BMC
Musculoskeletal Disorders

STUDY PROTOCOL

Open Access

Design of the Verbiest trial: cost-effectiveness of surgery versus prolonged conservative treatment in patients with lumbar stenosis

Gijsbert M Overdevest[1,2*], Pim AJ Luijsterburg[3], Ronald Brand[4], Bart W Koes[3], Sita MA Bierma-Zeinstra[3,5],
Just AH Eekhof[6], Carmen LAM Vleggeert-Lankamp[1], Wilco C Peul[1,2]

Abstract

Background: Degenerative changes of lumbar spine anatomy resulting in the encroachment of neural structures are often regarded progressive, ultimately necessitating decompressive surgery. However the natural course is not necessarily progressive and the efficacy of a variety of nonsurgical interventions has also been described. At present there is insufficient data to compare surgical and nonsurgical interventions in terms of their relative benefit and safety. Previous attempts failed to provide clear clinical recommendations or to distinguish subgroups that substantially benefit from a certain treatment strategy. We present the design of a randomized controlled trial on (cost-) effectiveness of surgical decompression versus prolonged conservative treatment in patients with neurogenic intermittent claudication caused by lumbar stenosis.

Methods/Design: The aim of the Verbiest trial is to evaluate the effectiveness of prolonged conservative treatment compared to decompressive surgery. The study is a multi-center randomized controlled trial with two parallel groups design. Patients (age over 50) presenting to the neurologist or neurosurgeon with at least 3 months complaints of neurogenic intermittent claudication and considering surgical treatment are eligible for inclusion. Participants are randomly allocated to either prolonged conservative treatment, receiving further treatment from their general practitioner and physical therapist, or allocated to surgery and operated within 4 weeks. Primary outcome measure is the functional assessment of the patient as measured by the Zurich Claudication Questionnaire at 24 months of follow-up. Data is analyzed according to the intention to treat principle.

Discussion: With a cost-effectiveness analysis the trade off between the costs of prolonged conservative treatment and delayed surgery in a smaller number of patients are compared with the current policy of surgical management. As surgery is expected to be inevitable in certain subgroups of patients, the distinction of and classification by predictive patient characteristics is most relevant to clinical practice.

Trial registration: Netherlands Trial Register (NTR): NTR2216

Background

Lumbar stenosis is defined as a reduction in the diameter of the spinal canal, lateral nerve canals, or neural foramina. Most frequently lumbar stenosis is the result of a degenerative disease process and consequently involves multiple levels of the lumbar spine or sites of lumbar stenosis. Degenerative lumbar stenosis is caused by bone hypertrophy, osteoarthritis of facet joints, ligamentous hypertrophy, disc protrusion, spondylolisthesis alone or any combination of these elements. Central canal stenosis results in compression of the cauda equina. Lateral recess stenosis and neuro foraminal stenosis result in encroachment of the nerve root. Degenerative spinal stenosis most commonly affects the L3-L4, and L4-L5 segments to cause cauda equina compression [1,2].

Symptoms related to lumbar stenosis range from numbness, fatigue to actual pain of the buttocks, thighs, and legs. Back pain and muscle weakness are also

* Correspondence: g.m.overdevest@lumc.nl
[1]Department of Neurosurgery, Leiden University Medical Center,
Albinusdreef 2, 2333 ZA Leiden, The Netherlands
Full list of author information is available at the end of the article

frequently reported. Symptoms may radiate from the buttocks to the distal lower extremities and are often accompanied with paraesthesias. In contrast to sciatica symptoms are generally bilateral and localized poorly [2]. A pathognomonic aspect of lumbar stenosis is the relation between symptoms and function. Symptoms are typically aggravated by standing and walking. However symptoms generally decrease with sitting or standing with lumbar flexion and with lying. As symptoms worsen patients become progressively limited in their activities and walking distance. This relationship is known as neurogenic intermittent claudication [3]. The clinical entity was first described by the Dutch neurosurgeon Verbiest in 1950, thus formerly also referred to as the Verbiest syndrome [4].

Degenerative lumbar stenosis is the most common reason for lumbar surgery in elderly people [5]. Taking into account that the mean age of the total population of the Netherlands and other Western countries will increase over the upcoming years and due to an increasing life expectancy in general, surgery rates are expected to increase even further. However the aging of the population does not explain the total increase of lumbar surgery rates observed over the past decades. Also there are significant geographic variations in spine surgery rates [5-7]. Despite extensive research in this area, various authors suggest a poor consensus on indications for surgery or the choice of particular surgical procedures. Patient demographic and clinical characteristics may be outweighed by surgeons and patients preferences concerning treatment allocation [5-8].

Although the condition is frequently diagnosed and various surgical and nonsurgical interventions for individuals with neurogenic intermittent claudication are widely accepted in clinical practice, there is limited evidence to support many of them, especially in terms of their relative benefit and risk compared with other options [9-11]. The long-term results of surgical management of spinal stenosis are good or excellent in 50-79% of patients [9]. Most cases of spinal stenosis can be treated with surgical decompression alone. In a minority of cases spinal stenosis is accompanied with a spinal deformity such as severe spondylolisthesis or scoliosis. In these instances the addition of spinal fusion to decompression may be necessary to maintain spinal stability and effectively decompress neural structures [9]. The available evidence suggests no advantage in routinely applying fusion techniques [11].

In a randomized controlled trial (RCT) comparing the effectiveness of early surgery to prolonged conservative treatment among patients with sciatica, early surgery favored a better short-term outcome compared to prolonged conservative management [12]. However, there was no significant nor clinically relevant difference

between treatment groups after 1 year. It is well known that the natural history of sciatica caused by lumbar disk herniation is favorable in most cases [13]. Spontaneous involution of herniated disk material can also be observed with imaging studies [14]. By contrast lumbar stenosis is caused by long-term degenerative changes affecting the structures surrounding the cauda equine or nerve roots. Unlike disc herniation these degenerative changes are not likely to resolve spontaneously. However the extent of narrowing of the central or lateral canal is not necessarily associated with symptom severity [2]. In addition, imaging studies demonstrate radiological significant lumbar stenosis can be found in up to 20 percent of asymptomatic elderly individuals [15,16].

Various studies observing the natural course of neurogenic intermittent claudication report improvement of symptoms and conclude that expectant observation could be an alternative to surgical treatment [9,17-21]. In a review by Watters et al. the effectiveness of nonsurgical therapy is discussed [9]. Favorable outcomes are achieved in approximately 70% of patients with moderate symptoms and in 33% with severe symptoms. Approximately 20-40% of patients with moderate symptoms who are treated conservatively require surgery in the long-term (2-10 years). However, overall observational results of surgical treatment consistently demonstrate better short-term and long-term results when compared to conservative therapy. Consequently, and despite the apparent superiority of surgical treatment, nonsurgical treatment can be an effective therapy in selected patients. Furthermore surgical intervention is equally effective when performed after failure of conservative treatment compared to primary surgical management [17].

Limited evidence is available regarding the efficacy for nonsurgical treatments in the treatment of lumbar stenosis [9,10]. Nonsurgical treatments have most frequently been compared to other treatments, rather than to the natural course of lumbar stenosis. Currently the choice of treatment is guided mainly by clinical judgment, observational studies and in analogy to other back conditions. A systematic review assessing the effectiveness of nonsurgical treatment of nonspecific low back pain suggests efficacy for a variety of nonsurgical interventions [10]. Treatment should focus on patient education, medications to control pain and exercise therapy [10,22].

Two previous RCTs comparing the effectiveness of surgery to nonsurgical treatment of degenerative lumbar stenosis have been performed [23,24]. Weinstein et al. reported favorable results for surgery throughout the 2 year follow-up period [23]. However the design of the trial was compromised due to large cross-over rates among both treatment groups, at least if one wants to estimate the true magnitude of the effect of surgery compared to nonsurgical treatment. Cross-over to some extent is to be

expected, as a subgroup of patients will have complaints refractory to nonsurgical management. However there was also considerable nonadherence among subjects allocated to surgery; only 67% of the surgical group underwent surgery during the two years of follow-up and non-adherence was even larger at earlier time points (54% at 6 months and 42% at 3 months). Possibly subjects reported significant symptom relief before surgery had been scheduled. More likely, however, these subjects did not have symptoms severe enough to justify surgical treatment or did not yet consider surgery as treatment. To address treatment nonadherence conclusions were based on an as treated analysis rather than an intention to treat analysis. These results must be interpreted cautiously as confounding is no longer controlled by random allocation. Although the analysis was corrected for known confounding factors, baseline comparability for other factors can not be assured. Further, the validity of conclusions is limited due to a patient selection bias since post-randomization eligibility for surgery in the initially conservative group and vice versa is clearly correlated to the condition of the patient prior to the cross-over. Malmivaara et al. and also reported favorable results for surgery as compared to conservative treatment [24]. Based on a previously defined minimal clinically imported difference (MCID) with regard to functional disability 63% of surgically managed subjects reported improvement compared to 30% of conservatively managed subjects [25]. In concordance with previous studies the initial advantage of surgery over conservative management was followed by a narrowing of the relative benefit over time [18,26]. Apart from the relatively small sample size, only subjects with moderate symptom severity were included. Final conclusions were further limited by the heterogeneity of treatments as surgical decompression was accompanied by fusion techniques in 20% of the subjects allocated to surgery. The distinction of surgical candidates in both RCT deserves scrutiny, especially since the generalizability of the findings depends on the ability of the study design to reflect the current clinical challenge.

At present there is still a paucity of high quality evidence to justify the timing of surgery and useful clinical parameters to identify subjects for the most appropriate treatment strategy are lacking. The primary goal of this study is to assess the (cost-) effectiveness of a policy of conventional surgical intervention as compared to prolonged conservative management in the treatment of degenerative lumbar stenosis. Further, patient characteristics and clinical findings are evaluated to define subgroups that substantially benefit from one of the two proposed treatment strategies.

Methods and design

The study is designed as a multi-center RCT with two parallel groups. The multi-center design is necessary to accrue the required amount of patients and obtain generalizable results. The trial is conducted in the hospitals, which previously collaborated in the Sciatica trial [12]. The medical research ethics committee of each hospital has approved the trial. If inclusion and exclusion criteria are met, patients are asked to participate in the proposed trial. Patients are randomized if the diagnosis lumbar stenosis is confirmed by imaging findings and an indication for surgery is confirmed by the neurosurgeon. Participants are allocated to either prolonged conservative treatment or surgery. As national consensus guidelines for the nonsurgical treatment of lumbar stenosis are lacking a clinical treatment protocol was developed for this trial. The purpose of such a protocol is to provide comparable nonsurgical treatment throughout the participants and aid adherence to treatment allocation. Surgery is scheduled within 4 weeks after randomization. As the clinical investigator and the research nurse offer counseling and evaluate possible negative health effects of treatment allocation, they cannot be blinded for randomization outcome. Throughout the study period participants fill out questionnaires and physical examination findings are recorded during consecutive visits to the outpatient clinic. The follow-up period will last 5 years. After six months secondary surgery is offered in case prolonged conservative treatment does not result in satisfactory symptom relief or functional improvement.

Participants and recruitment

Patients presenting to the neurologist or neurosurgeon in the participating hospitals with at least 3 months complaints of neurogenic intermittent claudication and considering surgical treatment are eligible for inclusion. The duration of these complaints is one of the main (a priori defined) factors to be evaluated for a possible treatment-effect-modification. Inclusion and exclusion criteria are listed in Table 1.

General practitioners and physical therapists in the vicinity of the participating hospitals have been informed at the start of the trial. They are directly involved as they provide primary care for subjects allocated to prolonged conservative treatment. Secondly, general practitioners are asked to refer potentially eligible patients with neurogenic intermittent claudication to the outpatient clinics for evaluation by the neurologist.

Prolonged conservative management

Conservative management is a prolonged nonsurgical treatment policy conducted by the general practitioner. The general practitioner provides additional information about the condition's causes, symptoms and treatment options. Further, the efficacy of the prescribed pain medication is reviewed and if necessary alterations are made, taking into account possible contra-indications in the

Table 1 Inclusion and exclusion criteria

Inclusion criteria	
	- Age ≥50 years old
	- ≥3 Months neurogenic intermittent claudication, as noted by leg/buttock/groin pain with or without back pain or fatigue in the legs provoked by walking. Leg/buttock/groin pain or fatigue needs to be relieved by lumbar flexion
	- Narrowed lumbar spinal canal, nerve root canal or intervertebral foramen at one or more levels confirmed by MRI
	- Regular indication for surgical decompression
	- Signed informed consent
Exclusion criteria	
	- Cauda equine syndrome
	- Paget's disease, severe osteoporosis or metastasis to the vertebrae
	- Significant scoliosis (Cobb angle > 25 degrees)
	- Previously laminectomy at the same level, has degenerative or lytic spondylolisthesis ≥ grade 2 (on a scale 1 to 4) at the affected level or significant instability of the lumbar spine
	- Severe comorbid conditions that increase the risk to the patient or interfere with the evaluability of this study (e.g. severe ischemic heart disease, musculoskeletal or neurological conditions impairing walking ability, cognitive impairment (MMSE <25 points)
	- Unable to read or write Dutch

elderly. Subjects are advised to stay active and if possible return to work and/or their leisure activities. The general practitioner prescribes physical therapy, which consists of active exercises to guide the patient in upgrading his or her activities according to an agreed time schedule. The guide is time, not the intensity of the pain. The protocol for the individual exercise therapy supervised by a physical therapist includes education, stretching, strengthening and conditioning exercises [27]. Education in proper posture, body mechanics for daily activities, and if necessary, the use of orthopedic aids is essential to maintain the gains made through the physical therapy sessions. The goal of stretching and strengthening is to decrease the extension forces on the lumbar spine attributable to agonist muscle tightness, antagonist weakness or both, which results in a decreased lumbar lordosis. Stretching exercises include hip flexor stretching, hamstring stretching and lumbar paraspinal stretching. The strengthening exercises primarily target the abdominal and lower extremity muscles. The conditioning exercises include walking on a treadmill, on level and inclined surfaces and/or riding a stationary bicycle, with the goal of increasing general fitness. A maximum of 9 sessions is allowed in the first 3 months. Three additional booster sessions take place in the fourth, fifth, and sixth month [10]. A physical therapy session in

primary care lasts about 30 minutes. The exercise therapy is discontinued if, according to the physical therapist, treatment goals have been achieved.

Surgical management

After induction of general endotracheal anaesthesia the patient is positioned prone or in knee-chest position. The level on which to operate is determined through anatomical landmarks and confirmed by fluoroscopy. After a midline posterior skin and subcutaneous tissue incision centered at the interspinous level to be decompressed, the dissection goes through the dorsolumbar fascia. The multifidi are detached from both sides of the spinous processes and laminar attachments. Paraspinal muscles are than retracted laterally. The interspinous ligament is resected to enable the removal of the inferior aspect of the cranial lamina and the superior aspect of the caudal lamina. Subsequent a flavectomy is performed. The decompression is extended laterally with an undercutting of the medial aspect of the facet joints and if necessary an extensive foraminotomy is performed. In case of severe bony stenosis a complete laminectomy is performed. No transpedicular fixation of the spine is performed during the first surgery. Hospital admission is 2 to 4 days, including the day of surgery. During the post-operative period the patients are mobilized accompanied by a physical therapist, starting 3 hours after surgery. If the patient regains the physical ability to manage basis activities of daily living, the patient is discharged. At home the patients visit their own physical therapist who guides them and provides exercises. If possible patients gradually resume their daily activities and work if applicable. The patient is allowed to resume all activities in case no physical limitations are present.

Study measurements

Clinical data is collected during consecutive follow-up visits to the outpatient clinics at baseline and at 12, 26, 52, 104, 260 weeks after randomization. Subjects fill in questionnaires at selected time points which are returned during the outpatient clinic visits or by mail (Table 2). Demographic data, physical examination findings information on pain intensity, imaging findings, illness related disability, societal costs and utilities and quality of life are collected. Further, subjects keep a diary on the financial consequences and health care consumption related to lumbar stenosis. Visits to the general practitioner, physical therapist, medical specialists, alternative health practice, use of analgesics, duration of sick leave from work and additional utility/mobility costs are recorded. The patient diaries are returned to the research nurse during the outpatient clinic visits or by mail.

Table 2 Study measurements over time

Time in weeks	-2	0	4	12	26	38	52	104	156	208	260
Inpatient visit: surgery			x								
Outpatient visit	x	x		x	x		x	x			x
Questionnaire		x	x	x	x	x	x	x	x	x	x
Patient demographics	x										
MMSE	x										
Basic physical examination	x										
Neurological examination	x		x	x			x	x			x
X-ray	x		x								
MRI	x										
ZCQ		x	x	x	x	x	x	x	x	x	x
Shuttle Walking Test	x			x			x	x			x
Accelerometry	x			x			x	x			x
SPPB	x			x	x		x	x			x
RMDQ		x	x	x	x	x	x	x	x	x	x
SF-36		x	x	x	x	x	x	x	x	x	x
Perceived Recovery			x	x	x	x	x	x	x	x	x
VAS for legs and back		x	x	x	x	x	x	x	x	x	x
EuroQol-5D		x	x	x	x	x	x	x	x	x	x
Patient diary				x	x		x	x	x	x	x
Complications			x	x	x		x	x			x
Re-operation			x	x	x		x	x			x

MMSE: mini-mental state examination; ZCQ: Zurich claudication questionnaire; SPPB: short physical performance battery; RMDQ: Roland Morris disability questionnaire; SF-36: short form 36; VAS: visual analogue scale.

Primary outcome measure
Zurich Claudication Questionnaire (ZCQ)
The ZCQ is a validated disorder-specific outcome measure specifically designed for lumbar spinal stenosis [28,29]. The questionnaire consists of 3 scales: a symptom severity scale, a physical performance scale and a patient satisfaction scale. Multiple validated questionnaires exist for the use in evaluating back conditions, however the ZCQ specifically addresses the symptoms and functional deficits associated with spinal stenosis. Further, this questionnaire provides additional information on patient's satisfactory with the results of surgery.

Secondary outcome measures
Shuttle Walking Test (SWT)
The functional status of the patient can be assessed by the SWT. The SWT has previously been used as an outcome measure in the evaluation of patients with neurogenic intermittent claudication [29]. Subjects are required to walk back and forth, turning around two cones placed 9 meters apart making the shuttle distance 10 meters long. The distance travelled during 15 minutes and/or the time until neurogenic intermittent claudication prohibits further walking is recorded.

Mobility examination
Allocation to prolonged conservative treatment may possibly result in subjects becoming less active. As the activity level decreases, situations as walking/standing that give rise to symptoms of neurogenic intermittent claudication diminish and possibly result in symptom/pain relief. Although physical activity, disability and symptom severity are assessed during follow-up examinations, typically the reliability and validity of the measurement of habitual physical activity by questionnaires is low [30]. A continuous 7-day measurement with triaxial accelerometers (GENEA®, Unilever) provides information about the total amount, the frequency, the intensity, and the duration of physical activity [31-33]. Further, a common geriatric assessment, the Short Physical Performance Battery (SPPB), is used to assess how well older persons perform simple movements that represent the building blocks of daily activities that require good lower extremity function [34].

Roland Morris Disability Questionnaire (RMDQ)
The RMDQ is one of the most frequently used functional disability questionnaires for back-related conditions. The RMDQ was originally designed as a back condition specific questionnaire but is also widely used for patients with sciatica and neurogenic intermittent claudication [18,35,36]. The RMDQ score ranges from 0 (no disability) to 23 (maximum disability). A validated Dutch language version of the RMDQ is available. Compared to the Oswestry disability, of which no validated Dutch version is available, the sensitivity to change is in favor of the RMDQ [37].

Short Form-36 (SF-36)
The SF-36 is generic health survey questionnaire that measures overall health status, functional status and health-related quality of life [38,39]. The questions are divided in to eight domains: physical functioning, physical role limitations, emotional role limitations, social functioning, physical pain, general mental health, vitality, general health perception. For each domain, item scores are coded, summed, and transformed on to a scale from 0 (worst health) to 100 (best health).

Perceived Recovery
Perceived recovery is assessed with a seven-point likert scale varying from 'completely recovered' to 'worse than ever'. The scales are completed by the patient, the research nurse and the neurosurgeon.

Visual analogue scale (VAS) for pain in back and leg
The pain intensity experienced by the patient during the past week is assessed on a horizontal 100 millimeters scale varying from 0 millimeter, "no pain", to 100 millimeters 'the worst pain I can imagine' [40,41]. Back and leg pain is assessed separately.

Societal costs and utilities

Cost-effectiveness is expressed as cost per QALY, based on the QALY difference between the surgery and prolonged conservative treatment group. QALYs are estimated by using mean health state values during consecutive follow-up assessments with the EuroQol-5 D. The EuroQol-5 D is a generic health related quality of life questionnaire frequently used for investigating cost-effectiveness [42,43]. Treatment costs and direct medical costs are estimated on basis of the cost centre method. Participants are requested to record direct medical costs (e.g. physiotherapy, visits to general practitioners and medical specialists, nursing care, and medication) and indirect cost (e.g. disability related loss of productivity and additional travel expenses) using a cost diary.

Complications and re-operation incidence

The neurosurgeon and research nurse systematically record direct and indirect surgical complications and re-operation incidence. In case of failure of prolonged conservative management considerations for delayed surgery are reviewed.

Sample size

The success rate of both treatments is determined through perceived recovery and by applying the concept of MCID on the self reported functional disability scores. Derived from previous studies the estimated success rate of surgical management is 60% [9,24,44] and conservative management 40% [9,18,20,24] at 2 years of follow-up. The sample size calculation is based on the comparison of two independent proportions and contains the following parameters: Alpha: 0.05 (2-sided), Beta: 0.10 (90% power), success rate surgery: 0.6, success rate prolonged conservative treatment: 0.4. These assumptions yield a required sample size of 260 patients. Based on an expected drop-out rate of around 10% 280 patients need to be accrued (i.e. 140 patients in each treatment group). A sample size calculation based on the RMDQ yields approximately the same result. A difference of 3 points has been recommended for sample size calculations for clinical trials [36,45]. With Alpha: 0.05, Beta 0.10, μ difference of 3, and a standard deviation of RMDQ scores of 7.6 [18,36] a sample size of 270 patients is required.

Statistical analysis

Differences between treatment groups at baseline are assessed using univariate analyses to check for a balanced randomization. In case of substantial differences that induce confounding of the main treatment effect, both an uncorrected and a corrected (multivariate) analysis are presented. Data analysis is performed according to the intention to treat principle. The primary hypothesis using the primary outcome measure is a comparison of the average ZCQ over a period of 24 months. All continuous outcome measures are assessed with a repeated-measures analysis of variance with a first-order autoregressive covariance matrix. Differences between treatment groups are assessed by computing both the main effect of the treatment and the interaction between treatment and time in a mixed linear model. The area under curve for the ZCQ scores and secondary treatment outcome variables during the follow-up period are compared both implicitly using the repeated measurements approach. A Kaplan-Meier survival analysis is used to estimate the proportion of patients reporting recovery as a function of time elapsed from randomization. The average difference is characterized by a Hazard Ratio derived from a Cox Proportional Hazards model and the accompanying likelihood ratio test.

Table 3 lists the predefined patient characteristics and clinical findings possibly related to treatment outcome. These patient characteristics and clinical findings are assessed in a series of multivariate models where each particular covariate (patient characteristic or clinical finding) is tested for treatment-effect modification by extending the main analysis models (either the repeated measurements models or the survival models) by the main effect and the treatment*covariate interaction, thus providing a formal testing framework.

One covariate is pre-specified as a primary effect-modifier: duration of complaints at randomization. This covariate is clinically crucial since it allows the estimation of a possible "timing-of-surgery" effect by using the natural variability among patients with respect to the

Table 3 Predefined patient characteristics and clinical findings possibly related to treatment outcome

Patient characteristics	
	- Age
	- Gender
	- Pain intensity
	- Duration of symptoms
	- Predominant symptom (back pain compared to leg pain)
	- Walking distance
	- Patient's preference for treatment
Clinical findings	
	- One versus multiple level stenosis
	- Extent of stenosis (diameter/surface area MRI)
	- Localization stenosis (lateral recess or central canal)
	- Concomitant spondylolisthesis
	- Neurological impairment (muscle weakness, sensory loss)

duration of complaints at the moment of the decision to perform surgery or to postpone it. Therefore each of the main analyses is accompanied with an extension to test for "duration-of-complaints" * "treatment" interaction. This interaction is tested at the nominal 5% level within the primary analysis. If the interaction is not significant, the primary analysis of this trial defaults to the standard model with main effects. If the interaction is significant, it follows that a fortiori a treatment effect exists (and it varies with the duration of complaints). The model without the interaction still gives the overall (average) effect of the main hypothesis with its associated confidence interval. For the remaining 11 variables in Table 3 a 10% level of significance is applied, due to the reduced power of these tests. First two models are fitted, one with all patient characteristics and their interactions, and one with all clinical factors and their interactions. The total model improvement by adding the set of interactions is tested with the appropriate degrees of freedom. Hence 2 statistical tests are performed and a Bonferroni correction to the alpha level of 5% is applied to test whether any characteristic or clinical finding exists which modifies the treatment effect. When significance in one or both models exists, each individual covariate is tested, using a Bonferroni-Holm correction based on the number of covariates in the respective collection of covariates. Any such factor eventually classified as a significant effect modifier (Bonferroni-Holm with 10% as a starting level of significance) leads to subgroup estimation of treatment effects, to be tabulated in the final analysis.

Data is stored via the internet-based secure data management system "ProMISe" of the department of Medical Statistics and Bio Informatics. The analyses are carried out using appropriate statistical software (e.g. SAS, SPSS).

Discussion

It has been previously suggested that conservative management should be the initial treatment of degenerative lumbar stenosis and decompressive surgery is appropriate to consider only in case of intolerable pain or disability. Despite these recommendations the overall observed clinical improvement after surgery is disappointing and a substantial proportion of surgical candidates managed conservatively may still report improvement. Taken in to account the unexplained large geographic variation of surgery rates for lumbar stenosis and the long-term narrowing of the relative benefit of surgery compared to conservative management, another critical appraisal of indications for surgery is necessary.

Abbreviations
ITT: intention to treat; MCID: minimal clinically important difference; MMSE: mini-mental state examination; RMDQ: Roland Morris disability questionnaire; MRI: magnetic resonance imaging; QALY: quality-adjusted life year; RCT: randomized controlled trial; SF-36: short form 36; SPPB: short physical performance battery; SWT: shuttle walking test; VAS: visual analogue scale; ZCQ: Zurich claudication questionnaire.

Author details
[1]Department of Neurosurgery, Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden, The Netherlands. [2]Department of Neurosurgery, Medical Center Haaglanden, Lijnbaan 32, 2512 VA The Hague, The Netherlands. [3]Department of General Practice, Erasmus MC, University Medical Center, 's-Gravendijkwal 230, 3015 CE Rotterdam, The Netherlands. [4]Department of Medical Statistics, Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden, The Netherlands. [5]Department of Orthopaedic Surgery, Erasmus MC, University Medical Center, 's-Gravendijkwal 230, 3015 CE Rotterdam, The Netherlands. [6]Department of General Practice, Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden, The Netherlands.

Authors' contributions
GO is the coordinator and principal investigator and drafted the manuscript. WP, PL, BK, SB and RB designed the study protocol. RB is responsible for the statistical analysis and data storage using the ProMISe data management system. CV is involved in the implementation of the trial in the participating hospitals, PL and SB consulted on the protocol for physical therapy, JE consulted on the protocol for the general practitioners. All authors participated in the study design and coordination. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

References
1. Arnoldi CC, Brodsky AE, Cauchoix J, Crock HV, Dommisse GF, Edgar MA, Gargano FP, Jacobson RE, Kirkaldy-Willis WH, Kurihara A, *et al*: **Lumbar spinal stenosis and nerve root entrapment syndromes. Definition and classification.** *Clin Orthop Relat Res* 1976, , **115:** 4-5.
2. Amundsen T, Weber H, Lilleas F, Nordal HJ, Abdelnoor M, Magnaes B: **Lumbar spinal stenosis. Clinical and radiologic features.** *Spine (Phila Pa 1976)* 1995, **20(10)**:1178-1186.
3. Evans JG: **Neurogenic Intermittent Claudication.** *Br Med J* 1964, **2(5415)**:985-987.
4. Verbiest H: **A radicular syndrome from developmental narrowing of the lumbar vertebral canal.** *J Bone Joint Surg Br* 1954, **36-B(2)**:230-237.
5. Ciol MA, Deyo RA, Howell E, Kreif S: **An assessment of surgery for spinal stenosis: time trends, geographic variations, complications, and reoperations.** *J Am Geriatr Soc* 1996, **44(3)**:285-290.
6. Birkmeyer NJ, Weinstein JN: **Medical versus surgical treatment for low back pain: evidence and clinical practice.** *Eff Clin Pract* 1999, **2(5)**:218-227.
7. Weinstein JN, Lurie JD, Olson PR, Bronner KK, Fisher ES: **United States' trends and regional variations in lumbar spine surgery: 1992-2003.** *Spine (Phila Pa 1976)* 2006, **31(23)**:2707-2714.
8. Deyo RA, Mirza SK, Martin BI, Kreuter W, Goodman DC, Jarvik JG: **Trends, major medical complications, and charges associated with surgery for lumbar spinal stenosis in older adults.** *JAMA* 2010, **303(13)**:1259-1265.
9. Watters WC, Baisden J, Gilbert TJ, Kreiner S, Resnick DK, Bono CM, Ghiselli G, Heggeness MH, Mazanec DJ, O'Neill C, *et al*: **Degenerative lumbar spinal stenosis: an evidence-based clinical guideline for the diagnosis and treatment of degenerative lumbar spinal stenosis.** *Spine J* 2008, **8(2)**:305-310.

10. van Tulder MW, Koes BW, Bouter LM: **Conservative treatment of acute and chronic nonspecific low back pain. A systematic review of randomized controlled trials of the most common interventions.** *Spine (Phila Pa 1976)* 1997, **22(18)**:2128-2156.
11. Gibson JN, Waddell G: **Surgery for degenerative lumbar spondylosis.** *Cochrane Database Syst Rev* 2005, , **4**: CD001352.
12. Peul WC, van Houwelingen HC, van den Hout WB, Brand R, Eekhof JA, Tans JT, Thomeer RT, Koes BW: **Surgery versus prolonged conservative treatment for sciatica.** *N Engl J Med* 2007, **356(22)**:2245-2256.
13. Saal JA, Saal JS: **Nonoperative treatment of herniated lumbar intervertebral disc with radiculopathy. An outcome study.** *Spine (Phila Pa 1976)* 1989, **14(4)**:431-437.
14. Bozzao A, Gallucci M, Masciocchi C, Aprile I, Barile A, Passariello R: **Lumbar disk herniation: MR imaging assessment of natural history in patients treated without surgery.** *Radiology* 1992, **185(1)**:135-141.
15. Jensen MC, Brant-Zawadzki MN, Obuchowski N, Modic MT, Malkasian D, Ross JS: **Magnetic resonance imaging of the lumbar spine in people without back pain.** *N Engl J Med* 1994, **331(2)**:69-73.
16. Boden SD, Davis DO, Dina TS, Patronas NJ, Wiesel SW: **Abnormal magnetic-resonance scans of the lumbar spine in asymptomatic subjects. A prospective investigation.** *J Bone Joint Surg Am* 1990, **72(3)**:403-408.
17. Amundsen T, Weber H, Nordal HJ, Magnaes B, Abdelnoor M, Lilleas F: **Lumbar spinal stenosis: conservative or surgical management?: A prospective 10-year study.** *Spine* 2000, **25(11)**:1424-1435, discussion 1435-1426.
18. Atlas SJ, Keller RB, Wu YA, Deyo RA, Singer DE: **Long-term outcomes of surgical and nonsurgical management of lumbar spinal stenosis: 8 to 10 year results from the maine lumbar spine study.** *Spine* 2005, **30(8)**:936-943.
19. Johnsson KE, Rosen I, Uden A: **The natural course of lumbar spinal stenosis.** *Clin Orthop Relat Res* 1992, , **279**: 82-86.
20. Simotas AC: **Nonoperative treatment for lumbar spinal stenosis.** *Clin Orthop Relat Res* 2001, , **384**: 153-161.
21. Benoist M: **The natural history of lumbar degenerative spinal stenosis.** *Joint Bone Spine* 2002, **69(5)**:450-457.
22. Atlas SJ, Delitto A: **Spinal stenosis: surgical versus nonsurgical treatment.** *Clin Orthop Relat Res* 2006, **443**:198-207.
23. Weinstein JN, Tosteson TD, Lurie JD, Tosteson AN, Blood E, Hanscom B, Herkowitz H, Cammisa F, Albert T, Boden SD, *et al*: **Surgical versus nonsurgical therapy for lumbar spinal stenosis.** *N Engl J Med* 2008, **358(8)**:794-810.
24. Malmivaara A, Slatis P, Heliovaara M, Sainio P, Kinnunen H, Kankare J, Dalin-Hirvonen N, Seitsalo S, Herno A, Kortekangas P, *et al*: **Surgical or nonoperative treatment for lumbar spinal stenosis? A randomized controlled trial.** *Spine* 2007, **32(1)**:1-8.
25. Hagg O, Fritzell P, Nordwall A: **The clinical importance of changes in outcome scores after treatment for chronic low back pain.** *Eur Spine J* 2003, **12(1)**:12-20.
26. Jonsson B, Annertz M, Sjoberg C, Stromqvist B: **A prospective and consecutive study of surgically treated lumbar spinal stenosis. Part II: Five-year follow-up by an independent observer.** *Spine (Phila Pa 1976)* 1997, **22(24)**:2938-2944.
27. Bodack MP, Monteiro M: **Therapeutic exercise in the treatment of patients with lumbar spinal stenosis.** *Clin Orthop Relat Res* 2001, , **384**: 144-152.
28. Stucki G, Daltroy L, Liang MH, Lipson SJ, Fossel AH, Katz JN: **Measurement properties of a self-administered outcome measure in lumbar spinal stenosis.** *Spine* 1996, **21(7)**:796-803.
29. Pratt RK, Fairbank JC, Virr A: **The reliability of the Shuttle Walking Test, the Swiss Spinal Stenosis Questionnaire, the Oxford Spinal Stenosis Score, and the Oswestry Disability Index in the assessment of patients with lumbar spinal stenosis.** *Spine* 2002, **27(1)**:84-91.
30. Shephard RJ: **Limits to the measurement of habitual physical activity by questionnaires.** *Br J Sports Med* 2003, **37(3)**:197-206, discussion 206.
31. Westerterp KR: **Assessment of physical activity: a critical appraisal.** *Eur J Appl Physiol* 2009, **105(6)**:823-828.
32. Matthews CE, Ainsworth BE, Thompson RW, Bassett DR Jr: **Sources of variance in daily physical activity levels as measured by an accelerometer.** *Med Sci Sports Exerc* 2002, **34(8)**:1376-1381.

33. Ward DS, Evenson KR, Vaughn A, Rodgers AB, Troiano RP: **Accelerometer use in physical activity: best practices and research recommendations.** *Med Sci Sports Exerc* 2005, **37(11 Suppl)**:S582-588.
34. Cornoni-Huntley J, Ostfeld AM, Taylor JO, Wallace RB, Blazer D, Berkman LF, Evans DA, Kohout FJ, Lemke JH, Scherr PA, *et al*: **Established populations for epidemiologic studies of the elderly: study design and methodology.** *Aging (Milano)* 1993, **5(1)**:27-37.
35. Roland M, Morris R: **A study of the natural history of low-back pain. Part II: development of guidelines for trials of treatment in primary care.** *Spine (Phila Pa 1976)* 1983, **8(2)**:145-150.
36. Patrick DL, Deyo RA, Atlas SJ, Singer DE, Chapin A, Keller RB: **Assessing health-related quality of life in patients with sciatica.** *Spine (Phila Pa 1976)* 1995, **20(17)**:1899-1908, discussion 1909.
37. Beurskens AJ, de Vet HC, Koke AJ: **Responsiveness of functional status in low back pain: a comparison of different instruments.** *Pain* 1996, **65(1)**:71-76.
38. Ware JE, Sherbourne CD: **The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection.** *Med Care* 1992, **30(6)**:473-483.
39. Brazier JE, Harper R, Jones NM, O'Cathain A, Thomas KJ, Usherwood T, Westlake L: **Validating the SF-36 health survey questionnaire: new outcome measure for primary care.** *BMJ* 1992, **305(6846)**:160-164.
40. Scott J, Huskisson EC: **Graphic representation of pain.** *Pain* 1976, **2(2)**:175-184.
41. Huskisson EC: **Measurement of pain.** *Lancet* 1974, **2(7889)**:1127-1131.
42. EuroQol–a new facility for the measurement of health-related quality of life. The EuroQol Group. *Health Policy* 1990, **16(3)**:199-208.
43. Lamers LM, Stalmeier PF, McDonnell J, Krabbe PF, van Busschbach JJ: **[Measuring the quality of life in economic evaluations: the Dutch EQ-5 D tariff].** *Ned Tijdschr Geneeskd* 2005, **149(28)**:1574-1578.
44. Turner JA, Ersek M, Herron L, Deyo R: **Surgery for lumbar spinal stenosis. Attempted meta-analysis of the literature.** *Spine (Phila Pa 1976)* 1992, **17(1)**:1-8.
45. Roland M, Fairbank J: **The Roland-Morris Disability Questionnaire and the Oswestry Disability Questionnaire.** *Spine (Phila Pa 1976)* 2000, **25(24)**:3115-3124.