Randomized Clinical Trial of Lumbar Instrumented Fusion and Cognitive Intervention and Exercises in Patients with Chronic Low Back Pain and Disc Degeneration

Jens Ivar Brox, MD, PhD,* Roger Sørensen, MD,* Astrid Friis, PT,* Øystein Nygaard, MD, PhD,† Aage Indahl, MD, PhD,‡ Anne Keller, MD,§ Tor Ingebrigtsen, MD, PhD,‖ Hege R. Eriksen, PhD,¶ Inger Holm, PT,* Anne Kathrine Koller, PT,* Rolf Riise, MD,* and Olav Reikerås, MD*

Study Design. Single blind randomized study.

Objectives. To compare the effectiveness of lumbar instrumented fusion with cognitive intervention and exercises in patients with chronic low back pain and disc degeneration.

Summary of Background Data. To the authors' best knowledge, only one randomized study has evaluated the effectiveness of lumbar fusion. The Swedish Lumbar Spine Study reported that lumbar fusion was better than continuing physiotherapy and care by the family physician.

Patients and Methods. Sixty-four patients aged 25–60 years with low back pain lasting longer than 1 year and evidence of disc degeneration at L4–L5 and/or L5–S1 at radiographic examination were randomized to either lumbar fusion with posterior transpedicular screws and postoperative physiotherapy, or cognitive intervention and exercises. The cognitive intervention consisted of a lecture to give the patient an understanding that ordinary physical activity would not harm the disc and a recommendation to use the back and bend it. This was reinforced by three daily physical exercise sessions for 3 weeks. The main outcome measure was the Oswestry Disability Index.

Results. At the 1-year follow-up visit, 97% of the patients, including 6 patients who had either not attended treatment or changed groups, were examined. The Oswestry Disability Index was significantly reduced from 41 to 26 after surgery, compared with 42 to 30 after cognitive intervention and exercises. The mean difference between groups was 2.3 (−6.7 to 11.4) ($P = 0.33$). Improvements inback pain, use of analgesics, emotional distress, life satisfaction, and return to work were not different. Fear-avoidance beliefs and fingertip-floor distance were reduced more after nonoperative treatment, and lower limb pain was reduced more after surgery. The success rate-according to an independent observer was 70% after surgery and 76% after cognitive intervention and exercises. The early complication rate in the surgical group was 18%.

Conclusion. The main outcome measure showed equal improvement in patients with chronic low back pain and disc degeneration randomized to cognitive intervention and exercises, or lumbar fusion. [Key words: chronic low back pain, disc degeneration, randomized, clinical trial, lumbar fusion, cognitive, exercises, fear-avoidance beliefs] **Spine 2003;28:1913–1921**

Lumbar fusion rates have increased greatly over the past 20 years, but the rates vary from country to country.[1] Fusion rates appear to vary markedly among individual surgeons, among small and large geographic regions in the nation, and between the United States and England.[2] Surgical investigations and interventions account for as much as one third of the health care costs for spinal disorders, but the scientific evidence for most procedures is still unclear.[3] The effect of using instruments, such as pedicle screws and cages, compared with pure osseous fusion has been compared in six randomized studies.[3] Only one of these studies reported a higher success rate in the instrumented group, but this was not a strictly randomized study, and the surgeon evaluated the outcome himself. The 1997 Volvo Award Winner concluded that lumbar posterolateral fusion with pedicle screw fixation increases the operation time, blood loss, and reoperation rate and leads to a significant risk of nerve injury.[4] Functional outcome and fusion rates were not different for fusion with compared to fusion without pedicle screws. The results contradict a hypothesized effect stated in a former European review, in which it was concluded that controlled clinical trials may prevent forthcoming medical licensing authorities from restricting the use of pedicle screw devices and dictating the practice of spinal surgery in Europe in the near future.[5]

One randomized study has compared lumbar fusion with nonoperative treatment.[6] This study found that lumbar fusion reduced pain and decreased disability in comparison with usual care within the primary health care system. The study was criticized for its design because fusion was compared with usual care and not with

From the *Department of Orthopedic Surgery, National Hospital, Oslo, the †Department of Neurosurgery, St. Olav's Hospital, Trondheim, the ‡Coast Hospital for Physical Medicine and Rehabilitation, Stavern, the §Department of Physical Medicine and Rehabilitation, Ullevaal University Hospital, Oslo, the ‖Department of Neurosurgery, University North Norway, Tromsø, and the ¶Department of Biological and Medical Psychology, University of Bergen, Norway.
Acknowledgment date: September 18, 2002. Acceptance date: October 28, 2002.
The manuscript submitted does not contain information about medical device(s)/drug(s). Federal and Foundation funds were received in support of this work. No benefits in any form have been or will be received from a commercial party related directly or indirectly to the subject of this manuscript.
Address reprint requests to Jens Ivar Brox, Department of Orthopaedics, National Hospital, 0027, Oslo, Norway. E-mail: jens.ivar.brox@rikshospitalet.no

a more comprehensive rehabilitation program, including exercise and behavior therapy.[7] According to systematic reviews, there is strong evidence that both exercise and behavior therapy are moderately effective compared with no treatment or usual treatment of patients with chronic low back pain (CLBP).[8,9]

The authors conducted a pilot study of 50 patients on a waiting list for surgery in order to adapt and estimate the preliminary effect of an exercise and behavior modification program on patients with a degenerative lumbar disorder selected for spinal fusion.[10] The goal was to let the patients experience that it is safe to move. At the 6-month follow-up visit, 42% wanted surgery, 16% were uncertain, and 36% no longer wanted surgery. However, the improvement on the Oswestry Disability Index (ODI) was minor, of about the same magnitude as that reported for the nonoperative group in the Swedish Lumbar Spine Study.[6] A previous 5-year follow-up study of patients with degenerative disc disease reported that the improvement in the ODI was about 10-fold compared with the present authors' pilot study.[11] Thus, the present authors anticipated better results after lumbar fusion than with exercise and behavior modification.

The immediate and long-term consequences of fear-avoidance in the initiation and maintenance of chronic pain and disability were outlined in a recent state-of-the-art review.[12] Pain-related fear and avoidance appear to be an essential feature of CLBP resulting in poor behavioral performance. In keeping with this review and the experience from the present authors' pilot study, assessment of fear-avoidance beliefs was included among the secondary outcome measures in the present study.[13]

The purpose of the present study was to compare instrumented lumbar fusion with cognitive intervention and exercises in patients with CLBP and disc degeneration at L4–L5 and/or L5–S1 on radiographic examination. The follow-up time was 1 year, and the ODI was the primary outcome measure. The study is reported according to the Consolidated Standards of Reporting Trials (CONSORT).[7]

■ Patients and Methods

Selection of Patients. Patients with CLBP, consecutively referred from the departments of orthopedic surgery, neurosurgery, and physical medicine and rehabilitation from all regions in Norway during the period 1997–2000 were eligible to participate in the study. At least one spine surgeon and one specialist in physical medicine and rehabilitation examined each patient. All patients underwent plain radiography, computed tomography, and/or magnetic resonance imaging of the lumbar spine.

The inclusion criteria were as follows:

Age 25–60 years.

Pain duration for at least 1 year.

A score of at least 30 of 100 points on the ODI.

Degeneration at L4–L5 and/or L5–S1 (spondylosis) on plain radiographs.

The exclusion criteria were as follows:

Widespread myofascial pain.

Spinal stenosis with reduced walking distance and neurologic signs.

Recurrent disc herniation or lateral recess stenosis with clinical signs of radiculopathy.

Inflammatory disease.

Previous spinal fracture.

Previous surgery of the spine.

Pelvic pain.

Generalized disc degeneration on plain radiographic examination.

Ongoing somatic or psychiatric disease that excluded either one or both treatment alternatives.

Registered medical abuse.

Reluctance to accept one or both of the treatment regimens of the study.

All eligible patients were given oral and written information about the study and the two treatment alternatives. They were told that they could withdraw from the study at any time without any further explanation. The ethics committee for medical research in health region I of Norway approved the study. Of 121 patients referred, 57 did not fulfill the criteria for randomization (Figure 1), and 64 gave signed, informed consent and were randomly allocated to the study groups. The baseline characteristics of the patients are shown in Table 1.

Randomization and Adherence to the Protocol. All patients referred for inclusion were examined at the Orthopedic Department at the National Hospital. When a patient was found eligible and had given an informed, signed consent, the project coordinator (A.F.) telephoned the randomization central at the University of Bergen and reported an identification number. Within an hour the patient was allocated to one of the intervention groups, the project coordinator was phoned back, and the patient was informed. The method of concealed random allocation was used. Simple randomization was conducted by a computer-generated random list. Blocks of patients were used to ensure fairly equal treatment numbers. The project coordinator was not aware of the block size and could not predict the group assignments. Treatments were started within 3 months after the randomization. The patients (n = 37) allocated to fusion underwent surgery at four different departments: Orthopedic Department, National Hospital (n = 26); Neurosurgical Department, St. Olav's Hospital (n = 3); Neurosurgical Department, University Hospital Northern Norway (n = 3); and Orthopedic Department, Ullevaal University Hospital (n = 1). Four patients did not receive the assigned treatment because they changed their minds after having been randomized to lumbar fusion. Patients (n = 27) allocated to cognitive intervention and exercises were treated (n = 25) at the Physiotherapy Department at the National Hospital. Two patients did not receive the assigned treatment because they changed their mind having been randomized to cognitive intervention/exercises.

Baseline Characteristics and Outcome Measures. Baseline characteristics were assessed by a standardized questionnaire. Comorbidity was assessed by a single question: "Do you have other diseases?" Visual analogue scales assessed pretreatment beliefs and expectancies about the effect of surgical and

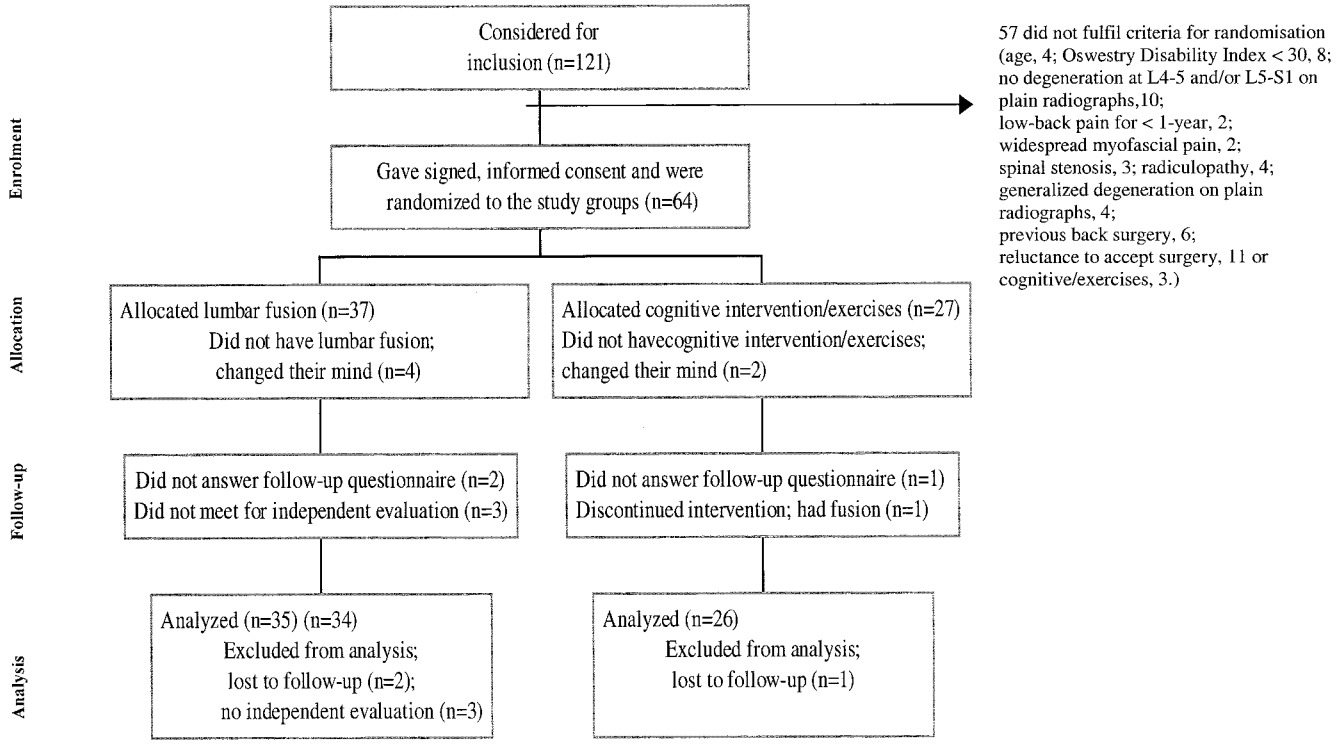

Figure 1. Recruitment and follow-up of study participants.

nonsurgical treatment, respectively. The scales ranged from 0 (no effect) to 100 (complete recovery).

A physical therapist and a specialist in physical medicine and rehabilitation carried out blind follow-up measurements 1 year after the first day of treatment. A nurse in the outpatient clinic always told the patients not to mention anything about their treatment to the independent observers.

The primary outcome measure was predefined in the study protocol. The Norwegian version of the original ODI (version 1.0) was used to evaluate condition-specific disability and pain.[14,15] This score has 10 questions about pain and pain-related disability in activities of daily life and social participation. Each question has six different response alternatives. The sum is calculated and presented as a percentage, wherein 0% represents no pain and disability, and 100% represents the worst possible pain and disability. In a current study evaluating several outcome measures in the target population, the ODI showed the highest test-retest reliability of the variables used in the questionnaire.[16] The interpretation of the results is that changes within 12 points using the ODI could be attributed to measurement error or random variation in a single patient with CLBP. A similar measurement error was reported in another study.[17]

Table 1. Baseline Characteristics of the Patients: Means (Standard Deviations) or Percentages

	Lumbar Fusion n = 37	Cognitive/Exercises n = 27
Age (yr)	44.1 (8.1)	42.4 (7.8)
Gender (% men)	43	33
Duration (yr)	9.6 (8.7)	12.5 (11.6)
Educational level (%)		
Primary school (9 yr)	37	44
High school (12 yr)	31	23
University/college	32	33
Work status* (%)		
Working	24	22
On sick leave	30	26
Rehabilitation	30	41
Disability pension	11	11
Homemaker	5	
Beliefs in surgery†	69.1 (17.6)	74.2 (24.2)
Beliefs in nonsurgical treatment†	40.2 (26.9)	36.1 (25.9)
Comorbidity (%)	32	22
Taking analgesics (%)	54	52
Smoking (%)	41	44
Married/living together (%)	86	81

* Two patients in the fusion group were unemployed.
† 0, no effect, 100, complete recovery.

Secondary outcome measures were assessed in a standardized questionnaire, which was completed before inclusion and at the 1-year follow-up visit. Patients scored the intensity of their back and lower limb pain on vertical visual analogue scales, ranging from 0 to 100, wherein 100 reflected the worst pain imaginable. Maximum pain, minimum pain, and current pain were scored on three different scales. The mean of the three measurements provided the pain index for back pain or lower limb pain, respectively.[6] The use of daily pain medication was registered at the last week before inclusion and the 1-year follow-up visit, respectively. The patients were asked to write down every single drug they used, and the exact dose. Medication was classified using anatomic therapeutic chemical codes and daily defined doses.[18,19] The anatomic therapeutic chemical classification system has been recommended by the World Health Organization for use in drug consumption studies.[20] The consumption of each drug was calculated with defined daily doses as a measurement unit and classified and presented by therapeutic group according to the anatomic therapeutic chemical system.[19] The therapeutic groups were as follows: analgesics; anxiolytics, hypnotics, and sedatives; antidepres-

sants; antiinflammatory agents; and muscle relaxants. One defined daily dose equals 3 g paracetamol or 0.5 g naproxen.

The General Function Score was used to measure back-related disability in activities of daily living.[17] Patients answered nine questions using one of the three alternatives: "can perform," "can perform with difficulty due to back complaints" and "cannot perform due to back complaints." The score was presented as a percentage wherein 100% represents maximum disability.

Emotional distress was rated by the Hopkins Symptom Check List-25).[21] The patients ranked their 25 symptoms from 1 (not at all) to 4 (extreme). A mean symptom score of 1.75 or more was observed in 20% of women and 9% of men in a large Norwegian epidemiologic study.[22] A score above 1.75 is a high predictor of current help-seeking, but seems to reflect illness or nonspecific distress more than psychiatric diagnoses.[23]

Waddel's Fear-Avoidance Belief Questionnaire (FABQ) was used to quantify fear-avoidance beliefs.[13] The FABQ used was a 10-item questionnaire with each item scored from 0 to 6, higher numbers indicating increased levels of fear-avoidance beliefs. Two subscales were used: a 4-item scale measuring fear-avoidance beliefs about physical activity (FABQ-PA) and a 6-item scale assessing fear-avoidance beliefs about work (FABQ-W). The FABQ-PA had a possible range of 0–24; the FABQ-W a possible range of 0–36. In a recent methodologic study, the present authors found considerably higher variance for this questionnaire compared with the results from Waddel *et al*.[16] An explanation for the agreement differences between the two studies may be the time frame used. In the study by Waddel *et al*, the interval between the two tests was 48 hours, compared to 2 weeks in the recent study.

Life satisfaction was scored on a vertical visual analogue scale from 1 (worst possible) to 10 (best possible).[24] This score showed a considerable variation when measurements were made twice within 2 weeks in patients with CLBP.[16]

Patients rated their overall function by the Global Back Disability question, answered only at the 1-year follow-up visit.[16] This is a single question designed to measure the patients' overall rating of their back disability today. It was designed by one of the authors (J.I.B.) based on previous experiences with the evaluation of patients after shoulder surgery.[25] The item had been pretested in a pilot study that included 50 patients with CLBP on a waiting list for spinal surgery. There were five response alternatives: "excellent, no or unimportant complaints," "good, occasionally bothered by back pain," "fair, some back pain and limited function," "poor, unchanged, considerable complaints and severe disability," and "miserable, worse, not self-reliant in activities of daily living." The authors have previously found that the reliability and construct validity of this question is good in the target population.[16]

Evaluation of work status included questions about paid work (full time, part time, not working) and status if not working (sick leave, rehabilitation, disability pension, unemployed, homemaker or student). Within the Norwegian social insurance system, persons who have been on sick leave for 1 year are entitled to a rehabilitation benefit or a disability pension. The aim was to obtain core data about work without obtaining permission from the National Insurance Office. The reliability (kappa) of work status was 0.94 for the first question and 0.61 for the second question in a previous study.[16]

A specialist in physical medicine and rehabilitation working at another hospital and a physical therapist not engaged in the treatments and working in another building at the National Hospital evaluated the patients at the 1-year follow-up visit. At the follow-up evaluation, the patients wore T-shirts to hide the scars from surgery that some had received. The physician rated the patient's overall functional and work status using the Prolo Scale.[26] This scale has been applied in large series to evaluate the results of lumbar discectomy.[27,28] It was compared with a questionnaire in another study, and[29] the results were similar with the two outcome measures. The Prolo Scale has two parts: economic status and functional status. The authors reported the results for functional status, which ranked pain responses and the effect of pain on activities of daily living in five categories: complete recovery, recurrence, perform all activities except sports, same as before operation, and worse.

The physical therapist supervised physical tests. Fingertip-floor distance was measured as described by Hyytiäinen, with the patient standing on a platform and bending forward.[30] The isokinetic trunk muscle test was used to assess maximal trunk extension muscle performance.[31] In addition, a radiologist measured back muscle size and density in a random sample before treatment and at the 1-year follow-up visit. The results and detailed methodology of measurements for back muscle strength and morphology are reported in a separate article.[32]

Radiographic fusion was assessed by experienced radiologists. Anteroposterior and lateral radiographs of the lumbar sacral spine were taken. Fusion was graded as fused or not fused. A previous study using kappa statistical analysis revealed only fair interobserver agreement in grading lumbar fusion status.[33] Recently a more detailed radiographic classification system has been presented. This system has good inter- and intraobserver agreement both with and without instrumentation.[34]

Treatments. Experienced orthopedic surgeons at the National Hospital or Ullevaal University Hospital in Oslo and experienced neurosurgeons at St. Olav's University Hospital in Trondheim and the University Hospital North Norway in Tromsø performed the instrumental lumbar fusions. The aim of the procedure is to stabilize the involved segments and to reduce pain. The standard treatment consisted of posterolateral fusion with transpedicular screws of the L4–L5 segment and/or the L5–S1 segment. Autologous bone was used in all cases. Postoperative rehabilitation was at the choice of the surgeon, not according to any study protocol. As a standard procedure, physical therapists at the respective departments gave advice about physical activity during the first 3 months after surgery. Patients had follow-up consultations by the surgeon at 3 and 6 months. At the 3-month follow-up visit, the surgeon customarily prescribed physiotherapy, including exercises. The numbers of prescribed sessions varied.

The cognitive intervention and exercises took place in the Physiotherapy Department at the National Hospital. The program had been evaluated in a pilot study.[10] The program was modified: the link between the information given and the content of the physical exercises was better coordinated, and the physical therapists had gained more confidence to challenge and encourage the patients. The first week, a specialist in physical medicine and rehabilitation (A.I.) gave a lecture to the patients to describe the pain receptors in the discs, facet joints, and muscles; the reflexive interplay between various structures; and the ability to suppress and reinforce various peripheral stimuli. The patients were given to understand that they could not do any harm to the disc (back) by engaging in the ordinary activities of daily life. The patients were told to use their backs, to bend them, and not to be too cautious.[35] This information

was reinforced every day during various physical activities and discussions. An attempt was made to give consistent information, and disagreements between the supervisors were discussed to reach agreement.

Recent evidence-based advice recommends that back patients stay active.[36,37] Nevertheless, patients with back pain are often told to be careful. It has been proposed that cultural changes have led to greater awareness and more disability regarding back symptoms over the past decade.[38] In the authors' experience, the patients included in the present study had previously been given restrictive advice about physical activity and work by physicians and physical therapists. The patients assigned to cognitive intervention and exercises were challenged in their thoughts about, and participation in, physical activities previously labeled as not recommended. This included vacuum cleaning, jumping, lifting, and ball games. According to knowledge gained from studies in applied work physiology, the patients were told to bend their backs while lifting light objects and to bend their knees while lifting heavy objects.[39]

Individual exercises were given for endurance and coordination. This part included a specific exercise intervention that advocates training the cocontraction of the deep abdominal muscles with the lumbar multifidus, performed according to the principles outlined by O'Sullivan.[40] In addition, individual goals for the rehabilitation process were established, based on the patient's answers to the comprehensive questionnaire (thoughts and feelings) and their test results (physical function and behavior).

The duration of the supervised treatment period was 1 week at first, followed by 2 weeks at home and another treatment period of 2 weeks. The intensity of the physical activities was gradually increased during the last 2 weeks. The average duration of the rehabilitation program was about 25 hours per week. Because the patients were recruited from all over Norway, most of them stayed at a patient hotel, and treatments were conducted in the outpatient clinic during the day. The groups consisted of 4 to 7 patients. Three daily workouts were performed: aerobics or outdoor activities, water gymnastics, and individual exercises. Additionally, individual consultations, group lessons, and discussions were given. One of the group lessons focused on imaging, and the patients' radiographs, computed tomography scans, and/or magnetic resonance imaging scans were demonstrated. Every group met a peer for exchanging experiences. The peer was a former participant in the program. All patients were offered a home program and a training diary. Follow-up consultations and tests were conducted at 3 and 6 months.

Sample Size and Statistical Analysis. The study was planned to detect the difference in change from base line to the 1-year follow-up visit of 10 points between groups on the ODI. After a pilot study, the standard deviation was estimated at 10 points.[10] With α set at 0.05 (type I error) and β at 0.05 (type II error), 26 patients were required for each treatment group to complete the trial.[41] The results were analyzed according to the method of intention-to-treat. The mean differences in change between groups and 95% confidence intervals were established. Multiple regression was used to measure point estimates for and confidence intervals for group differences in ODI from baseline to the 1-year follow-up visit after adjustment for gender and pretreatment beliefs in the surgery and nonsurgical treatment groups. The nonparametric Mann-Whitney rank sum test was applied to detect significant differences between groups from baseline to the 1-year follow-up visit (Table 2).

In addition, Bonferroni corrections were made for secondary outcome measures by simply multiplying secondary outcome *P* values by the number (n = 12) of significance tests. Categorical data for work and overall rating were dichotomized. According to a recent systematic review, success was defined as the three best grades (all responses except unchanged and worse) for the Prolo Scale and the Global Back Question.[3] For the overall ratings at the 1-year follow-up visit, the odds ratio and the number needed to treat were calculated, and the two groups were compared with the use of Fisher's exact test. Mathematically, the number needed to treat equals the reciprocal of absolute risk reduction.[42] Thus, the number needed to treat describes the number of patients needed to have surgery, according to the patients' overall rating, to expect one successful outcome compared with the nonsurgical intervention.

Additional analyses included the nonparametric Wilcoxon matched pairs signed rank sum test to assess improvement within groups from baseline to the 1-year follow-up visit. Also, a second analysis, which included only the patients who completed the study, paralleled the intention-to-treat analysis. Analyses were performed with the use of SPSS software, version 11.0.

■ Results

Patients

The 1-year follow-up rate was 97%. The percentage of men was lower in the group given cognitive intervention and exercise (33%) than in the group given lumbar fusion (43%) (Table 1). The groups did not differ in age, duration of disease, or occupational education. Twenty-four percent of the patients given surgery and 22% of the patients given cognitive intervention and exercise were working at the time of inclusion. Beliefs in surgery were about twice as high as beliefs in nonsurgical treatment in both groups (Table 1).

Compliance and Follow-up

Three patients in the fusion group and one patient in the cognitive/exercise group did not have an independent evaluation at the 1-year follow-up visit (Figure 1). Four patients randomized to lumbar fusion did not have surgery. Two patients randomized to cognitive intervention and exercises did not attend treatment. Additionally, 1 patient dropped out after the first treatment period and had surgery. The mean number of physiotherapy sessions given after the hospital stay was 31 (SD 28) in the surgical group and 3 (SD 7) for cognitive intervention and exercise.

Efficacy

The overall improvements in the ODI at the 1-year follow-up visit did not differ significantly between treatments (Table 2). The mean difference in change between groups was 2.3 (95% confidence interval −6.8 to 11.4), 2.6 (−6.5 to 11.7) after adjustment for gender, and 2.7 (−6.8 to 12.2) after adjustment for gender and pretreatment beliefs in the effect of surgical and nonsurgical treatment. The mean scores at baseline, 3 and 6 months,

Table 2. Primary and Secondary Outcome Measure

Outcome	Lumbar Fusion (n = 35)*	Cognitive/Exercises (n = 26)*	Mean Difference Between Groups from Baseline to 1 year (95% CI)†	*P* value†
Primary Disability Index				
Oswestry (0–100)				
Baseline	42.0 (11.0)	43.0 (13.0)		
1 year	26.4 (16.4)	29.7 (19.6)	2.3 (−6.8 to 11.4)	0.33
Secondary‡				
General Function Score (0–100)				
Baseline	35.9 (18.6)	44.6 (13.7)		
1 year	18.3 (17.3)	22.6 (18.9)	−4.1 (−14.9 to 6.7)	0.50
Back pain (0–100)				
Baseline	62.1 (14.5)	64.1 (13.7)		
1 year	39.4 (25.5)	48.7 (24.0)	8.6 (−3.0 to 20.1)	0.14
Lower limb pain (0–100)				
Baseline	43.5 (27.7)	34.0 (19.3)		
1 year	26.6 (28.1)	35.5 (30.6)	17.5 (4.3 to 30.7)	0.002
Medication (range) (DDD§)				
Baseline	0.2 (0 to 3.2)	0.3 (0 to 2.1)		
1 year	0.1 (0 to 12.8)	0.1 (0 to 1.7)	−0.4 (−1.1 to 0.5)	0.44
Emotional distress (1–4)				
Baseline	1.8 (0.5)	1.9 (0.6)		
1 year	1.5 (0.4)	1.6 (0.5)	0.1 (−0.2 to 0.3)	0.35
Life satisfaction (1–10)				
Baseline	5.4 (2.4)	4.8 (2.0)		
1 year	6.2 (2.4)	6.6 (2.1)	−0.8 (−2.1 to 0.5)	0.42
Fear-avoidance physical activity (0–24)				
Baseline	13.7 (4.8)	16.4 (5.3)		
1 year	11.5 (6.3)	6.5 (6.0)	−7.7 (−11.6 to −3.8)	<0.001
Fear-avoidance work (0–42)				
Baseline	26.8 (9.8)	27.1 (12.4)		
1 year	27.9 (12.0)	21.5 (14.4)	−8.3 (−13.7 to −3.0)	0.002
Fingertip-floor distance (cm)				
Baseline	15.1 (17.5)	22.5 (24.5)		
1 year	13.4 (13.5)	7.1 (14.7)	−13.7 (−21.6 to −4.2)	0.009
Patients overall rating: no. (%) success				
1 year	25 (71)	17 (63)	1.5 (0.5 to 4.3)§§	0.59
Independent observer: no. (%) success				
1 year	24 (71)	20 (77)	0.7 (0.2 to 2.3)§§	0.77
Work: no. (%)				
Baseline	9 (24)	6 (22)		
1 year	8 (22)	9 (33)		0.40

Means (standard deviations) at baseline and 1 year are given for lumbar fusion and cognitive/exercises unless stated otherwise. Means (95% confidence intervals) are given for the differences between groups from baseline to 1 year unless stated otherwise.

* Two patients in randomized surgery and one patient on randomized cognitive/exercises did not attend 1-year follow-up. One patient given surgery was not examined by the independent observer.

† *P* values for the difference in change from baseline to 1-year follow-up were calculated with the Mann-Whitney test. The comparisons of ordinal data were calculated with Fisher's exact test.

‡ Bonferroni correction for secondary outcome measure can be obtained by multiplying p-values with 12.

§ DDD = Daily Defined Doses based on ATC codes. §§ Odds ratio (95% CI) for success at 1-year.

and the 1-year follow-up visit are shown in Figure 2. Both groups had improved significantly on all outcome measures at the 1-year follow-up visit except for lower limb pain (cognitive/exercise group) and fear-avoidance beliefs and fingertip-floor distance (fusion group).

Fear-avoidance beliefs were significantly more reduced in the patients given cognitive intervention and exercise, and lower limb pain was more reduced in those given surgery. The reported use of analgesics varied substantially between individuals but was modest between groups.

The patients' overall ratings were not significantly different in the two groups. The number needed to treat was 12.5. Only a limited number of patients were working, and there was no difference between groups in terms of return to work.

The differences for ODI from baseline to the 1-year follow-up visit ranged from −4 to 20 (mean = 6) in the 3 patients who dropped out from surgery and from −2 to 46 (mean = 15) in the 3 patients who either did not attend cognitive intervention and exercise or underwent surgery. The difference between groups for ODI from baseline to the 1-year follow-up visit in the patients (n = 33) who adhered to their assigned treatment was 1.3 (−8.5 to 11.0).

The median difference for ODI from baseline to the 1-year follow-up visit was 13 (range −8 to 48) in patients operated on at the National Hospital (n = 26) and 10 (range −12 to 46) in patients operated on at the other hospitals (n = 7). The corresponding results were 19 (range −12 to 46) for single-level L5–S1 fusions and 9 (range −8 to 48) for two-level L4–S1 fusions.

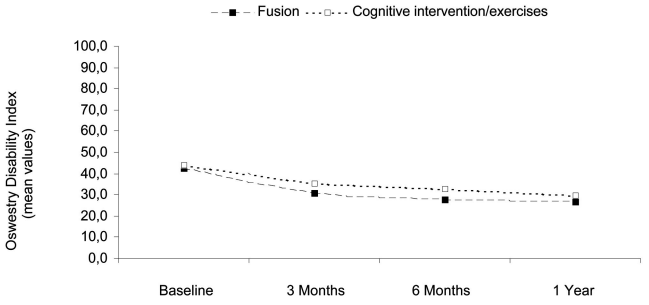

Figure 2. Mean values for Oswestry Disability Index (ODI) in each study group. The ODI was the predefined main outcome variable and the study was designed to detect a difference in change of at least 10 points between groups. This measure consists of 10 questions about pain, pain related disability of activities of daily life and social participation. The total score ranges from 0 (no pain and disability) to 100 (worst possible pain and disability).

The fusion rate was 84%, and the early complication rate after surgery was 18% (6/33 patients). There were two wound infections, two bleedings, one dural tear, and one venous thrombosis, but no late complications.

■ Discussion

The object of this study was to compare instrumented lumbar fusion with cognitive intervention and exercise in patients with CLBP and disc degeneration of L4–L5 and/or L5–S1 on radiographic examination. The observed difference was much less than the difference that the authors had considered clinically relevant before starting the study. Moreover, it was less than the measurement error estimated in another study of similar patients. The small size of the present study and the large variation between patients are reflected in the wide confidence intervals. The confidence interval of the main outcome measure included 10.0, which the authors had considered to be clinically relevant. In retrospect, the authors estimated the size of a study designed to detect the observed mean difference of 2.3 between groups. Based on the observed standard deviation for the difference from baseline to 1 year in the fusion group or the cognitive intervention/exercise group, and a power of 80%, it will take approximately 900 patients in each group to detect the difference observed in the present study as statistically significant.[43] Such a study is unrealistic to perform, and considering new technologic innovations it will have merely historical interest. For example, it took 6 years to include 289 patients in the Swedish Lumbar Spine Study.[6]

The lack of a "no treatment" group is a limitation of the present study. This raises the possibility that neither treatment was effective and that the modest improvements noted in the primary outcome measure simply reflected the natural history of the problem among subjects recruited. Our results for both interventions resemble the results in the fusion group in the Swedish Lumbar Spine Study. This suggests that both interventions are modestly superior compared with "no treatment." The short-term effects may be attributed to the increased attention and care provided by the two interventions, but the 1-year results for fear-avoidance beliefs compared with lumbar fusion suggests a change in pain behavior in the group assigned to cognitive intervention and exercise.

Like other painful conditions, CLBP offers few objective clinical variables for use in follow-up studies. The results described here are based on a validated predefined primary outcome measure and are supported by measures of secondary outcome variables. The calculation of the number needed to treat was based on patients' overall rating, and the observed result means that of every 12 or 13 patients who receive surgery, 1 will benefit, in comparison with cognitive intervention and exercise. However, this assumption is based on one of the secondary outcome measures. Replacing patients' overall rating by another secondary outcome measure, namely, the independent observer's evaluation, yields a number needed to treat in favor of cognitive intervention and exercise. This is in keeping with the main result that there was no benefit from fusion in the patients in the present study.

The differences in favor of surgery for lower limb pain may result from local denervation after surgery. Surgery carries strong implications of success.[44] For example, relief of sciatica and back pain has been reported in at least one third of back surgery patients who had no disc herniation.[45] Also, possible placebo effects may have been introduced in the cognitive intervention and exercise group because they may have received more attention over a longer time. However, the difference between groups was practically unaltered when pretreatment beliefs in the effect of surgical and nonsurgical treatment were controlled for. According to this, the authors have no reason to consider that expectations influenced the observed difference between groups.

Confrontation and avoidance are postulated as the two extreme responses to the fear of pain.[46] A consequence of avoidance is escape and avoidance behavior, resulting in poor performance and muscle reactivity, physical disuse, and guarded movements.[12] In the present study, the patients who were randomized to cognitive intervention and exercises were first given the understanding that they could not do any harm to the disc (back) by engaging in the ordinary activities of daily life. Then they were encouraged to participate in physical activities that were previously not recommended, to achieve confidence. This is in contrast with the message from classic back schools, where patients were told to keep their back straight, and not participate in regular activities without restriction. In a recent study, patients who had undergone disc surgery were recommended to go back to ordinary activity as soon as the stitches from surgery were removed.[47] The results were excellent, with no complications from the early activation. Most likely, patients undergoing lumbar fusion may also benefit from a cognitive approach, which emphasize understanding, early activation, and no restrictions.

Chronic lower back pain is often attributed to the intervertebral disc, but disc degeneration is observed in pain-free individuals.[48] Recent studies suggest that invasive procedures like discography are not helpful in the selection of patients for fusion unless information from psychologic testing is considered.[49,50]

Both the baseline characteristics of the patients and the results in this study are comparable with the improvements on identical outcome measures after fusion in the Swedish Lumbar Back Study.[6] The Swedish study was criticized for faulty design because it did not compare fusion with the best nonsurgical alternative but continued previous physiotherapy, thus resembling a waiting list.[51] The present study is the first study, as far as the authors are aware, to compare lumbar fusion with an evidence-based nonsurgical approach. Most likely, the difference in outcome in the two studies is related to the difference in the nonsurgical intervention. According to evidence-based medicine, behavior therapy or exercise is better than usual treatment in patients with CLBP.[12,16] The results of the present study suggest that the combination of these treatments emerges as an alternative to surgery in these patients.

In addition, the Swedish Lumbar Back Study estimated that treatment costs were three times as high for surgery as for conservative treatment.[52] Because of a net return to work after surgery, the researchers concluded that the total costs were lower for surgery. The present authors observed no advantage for surgery in terms of return to work. Only a few patients returned to work, and the tendency was in favor of cognitive intervention and exercise. According to a recent large Norwegian epidemiologic study, disability pension for back pain is strongly associated with educational level and socioeconomic position.[53] In keeping with this finding, the majority of patients in the present study had lower education and did not return to work. It is reasonable to assume that work rehabilitation and adult education could have improved return to work in the study population. The educational level of the patients was not reported in the Swedish Lumbar Spine Study. Based on the results of the present study, both costs for treatment and total costs are most likely considerably higher for surgery than for cognitive intervention and exercise.

It was concluded that after 1 year of follow-up, the difference between the groups given lumbar instrumented fusion and cognitive intervention and exercise was neither clinically important nor significant. Most cases of CLBP can be managed by cognitive intervention and exercise, with lumbar fusion as a more expensive alternative. The rehabilitation program that was compared with lumbar fusion in the present study is not widely available, but behavior intervention and exercise were better than usual physiotherapy in a previous randomized Swedish study.[54] It is suggested that future studies should evaluate strategies toward the implementation of cognitive intervention and exercise for CLBP. The content and comprehensiveness of behavior intervention and exercise programs should be assessed. In particular, the duration longer than 100 hours, as suggested by a recent systematic review of multidisciplinary rehabilitation for CLBP, *versus* content, should be addressed.[55]

Acknowledgments

The authors thank the patients who participated in the study, the nurses and nurse aids at the hospital departments and outpatient clinics, and the physicians at the departments who referred patients for the study. They thank Professor H. Ursin for randomization of the patients and comments on study design and the manuscript, back surgeons O. Grundnes and J. Lange for operating on the patients and contributing to their recruitment, and physiotherapists M.Aa. Fosdahl and T. Haakenstad for supervising the exercise program and participating in the cognitive intervention. They also thank the orthopedic surgeons and neurosurgeons who assisted in the lumbar fusions, and radiologists R. Gunderson and A.M. Finnanger for their assistance.

■ Key Points

- Recently one randomized study (Swedish Lumbar Spine Study) reported that lumbar fusion reduced pain and improved disability more efficiently than commonly used nonsurgical treatment of patients with chronic low back pain and disc degeneration.
- The Swedish Lumbar Spine Study was criticized because nonsurgical therapy continued previous physiotherapy and resembled a waiting list group.
- Using similar inclusion criteria and outcome measures, the authors compared instrumented lumbar fusion with a rehabilitation program emphasizing cognitive intervention and exercises.
- At the 1-year follow-up visit, the main outcome measure (Oswestry Disability Index) showed equal improvement in patients who were randomized cognitive intervention and exercise or to instrumented lumbar fusion.
- Fear-avoidance beliefs for physical activity were significantly reduced in the cognitive intervention and exercise group, compared with patients given lumbar instrumented fusion and postoperative physiotherapy.[55]

References

1. Cherkin DC, Deyo RA, Loeser JD, et al. An international comparison of back surgery rates. Spine 1994;19:1201–6.
2. Katz JN. Lumbar spinal fusion. Surgical rates, costs, and complications. Spine 1995;20:78.S-83S.
3. Gibson JN, Grant IC, Waddell G. The Cochrane review of surgery for lumbar disc prolapse and degenerative lumbar spondylosis. Spine 1999;24: 1820–32.
4. Thomsen K, Christensen FB, Eiskjær SP, Hansen ES, Fruensgaard S, Bünger CE. Volvo Award winner in clinical studies. The effect of pedicle screw instrumentation on functional outcome and fusion rates in posterolateral lumbar spinal fusion: a prospective, randomized clinical study. Spine 1997; 1997:22.:2813–22.
5. Boos N, Webb JK. Pedicle screw fixation in spinal disorders: a European view. Eur Spine J 1997;6:2–18.

6. Fritzell P, Hägg O, Wessberg P, Nordwall A. 2001 Volvo award winner in clinical studies: Lumbar fusion versus nonsurgical treatment for chronic low back pain: a multicenter randomized controlled trial from the Swedish lumbar spine study group. Spine 2001;26:2521–32.
7. Moher D, Schulz KF, Altman D. The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomised trials. Lancet 2001;357:1191–4.
8. van Tulder M, Malmivaara A, Esmail R, Koes B. Exercise therapy for low back pain: a systematic review within the framework of the Cochrane collaboration back review group. Spine 2000;25:2784–96.
9. van Tulder MW, Ostelo R, Vlaeyen JW, Linton SJ, Morley SJ, Assendelft WJ. Behavioral treatment for chronic low back pain: a systematic review within the framework of the Cochrane Back Review Group. Spine 2001;26:270–81.
10. Brox JI, Friis A, Holm I, et al. Patients with chronic degenerative spinal disease: Can conservative treatment reduce the waiting list for surgery? Tidsskr Nor Lœgeforen 1999;119:1784–7.
11. Buttermann GR, Garvey TA, Hunt AF, et al. Lumbar fusion results related to diagnosis. Spine 1998;23:116–27.
12. Vlaeyen JW, Linton SJ. Fear-avoidance and its consequences in chronic musculoskeletal pain: a state of the art. Pain 2000;85:317–32.
13. Waddel G, Newton M, Henderson I, et al. A fear-avoidance beliefs questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. Pain 1993;52:157–68.
14. Fairbank JCT, Davis J, Couper J, et al. Oswestry disability questionnaire. Physiotherapy 1980;66:271–3.
15. Roland M, Fairbank J. The Roland-Morris Disability Questionnaire and the Oswestry Disability Questionnaire. Spine 2000;25:3115–24.
16. Holm I, Friis A, Storheim K, et al. Measuring self-reported functional status and pain in patients with chronic low back pain by postal questionnaires: A reliability study. Spine 2003;28:828–33.
17. Hägg O, Fritzell P, Romberg K, et al. The General Function Score: A useful tool for measurement of physical disability Validity and reliability. Eur Spine J 2001;10:203–10.
18. Brox JI, Staff PH, Ljunggren AE, et al. Arthroscopic surgery compared with supervised exercises in patients with rotator cuff disease (stage II impingement syndrome). BMJ 1993;307:899–903.
19. WHO Collaborating Centre for Drug Statistics Methodology. ATC Index with DDDs. Oslo, Norway: Who Collaborating Centre, 2002.
20. Rønning M, Blix HS, Harbo BT, et al. Different versions of the anatomical therapeutic chemical classification system and the defined daily dose: Are drug utilisation data comparable? Eur J Clin Pharmacol 2000;56:723–7.
21. Derogatis LR, Lipman RS, Rickels K, et al. The Hopkins Symptom Checklist (HSCL): A self-report inventory. Behav Sci 1974;19:1–15.
22. Sandanger I, Nygård JF, Ingebrigtsen G, et al. Prevalence, incidence and age at onset of psychiatric disorders in Norway. Soc Psychiatry Psychiatr Epidemiol 1999;34:570–9.
23. Sandanger I, Moum T, Ingebrigtsen G, et al. The meaning and significance of caseness: the Hopkins Symptom Checklist-25 and the Composite International Diagnostic Interview: II. Soc Psychiatry Psychiatr Epidemiol 1999;34:53–9.
24. Andrews FM, Robinson JP. Measures of subjective well-being. In: Robinson JP, Shaver PR, Wrightsman LS, eds. Measures of Personality and Social Psychological Attitudes. San Diego: Academic Press, 1991.
25. Brox J, Gjengedal E, Uppheim G, et al. Arthroscopic surgery versus supervised exercises in patients with rotator cuff disease (stage II impingement syndrome): A prospective, randomized, controlled study in 125 patients with a 2 1/2 year follow-up. J Shoulder Elbow Surg 1999;8:102–11.
26. Prolo D, Oklund S, Butcher M. Toward uniformity in evaluating results of lumbar spine operations. Spine 1986;11:601–6.
27. Berger E. Late postoperative results in 1000 work-related lumbar spine conditions. Surg Neurol 2000;54:101–6.
28. Davis RA. A long-term outcome analysis of 984 surgically treated herniated lumbar discs. J Neurosurg 1994;80:415–21.
29. Wörtgen C, Holzschuh M, Rothoerl RD, et al. Does the choice of outcome scale influence prognostic factors for lumbar disc surgery? A prospective, consecutive study of 121 patients. Eur Spine J 1997;6:173–80.
30. Hyytiäinen K, Salminen JJ, Suvitie T, et al. Reproducibility of nine tests to measure spinal mobility and trunk muscle strength. Scand J Rehabil Med 1991;23:3–10.
31. Keller A, Hellesnes J, Brox JI. Reliability of the Isokinetic Trunk Extensor Test, Biering-Sorensen Test, and Åstrand Bicycle Test: Assessment of intraclass correlation coefficient and critical difference in patients with chronic low back pain and healthy individuals. Spine 2001;26:771–7.
32. Keller A, Brox JI, Gunderson R, et al: Trunk muscle strength, cross-sectional area and density in patients with chronic low back pain randomized to lumbar fusion and cognitive intervention and exercises. Spine 2003;28: submitted for publication.
33. Hamill CL, Simmons ED Jr. Interobserver variability in grading lumbar fusions. J Spinal Disord 1997;10:387–90.
34. Christensen FB, Laursen M, Gelineck J, et al. Interobserver and intraobserver agreement of radiograph interpretation with and without pedicle screw implants: The need for a detailed classification system in posterolateral spinal fusion. Spine 2001;26:538–43.
35. Indahl A, Velund L, Reikerås O. Good prognosis for low back pain when left untampered: A randomized clinical trial. Spine 1995;20:473–7 [see comments].
36. Buchbinder R, Jolley D, Wyatt M. 2001 Volvo award winner in clinical studies: Effects of a media campaign on back pain beliefs and its potential influence on management of low back pain in general practice. Spine 2001; 26:2535–42.
37. Deyo RA. Acute low back pain: A new paradigm for management: Limited imaging and early return to normal activities. BMJ 1996;313:1343–5.
38. Palmer KT, Walsh K, Bendall H, et al. Back pain in Britain: comparison of two prevalence surveys at an interval of 10 years. BMJ 2000;320:1577–8.
39. Hagen KB, Harms-Ringdahl K. Ratings of perceived thigh and back exertion in forest workers during repetitive lifting using squat and stoop techniques. Spine 1994;19:2511–7.
40. O'Sullivan PB, Twomey LT, Allison GT. Evaluation of specific stabilizing exercise in the treatment of chronic low back pain with radiologic diagnosis of spondylosis or spondylolisthesis. Spine 1997;22:2959–67.
41. Pocock SJ. Clinical Trials. New York: John Wiley, 1983.
42. Cook RJ, Sackett DL. The number needed to treat: A clinically useful measure of treatment effect. BMJ 1995;310:452–4.
43. Altman DG. Practical Statistics for Medical Research, 2nd ed. London: Chapman & Hall, 1991:456.
44. Turner JA, Deyo RA, Loeser JD, et al. The importance of placebo effects in pain treatment and research. JAMA 1994;271:1609–14.
45. Spangford EV. The lumbar disc herniation: A computer-aided analysis of 2504 operations. Acta Orthop Scand 1972;342(Suppl):1–95.
46. Lethem J, Slade PD, Troup JD, et al. Outline of a fear-avoidance model of exaggerated pain perception: I. Behav Res Ther 1983;21:401–8.
47. Carragee EJ, Han MY, Yang B, et al. Activity restrictions after posterior lumbar discectomy. Spine 1999;24:2346–51.
48. Boos N, Rieder R, Schade V, et al. Volvo award in clinical sciences: The diagnostic accuracy of magnetic resonance imaging, work perception, and psychosocial factors in identifying symptomatic disc herniations. Spine 1995;20:2613–25.
49. Carragee EJ, Chen Y, Tanner CM, et al. Provocative discography in patients after limited lumbar discectomy: A controlled, randomized study of pain response in symptomatic and asymptomatic subjects. Spine 2000;25:3065–71.
50. Carragee EJ, Tanner CM, Khurana S, et al. The rates of false-positive lumbar discography in select patients without low back symptoms. Spine 2000;25: 1373–80.
51. Mooney V. Point of view. Spine 2001;26:2532–3.
52. Fritzell P, Hägg O, 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 2002;27:1131–41.
53. Hagen KB, Holte HH, Tambs K, et al. Socioeconomic factors and disability retirement from back pain: A 1983–1993 population-based prospective study in Norway. Spine 2000;25:2480–7.
54. Lindström I, Öhlund C, Eek C, et al. Mobility, strength, and fitness after a graded activity program for patients with subacute low back pain. Spine 1992;17:641–9.
55. Guzman J, Esmail R, Karjalainen K, et al. Multidisciplinary rehabilitation for chronic low back pain: Systematic review. BMJ 2001;322:1511–6.