Rehabilitation following surgery for lumbar spinal stenosis

A Cochrane Review

Alison H McGregor[1], Katrin Probyn[1], Suzie Cro[2], Caroline J Doré[2], A Kim Burton[3], Federico Balagué[4], Tamar Pincus[5], Jeremy Fairbank[6]

[1]Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, London, UK
[2]Medical Research Council Clinical Trials Unit, London, UK
[3] Spinal Research Unit, University of Huddersfield, Huddersfield, UK
[4]Cantonal Hospital of Fribourg, Fribourg, Switzerland
[5]Department of Psychology, Royal Holloway University of London, Egham, UK
[6]Department of Orthopaedic Surgery, Nuffield Orthopaedic Centre, Oxford, UK

- Address for correspondence

Professor Alison McGregor
Department of Surgery and Cancer, Faculty of Medicine
Imperial College London
Charing Cross Hospital
London
W6 8RF
UK
E-mail: a.mcgregor@imperial.ac.uk

Abstract

Study Design

A systematic review of randomised controlled trials (RCTs)

Objective

To determine the effects of active rehabilitation on functional outcome following lumbar spinal stenosis surgery when compared with 'usual postoperative care'.

Summary of background data

Surgery rates for lumbar spinal stenosis have risen, yet outcomes remain suboptimal. Post-operative rehabilitation has been suggested as a tool to improve post-operative function but, to date, there is limited evidence to support its use.

Methods

CENTRAL (*The Cochrane Library*), the Cochrane Back Review Group Trials Register, MEDLINE, EMBASE, CINAHL and PEDro electronic databases were searched. Randomised controlled trials (RCTs) comparing the effectiveness of active rehabilitation with usual care in adults with lumbar spinal stenosis who had undergone primary spinal decompression surgery were included. Two authors independently selected studies, assessed the risk of bias, and extracted the data in line with the recommendations of the Cochrane Back Review Group. Study results were pooled in a meta-analysis when appropriate using functional status as the primary outcome, with secondary outcomes including measures of leg pain, low back pain, and global improvement/general health. The GRADE approach was used to assess the quality of the evidence.

Results

Our searches yielded 1,726 articles, of which three studies (N = 373 participants) were suitable for inclusion in meta-analysis. All included studies were deemed to have low risk of bias; no study had unacceptably high dropout rates. There was moderate evidence suggesting that active rehabilitation was more effective than usual care in improving both short- and long-term functional status following surgery. Similar findings were noted for secondary outcomes, including short-term improvement in low back pain and long-term improvement in both low back pain and leg pain.

Conclusions

We obtained moderate-quality evidence indicating that postoperative active rehabilitation after decompression surgery for lumbar spinal stenosis is more effective than usual care. Further work is required particularly with respect to the cost effectiveness of such interventions.

Key words: systematic review; Cochrane Back Review Group; post-operative management; meta-analysis; functional outcome; quality of life

Mini Abstract

This review sought to determine if there is any evidence supporting the use of rehabilitation after surgery for lumbar spinal stenosis. It concluded that there is moderate quality evidence to suggest that active rehabilitation can lead to improvements in function and pain.

Background

Spinal stenosis is a narrowing of the spinal canal leading to pressure on the nerve roots or spinal cord, causing pain, predominantly in the leg but also in the back. The causes of spinal stenosis are multi-faceted, but are associated with degenerative changes to the intervertebral disc, associated vertebrae and supporting ligamentous structures. The net result is narrowing of either the central or the lateral root canal (or both) leading to pressure on the nerve root and associated pain. Decompression surgery, which involves a posterior midline incision through the fascia and spinal muscles to obtain access to the compressed nerves, is often performed to relieve this leg pain. Constriction is reduced by removal of any excess bone, thickened ligaments, degenerate disc material and other fibrous tissue.

Many researchers have noted a rise in spinal decompression surgical rates in recent times attributing this to the growing elderly population around the world [Taylor et al 1994, Deyo et al 2005, Chou 2009; Rhee 2006; Stromqvist 2001], with Deyo [2010] commenting that spinal stenosis is now the most common indication for spinal surgery in those over 65 years of age. Further rises in decompression surgery are however, predicted in line with the anticipated growth by 59% of the population over 65 years by 2025 [Deyo 201].

Spinal stenosis does not always require surgery and there is some evidence that facet joint and epidural injections may be effective in its management [Manchikanti 2010]. However overall, surgery seems to be better than conservative interventions such as injections and rehabilitation (Atlas 2005; Chang 2005; Tran 2010). This was confirmed by a recent Cochrane review which noted that good quality trials into alternative conservative approaches for the management of spinal stenosis were lacking and that further research in this area was urgently required [Ammendolia et al 2013]. Various surgical techniques are used in decompression surgery, the most common being a decompression laminectomy, whereby the structures compressing the nerve root are removed (Genevay 2010). When multiple nerve roots are involved, this often necessitates a fusion procedure. The use of spinal fusion is still widely debated, and a range of approaches and techniques and outcomes have been described (Gibson 2005). Decompression is one of the most common types of spinal surgery; (Chou 2009; Rhee 2006; Stromqvist 2001; Taylor 1994), with the US Medicare system reporting that more than 37,500 surgical procedures were performed for this condition in 2007 (Deyo 2010).

However, a sizeable proportion of participants do not regain good function after surgery, and the outcome of spinal decompression surgery is not ideal. 'Success' rates for decompression surgery vary

considerably, with functional improvement ranging between 58% and 69% (Gunzburg 2003; Stromqvist 2001; Turner 1992), participant satisfaction ranging from 15% to 81% (Atlas 2005; McGregor 2002; Yee 2008) and gain in function and pain varying between studies (McGregor 2002; Yorimitsu 2001). Evidence of trunk muscle dysfunction has been noted in people with back problems (Hides 1994), and muscles are known to be damaged by surgery (Taylor 2002); thus rehabilitation would appear to be a promising approach to improving outcomes.

Postoperative care following spinal surgery is variable, with major differences reported between surgeons in the type and intensity of rehabilitation provided and in restrictions imposed and advice offered to participants (McGregor 2006). Postoperative management may include education (McGregor 2007), rehabilitation (Erdogmus 2007; McGregor 2010; Nielsen 2008; Ostelo 2009), exercise (Kim 2010), behavioural graded training (Ostelo 2004), neuromuscular training (Millisdotter 2007) and stabilisation training (Mannion 2007). Evidence is currently insufficient for researchers to determine best clinical practice, although indications suggest that some form of exercise or rehabilitation intervention may be beneficial. This review was therefore undertake to determine whether active rehabilitation programmes following primary surgery for lumbar spinal stenosis have an impact on functional outcomes and whether such programmes are superior to 'usual postoperative care'. This article is adapted from the Cochrane review “McGregor AH, Probyn K, Cro S, Doré CJ, Burton AK, Balagué F, Pincus T, Fairbank J. Rehabilitation following surgery for lumbar spinal stenosis. Cochrane Database of Systematic Reviews 2013, Issue 12. Art. No.: CD009644. DOI: 10.1002/14651858.CD009644.pub2.”

Methods

The objective of this systematic review was to determine whether active rehabilitation or specific advice to stay active has an impact on the functional outcome of primary decompression surgery for lumbar spinal stenosis as compared to 'usual care', which includes no post-operative intervention or deliberately delivered 'therapeutic' advice.

Types of studies: Only randomised controlled trials (RCTs) were included in the review.

Types of participants: Adults 18 years of age or older who had spinal decompression surgery for central or lateral stenosis at single or multiple levels were included. Stenosis had to be confirmed through imaging and clinical assessment, and the surgery performed had to be primary decompression surgery for stenosis (as distinct from surgery for disc herniation). All surgical decompression procedures, with or without vertebral fusion, were included.

Types of interventions: This review examines the delivery of active rehabilitation versus usual care after surgery. Interventions were classified as active rehabilitation or usual care. Postsurgical active rehabilitation interventions included all forms of active rehabilitation treatment that aimed to restore or improve function. This encompasses all forms of group or therapist-led exercise training or stabilisation training involving muscle-strengthening exercises and flexibility training, as well as educational materials encouraging activity. Usual care ranged from limited advice provided postoperatively to stay active to a brief general programme of exercises with the primary aim of preventing deep vein thrombosis.

Types of outcome measures: Trials were included if they utilised one or more of the standardised outcome measures recommended by the Cochrane Back Review Group (Furlan 2009). These include; Disease-specific measures of functional and/or disability status (eg Oswestry Disability Index, the Roland Morris Disability Questionnaire) which were considered as primary outcomes; and measures of global health (eg SF36, EQ-5D); global improvement measures (eg proportion of participants that recovered); pain severity (eg visual analogue scale (VAS)); and work absenteeism which were considered as secondary outcomes. Work absenteeism was poorly or inconsistently reported and therefore was not considered in the analysis. Outcomes were considered within 6 months of surgery (short term) and 12 months after surgery (long term).

Search methods for identification of studies: We searched the following databases from their first issues to March 2013: CENTRAL (*The Cochrane Library,* most recent issue (March 2013), which

includes the CBRG Trials Register; MEDLINE; EMBASE; CINAHL; and PEDro (Appendix 1)). In addition the reference lists of all relevant papers were screened as well as personal biographies and communications of known experts in the field.

Data collection and analysis

Selection of studies: Two review authors (AMcG and KP) independently screened the search results by reading titles and abstracts. Potentially relevant studies were obtained in full text and were independently assessed for inclusion by the same authors. Disagreement was resolved through discussion. A third, fourth and fifth review author (AKB, JF and FB) adjudicated unresolved disagreements. Authors of individual trials were excluded from any decisions regarding inclusion/exclusion, data extraction and risk of bias.

Data extraction and management: Basic information was obtained for each study concerning methods (study design, sample size, etc.), participants (selection criteria and diagnoses, age, gender, etc.), type of surgery, intervention, control treatment and outcome variables with results recorded onto a separate pre-developed form. Data extraction was performed independently by two review authors (KP and SC); inconsistencies were resolved through a 3rd author where necessary.

Risk of Bias Assessment: This was determined using the criteria recommended by the Cochrane Back Review Group with the additional criteria, 'Other sources of bias' (Furlan 2009; Higgins 2011). For each study, each criterion was rated as 'low risk', 'high risk' or 'unclear risk'. Studies with a low risk of bias were defined as RCTs that satisfied six or more of the low risk of bias criteria and that had no serious flaws (Furlan 2009). Serious flaws were predefined to include unacceptably high dropout rates (e.g. greater than 50% at first and subsequent time points); unacceptably unbalanced dropout rates (e.g. 40% greater dropout rate in one group); unacceptably low adherence rates (e.g. less than 50% with total or nearly total non-adherence to the protocol); and clear, significantly unbalanced baseline differences for the primary outcome (functional status) that were not accounted for in the analysis.

Measures of treatment effect: Identified studies were evaluated as clinically homogeneous regarding study populations, types of interventions and types of follow-up and outcomes, allowing us to perform meta-analysis to pool treatment effects. For continuous outcomes, a pooled mean difference (MD) was calculated when the same measurement scale was used and a standardised mean difference (SMD) when different measurement scales were used. For each pooled outcome, an associated 95% confidence interval (95% CI) was computed. When continuous outcome data

were positively skewed, the meta-analysis was conducted on a log-scale. Pooled mean differences on the log-scale were converted back to the original measurement scale using the anti-log, EXP(), to give a ratio of geometric means on the unlogged scale. Ratios are also expressed as percentage differences to aid interpretation of relative differences in original untransformed outcome variables between intervention groups (Bland 1996). The clinical relevance of each included study was independently assessed by the review authors using Furlan's (2009) approach. We evaluated the statistical importance and the clinical importance of pooled results. Effect sizes were assessed and interpreted using Cohen's levels (Higgins 2011).

Unit of analysis issues: The unit of analysis was the participant. One of the included studies (Mannion 2007) compared two treatment groups against one usual care group. This raised a unit of analysis problem, as in a meta-analysis, every individual must appear only once in every comparison. So that all individuals and both of the treatment groups could be included, the two treatment groups were combined into one treatment group (and compared with one control group). This is the approach recommended in the Cochrane Handbook (Higgins 2011).

Missing data: In two of the papers selected (McGregor 2010 and Mannion 2007), only subgroups were suitable for inclusion in the review; relevant data from these subgroups were not published in the papers but were obtained directly from the authors.

Assessment of heterogeneity: This was determined by examining characteristics of study participants, types of interventions, comparisons, follow-up and assessment of primary and secondary outcomes. The Chi^2 test and the I^2 statistic were used to assess the statistical heterogeneity of studies deemed clinically homogeneous. A P value for the Chi^2 test of less than 0.05 or $I^2 > 50\%$ was considered to indicate significant statistical heterogeneity. Forest plots were also used to assess heterogeneity visually.

Assessment of reporting biases: Trial registers and published reports of trials were searched to identify any inconsistencies between published trials and registered trials. The presence of publication bias and heterogeneity was assessed using funnel plots.

Data synthesis: As a sufficient number of clinically similar studies were available, the results were pooled in meta-analyses using Review Manager software (Review Manager 2011). A fixed-effect inverse variance model was used to pool results when no substantial evidence of statistical heterogeneity was found (Chi^2 $P > 0.05$, $I^2 \leq 50\%$). When substantial statistical heterogeneity was detected, a random-effects inverse variance model was used as an alternative.

We assessed the overall quality of the evidence for each outcome using the GRADE approach, as recommended in the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2011) and adapted in the *Updated Method Guidelines for Systematic Reviews in the Cochrane Back Review Group* (Furlan 2009).

Sensitivity analysis: Sensitivity analysis was planned to explore the robustness of the review findings, however, due to the small number of identified studies (3) this was not undertaken.

Results

In total, the search, after duplicates were removed, yielded 1,726 articles. Titles and abstracts were screened, and 1,712 records were excluded, primarily for one of the following reasons: incorrect type of surgery (e.g. discectomy), comparison of surgical interventions, comparison of surgery with conservative interventions, investigation of conservative treatment for low back pain, focus on rehabilitation for other conditions such as knee replacement, cardiovascular illness, etc., or they were review papers. Fourteen records were retrieved in full text, and 11 of them were excluded; 2 were not formal RCTs (Pons et al 2011, Sogaard et al 2008); 3 were not the correct population group (Canbulat et al 2011, Jeric et al 1991, Mannion et al 2010); and the remainder were not the correct intervention (Abbott et al 2010, Brox et al 2003, Christensen et al 2003,Hagg et al 2004, Nielsen et al 2010, Wu 2005). The remaining three studies were included in the review. All three trials were clinically homogeneous regarding the baseline characteristics of participants and outcome measures. Interventions were comparable regarding starting point, type, intensity and duration of treatment. Interventions were started at six weeks postoperatively (McGregor 2010), two months postoperatively (Mannion 2007) and three months postoperatively (Aalto 2011). Intervention duration ranged from six weeks, with twice-weekly 60-minute sessions (McGregor 2010), to 12 weeks, with a 90-minute session once a week (Aalto 2011), or 12 weeks, with twice-weekly 30-minute sessions (Mannion 2007). The control groups in all trials were comparable, insofar as they did not include specific postoperative interventions and were treated with either "usual care" or "self-management" and were given either advice postoperatively to “stay active” or a brief general programme of exercises.

Characteristics of included studies: The three studies included a total of 373 participants; however, as not all of the participants attended the short-term follow-up assessment in McGregor 2010, fewer participants (340) are included in the meta-analysis of the short-term follow-up outcomes than for the long-term follow-up outcomes. Although McGregor 2010 had higher follow-up at 12 months (the study's primary endpoint) than at three months, sensitivity analyses conducted by McGregor et al (2010) verified consistent findings with different missing data assumptions, giving us no reason to exclude this study.

The included studies were similar with regard to baseline characteristics (Table 1); however, some unexplained heterogeneity with respect to gender and age was noted. Mannion 2007 analysed a greater number of male participants (59% vs 41% and 51.5%) who on average were five years older than those in the other two studies. For the purposes of this review, it was decided that the studies were sufficiently similar to permit pooling of data for the meta-analysis.

Postoperative baseline values of outcome variables at the start of the intervention (Table 1) were also assessed for each included study and were found to be similar between groups. Baseline values for functional outcome in Mannion 2007 were slightly lower than those in the other two studies. This may be due to the slightly later starting point of the intervention in this trial.

Risk of bias in included studies: All included studies were rated as having low risk of bias, Figure 1 summarises the risks identified for each study. Within all three included studies there was a high risk of performance bias since participants and care providers had knowledge of the allocated interventions during the studies. Blinding of participants and care providers across studies was not feasible due to the nature of the intervention.

Effects of interventions

All of the studies reported functional status as the primary outcome measure, with secondary outcomes including leg and low back pain. In addition, two of the included studies also reported general health; reporting of work status was poor in all studies and could therefore not be included in subsequent analysis. None of the included studies reported any relevant adverse events. McGregor 2010 reported short-term outcomes at three months, Mannion 2007 at five months, and Aalto 2011 at six months. All three trials reported long-term outcomes at 12 months postoperatively. We judged the three identified studies to be clinically homogeneous regarding study population, types of interventions, comparisons, follow-up and outcome, allowing us to perform a meta-analysis to pool treatment effects across all three studies.

Short term outcomes (within six months postoperative): Moderate-quality evidence from three RCTs (340 participants) indicates that active rehabilitation is more effective than usual care for functional status (log SMD -0.22, 95% CI -0.44 to 0.00, corresponding to an average percentage improvement (reduction in standardised functional score) of 20%, 95% CI 0% to 36%; Figure 2), and more effective than usual care for reported low back pain (log MD -0.18, 95% CI -0.35 to -0.02, corresponding to an average percentage improvement (reduction in VAS score) of 16%, 95% CI 2% to 30%; Figure 2).

Low-quality evidence from three RCTs (340 participants) shows no statistical difference in leg pain (log MD -0.17, 95% CI -0.52 to 0.19, corresponding to an average percentage improvement (reduction in VAS score) of 16%, 95% CI 21% worsening (increase in VAS) to 41% improvement (decrease in VAS)) between participants who received active rehabilitation and those who received usual care (Figure 2). Low-quality evidence from two RCTs (238 participants) reveals no statistically significant difference in general health (MD 1.30, 95% CI -4.45 to 7.06) between participants who received rehabilitation and those who received usual care (Figure 2).

Long term outcomes (12 months postoperative): Moderate-quality evidence from the three RCTs (373 participants) indicates that active rehabilitation is more effective than usual care for functional status (log SMD -0.26, 95% CI -0.46 to -0.05, corresponding to an average percentage improvement (reduction in standardised functional score) of 23%, 95% CI 5% to 37%; Figure 3.) There is moderate-quality evidence (373 participants) indicating that active rehabilitation is more effective than usual care for low back pain (log MD -0.20, 95% CI -0.36 to -0.05, corresponding to an average percentage improvement (reduction in VAS score) of 18%, 95% CI 5% to 30%; Figure 3). Similarly there is moderate-quality evidence (373 participants) to suggest that active rehabilitation is more effective than usual care for leg pain (log MD -0.24, 95% CI -0.47 to -0.01, corresponding to an average percentage improvement (reduction in VAS score) of 21%, 95% CI 1% to 37%; Figure 3). Finally there was low-quality evidence from two RCTs (273 participants) that there was no statistical differences in general health (MD -0.48, 95% CI -6.41 to 5.4) between participants who receive rehabilitation and those who receive usual care (Figure 3).

Clinically Relevant Effect Size Estimates

Established predefined outcome-specific minimal clinically important differences were employed to interpret effect sizes (Furlan 2009). Consideration of the magnitude of the effects (differences between groups when the data are analysed on the raw scale, or relative differences between groups when the data are analysed on the logged scale) based on Cohen's levels and predefined

outcome-specific clinically relevant effect sizes (Furlan 2009, Copay 2008) indicates that in the short term, a clinically significant medium effect of rehabilitation on functional status is noted (above the predefined relative functional outcome threshold of 8% to 12% for clinical relevance). A medium effect of rehabilitation on low back pain has been observed; however, this finding is not clinically significant, as it is below the predefined clinically relevant difference for low back pain of 30%. The effects of rehabilitation on leg pain and general health are small and are neither statistically nor clinically significant.

With respect to long term outcome, on average there is a clinically significant medium effect of rehabilitation on functional status. A medium effect of rehabilitation on low back pain was noted, but this is not clinically significant because it is below the predefined clinically relevant difference of 30% for low back pain. The effects of rehabilitation on leg pain and on general health are small and statistically significant but are not clinically significant.

The main findings of this review are summarised in table 3.

Discussion

This review sought to determine whether active rehabilitation following surgery for lumbar spinal stenosis has an impact on functional outcomes. Although our searches yielded 1,726 results, only three randomised controlled trials were suitable for inclusion in this review, and for two of these, only subgroups of the original study population met the inclusion criteria. Whilst the diagnosis of spinal stenosis was consistent between studies, local variations in surgical procedures may have occurred. However, in all studies the surgery was intended to relieve nerve root compression and did not require a fusion procedure.

The results of our subsequent meta-analysis based on the 3 studies and 373 participants indicate that active rehabilitation is clinically more effective than usual care in improving both short-term and long-term functional status and this is supported by moderate-quality evidence. Similarly, moderate-quality evidence suggests that active rehabilitation is more effective than usual care for short-term (within six months postoperatively) improvement in low back pain and for long-term (12 months postoperatively) improvement in low back pain and leg pain. However, observed pooled differences in low back pain were smaller than the predefined clinically relevant difference for low back pain improvement in both the short-term and long-term follow-up. The observed pooled difference for leg pain in the long term was also smaller than the predefined clinically relevant difference. With respect to changes in other secondary outcomes, active rehabilitation could not be confirmed to be

more effective than usual care in either short-term or long-term follow-up; however, these results should be interpreted cautiously because the quality of the evidence was low.

The findings of this review are in accordance with a similar Cochrane review of postoperative rehabilitation after surgery to relieve disc herniation (Ostelo 2009) which indicated that post-operative exercise programmes instigated 4-6 weeks following surgery lead to a faster decrease in pain and disability than no treatment. No adverse events were noted and there was no indication that active programmes lead to an increase in reoperation rates. Rushton et al's (2011) systematic review of post-operative physiotherapy noted very low-quality evidence supporting the use of postoperative rehabilitation and highlighted that best practice in relation to rehabilitation was unclear. In relation to this review and other new RCTs in relation to spinal discectomy this Cochrane review is currently being updated.

Recent work by McGregor 2010 suggested that outcomes from postoperative interventions may be influenced by underlying pathology; arguing that spinal stenotic patients were older, with poor preoperative function and a higher risk of comorbidities and thus a greater need for rehabilitation to improve surgical outcome. Nielsen 2010 explored the outcomes of spinal decompression surgery in participants with degenerative disc disease. However, this study included both preoperative and postoperative rehabilitation, which consisted of exercise, dietary changes and general advice. Outcomes were positive compared with those following usual care, suggesting a beneficial impact on functional recovery and on hospital length of stay, and lending further support to the benefits of rehabilitation provided to spinal surgery patients. However, this study was excluded from the current review because of the combined complex intervention.

In this review, we excluded the Abbott 2010 study, which compared exercise therapy after fusion surgery with cognitive-behavioural therapy because no usual care control arm was included in this study. Nevertheless, it is worth noting that this work suggested that additional improvements could be achieved through the inclusion of psychomotor therapy. Similarly, Christensen 2003 indicated that simple support provided through a back-café group achieved greater improvements in physical function than were attained through regular exercise classes. The intervention was instigated only three months after surgery; again, no control arm was included, but study findings do support the inferences of the Abbott 2010 study and warrant consideration in the design of future rehabilitation strategies.

Clearly, further research is required to consolidate the findings of this current Cochrane review primarily due to the low number of trials eligible for inclusion. Future studies should also include a

cost-benefit analysis as in the present review such data were available for only one of the three studies. Other issues highlighted in this review include the timing of the intervention after surgery and, as indicated in Rushton 2012, the content of the rehabilitation package. At the moment, little consensus has been reached on what constitutes an appropriate active rehabilitation programme, when it should be delivered, how intense it should be, how long it should be delivered for, or how frequently, and, of course, whether a group format for delivery is preferable. We know that compliance can be an issue for patients (Johnson 2007; McGregor 2010); thus future work is needed in this area to explore these issues. This work should factor in the need for clear educational materials and the growing emphasis on self-management strategies.

Nielsen 2010 has suggested that there may be a role for preoperative rehabilitation. It would be useful to look at the care pathway and to view interventions in a more holistic way, rather than simply focusing on the surgical intervention. This would necessitate greater consideration of patient preferences and experiences and the need to tailor care at a more individual level.

To summarise, this review has revealed moderate-quality evidence indicating that postoperative active rehabilitation after decompression surgery for lumbar spinal stenosis is more effective than usual care in improving both short-term and long-term (back-related) functional status. Similar findings were noted for secondary outcomes, including short-term improvement in low back pain and long-term improvement in both low back pain and leg pain.

Acknowledgements

We thank Teresa Marin, Rachel Couban and Allison Kelly from the Cochrane Back Review Group for their help.

Contributions of authors

All authors have contributed to the development of this review. AMcG and KP screened the search results and selected studies. AKB, FB, TP and JF were involved in the final decision making. AMcG, KP and SC extracted the data (AMcG was not involved in data extraction from the trial McGregor 2010, for which she served as first author). Risk of bias assessment was completed by KP and SC. GRADE assessment and assessment of clinical relevance were completed by KP and SC. Analysis was done by KP and SC, and CJD was also involved in the process. AMcG, KP and SC were the main authors of the text. All review authors contributed to the review writing process.

Table 1 Baseline characteristics of included studies

	Aalto 2011		Mannion 2007		McGregor 2010	
	Reha mean (SD)	Control mean (SD)	Reha mean (SD)	Control mean (SD)	Reha mean (SD)	Control mean (SD)
Age, years	62.5 (34 to 86; 11.1)		67.1 (10.6)		62 (15)	
BMI	29.5 (4.0)		27 (4.5)		27(5)	
Gender: female/male	59%/41%		41%/59%		49.5%/51.5%	
Functional status (Oswestry Disability Index (ODI 0 to 100%) or Roland Morris (0 to 24))	24.3 (15.9)	29.7 (20,5)	10.9 (4.9)	10.6 (4.7)	30(18)	32(21)
Low back pain (VAS 0 to 100) (Where 0 is no pain and 100 is worst pain)	16 (19)	20 (26)	24.6 (19.8)	29 (21)	35 (26)	35 (29)
Leg pain (VAS 0 to 100) (Where 0 is no pain and 100 is worst pain)	27 (26)	32 (28)	29.5 (22.9)	22 (24)	33 (27)	32 (28)
General health (VAS 100 to 0) Where 100 is worst and 0 is best general health status)	-	-	68.6 (21)	69 (26)	69 (22)	66 (26)

Table 2 Summary of findings

Short term outcomes:

Rehabilitation following surgery for lumbar spinal stenosis—short-term outcomes						
Patient or population: participants with lumbar spinal stenosis **Settings:** hospital **Intervention:** rehabilitation after surgery						
Outcomes	**Illustrative comparative risks* (95% CI)**		**Relative effect (95% CI)**	**No. of participants (studies)**	**Quality of the evidence (GRADE)**	**Comments**
	Assumed risk	**Corresponding risk**				
	Control	**Rehabilitation after surgery**				
Functional	Mean	Mean functional		340	⊕⊕⊕⊖	SMD on the log-

status, short term Oswestry Disability Index (ODI 0 to 100%) or Roland Morris (0 to 24) Follow-up: three to six months	functional status short term ranged across control groups from **1.98 to 3.32 on log-scale**	status short term in the intervention groups was **0.22 standard deviations lower** (0.44 lower to 0 higher)		(three studies)	**Moderate**[1]	scale corresponds to 20% improvement (0% to 36% improvement) in the rehabilitation group. This difference is clinically relevant[2]
Leg pain, short term Visual analog scale (VAS 0 to 100) Follow-up: three to six months	Mean leg pain short term ranged across control groups from **2.88 to 3.42 on log-scale**	Mean leg pain short term in the intervention groups was **0.17 lower** (0.52 lower to 0.19 higher)		340 (three studies)	⊕⊕⊖⊖ **Low**[3,4]	MD on the log-scale corresponds to 16% improvement (21% worsening to 41% improvement). This difference is not statistically or clinically relevant[5]
Low back pain, short term Visual analogue scale (VAS 0 to 100) Follow-up: three to six months	Mean low back pain short term ranged across control groups from **2.50 to 3.51 on log-scale**	Mean low back pain short term in the intervention groups was **0.18 lower** (0.35 to 0.02 lower)		340 (three studies)	⊕⊕⊕⊖ **Moderate**[6]	MD on the log-scale corresponds to 16% improvement (2% to 30% improvement) in low back pain. This difference is not clinically relevant[5]
General health, short term Visual analogue scale (VAS 100 to 0) Follow-up: three to five months	Mean general health short term ranged across control groups from **66 to 74**	Mean general health short term in the intervention groups was **1.3 higher** (4.45 lower to 7.06 higher)		238 (two studies)	⊕⊕⊖⊖ **Low**[4,7]	Mean difference is not statistically significant and is not clinically relevant
Adverse Events - not reported	See comment	See comment	Not estimable		See comment	None of the included studies reported any relevant adverse events

*The **corresponding risk** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).
CI: Confidence interval.

GRADE Working Group grades of evidence.
High quality: Further research is very unlikely to change our confidence in the estimate of effect.
Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low quality: We are very uncertain about the estimate.

Footnotes

[1]Serious Inconsistency: due to direction. One of the three studies reported an average effect size that favoured the control; the other two favoured rehabilitation.
[2]This difference is clinically relevant because it is above the predefined clinically relevant relative difference of 8% to 12%.
[3]Serious Inconsistency: due to direction and statistical heterogeneity (P = 0.01). Two of the trials reported an average effect size that favoured rehabilitation. The other trial average effect size favoured the control.
[4]Serious Imprecision: 95% CI for the pooled intervention effect could support the rehabilitation group or the control group.
[5]This difference is not clinically relevant because it is below the predefined clinically relevant difference of 30%.
[6]Serious inconsistency: due to direction. Two of the trials reported an average effect size that favoured rehabilitation. The other trial average effect size favoured the control.
[7]Serious inconsistency: Average effects of the two included trials differ in direction.

Long term outcomes:

Rehabilitation following surgery for lumbar spinal stenosis—long-term outcomes						
Patient or population: participants with lumbar spinal stenosis **Settings:** hospital **Intervention:** rehabilitation after surgery						
Outcomes	**Illustrative comparative risks* (95% CI)**		**Relative effect (95% CI)**	**No. of participants (studies)**	**Quality of the evidence (GRADE)**	**Comments**
	Assumed risk	**Corresponding risk**				
	Control	**Rehabilitation after surgery**				
Functional status, long term Oswestry Disability Index (ODI 0 to 100%) or Roland Morris (0 to 24) Follow-up: 12 months	Mean functional status long term ranged across control groups from **2.04 to 3.32 on log-scale**	Mean functional status long term in the intervention groups was **0.26 standard deviations lower** (0.46 to 0.05 lower)		373 (three studies)	⊕⊕⊕⊖ **Moderate**[1]	SMD on the log-scale corresponds to 23% improvement (5% to 37% improvement) in functional status. This difference is clinically relevant[2]
Leg pain, long term Visual analogue scale (VAS 0 to 100) Follow-up: 12 months	Mean leg pain long term ranged across control groups from **3.20 to 3.56 on log-scale**	Mean leg pain long term in the intervention groups was **0.24 lower** (0.47 to 0.01 lower)		373 (three studies)	⊕⊕⊕⊖ **Moderate**[1]	MD on the log-scale corresponds to 21% improvement (1% to 37% improvement) in leg pain. This difference is not clinically relevant[3]
Low back pain, long term Visual analogue scale (VAS 0 to 100)	Mean low back pain long term ranged across control groups	Mean low back pain long term in the intervention groups was **0.2 lower** (0.36 to 0.05 lower)		373 (three studies)	⊕⊕⊕⊖ **Moderate**[1]	MD on the log-scale corresponds to 18% improvement (5% to 30% improvement) in leg pain. This

Follow-up: 12 months	from **2.79 to 3.54 on log-scale**					difference is not clinically relevant[3]
General health, long term Visual analogue scale (VAS 100 to 0) Follow-up: 12 months	Mean general health long term ranged across control groups from **64 to 70**	Mean general health long term in the intervention groups was **0.48 higher** (5.44 lower to 6.41 higher)		273 (two studies)	⊕⊕⊖⊖ **Low**[4,5]	Mean difference is not statistically significant or clinically relevant
Adverse event—not reported	See comment	See comment	Not estimable		See comment	None of the included studies reported any relevant adverse events

*The **corresponding risk** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).
CI: Confidence interval.

GRADE Working Group grades of evidence.
High quality: Further research is very unlikely to change our confidence in the estimate of effect.
Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low quality: We are very uncertain about the estimate.

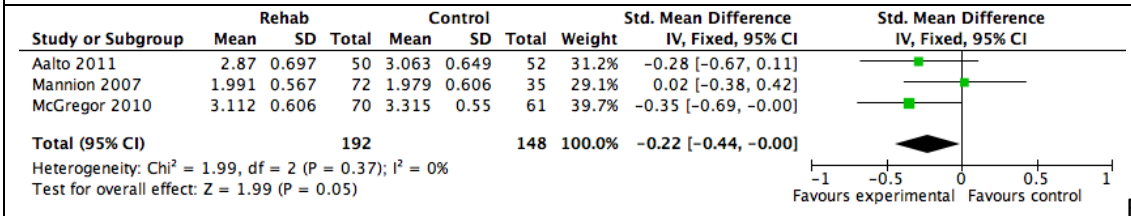
Footnotes

[1]Serious Inconsistency: All studies were agreeable on the direction of the average effect, but only one study identified this to be a significant effect.
[2]This difference is clinically relevant because it is above the predefined clinically relevant relative difference of 8% to 12%.
[3]This difference is not clinically relevant because it is below the predefined clinically relevant difference of 30%.
[4]Serious inconsistency: The average effects of the two included trials differ in direction.
[5]Serious imprecision: 95% CI for the pooled intervention effect could support the rehabilitation group or the control group.

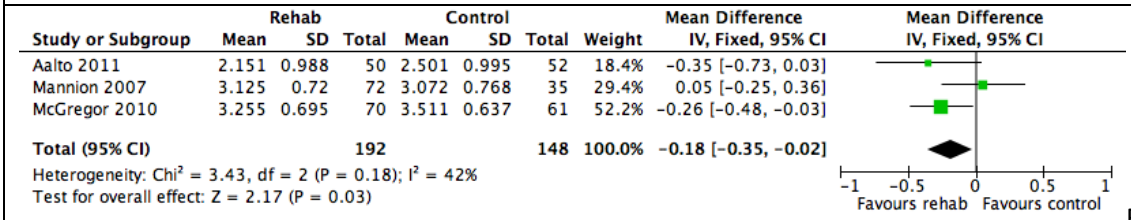
Figure 1

Caption: Risk of bias summary: review authors' judgements about each risk of bias item for each included study.

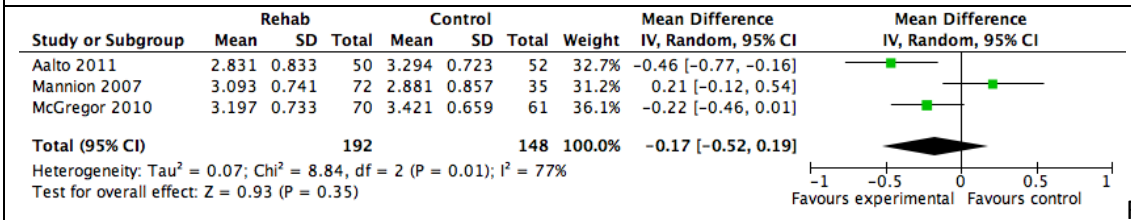
Figure 2 Short term outcomes



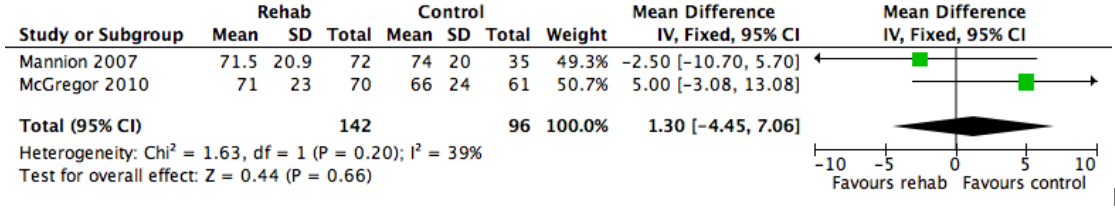
Forest plot of comparison: 1 Short term, outcome: 1.1 Functional status short term on log-scale



Forest plot of comparison: 1 Short term, outcome: 1.3 Low back pain short term on log-scale.

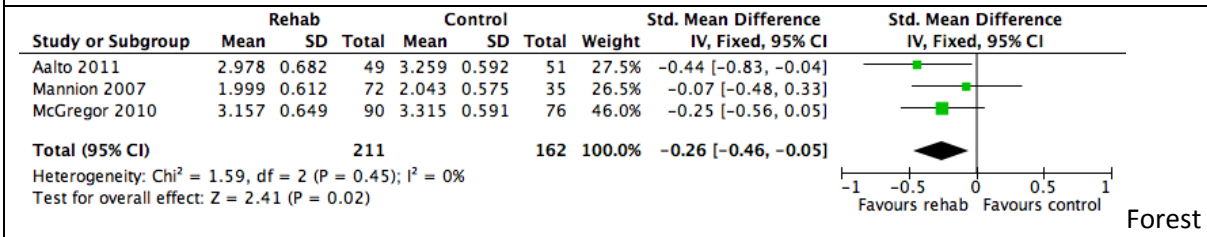


Forest plot of comparison: 1 Short term, outcome: 1.2 Leg pain short term on log-scale.



Forest plot of comparison: 1 Short term, outcome: 1.4 General health

Figure 3 Long term outcomes



Forest plot of comparison: 2 Long term, outcome: 2.1 Functional status long term on log-scale