ELSEVIER

The Spine Journal 5 (2005) 36–44

THE SPINE JOURNAL

2004 Outstanding Paper Award: Surgical Science

The long-term effect of posterolateral fusion in adult isthmic spondylolisthesis: a randomized controlled study

Per Ekman, MD[a,*], Hans Möller, MD, PhD[b], Rune Hedlund, MD, PhD[b]

[a]*Stockholm Söder Hospital, Department of Orthopedic Surgery, Stockholm Söder Hospital, 118 83, Stockholm, Sweden*
[b]*Karolinska University Hospital, Huddinge, Department of Orthopedic Surgery, 141 86 Stockholm, Sweden*

Received 12 January 2004; accepted 14 May 2004

Abstract

BACKGROUND: Today there is some evidence-based medicine support for a positive short-term treatment effect of fusion in chronic low back pain in spondylolisthesis and in nonspecific degenerative lumbar spine disorders. The long-term effect is, however, unknown.

PURPOSE: To determine the long-term outcome of lumbar fusion in adult isthmic spondylolisthesis.

STUDY DESIGN: Prospective, randomized controlled study comparing a 1-year exercise program with instrumented and non-instrumented posterolateral fusion with average long-term follow-up of 9 years (range, 5–13).

PATIENT SAMPLE: 111 patients aged 18 to 55 years with adult lumbar isthmic spondylolisthesis at L5 or L4 level of all degrees, and at least 1-year's duration of severe lumbar pain with or without sciatica.

OUTCOME MEASURES: Pain and functional disability was quantified by pain (VAS), the Disability Rating Index (DRI), the Oswestry Disability Index (ODI) work status, and global assessment of outcome by the patient into much better, better, unchanged or worse. Quality of life was assessed by the SF-36.

METHODS: The patients were randomly allocated to treatment with 1) a 1-year exercise program (n=34), 2) posterolateral fusion without pedicle screw instrumentation (n=37), or 3) posterolateral fusion with pedicle screw instrumentation (n=40). Long-term follow-up was obtained in 101 (91%) patients. Nine patients in the exercise group were eventually operated on.

RESULTS: Longitudinal analysis: At long-term follow-up pain and functional disability were significantly better than before treatment in both surgical groups. No significant differences were observed between instrumented and non-instrumented patients in any variable studied. In the exercise group the pain was significantly reduced but not the functional disability. Compared with the 2-year follow-up a significant increase in functional disability was observed, as measured by the DRI, but not the ODI, in the surgical group at long term. In the exercise group no significant changes were observed between the 2-year and the long-term follow-up. Cross-sectional analysis: Between the surgical and conservative group no significant differences were observed in any outcome measurement at long-term follow-up except for global assessment, which was significantly better for surgical patients. Of surgical patients 76% classified the overall outcome as much better or better compared with 50% of conservatively treated patients (p=0.015). Quality of life as estimated by the SF-36 at long term was not different between treatment groups in any of the eight domains studied but was considerably lower than for the normal population.

CONCLUSIONS: Posterolateral fusion in adult lumbar isthmic spondylolisthesis results ina modestly improved long-term outcome compared with a 1-year exercise program. Although the results

FDA device/drug status: approved for this indication (CD pedicle screw instrumentation).

Supported by The Swedish Society of Spine Surgery, Falun, Sweden. Nothing of value received from a commercial entity related to this research.

* Corresponding author. Department of Orthopedic Surgery, Stockholm Söder Hospital, 118 83, Stockholm, Sweden. Tel.: (+46) 8 6161000; fax: (+46) 8 6162804.

E-mail address: per.ekman@sos.sll.se (P. Ekman)

1529-9430/05/$ – see front matter
doi:10.1016/j.spinee.2004.05.249

show that some of the previously reported short-term improvement is lost at long term, patients with fusion still classify their global outcome as clearly better than conservatively treated patients. Furthermore, because the long-term outcome of the patients conservatively treated most likely reflects the natural course, one can also conclude that no considerable spontaneous improvement should be expected over time in adult patients with symptomatic isthmic spondylolisthesis. Substantial pain, functional disability and a reduced quality of life will in most patients most likely remain unaltered over many years.

Keywords: Spondylolisthesis; Fusion; Exercise; Outcome; Long-term

Introduction

Two randomized studies previously showed an improved short-term outcome of lumbar spine fusion compared with conservative treatment. In the first study that reported on the effect of fusion of the degenerative lumbar spine and used a randomized controlled design, we showed that posterolateral fusion in adult isthmic spondylolisthesis resulted in a significantly improved outcome at 2-year follow-up compared with an exercise program [1]. In surgical patients the pain index (visual analogue scale [VAS], 0–100) on average improved from 63 to 37, compared with 65 to 56 in patients conservatively treated. A functional index (Disability Rating Index, DRI) showed the same results. Similarly, the global assessment of outcome was clearly better in surgical patients, 74% classified their outcome as excellent or good, compared with 26% for patients after an intensive exercise program.

Adult isthmic spondylolisthesis can be viewed as a radiologically verifiable model of accelerated spinal degeneration; the symptoms and clinical findings are undistinguishable from nonspecific degenerative lumbar spine disorders [2]. The positive effect of fusion on low back pain was confirmed by the Swedish lumbar spine study (SLSS), which showed very similar short-term results in nonspecific degenerative low back pain with an improved outcome in patients with fusion compared with physiotherapy [3].

We found no change between the outcome at 1 and 2 years, suggesting a stable early situation. In the SLSS, however, a significant increase was observed in low back pain and also a tendency toward increased leg pain between the 1- and 2-year follow-up. This raises the question of the long-term effect of fusion. The possibility of secondary accelerated degeneration in adjacent motion segments is much discussed. Whether this phenomenon occurs and whether it is of clinical importance is, however, controversial [4–7].

Except for a Danish prospective study without nonoperated control subjects, with an initial 2-year [8] and a subsequent 5-year follow-up [9], only retrospective data of limited scientific validity are published on long-term effects [10–12]. The Danish study [9] reported a further improvement of a subgroup of posterolaterally fused patients between the 2nd and the 5th year follow-up. This was, however, only demonstrated for the subgroup of non-instrumented fusions, and not for the total material of fused patients.

Thus, although there is some evidence for a positive treatment effect of fusion in chronic low back pain, the effect is not dramatic, and whether the improvement remains over time has not been demonstrated. The purpose of the present study was to determine the long-term effects on pain, functional disability, global outcome and quality of life (QOL) after posterolateral lumbar spine fusion.

Patients and methods

The original study with a 2-year follow-up of the present patient material was previously published by Moller and Hedlund in 2000 [1]. The present study reports the long-term outcome of the same patient material. From 1990 through 1995 a total of 116 patients fulfilled the inclusion criteria of lumbar isthmic spondylolisthesis of all grades with at least 1 year of low back pain with or without sciatica and a severely restricted functional ability in individuals aged 18 to 55 years.

Exclusion criteria were previous spine surgery, alcohol or drug abuse or only mild symptoms. The main complaint of all patients was low back pain with or without sciatica. The patients were randomly assigned into three groups: 1) exercise, 2) posterolateral fusion with transpedicular fixation, and 3) posterolateral fusion without transpedicular fixation. The patients were all referred to the spine units of Huddinge University Hospital (n=81) and Linköping University Hospital (n=30).

Randomization without stratification was used. For each patient three different notes were each marked with one of the three different treatment modalities, and one note was blindly chosen by the attending nurse at the outpatient ward. The procedure gives each patient the same chance to each treatment group. The type of treatment was unknown to the patient and the doctor until after the patient had given consent. The randomization resulted in similar distribution of age, symptoms, level and grade of slip and lifestyle factors between the groups (Table 1).

Two patients refused randomization. One patient randomly assigned to fusion had surgery arranged at another hospital because of long waiting time for surgery. Two further patients, randomly assigned to fusion improved spontaneously, and at admission surgery was no longer considered necessary. These five patients were excluded, leaving 111 patients who entered the study, 34 patients in the exercise group, 37 in the fusion with instrumentation group and 40 patients in the fusion without instrumentation group.

Table 1
Demographics, symptoms, level and grade of slip and life style factors, in percent (except age and sick-leave period), before treatment according to type of randomized treatment (observed numbers within brackets).

	All n=111	Non-instrumented n=40	Instrumented n=37	Exercise n=34
Mean age (years)	39	39	39	37
Mean age at onset of symptoms (years)	26	25	29	25
Women	49 (54)	45 (18)	57 (21)	44 (15)
Men	51 (57)	55 (22)	43 (16)	56 (19)
Low back pain only	31 (33)	25 (10)	30 (11)	39 (12)
Low back pain+sciatica	62 (67)	68 (27)	62 (23)	55 (17)
Sciatica only	7 (8)	8 (3)	8 (3)	6 (2)
Level L5	85 (94)	83 (33)	84 (31)	88 (30)
Level L4	13 (14)	15 (6)	14 (5)	9 (3)
Levels L4 and L5	3 (3)	3 (1)	3 (1)	3 (1)
Grade 1 slip	60 (67)	68 (27)	54 (20)	59 (20)
Grade 2 slip	38 (42)	30 (12)	43 (16)	41 (14)
Grade 3 slip	2 (2)	3 (1)	3 (1)	0
Sick-leave or disability pension	71 (79)	68 (27)	84 (31)	62 (21)
Mean sick-leave before treatment (months)	16	15	14	18
Blue collar	80 (87)	90 (36)	75 (27)	73 (24)
Immigrants	32 (35)	30 (12)	27 (10)	38 (13)
Married	74 (82)	75 (30)	76 (28)	71 (24)
Smokers	54 (60)	63 (25)	57 (21)	41 (14)
Medication for other than back pain	21 (23)	30 (12)	16 (6)	15 (5)

Of the 111 patients randomly assigned 2-year outcome data were earlier obtained from 106 patients and reported by Möller and Hedlund [1]. The 1- and 2-year outcome data were collected at outpatient visits by an unbiased observer: an independent surgeon, the main surgeon's assistant or by the physiotherapist leading the exercise program. To avoid bias, at follow-up some of the surgical patients were seen by a physiotherapist and some of the patients receiving the exercise program were seen by a surgeon.

All patients were followed at regular intervals, 2, 6, 12 and 24 months after start of exercise or surgery. All patients completed questionnaires about functional disability and pain before treatment and at 1- and 2-years follow-up at outpatient visits to the unbiased observer. After the 2-year follow-up the patients were not followed at regular intervals, and the study protocol did not include any type of treatment after the 2-year follow-up. Therefore, the patients in the surgical as well as exercise group could have received non-documented conservative treatment outside the context of the study. Possible further surgical treatment was documented at long-term follow-up.

The 106 patients with a completed 2-year follow-up were invited by mail to take part in the long-term study. Questionnaires were mailed to the patients, and answers by mail were obtained from 101 patients, resulting in a long-term total follow-up rate of the original 111 patients of 91% (101 of 111).

In the exercise group 9 patients were operated on with fusion because of worsening or continuation of severe symptoms. They were kept in their original conservative group according to the "intention to treat" principle and analyzed as part of the exercise group. Furthermore, statistical analysis of the long-term results was performed with, as well as without, these 9 patients, and also selectively for the 9 patients.

The mean long-term follow-up time was 9 years (range, 5–13). The three treatment-groups had less than a year's difference in follow-up time. The mean age of the patients at inclusion in the study was 39 years (range, 18–55) and at long-term follow-up 47 years (range, 28–68). There were 54 women and 57 men. The level of the slip was L5 in 94 patients, L4 in 14 patients and both L4 and L5 in three patients. The majority of patients were grade 1 (67) or 2 (42) slips, and only two patients had a grade 3 slip (Table 1). All 77 surgical patients had a posterolateral fusion in situ with autologous bone transplantation harvested from the right iliac crest. The senior author performed 85% of the operations. No instrumentation was used in 40 patients, and pedicle screw fixation (Cotrel-Debousset, Sofamor, Paris, France) was used in 37 patients. The non-instrumented patients wore a daytime lumbar brace for 6 months after surgery. The surgical patients did not receive a postoperative exercise or physiotherapy program.

The 34 patients randomly assigned to the exercise program were referred to a physiotherapist with a special interest in spondylolisthesis. The exercise program was based on strength and postural training with emphasis on back and abdominal muscle exercises. Twelve different exercises were performed. To allow exercises at home, eight of the exercises did not demand any specific training equipment. Four exercises included a pulley machine and a leg press machine.

These patients exercised three times a week the first 6 months, and twice a week between 6 and 12 months. The duration of the exercise program was approximately 45 minutes and was supervised by a physiotherapist. After 1 year

the patients were instructed to continue with a home program, which consisted of the eight exercises that did not need special equipment. Two thirds of the patients complied with the full program during the first year. After the first year it is not known to what extent the patients continued with the recommended exercises. Most likely most patients largely discontinued regular back-strengthening exercises.

Outcome measurements

Pain and functional disability was quantified by pain-index (VAS), the DRI and work status before treatment, at 1 year, 2 years and at long-term follow-up. A global assessment by the patient was performed at 2-years and at long-term follow-up, with classification of results into much better, better, unchanged or worse The Oswestry Disability Index (ODI) was obtained at 2-year and long-term follow-up. The SF-36 was obtained only at long-term follow-up. The same questionnaires as used at 1 and 2-years, with the addition of the ODI and SF-36 questionnaires, were mailed to the patients.

Pain was quantified by one VAS for "pain right now" and one VAS for "worst pain last week," with anchor points; 0=no pain and 100=intolerable pain. The mean of the two pain scores provided the pain index.

DRI is a validated disease-specific functional outcome instrument composed of 12 VAS items, ie, dressing, outdoor walking, climbing stairs, sitting for longer time, standing bent over a sink, carrying a bag, making a bed, running, light work, heavy work, lifting heavy objects and participation in exercise or sports [13]. The mean of the 12 functional VAS items provides the DRI. The form is self-administered according to oral instructions. In the VAS the patient marks on a 100-mm scale the ability to perform the respective activity, with anchor points; 0=without difficulty and 100=impossible.

ODI, a validated disease-specific instrument for assessment of spinal disorders, is a 10-item ordinal scale instrument with six response alternatives for each item [14]. The total score ranges from 0 to 100, whereby 100 is worst disability. The items are pain intensity, personal care, ability to lift, walk, sit, stand, sleep, sex life, social and travelling. Normal function is 0 and worst disability for each item is 5. The sum of the 10 items multiplied by 2 constitutes the ODI (0–100).

Global outcome was assessed by the patients who classified the overall results into "much better," "better," "unchanged" or "worse". Work status was documented. The QOL was estimated with the SF-36 [15], which includes 36 questions arranged into eight domains: physical function, role physical, bodily pain, general health, social function, role emotional, mental health and vitality. Results are presented as a profile.

All patients completed a pain drawing [16]. Sciatica was defined as pain symbols below the knee in the pain drawing. According to the pain drawing 33 patients had only low back pain, 67 patients had low back pain and sciatica and 8 patients had only sciatica. In 3 patients no pain drawing was performed. In the surgical group patients with sciatica were investigated with MRI or CT-myelography for documentation of nerve root compression. No patients showed any radiological sign of disc prolapse or central spinal stenosis.

Statistical method

The appropriate sample size sufficient to detect a clinical relevant difference in outcome, if it exists, was calculated according to Altman [17]. The risk of a type I error was set to 5% (significance level, 0.05). Type II error was set to 10% (90% power). The standard deviation of the DRI was 18, and the clinical difference of relevance was set to 15, resulting in a necessary sample size of 30 in each group.

To test for differences between pain index, DRI, ODI and SF-36 the nonparametric Mann Whitney *U* test was used for unpaired data, ie, comparisons between the two groups, and the Wilcoxon signed rank test was used for paired data, ie, longitudinal comparison within each group. In addition to the nonparametric tests, 95% confidence intervals, based on standard error of the mean, are also given. The chi-square test was used to determine differences in the work ability and global outcome between groups. The global outcome was assessed after collapse of the much better and better categories into one cell, and the unchanged and worse into another. In addition the chi-square test for trend was used for comparison between treatments, keeping the four cells of global outcome intact. The Yates continuity correction was used in the chi-square analysis. $P<0.05$ was considered as statistically significant.

The study was approved by the Medical Ethical Committee of Huddinge University Hospital.

Results

Cross-sectional analysis

At long-term follow-up the pain index, DRI, ODI, global assessment, SF 36 and work ability were similar with no significant differences between surgical patients with posterolateral fusion with and without instrumentation (Tables 2, 3). Therefore, surgical patients were analyzed together, to increase the power of the study.

In the surgical group 11 patients (14%) had more than one lumbar spine operation for the following reasons: two nerve root injuries in instrumented fusions, one pseudarthrosis, one discectomy and removal of implants because of possible local irritation in 7 patients. There were no early or late deep infections.

The pain index, DRI, ODI and work ability were not significantly different between surgically (combined surgical group) and conservatively treated patients at long-term

Table 2
Mean pain index, DRI, ODI and proportion of patients at work at long-term follow-up in non-instrumented and instrumented patients

Long-term	Non-instrumented	Instrumented	p
Pain index	45	36	0.27
DRI	36	30	0.31
ODI	30	27	0.79
At-work	50%	52%	0.90

follow up (Table 4; Fig. 1, 2). Similarly, no significant differences were observed between the eight domains in the SF-36 between the surgical and conservative group (Fig. 3). The global assessment, however, was significantly better for the surgical group. This was true for the different categories analyzed as separate cells ($0.02<p<0.05$) as well as after collapse of much better and better into one cell, and unchanged and worse into another ($p=0.015$) (Table 5). The mean pain index, DRI and ODI among the much better patients in the surgical group were clearly lower than the scores for the much better patients in the exercise group. The mean scores for the pain and DRI and ODI were 18, 13, and 8 compared with 28, 22, and 20 in the much better surgical and conservative group, respectively. The differences were not significant for pain and DRI, but almost significant for ODI ($p=0.078$).

There were no significant differences in long-term outcome between the nine patients in the conservative group who had been eventually operated on compared with the original conservative group. Furthermore, exclusion of these 9 patients from the conservative group did not change the overall results. There were still no significant differences compared with the surgical group at long-term follow-up, except for global outcome. There was no correlation between outcome and the variation in follow-up time (5–13 years).

Longitudinal analysis

From start of study to long-term follow-up

In the surgical group the mean DRI improved from 48 to 33 ($p< 0.001$) and the pain index from 63 to 40 ($p<0,0001$) from before treatment to long-term follow-up. In the surgical group 25% worked before the study compared with 51% at long term ($p<0.0001$) (Table 4; Fig. 1, 2). There were no differences between instrumented and non-instrumented patients in the longitudinal analysis; all scores improved similarly for both groups.

Table 3
Global outcome in the instrumented and the non-instrumented group at long-term follow-up, as classified by the patients

	Instrumented (%)	Non-instrumented (%)
Much better	33	44
Better	49	22
Unchanged	9	22
Worse	9	12

In the conservative group the mean DRI improved non-significantly from 44 to 38 ($p=0.13$), and the pain index from 65 to 49 ($p=0.013$). The proportion of patients that worked in the conservative group increased from 38% to 46% at long term (NS).

Between the 2-year and long-term follow-up

Between the 2-year and long-term follow-up the mean DRI of the surgical group worsened significantly from 29 to 33 ($p=0.049$). Similar marginal deteriorations were observed for the pain index and ODI, however non-significantly (Table 4). The global assessment was unchanged between the 2-year and long-term follow-up. There were no differences between instrumented and non-instrumented patients.

In the conservative group all scores except the ODI improved non-significantly between the short-term and long-term follow-up. The ODI worsened from 28 to 31 (NS) (Table 4).

Discussion

This is the first long-term randomized study on lumbar fusion that compared surgery with conservative treatment. At an average follow-up of 9 years surgical patients with fusion scored better than the conservative group for all variables documented. However, the global outcome was the only variable that was significantly better in the surgical group than in the conservative group. In the surgical group 76% of patients classified their result as much better or better compared with 50% in the conservative group ($p=0.015$). For all other variables studied the differences were small and not statistically significant. Thus, in contrast to the short-term study of the same patient material, in which fusion was significantly better than conservative treatment for all variables studied, a significant long-term difference remained only for the global outcome variable [1].

The reduced differences between the groups were the combined effect of two trends between the short-term and the long-term follow-up: the conservative group showed a slight nonsignificant improvement, whereas the surgical group showed a slight deterioration, significant only for the DRI. Because the exercise program only lasted 1 year, and few patients most likely changed their normal exercise habits after discontinuation of the program, the results for the exercise group reflect the natural course of degenerative disorders of the lumbar spine, Although the results of the conservative group showed a slight spontaneous improvement over time, it cannot be excluded that the results simply reflect the phenomenon of “regression toward the mean.” All patients were recruited into the study at a time when their symptoms were severe, and because chronic low back pain is a fluctuating disorder, over time the group will tend to score on average better at later follow-ups, not scheduled after severity of symptoms.

Although 1 year of exercise followed by no regular treatment for on average 8 years resulted in a significant

Table 4
Mean value of pain index, DRI, and percentage of patients at work before treatment and at 1- and 2-year follow-up and at long-term follow-up, for the surgical group and the exercise group (95% confidence intervals within brackets).

		Pretreatment n=106	1-year n=98	2-year n=106	Long-term n=101	p†
Surgery	Pain index	63 (58.5–67.7)	35 (28.7–42.2)	37 (29.6–43.8)	40 (34.0–47.1)	<0,0001
	DRI	48 (43.9–52.3)	29 (23.0–34.6)	29 (23.5–34.9)	33 (27.8–38.8)	<0,0001
	ODI			26 (18.1–31.6)	28 (23.0–33.0)	0,223
	At-work	25%	46%	54%	51%	<0.001*
Exercise	Pain index	65 (57.3–71.9)	54 (44.7–63.7)	56 (48.7–63.8)	49 (38.4–58.8)	0,013
	DRI	44 (38.2–50.3)	45 (36.4–53.7)	44 (36.5–50.9)	38 (29.1–47.7)	0,131
	ODI			28 (20.5–35.0)	31 (24.0–37.4)	0,887*
	At-work	38%	48%	55%	46%	n.s.

† p-value refers to comparison between the 2 years and the long-term follow-up
* The ODI was only measured at 2-year follow-up and at long-term follow-up. The worst possible status is 100 and the best possible status is 0.

improvement of pain (from VAS 64 to 49), considerable pain and disability remained at long term as reflected by only half of the patients working and a subnormal level of QOL. From the results of the study it seems clear that the natural course of symptomatic adult isthmic spondylolisthesis generally does not include any major changes over time. Thus, one should not expect that disabling low back pain spontaneously improves over time in patients with symptoms severe enough to qualify for fusion.

The cause of the slight deterioration of outcome of the surgical group with time can only be speculated on, particularly because radiologic follow-up is not yet completed. Adjacent segment degeneration after lumbar fusion has been reported by several investigators [4,7,10,18].

Although the present results lend no support to the notion of a negative effect of fusion on the outcome in the long term because of adjacent segment degeneration, such an effect cannot be excluded. It could be that adjacent segment degeneration gradually reduces the positive effects of fusion and that even longer follow-up will show a worse outcome in fused patients than in conservatively treated patients.

Seitsalo et al. [5] compared surgically and conservatively treated patients and reported no increase in adjacent segment degeneration on average 15 years after lumbar fusion in isthmic spondylolisthesis in patients younger than 20 years. However, in a comparison between fused and nonfused patients treated surgically for degenerative disc disease Kumar et al. [10] found a twice as high incidence of degenerative changes at levels above the previous operation. However, no difference was observed in outcome between the groups. Similarly, Lehmann et al. [4] reported a high incidence of adjacent segment instability (45%) on average 33 years after

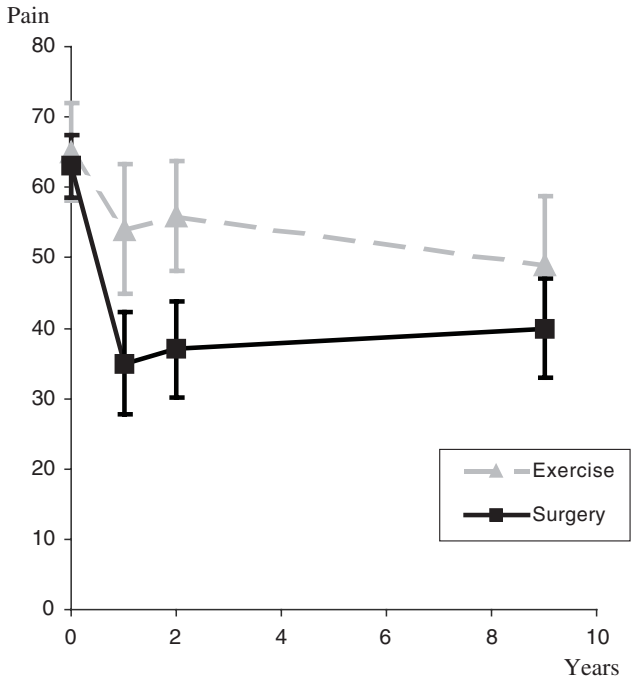

Fig. 1. The mean value of pain index before treatment and at 1- and 2-year follow-up and at long-term follow-up for the surgical group and the exercise group. 95% confidence intervals are indicated.

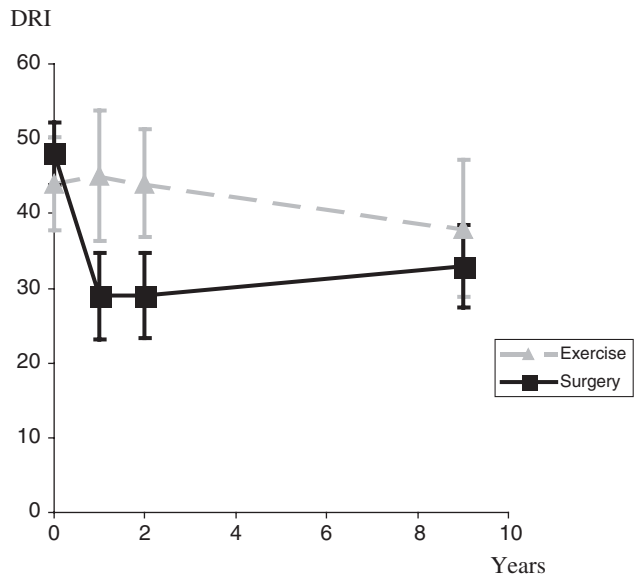

Fig. 2. The mean value of DRI before treatment and at 1- and 2-year follow-up and at long-term follow-up for the surgical group and the exercise group. 95% confidence intervals are indicated.

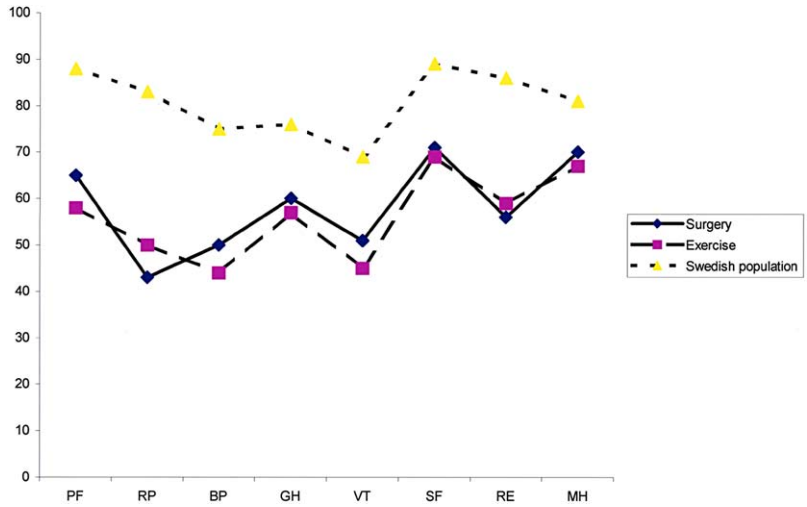

Fig. 3. The mean value for the eight subscales of the SF-36 at long-term follow-up for the surgical group and the exercise group, as well as for the average Swedish population. No significant differences were observed between the two treatment groups in any of the subscales. PF=physical function; RP=role physical; BP=bodily pain; GH=general health; VT=vitality; SF=social functioning; RE=role emotional; MH=mental health.

lumbar fusion, but no correlation between the radiographic condition and symptoms. The same finding was also reported by Ishihara et al. [6] who observed adjacent segment degenerative changes in more than 50% of patients over 10 years after anterior lumbar interbody fusion for isthmic spondylolisthesis, but no correlation to clinical outcome. Because no randomized controlled study has addressed the question of the occurrence and possible clinical relevance of adjacent segment degeneration after fusion, the issue remains controversial. Except for accelerated degenerative changes, it cannot be excluded that the worsening of symptoms in the surgical group can be an effect of a diminishing placebo effect.

Because the global outcome shows rather clear differences between the groups, in contrast to the scores for pain, disability and QOL scales are of interest. It could be that important factors related to the disorder are not included in the disease-specific (ODI, DRI, VAS-pain) as well as in the QOL questionnaires, or that these variables are not sensitive to change. We found the same tendency in our previous short-term report of the same patient material, the difference between the groups was most obvious in the global outcome assessment. Hägg et al. [19] showed that global outcome well reflects several validated outcome measurements and suggested it as an important outcome measurement in the degenerative lumbar spine. We agree that global outcome could be the most important outcome measurement, presupposing that the study design is randomized. It does not seem possible to assess treatment effects from nonrandomized studies with global outcome as the only end point.

Table 5
Global outcome in the surgical group and the exercise group at long-term follow-up, as classified by the patients

	Surgery (%)	Exercise (%)
Much better	39	27
Better	37	23
Unchanged	13	27
Worse	10	23

One could expect that the more general questionnaires addressing several areas of everyday life, ie, the SF-36 and the ODI, would reflect global outcome better than DRI (a pure functional score) and VAS-pain. This was, however, not the case in the present study.

The ODI was very similar in both groups with only a marginally better score for surgical patients, and the SF-36 was for some (2 of 8) domains even better for the conservatively treated patients. The results suggest that ODI is less responsive to change than VAS-pain and the DRI, and also that differences between different treatment outcomes in the degenerative lumbar spine may not necessarily be reflected in QOL-instruments.

The SF-36 scores of conservatively as well as surgically treated patients were clearly below the average score in all eight areas for the normal Swedish population (Fig. 3). Thus, neither surgical treatment nor the long-time natural course seems to normalize QOL in the adult degenerative spine. The lack of expression of the global outcome in the SF-36 can only be speculated on, but an obvious explanation is of course that QOL could be a too crude measurement to reflect rather subtle differences between different treatment effects in patients with low back pain.

Work status is influenced by the general economical situation, which in Sweden in the early years of 2000 was less positive with higher unemployment rates than in the mid-1990s. This makes the development over time difficult to interpret. In the short-term analysis (2 years) no difference was observed in work status, 45% and 46% were working in the surgical and conservative group, respectively. Interestingly, this proportion was largely unchanged at long term,

51% and 46% were working in the two groups, respectively. Although the increase compared with pretreatment in work ability was significant only for the surgical group, the study does not demonstrate any serious effect on work ability by lumbar fusion. In the Swedish lumbar spine study, however, more patients went back to work among surgical patients [3].

However, because several outcome measures showed only nonsignificant differences compared with the natural course, we strongly emphasize the need of careful patient selection. Note that the patients in the much better category among fused patients scored much better for pain, DRI and ODI than the much better patients in the conservative group.

Although this was not a statistically significant finding, possibly a reflection of the limited number of much better patients in the conservative group (n=7), it is in accordance with our clinical experience. Totally asymptomatic patients are much more often found among fused patients. As stated by many investigators, the challenge for the surgeon is how to select patients for fusion who will show such a much better long-term result. Unfortunately, we have few means for such predictions today.

The drawback of the study is the limited number of conservatively treated patients, a problem that is aggravated by the fact that 9 of the 29 patients in the conservative group were eventually operated on with fusion. This, of course, could influence the results of the conservative group. However, irrespective of whether these patients were included in their original treatment group, according to the intention to treat analysis principle, or excluded, the results were not influenced. Furthermore, the outcome of these 9 patients was not significantly worse than for the clean conservative group. Thus, the conservative group analysed according to the intention to treat principle was not "falsely improved" by keeping these 9 fused patients in their original conservative group, and one must conclude that this source of error does not invalidate the results.

Another source of error is the limited patient material. When the study was designed in the late 1980s, a treatment difference of 15 DRI points was chosen as the minimal clinically relevant difference. Thus, the study was designed to identify clinically unequivocal differences in treatment effects rather than to demonstrate possible marginal differences in outcome. A study of at least 90% power on the 5% significance level resulted in a calculated sample size of 30 subjects in each of the three groups. Because all analyses show no differences between the two surgical groups, they are analyzed together, which results in an improved power. The comparisons between conservative and surgical treatment with a final follow-up of 71 surgically and 29 conservatively treated patients results in a power of 95%. If the 9 patients in the conservative group are excluded, we are left with only 20 patients in the conservative group, which reduces the power to 89%.

It can, of course, correctly be argued that smaller differences than 15 points on these scores could be of some importance to the patient, and that the power of the present study may not be sufficient to detect such smaller differences. The power of the present study to detect 10 points of difference on the 5% level is only 70%; ie, the study is associated to the risk of not being able to demonstrate minor differences between treatments. When the study was designed, it was believed that the potential treatment effects as measured on pain and functional scales would be more obvious than what has turned out to be the case. Several prospective studies have now shown that the effect of fusion on commonly used outcome measurements are rather marginal [1,3,8], and, consequently, larger patient materials are needed if minor differences in treatment effects are of interest.

Most randomized controlled trial (RCT) studies failed to document any difference in clinical outcome between instrumented and non-instrumented fusions of the lumbar spine [8,20–23]. All studies are short term, with the exception of the present study and Bjarke Christensen et al. [9], who reported a better outcome at 5-year follow-up in a subgroup of patients with non-instrumented spondylolisthesis compared with instrumented spondylolisthesis. In accordance with previous short-term studies, however, we found a similar long-term outcome for instrumented and non-instrumented patients. The reason for the discrepancy between the two Scandinavian studies is unknown, but from a methodological point of view conclusions from subgroup analysis must be made with caution, particularly if the number of patients is very limited, as in the Danish study. Furthermore, no difference between instrumented and non-instrumented patients was found in the main Danish patient material, composed of mixed diagnoses. Thus, the previous evidence of no short-term benefit of instrumentation on the outcome of one or two level lumbar fusion can now be expanded also to the long-term situation. However, we found no evidence of a negative long-term effect of instrumentation. A negative effect on adjacent level facet joints by pedicle screws has been discussed, as well as a negative effect on adjacent segments because of a stiffening effect on the fused area by the instrumentation [18].

In view of the limited improvement of the exercise in particular, but also of fusion, there is a definite need for alternative treatment models. Although evidence for its effectiveness are lacking, disc prosthesis procedures have been increasingly popular the past few years. Only one RCT is available on disc replacement surgery, and it showed no superiority compared with fusion at 6 months [24]. Several short-term and long-term RCTs are, however, mandatory before the place of disc replacement in the treatment of degenerative lumbar disorders can be properly assessed.

In conclusion, the significantly better global outcome of fused patients compared with conservatively treated patients supports a limited but positive long-term effect on fusion in adult isthmic spondylolisthesis. The study also shows, however, a slight deterioration of the effect of fusion with time. In addition, the results clearly show that instrumentation neither positively nor negatively affects the long-term

results. Finally, the study also shows that the natural course of symptomatic adult spondylolisthesis is not one of spontaneous improvement, but rather of continuing disability affecting many areas of life for many years.

References

[1] Möller H, Hedlund R. Surgery versus conservative treatment in adult isthmic spondylolisthesis. A prospective randomized trial. Part 1. Spine 2000;25:1711–5.

[2] Möller H, Sundin A, Hedlund R. Symptoms, signs and functional disability in adult spondylolisthesis. Spine 2000;25:683–9.

[3] Fritzell P, Hägg O, Wessberg P, Nordwall A. Lumbar fusion versus nonsurgical treatment for chronic low back pain. Spine 2001;23: 2521–34.

[4] Lehmann TR, Spratt KF, Tozzi J, et al. Long-term follow-up of lumbar fusion patients. Spine 1987;12:97–104.

[5] Seitsalo S, Schlenzka D, Poussa M, Osterman K. Disc degeneration in young patients with isthmic spondylolisthesis treated operatively or conservatively: a long-term follow-up. Eur Spine J 1997;6:393–7.

[6] Ishihara H, Osada R, Kanamori M, et al. Minimum 10-year follow-up study of anterior lumbar interbody fusion for isthmic spondylolisthesis. J Spinal Disord 2001;14:91–9.

[7] Gillet P. The fate of adjacent motion segments after lumbar fusion. J Spinal Disord 2003;16:338–45.

[8] Thomsen K, Christensen FB, Eiskjaer SP, Hansen ES, Fruensgaard S, Bunger CE. The effect of pedicle screw instrumentation on functional outcome and fusion rates in posterolateral lumbar fusion. Spine 1997;24:2813–22.

[9] Bjarke Christensen F, Stender Hansen E, Laursen M, Thomsen K, Bunger CE. Long-term functional outcome of pedicle screw instrumentation as a support for posterolateral spinal fusion: randomized clinical study with a 5-year follow-up. Spine 2002;27:1269–77.

[10] Kumar MN, Jacquot F, Hall H. Long-term follow-up of functional outcomes and radiographic changes at adjacent levels following lumbar spine fusion for degenerative disc disease. Eur Spine J 2001; 10:309–13.

[11] Kuroki H, Tajima N, Kubo S. Clinical results of posterolateral fusion for degenerative lumbar spinal diseases: a follow-up study of more than 10 years. J Orthop Sci 2002;7:317–24.

[12] Glaser J, Stanley M, Sayre H, Woody J, Found E, Spratt K. A 10-year follow-up evaluation of lumbar spine fusion with pedicle screw fixation. Spine 2003;28:1390–5.

[13] Salén BA, Spangfort EV, Nygren ÅL, Nordemar R. The disability rating index: an instrument for the assessment of disability in clinical settings. J Clin Epidem 1994;47:1423–34.

[14] Fairbank JC, Couper J, Davies JB, O'Brien JP. The Oswestry low back pain disability questionnaire. Physiotherapy 1980;66:271–3.

[15] Ware J. SF-36 health survey update. Spine 2000;25:3130–9.

[16] Udén A, Åström M, Bergenudd H. Pain drawings in chronic back pain. Spine 1988;13:389–92.

[17] Altman DG, Practical statistics for medical research. London: Chapman & Hall, 1991.

[18] Shah RR, Mohammed S, Saiffuddin A, Taylor BA. Radiologic evaluation of adjacent superior segment facet joint violation following transpedicular instrumentation of the lumbar spine. Spine 2003;28: 272–5.

[19] Hägg O, Fritzell P, Romberg K, Nordwall A. The General Function Score: a useful tool for measurement of physical disability. Validity and reliability. Eur Spine J 2001;10:203–10.

[20] Möller H, Hedlund R. Instrumented and noninstrumented posterolateral fusion in adult spondylolisthesis. A prospective randomized study: part 2. Spine 2000;13:1716–21.

[21] Fischgrund JS, Mackay M, Herkowitz HN, Brower R, Montgomery DM, Kurtz LT. 1997 Volvo Award winner in clinical studies. Degenerative lumbar spondylolisthesis with spinal stenosis: a prospective randomized study comparing decompression laminectomy and arthrodesis with and without spinal instrumentation. Spine 1997;22: 2807–12.

[22] France JC, Yaszemski MJ, Lauerman WC, et al. A randomized prospective study of posterolateral lumbar fusion. Outcomes with and without pedicle screw instrumentation. Spine 1999;24:553–60.

[23] McGuire RA, Amundson GM. The use of primary internal fixation in spondylolisthesis. Spine 1993;18:1662–72.

[24] Delamarter RB, Fribourg DM, A Kanim LE, Bae H. ProDisc artificial total lumbar disc replacement: introduction and early results from the United States clinical trial. Spine 2003;28:S167–75.