ELSEVIER

The Spine Journal ■ (2013) ■

THE SPINE JOURNAL

Clinical Study

One-year outcomes of surgical versus nonsurgical treatments for discogenic back pain: a community-based prospective cohort study

Sohail K. Mirza, MD, MPH[a,*], Richard A. Deyo, MD, MPH[b,c,d,e], Patrick J. Heagerty, PhD[f], Judith A. Turner, PhD[g], Brook I. Martin, PhD[a], Bryan A. Comstock, MS[f]

[a]*Department of Orthopaedic Surgery and The Dartmouth Institute for Health Policy and Clinical Practice, Geisel School of Medicine at Dartmouth, Hanover, NH 03755, USA*

[b]*Department of Family Medicine, Oregon Health and Science University, 3181 SW Sam Jackson Park Rd, Portland, OR 97239-3098, USA*

[c]*Department of Medicine, Oregon Health and Science University, 3181 SW Sam Jackson Park Rd, Portland, OR 97239-3098, USA*

[d]*Department of Public Health and Preventive Medicine, Oregon Health and Science University, 3181 SW Sam Jackson Park Rd, Portland, OR 97239-3098, USA*

[e]*Center for Research in Occupational and Environmental Toxicology, Oregon Health and Science University, 3181 SW Sam Jackson Park Rd, Portland, OR 97239-3098, USA*

[f]*Department of Biostatistics, University of Washington, Box 357232, 1959 NE Pacific St, Seattle, WA 98195, USA*

[g]*Department of Psychiatry and Behavioral Sciences, University of Washington, Box 356560, 1959 NE Pacific St, Seattle, WA 98195, USA*

Received 2 March 2012; revised 20 November 2012; accepted 24 May 2013

Abstract

BACKGROUND CONTEXT: The clinical entity ''discogenic back pain'' remains controversial at fundamental levels, including its pathophysiology, diagnostic criteria, and optimal treatment. This is true despite availability of four randomized trials comparing the efficacy of surgical and nonsurgical treatments. One trial showed benefit for lumbar fusion compared with unstructured nonoperative care, and three others showed roughly similar results for lumbar surgery and structured rehabilitation.

PURPOSE: To compare outcomes of community-based surgical and nonsurgical treatments for patients with chronic back pain attributed to degeneration at one or two lumbar disc levels.

DESIGN: Prospective observational cohort study.

PATIENT SAMPLE: Patients presenting with axial back pain to academic and private practice orthopedic surgeons and neurosurgeons in a large metropolitan area.

OUTCOME MEASURES: Roland-Morris back disability score (primary outcome), current rating of overall pain severity on a numerical scale, back and leg pain bothersomeness measures, the physical function scale of the short-form 36 version 2 questionnaire, use of medications for pain, work status, emergency department visits, hospitalizations, and further surgery.

METHODS: Patients receiving spine surgery within 6 months of enrollment were designated as the ''surgical treatment'' group and the remainder as ''nonsurgical treatment.'' Outcomes were assessed at 3, 6, 9, and 12 months after enrollment.

RESULTS: We enrolled 495 patients with discogenic back pain presenting for initial surgical consultation in offices of 16 surgeons. Eighty-six patients (17%) had surgery within 6 months of enrollment. Surgery consisted of instrumented fusion (79%), disc replacement (12%), laminectomy,

FDA device/drug status: Approved (Instrumentation used with spinal fusion surgery).

Author disclosures: ***SKM:*** Grant: NIH/NIAMS/NIA/AHRQ (I, Paid directly to institution/employer). ***RAD:*** Grant: NIH (F, Paid directly to institution/employer); Consulting: UptoDate (A); Board of Directors: Foundation for informed Medical Decision Making (B); Scientific Advisory Board: National Advisory Committee for the Robert Wood Johnson Foundation Physician Faculty Scholars Program (B); Endowments: Kaiser Permanente professorship in Evidence-Based Family Medicine (H, Paid directly to institution/employer); Grants: NIH research grants (I, Paid directly to institution/employer). ***PJH:*** Grant: NIH NIAMS (I, Paid directly to institution/employer). ***JAT:*** Nothing to disclose. ***BIM:*** Grant: NIAMS (Partial salary support, Paid directly to institution/employer); Other: Stryker Corporation (Partial salary support, Paid directly to institution/employer); Consulting: University of Washington and University of Missouri—Kansas City (D); Speaking and/or Teaching Arrangements: Ohio State University (B); Research Support (Investigator Salary): National Bureau of Economic Research (B); Grants: NIA, AHRQ (Partial salary support, Paid directly to institution/employer). ***BAC:*** Nothing to disclose.

The disclosure key can be found on the Table of Contents and at www.TheSpineJournalOnline.com.

* Corresponding author. Department of Orthopaedics, Dartmouth-Hitchcock Medical Center, Lebanon, NH 03756, USA. Tel.: (603) 653-6090; fax: (603) 653-3581.

E-mail address: sohail.k.mirza@dartmouth.edu (S.K. Mirza)

http://dx.doi.org/10.1016/j.spinee.2013.05.047

or discectomy (9%). Surgical patients reported more severe pain and physical disability at baseline and were more likely to have had prior surgery. Adjusting for baseline differences among groups, surgery showed a limited benefit over nonsurgical treatment of 5.4 points on the modified (23-point) Roland disability questionnaire (primary outcome) 1 year after enrollment. Using a composite definition of success incorporating 30% improvement in the Roland score, 30% improvement in pain, no opioid pain medication use, and working (if relevant), the 1-year success rate was 33% for surgery and 15% for nonsurgical treatment. The rate of reoperation was 11% in the surgical group; the rate of surgery after treatment designation in the nonsurgical group was 6% at 12 months after enrollment.

CONCLUSIONS: The surgical group showed greater improvement at 1 year compared with the nonsurgical group, although the composite success rate for both treatment groups was only fair. The results should be interpreted cautiously because outcomes are short term, and treatment was not randomly assigned. Only 5% of nonsurgical patients received cognitive behavior therapy. Nonsurgical treatment that patients received was variable and mostly not compliant with major guidelines.

Keywords: Discogenic back pain; Surgery; Fusion; Nonsurgical treatment; Outcome

Introduction

Back pain associated with disc degeneration may be the most controversial subject in spine care and perhaps the one most in need of further clinical research. Because chronic back pain is so disabling and so common [1,2], a large population of vulnerable patients is yearning for any promise of relief. These patients are attracted to an expanding range of costly diagnostic and therapeutic interventions [2,3]. They may be unaware of the controversy surrounding many aspects of a diagnosis of ''discogenic back pain [4].'' This diagnosis does not have well-established criteria. It generally refers to back pain in patients without radicular symptoms and without structural abnormalities other than lumbar disc degeneration. Professional societies provide some guidelines on this condition [5–8], but no consensus exists on whether or how to follow the recommendations [9–12]. It remains unclear, for example, whether there is a localized peripheral generator of back pain [13] (ie, the intervertebral disc), whether imaging studies such as magnetic resonance imaging (MRI) [14] or diagnostic tests such as discography [15] can distinguish painful from nonpainful discs, or whether such a pain source can be eliminated by nonsurgical treatment, excision, fusion, or artificial disc replacement.

Lack of consensus regarding the efficacy of lumbar spinal fusion for discogenic back pain is particularly troubling because four European randomized trials have compared fusion with nonsurgical care [16–19]. Fusion showed a small benefit for back disability compared with nonstandardized nonsurgical care but roughly similar benefit compared with intensive rehabilitation incorporating cognitive behavior therapy [20]. The Food and Drug Administration investigational device exemption artificial disc approval studies showed disc replacement to have less than 60% success rate for a composite outcome and even lower success for the comparator lumbar fusion [21–23]. In the context of rising use and costs of lumbar fusion [24,25], these results have invited scrutiny from payers [26,27]. Independent evidence reviews commissioned by the Centers for Medicare and Medicaid Services Coverage and Advisory Committee [28] and the Washington Healthcare Technology Assessment Program [29] concluded that lumbar fusion for degenerative disc disease lacked sufficient evidence of efficacy and safety to justify unconditional coverage, and private payers have reached similar conclusions for artificial disc replacement [30]. In contrast, experienced surgeons claim that ''properly selected patients'' have successful outcomes with surgery, without specifying the selection criteria [31].

The randomized trials were performed in countries with nationalized health care. We considered the possibility that patients in United States with discogenic back pain who receive surgery in a community-based practice setting may differ in important measurable characteristics from those who do not receive surgery; they may also have different outcomes. They may have different expectations, different access to services, and different social support options. Also, in a controlled trial, the strict eligibility criteria and standardized interventions may not reflect real-world practice. An observational study to describe the demographics, baseline features, and treatment use among US patients with discogenic back pain would add new and complementary information to these trials. We designed an observational prospective cohort study to address these hypotheses [32]. Our goals were to select patients presenting for initial surgical consultation for axial back pain associated with disc degeneration, select those patients the treating surgeon identified as having discogenic back pain, identify baseline characteristics associated with receiving surgery, and compare outcomes of surgical versus nonsurgical treatment. In contrast with a controlled trial, we did not interfere with the diagnosis and care process; we simply did our best to record what was done and how the patients' pain and function changed.

Methods

Study design

We conducted a prospective cohort study of patients with axial back pain seeking surgical consultation. The detailed study protocol has been previously published [32]. To obtain a representative sample of community practice, we enrolled patients from five sites: a county hospital, an academic medical center, and three community hospitals. Orthopedic surgeons (n=12) and neurosurgeons (n=4) participated. The study protocol was approved by the institutional review boards of University of Washington and participating community hospitals, and all study participants provided written informed consent. Study participants were followed for 12 months after enrollment.

Patient selection

Our goal was to identify patients with discogenic back pain consulting a surgeon for the first time to discuss surgical treatment options. Because no established criteria exist for this diagnosis, we relied on the judgment of the treating surgeon regarding interpretation of the patient's clinical presentation and imaging studies. We required that patients have low back pain as the primary symptom and an MRI scan confirming disc degeneration at only one or two lumbar discs. Enrollment criteria did not restrict the MRI to be within any specific time interval. The diagnosis of discogenic back pain was established by the surgeon. Investigators reviewed the radiologist's report to confirm that no specific structural abnormalities were reported by the radiologist and degeneration was limited to one or two levels. Because discography remains controversial [33], we did not require discography as an enrollment criterion. We required symptom duration of at least 6 months, as surgeons rarely consider surgery for discogenic pain of shorter duration. We did not ask whether the pain varied during the duration interval (eg, pain everyday or most days).

Research coordinators screened records of all patients presenting with back pain at the participating sites. Coordinators used a prespecified list of exclusion criteria. Patients could be excluded at screening, baseline interview, or subsequent confirmation of enrollment criteria and consent. We excluded patients with neurological deficit or predominantly nerve root symptoms, motor deficits, abnormal electrodiagnostic studies (if performed), structural spine deformity such as stenosis or spondylolisthesis (>Grade 1), inflammatory disease, spinal malignancy, instability on radiographs (if performed), pregnancy, other specific causes for back pain, or severe comorbidity that would contraindicate surgery. We also excluded patients older than 65 years because of the high prevalence of multilevel disc degeneration and spinal stenosis in this age group.

Definition of treatment groups

In this observational study, we did not specify the treatments patients received. At each assessment, we asked patients about treatments they had received. We also recorded treatments listed in their medical records. Patients who underwent surgery within six calendar months after study enrollment date were designated as ''surgical'', and those who did not were designated as ''nonsurgical,'' even if they had surgery later in the study period. Patients could undergo multiple treatments concurrently during the study period. Any additional or concurrent nonoperative treatments received by surgical patients during the initial 6 months after enrollment were considered co-interventions.

Outcome measures

Our primary outcome was the modified Roland-Morris back disability score as measured by the number of ''yes'' responses to 23 statements describing activity limitations related to back pain [34]. Higher scores indicate greater disability. A five-point or 30% reduction from baseline score [35,36] is considered the minimal clinically important change in this score.

Secondary outcome measures included patient current rating of overall pain severity on a numerical scale of 0=''no pain'' to 10=''worst possible pain,'' back and leg pain bothersomeness measures [37–39], the physical function scale of the short-form 36 version 2 (SF-36v2) questionnaire [40], use of medications for pain, and work status.

We also examined success rate using a composite definition of success: 30% improvement from baseline in the Roland score, 30% improvement from baseline in current pain rating, no opioid medication use within the past 3 months, and working (for patients for whom work was relevant, that is, not retired, working in the home, or receiving disability compensation before surgery).

Outcomes were assessed in person at baseline and by telephone interviews or mailed questionnaires at 3, 6, 9, and 12 months after enrollment. Interviewers were blinded to the subject's treatment group designation. Study participants were considered lost for a particular follow-up after a minimum of 12 dispersed unsuccessful telephone interview attempts and no response to 3 sequential mailings during the follow-up time window.

Baseline measures

At baseline, we assessed a variety of patient characteristics to describe the sample. These included patient sociodemographic characteristics, history of symptoms and treatments, work status, work disability compensation status, and litigation [32]. We recorded medical comorbidity using a questionnaire based on the Charlson comorbidity index [41] and asked about smoking, alcohol, and drug use [42]. Psychological measures included the symptom checklist 90 depression and somatization scales [43], the

SF-36v2, mental health scale, and the pain catastrophizing scale [44]. Psychological measures were also administered at follow-up during the first year of the study, but patient complaints of questionnaire burden required us to modify our protocol and collect these only at baseline.

Assessment of therapeutic safety

Because patients were recruited from multiple practices in this community-based study, we chose to evaluate therapeutic safety primarily through information obtained by patient interviews. We selected three adverse outcomes common to both surgical and nonsurgical treatment groups: emergency department visits, hospitalizations, and repeat surgery in the surgical group and any surgery after the treatment group designation period (first 6 months after enrollment) for the nonsurgical group. These measures were ascertained uniformly for both groups through patient interviews. We reviewed operation reports and hospital discharge summaries of the surgical patients. We also performed detailed safety surveillance of surgical patients at two hospitals [45], but we did not have sufficient study personnel for this type of direct hospitalization surveillance of all study participants.

Statistical analysis

Our sample size calculations indicated that we needed approximately 95 patients in each treatment arm to detect a 3.0-point difference in the Roland score (primary outcome), based on a standard deviation of 7.37 from the Maine Lumbar Spine Study [38,46], two-sided alpha=0.05, and power=0.80 [32]. Because observational analyses have a strong potential for meaningful biases, we present confidence levels and sensitivity analyses for observed differences for clinically important baseline features and outcome measures.

To compare the primary outcome (1-year modified Roland score) in the two treatment groups, we used a linear mixed-effects regression model with treatment status entered as an independent variable. The model used available information to account for missing data and patients lost to follow-up. We included random effects to account for correlated measures collected on the same individuals over time with unstructured covariance. Using bivariate and multivariate logistic regression models, we examined the associations between baseline characteristics and having surgery to assess the risk of confounding in this nonrandomized study. We adjusted the linear mixed-effects model for potentially confounding baseline factors that were associated with receipt of surgery.

The inclusion of covariates for the linear mixed model using $p<.05$ for prediction of surgery may be too strong and exclude some other factors that are indeed associated with surgery, for which adjustment could impact estimates. Therefore, as a sensitivity measure for the primary outcome, we also performed an alternative analysis using a stepwise selection logistic regression method that included baseline variables for outcomes (Roland, SF-36 [eight domains], catastrophizing, depression, somatization, helplessness and rumination, back and leg pain), patient characteristics (gender, education, race, age, work status, back pain duration, body mass index, comorbidity, smoking, alcohol, marital status, disability, and lawyer help), and resources (days cut down on activity, bed days, missed work, prescriptions). We set the probability to remove at $>.20$. The following variables remained in the model: SF-36 vitality, social function, role physical, and general health domains, leg weakness, work status, cut down on activities, somatization, catastrophizing, helplessness, recruitment site, symptom duration, and previous surgery.

The number of interventions received during the treatment period was compared among groups using a trend test (ptrend in Stata). Proportions for the categorical outcome measures (safety and the composite definition of success) were compared using logistic regression with adjustment for covariates. All analyses were performed using Stata version 11.0 (StataCorp LP, College Station, TX, USA). All tests were two sided and p values $<.05$ were considered to be significant.

Results

Study participants

Fig. 1 shows the study flow. We screened 7,344 patients who had back pain and were referred to the study. Of these, 495 met the study inclusion criteria and agreed to participate. Reasons for exclusions are listed in Table 1. The most frequent exclusions were for age greater than 65 years (922, 13.5%), presence of radiculopathy (848, 12.4%), problem not related to lumbar spine (827, 12.1%), and prior fusion (783, 11.4%). Our sample size target was 95 in each treatment group, but after nearly 5 years of enrollment, only 86 (17%) of 495 study participants received surgery within 6 months of the enrollment date. Although more patients in the surgical group would have been optimal, we stopped further enrollment because of limits in our project funding.

Twelve-month post-enrollment interviews were completed in 70 (81%) surgical patients and 336 (82%) nonsurgical patients (Fig. 1). Patients lost at the 12-month post-enrollment interview were significantly more likely to have baseline characteristics generally associated with worse outcomes: more prior surgery, more severe pain, greater physical disability, and more worker's compensation claims.

Baseline comparisons of surgical and nonsurgical patients

Surgical and nonsurgical patients were similar in most of the characteristics we measured at baseline, including

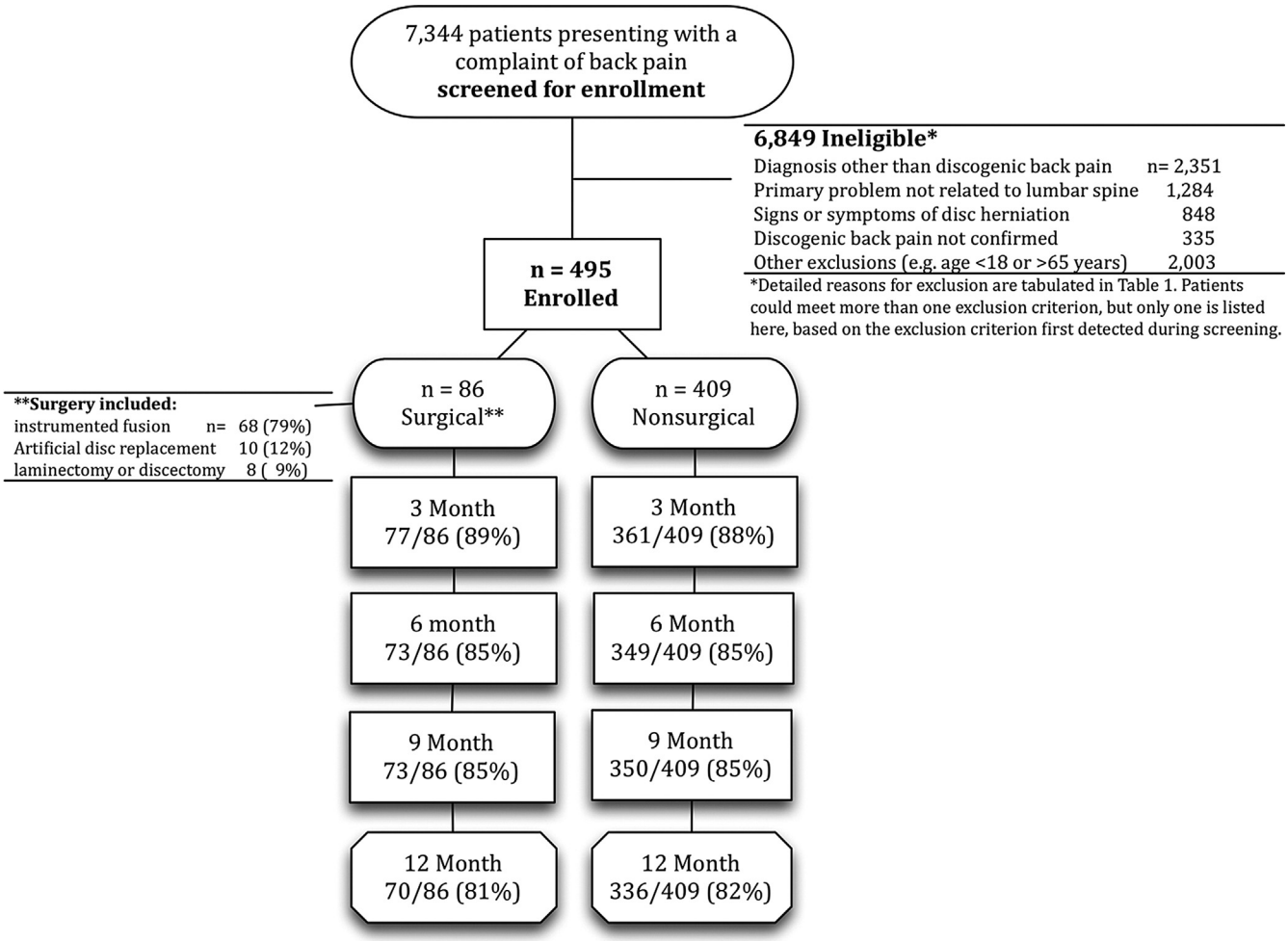

Fig. 1. Flow diagram of patient screening, eligibility, enrollment, and follow-up.

measures of psychological distress (Table 2). However, patient with prior lumbar decompression surgery, greater baseline back and leg pain bothersomeness, and greater back-related physical disability were more likely to receive surgery in the next 6 months (Table 2). In the multivariate model, after adjusting for other important baseline characteristics, prior surgery, greater back-related physical disability, and being seen at a private practice site (as compared with an academically affiliated hospital) were associated with receiving surgery. We also found a trend toward a significantly lower chance of receiving surgery for patients who were smokers and a trend toward receiving surgery for patients with greater leg pain bothersomeness, controlling for other factors (not shown).

There was a wide range of scores on each baseline measure. On average, study participants scored almost 1 standard deviation below (ie, worse than) the general population mean on the SF-36 mental health scale, showed moderate levels of depressive symptoms, and reported multiple nonspecific physical symptoms. Mean pain-related catastrophizing scores were similar to those reported among patients seen in outpatient pain clinics [47,48].

Nature of the surgical treatments and co-interventions

Surgical treatment varied, consisting of instrumented fusion in 68 patients (79% of the surgical patients), artificial disc replacement in 10 (12%), and laminectomy or discectomy in 8 (9%). Enrollment criteria were confirmed in these patients; we cannot explain the rationale for decompression surgery. Surgical patients also received multiple nonsurgical co-interventions during the first 6 months after enrollment (Table 3). The number of nonoperative treatments was greater in the surgical group than in the nonsurgical group within the first 6 months ($p=.001$, Table 3). Surgery was performed at a mean of 2.4 months after enrollment.

Nature of nonsurgical treatments

Nonsurgical treatments received by patients during the first 6 months of the study were similar to the treatments patients reported receiving before enrollment, although rates of use were lower (Table 3). A substantial minority of patients (13%) reported receiving no treatment during

Table 1
Patients excluded after screening assessment (n=6,849)*

Reason for exclusion	n	%
Signs or symptoms of disc herniation		
Disc extrusion/protrusion/bulge	447	6.5
Pain radiating below knee	163	2.4
Lumbar radiculopathy	149	2.2
Leg pain more severe than back pain	78	1.1
Nerve root impingement	6	0.1
Abnormal electrodiagnostic test	5	0.1
Discogenic back pain not confirmed		
Not confirmed with imaging	204	3.0
Not discogenic by surgeon assessment	115	1.7
Not referred by surgeon for study for other reasons	16	0.2
Primary problem not related to lumbar spine		
Neck	827	12.1
Problem not back (hip, leg, ankle)	212	3.1
Thoracic spine	207	3.0
Sacral	28	0.4
Shoulder	10	0.1
Diagnosis other than discogenic back pain		
Prior fusion	783	11.4
Fracture	378	5.5
Planned surgery not fusion or disc replacement	421	6.1
Scoliosis	239	3.5
Disc degeneration at more than two levels	223	3.3
Lumbar spinal stenosis	147	2.1
Lumbar spondylolisthesis >25% (Grade 1)	86	1.3
Previous multilevel laminectomy	22	0.3
Lesion or cyst	13	0.2
Malignancy or infection	15	0.2
Kyphosis	9	0.1
Developmental deformity	7	0.1
Spondylolysis	6	0.1
Neuropathy	1	<0.1
Instability on flexion-extension X-rays	1	<0.1
Other exclusions		
Age >65 y	922	13.5
Postoperative/wound management	627	9.2
Age <18 y	332	4.8
Did not speak English	77	1.1
Pain duration <6 mo	25	0.4
Medical comorbidity contraindicating surgery	14	0.2
Pregnancy	5	0.1
No phone	1	<0.1

* Patients could meet more than one exclusion criterion, but only one is listed here, based on the exclusion criterion first detected during screening.

the first 6 months after enrollment, the treatment designation period for the study.

Outcomes of surgical and nonsurgical treatment

The primary outcome, back-specific disability, showed advantage for surgery (Fig. 2). Based on linear mixed models adjusting for baseline measures associated with receipt of surgery and for loss to follow-up, patients who received surgery in the first 6 months of the study, on average, had Roland score that was 5.4 points (95% confidence interval 3.9–6.9, p<.001) lower than those of patients in the nonsurgical group at 1 year (6–12 months after surgery, mean 9.6 months after surgery).

Alternative analysis using a stepwise model suggested a slightly greater improvement in the surgical group than the original model, but this analysis was not as parsimonious, and lost some subjects because of missing baseline variables. The Roland score for the surgical group at 12 months was 6.07 points lower (95% confidence interval 4.4–7.8) compared with the nonsurgical group (p<.001), overlapping with the estimate from the original model. However, the stepwise model included many more terms and only 399 subjects because of missingness; the original model included 495 subjects and fewer parameters. The statistical tests between surgery and nonsurgery were not otherwise different, and the conclusions remain the same.

Secondary outcomes also showed advantage for surgery, including overall pain intensity rating (Fig. 2), composite measures of success (Table 4), and other physical and mental health measures (Table 5). Patients showed variable improvement on both the overall pain rating and Roland back disability scales, but the surgical group had greater improvement than the nonsurgical group at all potential cut-off thresholds for defining success (Fig. 3). Using a composite measure of success defined as 30% improvement from baseline in the Roland score, 30% improvement from baseline in pain rating, return to work for eligible workers, and no opioid pain medication use, the 1-year post-enrollment success rate was 33% in the surgery group and 15% in the nonsurgical group (p<.001, Table 4). Some patients did not improve: approximately 25% of nonsurgical and 15% of surgical patients reported worse function and increased pain at 12 months after enrollment compared with baseline (Fig. 3).

Patient-reported measures of safety

Between 6 and 12 months after enrollment, emergency department visits occurred with similar frequency in the two treatment groups: 5/76 (7%) of surgical patients and 41/366 (11%) nonsurgical patients (p=.23). Overnight hospitalizations also occurred with similar frequency in both groups: 1/76 (1%) surgical and 13/366 (4%) nonsurgical (p=.31). Repeat surgery occurred in 8/76 (11%) of surgical patients. Also, 22/366 (6%) of patients in the nonsurgical group received surgery between 6 and 12 months after enrollment.

Discussion

This community-based comparative effectiveness study showed only fair outcomes for both surgical and nonsurgical treatments of discogenic back pain. Patients with chronic back pain who seek surgical consultation and are found to have discogenic back pain presented, on average, with moderate levels of pain, physical disability, and psychological distress. When assessed 12 months after enrollment, patients who had surgery combined with nonsurgical

Table 2
Baseline comparison of patients who received surgery within 6 months of enrollment and those who did not

Factor	Level	Nonsurgical (n=409)	Surgical (n=86)	p Value among groups
Demographics				
Sex, %	Male	48	45	.67
	Female	52	55	
Education, %	High school or less	27	25	.62
	Some college	40	46	
	College degree	32	29	
Race, %	White	84	87	.45
	Other	16	13	
Age (y), mean (SD)		42.7 (9.3)	42.1 (8.7)	.60
Work status, %	Working full or part time	50	47	.06
	On leave, unemployed	24	29	
	Homemaker, student, retired	9	15	
	Disabled	17	8	
Clinical characteristics				
Duration, %	<12 mo	17	14	.67
	1–5 y	47	52	
	5+ y	35	34	
Previous surgery, %	Yes	21	36	.004
BMI, %	<24.9	36	29	.51
	25.9–29.9	36	41	
	30.0+	28	29	
Comorbidity in Charlson index, %	Any	36	41	.43
Smoker, %	Yes	29	21	.12
Excessive alcohol/drug screen, %	Positive*	12	13	.83
Setting				
Enrollment site, %	County medical center	37	26	.08
	Academic affiliate	22	22	
	Private hospital affiliate	41	52	
Baseline physical health measures†				
Roland score, mean (SD)	0–23	15.9 (5.4)	17.7 (4.2)	.003
SF-36v2 physical function, mean (SD)	Norm based (0–100)	32.9 (10.7)	28.8 (9.2)	<.001
Overall pain rating, mean (SD)	0–10	6.1 (2.3)	6.5 (1.9)	.10
Back pain bothersomeness, mean (SD)	1–5	4.0 (0.9)	4.4 (0.7)	.03
Leg pain bothersomeness, mean (SD)	1–5	3.0 (1.4)	3.5 (1.3)	.01
Baseline mental health measures†				
SF-36v2 mental health, mean (SD)	Norm based (0–100)	42.2 (12.1)	42.2 (12.3)	.96
Pain catastrophizing, mean (SD)	Raw score (0–52)	23.2 (13.4)	25.3 (13.3)	.20
SCL-90 somatization, mean (SD)	Raw score (0–4)	1.2 (0.6)	1.1 (0.6)	.80
SCL-90 depression, mean (SD)	Raw score (0–4)	1.1 (0.9)	1.1 (0.9)	.58

BMI, body mass index; SF-36v2, short-form 36 version 2; SD, standard deviation; SCL-90, symptom checklist 90.

* Positive alcohol/drug screen was ''yes'' to either of two questions: excessive use within the past year or desire to cut down.

† Higher scores indicate worse pain and function on all outcome measures except the SF-36 physical function and mental health scales, where higher scores indicate better function.

co-interventions showed modest but significantly greater improvement in self-reported disability, back pain, generic physical function, and composite success measures compared with patients who had continuation of unstructured nonsurgical care. Surgical patients also concurrently received more intensive cotreatments than the nonsurgical group. Although surgery combined with various additional nonsurgical treatments showed advantage over nonsurgical treatment alone, only one-third of surgical patients attained a successful result defined by stringent criteria of clinically important improvement in pain and function, no opioid medication use, and return to work for eligible workers. The rate of activity restriction and opioid use was significantly greater in the surgical group along with greater use of corsets and bed rest. Surgery did not reduce the frequency of emergency room visits or overnight hospitalizations, and 11% of patients had repeat surgery within the first year postoperatively.

Patients, on average, showed minimal improvement after 12 months of continued nonsurgical care as currently provided in the United States. Nonsurgical treatments that the patients received varied widely, were used in an unstructured manner, and mostly did not comply with conservative care guidelines. For example, only 5% received cognitive behavior therapy. Outcomes in the nonsurgical group may have improved more if treatment had adhered to recommendations from clinical practice guidelines [6].

The poor prognosis we observed for discogenic back pain has policy implications. The nonsurgical patients in our study provide somewhat of a natural history of discogenic back pain because they received minimal new therapy after enrollment. In contrast with the commonly held view

Table 3
Therapeutic interventions reported at baseline and during the treatment designation period

Intervention	Baseline*			Treatment period†		
	Nonsurgical group (n=409), %	Surgical group (n=86), %	p Value	Nonsurgical group (n=409), %	Surgical group (n=86), %	p Value
Surgery	21	36	<.001	0	100	NA
Exercise	80	85	.340	73	78	.37
Activity restriction	74	80	.322	66	88	<.001
Nonsteroidal anti-inflammatory medications	90	91	.695	65	47	.27
Opioid pain medications	80	88	.072	64	89	<.001
Physical therapy/occupational therapy	88	95	.073	52	56	.61
Bed rest	68	73	.317	52	66	.018
Massage	55	61	.319	33	28	.33
Brace or corset	42	46	.544	30	50	.001
Spinal injections	63	81	.002	29	26	.66
Pain program	17	29	.010	17	13	.34
Chiropractic care	45	52	.272	15	10	.23
Transcutaneous electrical nerve stimulation	25	32	.244	15	19	.39
Ultrasound	34	44	.082	12	20	.054
Acupuncture	21	27	.270	7	6	.74
Cognitive-behavioral therapy	5	8	.457	6	4	.48
Intradiscal electrothermal coagulation	8	11	.310	4	3	.57
Number of co-interventions: different nonsurgical interventions received during the treatment designation period only						
0				13	8	.001
1–3				41	29	
4–6				39	54	
>6				7	9	

NA, not applicable.

* Subjects were asked if they received the treatment at any time in the past before enrollment.

† The initial 6 months after enrollment was the time interval during which treatment group designation was determined. Those patients who received surgery during that period comprised the surgical group. Values indicate any endorsement of that treatment at either the 3- or the 6-month post-enrollment assessment.

that nonspecific back pain has a benign course, those who sought surgical consultation had marked pain and functional limitations at enrollment and remained essentially unchanged during 12 subsequent months of surveillance. The fundamental concepts underlying the mechanism of pain in these patients, the treatment options offered to them, and policies governing these treatments need re-evaluation. New diagnostic and surgical technologies are readily

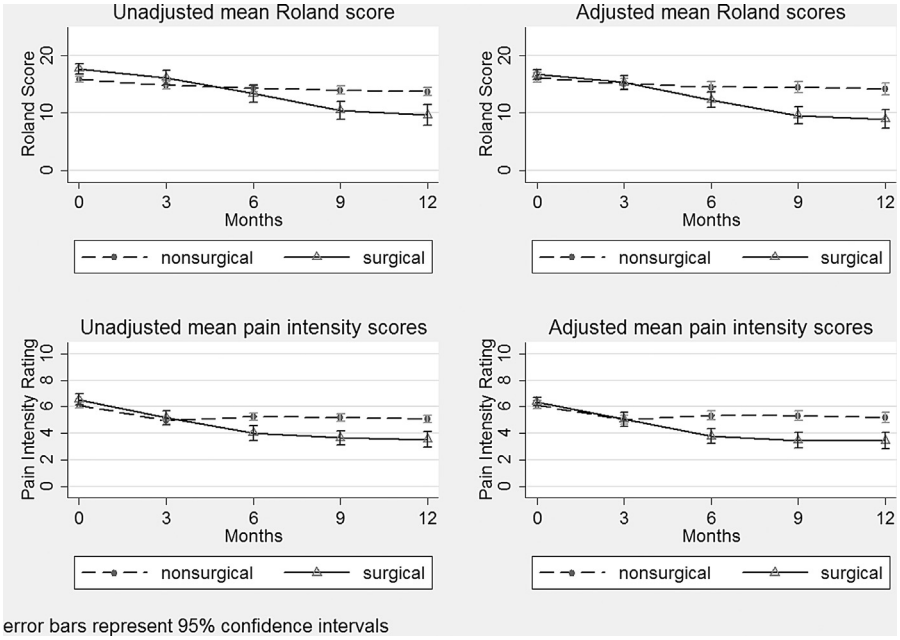

Fig. 2. Adjusted and unadjusted Roland disability (primary outcome) and back pain intensity at each follow-up time point measured from enrollment date.

Table 4
Success rates at 1 year after enrollment for various definitions of success

Criteria for success at 1 year	Nonsurgical group (%)	Surgical group (%)	p Value*
30% improvement (from baseline) in pain intensity	35	71	<.001
30% improvement (from baseline) in Roland score	25	57	<.001
Working (among those for whom work is relevant)†	57	59	.92
No opioid pain medications in the past 3 mo	51	47	.51
30% improvement in pain intensity and 30% improvement in Roland	19	51	<.001
30% improvement in pain intensity, 30% improvement in Roland, and working at 12 months†	18	46	<.001
30% improvement in pain intensity, 30% improvement in Roland, working at 12 months†, and no opioid pain medications in the past 3 mo	15	33	<.001

* p Values based on logistic regression controlling for age, sex, education, previous surgery, alcohol use, smoking, body mass index, race, work status, Charlson comorbidity index, overall pain intensity, back pain bothersome, leg pain bothersome, symptom duration, and study recruitment site.

† Excluding those patients who at baseline reported being a student, homemaker, retired, or on permanent disability.

available in community practice, but comprehensive rehabilitation and cognitive behavior therapy are difficult to find and frequently not covered by insurance programs.

We conducted an observational study to obtain a ''real-world'' or pragmatic view of community-based practice and outcomes for discogenic back pain. We acknowledge that observational studies comparing treatment effectiveness have many limitations; results should be interpreted with caution. Despite statistical adjustments, unmeasured confounding factors persist and bias observed associations. The population chosen for this study was likely biased in favor of surgery because it had already made the decision to seek a surgical consultation. Only a small fraction of patients with back pain seeking surgical opinion were judged by the treating surgeon to have discogenic back pain (ie, chronic back pain, disc degeneration at only one or two lumbar levels, and no other focal abnormalities).

In contrast with the standardization often imposed in randomized trials, this study shows that under natural conditions, patients with chronic back pain often mix multiple treatment interventions concurrently. Surgical patients simultaneously received multiple nonsurgical co-interventions. In fact, patients who underwent surgery received more nonoperative treatments than did patients in the nonsurgical group. We enrolled patients after the studies by Fritzell et al. [19] and Brox et al. [17] were published. During orientation of surgeons participating in our study, we reviewed that data showing intensive rehabilitation incorporating cognitive behavior therapy were just as effective as surgery [20]. Despite availability of this knowledge, the nonsurgical care received by patients in our study remained haphazard.

Our study reports early short-term results. Follow-up for surgical patients averaged 9.6 months postoperatively. Lumbar arthrodesis procedures can require that duration for healing, placebo effects may be particularly strong in the early postoperative period [49], and the advantage for surgery may diminish with time. Surgical advantage was greatest at 1 year and diminished by 2 years in a Swedish randomized trial comparing lumbar fusion to nonoperative care [19]; outcomes for surgical and nonsurgical groups were similar at 5-year follow-up [50]. A small Japanese trial also showed early advantage for surgery [51].

Although it is widely accepted that back pain in some patients may be caused by lumbar disc degeneration, the diagnosis of discogenic back pain lacks a firm biological basis and clear clinical description. It is uncertain whether the structural and physiological intervertebral disc changes associated with aging alone can be distinctly separated from changes that cause low back pain. Because disc degeneration is almost ubiquitous beyond the age of 50 years and because back pain is very common, this ambiguity has important clinical implications. Some physicians believe that individual discs can be identified as sources of pain in individual patients and infer that surgery to immobilize or replace the disc will help eliminate back pain [52]. Others believe it is nearly impossible to identify specific discs as a cause of pain in individual patients [52]. A Combined Task Force of the North American Spine Society, American Society of Spine Radiology, and American Society of Neuroradiology for Nomenclature and Classification of Lumbar Disc Pathology acknowledged the difficulty of distinguishing disc degeneration from normal aging [53].

Efficacy cannot be proven by nonrandomized studies such as ours [54]. We address group differences by reporting multiple comparisons of baseline measures, interventions, and outcomes. However, confounding by unmeasured factors may account for the associations observed between surgery and outcomes. Surgical patients in our study also received more intensive nonoperative co-interventions compared with the nonsurgical cohort, none of whom received structured state-of-the-art rehabilitation, and some received no treatment. However, if patient preferences, the treating surgeon's patient selection biases, co-interventions, and other unmeasured factors associated with receiving surgery are considered as a bundled package (ie, ''use effectiveness'' [55]), surgical patients had better—but still poor—outcomes compared with continued unstructured nonsurgical care.

Our study shows that patients whom a spine surgeon labeled as having discogenic back pain have severe functional limitations at baseline, have multiple comorbidities, and receive multiple concurrent treatments in the course of usual care. In contrast with the general perception that patients with nonspecific back pain improve with minimal treatment, we found these patients continue to have severe

Table 5
Linear mixed models to estimate treatment effects on primary and secondary outcomes for subjects who received surgery compared with those who received only nonsurgical treatment*

Outcome	Time point (mo)	Nonsurgical mean (95% CI)	Surgical mean (95% CI)	p Value
Physical health measures				
Roland disability index	3	15.1 (14.6–15.5)	15.3 (14.3–16.3)	.52
	6	14.5 (14.0–15.0)	12.3 (11.2–13.4)	<.001
	9	14.4 (13.9–15.0)	9.6 (8.4–10.8)	<.001
	12	14.1 (13.5–14.8)	8.9 (7.5–10.3)	<.001
Overall pain rating	3	5.0 (4.8–5.2)	5.1 (4.6–5.5)	.43
	6	5.3 (5.1–5.5)	3.8 (3.3–4.3)	<.001
	9	5.3 (5.1–5.6)	3.5 (2.9–4.0)	<.001
	12	5.2 (4.9–5.4)	3.4 (2.9–4.0)	<.001
SF-36v2 physical function	3	34.6 (33.8–35.3)	35.0 (33.2–36.7)	.38
	6	35.0 (34.1–35.8)	37.5 (35.5–39.4)	.005
	9	34.9 (34.0–35.8)	43.8 (41.7–45.9)	<.001
	12	35.3 (34.3–36.3)	43.3 (40.9–45.6)	<.001
Leg pain bothersomeness	3	2.8 (2.7–2.9)	2.6 (2.4–2.9)	.24
	6	2.7 (2.6–2.9)	1.7 (1.5–2.0)	<.001
	9	2.8 (2.6–2.9)	1.5 (1.2–1.8)	<.001
	12	2.7 (2.5–2.8)	2.0 (1.6–2.3)	<.001
Back pain bothersomeness	3	3.8 (3.7–3.9)	3.5 (3.3–3.7)	.052
	6	3.6 (3.5–3.7)	2.9 (2.7–3.2)	<.001
	9	3.6 (3.5–3.8)	2.7 (2.4–2.9)	<.001
	12	3.4 (3.3–3.6)	2.8 (2.5–3.0)	<.001
Mental health measures				
SF-36v2 mental health	3	43.9 (42.8–45.1)	45.5 (43.1–47.9)	.26
	6	43.5 (42.3–44.7)	47.0 (44.4–49.6)	.017
	9	43.1 (41.8–44.3)	50.0 (47.2–52.8)	<.001
	12	43.8 (42.4–45.1)	49.1 (46.1–52.2)	.002
Pain catastrophizing	3	21.2 (20.1–22.4)	18.8 (16.3–21.2)	.087
	6	20.7 (18.4–23.0)	11.5 (6.2–16.8)	.002
	9	20.5 (17.7–23.3)	9.6 (2.7–16.5)	.004
	12	20.1 (18.9–21.4)	12.1 (9.2–15.0)	<.001
SCL-90 somatization	3	1.06 (1.00–1.12)	1.15 (1.02–1.29)	.200
	6	1.14 (1.01–1.27)	0.76 (0.45–1.07)	.025
	9	1.27 (1.10–1.44)	0.95 (0.50–1.40)	.19
	12	1.14 (0.89–1.39)	1.05 (0.44–1.65)	.78
SCL-90 depression	3	1.01 (0.92–1.09)	0.85 (0.67–1.04)	.13
	6	1.00 (0.84–1.15)	0.81 (0.45–1.16)	.32
	9	1.26 (1.07–1.45)	0.70 (0.24–1.17)	.029
	12	1.10 (1.01–1.19)	0.70 (0.50–0.90)	<.001

CI, confidence interval; SF-36v2, short-form 36 version 2; SCL-90, symptom checklist 90.

* Controlling for age, sex, education, previous surgery, alcohol use, smoking, body mass index, race, work status, Charlson comorbidity index, overall pain intensity, back pain bothersome, leg pain bothersome, symptom duration, and study recruitment site.

pain and functional limitations. The degree of improvement we observed in surgical patients was marginal despite intensive concurrent treatments. These findings are relevant to guiding policy and practice for this patient population in the United States. Facilitating effective treatments based on trustworthy clinical practice guidelines may yield large benefits at the population level.

Access to patient-level data

The Center for Surgical Innovation at Dartmouth (CSI) website (http://dartmouth-hitchcock.org/csi) will provide de-identified study data to any researchers interested in evaluating new hypotheses or confirming hypotheses within this paper, or professors wishing to use the database for classroom purposes. The website provides the data dictionary for the study cohort, an accompanying document that describes the data structure, and general guidelines regarding how to request the data. The Methodology Core for the Dartmouth/NIH NIAMS Multidisciplinary Clinical Research Center for Musculoskeletal Diseases (MCRC) will evaluate submitted applications and, when necessary, work with the applicants to clarify the project aims. Upon approval, the MCRC will provide the applicant with a secure data set that includes all of the variables the researchers and the MCRC evaluation committee determine are necessary and a data entry program in the statistical package (SAS, STATA, SPSS, or R) selected by users. The CSI website will track the number of applicant requests, projects that have been summited, projects that have been approved, the project title and involved researchers for approved

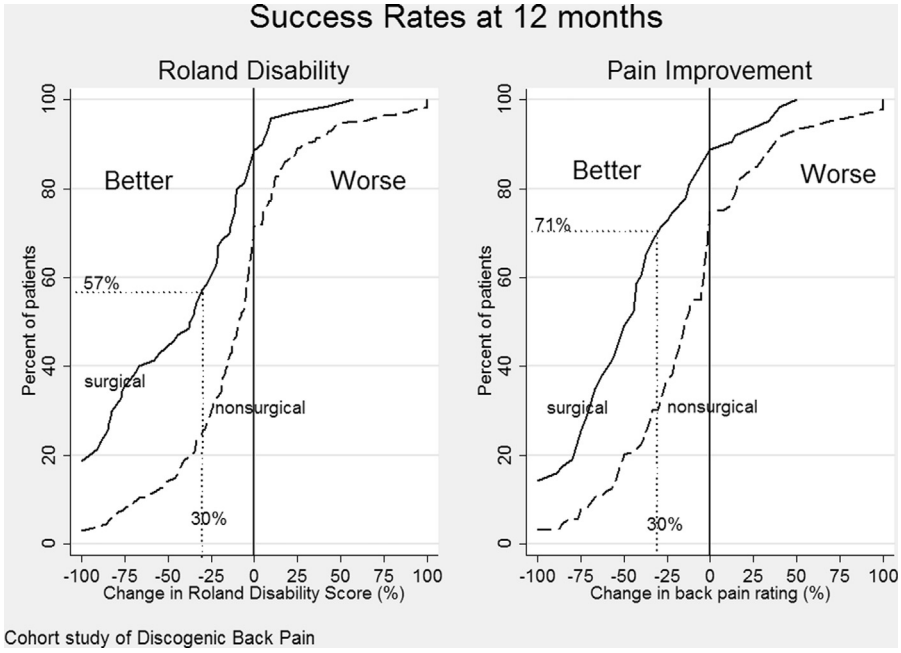

Fig. 3. Cumulative proportion of responders for the 12-month change (Change % = [(Final - Baseline) / Baseline] * 100) in the Roland disability score (primary outcome) and overall pain intensity rating over the entire range of possible cut-off points for defining success. The graphs allow comparison of the treatment groups at any response level that a surgeon or patient may consider important. For example, if 30% or more improvement from baseline is considered as the criterion for success, 57% of surgical patients achieved it on the Roland Scale compared with 25% of nonsurgical patients. For overall pain (0–10 numerical rating scale), 71% of the surgical patients and 30% of the nonsurgical patients achieved 30% or more improvement compared with baseline scores.

projects, and, once completed, a structured abstract of the project.

Acknowledgments

This work was supported by grants from the National Institutes of Health/National Institute of Arthritis, Musculoskeletal, and Skin Disorders 5K23AR48979, 5P60-AR48093, and RC1AG036268. It was also supported in part by the Spine End-Results Research Fund at the University of Washington established through a gift from Synthes Spine (Paoli, PA, USA), a manufacturer of spinal surgery implants. The study was also funded by grants NIH/NIAMS 60AR 048093 and K23AR048979 and NIH/NIA RC1AG036268. We thank Dr Kevin Spratt, in the methodology core of the Dartmouth NIH/NIAMS MCRC, for managing access to study data.

We are grateful to the following surgeons and their office staff for referring patients to the study and providing valuable feedback: Bone and Joint Center, University of Washington: Dheera Ananthakrishnan, Rick Bransford, Jens Chapman, Todd Jarosz, and Ted Wagner; Evergreen Hospital: Reginald Knight; Harborview Medical Center: Carlo Bellabarba, Jens Chapman, Andrew Dailey, and Alex West; Providence Medical Center: David Hanscom and Paul Schwaegler; Swedish Medical Center: Jeff Garr and Jay Williams; University of Washington Medical Center: Christopher Shaffrey and Trent Tredway.

References

[1] Deyo RA, Mirza SK, Martin BI. Back pain prevalence and visit rates: estimates from U.S. national surveys, 2002. Spine 2006;31:2724–7.

[2] Martin BI, Deyo RA, Mirza SK, et al. Expenditures and health status among adults with back and neck problems. JAMA 2008;299:656–64.

[3] Ivanova JI, Birnbaum HG, Schiller M, et al. Real-world practice patterns, health-care utilization, and costs in patients with low back pain: the long road to guideline-concordant care. Spine J 2011;11:622–32.

[4] Groopman J. A knife in the back: is surgery the best approach to chronic back pain? New York, NY: The New Yorker, 2002.

[5] Chou R, Loeser JD, Owens DK, et al. Interventional therapies, surgery, and interdisciplinary rehabilitation for low back pain: an evidence-based clinical practice guideline from the American Pain Society. Spine 2009;34:1066–77.

[6] Chou R, Atlas SJ, Stanos SP, Rosenquist RW. Nonsurgical interventional therapies for low back pain: a review of the evidence for an American Pain Society clinical practice guideline. Spine 2009;34:1078–93.

[7] Chou R, Baisden J, Carragee EJ, et al. Surgery for low back pain: a review of the evidence for an American Pain Society Clinical Practice Guideline. Spine 2009;34:1094–109.

[8] Resnick DK. Evidence-based guidelines for the performance of lumbar fusion. Clin Neurosurg 2006;53:279–84.

[9] Chou R. Letter to the editor: Re: Chou R, Loeser JD, Owens DK, et al. Interventional therapies, surgery and interdisciplinary rehabilitation for low back pain: an evidence-based clinical practice guideline from the American Pain Society. Spine 2009;34:1066–77. Spine 2011;36:590.

[10] Datta S, Manchikanti L. Re: Chou R, Loeser JD, Owens DK, et al. Interventional therapies, surgery, and interdisciplinary rehabilitation for low back pain: an evidence-based clinical practice guideline

from the American Pain Society. Spine 2009;34:1066–77. Spine 1826;35. author reply 1826–1827.
[11] Manchikanti L, Datta S, Derby R, et al. A critical review of the American Pain Society clinical practice guidelines for interventional techniques: part 1. Diagnostic interventions. Pain Physician 2010;13:E141–74.
[12] Furlan AD, Pennick V. Limitations of guidelines for low back pain therapy. Nat Rev Rheumatol 2009;5:473–4.
[13] Giesecke T, Gracely RH, Grant MA, et al. Evidence of augmented central pain processing in idiopathic chronic low back pain. Arthritis Rheum 2004;50:613–23.
[14] Carragee E, Alamin T, Cheng I, et al. Are first-time episodes of serious LBP associated with new MRI findings? Spine J 2006;6: 624–35.
[15] Carragee EJ, Alamin TF, Carragee JM. Low-pressure positive discography in subjects asymptomatic of significant low back pain illness. Spine 2006;31:505–9.
[16] Brox JI, Reikeras O, Nygaard O, et al. Lumbar instrumented fusion compared with cognitive intervention and exercises in patients with chronic back pain after previous surgery for disc herniation: a prospective randomized controlled study. Pain 2006;122:145–55.
[17] Brox JI, Sorensen R, Friis A, et al. Randomized clinical trial of lumbar instrumented fusion and cognitive intervention and exercises in patients with chronic low back pain and disc degeneration. Spine 2003;28:1913–21.
[18] Fairbank J, Frost H, Wilson-MacDonald J, et al. Randomised controlled trial to compare surgical stabilisation of the lumbar spine with an intensive rehabilitation programme for patients with chronic low back pain: the MRC spine stabilisation trial. BMJ 2005;330: 1233.
[19] Fritzell P, Hagg O, Wessberg P, Nordwall A. 2001 Volvo Award Winner in Clinical Studies: lumbar fusion versus nonsurgical treatment for chronic low back pain: a multicenter randomized controlled trial from the Swedish Lumbar Spine Study Group. Spine 2001;26:2521–32; discussion 32–34.
[20] Mirza SK, Deyo RA. Systematic review of randomized trials comparing lumbar fusion surgery to nonoperative care for treatment of chronic back pain. Spine 2007;32:816–23.
[21] Blumenthal S, McAfee PC, Guyer RD, et al. A prospective, randomized, multicenter Food and Drug Administration investigational device exemptions study of lumbar total disc replacement with the CHARITE artificial disc versus lumbar fusion: part I: evaluation of clinical outcomes. Spine 2005;30:1565–75; discussion E387–E391.
[22] McAfee PC, Cunningham B, Holsapple G, et al. A prospective, randomized, multicenter Food and Drug Administration investigational device exemption study of lumbar total disc replacement with the CHARITE artificial disc versus lumbar fusion: part II: evaluation of radiographic outcomes and correlation of surgical technique accuracy with clinical outcomes. Spine 2005;30:1576–83; discussion E388–E390.
[23] Zigler J, Delamarter R, Spivak JM, et al. Results of the prospective, randomized, multicenter Food and Drug Administration investigational device exemption study of the ProDisc-L total disc replacement versus circumferential fusion for the treatment of 1-level degenerative disc disease. Spine 2007;32:1155–62; discussion 63.
[24] Deyo RA, Gray DT, Kreuter W, et al. United States trends in lumbar fusion surgery for degenerative conditions. Spine 2005;30:1441–5; discussion 6–7.
[25] Deyo RA, Nachemson A, Mirza SK. Spinal-fusion surgery—the case for restraint. N Engl J Med 2004;350:722–6.
[26] CMC/HHS. Medicare Program; Meeting of the Medicare Coverage Advisory Committee—November 30, 2006. Available at: http://www.cms.gov/medicare-coverage-database/details/medcac-meeting-details.aspx?MEDCACId=37&fromdb=true. Accessed November 27, 2007. Federal Register 2006;71.
[27] Washington State Health Care Authority, Health Technology Assessment, Health Technology Committee. Health Technology Clinical Committee Meeting Schedule November 16, 2007 lumbar fusion and discography. Available at: http://www.hta.hca.wa.gov/documents/adopted_minutes_021508_final.pdf. Accessed November 27, 2007. Washington State Healthcare Authority.
[28] Washington State Healthcare Authority, Health Technology Assessment. Spinal fusion for treatment of degenerative disease affecting the lumbar spine. Rockville, MD: Agency for Healthcare Research and Quality, 2006. Available at: http://www.cms.hhs.gov/determinationprocess/downloads/id41ta.pdf. Accessed November 27, 2007.
[29] Washington State Healthcare Authority, Health Technology Assessment. Spinal fusion and discography for chronic low back pain and uncomplicated lumbar degenerative disc disease. Olympia, WA: Washington Health Technology Assessment. Available at: http://www.hta.hca.wa.gov/documents/spinal_fusion_discography_final_101907.pdf. Accessed November 27, 2007.
[30] BCBS. Artificial lumbar disc replacement. Technol Eval Cent Asses Program Exec Summ 2007;22:1–3.
[31] Errico TJ, Gatchel RJ, Schofferman J, et al. A fair and balanced view of spine fusion surgery. Spine J 2004;4:S129–38.
[32] Deyo RA, Mirza SK, Heagerty PJ, et al. A prospective cohort study of surgical treatment for back pain with degenerated discs; study protocol. BMC Musculoskelet Disord 2005;6:24.
[33] Wichman HJ. Discography: over 50 years of controversy. WMJ 2007;106:27–9.
[34] 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.
[35] Ostelo RW, Deyo RA, Stratford P, et al. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. Spine 2008;33:90–4.
[36] Jordan K, Dunn KM, Lewis M, Croft P. A minimal clinically important difference was derived for the Roland-Morris Disability Questionnaire for low back pain. J Clin Epidemiol 2006;59: 45–52.
[37] Keller RB, Atlas SJ, Singer DE, et al. The Maine Lumbar Spine Study, Part I. Background and concepts. Spine 1996;21:1769–76.
[38] 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 1996;21:1777–86.
[39] 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 94–95.
[40] Ware JE Jr. SF-36 health survey update. Spine 2000;25:3130–9.
[41] Katz JN, Chang LC, Sangha O, et al. Can comorbidity be measured by questionnaire rather than medical record review? Med Care 1996;34:73–84.
[42] Brown RL, Leonard T, Saunders LA, Papasouliotis O. A two-item conjoint screen for alcohol and other drug problems. J Am Board Fam Pract 2001;14:95–106.
[43] Derogatis LR, Rickels K, Rock AF. The SCL-90 and the MMPI: a step in the validation of a new self-report scale. Br J Psychiatry 1976;128:280–9.
[44] Sullivan MJL, Bishop SR, Scott R, Pivik J. The pain catastrophizing scale: development and validation. Psychol Assess 1995;7:524–32.
[45] Mirza SK, Deyo RA, Heagerty PJ, et al. Towards standardized measurement of adverse events in spine surgery: conceptual model and pilot evaluation. BMC Musculoskelet Disord 2006;7:53.
[46] Atlas SJ, Deyo RA, van den Ancker M, et al. The Maine-Seattle back questionnaire: a 12-item disability questionnaire for evaluating patients with lumbar sciatica or stenosis: results of a derivation and validation cohort analysis. Spine 2003;28:1869–76.

[47] Osman A, Barrios FX, Gutierrez PM, et al. The Pain Catastrophizing Scale: further psychometric evaluation with adult samples. J Behav Med 2000;23:351–65.

[48] Sullivan MJ, Stanish W, Waite H, et al. Catastrophizing, pain, and disability in patients with soft-tissue injuries. Pain 1998;77:253–60.

[49] Flum DR. Interpreting surgical trials with subjective outcomes: avoiding UnSPORTsmanlike conduct. JAMA 2006;296:2483–5.

[50] Fritzell P. Abstract. Annual meeting of the International Society for Study of the Lumbar Spine 2004. Porto, Portugal.

[51] Ohtori S, Koshi T, Yamashita M, et al. Surgical versus nonsurgical treatment of selected patients with discogenic low back pain: a small-sized randomized trial. Spine 2011;36:347–54.

[52] Nachemson A, Zdeblick TA, O'Brien JP. Lumbar disc disease with discogenic pain. What surgical treatment is most effective? Spine 1996;21:1835–8.

[53] Fardon DF, Milette PC. Nomenclature and classification of lumbar disc pathology. Recommendations of the combined task Forces of the North American Spine Society, American Society of Spine Radiology, and American Society of Neuroradiology. Spine 2001;26:E93–113.

[54] Weiss NS. Clinical epidemiology: the study of the outcome of illness. 2nd ed. New York, NY: Oxford University Press, 1996.

[55] Meier P. Comment on ''Compliance as an explanatory variable in clinical trials'' by Efron and Feldman. J Am Stat Assoc 1991;86:19–22.