SPINE Volume 35, Number 6, pp 672–683

Appropriateness Criteria for Surgery Improve Clinical Outcomes in Patients With Low Back Pain and/or Sciatica

Nadia Danon-Hersch, MD,* Dino Samartzis, DSc, MSc,† Vincent Wietlisbach, François Porchet, MD,‡ and John-Paul Vader, MD, MPH*

Study Design. Prospective, controlled, observational outcome study using clinical, radiographic, and patient/physician-based questionnaire data, with patient outcomes at 12 months follow-up.

Objective. To validate appropriateness criteria for low back surgery.

Summary of Background Data. Most surgical treatment failures are attributed to poor patient selection, but no widely accepted consensus exists on detailed indications for appropriate surgery.

Methods. Appropriateness criteria for low back surgery have been developed by a multispecialty panel using the RAND appropriateness method. Based on panel criteria, a prospective study compared outcomes of patients appropriately and inappropriately treated at a single institution with 12 months follow-up assessment. Included were patients with low back pain and/or sciatica referred to the neurosurgical department. Information about symptoms, neurologic signs, the health-related quality of life (SF-36), disability status (Roland-Morris), and pain intensity (VAS) was assessed at baseline, at 6 months, and at 12 months follow-up. The appropriateness criteria were administered prospectively to each clinical situation and outside of the clinical setting, with the surgeon and patients blinded to the results of the panel decision. The patients were further stratified into 2 groups: appropriate treatment group (ATG) and inappropriate treatment group (ITG).

Results. Overall, 398 patients completed all forms at 12 months. Treatment was considered appropriate for 365 participants and inappropriate for 33 participants. The mean improvement in the SF-36 physical component score at 12 months was significantly higher in the ATG (mean: 12.3 points) than in the ITG (mean: 6.8 points) ($P = 0.01$), as well as the mean improvement in the SF-36 mental component score (ATG mean: 5.0 points; ITG mean: −0.5 points) ($P = 0.02$). Improvement was also significantly higher in the ATG for the mean VAS back pain (ATG mean: 2.3 points; ITG mean: 0.8 points; $P = 0.02$) and Roland-Morris disability score (ATG mean: 7.7 points; ITG mean: 4.2 points; $P = 0.004$). The ATG also had a higher improvement in mean VAS for sciatica (4.0 points) than the ITG (2.8 points), but the difference was not significant ($P = 0.08$). The SF-36 General Health score declined in both groups after 12 months, however, the decline was worse in the ITG (mean decline: 8.2 points) than in the ATG (mean decline: 1.2 points) ($P = 0.04$). Overall, in comparison to ITG patients, ATG patients had significantly higher improvement at 12 months, both statistically and clinically.

Conclusion. In comparison to previously reported literature, our study is the first to assess the utility of appropriateness criteria for low back surgery at 1-year follow-up with multiple outcome dimensions. Our results confirm the hypothesis that application of appropriateness criteria can significantly improve patient outcomes.

Key words: spine, lumbar, low back surgery, sciatica, appropriateness criteria, RAND method, treatment, quality of care, outcome, health-related, quality of life. **Spine 2010;35:672–683**

The majority of surgical treatment failures in spine surgery are attributed to poor patient selection,[1,2] but no widely accepted consensus exists on detailed indications that regard surgical intervention as "appropriate."[3–5] It is estimated that among all patients with sciatica, only about 10% ultimately undergo surgery.[6] Without question, proper indications in the decision-making armamentarium of the treating physician are essential to dictate appropriate patient management. However, despite decades of controversy regarding the optimal care for the treatment of sciatica in consideration of disc surgery to conventional care, no clear consensus has been reached among health-care practitioners.[7]

The RAND-UCLA appropriateness method is a standardized procedure to develop explicit criteria for the appropriate use of medical interventions combining a systematic review of available evidence and collective expert opinion.[8,9] We used this method to develop detailed and explicit criteria for surgical indications concerning low back pain and/or sciatica patients[10]: a procedure was considered appropriate when "the medical benefit in terms of quality of life or life-span, reduction in pain and improved function were thought to exceed by a sufficiently wide margin, the medical risks of mortality, morbidity, and anxiety caused by the procedure."[9,11] Although previous studies have described appropriateness of treatment in patient populations using such criteria[12]

From the *Institute of Social and Preventive Medicine (IUMSP), University of Lausanne, Lausanne, Switzerland; †Department of Orthopaedics and Traumatology, University of Hong Kong, Hong Kong SAR, China; and ‡Neurosurgical Department, Schulthess-Clinic, Zürich, Switzerland.
Acknowledgment date: November 25, 2008. First revision date: February 26, 2009. Second revision date: April 27, 2009. Acceptance date: April 27, 2009.
The manuscript submitted does not contain information about medical device(s)/drug(s).
Foundation 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.
Supported by the Swiss National Science Foundation, Grant 3200–05925.97; the 450th Anniversary University of Lausanne Foundation; and the Fond de performance, project No. 11, Etat de Vaud.
Vincent Wietlisbach is deceased.
Address correspondence and reprint requests to Nadia Danon-Hersch, MD, Institute of Social and Preventive Medicine (IUMSP), Centre Hospitalier Universitaire Vaudois and University of Lausanne, Bugnon 17, 1005 Lausanne, Switzerland; E-mail: nadiadanon@yahoo.fr; or Nadia.Danon@chuv.ch

and documented the reproducibility of such criteria,[13] no previous study has examined the clinical validity of criteria for back surgery developed by the RAND-UCLA appropriateness method, by assessing the comparative outcomes of appropriately *versus* inappropriately treated patients; and although such validation studies have been carried out for appropriateness criteria for other procedures,[14,15] follow-up was limited to 3 or 6 months.

To fill this gap and extend previous research from other fields, we carried out this prospective, observational study to test the hypothesis that the use of appropriateness criteria can improve clinical outcome at 12 months in surgery for low back pain and/or sciatica. Appropriateness criteria were applied prospectively to each clinical situation and outside of the clinical setting, so as not to influence the therapeutic decision.

■ Materials and Methods

Development of Appropriateness Ratings

A multispecialty panel was created including 2 orthopedic surgeons, 3 neurosurgeons, 1 neurologist, 1 rheumatologist, 1 general physician, and 1 internist. The panel followed the standardized procedure of the RAND-UCLA appropriateness method[9,13,16] to develop explicit criteria for surgical indication concerning patients with low back pain and/or sciatica. The work of the expert panel was carried out before the recruitment of the patients to the present study.

First, a summarized literature review[17] was provided to the panelists covering the pertinent literature of low back disorders. The panelists were instructed to consider the conclusions of the literature and their clinical expertise in judging the appropriateness of operative intervention for the theoretical scenarios presented. A list of 1000 theoretical clinical scenarios covering all potential surgical indications for low back pain and/or sciatica was designed and submitted to the expert panel. The variables of the scenarios were the patients' symptoms and duration, the radiologic diagnosis of disc pathology, the neurologic signs, the former applied treatments and the disability status (Table 1). Each clinically specific scenario was first rated separately by all 9 panelists and based on a 9 point scale as to the appropriateness of surgery (1 = extremely inappropriate, 5 = equivocal, 9 = extremely appropriate). Afterwards, the panelists came together and were provided with reports showing their initial ratings and the distribution of the other panelist's ratings. Indications were discussed and the panelists then again assigned individual ratings of appropriateness to each scenario. It was this second rating which was used to determine the appropriateness of indications for disc surgery.[13] Using the median panel vote, and a statistical definition of disagreement, each indication was categorized as "appropriate," "equivocal," or "inappropriate."

Study Population

The study cohort consisted of all consecutive patients between 18 and 75 years old who presented with low back pain and/or sciatica at our outpatient clinic, with reasonable assurance of a possible 12 months follow-up. During a 1-year period, all patients who were referred for radicular pain to the Neurosurgical Department were eligible for inclusion into the study (n = 601); their demographic and pathologic characteristics were described elsewhere.[18] Some of the patients were already hospitalized because of another concomitant disease in the University Hospital of Lausanne in Lausanne, Switzerland. Written informed consent was obtained from all patients before enrolment. Exclusion criteria were previous lumbar surgery, spinal stenosis, spondylolisthesis and prior lumbar fractures. The physical evaluation included a detailed neurologic examination, assessment of pain intensity, disability, health-related quality of life, functional and economic capacity, and the presence of neuroradiological disc disease.

Table 1. Variables Involved in the Description of the Clinical Scenarios Rated by the Expert Panel and Their Categories

Variable	Categories
Major symptom categories	Sciatica/cruralgia
	Back pain only
	Symptoms of central spinal stenosis not due to spondylolisthesis
	Symptoms of lateral spinal stenosis
	Spondylolisthesis
	Miscellaneous (progressive motor weakness, cauda equine syndrome)
	Repeat laminectomy within 3 mo of prior laminectomy
Symptom duration	Acute (<6 wk duration)
	Subacute (6 wk to 6 mo duration)
	Chronic (>6 mo)
Radiological finding	No herniated disc
	Disc herniation
	Disc herniation with free fragment
	No central/lateral stenosis
	Central/lateral stenosis
	Grade I–II isthmic spondylolisthesis
	Grade I–II degenerative spondylolisthesis
Neurological signs	No neurological findings
	Root tension sign only
	Minor neurologic abnormality without root tension sign
	Minor neurologic abnormality with root tension sign
	Major weakness
Previous nonoperative treatments	None
	One
	≥2
If ≥1 nonoperative modality	Supervised exercise not included
	Supervised exercise included
Degree of disability	Mild (limits sports)
	Moderate (limits work)
	Severe (unable to work)
	Bedbound (or hospitalized)
Pending insurance claim	Yes
	No

Example of clinical scenario rated by the expert panel: patient presenting with a subacute sciatica, disc herniation on imaging, minor neurologic abnormality with root tension sign, having received at least 2 nonoperative treatment modalities, and having a severe disability.

Clinical Records Data

After patient enrolment the consulting neurosurgeon determined the treatment plan based on a free clinical decision, which means that he was blind to the expert panel treatment appropriateness criteria. The appropriateness criteria were applied in parallel in a prospective manner to each clinical situation, outside of the clinical setting.

The patients were treated either conservatively or surgically based on the clinical judgment of the enrolling physician and followed up prospectively at 6 and 12 months. The surgical

treatment could be a simple discectomy, a discectomy with decompression or a spondylodesis. Outcome was assessed in accordance to the SF-36 quality of life score,[19–21] the modified Roland-Morris disability scale,[20,22,23] the visual analog scale (VAS) concerning back and leg pain,[24] and the Functional and Economic Outcome Rating scale of Prolo.[25] The root tension sign was also assessed since it is an important preoperative sign of irritation of the nerve roots. Every patient was examined neurologically at baseline, at 6 and at 12 months post-treatment. In addition, the patients were blinded to the appropriateness of their treatment with respect to the panel criteria.

A nurse study coordinator was responsible for collecting the complete data of all included patients. The required variables (listed in Table 1) describing the clinical presentation, the neurologic examination, the neuroradiologic imaging, and the type and results of any treatment related to the disease, were entered into a computer program to assess appropriateness of treatment. This computer program was specifically developed by the Department of Theoretical Informatics of the Swiss Federal Institute of Technology in order to match the clinical presentation of each study patient to 1 of the 1000 theoretical scenarios rated by the expert panel.[26] Findings from the computer analyses of the appropriateness were not communicated to either the consulting physician or the operating neurosurgeon.

Outcome Measures

The SF-36[19–21] is an instrument for measuring health-related quality of life. Its 36 questions address 8 domains (physical functioning, role physical, bodily pain, general health, vitality, social functioning, role emotional, and mental health) and has 2 main components (physical and mental). Each of these 10 SF-36 scales obtain a score ranging from 0 (poor health) to 100 (optimal health). A single health transition item is also included, asking the patient to rate his health compared to 1 year before. The SF-36 has been translated and validated in French.[27]

The modified Roland-Morris disability scale[20,22,23] is a scale ranging from 0 (no disability) to 23 (severe disability). The VAS measures intensity of back pain and sciatica on a scale ranging from 0 (no pain at all) to 10 (worst pain ever).[24] Two separate VAS scales were used for low back pain and sciatica.

The Functional and Economic Outcome Rating scale of Prolo[25] measures the functional and the economic capacity of the individuals, with a scale ranging from 1 to 5 for each of both dimensions. For example an economic capacity of 1 on this scale means that the patient is completely invalid, and a score of 5 on this scale means that the patient is able to work at his previous occupation with no restrictions of any kind. The Prolo score has been shown to be highly correlated with the severity of disc disease as assessed by radiologic imaging in the same study population.[18] For ease of comprehension, a table will summarize the improvement of scores at 12 months instead of their absolute value in each respective scale.

Neuroradiological Assessment

Disc disease was assessed in all patients according to the criteria proposed by Jensen *et al*,[28] Modic,[29] and Spengler *et al*[30] and was categorized as bulging, protrusion, extrusion, and sequestration. To assess the possible variation of interpretation of neuroradiologic imaging by the neurosurgeons, a random sample of 50 imaging studies was double-read by an independent neuroradiologist. The statistical analysis showed excellent agreement (agreement rate 96%, Kappa value of 0.91).[31,32]

Statistical Analyses

We did not have explicit criteria for the appropriateness of conservative treatment. We therefore assumed that if operative treatment was inappropriate, then conservative treatment was appropriate. It was hypothesized that compared to the baseline measurements, patients treated (surgically or nonsurgically) in accordance to the panel appropriateness criteria would have a better quality of life (SF-36), less pain (VAS) and would be less disabled (Roland-Morris) compared to patients treated in disagreement to the criteria. As is customary with the RAND method for binary analyses, "equivocal" care was considered as "appropriate" in order to give the clinician the benefit of the doubt.[33] A confirmed significant measurable better outcome in patients treated in accordance to the appropriateness criteria would be considered as evidence of validation. On the basis of previous studies,[34] minimally clinically important differences were a priori set at 3 of 24 points for the Roland-Morris disability scale,[35] 10% for the SF-36 quality of life score,[36] 2 of 10 points for the VAS of back or leg pain,[37] and 1.5 of 5 points for the Prolo economic score.[1]

Statistical analyses were performed using STATA 10 (STATA Corp., College Station, TX). Results were expressed as means and 95% confidence intervals or as percentages. Bivariate comparisons of appropriateness treatment groups at baseline were performed using the Student t test or χ^2 test for contingency tables, for continuous or categorical variables respectively. The effect of treatment appropriateness on the outcome variables was assessed in 2 ways that were taking into account the longitudinal dependence of the data of the same patients. First, for each outcome, the mean score improvement from baseline to 12 months post-treatment was compared between appropriateness treatment groups using t tests. Second, for each outcome, mixed effects longitudinal regression models were used to estimate the effects of treatment appropriateness, time (binary variable: visit at baseline *vs.* at 12 months post-treatment) and the interaction between time and treatment appropriateness. In these models, the presence of a significant interaction means that an outcome's change over time is different between treatment appropriateness groups. These models also allowed adjustment for potential confounders (age, sex, and surgical treatment). Linear mixed models were carried out to compare continuous outcomes,[38] an ordinal mixed-effects logistic regression was carried out for comparing the Prolo economic scale (ordinal variable)[39] and a mixed-effects logistic regression model was run for comparing the prevalence of the root tension sign. All regression models were first applied to the outcome variables alone and then repeated including the adjustment for age, sex, and surgical treatment. Statistical significance was established with P-values less than 0.05 and 95% confidence interval (CI) bounds were assessed for significance and precision.

Ethical Approval

The present project had received ethical approval from the Ethical Commission of the University Hospital of Lausanne.

Results

Sample Characteristics

As indicated in Figure 1, 601 patients presented with low back pain and/or sciatica at the outpatients clinics. Forty-nine patients refused to participate, 7 were excluded because they had exclusion criteria, 66 were lost to fol-

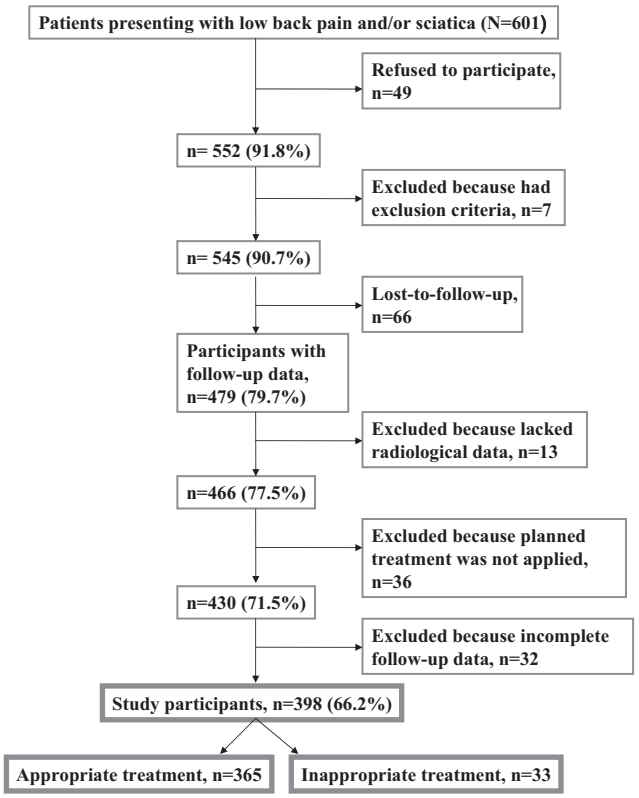

Figure 1. Flow chart of study participants.

low-up, 13 patients were excluded because they lacked radiologic data (which were needed for determining the appropriateness of surgery), 36 patients were excluded because the planned treatment was not applied (8 planned operations were not performed, and 28 operations were performed in patients for whom a conservative treatment had been planned), and 32 patients with missing follow-up data were excluded.

Overall, 398 patients presenting with back pain and/or radicular pain were available for the study after having signed the informed consent form. The male population was predominant with 246 (62%) men.

Of the 1000 theoretical clinical scenarios rated by the expert panel, 137 (13.7%) were actually seen in this study. Descriptions of the 10 most common scenarios are listed in the Appendix. These 10 scenarios accounted for 36.9% of the 398 patients.

With regards to treatment, 197 (49.5%) patients were treated by a surgical procedure and 201 (50.5%) patients were treated conservatively.

Appropriateness of Treatment

Overall, 33 patients (8.3%) were treated inappropriately according to the panel criteria, whereas 365 patients were treated appropriately. Treatment received by participants in the appropriate treatment group (ATG) was surgical in 51% of cases, whereas it was surgical in only 33% of cases in the inappropriate treatment group (ITG; χ^2 (1df) = 3.8; $P = 0.052$). Among the 28 patients with unplanned surgery (and excluded from the statistical analyses), 4 (14.3%) had an inappropriate treatment, and 24 (85.7%) had an appropriate or an equivocal treatment. In comparison, among the 197 patients included in the analysis with planned surgery, 5.6% had an inappropriate treatment.

Table 2 illustrates the baseline health and social characteristics of the participants in each group. The participants of each group were almost equally distributed with respect to age, sex, number of previous episodes of low back pain or sciatica, smoking status, educational level, civil status, and insurance status.

Scales at Baseline

Table 3 shows the baseline scores in the inappropriate and in the appropriate treatment group. There is no statistically significant difference in any of the baseline scores between both groups. The ATG has a higher mean (95% CI) Roland-Morris score (17.0 [16.6–17.5]) than the ITG (15.7 [14.2–17.2]), indicating a somewhat (but nonsignificantly) higher disability level at baseline.

Outcome at 12 Months

Table 4 aims at answering 3 different questions: (1) whether there is a statistically significant improvement of scores from baseline to the visit at 12 months follow-up in each treatment appropriateness group, (2) whether there is a clinically significant improvement in each group, and (3) whether there is a statistically significant difference in the improvement between both groups. Overall, the entire study population had an improvement at 12 months in all outcome scales except for the SF-36 General Health domain, with a worsening of 8.2 points in the ITG and of 1.2 points in the ATG. The percentage of patients with a positive root tension sign decreased significantly in both groups, and although this decrease was of higher magnitude in the ATG than in the ITG, the difference between both groups was not statistically significant.

The mean (95% CI) improvement of SF-36 physical functioning at 12 months was significantly higher in the ATG (35.6 [32.4–38.8] points) than in the ITG (21.2 [11.6–30.8] points). The mean value of VAS decreased 2.3 (2.0–2.7) points in the ATG compared to 0.8 (0.4–2.1) points in the ITG for back pain and 4.0 (3.6–4.4) points in the ATG compared to 2.8 (1.4–4.1) points in the ITG for sciatica.

A significantly higher improvement was found in the ATG for the SF-36 bodily pain (32.5 [29.6–35.3] points *vs.* 14.8 [4.3–25.3] points in the ITG) and Roland-Morris disability score (7.7 [7.0–8.3] points *vs.* 4.2 [2.2–6.3] points in the ITG). Patients in the ITG showed an improvement in most outcome scores at 6 months (results not shown), but these improvements were not maintained at 12 months. Overall, appropriately treated patients had a significantly better outcome at 12 months.

Figures 2 and 3 show the mean scores at baseline and at 12 months follow-up in each group for the main outcome measures. In the ATG, an improvement can be seen at 12 months follow-up in the Roland-Morris disability score, the VAS for sciatica, the VAS for back pain, the

Table 2. Baseline Health and Social Characteristics of the Participants in the Inappropriate Treatment Group (n = 33) and in the Appropriate Treatment Group (n = 365)

Treatment	Inappropriate (n = 33) Number (%)	Appropriate (n = 365) Number (%)	χ^2	*P*
Age category				
<35 yr	5 (15.2)	79 (21.6)	3.070	0.381
35–44 yr	11 (33.3)	97 (26.6)		
45–54 yr	10 (30.3)	78 (21.4)		
≥55 yr	7 (21.2)	111 (30.4)		
Sex				
Men	20 (60.6)	226 (61.9)	0.022	0.882
Women	13 (39.4)	139 (38.1)		
Disc pathology				
No disc herniation	5 (15.2)	41 (11.2)	1.745	0.782
Bulging disc	2 (6.1)	13 (3.6)		
Disc protrusion	5 (15.2)	48 (13.2)		
Disc extrusion	18 (54.6)	238 (65.2)		
Disc sequestration	3 (9.1)	25 (6.9)		
Level of disc disease*				
L2–L3	1 (3.7)	6 (1.9)	0.672	0.880
L3–L4	4 (14.8)	40 (12.5)		
L4–L5	9 (33.3)	122 (38.0)		
L5–S1	13 (48.2)	153 (47.7)		
Lifetime no. of previous episodes of low back pain or sciatalgia, which have completely disappeared				
1–5 episodes	10 (30.3)	145 (39.7)	3.007	0.222
>5 episodes	15 (45.5)	170 (46.6)		
None before the current episode	8 (24.2)	50 (13.7)		
Time since first medical visit for low back pain or sciatalgia				
<1 yr	9 (27.3)	104 (28.5)	0.283	0.963
1–4 yr	7 (21.2)	89 (24.4)		
5–9 yr	6 (18.2)	58 (15.9)		
≥10 yr	11 (33.3)	114 (31.2)		
Pain irradiating to legs	31 (93.9)	341 (93.4)	0.013	0.909
Cardio-vascular or respiratory problem	8 (24.2)	86 (23.6)	0.008	0.930
Smoking status				
Never smoker	9 (27.3)	116 (31.8)	0.923	0.820
No, stopped >6 mo ago	7 (21.2)	81 (22.2)		
No, stopped <6 mo ago	1 (3.0)	19 (5.2)		
Yes	16 (48.5)	149 (40.8)		
Highest educational level completed				
Minimal required no. of school years	11 (33.3)	88 (24.1)	2.789	0.248
Apprenticeship	19 (57.6)	206 (56.4)		
Technical school, college or University	3 (9.1)	71 (19.5)		
Marital status				
Married or living with someone	24 (72.7)	259 (71.0)	0.046	0.830
Widowed, divorced, or never married	9 (27.3)	106 (29.0)		
Currently applying for insurance compensation for back problems or sciatica	0 (0.0)	16 (4.4)	1.507	0.220
Already receiving insurance compensation for back problems or sciatica	1 (3.0)	6 (1.6)	0.337	0.562

*Level of disc pathology was known for 27 patients with inappropriate treatment, and 321 patients with appropriate treatment (total = 348).

SF-36 physical, and mental component scores. The improvement was observed in the same scales in the ITG, except for the mental component score, which had slightly worsened. Compared with conservative treatment, surgery resulted in a statistically significant higher improvement in mean VAS for sciatica and SF-36 bodily pain (results not shown, $P < 0.001$).

The appropriateness of treatment of 30 of the 32 patients with incomplete follow-up data was documented: 25 of them had been appropriately treated, and 5 of them inappropriately. A sensitivity analysis has been done for the mean improvement in the 6 outcome variables selected for Figures 2 and 3. We calculated the effect of a "worse case scenario" on the mean improvement by appropriateness group: in the worse case scenario, the 25 missing, but appropriately treated patients had the mean improvement of the inappropriate treatment group for each outcome, and the 5 missing, but inappropriately treated patients had the mean improvement of the appropriate treatment group. Overall, the mean improvement remained unchanged (results not shown), confirming the robustness of our results.

Table 5 illustrates that after adjustment for age, sex, and surgical treatment, a statistically significant interaction between time and treatment appropriateness was observed for the Roland-Morris scale, the VAS back pain (with VAS sciatica close to statistical significance), the SF-36 physical functioning, physical role, bodily pain,

Table 3. Baseline Scores in the Inappropriate (ITG) and in the Appropriate Treatment Group (ATG)

Treatment	Inappropriate (n = 33) Mean (95% CI)	Appropriate (n = 365) Mean (95% CI)	*t*-test *P**
Roland-Morris (0–23)	15.7 (14.2–17.2)	17.0 (16.6–17.5)	0.096
Prolo economic status (1–5)	2.1 (1.6–2.6)	2.2 (2.1–2.4)	0.550
VAS back pain (0–10)	6.5 (5.4–7.6)	6.2 (5.9–6.5)	0.583
VAS sciatica (0–10)	6.8 (5.7–7.9)	7.0 (6.7–7.3)	0.708
Root tension sign (%)†	69.2 (50.2–88.2)	64.3 (58.8–69.8)	0.614
SF-36 (0–100)			
Physical functioning	41.4 (31.9–50.8)	34.6 (32.0–37.1)	0.135
Physical role	7.6 (0.4–14.8)	7.1 (5.1–9.0)	0.883
Bodily pain	24.5 (18.7–30.3)	19.3 (17.6–21.0)	0.086
General health	59.0 (51.1–67.0)	62.0 (59.8–64.1)	0.441
Vitality	35.2 (29.1–41.2)	36.3 (34.4–38.1)	0.733
Social functioning	50.4 (40.3–60.5)	42.4 (39.6–45.2)	0.110
Emotional role	42.4 (26.4–58.4)	41.9 (37.4–46.5)	0.950
Mental health	54.4 (47.7–61.1)	52.6 (50.6–54.6)	0.611
Physical component score	30.4 (27.2–33.5)	29.0 (28.2–29.8)	0.345
Mental component score	41.7 (37.2–46.2)	41.2 (40.0–42.4)	0.821

**t*-test null hypothesis: mean baseline score in the ITG = mean baseline score in the ATG.
†Root tension sign: ITG (n = 26), ATG (n = 294), total (n = 320).

general health (with vitality and emotional role close to statistical significance), social functioning, mental health, physical component score and mental component score. Thus, although general clinical improvement was achieved for the entire study population, added improvement was achieved by use of a treatment that was appropriate according to the panel criteria.

There was no interaction for the Prolo economic scale or for the root tension sign. Further adjustment for marital status did not change the estimates (results not shown).

As shown in Table 4, the differences between the baseline scores and the scores at 12 months in each treatment appropriateness group were not only statistically significant for most outcome scales, but also clinically significant according to the a priori objectives set in the methods.

With regards to the SF-36 item for self-rated general health (In general, would you say your health is: excellent/very good/good/fair/poor?), when considering the difference between self-rated general health at baseline and self-rated general health at 12 months follow-up (with dichotomized variables, *i.e.*, health self-rated as better *vs.* the same or worse), 78% of participants in the ATG but only 64% of participants in the ITG had reported an improved general health (χ^2 (1df) = 3.4; P = 0.065). However, when considering the self-reported health transition item itself (Compared to 1 year ago, how would you rate your health in general now? Much better/Somewhat better/About the same/Somewhat worse/Much worse now than 1 year ago), at 12 months follow-up, only 35% of study participants rated their general health as somewhat better or much better than the year before. There was no statistically significant difference between the ATG and the ITG ($P > 0.05$).

■ Discussion

In the present study, patients who had received appropriate treatment according to the expert panel criteria had significantly better outcomes at 12 months than patients who had received inappropriate treatment for almost all SF-36 scales, the Roland-Morris disability score and the VAS scale for back pain. The outcomes were considered as clinically significant according to goals established before the study. The consistency of the findings for several outcomes provides strong evidence that

Table 4. Mean Improvement of Scores From 0 to 12 Month in the Inappropriate (ITG, n = 33) and in the Appropriate Treatment Group (ATG, n = 365)

Treatment	Inappropriate Mean (95% CI)	Appropriate Mean (95% CI)	Difference Between Mean Improvement of ATG and ITG Mean (95% CI)	*t*-test *P**
Roland-Morris (0–23)	4.2 (2.2–6.3)†	7.7 (7.0–8.3)†	3.4 (1.1–5.8)	0.004
Prolo economic status (1–5)	0.9 (0.3–1.4)	1.2 (1.0–1.4)	0.3 (−0.3–0.9)	0.275
VAS back pain (0–10)	0.8 (0.4–2.1)	2.3 (2.0–2.7)†	1.5 (0.2–2.7)	0.019
VAS sciatica (0–10)	2.8 (1.4–4.1)†	4.0 (3.6–4.4)†	1.3 (−0.1–2.6)	0.077
Root tension sign (%)	42.3 (19.0–65.6)	44.9 (38.4–51.4)	2.6 (−20.3–25.5)	0.824
SF-36 (0–100)				
Physical functioning	21.2 (11.6–30.8)†	35.6 (32.4–38.8)†	14.4 (3.3–25.4)	0.011
Physical role	26.5 (10.9–42.2)†	43.5 (39.1–47.9)†	17.0 (1.6–32.4)	0.031
Bodily pain	14.8 (4.3–25.3)†	32.5 (29.6–35.3)†	17.7 (7.7–27.6)	0.001
General health	−8.2 (−15.9 to −0.5)	−1.2 (−3.1–0.7)	7.0 (0.3–13.7)	0.042
Vitality	6.2 (−1.6–14.1)†	13.2 (11.2–15.3)†	7.0 (−0.2–14.3)	0.057
Social functioning	2.3 (−8.7–13.3)	25.8 (22.3–29.2)†	23.5 (11.6–35.4)	<0.001
Emotional role	11.1 (−9.3–31.5)†	28.9 (23.3–34.4)†	17.7 (−1.5–37.0)	0.071
Mental health	2.7 (−5.4–10.7)	11.6 (9.6–13.6)†	8.9 (1.9–16.0)	0.013
Physical component score	6.8 (2.9–10.7)†	12.3 (11.2–13.5)†	5.6 (1.5–9.6)	0.007
Mental component score	−0.5 (−6.0–5.0)	5.0 (3.8–6.3)†	5.5 (1.0–10.0)	0.016

**t*-test null hypothesis: mean improvement in the score of the ITG = mean improvement in the score of the ATG.
†Clinically significant improvement. For the SF-36 quality of life scores, an improvement of 10% was expected.

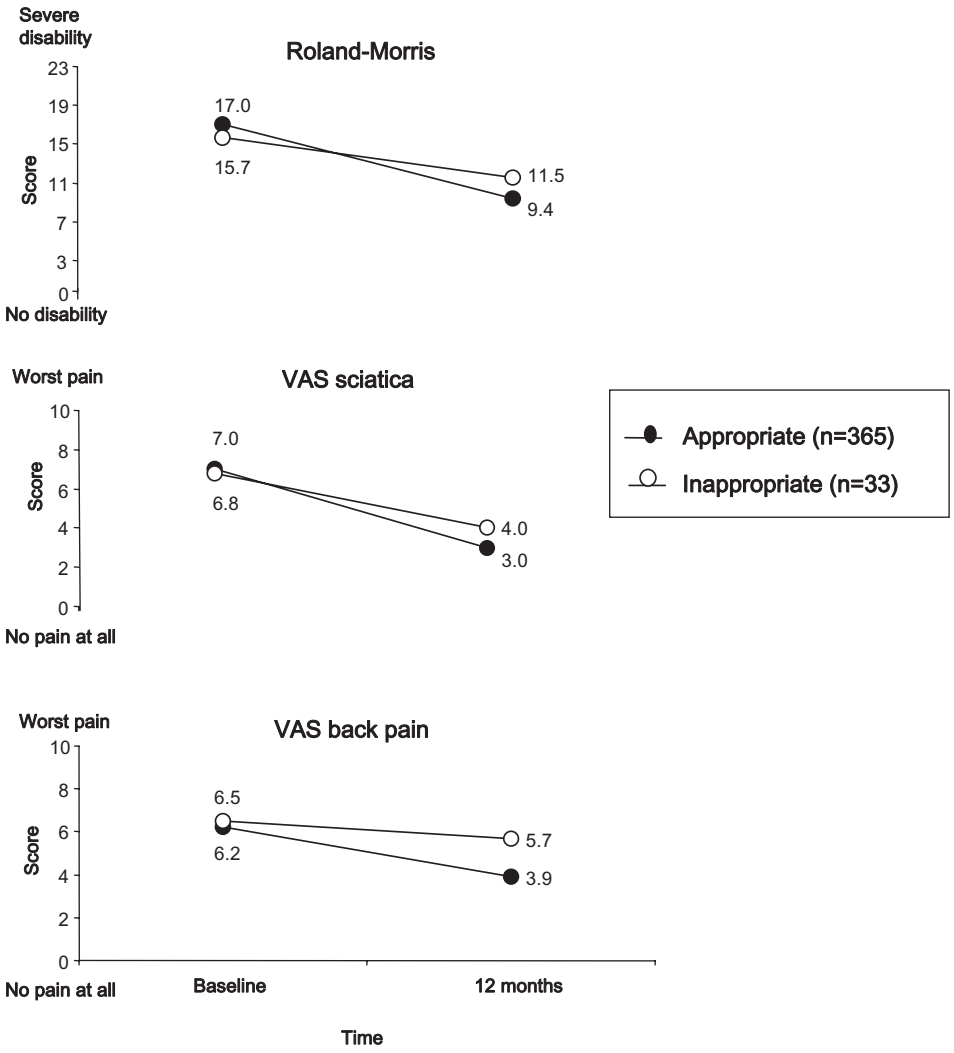

Figure 2. Disability and pain scores at baseline and 12 months follow-up, by treatment appropriateness.

appropriate criteria indeed improve outcome in low back surgery. To our knowledge, no previous study has analyzed outcome of low back surgery using appropriateness criteria.

Like in several trials or cohorts,[40–42] which reported that relief of leg pain was the most striking and consistent improvement with surgery, surgery in our study was significantly associated with higher improvement than conservative treatment for VAS for sciatica and SF-36 in bodily pain.

Although our study was prospective and controlled, inherent limitations existed. For example, there was no randomization. While it does not exclude some degree of selection bias, the fact that no significant differences were identified between the 2 groups (Tables 2 and 3) at baseline is reassuring as to the validity of our findings. There might have been residual confounding by unmeasured confounders, such as body mass index or occupation category. It has been reported that patients with sedentary work and housewives had better economic Prolo scores than those who did strenuous work.[1] In our study, the level of work strenuousness was not measured, so that the distribution of this variable in both comparison groups is not known.

Assessments of outcomes were self-reported. Patient perception of symptoms might have been influenced by their expectations of treatment results. Likewise, as demonstrated for the 197 participants who underwent surgery,[34] more optimistic physician expectation could have positively influenced outcome of patients with inappropriate treatment. Neither the patients nor the investigators assessing outcomes were blinded to the treatment, which could only have been accomplished by performing sham surgical procedures that may have been deemed unethical. But they were blinded to the treatment appropriateness ratings of the expert panel. The last outcome evaluation was at 12 months of follow-up and not more long-term. A controlled trial that compared surgery with conservative treatment for patients with sciatica showed that while surgical intervention had better results after 1 year, no significant differences were found after 4 and 10 years of follow-up.[43] Since 32 participants with incomplete follow-up data have been excluded from our study, a selection bias might have occurred. However, they only represent 7% of all followed-up participants and our sensitivity analysis confirmed the robustness of our results.

This study evaluated the use of appropriateness criteria for low back surgery at 12 months follow-up, which is a longer follow-up than previous studies on appropriateness methods.

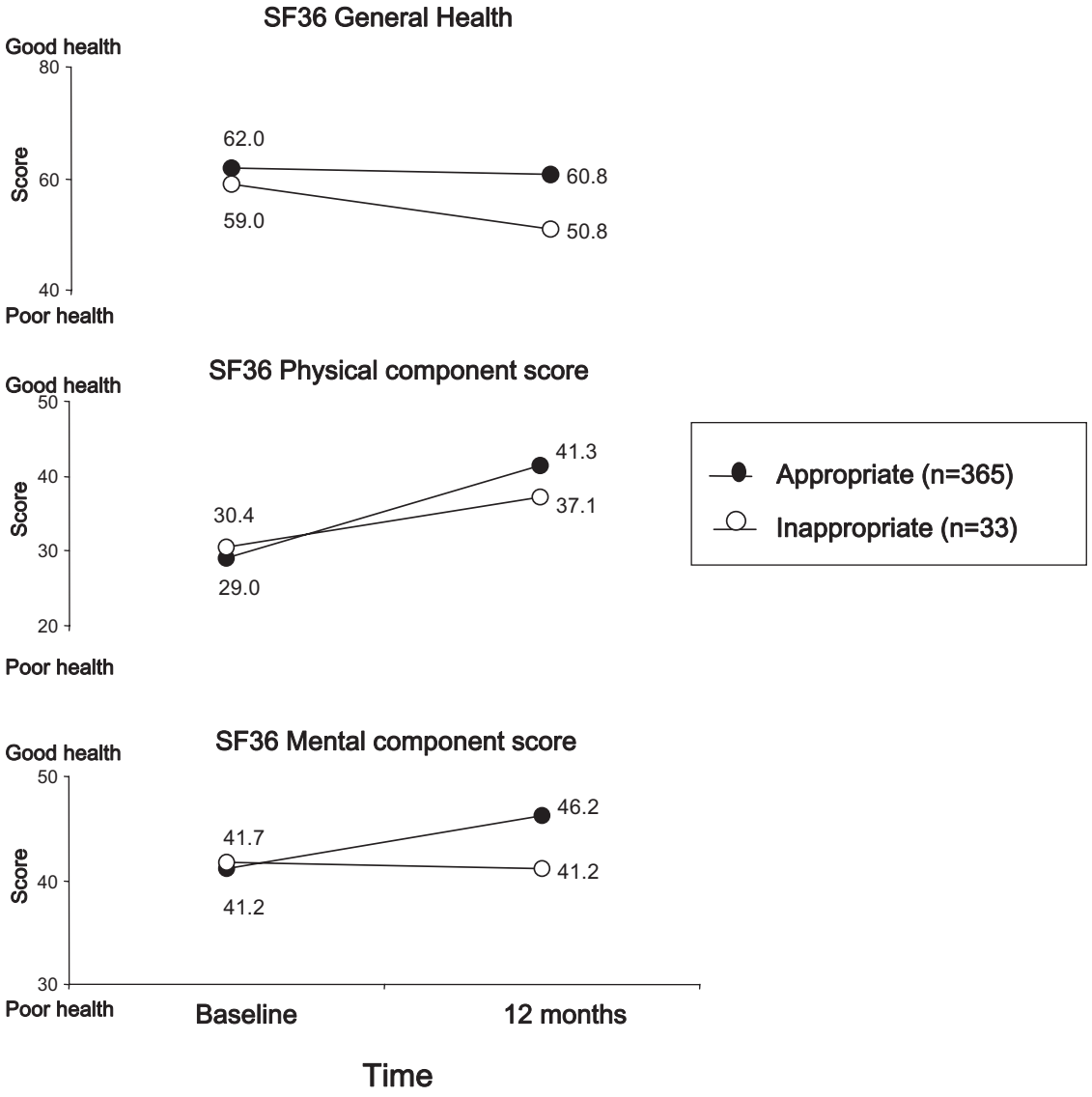

Figure 3. Main health-related quality of life scores at baseline and 12 months follow-up, by treatment appropriateness.

We used valid patient-oriented outcome measures that had been previously recommended for clinical research on the treatment of low back pain,[44] allowing direct comparison with other studies. Moreover, appropriateness criteria issued by experts could thus be linked with patient perspectives, which is important when keeping in mind that the main objective of treatment is to improve the patient's quality of life.

Different outcomes were measured, allowing evaluation of the consistency of the findings. Cut-off values for considering improvement at 12 months as "clinically meaningful" were set pre hoc for each outcome score (at the stage of the calculation of the study sample size). Despite a low statistical power due to the small size of the inappropriate treatment group (n = 33), appropriately treated patients had statistically and clinically significant higher improvement at 12 months than inappropriately treated patients for almost all outcome scores, which provided evidence for a "real" additional effect of treatment appropriateness on clinical outcome.

While some authors[45] advocate the use of the Oswestry Disability Index for the assessment of disability, the Roland-Morris disability scale was used in our study. However, the Roland-Morris scale was only 1 among 15 outcome measures in the present study, and highly consistent results have been obtained between them.

Low back pain accounts for 2.3% of physician visits[46] and more than 1 million spinal procedures are performed yearly in the United States.[47] Failure rates for lumbar back surgery range from 5% to 50%,[1,48,49] depending on the setting and definition of "failure rate," leading 1 author to conclude that "ill-advised spine surgery creates more problems for patients than it solves."[47] Expression of such concerns suggests that appropriateness criteria for low back surgery are still much needed.

Although the panel appropriateness criteria were up to date at the time of the study, they were developed in 1995 and might now be partly outdated. As an example, the US National Guidelines Clearing House (www.guideline.gov) uses the limit of 5 year for the deactivation of guidelines

Table 5. Outcome Scores According to Treatment Appropriateness, Time (Labelled "Visit") and Interaction Visit* Treatment Appropriateness, Using a Mixed Linear Model, Unless Otherwise Indicated

	Crude Analysis		Multivariate Analysis*			
Outcome Scores	Coefficient	*P*	Coefficient	Lower 95% Limit	Upper 95% Limit	*P*
Roland-Morris (0–23)						
Appropriate	3.05	0.033	3.11	0.30	5.92	0.030
Visit	−2.11	<0.001	−2.11	−3.22	−1.00	<0.001
Visit* appropriate	−1.72	0.004	−1.72	−2.88	−0.56	0.004
Prolo economic score (1–5)†						
Appropriate	1.02	0.975	1.14	0.34	3.83	0.828
Visit	2.03	0.003	2.02	1.27	3.23	0.003
Visit* appropriate	1.33	0.252	1.33	0.82	2.16	0.248
VAS back pain (0–10)						
Appropriate	0.43	0.573	0.55	−0.94	2.04	0.471
Visit	−0.42	0.155	−0.42	−1.01	0.16	0.155
Visit* appropriate	−0.74	0.018	−0.74	−1.35	−0.13	0.018
VAS sciatica (0–10)						
Appropriate	0.83	0.314	0.79	−0.82	2.40	0.334
Visit	−1.38	<0.001	−1.38	−2.04	−0.71	<0.001
Visit* appropriate	−0.63	0.077	−0.63	−1.32	0.07	0.077
Root tension sign (%)‡						
Appropriate	0.84	0.842	0.70	0.14	3.54	0.669
Visit	0.31	0.001	0.32	0.16	0.64	0.001
Visit* appropriate	0.89	0.745	0.86	0.42	1.75	0.670
SF-36						
Physical functioning						
Appropriate	−13.98	0.034	−13.32	−26.18	−0.47	0.042
Visit	10.61	<0.001	10.61	5.33	15.88	<0.001
Visit* appropriate	7.18	0.011	7.18	1.67	12.69	0.011
Physical role						
Appropriate	−9.01	0.316	−9.28	−26.92	8.36	0.302
Visit	13.26	<0.001	13.26	5.90	20.61	<0.001
Visit* appropriate	8.49	0.030	8.49	0.81	16.17	0.030
Bodily pain						
Appropriate	−13.99	0.018	−14.06	−25.62	−2.50	0.017
Visit	7.41	0.002	7.41	2.65	12.16	0.002
Visit* appropriate	8.84	<0.001	8.84	3.87	13.80	<0.001
General health						
Appropriate	−0.55	0.911	−1.85	−11.53	7.82	0.708
Visit	−4.09	0.012	−4.09	−7.30	−0.88	0.012
Visit* appropriate	3.49	0.041	3.49	0.14	6.84	0.041
Vitality						
Appropriate	−2.39	0.611	−2.74	−11.99	6.50	0.561
Visit	3.11	0.078	3.11	−0.35	6.56	0.078
Visit* appropriate	3.52	0.056	3.52	−0.09	7.13	0.056
Social functioning						
Appropriate	−19.74	0.007	−19.85	−34.16	−5.53	0.007
Visit	1.14	0.696	1.14	−4.56	6.83	0.696
Visit* appropriate	11.76	<0.001	11.76	5.81	17.70	<0.001
Emotional role						
Appropriate	−9.38	0.418	−10.14	−32.86	12.59	0.382
Visit	5.56	0.236	5.56	−3.64	14.75	0.236
Visit* appropriate	8.87	0.070	8.87	−0.73	18.48	0.070
Mental health						
Appropriate	−6.25	0.188	−6.60	−15.90	2.70	0.164
Visit	1.33	0.438	1.33	−2.04	4.70	0.438
Visit* appropriate	4.46	0.013	4.46	0.95	7.98	0.013
Physical component score						
Appropriate	−4.13	0.088	−4.13	−8.89	0.63	0.089
Visit	3.39	0.001	3.39	1.45	5.32	0.001
Visit* appropriate	2.78	0.007	2.78	0.76	4.80	0.007
Mental component score						
Appropriate	−3.24	0.261	−3.53	−9.19	2.13	0.221
Visit	−0.25	0.822	−0.25	−2.39	1.90	0.822
Visit* appropriate	2.76	0.016	2.76	0.52	4.99	0.016

*Coefficients and their 95% confidence limits are adjusted for age, sex, and surgical treatment.
†Ordinal mixed-effects logistic regression model.
‡Multilevel mixed-effects logistic regression model.

which have not been revisited to ensure that they are still up to date. With this in mind, the appropriateness criteria used in this study should be revisited and updated before they are made more widely available. An important change introduced into the clinical guidelines for low back pain since 1995 is the recommendation to avoid bed rest.[50] Still, the controversy about which cases to operate persists over time,[51,52] underscoring the need for explicit criteria and

studies such as this one. More generally, the present study confirms that the RAND appropriateness method can yield higher clinical improvement.

■ Conclusion

The consistently higher improvement in appropriately treated patients for most clinical outcomes provides strong evidence that appropriate criteria indeed improve outcome in low back surgery. However, before wider dissemination, these appropriateness criteria should be updated to take into account current state of the art. Further, larger studies involving different study populations should be carried out to evaluate the more long-term effect of the use of appropriateness criteria on various outcome dimensions. While randomization could overcome the inevitable risk of bias and confounding inherent in observational studies, the numerous practical difficulties are not likely to be easily overcome. For example, it seems that both patients and surgeons generally prefer surgery over conservative treatment. They might thus not accept randomization; and it would not be possible for the patients to be blinded to whether they have or not received surgery. Overall, our results clearly support the hypothesis that application of appropriateness criteria can significantly improve patient outcomes.

■ Key Points

- A multispecialty expert panel used the RAND appropriateness method to develop appropriateness criteria for low back surgery.
- In this prospective, controlled study addressing surgery for patients with low back pain and/or sciatica, patients appropriately treated according to the expert panel had a significantly higher improvement at 12 months than inappropriately treated patients.
- The higher improvement experienced by appropriately treated patients was clinically meaningful.
- The study suggests that application of appropriateness criteria for patients with low back pain and/or sciatica can significantly improve patient outcomes.

Acknowledgments

The authors thank Luca Regli, MD, Olivier Vernet, MD, Claudio Pollo, MD, Jocelyne Bloch, MD, and Marie Ruiz for their support in the recruitment and follow-up of patients. The authors also thank Pierluigi Ballabeni, PhD, for assistance with statistical analyses.

References

1. Davis RA. A long-term outcome analysis of 984 surgically treated herniated lumbar discs. *J Neurosurg* 1994;80:415–21.
2. Fager CA, Freidberg SR. Analysis of failures and poor results of lumbar spine surgery. *Spine* 1980;5:87–94.
3. Atlas SJ, Keller RB, Wu YA, et al. Long-term outcomes of surgical and nonsurgical management of sciatica secondary to a lumbar disc herniation: 10 year results from the maine lumbar spine study. *Spine* 2005;30:927–35.
4. Jarvik JG, Deyo RA. Diagnostic evaluation of low back pain with emphasis on imaging. *Ann Intern Med* 2002;137:586–97.
5. Osterman H, Seitsalo S, Karppinen J, et al. Effectiveness of microdiscectomy for lumbar disc herniation: a randomized controlled trial with 2 years of follow-up. *Spine* 2006;31:2409–14.
6. Frymoyer JW. Back pain and sciatica. *N Engl J Med* 1988;318:291–300.
7. Koes BW, van Tulder MW, Peul WC. Diagnosis and treatment of sciatica. *BMJ* 2007;334:1313–7.
8. Park RE, Fink A, Brook RH, et al. Physician ratings of appropriate indications for six medical and surgical procedures. *Am J Public Health* 1986;76:766–72.
9. Brook RH, Chassin MR, Fink A, et al. A method for the detailed assessment of the appropriateness of medical technologies. *Int J Technol Assess Health Care* 1986;2:53–63.
10. Porchet F, Vader JP, Larequi-Lauber T, et al. The assessment of appropriate indications for laminectomy. *J Bone Joint Surg Br* 1999;81:234–9.
11. Leape LL, Hilborne LH, Park RE, et al. The appropriateness of use of coronary artery bypass graft surgery in New York State. *JAMA* 1993;269:753–60.
12. Larequi-Lauber T, Vader JP, Burnand B, et al. Appropriateness of indications for surgery of lumbar disc hernia and spinal stenosis. *Spine* 1997;22:203–9.
13. Vader JP, Porchet F, Larequi-Lauber T, et al. Appropriateness of surgery for sciatica: reliability of guidelines from expert panels. *Spine* 2000;25:1831–6.
14. Quintana JM, Escobar A, Azkarate J, et al. Appropriateness of total hip joint replacement. *Int J Qual Health Care* 2005;17:315–21.
15. Quintana JM, Cabriada J, Arostegui I, et al. Health-related quality of life and appropriateness of cholecystectomy. *Ann Surg* 2005;241:110–8.
16. Wietlisbach V, Vader JP, Porchet F, et al. Statistical approaches in the development of clinical practice guidelines from expert panels: the case of laminectomy in sciatica patients. *Med Care* 1999;37:785–97.
17. Stevens CD, Dubois RW, Larequi-Lauber T, et al. Efficacy of lumbar discectomy and percutaneous treatments for lumbar disc herniation. *Soz Praventivmed* 1997;42:367–79.
18. Porchet F, Wietlisbach V, Burnand B, et al. Relationship between severity of lumbar disc disease and disability scores in sciatica patients. *Neurosurgery* 2002;50:1253–9.
19. Brazier JE, Harper R, Jones NM, et al. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *BMJ* 1992;305:160–4.
20. Patrick DL, Deyo RA, Atlas SJ, et al. Assessing health-related quality of life in patients with sciatica. *Spine* 1995;20:1899–908.
21. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). Part I: conceptual framework and item selection. *Med Care* 1992;30:473–83.
22. Stansfeld SA, Smith GD, Marmot M. Association between physical and psychological morbidity in the Whitehall II Study. *J Psychosom Res* 1993;37:227–38.
23. Stratford PW, Binkley J, Solomon P, et al. Defining the minimum level of detectable change for the Roland-Morris questionnaire. *Phys Ther* 1996;76:359–65.
24. Haefeli M, Elfering A. Pain assessment. *Eur Spine J* 2006;15(suppl 1):S17–24.
25. Prolo DJ, Oklund SA, Butcher M. Toward uniformity in evaluating results of lumbar spine operations: a paradigm applied to posterior lumbar interbody fusions. *Spine* 1986;11:601–6.
26. Ebel N, Jeannot J, Rekik Y, et al. *CADAM: Towards a Web-Based Information and Decision Support System for Appropriateness in Medicine. Europia '98*. Paris, France: Cyberdesign; 1998.
27. Perneger TV, Leplege A, Etter JF, et al. Validation of a French-language version of the MOS 36-Item Short Form Health Survey (SF-36) in young healthy adults. *J Clin Epidemiol* 1995;48:1051–60.
28. Jensen MC, Brant-Zawadzki MN, Obuchowski N, et al. Magnetic resonance imaging of the lumbar spine in people without back pain. *N Engl J Med* 1994;331:69–73.
29. Modic MT. Advances in spinal imaging. *Clin Neurosurg* 1992;38:97–111.
30. Spengler DM, Ouellette EA, Battie M, et al. Elective discectomy for herniation of a lumbar disc: additional experience with an objective method. *J Bone Joint Surg Am* 1990;72:230–7.
31. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
32. Kundel HL, Polansky M. Measurement of observer agreement. *Radiology* 2003;228:303–8.
33. Shekelle PG, Kahan JP, Bernstein SJ, et al. The reproducibility of a method to identify the overuse and underuse of medical procedures. *N Engl J Med* 1998;338:1888–95.
34. Graz B, Wietlisbach V, Porchet F, et al. Prognosis or "curabo effect?": physician prediction and patient outcome of surgery for low back pain and sciatica. *Spine* 2005;30:1448–52.

35. Hagg O, Fritzell P, Nordwall A. The clinical importance of changes in outcome scores after treatment for chronic low back pain. *Eur Spine J* 2003;12: 12–20.
36. Angst F, Aeschlimann A, Stucki G. Smallest detectable and minimal clinically important differences of rehabilitation intervention with their implications for required sample sizes using WOMAC and SF-36 quality of life measurement instruments in patients with osteoarthritis of the lower extremities. *Arthritis Rheum* 2001;45:384–91.
37. Bombardier C, Hayden J, Beaton DE. Minimal clinically important difference. Low back pain: outcome measures. *J Rheumatol* 2001;28:431–8.
38. Vangeneugden T, Laenen A, Geys H, et al. Applying concepts of generalizability theory on clinical trial data to investigate sources of variation and their impact on reliability. *Biometrics* 2005;61:295–304.
39. Scott SC, Goldberg MS, Mayo NE. Statistical assessment of ordinal outcomes in comparative studies. *J Clin Epidemiol* 1997;50:45–55.
40. 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.
41. Weinstein JN, Lurie JD, Tosteson TD, et al. Surgical vs. nonoperative treatment for lumbar disk herniation: the Spine Patient Outcomes Research Trial (SPORT) observational cohort. *JAMA* 2006;296:2451–9.
42. Weinstein JN, Tosteson TD, Lurie JD, et al. Surgical vs. nonoperative treatment for lumbar disk herniation: the Spine Patient Outcomes Research Trial (SPORT): a randomized trial. *JAMA* 2006;296:2441–50.
43. Weber H. Lumbar disc herniation: a controlled, prospective study with ten years of observation. *Spine* 1983;8:131–40.
44. Deyo RA, Andersson G, Bombardier C, et al. Outcome measures for studying patients with low back pain. *Spine* 1994;19:2032S–6S.
45. Frost H, Lamb SE, Stewart-Brown S. Responsiveness of a patient specific outcome measure compared with the Oswestry Disability Index v2.1 and Roland and Morris Disability Questionnaire for patients with subacute and chronic low back pain. *Spine* 2008;33:2450–7.
46. Deyo RA, Mirza SK, Martin BI. Back pain prevalence and visit rates: estimates from US national surveys, 2002. *Spine* 2006;31:2724–7.
47. Burton CV. Failed back surgery patients: the alarm bells are ringing. *Surg Neurol* 2006;65:5–6.
48. Asch HL, Lewis PJ, Moreland DB, et al. Prospective multiple outcomes study of outpatient lumbar microdiscectomy: should 75% to 80% success rates be the norm? *J Neurosurg* 2002;96:34–44.
49. Slipman CW, Shin CH, Patel RK, et al. Etiologies of failed back surgery syndrome. *Pain Med* 2002;3:200–14.
50. Allen C, Glasziou P, Del MC. Bed rest: a potentially harmful treatment needing more careful evaluation. *Lancet* 1999;354:1229–33.
51. Melloh M, Roder C, Elfering A, et al. Differences across health care systems in outcome and cost-utility of surgical and conservative treatment of chronic low back pain: a study protocol. *BMC Musculoskelet Disord* 2008;9:81.
52. Peul WC, van den Hout WB, Brand R, et al. Prolonged conservative care versus early surgery in patients with sciatica caused by lumbar disc herniation: two year results of a randomised controlled trial. *BMJ* 2008;336: 1355–8.

■ Appendix

The 10 most frequent clinical scenarios seen in this study (n = 147, 36.9%).

1. Patient with a subacute sciatica/cruralgia (6 weeks to 6 months duration), disc herniation on imaging, no insurance claim, minor neurologic abnormality with root tension sign, previous treatment with 2 or more nonoperative methods, and a severe degree of disability (unable to work). According to the expert panel rating, surgery was the appropriate treatment for this scenario (n = 26, 6.5%).
2. Patient with a subacute sciatica/cruralgia (6 weeks to 6 months duration), disc herniation on imaging, no insurance claim, minor neurologic abnormality without root tension sign, previous treatment with 2 or more nonoperative methods, and a severe degree of disability (unable to work). According to the expert panel rating, surgery was the appropriate treatment for this scenario (n = 19, 4.8%).
3. Patient with a chronic sciatica/cruralgia (more than 6 months duration), disc herniation on imaging, no insurance claim, minor neurologic abnormality with or without root tension sign, previous treatment with 2 or more nonoperative methods, supervised exercise included, and a severe degree of disability (unable to work). According to the expert panel rating, surgery was the appropriate treatment for this scenario (n = 19, 4.8%).
4. Patient with acute sciatica/cruralgia (less than 6 weeks duration), disc herniation on imaging, minor neurologic abnormality with root tension sign, previous treatment with 2 or more nonoperative methods, and a severe degree of disability (unable to work). According to the expert panel rating, the appropriateness of surgery was equivocal for this scenario (n = 18, 4.5%).
5. Patient with acute sciatica/cruralgia (less than 6 weeks duration), disc herniation on imaging, minor neurologic abnormality without root tension sign, previous treatment with 2 or more nonoperative methods, and a severe degree of disability (unable to work). According to the expert panel rating, the appropriateness of surgery was equivocal for this scenario (n = 16, 4.0%).
6. Patient with a subacute sciatica/cruralgia (6 weeks to 6 months duration), disc herniation on imaging, no insurance claim, minor neurologic abnormality without root tension sign, previous treatment with 2 or more nonoperative methods, and a moderate degree of disability (limiting work). According to the expert panel rating, the appropriateness of surgery was equivocal for this scenario (n = 13, 3.3%).
7. Patient with acute sciatica/cruralgia (less than 6 weeks duration), disc herniation on imaging, minor neurologic abnormality without root tension sign, previous treatment with 2 or more nonoperative methods, and a moderate degree of disability (limiting work). According to the expert panel rating, surgery was inappropriate for this scenario (n = 9, 2.3%).
8. Patient with a subacute sciatica/cruralgia (6 weeks to 6 months duration), disc herniation on imaging, no insurance claim, minor neurologic abnormality without root tension sign, previous treatment with 1 nonoperative modality, and a moderate degree of disability (limiting work). According to the expert panel rating, the appropriateness of surgery was equivocal for this scenario (n = 9, 2.3%).
9. Patient with a subacute sciatica/cruralgia (6 weeks to 6 months duration), disc herniation on imaging,

no insurance claim, minor neurologic abnormality with root tension sign, previous treatment with 2 or more nonoperative methods, and a moderate degree of disability (limiting work). According to the expert panel rating, the appropriateness of surgery was equivocal for this scenario (n = 9, 2.3%).

10. Patient with symptoms of central spinal stenosis, not due to spondylolisthesis, central stenosis on imaging, an abnormal neurologic examination and a severe degree of disability. According to the expert panel rating, surgery was the appropriate treatment for this scenario (n = 9, 2.3%).