. *Spine* 1990;15:741–50.