SPINE Volume 26, Number 23, pp 2521–2534

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

Peter Fritzell, MD,* Olle Hägg, MD,† Per Wessberg, MD,†
Anders Nordwall, MD, PhD,† and the Swedish Lumbar Spine Study Group‡

Study Design. A randomized controlled multicenter study with a 2-year follow-up by an independent observer.

Objectives. To determine whether fusion of the lower lumbar spine could reduce pain and diminish disability more effectively when compared with nonsurgical treatment in patients with severe chronic low back pain (CLBP).

Summary of Background Data. The reported results after fusion surgery on patients with CLBP vary considerably, and the evidence of treatment efficacy is weak in the absence of randomized controlled studies.

Patients and Methods. A total of 294 patients referred to 19 spinal centers from 1992 through 1998 were randomized blindly into four treatment groups. Patients aged 25–65 years with CLBP for at least 2 years and with radiologic evidence of disc degeneration at L4–L5, L5-S1, or both were eligible to participate in the study. The surgical group (n=222) included three different fusion techniques, not analyzed separately in this study. Patients in the nonsurgical group (n=72) were treated with different kinds of physical therapy. The surgical group comprised 49.5% men, and the mean age was 43 years. The corresponding figures for the nonsurgical group were 48.6% and 44 years. The patients had suffered from low back pain for a mean of 7.8 and 8.5 years and been on sick leave due to back pain for a mean of 3.2 and 2.9 years, respectively. The Visual Analogue Scale (VAS) was used to measure pain. The Oswestry Low Back Pain Questionnaire, the Million Score and the General Function Score (GFS) were used to measure disability. The Zung Depression Scale was used to measure depressive symptoms. The overall result was assessed by the patient and by an independent observer. Records from the Swedish Social Insurance were used to evaluate work disability. Patients who changed groups were included in the analyses of significance according to the intention-to-treat principle.

Results. At the 2-year follow-up 289 of 294 (98%) patients, including 25 who had changed groups, were examined. Back pain was reduced in the surgical group by 33% (64 to 43), compared with 7% (63 to 58) in the nonsurgical group (P=0.0002). Pain improved most during the first 6 months and then gradually deteriorated. Disability according to Oswestry was reduced by 25% (47 to 36) compared with 6% (48 to 46) among nonsurgical patients (P=0.015), according to Million by 28% (64 to 46) compared with 8% (66 to 60) (P=0.004), and according-toGFS by 31% (49 to 34) compared with 4% (48 to 46) (P=0.005). The depressive symptoms, according to Zung, were reduced by 20% (39 to 31) in the surgical group compared with 7% (39 to 36) in the nonsurgical group (P=0.123). In the surgical group 63% (122/195) rated themselves as "much better" or "better" compared with 29% (18/62) in the nonsurgical group (P<0.0001). The "net back to work rate" was significantly in favor of surgical treatment, or 36% vs. 13% (P=0.002). The early complication rate in the surgical group was 17%.

Conclusion. Lumbar fusion in a well-informed and selected group of patients with severe CLBP can diminish pain and decrease disability more efficiently than commonly used nonsurgical treatment. [Key words: chronic low back pain, degenerative disc disease, lumbar spinal fusion, nonsurgical treatment, clinical outcome, multicenter randomized controlled trial] **Spine 2001;26:2521–2534**

Approximately 70-85% of all people experience back pain at some time in life,[2] and patients with chronic low back pain (CLBP), *i.e.*, pain duration more than 3 months, use the health services more often than most other patient groups.[21,50] The spontaneous recovery of these patients is slow and uncertain, and the return-to-work rate after 2 years of absence from work due to low back pain has been shown to be close to zero.[41] There is a lack of consensus regarding treatment of patients with CLBP,[17,34,36,52] but physical therapy and different forms of exercise are usually the first choices. Some of these regimens have been demonstrated to give satisfactory results, at least in the short term.[30,49] When nonsurgical treatments fail, spine surgeons may occasionally perform fusion surgery (arthrodesis) on selected patients, with the aim of reducing pain and decreasing disability.[22,54] However, surgical treatment is controversial. Several authors have claimed that CLBP is basically a multifactorial entity and that surgical treatment so far has failed to demonstrate any beneficial effects when compared with nonsurgical treatment, placebo, or natural history.[17] Reported results from several surgical studies—mostly retrospective and conducted on heterogeneous patient groups without a control group—have also been inconsistent.[18,46]

Consequently, indications for fusion surgery vary between different countries and areas within countries,

From the *Department of Orthopedic Surgery, Falun Hospital, Falun, and the †Department of Orthopedic Surgery, Sahlgrenska University Hospital, Göteborg, Sweden. Members of the Swedish Lumbar Spine Study Group are listed in the Appendix at the end of this article.
Financial support was granted by Acromed Corporation, Raynham, Massachusetts, USA, and Ossano Scandinavica AB, Stockholm, Sweden.

even among surgeons at the same institution.[1,7,8,10,13,15]. Taylor et al have reported that surgical treatment, especially lumbar fusion, is increasing in frequency, while at the same time hospitalization for most low back pain disorders is decreasing.[45] However, Elam et al have shown that it is possible to reverse the increasing trend of fusion surgery by using strict guidelines.[13]

This means there is a need for randomized controlled studies comparing fusion surgery to nonsurgical treatment in well-defined patient groups[6,11,17,33–36,40,46] where confounding factors,[18] the natural history of the condition,[51] the regression to the mean,[53] and, if possible, the placebo effects[47] can be controlled.

To meet these needs, the Swedish Lumbar Spine Study Group conducted a multicenter randomized controlled trial (RCT) to evaluate whether lumbar fusion could reduce pain and decrease disability more effectively than commonly recommended nonsurgical treatment on patients with severe CLBP.

■ Patients and Methods

Patients aged 25–65 years with CLBP, consecutively referred from primary care physicians and other clinicians to a spine surgeon at 19 orthopedic departments during the period 1992–1998, were eligible to participate in the study. If the patient fulfilled the inclusion criteria of the study, he or she was given oral and written information by the treating surgeon about treatment options and about the Swedish Lumbar Spine Study. The patient was told that no treatment method, as far as was known, was superior to any other. If the patient chose to participate, self-administered questionnaires covering subjective, clinical, paraclinical, and sociodemographic variables, as suggested by the Quebec Task Force,[41] were completed. All data obtained were sent for a second opinion from a spine surgeon who also participated in the study. Both surgeons had to agree on the eligibility and the inclusion and exclusion criteria for the patient to be included.

The ethics committees at the universities in all participating areas approved the study. Before randomization, all patients gave informed consent, and all pretreatment questionnaires and protocols were completed.

Inclusion criteria were:

- Patients aged 25–65 years and of both sexes with severe CLBP.
- Pain duration at least 2 years.
- Back pain more pronounced than leg pain and no signs of nerve root compression.
- The treating surgeon should interpret the pain as emanating from L4–L5 and/or L5–S1 using the patients' history, physical examination, and radiographic signs.
- The patient must have been on sick leave (or have had "equivalent" major disability) for at least 1 year, and nonsurgical treatment efforts should have been unsuccessful.
- A score of at least 7 of 10 points for 10 questions reflecting "Function and Working Disability," where 10 was equivalent to "severe pain, no function" in combination with "total handicap, no working ability" (Table 1).
- Degenerative changes at L4–L5 and/or L5–S1 ("spondylosis") on plain radiographs and/or computed tomography (CT), and/or magnetic resonance imaging (MRI). The presence of a herniated disc was allowed in the absence of clinical signs of nerve root compression.
- Good understanding of the Swedish language.

Exclusion criteria were:

- Obvious ongoing psychiatric illness.
- Previous spine surgery except for successful removal of a herniated disc more than 2 years before entering the study and with no persistent nerve root symptoms.
- Specific radiologic findings, such as spondylolisthesis, new or old fractures, infection, inflammatory process, or neoplasm.
- Obvious painful and disabling arthritic hip joints and also anamnestic and radiologic signs of spinal stenosis.

Table 1. Function-Working Disability Score*

Points	FUNCTION
1	No pain, can do anything, even sports
2	Occasional pain, little decrease in function
3	Moderate pain, can do many things, not sports
4	Severe pain, can manage ADL
5	Severe pain, no function
	WORKING ABILITY
1	No restrictions in working ability
2	Some restrictions in working ability
3	Can work, but with some restrictions
4	On sick leave, can manage light work
5	Completely disabled, no working ability

* A sum of 7 or more served as an inclusion criterion into the study. No patient had less than 3 in any scale.

Other comorbidity, as described by the patient, was permitted where deemed appropriate by the treating surgeon (Table 2). Both employed and unemployed patients could participate, and neither compensation nor ongoing litigation claims were used as criteria for exclusion. A detailed comparison of the study patients with an age and sex-matched sample of the Swedish population is reported separately.[24]

Discography, external fixation, facet blocks, and external support (corset) were used as diagnostic procedures after the randomization process in different subgroups of the study population. The feasibility and prognostic value of these diagnostic methods, often discussed but not routinely used, will be detailed in a separate report.

Power Calculation. The size of the study was based on the following assumptions: 1) The primary aim was to compare nonsurgical and surgical treatment, with a secondary aim of comparing three different surgical techniques with one another and with nonsurgical treatment. As a result, the two main groups (surgery and no surgery) were asymmetrical in size; 2). Outcome was estimated as either an essential improvement (1) or as no essential improvement (0) after 2 years. This was analogous to one of the primary outcome measures used in the study, where the result was assessed by the patient as an improvement expressed as either "much better" or better"(1), or as no improvement expressed as "unchanged" or "worse"(0); 3) "Essential improvement" was assumed to occur in 5% of the patients in the nonsurgically treated group; 4) The probability of detecting a statistically significant difference (power) between the main groups should be at least 80%; 5) Calculations were performed with the Fisher's exact test (two-sided) and a

Table 2. Sociodemographic Characteristics at the Time of Randomization*†

	Surgical group (n = 222)		Nonsurgical group (n = 72)		
Age (range)	43 (25–64)		44 (26–63)		
Mean pain duration, years (range)	7.8 (2–34)		8.5 (2–40)		
Mean time of sick leave, years (range)	3.2 (0.1–18)		2.9 (0.1–8)		
	%	Missing	%	Missing	*P* value
Men	49.5		48.6		NS
Married/living together	77.3	2	76.4		NS
Comorbidity	39.1	7	23.5	4	0.020
Smoking	40.6	3	49.3	1	NS
Earlier (>2 years) successful removal of herniated disc	18.6	2	19.4		NS
Employed	73.8	1	66.7		NS
Changed work because of back problems	29.4	4	35.2	1	NS
On sick leave because of back problems	58.8		54.2		NS
Disability pension because of back problems	20.3		22.2		NS
Working part or full time	20.9		23.6		NS
Litigation or compensation because of back problems	60.4		64.5		NS
Patient consider problems as back work injury	54.3	12	64.2	5	NS

* Statistical significance was calculated with Fisher's Exact test.
† Patients on disability pension were excluded when the number of patients working were calculated.

significance level of 0.05; and 6) We decided to include 75 patients in each treatment arm (225 in the surgical group with three different procedures and 75 in the control group). This size basically implied that we would achieve a significant difference (with 80% power) even if only 18% of the patients in a surgical group compared with 5% in a nonsurgical group should be essentially improved. Thus, we considered the suggested sample size being sufficient to detect an assumed clinically relevant difference or 13% (18-5) expressed as "essential improvement" between the two main groups.

Randomization. After inclusion in the study the patients were randomized blindly into one of four treatment groups using a computer-generated random list in which the individual numbers were kept in sealed envelopes outside the participating departments. The different treatments are presented below. In this article the three surgical groups are referred to as *the surgical group*.

- Group 1 (surgical group)
  - *Group 1a (n=73) posterolateral fusion (PLF).* These patients wore a rigid plastic brace during the first 5 postoperative months.[26]
  - *Group1b (n=74).* PLF as in Group 1a, and in addition an internal fixation device, the Variable Screw Placement (VSP), with pedicle screws and plates manufactured by DePuy Acromed (Raynham, MA).[42] A reinforced canvas corset was used for 5 postoperative months.
  - *Group1c (n=75).* The "circumferential" group. As in Group 1b, with additional interbody bone graft either as an anterior lumbar interbody fusion (ALIF)[37] or a posterior lumbar interbody fusion (PLIF),[5] according to the preference of the surgeon. A reinforced canvas corset was used as in Group 1b.
- Group 2 (n=72). Nonsurgical treatment, mainly commonly used physical therapies as suggested in the study protocol and with the possibility of local modifications and variations. This group served as a control.

Because of the multicentricity of this study, we have no means of being sure of exactly how many patients refused to participate before randomization, but three were reported to have done so.

Treatment Procedures.

Surgical Group. Twenty-six experienced spine surgeons who participated in the study carried out the surgical procedures. Autologous bone[27] harvested from the iliac crest was used in all cases. In the PLF procedure, bone was applied to decorticated transverse processes and facet joints. When VSP pedicle screws were used, they were inserted using fluoroscopy as a guide. For the interbody procedure, bone blocks were cut from the iliac crest and inserted between the intervertebral bodies from the posterior (PLIF) or from the anterior (ALIF). All patients were fused *in situ* with no intention of decompression, and only the segments of L4–L5 and/or L5–S1 were addressed. The rehabilitation focused on early activity but not according to any study protocol. The three surgical subgroups will be compared in a separate study.[16]

Nonsurgical Group (Control Group). A nonsurgical treatment program was constructed on a consensus basis to serve as a guideline within the study. The main component was physical therapy, which could be supplemented with other forms of treatment, such as information and education, treatment aimed at pain relief (TENS, acupuncture, injections), cognitive and functional training, and coping strategies. Thus, the treatment could vary within broad but commonly used limits reflecting the nonsurgical treatment policy in the society.

Collected Data and Outcome Measures. Pain, disability, global self-rating by the patient, and back-to-work were used as "primary outcome measures" in the study, and other measurements were used as well:

1. Sociodemographic characteristics were collected by means of standardized protocols (Table 2).
2. Clinical characteristics were also collected using standardized protocols.
3. Pain. Back and leg pain intensity was measured on a vertical Visual Analogue Scale (VAS),[23] ranging from

0–100 mm, where 100 reflected the worst pain imaginable. The measurements were done in three dimensions: "maximum pain," " minimum pain," and "current pain." The mean of the three measurements provided the pain index.

4. Disability. Pain-related disability was measured using two different disease-specific health scores and one reflecting pure functional disability:

• The Oswestry Low Back Pain Questionnaire,[14] which describes back-related disability as a combination of physical and social restriction through 10 questions covering different dimensions of daily living. A sum is calculated and presented as a percentage, where 0% represents no disability and 100% the worst possible disability.
• The Million Visual Analogue Score,[31] consisting of 15 questions reflecting back-related issues. The answer is expressed on a horizontal VAS, ranging from 0-100 mm, where 100 represents maximum disability.
• The General Function Score (GFS),[25] serving as an instrument strictly focused on physical activities of daily living and thus giving a picture of the purely physical disability inflicted by a disorder. The patient answers nine questions using one of three alternatives: "Can perform," "Can perform with difficulty," and "Cannot perform." The score is presented as a percentage where 100% represents maximum disability. This instrument was validated in a separate study.[24]

5. Depressive symptoms. The Zung Depression Scale[55] was used to measure the depressive symptoms as reflected on a score. It consists of 20 statements reflecting different aspects of well-being, each with four choices ranking the severity of each statement. The result is reported as a number between 20 and 80 but was transformed in this study to a scale ranging from 0 to100%, where 100% means most profound depression. In addition the psychological profile was investigated by two additional methods. Personality traits were assessed by means of the Karolinska Scales of Personality (KSP),[38] and personality disorders were assessed with the Structured Clinical Interview for DSM-III-R (SCID II) screen questionnaire.[12] The results from these three measurements showed no significant difference regarding depressive symptoms, the psychological profile, or psychological disturbances between the treatment groups at the time of randomization. The association between treatment outcome and personality traits and personality disorders will be discussed in separate reports.
6. Patient overall assessment. The result, according to the patient's own opinion, was reflected in response to the question, "Compared with before treatment my back problems are: 'much better, better, unchanged, or worse'."
7. Work status. This analysis was carried out after excluding all patients receiving full disability pension at the time of randomization, as it is known that in these cases almost no one returns to work.[2,17,33] For data collection we used records from the Swedish Social Insurance (SSI), which provides information on sick leave and economic compensation for Swedish residents. All hospital patient records were controlled as well, and in cases of doubt, a secretary interviewed the patient by telephone. The data from the SSI were interpreted as follows: If a patient was not on full-time sick leave or on full-time disability pension, he or she was considered to be working, *i.e.*, the patient could work part time or with modified duties. A patient who was not working at the baseline but had returned to work after 2 years was regarded as "back to work," and patients moving in the opposite direction as "stopped working." The difference between the two study groups both regarding "back to work" and what we called *net back to work* ("back to work" minus "stopped working") were calculated.
8. Compensation. The correlation between the "overall result" as assessed by the patient and the compensation status was reported.
9. Independent observer overall assessment. The result according to an independent observer, a spine surgeon not engaged in the study, was assessed using a well-described ordinal scale: excellent (no pain, no functional restrictions, no analgesics), good (sporadic pain, slight restriction of function, occasional analgesics), fair (moderate pain, moderate restriction, no sports, daily analgesics), and poor (moderate daily or occasional severe pain, function restricted to ADL, use of strong analgesics.)
10. Go through again. Two years after the start of treatment, the patients answered the question "Knowing the result, would you go through the treatment again?"
11. Radiographic fusion. One experienced radiologist, blinded to the clinical outcome, assessed the fusion rate. Both the posterolateral and the intercorporal fusion were assessed on a scale of 1 to 4, which was converted into three levels for the presentation (1, definitely fused; 2, uncertain fusion; 3, definitely pseudarthrosis).

Anteroposterior and lateral radiographs of the lumbosacral spine were obtained. Fusion was defined as trabeculae crossing the graft–transverse process interface on both sides, with evidence of increasing density of the graft with cortication. In case of an interbody procedure, fusion was defined as trabeculae crossing the graft–vertebral body interface on both sides of the graft. Fusion was accepted only if both posterolateral areas were fused in Group 1 and Group 2, but in Group 3 it was considered sufficient if the interbody fusion was successful as defined by Greenough et al.[19]

12. Complications were reported.

Follow-up. Pain and complications were measured after 6 and 12 months in addition to the 2-year follow-up. This was done partly for ethical reasons to control for unexpected results in any group, in which case the study could have been stopped. Two years after the onset of treatment the routines of data collection before randomization were repeated by means of the same questionnaires and protocols in both groups.

"Group Changers". All patients who for various reasons did not complete treatment according to the randomization, "group changers" (n=25), were examined at the 2-year follow-up. These patients did not differ from the main study population in terms of sociodemographic and clinical characteristics. The results from patients who changed groups have been included in the statistical analyses in accordance with the intention-to-treat principle, *i.e.,* their outcome results were attributed to the group to which they were randomized and not to the group in which they were actually treated.

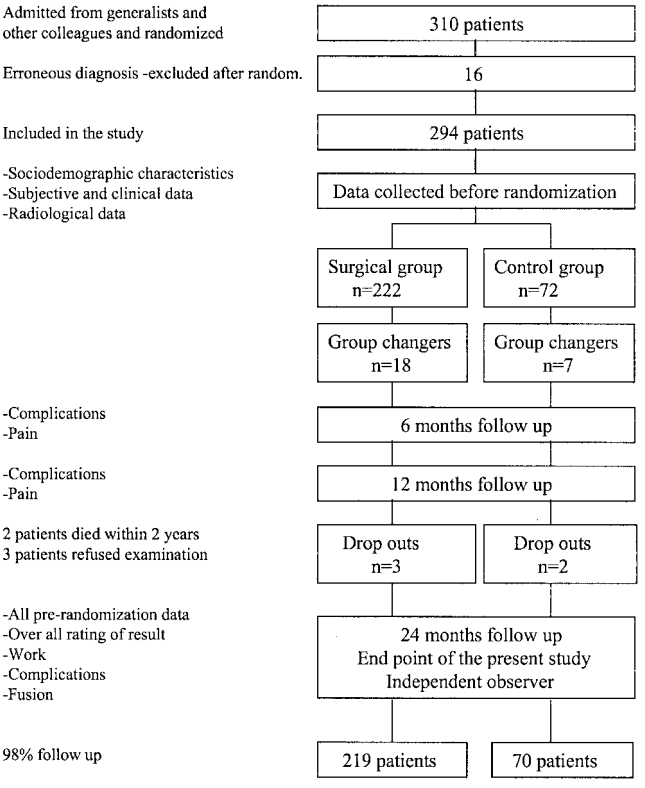

Figure 1. Flow diagram describing the patients during the study period.

Dropouts. Dropouts (n=5) were randomized patients who for various reasons could not be followed up after 2 years.

Statistical Methods. All data were entered into the SPSS statistic program (version 10 SPSS, Inc., Chicago, IL). Because our variables were basically ordinal data and continuous variables (Visual Analogue Scales) are not exactly normally distributed, nonparametric methods were used for calculation of statistical significance. The Mann-Whitney test, Fisher's Exact test, and the test for trends in contingency table were considered to be appropriate to test for the differences between groups, and the Wilcoxon Signed Rank test was used for differences within groups. The Spearman's Rank Correlation test was used to test the correlation between variables. The "intention-to-treat principle" was used in the analyses to assess the significance of the difference between surgical and nonsurgical treatment. Two-tailed tests were used.

■ Results

A total of 310 patients were randomized into the study during 1992 through 1998. After the randomization procedure, 16 patients were excluded because of erroneous diagnoses (spondylolisthesis and spinal stenosis), leaving 294 patients allocated to either surgical treatment (n=222) or nonsurgical treatment (n=72). The follow-up rate after 2 years was 98% (Figure 1). Men comprised 49.5% of the surgical group and 48.6% of the nonsurgical group. The mean age was 43 years in the surgical and 44 years in the nonsurgical group. The mean duration of CLBP was 7.8 years (range 2–34) in the surgical group and 8.5 years (range 2–40) in the nonsurgical group. Patients in the surgical group had been on sick leave because of back pain for a mean duration of 3.2 years (range 0.1–18), and in the nonsurgical group for 2.9 years (range 0.1–8).

In all, fusion was performed on 211 of 222 patients (18 patients from the surgical group changed to nonsurgical treatment, and 7 patients changed in the opposite direction). Of these 211 patients, 122 were fused at a single level and 89 at 2 levels. A total of 668 pedicle screws were inserted in 140 patients. Patients in the surgical group spent an average of 9.5 days in the hospital after surgery. The number of patients who went on to repeat unintended surgery was 16 of 211 (7.8%).

Sociodemographic characteristics. There were no significant differences between the groups regarding the sociodemographic characteristics, except for comorbidity, which was significantly more frequent in the surgical group ($P=0.020$) (Table 2).

Clinical characteristics. There were no differences between the groups regarding clinical characteristics, *e.g.*,

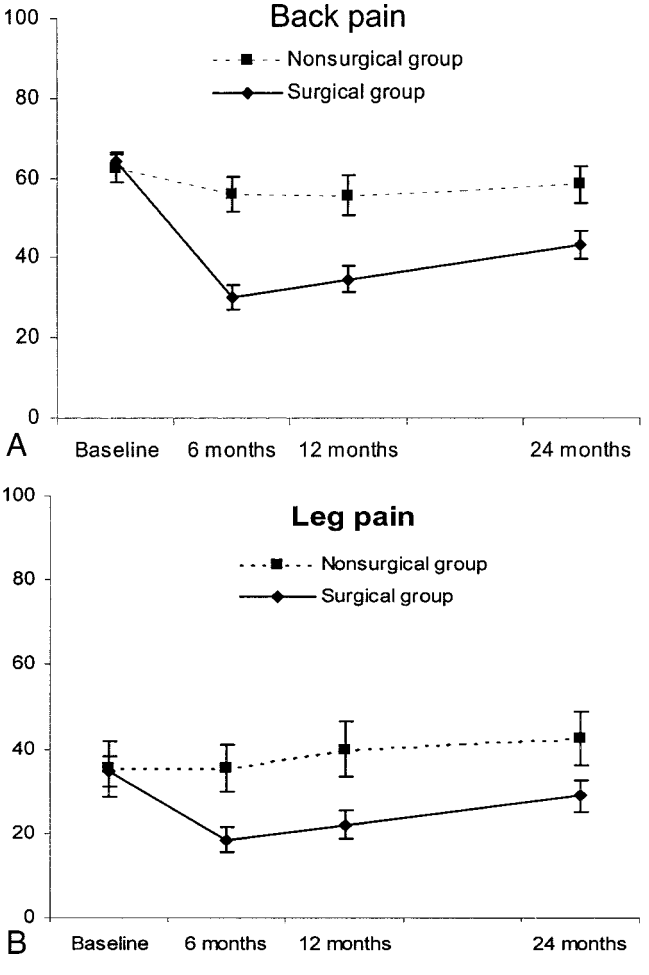

Figure 2. **A, B,** Back and leg pain expressed on a visual analoguescale (VAS) at baseline, 6, 12, and 24 months. 100 represents maximal experienced pain. Mean values with 95% confidence interval. The *P* values refer to the differences between groups, and were calculated with the Mann-Whitney U-test*. Back pain increased significantly between 1 and 2 years ($P < 0.0001$), Wilcoxon Sign Rank test.

Table 3. Back and Leg Pain, Disability, and Depressive Symptoms

	Surgical group (n = 201)				Nonsurgical group (n = 63)				Diff baseline-2 years			
	Baseline	2 years fu	Diff %	*P*-value*	Missing	Baseline	2 years fu	Diff %	*P*-value*	Missing	Surgery	Non-surg	Diff *P*-value
VAS Back	64.2 (14.3)	43.2 (25.2)	32.7	<0.0001	6	62.6 (14.3)	58.3 (18.8)	6.8	0.017	1	21.0	4.3	0.0002
VAS Leg	35.3 (25.4)	29.0 (27.0)	17.8	0.002	9	35.6 (25.2)	42.6 (24.8)	−20.5	0.219	4	6.3	−7.0	0.005
ODI	47.3 (11.4)	35.7 (18.0)	24.5	<0.0001	4	48.4 (11.9)	45.6 (16.1)	5.8	0.025	1	11.6	2.8	0.015
MVAS	63.7 (11.3)	45.6 (23.1)	28.4	<0.0001	4	65.5 (11.5)	60.4 (17.2)	7.8	0.021	2	18.1	5.1	0.004
GFS	49.1 (15.9)	34.1 (22.4)	30.5	<0.0001	7	47.6 (16.3)	45.5 (20.3)	4.4	0.073	2	15.0	2.1	0.005
Zung	39.1 (13.3)	31.4 (15.2)	19.7	<0.0001	10	39.4 (13.9)	36.7 (14.5)	6.9	0.041	2	7.7	2.7	0.123

Group changers + dropouts (18 + 3 in the surgical group and 7 + 2 in the nonsurgical group) were excluded from the illustrated results, but group changers were included in the analyses of difference, consistent with the intention to treat principle.
The difference between the baseline value and the value after two years (diff %) was calculated on group level, and not on an individual level.
Pain was measured in mm on a vertical Visual Analogue Scale (VAS).
Disability was measured with the Oswestry Low Back Pain Questionnaire, the Million Score (MVAS), and General Function Scale (GFS).
Depression was measured with the Zung Depression Scale. All scales ranging from 0 to maximum severity 100.
All values are illustrated as means (Standard deviation within brackets), but nonparametric tests were used for statistical analyses.
The significance of the difference between baseline and 2 years follow-up within each group was calculated with the Wilcoxon Signed Rank test.*
The difference between the change in the two groups after 2 years was analysed with the Mann Whitney U Test.

reflexes, sensibility, muscular strength, or straight leg raising test. All patients recognized their maximum pain location on provocation (springing test) over L4–L5 and/or L5–S1.

Pain. At the 2-year follow-up, the reduction in both back and leg pain was significantly greater in patients treated surgically compared with those treated nonsurgically ($P=0.0002$ and $P=0.005$, respectively) (Table 3). In the analysis including the follow-up at 6 and 12 months, there was a highly significant reduction of both back and leg pain in the surgical group on all measuring points compared with the preoperative baseline values and also compared with the nonsurgical group. The back pain component in the nonsurgical group also improved significantly, but to a much lesser extent. Back pain increased significantly between 1 and 2 years in the surgical group ($P<0.0001$) (Figure 2).

Disability. According to all instruments used, the decrease in disability was significantly greater in the surgical group after 2 years: Oswestry ($P=0.015$), Million ($P=0.004$), and the GFS ($P=0.005$) (Table 3).

Depressive symptoms. The depression score was significantly reduced in the surgical group ($P<0.0001$) and also significantly in the nonsurgical group ($P=0.041$). The difference of reduction between the groups was not significant ($P=0.123$) (Table 3).

Patient overall assessment. The patients rated the "overall result" as improvement ("much better" or "better") or no improvement ("unchanged" or "worse") of back problems. The result was significantly better in the surgical group, with 63% reporting to be improved compared with 29% in the nonsurgical group ($P<0.0001$) (Table 4). The intention-to-treat principle was used in analyzing the significance of the difference. In addition, when a worst-case analysis was performed where those patients who did not answer this question in the surgical group (n=14) were regarded as "worse," and those who did not answer in the nonsurgical group (n=4) were regarded as improved, the result was 56% improved in the surgical group compared with 38% in the nonsurgical group ($P=0.01$).

Work status. At the time of randomization, 21% (63/294) of the patients received full pension or supplementary disability pension (21% [47/222] in the surgical group, and 22% [16/72] in the control group). These patients were excluded from the analysis because only three resumed part-time work after 2 years. Drop outs (n=5) were also excluded in the analysis, whereas group changers were included according to intention-to-treat. Using the χ^2 test, there was a significant difference in favor of surgery expressed as "net back to work" ($P=0.002$) and also as "back to work" ($P=0.049$) (Table 6).

Compensation. Of those patients not on litigation or compensation in the surgical group at the time of randomization, 70% (47/67) considered themselves as "much better" or "better" after 2 years compared with 58% (57/98) of those on litigation or compensation ($P=0.085$). The corresponding figures for being "much better" or "better" in the nonsurgical group was 50% (10/20) among those not on litigation/compensation and

Table 4. Patient Overall Rating of the Result After Two Years

	Surgical group n = 195 (missing 6)	Nonsurgical group n = 62 (missing 1)
	%	%
Much Better	28.8	14.5
Better	33.8	14.5
Unchanged	23.6	45.2
Worse	13.8	25.8
Total	100	100

$P < 0.0001$*.
(Group changers + dropouts [18 + 3 in the surgical group and 7 + 2 in the nonsurgical group], were excluded from the illustrated results but group changers were included in the analysis of difference.
Differences between the groups was tested by the test for trend in contingency table*.

Table 5. Independent Observer Overall Rating of the Result After Two Years

	Surgical group n = 200 (missing 1)	Nonsurgical group n = 62 (missing 1)
	%	%
Excellent	16.5	1.6
Good	29.0	16.1
Fair	32.5	40.3
Poor	22.0	41.9
Total	100	100

$P < 0.0001$*.
Group changers + dropouts (18 + 3 in the surgical group and 7 + 2 in the nonsurgical group), were excluded from the illustrated results but group changers were included in the analysis of difference.
Differences between the groups was tested by the test for trend in contingency table*.

18% (6/34) among those on litigation/compensation (P=0.043).

Independent observer overall assessment. The independent observer rated his overall impression of the results significantly in favor of surgery. When the assessments "excellent" and "good" were combined, 45% in the surgical group fell into this category compared with 18% in the nonsurgical group (P=0.005). When "fair" was included, 78% in the surgical group and 58% in the nonsurgical group were assessed to belong to this combined category (P=0.003) (Table 5).

Go through again. In the surgical group, 75% (145/193) of the patients were willing to go through the treatment again knowing the result after 2 years. In the nonsurgical group, 53% (30/57) would do so (P=0.002).

Radiographic fusion. Of the surgically treated patients, 83% were assessed as having a radiologic fusion. There was no significant correlation between radiographic fusion and the patient's overall rating of the result or the improvement in pain and disability.

Complications. There were no deaths in connection with surgery. A total of 17% of the patients in the surgical group suffered an early complication (Table 7). Most complications were handled with no obvious sequelae for the patient, but 9 of 140 (6%) patients in the 2 groups where pedicle screws were used experienced a new sensation of nerve root pain of various intensity and duration. In three of these patients a screw was subsequently removed, with immediate improvement of symptoms in one case. Three deep (1.4%) and two superficial infections were successfully handled with debridement. In two cases the implant was involved but was left in place with no consequence for a successful healing. One patient suffered from persistent wing scapula and shoulder weakness after surgery due to an injury to the long thoracic nerve, which was attributed to the intraoperative positioning. One patient had severe and persistent pain problems due to an injury to the nervus cutaneus femoris lateralis at the donor site. Complications after the postoperative period (2 weeks) were referred to as *late*. After 6 and 12 months, two patients suffered a late implant-related infection, with subsequent removal of the implant. Surgery and re-fusion were performed on two patients having a pseudarthrosis within 1 year, though without any obvious benefit to the patient. One patient was operated at the wrong level and was reoperated at the intended level 6 months later. Radiologic examination showed both levels were healed after 2 years. The result was included and reported 2 years after the second operation. The patient received compensation according to the regulations of the Swedish patient insurance company. There were no obvious complications in the nonsurgical group, but three patients threatened to commit suicide if they did not have an operation.

"Group Changers"

All group changers were examined after 2 years. The results in patients who changed from surgical treatment to the nonsurgical group (n=18) were on the whole not better than the average results in this group. On the other hand, six of the seven patients who changed from nonsurgical to the surgical group did far better than the average patient in that group. In this subgroup of seven patients, including three patients who threatened to commit suicide if not operated, back pain according to VAS was reduced by 61% (67 to 26) and leg pain was reduced by 75% (39 to 10). Disability according to Oswestry decreased by 70% (55 to 16), according to Million by

Table 6. Work Status

Individual work status baseline/ follow up	Surgical group		Nonsurgical group		
	n patients	%	n patients	%	*P*-value*
1. Work/work	12	8	5	10	NS
2. Not work/not work	79	49	27	57	NS
3. Work/not work (stopped working)	6	4	5	10	NS
4. Not work/work (back to work)	63	39	11	23	0.049
Total	160	100	48	100	
"Net back to work"#	57	36	6	13	0.002

The subgroup of patients who stopped working were subtracted from those who went back to work (63-6 in the surgical group and 11-5 in the nonsurgical group).
Patients who were on full-time pension/disability pension at baseline were excluded (surgical group n = 47 and nonsurgical group n = 16) together with drop outs (n = 5). Group changers (n = 25) were excluded from the illustration but included in the analysis according to interntion to treat. Fischers exact test* was used to analyze differences between the groups.

Table 7. Complications

	Major	Minor
Early	6 New sensation of nerve root pain, no reoperation	3 Gastro-intestinal bleeding
	*3 Nerve root hit by pedicle screw, reoperation	2 Laterally placed screw
	*3 Deep wound infections, two implant related	*2 Hematomas at the donor site
	2 Major bleedings during surgery (ALIF)	2 Sympathetic cord damage with symptoms
	1 Thrombosis + pulmonary embolus	*2 Superficial wound infection
	1 Thrombosis	2 Skin problem after surgery
	1 Aspiration-sepsis-ARDS	1 Dural tear
	1 Pulmonary oedema	1 Wingscapula after surgery
	1 Heart failure + GI bleeding	1 Pain in the arm after surgery
	*1 Patient operated on wrong level	*1 Injury of the nervus cut fem lat
Late	*2 Deep wound infections that led to surgery	9 Donor site pain
	*2 Pseudarthroses that led to new surgery	

Complications were classified as early and late, and major and minor. In the surgical group 37/211 patients experienced an early (within two weeks) complication (18%). 13 patients experienced a late (after more than 2 weeks) complication (6%). All patients with a new sensation of nerve root pain (9/140) were in the pedicle screw groups. There were 16 unintended reoperations* (7.8%).

70% (68 to 20), and according to GFS by 71% (61 to 18). The Zung depression score was reduced by 44% (33 to 19). Of the group changers who changed from the nonsurgical group to surgery, 4 of 7 assessed themselves as "much better," whereas 2 of 18 of the group changers that changed from the surgical group to the nonsurgical group did so ($P=0.032$), (Tables 8, 9, and 10).

Drop Outs

Five patients dropped out of the study. One patient in the surgical group and two patients in the nonsurgical group refused to answer the questionnaires at the 2-year follow-up, and two patients in the surgical group died within 2 years after treatment start. One patient had a fatal cardiac infarction 1 year after surgery, and one patient committed suicide 2 years after surgery for reasons that according to the treating surgeon were not back related.

■ Discussion

Statistically, all primary outcome measures in the study were significantly in favor of surgery. The RCT setting made it possible to control for confounders such as age, sex, duration of pain, selection bias, and compensation, and also the natural history and regression to the mean, which could have distorted the results. This, together with the comparably homogeneous patient group, the blinded randomization procedure, the standardized treatment and data collection routines, the 98% follow-up rate by an independent observer, and the use of the intention-to-treat principle in the data analysis should make the results in this study valid.

The use of the intention-to-treat principle in the analysis (*i.e.,* the results of those who changed groups were included in the group to which they were originally randomized and not in the actual treatment group) meant that the results from the 18 patients who changed from the surgical to the nonsurgical group were approximately the same as the average results in that group. However, the group of seven patients who changed from nonsurgical to surgical treatment did far better than the average patient in the nonsurgical group to which their results were attributed. Consequently, the use of the intention-to-treat principle improved the results in the nonsurgical group as compared with the surgical group,

Table 8. Group Changers

	From surgical to non-surgical group (NS) n = 18					From non-surgical to surgical group (S) n = 7					Diff baseline—2 years†		
	Baseline	2 years fu	Diff %	*P*-value*	Missing	Baseline	2 years fu	Diff %	*P*-value*	Missing	NS	S	Diff *P*-value
VAS Back	65.4 (15.8)	50.4 (23.3)	22.9	0.233	10	67.2 (17.3)	26.1 (28.5)	61.2	0.046	1	15	41.1	0.228
VAS Leg	36.2 (26.4)	37.5 (29.2)	−3.6	0.833	10	39.4 (24.7)	10.0 (18.4)	74.6	0.041	1	−1.3	29.4	0.043
ODI	44.1 (11.8)	44.9 (21.1)	−1.8	0.799	10	54.6 (11.6)	16.3 (22.9)	70.1	0.028		−0.8	38.3	0.009
MVAS	58.5 (12.1)	57.4 (17.8)	1.9	0.878	7	67.6 (18.2)	20.4 (29.2)	69.8	0.028		1.1	47.2	0.015
GFS	42.1 (14.7)	42.5 (19.6)	−0.9	0.286	6	60.7 (11.7)	17.5 (22.3)	71.2	0.018		−0.4	43.2	0.001
Zung	36.1 (15.6)	38.5 (19.8)	−6.6	0.755	7	33.3 (17.9)	18.8 (12.8)	43.5	0.042		−2.4	14.5	0.055

Dropouts were excluded from the analysis (1 in the surgical group and 1 in the nonsurgical group).
The difference between the baseline value and the value after two years, diff %, was calculated on group level, and not on an individual level.
Back and Leg Pain, Disability and Depression. Pain was measured on a vertical Visual Analogue Scale (VAS).
Disability was measured with the Oswestry Disability Index (ODI), the Million Scale (MVAS), and General Function Score (GFS).
Depression was measured with the Zung Depression Scale. All scales ranging from 0 to maximum severity (100).
All values are illustrated as means (Std within brackets), but nonparametric tests were used for statistical analyses.
The significance of the difference between baseline and 2-year follow up within each group was calculated with the Wilcoxon Signed Rank Test*.
The difference between the change in the two groups after 2 years was analysed with the Mann Whitney U Test†.

Table 9. Patient Overall Rating of the Result After Two Years Among Those Who Changed Groups From Surgical to Nonsurgical Group (NS), From Nonsurgical to Surgical Group (S)

	NS n = 13 (missing 5)	S n = 6 (missing 1)
	%	%
Much Better	15.4	66.7
Better	7.7	16.7
Unchanged	53.8	16.7
Worse	23.1	0

* $P = 0.024$.
Test for trends in contigency table* was used to calculate the difference between the groups.

and since despite this fact, the results in the surgical group were significantly superior, we conclude that the difference between the groups was statistically relevant. In addition to the use of the intention-to-treat principle, we also performed a worst-case analysis without any change in the significance of the results. We would achieve the desired 80% power (chance not to make a Type 2 error, *i.e.*, missing a significant difference) if 18% in the surgical group would be "much better" or "better" and the improvement in the nonsurgical group was at most 5%, but the number of patients who were improved was in fact 63% *versus* 29%, resulting in a power according to self-rated improvement of 99.7%. Power according to the other primary outcome measurements was also calculated: VAS, 99.6%; Oswestry, 61%; Million, 81.5%; GFS, 99.6%; and net back to work, 86%.

The increased comorbidity in the surgical group at the time of treatment start ($P=0.020$) occurred despite the randomization procedure, but we do not believe that this changed the outcome in favor of surgery, as others have reported that patients with increased comorbidity also have an increased disability.[24] The placebo effect after surgery is known to be profound[47] and more powerful than that of nonsurgical treatment methods. This can probably explain some of the difference between the groups regarding pain after 6 and 12 months (Figure 2), but we do not believe that this effect can explain the differences between the groups after 2 years.

In the analysis of pain that also included information at 6 and 12 months, the reduction of back pain in the surgical group was superior on all occasions compared with the nonsurgical group. An interesting observation was that pain in both groups increased between 1 and 2 years. This was significant ($P<0.0001$) only in the surgical group, possibly because the improvement in the nonsurgical group in the first place was too small. This finding indicates the importance of these measuring points and supports the view that longer observation times are necessary for this group of patients. The results after 5 years will be presented in a coming report. The leg pain component improved by 18% in the surgical group, whereas it worsened by 21% in the nonsurgical group. We believe that this nonradiculopathy leg pain component is referred from the back, and that both improve through the stabilizing effect of the fusion procedure.

Before treatment, patients in the present study were shown to have more depressive symptoms than a sex- and age-matched control group,[24] and patients with depression have been shown to do worse after surgery[43,48] than nondepressed patients. The improvement of the depression score was however significant in both the surgical ($P<0.0001$) and the nonsurgical ($P=0.041$) groups. Consequently, depressive symptoms did not seem to constitute a contraindication to lumbar fusion in this selected group of patients.

Satisfactory results after different surgical fusion techniques used in patients with low back pain has been shown in the literature to vary between 16–95% and the reported "overall" improvement in the present surgical group of 63% is comparable to the results reported in other prospective studies; retrospective studies tend to report better results.[46] In the present study, patients in the surgical group were rehabilitated according to local preferences, with a focus on early mobilization and informational support. As no specific or extensive exercise program was used routinely, it is probable that the result in this group is not particularly influenced by additional nonsurgical treatment but rather is reflective of surgery alone.

In the beginning of the study, the patients randomized to the nonsurgical group were thought to be a threat to the study compliance. The concern was that a patient who had tried most conservative treatment modalities would not agree to participate if randomized to this group. However, this did not prove to be a problem. Maybe surprisingly, patients allocated to surgery were as prone to changing groups as patients allocated to nonsurgical treatment. In the surgical group, 18 of 222 (8%) changed to nonsurgical treatment, and 7 of 72 (10%) changed from nonsurgical treatment to the surgical group.

The treatment in the nonsurgical group was physical therapy according to the study protocol. This could be carried out according to local preferences and according

Table 10. Independent Observer Overall Rating of the Result After Two Years Among Patients Who Changed Groups From Surgical to Nonsurgical Group (NS), From Nonsurgical to Surgical Group (S)

	NS (n = 18)	S (n = 7)
	%	%
Excellent	0	28.6
Good	33.3	42.9
Fair	38.9	14.3
Poor	27.8	14.3

* $P = 0.09$.
Test for trends in contigency table* was used to calculate the difference between the groups.

to the earlier experience of each patient. Consequently, no homogeneous physical therapy or exercise program was used. It can be argued that continuing nonsurgical therapy of the same kind that patients received before has little chance to improve the situation and that these patients might resemble a waiting list control group where no improvement is to be expected. There is a point in this comparison, but we do not believe that our control group is quite analogous to waiting list controls because in 46% of the cases it was explicitly stated in the patient records that special rehabilitation was initiated as a consequence of the study and that results in the control group in fact were significantly better after 2 years. We obviously did not have full control over the treatment given in this group, but we consider that it is appropriate to say that these patients (on a group level) received nonsurgical treatment above normal compared with what patients not participating in the study would have received, and in that respect resembled an extensive "existing practice" in society during the study period.

In a Cochrane meta-analysis from 1997, van Tulder et al[49] found strong scientific evidence for different exercise programs being better than placebo and natural history, at least in the short run. Further research is therefore needed to compare long-term outcome after surgical fusion to those and other well-described nonsurgical methods.[29]

In the surgical group, 63% (122/195) rated themselves as "much better" or "better," but 75% (145/193) were willing to go through treatment again knowing the result. In the nonsurgical group, 29% (18/62) rated themselves as "much better" or "better," and 53% (30/57) were willing to go through the treatment again. Thus, even in the absence of improvement, 12% in the surgical group and 24% in the nonsurgical group did not regret the treatment. This might reflect a feeling of having been taken care of, as is suggested by Möller and Hedlund in their randomized study comparing fusion with physiotherapy in patients with spondylolisthesis.[32] In joining a scientific study some patients might also experience recognition in relation to their suffering, both by the profession and by society. Several patients who reported they were "unchanged or worse" in both groups actually expressed a feeling of satisfaction because they were now being "taken seriously." In fact 22% (6/27) patients in the surgical group who assessed themselves as "worse" were willing to go through the treatment again; the corresponding figure in the nonsurgical group was 31% (5/16). On the other hand, 5% (6/121) of the patients in the surgical group reporting they were "much better" or "better," and correspondingly 17% (3/18) in the nonsurgical group did not want to go through the treatment again. Consequently, the answers given to the question "Would you, knowing the result, go through the same treatment again?" was hard to interpret and did not correlate well to the subjective overall rating of the outcome.

The early complication rate of 17% in the surgical group is comparable to other reports.[9] Most complications were reversible, but 9 of 140 (6%) patients in whom pedicle screws had been used had a new sensation of nerve root pain of various intensity and duration. In six cases there were no visible malpositioning of a screw on plain films. The exact etiology of this postoperative leg pain was not always clear, but it must be suspected that a screw caused it, at least in those patients in whom the spinal canal was not exposed (n=7). The importance of the patients being properly positioned during what were often lengthy surgical procedures was demonstrated by the fact that one patient had persistent shoulder weakness due to a nerve injury caused by overstretching, and two patients had reversible skin pressure injuries.

The results from the present study were achieved in a subgroup of patients with severe CLBP and a mean duration of pain of about 8 years and on sick leave or with equivalent disability for at least the last preceding year (the mean total, but sometimes interrupted, length of sick leave due to low back pain was about 3 years in both groups). This means that the patients were not ideal from a therapeutic standpoint. It has been reported that patients who have been out of work because of back pain for 2 years have a very slim chance ("close to zero") of returning to work.[33,41] In this light we think that the "net back to work" in the surgical group of 57 of 160 patients is satisfactory. We believe, however, that it may be difficult to use the work status as a measure of success. Since a variety of circumstances of both socioeconomic and psychological nature may influence the work status[39,44] it is possible that the return to work rate will differ, even between randomized groups, depending on changes in the socioeconomic situation. For example, if jobs are readily available, patients with less successful outcome after treatment will have the opportunity to work, whereas if there is a job shortage, only patients with a highly successful outcome will have that opportunity.

We used plain radiographic films to estimate the fusion rate. This method has been shown to be inexact,[4,28] partly because the stainless steel implant makes it hard to visualize all transplanted areas but also because it is sometimes impossible to visualize a small defect in transplantation mass. The implant made the use of CT difficult, at least as far as the posterolateral fusion area was concerned, and the use of MRI impossible. The reported radiographic fusion rate after using plain radiographic images varies in the literature,[3,20] and we cannot claim that the 83% fusion rate reported in this study is a "true figure."

There was still a considerable amount of both pain and disability reported 2 years after treatment start even in the surgical group, and it is important to realize that lumbar fusion in this patient category very seldom cures the patient. In this study: "only" 29% assessed themselves as "much better" in the surgical group compared with 14% in the nonsurgical group ($P=0.029$). The goal

instead must be to reduce pain and disability to such extent that it makes an important difference to the patient. In the surgical group the results reflecting both these variables indicated a substantial improvement averaging 25–33%, and we consider that these results could be interpreted clinically as better than in the nonsurgical group where the corresponding improvements ranged 4–8%. In addition, the depression score was reduced by 20% in the surgical group *versus* 7% in the nonsurgical group, although this difference was not statistically significant. The findings that 63% in the surgical group assessed themselves as "much better" or "better" compared with 29% in the nonsurgical group and that significantly more patients from the surgical group returned to work strengthened the impression of a clinically important improvement as a result of surgery.

When the first 100 patients were compared with the last 100 reflecting treatment from 1992 through 1994 compared with 1995 through 1998, there were no significant differences regarding pain, disability, depressive symptoms or global rating by the patient or the independent observer in any group. Thus, there was in this series no detectable "learning curve effect" in the surgical group. All participating physicians were, however, trained spinal surgeons already when the study commenced.

■ Summary

In this multicenter, randomized, controlled trial conducted on a comparably well-defined patient group with severe CLBP and with radiographic signs of disc degeneration and spondylosis, the improvement of pain and disability after surgical fusion was significantly superior to that of the nonsurgical treatments used. We conclude that lumbar fusion can be used to reduce pain and decrease disability in carefully selected and well-informed patients suffering from CLBP.

■ Appendix

The Swedish Spine Study Group: O. Andréen, MD, PhD; G. Appelgren, MD, Halmstad; S. Berg, MD, Uppsala; B. Branth, MD, Stockholm; C. G. Cederlund, MD, Göteborg; P. Elkan, MD, Stockholm; P. Fritzell, MD, Falun; R. Hedlund, MD, PhD, Stockholm; O.Hägg, MD, Göteborg; H. Kogler, MD, Örebro; C. Leufvén MD, Eskilstuna; G. Németh, MD, PhD, Stockholm; P. Neumann, MD, PhD, Göteborg; M. Nilsson, MD, PhD, Stockholm; K. Nordenström, MD, PhD, Karlstad; A. Nordwall, MD, PhD, Göteborg; A. Ohlin, MD, PhD, Malmö; G. Ordeberg, MD, PhD, Uppsala; T. Reigo, MD, PhD, Linköping; T. Sahlstrand, MD, PhD, Helsingborg; R. Sandberg, MD, PhD, Karlstad; L. Skogland, MD, PhD, Oslo; B. Strömqvist, MD, PhD, Lund; H. Tropp, MD, PhD, Norrköping; T. Tullberg, MD, PhD, Stockholm; T. Wikström, MD, Sundsvall; J. Willén, MD, PhD, Göteborg.

Acknowledgment

The authors thank Anders Odén, PhD, statistical advisor, Romelanda, Sweden.

■ Key Point

• This multicenter, randomized, controlled trial compares surgical fusion of the lower lumbar spine (L4–L5 and/or L5–S1) with nonsurgical treatment in 294 patients suffering from chronic low back pain. Using scientifically valid methods it is shown that lumbar fusion improves both pain and disability and also compares favorably with not specified but commonly used nonsurgical treatments.

References

1. Abraham DJ, Herkowitz HN, Katz JN. Indications for thoracic and lumbar spine fusion and trends in use. Orthop Clin North Am 1998;29:803.
2. Andersson GB. Epidemiological features of chronic low-back pain. Lancet 1999;14;354(9178):581–5.
3. Blumenthal SL, Gill K. Can lumbar spine radiographs accurately determine fusion in postoperative patients? Correlation of routine radiographs with a second surgical look at lumbar fusions. Spine 1993;18:1186–9.
4. Brodsky AE, Kovalsky ES, Khalil MA. Correlation of radiologic assessment of lumbar spine fusions with surgical exploration. Spine 1991;16(Suppl): S261–5.
5. Cloward RB. Lesions of the intervertebral discs and their treatment by interbody fusion methods. Clin Orthop 1963;27:51–77.
6. Cochrane AL. Effectiveness and efficiency: random reflections on health services. Cambridge, UK: The Nuffield Provincial Hospitals Trust, 1972.
7. Davis H. Increasing rates of cervical and lumbar spine surgery in the United States 1979–1990. Spine 1994;19:1117–23, discussion 1123.
8. Deyo RA, Cherkin D, Conrad D, et al. Cost, controversy, crisis: low back pain and the health of the public. Annu Rev Public Health 1991;12;141–56.
9. Deyo RA, Cherkin DC, Loeser JD, et al. Morbidity and mortality in association with operations on the lumbar spine: the influence of age, diagnosis, and procedure. J Bone Joint Surg Am 1992;74:536–43.
10. Deyo RA, Ciol MA, Cherkin DC, et al. Lumbar spinal fusion: a cohort study of complications, reoperations, and resource use in the Medicare population. Spine 1993;18:1463–70.
11. Deyo RA. Practice variations, treatment fads, rising disability: do we need a new clinical research paradigm? Spine 1993;18:2153–62.
12. Ekselius L, Lindstrom E, von Knorring L, et al. SCID II interviews and the SCID Screen questionnaire as diagnostic tools for personality disorders in DSM-III-R. Acta Psychiatr Scand 1994;90:120–3.
13. Elam K, Taylor V, Ciol MA, et al. Impact of a worker's compensation practice guideline on lumbar spine fusion in Washington State. Med Care 1997;35:417–24.
14. Fairbank JC, Couper J, Davies JB, et al. The Oswestry Low Back Pain Disability Questionnaire. Physiotherapy 1980;66:271–3.
15. Franklin GM, Haug J, Heyer NJ, et al. Outcome of lumbar fusion in Washington state workers' compensation [see comments]. Spine 1994;19:1897–903; discussion, 1904.
16. Fritzell PF, Hägg OH, Wessberg P, et al. Chronic Low Back Pain and Fusion - A Comparison of Three Surgical Techniques. A Prospective Multicenter Randomized Study from the Swedish Lumbar Spine Study Group. Spine, in press.
17. Gibson JN, Grant IC, Waddell G. The Cochrane review of surgery for lumbar disc prolapse and degenerative lumbar spondylosis. Spine 1999;24: 1820–32.
18. Greenough CG. Results of treatment of lumbar spine disorders: effects of assessment techniques and confounding factors. Acta Orthop Scand Suppl 1993;251:126–9.
19. Greenough CG, Taylor LJ, Fraser RD. Anterior lumbar fusion: a comparison of noncompensation patients with compensation patients. Clin Orthop 1994;300:30–7.
20. Greenough CG, Peterson MD, Hadlow S, et al. Instrumented posterolateral lumbar fusion: results and comparison with anterior interbody fusion. Spine 1998;23:479–86.
21. Hart LG, Deyo RA, Cherkin DC. Physicians office visits for low back pain:

frequency, clinical evaluation, and treatment patterns from a US national survey. Spine 1995;20:11–9.
22. Herkowitz H. Lumbar spine fusions in the treatment of degenerative conditions: current indications and recommendations. J Am Acad Orthop Surgeons 1995;3:123–35.
23. Huskisson EC. Measurement of pain. Lancet 1974;2:1127–31.
24. Hägg O, Fritzell, P, Nordwall A, et al. Characteristics of patients with chronic low back pain selected for surgery. A comparison with the general population. A report from the Swedish Lumbar Spine Study. Submitted to Spine.
25. Hägg O, Fritzell P, Romberg K, et al. The general function score: a useful tool for measurment of physical disability. Validity and reliability. Accepted for publication in European Spine Journal.
26. Johnsson R, Stromqvist B, Axelsson P, et al. Influence of spinal immobilization on consolidation of posterolateral lumbosacral fusion. a roentgen stereophotogrammetric and radiographic analysis. Spine 1992;17:16–21.
27. Jörgenson SS, Lowe TG, France J, et al. A prospective analysis of autograft versus allograft in posterolateral lumbar fusion in the same patient: a minimum of 1-year follow-up in 144 patients [see comments]. Spine 1994;19: 2048–53.
28. Kant AP, Daum WJ, Dean SM, et al. Evaluation of lumbar spine fusion: plain radiographs versus direct surgical exploration and observation. Spine 1995; 20:2313–7.
29. Linton SJ. Utility of cognitive-behavioural psychological treatments. In: Neck and Back Pain. Philadelphia: Lippincott Williams & Wilkins, 2000: 361–381.
30. Mannion AF, Muntener M, Taimela S, et al. A randomized clinical trial of three active therapies for chronic low back pain. Spine 1999;24:2435–48.
31. Million R, Hall W, Nilsen KH, et al. Assessment of the progress of the back-pain patient: 1981 Volvo Award in Clinical Science. Spine 1982;7: 204–12.
32. Möller H, Hedlund R. Surgery versus conservative tratment in adult isthmic spondylolisthesis. A prospective study. Part 1. Spine 2000;25:1711–5.
33. Nachemson A. Ont i ryggen - orsaker, diagnostik och behandling. Anonymous. 108, SBU Stockholm 1991.
34. Nachemson AL. Evaluation of results in lumbar spine surgery. Acta Orthop Scand Suppl 1993;251:130–3.
35. Nachemson A. Instrumental fusion of the degenerative lumbar spine. Philadelphia: Lippincott, 1996.
36. Owens DK. Spine update: patient preferences and the development of practice guidelines. Spine 1998;23:1073–9.
37. Raugstad TS, Harbo K, Oogberg A, et al. Anterior interbody fusion of the lumbar spine. Acta Orthop Scand 1982;53:561–5.
38. Schalling D. The development of the KSP inventory. University of Stockholm: Department of Psychology: Report no 64 from the project Individual Development and adjustment, 1986.
39. Scheer SJ, Watanabe TK, Radack KL. Randomized controlled trials in industrial low back pain. Part 3. Subacute/chronic pain interventions. Arch Phys Med Rehabil 1997;78:414–23.
40. Schultz KF, Chalmers I, Hayes RJ, et al. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995;273:408–12.
41. Spitzer WO, LeBlanc FE, DuPuis M. Scientific approach to the assessement and management of activity-related spinal disorders: a monograph for clinicians. Spine 1987;12(Suppl.):S5–59.
42. Steffee AD, Biscup RS, Sitkowski DJ. Segmental spine plates with pedicle screw fixation: a new internal fixation device for disorders of the lumbar and thoracolumbar spine. Clin Orthop 1986;203:45–53.
43. Tandon V, Campbell F, Ross ER. Posterior lumbar interbody fusion: association between disability and psychological disturbance in noncompensation patients. Spine 1999;24:1833–8.
44. Taylor ME. Return to work following back surgery: a review. Am J Ind Med 1989;16:79–88.
45. Taylor VM, Deyo RA, Cherkin DC, et al. Low back pain hospitalization: recent United States trends and regional variations. Spine 1994;19:1207–12; discussion, 13.
46. Turner JA, Ersek M, Herron L, et al. Patient outcomes after lumbar spinal fusions. JAMA 1992;268:907–11.
47. Turner JA, Deyo RA, Loeser JD, et al. The importance of placebo effects in pain treatment and research [see comments]. JAMA 1994;271:1609–14.
48. Van Susante J, Van de Schaaf D, Pavlov P. Psychological distress deteriorates the subjective outcome of lumbosacral fusion. a prospective study. Acta Orthop Belg 1998;64:371–7.
49. van Tulder MW, Koes BW, Bouter LM. Conservative treatment of acute and chronic nonspecific low back pain: a systematic review of randomized controlled trials of the most common interventions [see comments]. Spine 1997; 22:2128–56.
50. Von Korff M, Wagner EH, Dworkin SF, et al. Chronic pain and use of ambulatory health care. Psychosom Med 1991;53:61–79.
51. Von Korff M. Studying the natural history of back pain. Spine 1994; 19(Suppl):S2041–6.
52. Waddell G. Low back pain: a twentieth century health care enigma. Spine 1996;2:2820–5.
53. Whitney CW, Von Korff M. Regression to the mean in treated versus untreated chronic pain. Pain 1992;50:281–5.
54. Zucherman J, Hsu K, Picetti G, et al. Clinical efficacy of spinal instrumentation in lumbar degenerative disc disease. Spine 1992;17:834–7.
55. Zung WW, Richards CB, Short MJ. Self-rating depression scale in an outpatient clinic: further validation of the SDS. Arch Gen Psychiatry 1965;13: 508–15.

Address correspondence to

Peter Fritzell, MD
Department of Orthopedic Surgery
Falun Hospital
79182 Falun, Sweden
E-mail: pfritzell@hotmail.com
Anders Nordwall
Department of Orthopedic Surgery
Sahlgrenska University Hospital
Göteborg, Sweden
E-mail: anders.nordwall@orthop.gu.se

Point of View

Vert Mooney, MD

Of course, this is a very commendable study with a 2-year follow-up assessment for 98% of the patients treated. It is a real achievement of the Swedish Lumbar Spine Group to have such a high degree of organization for treating patients in such a standardized manner that valid comparisons could be made. These comparisons were measured objectively in the pre- and posttreatment testing by well-accepted standardized test instruments. By appropriate statistical analysis, significant differences can be demonstrated.

My dissatisfaction, however, is with the missed opportunity to learn something. The surgical patients were compared with usual nonsurgical treatment in Sweden. This is the treatment by which they had failed to become surgical candidates. How could anyone expect a satisfactory result from continuation of this treatment? It is a

real accomplishment of the Swedish health care system that the unlucky patients randomized to the nonsurgical group were docile enough to persist with the treatment that already had been demonstrated as ineffective. What a surprise that it did not work!

The patients selected for this study were restricted to those with significant degenerative changes at one or two of the lowest lumbar levels. Theoretically, these motion segments had become incompetent and at least microscopically unstable for these patients to become surgery candidates. Why would it not be reasonable to conduct a trial for physiologic stabilization by comparing active resistance exercises with surgically induced stabilization? Although exercises were no doubt incorporated into the nonoperative treatment program, the therapist apparently was left to choose how they were to be provided. Therefore, the specific dose of treatment is unknowable. The treatment certainly was not as consistent for the nonsurgical group as it was for the three surgical groups.

The endpoint of the surgical treatment also was well defined from a physiologic standpoint in that all the surgical patients were evaluated for fusion status. How were the nonsurgical patients physiologically evaluated? What percentage strength gain did they demonstrate? Of course, compliance was unavoidable with the surgical patients, but what was the compliance with treatment in the nonsurgical patients? The postoperative morbidity was no doubt several months in the surgical group. Were all the patients in the nonsurgical group tested for a similar length of time?

I do not believe these questions concerning the nonsurgical group can be answered unless some consistent specific piece of equipment is used to measure strength change and provide a standardized exercise routine. The cost of such equipment is minor compared with the costs of surgery. Or is it the belief of investigators that strength is irrelevant? If so, some other consistent passive piece of equipment should have been used.

Thus, this well-organized randomized clinical trial has shown that surgical care is more effective than the usual nonsurgical care for the chronic degenerative low back problem. It is such exciting new information, that the study received the prestigious Volvo Award. Does anyone know what usual nonsurgical care means? My real concern with the results of this study is that it will be taken as justification for even more fusions of the benign degenerative back without even an attempt at a valid physical training program. My current belief is that such training can be effective. If only this group had taken the time to define a consistent physical treatment program for comparison with surgery, we might really have learned something. I could change my belief with some evidence.

Response to Dr. Mooney's Point of View

Peter Fritzell, MD,*
Olle Hägg MD,†
Per Wessberg MD,†, and
Anders Nordwall, MD, PhD†

We agree with Dr. Mooney that the choice of control group strongly influences what can be learned from a study. The control group used in this study enables us to draw important conclusions about treating chronic low back pain patients with surgical fusion.

The deliberate aim of the study was to compare fusion surgery with commonly used nonsurgical methods as practiced in society, in a close to "real life situation" regarding average nonsurgically treated patients. The aim was not to investigate the effects of a specific physical therapy. The reason for this was that we, after thorough discussions during the preparation of the study in 1991, finally considered that there was no scientific evidence of the superiority of any specific physical treatment (still contradictory results).[1–3]

To guide the treatment in the nonsurgical group, a study protocol was developed on a consensus basis, including: behavioral training, coping strategies, back school, physical training, and symptomatic treatment. The nonsurgical treatment given to the patients in the control group thus represented the spectrum of prevailing conservative management in Sweden during the study. The treatment could be carried out either on an indoor or an outdoor basis, and the periods suggested were: weeks 1 through 4, 30 hours; weeks 5 through 12, 16 hours; months 4 through 12, 18 hours; and months 13 through 24, 6 hours. Both alternatives included 70 supervised hours of training during two years following randomization. Regular check-ups by the surgeon were performed on the same intervals as those of the surgical group.

To secure compliance in the control group, patients were informed before randomization that those who were allocated to the nonsurgical group could, if desir-

From the *Department of Orthopaedics, Falun Hospital, Falun, Sweden, and the †Department of Orthopaedics, Sahlgrenska University Hospital, Göteborg, Sweden.

able, be treated with methods that differed from what they had encountered previously. This strategy resulted in a group changing frequency of only 9% and a drop out rate of less than 3%. Also, 29% of those Dr. Mooney described as "unlucky patients" in the control group improved, and close to 60% would wish to go through the same treatment again.

In the literature today, there is little to favor a positive association between both muscular strength and range of motion, and the amount of success (measured as pain, disability, and return to work) after treatment for chronic low back pain, whether it is surgical[4,5] or nonsurgical.[6–10] Nor is muscular strength a valid diagnostic tool in lumbar disorders.[11] Because our primary interest was to estimate the effects of treatment on pain and disability, these other measurements were not included in the study. However, we hope that further studies will report on the relation between isokinetic muscle strength and disability/pain in the light of two seemingly contradictory results: reduced pain and disability after fusion in the Swedish study, and permanently (?) reduced muscle function after lumbar surgery reported by others.[12,13] Also, with the knowledge gained from other researchers during the course of our study, it would be interesting to compare fusion surgery with "intensive multidisciplinary rehabilitation," as there are now indications in the literature that this might be superior to "usual nonsurgical care."[14]

We do not fully understand Dr. Mooney's suggestion of radiologic fusion in the surgical group and muscle strength in the nonsurgical group as comparable physiologic endpoints. Our primary outcome measures applied in both treatment groups, as described in the "Material and Methods" section, were pain, disability, and work status at follow-up. We believe this is the proper way of comparing treatments aimed at reduction of symptoms from a lumbar disorder.

Dr Mooney says his real concern is that this study will lead to an increase of lumbar spine surgery for "benign degenerative back." This is also a concern to us, but we believe that the strict inclusion and exclusion criteria used is making it clear that this is a highly selected group of patients who were operated on because of severe and therapy-resistant pain (which was not considered as a benign condition) and not because of degenerative changes on X ray. In the "Summary," it is also pointed out that fusion should be offered only to "carefully selected and well-informed patients."

The present study is the result of an effort to replace belief with knowledge. It does not give answers to all our questions and is but one step in a continuous scientific process. The Swedish healthcare system is working in a definite direction toward scientifically validated treatment modalities. Today, we can claim that we know that fusion surgery can be a valid option to patients with longstanding lumbar pain. Those who believe that physiologic stabilization and physical training programs using different equipments are effective in the treatment of severe CLBP should try to prove this through scientifically valid studies.

References

1. Furlan AD, Clarke J, Esmail R, et al. A critical review of reviews on the treatment of chronic low back pain. Spine 2001;26:E155–62.
2. Bentsen H, Lindgarde F, Manthorpe R. The effect of dynamic strength back exercise and/or a home training program in 57-year-old women with chronic low back pain. Results of a prospective randomized study with a 3-year follow-up period. Spine 1997;22:1494–500.
3. Kuukkanen T, Malkia E. Muscular performance after a 3 month progressive physical exercise program and 9 month follow-up in subjects with low back pain. A controlled study. Scand J Med Sci Sports 1996;6:112–21.
4. Mayer TG, Mooney V, Gatchel RJ, et al. Quantifying postoperative deficits of physical function following spinal surgery. Clin Orthop 1989;Jul(244): 147–57.
5. Mayer TG, Vanharanta H, Gatchel RJ, et al. Comparison of CT scan muscle measurements and isokinetic trunk strength in postoperative patients. Spine 1989;14:33–6.
6. Mannion AF, Junge A, Taimela S, et al. Active therapy for chronic low back pain: part 3. Factors influencing self-rated disability and its change following therapy. Spine 2001;;26:920–9.
7. Cox ME, Asselin S, Gracovetsky SA, et al. Relationship between functional evaluation measures and self-assessment in nonacute low back pain. Spine 2000;25:1817–26.
8. Sullivan MS, Shoaf LD, Riddle DL. The relationship of lumbar flexion to disability in patients with low back pain. Phys Ther 2000;80:240–50.
9. Nattrass CL, Nitschke JE, Disler PB, et al. Lumbar spine range of motion as a measure of physical and functional impairment: an investigation of validity. Clin Rehabil 1999;13:211–8.
10. Mannion AF, Muntener M, Taimela S, et al. A randomized clinical trial of three active therapies for chronic low back pain. Spine 1999;24:2435–48.
11. Rantanen P. Physical measurements and questionnaires as diagnostic tools in chronic low back pain. J Rehabil Med 2001;33:31–5.
12. Gejo R, Matsui H, Kawaguchi Y, et al. Serial changes in trunk muscle performance after posterior lumbar surgery. Spine 1999;24:1023–8.
13. Tiusanen H, Hurri H, Seitsalo S, et al. Functional and clinical results after anterior interbody lumbar fusion. Eur Spine J 1996;5:288–92.
14. Guzman J, Esmail R, Karjalainen K, et al. Multidisciplinary rehabilitation for chronic low back pain: systematic review. BMJ 2001;322:1511–6.