

R. W. J. G. Ostelo
H. C. W. de Vet
M. W. Berfelo
M. R. Kerckhoffs
J. W. S. Vlaeyen
P. M. J. C. Wolters
P. A. van den Brandt

Effectiveness of behavioral graded activity after first-time lumbar disc surgery: short term results of a randomized controlled trial

Received: 29 June 2002
Revised: 20 January 2003
Accepted: 21 March 2003
Published online: 23 September 2003
© Springer-Verlag 2003

The research for this study was carried out at the Department of Epidemiology, Maastricht University. The research was supported by “Profileringsfonds” (PF-57) of the Maastricht University Hospital and “Stichting Annafonds Leiden”

R. W. J. G. Ostelo (✉) · H. C. W. de Vet
Institute for Research in Extramural
Medicine (EMGO Institute),
VU University Medical Center,
van der Boechorststraat 7,
1081 BT, Amsterdam, The Netherlands
Tel.: +31-20-4448180,
Fax: +31-20-4448181,
e-mail: r.ostelo.emgo@med.vu.nl

R. W. J. G. Ostelo · M. R. Kerckhoffs
P. M. J. C. Wolters
Department of Physiotherapy,
Hogeschool Zuyd, Heerlen, The Netherlands

R. W. J. G. Ostelo · M. R. Kerckhoffs
P. A. van den Brandt
Department of Epidemiology,
Maastricht University, The Netherlands

M. W. Berfelo
Department of Neurosurgery,
University Hospital Maastricht,
The Netherlands

M. R. Kerckhoffs
Department of Physiotherapy,
University Hospital Maastricht,
The Netherlands

J. W. S. Vlaeyen
Department of Clinical
and Experimental Psychology,
Maastricht University, The Netherlands

Abstract Behavioral approaches to treating patients following lumbar disc surgery are becoming increasingly popular. The treatment method is based on the assumption that pain and pain disability are not only influenced by somatic pathology, if found, but also by psychological and social factors. A recent study highlighted the effectiveness of cognitive-behavioral interventions, as compared to no treatment, for chronic low back patients. However, to the authors' knowledge, there is no randomized controlled trial that evaluates a behavioral program for patients following lumbar disc surgery. The purpose of this study was to assess the effectiveness of a behavioral graded activity (BGA) program compared to usual care (UC) in physiotherapy following first-time lumbar disc surgery. The BGA program was a patient-tailored intervention based upon operant therapy. The essence of the BGA is to teach patients that it is safe to increase activity levels. The study was designed as a randomized controlled trial. Assessments were carried out before and after treatment by an observer blinded to treatment allocation. Patients suffering residual symptoms restricting their activities of daily living and/or work at the 6 weeks post-surgery consultation by the neurosurgeon were included. The exclusion criteria were: complications during surgery, any relevant underlying pa-

thology, and any contraindication to physiotherapy or the BGA program. Primary outcome measures were the patient's Global Perceived Effect and the functional status. Secondary measures were: fear of movement, viewing pain as extremely threatening, pain, severity of the main complaint, range of motion, and relapses. Physiotherapists in the BGA group received proper training. Between November 1997 and December 1999, 105 patients were randomized; 53 into the UC group and 52 into the BGA group. The unadjusted analysis shows a 19.3% (95% CI: 0.1 to 38.5) statistically significant difference to the advantage of the UC group on Global Perceived Effect. This result, however, is not robust, as the adjusted analyses reveal a difference of 15.7% (95% CI: -3.9 to 35.2), which is not statistically significant. For all other outcome measures there were no statistically significant or clinically relevant differences between the two intervention groups. In general, the physiotherapists' compliance with the BGA program was satisfactory, although not all treatments, either in the BGA or the UC group, were delivered exactly as planned, resulting in less contrast between the two interventions than had been planned for. There was one re-operation in each group. The BGA program was not more effective than UC in patients following first-time lumbar disc surgery. For Global Per-

ceived Effect there was a borderline statistically significant difference to the advantage of the UC group. On functional status and all other outcome measures there were no rele-

vant differences between interventions. The number of re-operations was negligible, indicating that it is safe to exercise after first-time disc surgery.

Keywords Lumbar disc surgery · Rehabilitation · Randomized controlled trial · Physiotherapy · Behavioral treatment

Introduction

In the Netherlands 10,000 operations are performed each year for lumbosacral radicular syndrome, which is based on a herniated lumbar disc [21, 29]. An international comparison of back surgery rates showed quite large differences between countries [5]. The published figures on the results of surgery for a lumbar herniated disc, however, vary widely, with the percentage of patients reporting residual sciatica after surgery ranging from 22 to 45%, and the percentage reporting residual low back pain ranging from 30 to 70% [10, 31, 42, 43]. Other studies report success rates ranging from 60 to 90% [1, 14, 16, 19, 20, 26, 30]. Persisting symptoms mainly consist of pain, motor deficits, and a decreased functional status. In 2–19% of patients who undergo surgery, a recurrent herniated lumbar disc occurs, 74% occurring within 6 months after the first surgery [7, 36]. If patients still have these symptoms despite surgery, they are often referred to physiotherapy. The content of post-surgery treatment ranges from advice, through normal physical training, up to total rehabilitation programs [13, 22, 23, 30]. It is the practice in some hospitals to treat all patients immediately after surgery, but often treatment is reserved for patients who still have symptoms after some time. In the Netherlands many physiotherapists work according to a biomechanical model of disease, thereby assuming a causal relationship between tissue damage and pain. From this perspective, pain (or the reaction of patients to the prior treatment) is used as guidance to determine the intensity of recommended exercises and advice concerning activities of daily living (ADL). More recently, treatment is being guided by the principles of the biopsychosocial model, generally referred to as cognitive-behavioral therapy [35, 38]. The main assumption of a behavioral approach is that pain and pain disability are influenced not only by somatic pathology, if found, but also by psychological and social factors. A recent study highlighted the effectiveness of cognitive-behavioral interventions, as compared to no treatment, for chronic low back patients [37]. As far as we know there is no randomized controlled trial that evaluates a behavioral program for patients following lumbar disc surgery. In general, three behavioral treatment approaches can be distinguished: operant, cognitive and respondent therapies. Operant therapy is the most relevant therapy to be applied by physiotherapists. It aims to increase health behaviors using graded activity and positive reinforcement and to decrease pain behaviors and increase tolerance levels [11, 39]. Based on recent studies [37, 38],

we hypothesized that operant therapy alters fear of movement and other mediators, which would subsequently lead to an improved functional status and a higher rate of recovery. So the aim of this study was to assess whether this operant therapy is more effective than usual care following first-time lumbar disc surgery.

Materials and methods

If patients did not respond well to conservative treatment, surgery was considered. The main indication for surgery was radicular leg pain with conclusive imaging findings on magnetic resonance imaging or radiography. Six weeks after surgery, patients were scheduled for a routine appointment with the neurosurgeon. If symptoms persisted (severe leg or back pain, motor deficits, or restriction of their ADL and/or work), they were referred to physiotherapy and received oral and written information. The research assistant then provided further details about the study and re-evaluated their eligibility.

Inclusion criteria were:

- Age between 18 and 65 years
- First-time disc surgery (one level only)
- Complaints (e.g. residual leg or back pain) restricting ADL and/or work

Exclusion criteria were:

- Complications during surgery (loss of cerebrospinal fluid, nerve root lesion, blood loss of more than 600 ml) to be judged by the neurosurgeon
- Confirmed and relevant underlying diseases that influenced ADL (e.g., stenosis, malignancies, M. Bechterew, M. Scheuerman)
- Contraindication for one of the treatments (e.g. because of respiratory complaints)

If patients were eligible and willing to participate, an informed consent form was signed. The effectiveness of behavioral graded activity (BGA) in comparison to usual care (UC) following first-time lumbar disc surgery was assessed in a randomized controlled trial. An extensive description of the design, background and outcome measures is published elsewhere [24]. The Medical Ethics Committee of the University Hospital Maastricht approved the study protocol. By using opaque, sealed and coded randomization envelopes (prepared by an independent person according to computer-generated random tables), the outcome assessor (M.R.K.) was blinded. To assess the success of randomization, several important prognostic factors (Table 1) were measured at baseline, as well as scores for all outcome measures. At the end of treatment (3 months after randomization), all outcome measures were compared to detect post-treatment effects.

Interventions

Behavioral graded activity (BGA) is an operant therapy using graded activity and positive reinforcement in order to increase health behaviors and decrease pain behaviors [11, 39]. It is based upon time-contingency management, as described in more detail

by Fordyce et al. [11, 12] and applied by Lindström et al. [17]. The term “behavioral graded activity” for this program emphasizes the behavioral component, rather than merely physical training principles, and is described extensively elsewhere [24]. Primary care physiotherapists who attended a 2-day practical training course and two refresher meetings during the study provided the treatment. The essence of BGA was to establish individually graded exercise training, based on baseline measurements performed at intake, to teach patients that it is safe to move while increasing activity levels. During initial baseline measurements, patients were asked to perform activities (selected by the patients themselves) or exercises until they reached their (pain) tolerance, upon which patients set their own, individual, treatment goals. The next step was to set quotas (time contingent), which were systematically increased towards the pre-set goal. Quotas were not to be over-performed or under-performed. First quotas were slightly under baseline level, to ensure that patients’ initial experiences, while performing exercises, were successful, which enhances motivation: positive reinforcement is one of the key principles in the operant conditioning theory. In this way, a patient-tailored, individual BGA program was developed. Patients had to practise at home. Activities or exercises were to be documented on performance charts, which were discussed with the physiotherapist.

The content of usual care (UC) was determined after extensive interviews and discussions with the participating physiotherapists. In general, the whole spectrum of techniques used by physiotherapists within these patients were included which, in our opinion, is sensible when investigating usual care. There were only a few restrictions. Specific BGA components were not allowed, and therapies such as acupuncture, osteopathic techniques and all kinds of other “alternative” techniques were excluded. Both sets of treatment were individually based.

Physiotherapists in both sets of treatment (maximum of 18 30-min sessions within 3 months) documented every session on treatment registration forms. UC physiotherapists were allowed to stop treatment if patients no longer had complaints and treatment goals had been achieved, thus complying with usual care principles. Patients in the BGA treatment group had to complete the full program.

A priori we identified important features distinguishing the BGA from the UC treatment. First, BGA is based on systematically performed baseline measurements, whereas UC is based on anamnesis and physical examination. Second, BGA management is time contingent once quotas have been set, whereas UC evaluates reactions of patients to previous treatments and possibly adapts treatment intensity based on this evaluation, which is pain contingent. Third, in BGA, specific behavioral components were used: goal setting by patients, a performance chart, systematic appraisal for health behaviors and extinction of pain behaviors. To assess the extent to which the two treatment methods differed from one another in practice, three blinded experts scored a random sample of audiotapes recorded in a selection of patients from both treatment groups. All assessments were scored on a visual analog scale (VAS) and for purpose of clarity these scores were afterwards categorized into three-point scales (0–3.3; 3.4–6.7; 6.8–10). Firstly, experts rated their overall impression for every sound sample, marking on the VAS how close the treatment session matched up to an optimal BGA treatment session. Secondly, for every sound sample there was a quality assessment of the three aforementioned BGA characteristics taken separately: extinction of pain behavior, reinforcement of health behaviors and providing information about prognosis and symptoms from a biopsychosocial perspective. Three pre-recorded sound samples, in our opinion containing the optimal quality for the BGA characteristics, were also included, in order to evaluate the scoring system.

Prognostic factors and outcome measurements

Demographics and clinical information were recorded from patients’ records. At baseline, the duration of complaints, medication, previous treatments and kind of job were documented. Information was also sought, using three methods, on the extent to which patients believed in recovery from their symptoms and in therapy. Firstly, patients were asked at baseline about their level of confidence, in general (regardless of therapy), about their recovery (much, moderate, no confidence in recovery, don’t know). Secondly, patients’ expectations were measured according to Vlaeyen et al. [40]. After two treatment sessions, patients were asked to what extent they believed that the allocated treatment would be beneficial (ten-point Likert scale: 0=expects no benefit at all, 10=absolutely convinced of benefit). Thirdly, negative affectivity was measured by the Negative Emotionality (NEM) subscale (14 items, two-point scale) of the Multidimensional Personality Questionnaire ([33], Tellegen, University of Minnesota, unpublished manual, 1982), that quantifies negative affect (high NEM scores) of patients.

Primary outcome measures were: Global Perceived Effect (GPE) rated on a seven-point scale (1=completely recovered, 7=worse than ever). These ratings were dichotomized into improved (“completely recovered” and “much improved”) versus not improved (“slightly improved”, “not changed”, “slightly worsened”, “much worsened”, “worse than ever”). The Roland Disability Questionnaire (RDQ) [28] was used to measure low back specific functional status.

Secondary outcome measures were: fear of movement, measured using the Tampa Scale for Kinisophobia (TSK; Miller, Kori and Todd unpublished report, 1991); the Pain Catastrophizing Scale (PCS) [34], which measured catastrophizing (viewing pain as extremely threatening); and the intensity of low back pain or sciatica, which was scored on a VAS. (The relevance, validity and reliability of the VAS are commonly accepted in the area of low back pain [4, 27, 32].) In addition, at baseline, patients selected two important ADL activities that were severely hampered by their symptoms, in a standardized way. These were called main complaints (MC) [2], and severity was scored on a VAS. General health and social functioning were evaluated by using the corresponding subscales of the SF-36 [41]. Range of motion (ROM) (flexion, extension) of the lumbar spine was measured by the Cybex EDI-320, which has been shown to have a satisfactory reproducibility, especially for flexion [3, 6, 15]. Occurrences of re-operations were recorded.

Analysis

In a consensus meeting we examined the treatment registration forms for protocol deviations. In the BGA program protocol, deviations were defined as: use of passive treatment modalities, more than twice not fulfilling quotas, co-interventions by other health care providers (e.g., neurosurgeon or general practitioner). In the UC group, only significant co-interventions (e.g., ceasing treatment on the advice of the neurosurgeon) were recorded as protocol deviations. Statistical analyses were carried out according to the intention to treat principle. The cause of dropping-out determined the replacement procedure:

1. Patients were deleted from the analysis if there was no association with allocated treatment (e.g., patients moved out of the catchment area)
2. Patients received a negative score if they had more pain, or when the neurosurgeon advised them to stop treatment because of strong indications of a (new) herniated disc
3. Patients received a positive score if they returned to work completely or there were other indications that justified a positive score.

For substitution of negative or positive scores, we used the 10th or 90th percentile score of the total group. An expert panel, blinded for treatment allocation, assigned replacement values independently. If two out of three attributed the same substitution values, this value was used. In addition, a per-protocol analysis was performed that was restricted to patients who were compliant with the treatment protocol. For all analyses, SPSS 9.0 for Windows (SPSS Inc. North Michigan Avenue, Chicago, Ill.) was used. For outcome measures collected at baseline, the difference between the baseline and post-treatment values was calculated for each individual, and these change-scores were compared for the two treatment groups using Student's *t*-test for statistical significance. For outcome measures without baseline measurement (e.g., GPE), differences between groups at the post-treatment stage were analyzed. Group differences and two-tailed 95% confidence intervals were calculated for all outcome measures. In order to adjust for possible baseline differences, a multiple linear regression analysis for continuous outcome measures was performed with the change scores as dependent variable, treatment as independent variable, and baseline scores of the prognostic variables as co-variables. With regard to the audiotapes for assessing the contrast, first agreement between the three experts was calculated on the original VAS score by means of Pearson's correlation coefficient *r*. Then, for each characteristic, the percentage correctly classified was calculated.

Results

Between November 1997 and December 1999 a total of 671 patients were screened by nine neurosurgeons in the four participating hospitals in the south of the Netherlands. Of these patients, 382 (57%) suffered no substantial residual complaints. In total 105 patients (16%) suffered substantial residual complaints and were eligible for the present study, and signed informed consent forms. Figure 1 summarizes patient flow through the study.

In the UC group 70% underwent a standard discectomy, versus 78% in the BGA group. In the BGA group one pa-

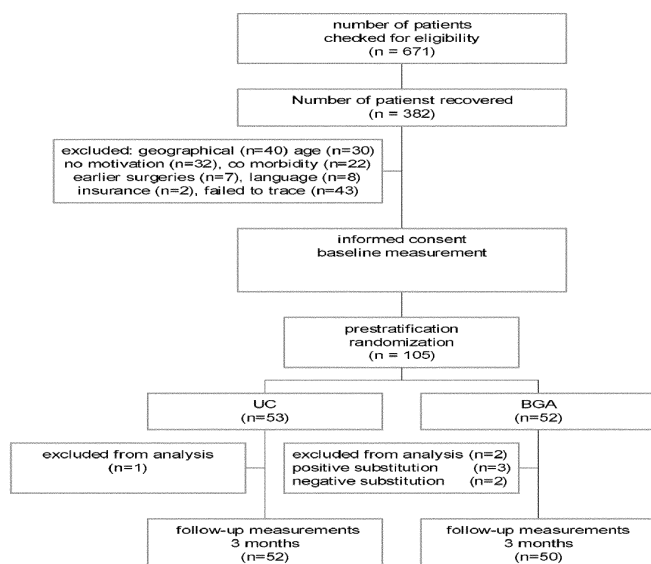


Fig. 1 Patient flow through the study

tient had a laminectomy and one a foraminectomy, while one patient in the UC group had a facetectomy. In the remaining patients neurosurgeons applied various combinations of surgical techniques (e.g., a standard discectomy in combination with a partial foraminectomy). There was one patient with complications during surgery in the UC group (nerve root lesion) and one in the BGA group (loss of cerebrospinal fluid). No included patients had a root block or bracing. After 2–3 days in bed, in order to recover from surgery, physiotherapists instructed patients with regard to low-back exercises and ADL functions. The distributions of baseline characteristics of both groups are presented in Table 1.

In the UC group there were on average 15.5 treatments, versus 14.8 in the BGA group. Eight patients dropped out: one in the UC group and seven in the BGA group. The UC patient disappeared after two sessions without stating any reason, and was therefore excluded from the analysis. Two BGA patients dropped out because of aggravated symptoms. Negative scores were therefore substituted for their values. One BGA patient had exacerbation of symptoms before the first session, and another patient had a flare-up of rheumatic disorders not reported before randomization. The reasons for these patients dropping out were considered independent of treatment and the patients were therefore excluded from further analysis. One patient reported himself completely pain free after two sessions and was no longer motivated to continue. Another patient reported withdrawing from the treatment because of lack of time and motivation following complete resumption of work (and no more symptoms). One patient withdrew because of personal circumstances, but had largely recovered after five treatment sessions. The values of these three patients were substituted by positive values.

Table 2 presents the results. In the UC group, 67% of the patients rated themselves as "recovered" on dichotomized Global Perceived Effect, versus 48% of the BGA patients. This 19.3% difference (95% CI: 0.1 to 38.5) to the advantage of the UC group is statistically significant. However, the adjusted analyses revealed only a 15.7% difference (95% CI: 3.9 to 35.2), which is no longer statistically significant. The RDQ scores improved significantly within both groups (5.6 points in the UC group vs 5.2 in the BGA), but there was no statistically significant or clinically relevant difference between the BGA and UC groups. This pattern was identical for the scores on the main complaint, pain (leg and back), range of motion and social functioning: statistically significant improvements were seen within groups but differences between groups were neither clinically relevant nor statistically significant. On the Pain Catastrophizing scale and the Tampa scale, as well as on the General Health and Social Functioning subscales of the SF-36, no substantial improvements were recorded.

The per-protocol analysis was restricted to the 78 patients who complied with their respective treatment protocols: 45 in the UC group and 33 in the BGA. The prog-

Table 1 Comparability of usual care (UC) and behavioral graded activity (BGA) treatment groups at baseline. Values are means (standard deviations) unless stated otherwise (RDQ Roland Disability Questionnaire, PCS Pain Catastrophizing Scale, TSK Tampa Scale for Kinisophobia)

Characteristics	UC (n=53)	BGA (n=52)
Age (years)	43.7 (8.8)	42.8 (8.8)
No. (%) of women	19 (35.8)	26 (50.0)
Type of surgery: standard discectomy (%)	70	78
Level of surgery (%)		
L4-L5	43.1	49.0
L5-S1	52.9	44.9
Surgical findings (%)		
Sequester	38.5	38.8
Protrusion	51.9	49.0
Days of hospitalization	7.7 (2.7)	7.0 (1.3)
Duration of these complaints before surgery: n (%)		
≤ 3 months	9 (17)	7 (13)
>3 months	44 (83)	45 (87)
Physiotherapy immediately after surgery (%)	33	42
No. (%) taking pain medication	18 (34)	20 (38)
No. (%) with physiotherapy before surgery	30 (56.6)	32 (60.4)
No. (%) with other treatment before surgery	16 (30.2)	13 (23.9)
No. (%) with paid jobs	37 (69.8)	47 (90.4)
Expectation of allocated treatment (0–10 points) ^a	6.85 (1.6)	6.9 (1.1)
Confidence with regard to recovery in general: n (%)		
Much confidence in recovery	22 (41.5)	21 (40.4)
Moderate confidence in recovery	18 (34.0)	24 (46.2)
No confidence in recovery	3 (5.7)	3 (5.8)
Don't know	10 (18.9)	4 (7.7)
Negative affectivity (0–14 points)	3.7 (4.0)	4.2 (4.1)
Outcome measures		
RDQ (0–24 points)	13.5 (4.5)	14.5 (3.7)
TSK (17–68 points)	36.9 (6.8)	35.9 (6.3)
PCS (0–52 points)	16.9 (11.7)	17.1 (10.2)
Main complaint (0–100)	67.4 (15.4)	71.1 (16.5)
No. (%) of patients with pain in back	45 (85)	48 (92)
Severity of pain in back (0–100)	46.7 (27.3)	43.40 (30.0)
No. (%) of patients with sciatica	52 (98)	50 (96)
Severity of sciatica (0–100)	41.3 (30.8)	39.0 (28.2)
Range of motion	81.2 (22.7)	78.1 (22.6)
General Health (Subscale SF-36)	65.6 (20.0)	68.2 (18.4)
Social Functioning (Subscale SF-36)	59.4 (25.3)	56.7 (26.6)

^a Administered after two treatments

nostic comparability between intervention groups that qualified for per-protocol analysis was quite similar to that for all the study subjects, as shown in Table 1. In general, restricting the analysis to patients who complied with their respective treatment protocols resulted in slightly larger improvements within groups, but between-group differences did not change in any substantial way.

Integrity check

The final master tape consisted of 24 sound samples (13 BGA samples and 11 UC samples) plus the three “gold standard” samples. The Pearson’s *r* for the assessments was

on average 0.65, with a range from 0.55 to 0.82 on the various characteristics. Overall, the agreement was satisfactory. Furthermore, the three gold standard samples were in all cases scored as expected, thereby establishing a certain extent of validity with regard to our scoring system. Calculating the percentages of the scores on the three-point scales for both treatment conditions separately showed that, on average, 70–80% of the sound samples were scored in expected categories on the various quality assessments and 20–30% were not classified in the expected direction (e.g., UC samples were scored as “no usual care characteristics”, or identified BGA characteristics were rated as “poor quality” while UC samples were rated as “good quality” with regard to BGA characteristics).

Table 2 Results at post-treatment. All outcome measures are presented as mean values and standard deviations, unless otherwise stated. Differences are presented as mean values with 95% confidence intervals, unless otherwise stated

	UC Improvement (<i>n</i> =52)	BGA Improvement (<i>n</i> =50)	Difference between groups BGA – UC	Adjusted ^e difference between groups BGA – UC
Global Perceived Effect: <i>n</i> (%) ^a	35 (67%)	24 (48%)	-19.3% [-38.5%; -0.1%]	-15.7% [-35.2%; 3.9%]
Roland Disability Questionnaire (0–24 points) ^b	-5.6 (5.3)	-5.2 (5.9)	0.4 [-1.8; 2.6]	0.95 [-1.4; 3.3]
Pain Catastrophizing Scale (0–52 points) ^b	-1.9 (9.5)	-2.9 (8.0)	-1.0 [-4.4; 2.5]	-0.8 [-4.2; 2.7]
Tampa Scale (17–68 points) ^b	-0.7 (6.8)	-0.7 (6.7)	0.0 [-2.7; 2.6]	-0.8 [-3.5; 1.8]
Main Complaint (0–100) ^b	-30.3 (25.2)	-32.3 (27.7)	-2.1 [-12.5; 8.3]	-0.2 ^f [-11.1; 10.6]
Severity of pain in back (0–100) ^b	-16.0 (25.3)	-9.3 (27.8)	6.6 [-3.8; 17.1]	5.0 [-4.9; 14.9]
Severity of sciatica (0–100) ^b	-14.2 (23.9)	-11.6 (31.3)	2.6 [-8.3; 13.5]	1.8 [8.2; -11.9]
Range of Motion	12.7 (22.0) ^c	12.7 (19.6) ^d	0.1 [-8.5; 8.7]	3.6 [-10.7; 3.4]
General Health (Subscale SF-36)	2.9 (14.2)	0.7 (12.5)	2.2 [-7.5; 3.1]	1.8 [-3.5; 7.2]
Social Functioning (Subscale SF-36)	18.3 (20.2)	19.5 (27.5)	1.2 [-10.7; 8.2]	0.7 [-7.2; 8.8]

^a Ratings of Global Perceived Effect on a seven-point scale were dichotomized (see Materials and Methods section)

^b Negative values denote positive results for patients

^c Nine data entries missing due to technical problems

^d One data entry missing due to technical problems

^e Adjusted for the following baseline variables: RDQ, duration of complaints, confidence in recovery, TSK, NEM, PEM (Negative/Positive Emotionality subscale of the Multidimensional Personality Questionnaire), back pain and PCS

^f Also adjusted for main complaint at baseline measurement

Discussion

In a single-blind randomized controlled trial the effectiveness of a behavioral graded activity (BGA) program compared to usual care (UC) was assessed. On the Global Perceived Effect the UC performed statistically significantly better than the BGA (19.3% difference, 95% CI: 0.1 to 38.5). However, the adjusted analyses reveal a 15.7% difference (95% CI: 3.9 to 35.2), which is no longer statistically significant.

All other outcome measures did not reveal any statistically significant nor clinically relevant differences between the two groups. Although there were more drop-outs from the BGA group, per-protocol analysis produced results similar to the intention to treat analysis. Adjustment for baseline characteristics did not alter results substantially, except for the GPE. Only one re-operation occurred in each group. Analysis of the audiotapes showed that experts classified 70–80% of the samples correctly.

The results regarding the GPE were dependent on whether the analyses were performed unadjusted or adjusted, indicating that the results for the GPE are not robust. One reason for this could be a lack of power, but for all other outcome measures there were no substantial differences between the different methods of analysis. Moreover, there were no clinically relevant differences between the two groups. Although there were no differences between the two groups for RDQ scores, the within-group improvement was considerable (5.2 for the BGA group and 5.6 for the UC). Because we lacked a no treatment control group, we are not able to say whether this improvement was attributable to either treatment (BGA or UC) or simply to the passing of time. The reason for not

including a control group was that it was considered inappropriate to withhold treatment from patients suffering residual symptoms 6 weeks post-surgery. A standard prescription for physiotherapy for these patients is considered usual care. Although there was considerable improvement in both groups, post-treatment RDQ scores remain high (7.9 in the UC group and 9.3 in the BGA group). Also, patients' main complaint improved in both groups, falling to almost half its pre-treatment value on the VAS, but complaints remained post-treatment. One reason for these high scores, irrespective of treatment, may be that this study was restricted to patients with substantial residual complaints 6 weeks post-surgery. As Fig. 1 shows, this was only 16% of all patients consulting the neurosurgeon for the routine 6 weeks post-surgery consultation. Of these 16%, the majority suffered, among other complaints, from low back pain. This is in line with results in the literature [10, 31, 42, 43]. The residual complaints of these patients may also account for longer duration (mean 7 days) of post-surgery hospitalization in this specific population. Whether these results hold true for *all* patients following lumbar disc surgery can not be concluded from this study.

Drop-outs were unevenly distributed between the two groups: only one UC patient versus seven BGA patients. However, there are no indications that this biased the results, as only two BGA patients suffered exacerbation. Moreover, three patients clearly showed improvements. Although it can be argued that the information about reasons for drop-out is subjective, it is more sensible to replace missing values according to reasons of drop-out than to assign mean values, because, in our opinion, bias of drop-out is better anticipated in this way.

A priori we hypothesized that the BGA treatment would reduce catastrophizing and fear avoidance (mediators), which would lead to an improvement in patients' overall rating of recovery and an improvement in their ADL. However, the stability of both catastrophizing and fear avoidance is remarkable, because these mediators were significantly altered neither in the UC nor in the BGA. Several reasons for these unexpected findings can be posited. Firstly, interventions may not be delivered as expected. Secondly, existing literature may have been too optimistically interpreted. Thirdly, this specific population may benefit from behavioral approaches.

Despite the 2-day training course and refresher meetings, the BGA protocol may still not have been delivered as planned. Changing the behavior of caregivers may be as difficult as changing the behaviors of patients. However, 70–80% of the sound samples were classified correctly, meaning that there was an overlap of 20–30%, resulting in less contrast between the two protocols than we had planned and hoped for. Furthermore, these results show that the UC sessions incorporated specific BGA characteristics (e.g. reinforcement of health behaviors) and that BGA therapists continue to apply some characteristics of the UC treatment. This is in concordance with a survey that showed that even if physiotherapists are trained in biopsychosocial approaches, attitudes and beliefs concerning details about specific treatment issues differ widely [25]. As 70–80% of sound samples were classified correctly, and the BGA treatment did not show the slightest sign of being more effective than UC, we do not think that this overlap concealed any possible effect of BGA. In conclusion, there are no reasons to apply BGA in primary care for patients following first-time lumbar disc surgery.

In a recent study, Danielsen et al. [8] concluded that vigorous medical exercise therapy started 4 weeks after surgery for lumbar disc herniation reduced disability and pain after surgery, and that there was hardly any danger associated with early and vigorous training. This was in line with a previous study which concluded that high-intensity training started 5 weeks after surgery had favorable effects [18]. This was mainly based on the finding that the high-intensity training group had no more exacerbations or re-operations than the mild exercise group, and both groups had a comparable improvement with regard to disability. Dolan et al. [9] concluded that exercise therapy improved the outcome within the exercise group, but no between-group results were presented. Our results regarding outcome on physical measures and disability also showed improvement over time. However, there are differences between our study and the others. First, the results of two of the aforementioned studies [8, 9] include the outcome of surgery, because the follow-up measures of these studies are compared with the pre-surgery levels of these outcome measures. Furthermore, these studies did not incorporate specific behavioral interventions. Another difference is that the current study was restricted to patients with residual complaints 6 weeks post-surgery. Therefore it can be argued that the more severe cases entered our trial. These differences make a direct comparison of the results of the studies by Danielsen et al. and Dolan et al. with our trial difficult. However, in line with these other studies we also conclude that activity after first-time lumbar disc surgery is safe, as the re-operation rate was very low, and therefore it is not necessary for patients to remain passive after surgery.

References

1. Barrios C, Ahmed M, Arroategui JJ, et al (1990) Clinical factors predicting outcome after surgery for herniated lumbar disc: an epidemiological multivariate analysis. *J Spinal Disord* 3: 205–209
2. Beurskens AJ, Vet de HC, Köke AJ, et al (1999) A patient-specific approach for measuring functional status in low back pain. *J Manipulative Physiol Ther* 22:144–148
3. Bø K, Hilde G, Storheim K (1997) Intra- and interobserver reproducibility of Cybex EDI-320 measuring spinal mobility. *Scand J Med Sci Sports* 7:140–143
4. Carlsson A M (1983) Assessment of chronic pain. 1. Aspects of the reliability and validity of the visual analogue scale. *Pain* 16:87–101
5. Cherkin DC, Deyo RA, Loeser JD, et al (1994) An international comparison of back surgery rates. *Spine* 19:1201–1206
6. Chiarello CM, Savidge R (1993) Interrater reliability of the Cybex EDI-320 and fluid goniometer in normals and patients with low back pain. *Arch Phys Med Rehabil* 74:32–37
7. Connolly ES (1993) Management of persistent or recurrent symptoms and signs in the postoperative lumbar disc patient. *Neurosurg Clin N Am* 4:161–166
8. Danielsen JM, Johnsen R, Kibsgaard SK, et al (2000) Early aggressive exercise for postoperative rehabilitation after discectomy. *Spine* 25:1201–1206
9. Dolan P, Greenfield K, Nelson RJ, et al (2000) Can exercise therapy improve the outcome of microdiscectomy? *Spine* 25:1523–1532
10. Dvorak J, Valach L, Fuhriemann P, et al (1988) The outcome of surgery for lumbar disc herniation. A 4–17 years' follow-up with emphasis on psychosocial aspects. *Spine* 13:1423–1427
11. Fordyce WE (1976) Behavioral methods for chronic pain and illness. Mosby, St Louis
12. Fordyce WE, Fowler RS, Lehmann JF, et al (1973) Operant conditioning in the treatment of chronic pain. *Arch Phys Med Rehabil* 5:399–408
13. Hansen JW (1964) Postoperative management in lumbar disc protrusions. *Acta Orthop Scand* 71 [Suppl]:1–47
14. Hurme M, Alaranta M (1987) Factors predicting the result of surgery for lumbar intervertebral disc herniation. *Spine* 12:933–938
15. Koes B, Mameren v H, Bouter L, et al (1989) The reproducibility of measuring spinal mobility with the EDI-320 (in Dutch). *Neder Tijdschr Fysiother* 99:207–211
16. Korres DS, Loupassis G, Stamos K (1992) Results of lumbar discectomy: a study using 15 different evaluation methods. *Eur Spine J* 1:20–24

17. Lindström I, Öhlund C, Eek v C, et al (1992) The effect of graded activity on patients with subacute low back pain. *Phys Ther* 72:279–293
18. Manniche C, Skall HF, Braendholt L, et al (1993) Clinical trial of postoperative dynamic back exercises after first lumbar discectomy. *Spine* 18:92–97
19. Manniche C, Asmussen KH, Vinterberg H, et al (1994) Analysis of preoperative prognostic factors in first-time surgery for lumbar disc herniation, including Finneson's and modified Spengler's score systems. *Dan Med Bull* 41:110–115
20. Manniche C, Asmussen KH, Vinterberg H, et al (1994) Back pain, sciatica and disability following first-time conventional hemilaminectomy for lumbar disc herniation. *Dan Med Bull* 41:103–106
21. Netherlands' Health Council (1999) Management of the lumbosacral radicular syndrome (sciatica) 1999/18. Netherlands Health Council, The Hague
22. Noordzij CM, Dolder v R (1991) Physiotherapy for patients following lumbar disc surgery (in Dutch). *Neder Tijdschr Fysiother* 101:160–162
23. Oerlemans H (1988) Physiotherapy following lumbar disc surgery, (in Dutch). *Neder Tijdschr Fysiother* 98: 245–251
24. Ostelo RW, Köke AJ, Beurskens AJ, et al (2000) Behavioral-graded activity compared with usual care after first-time disk surgery: considerations of a design of a RCT. *J Manipulative Physiol Ther* 23:312–319
25. Ostelo RW, Stomp-van den Berg SG, Vlaeyen JW, et al (2003) Healthcare providers' attitudes and beliefs regarding chronic low back pain. *Manual Therapy* (in press)
26. Pappas CT, Harrington ET, Sonntag VK (1992) Outcome analysis in 654 surgically treated lumbar disc herniations. *Neurosurgery* 30:862–866
27. Revill SI, Robinson JO, Rosen M, et al (1976) The reliability of a linear analogue for evaluating pain. *Anaesthesia* 31:1191–1198
28. Roland M, Morris R (1983) A study of the natural history of back pain. 1. Development of a reliable and sensitive measure of disability in low-back pain. *Spine* 8:141–144
29. Slebus FG, Maas vd PJ, Braakman R, et al (1989) The effectiveness of imaging diagnostics and surgery of the lumbar disc herniation (in Dutch). Rapport Ministerie van WVC
30. Smulders WL, Loon v G (1993) Treatment following a lumbar disc surgery or chemonucleolyse. An 'out-patient' survey. *Fysiopraxis* 2:4–6
31. Spangfort EV (1972) The lumbar disc herniation. A computer-aided analysis of 2,504 operations. *Acta Orthop Scand Suppl* 95:1421–1495
32. Sriwatanakul K, Kelvie W, Lasagna L, et al (1983) Studies with different types of visual analog scales for measurement of pain. *Clin Pharmacol Ther* 34:234–239
33. Stegen K, Neujens A, Crombez G, et al (1998) Negative affect, respiratory reactivity, and somatic complaints in a CO₂ enriched air inhalation paradigm. *Biol Psychol* 49:109–122
34. Sullivan MJ, Bishop SR, Pivik J (1995) The Pain Catastrophizing Scale: development and validation. *Psychol Assess* 7:524–532
35. Turk DC (1984) Etiological theories and treatments for chronic back pain. 2. Psychological models and interventions. *Pain* 19:209–233
36. Van Burken MM, Sanders EA (1992) Positive results of re-operation for recurrent lumbar disc herniation following a first-time lumbar disc surgery (in Dutch). *Ned Tijdschr Geneesk* 136: 692–695
37. Van Tulder MW, Ostelo RW, Vlaeyen JW, et al (2001) Behavioral treatment for chronic low back pain: a systematic review within the framework of the Cochrane Back Review Group. *Spine* 26:270–281
38. Vlaeyen JW, Linton SJ (2000) Fear-avoidance and its consequences in chronic musculoskeletal pain: a state of the art. *Pain* 85:317–332
39. Vlaeyen JW, Haazen IW, Schuerman JA, et al (1995) Behavioral rehabilitation of chronic low back pain. *Br J Clin Psychol* 34:95–118
40. Vlaeyen JW, Kole-Snijders AM, van Eek H (1996) Chronic pain and rehabilitation (in Dutch). Bohn Stafleu Van Loghum, Hoensbroek
41. Ware JE Jr, Sherbourne CD (1992) The MOS 36-item short-form health survey (SF-36). 1. Conceptual framework and item selection. *Med Care* 30: 473–483
42. Weber H (1983) Lumbar disc herniation. A controlled, prospective study with ten years of observation. *Spine* 8: 131–140
43. Yorimitsu E, Chiba K, Toyama Y, et al (2001) Long-term outcomes of standard discectomy for lumbar disc herniation: a follow-up study of more than 10 years. *Spine* 26:652–657