Is surgery more effective than non-surgical treatment for spinal stenosis and which non-surgical treatment is more effective? a systematic review

MAY, Stephen and COMER, C

Available from Sheffield Hallam University Research Archive (SHURA) at:
http://shura.shu.ac.uk/6976/

This document is the author deposited version. You are advised to consult the publisher's version if you wish to cite from it.

Published version

MAY, Stephen and COMER, C (2013). Is surgery more effective than non-surgical treatment for spinal stenosis and which non-surgical treatment is more effective? a systematic review. Physiotherapy, 99 (1), 12-20.

Copyright and re-use policy

See http://shura.shu.ac.uk/information.html
Stephen May¹
Christine Comer²

Is surgery more effective than non-surgical treatment for spinal stenosis; and which non-surgical treatment is more effective? A systematic review.

¹ = corresponding author
Sheffield Hallam University, Collegiate Campus, Sheffield, S 10 2BP, UK
s.may@shu.ac.uk
0044 (0)114 2252370 (T) 0044 (0)114 225 5972 (F)

²
Section of Musculoskeletal Disease, Faculty of Medicine and Health, University of Leeds, Chapel Allerton Hospital, Chapeltown Road, Leeds, LS7 4SA, UK
c.comer@nhs.net

Word count: 3,132
Abstract

**Background**  Spinal stenosis can be treated both conservatively and with decompression surgery.

**Objectives**  This review explored the effectiveness of surgery versus conservative treatment and of conservative interventions for spinal stenosis.

**Data sources**  Medline, CINAHL, AMED, PEDro, and Cochrane databases, as well as the reference lists of retrieved studies, were searched.

**Study selection**  The search included non-English studies, and all conservative interventions were included.

**Study appraisal**  The PEDro scale was used to assess quality and levels of evidence were used to synthesise studies where possible.

**Results**  Thirty-one studies met the inclusion criteria; eighteen were high quality. Decompression surgery was more effective than conservative care in 4 out of 5 studies, but only one of these was high quality. In 6 high quality studies there was strong evidence that steroid epidural injections were not effective; in 4 out of 5 studies (2 high quality) there was moderate evidence that calcitonin was not effective. There was no evidence for the effectiveness of all other conservative interventions.

**Limitations**  Further research is needed to determine if decompression surgery is really more effective than conservative care, and which conservative care is most effective.

**Conclusion and implications**  Today there is no evidence that favours the effect of any conservative management for spinal stenosis. At present there is
urgent need to see if any conservative treatment can change pain and functional outcomes in spinal stenosis.

**Funding**  No funding was received for this systematic review.

**Keywords**  Spinal stenosis, Neurogenic claudication, Conservative treatments, Non-surgical management, Systematic review
Introduction

Spinal stenosis is a degenerative condition that affects the lumbar spine, which can be an incidental finding [1], but can cause back and leg symptoms, or neurogenic claudication. The prevalence of symptomatic spinal stenosis is likely to increase with growing cohorts of the old and very old. Stenosis in the lumbar spine is considered to have both a structural and a dynamic component [2]. With walking causing further narrowing of the spinal canals, and an increase in epidural pressure [3, 4]. Patients are typically more than 50 years old, with a long history of back pain, extensive degenerative changes on radiography, with major loss of extension that may provoke symptoms, and neurological deficit in some [1, 5, 6, 7].

Surgery for spinal stenosis has been on the increase, especially in the USA [8, 9]. Some good outcomes from surgical interventions have been demonstrated [10, 11, 12, 13, 14-16], but outcomes vary widely, and complications and re-operations have been reported [11, 17, 18, 19, 20, 21]. It has been concluded that there was limited evidence for decompression or fusion for spinal stenosis [22, 23].

The natural history of spinal stenosis and response to non-surgical care can be favourable [24-26], with numerous therapies being proposed [5, 25, 27, 28, 29, 30]. Physiotherapists in the UK favoured advice, and flexion and stabilisation exercises [31].
There are a number of systematic reviews into conservative care for spinal stenosis that are available, but most of these had limited remits. An evidence-based clinical guideline [32] did not give details about the studies on which it was based; two systematic reviews looked at calcitonin only [33,34]; one looked at chiropractic treatment, but with no randomised controlled trials (RCT) [35]; and one looked at manual therapy, with only one RCT included [36]. Two systematic reviews were published whilst we were preparing ours; one only included 13 RCTs when we had already located nearly 30 relevant trials [37]; and the other only compared surgery to conservative treatment for spinal stenosis, and included 5 RCTs [38]. Therefore the aim of our review was to conduct a systematic review of surgery versus non-surgical treatment and all non-surgical treatments for spinal stenosis.

Methods

Identification of studies

The following databases were searched: Medline (1966 to August 2010), Cinahl (1982 to August 2010), AMED (1985 to August 2010), PEDro (August 2010), Cochrane Central Register of Controlled trials (1st quarter 2010). All articles were identified through electronic searching using single terms, and then combinations of terms using Boolean operators. The reference lists of all retrieved articles were also searched. Search terms and combinations were as follows: spinal stenosis OR neurogenic claudication, AND non-operative treatment OR conservative treatment OR physical therapy OR physiotherapy OR rehabilitation, AND randomized controlled trial OR clinical trial. Abstracts and titles were scanned by both authors and decisions were independently
made whether to include article or not according to the inclusion / exclusion criteria. All selected articles went through to the next stage, when full articles were viewed, and again independent judgements were made about whether to include or exclude in the review. Figure 1 provides a flow diagram of the selection process.

Inclusion criteria were as follows:

- Human subjects
- Subjects with spinal stenosis and back (leg) pain
- Clinical or radiological diagnosis of spinal stenosis
- Randomized controlled trial
- Use of non-surgical treatment in one group.

**Study quality**

Study quality was judged using the PEDro scale, which is an 11-item scale designed to assess the methodological quality of randomized controlled trials (RCTs) (see foot of table 1 for criteria). It was decided to use the scores from the PEDro database, where points are only awarded if a criterion is clearly satisfied; if an item is not reported no points are awarded. The first criterion is not included in the final score, which is normal practice and so the total is out of 10; studies scoring 6 points or more were considered to be high quality [39] (see [http://www.pedro.fhs.usyd.edu.au/scale_item.html](http://www.pedro.fhs.usyd.edu.au/scale_item.html) for further detail).

When available, we used the score on the PEDro database; otherwise the studies were independently rated, and agreement and kappa value between reviewers were calculated.
**Data extraction and analysis**

Data extraction was done independently by the two reviewers using a pre-determined data extraction sheet. Differences were resolved at a consensus meeting. Pain and disability were determined to be the key outcome measures. Outcomes were considered short-term (less than 3 months after randomization), medium-term (between 3 and 12 months follow-up), and long-term (at least 12 months follow-up).

If there was insufficient homogeneity, in terms of interventions and outcomes, to justify a meta-analysis a qualitative synthesis using the following levels of evidence would be adopted from the Cochrane Collaboration Back Review Group [40]:

- **Strong** – consistent findings amongst multiple (3 or more) high-quality RCTs
- **Moderate** – consistent findings amongst multiple (3 or more) low-quality RCTs and/or controlled clinical trials (CCTs) and/or one high-quality RCT
- **Limited** – one low-quality RCT and/or CCT
- **Conflicting** – inconsistent findings amongst multiple trials (RCTs and/or CCTs)
- **No evidence** – no trials.

Initially 653 abstracts and titles were reviewed; 512 were clearly not relevant, and a further 103 abstracts did not meet inclusion criteria; leaving 43 articles for full inspection, of which 2 reviewers independently selected 32 for final
inclusion in the review (figure 1). Two exclusions (one in German and one in Turkish) followed discussion with German and Turkish colleagues. Two of those included were in Japanese and 2 were in Finnish. Colleagues translated the relevant sections, and one of the Finnish articles appeared to provide outcome data at the same points as a publication of the same study in English, so this was excluded; thus a total of 31 articles were included [10, 12-16, 41-65]. Seven articles had been reviewed by PEDro; 21 articles were independently scored by the 2 reviewers – agreement was 86%, kappa 0.67, with differences easily agreed at the consensus meeting. Two Japanese articles were scored by one reviewer only in Japanese [47, 60], and 1 Finnish article by one reviewer only in Finnish [57].

Results

Studies

Median PEDro score was 6.0 (interquartile range 5.0, 8.0), and 18 articles were deemed to be high quality (see Table 1 for PEDro scores). The commonest weaknesses were lack of blinding of clinicians and patients, lack of intention-to-treat analysis, and lack of concealed treatment allocation. Four articles were longer-term follow-up of research projects reported elsewhere [15, 16, 57, 65]; thus 27 separate trials were included in total. Ten trials had less than 3 months follow-up; 10 between 3 and 12 months; and 7 had follow-up greater than a year. The non-surgical interventions involved a wide range of physical therapy, medication and injection-related conservative care alternatives; likewise the control groups varied. After data extraction it was
determined that there was insufficient homogeneity in terms of interventions controls, patients, and outcomes to justify a meta-analysis.

Eight studies involved various physical therapy interventions. One study involved orthosis, back school and general physical training [10]; one NSAID, back school and individualised physical therapy [48]; one bed rest, orthosis and physical therapy [13]; one study involved use of a walking stick [41]; one ultrasound with exercise [45]; 2 studies included body-weight sustained treadmill [55,62]; one study used ‘usual care’ as a control group for decompression surgery [14-16].

Five studies involved calcitonin [43, 53, 54, 56, 58]; three studies involved prostaglandin [47, 51, 60]; one study involved methylcobalamin, a vitamin B12 [61], and one study involved gabapentin as an additional treatment on top of exercises, traction, corset and NSAID [63]. Nine studies involved different methods of epidural steroid injections with or without an anaesthetic [42, 44, 46, 48, 49, 50, 52, 59, 64, 65].

In five studies the conservative treatment was the control group for a surgical intervention [10, 12-16, 57, 64, 65]. Some studies used a minimal or no intervention, usual care or placebo as a control group [26, 41, 43, 45, 46, 53, 54], but the rest of the studies compared two or more conservative interventions. (see Table 2 for description and results of studies).

Outcomes
In comparisons between decompression surgery and a range of conservative interventions, the former produced statistically significant better results in 3 trials [12, 13, 64, 65]. There was no statistical analysis in Amundsen et al. [10], who reported on percentage improvements only, which were substantially better in the surgery group at all time points. Only one of these trials was high quality on the PEDro scale [12]. There were no statistical differences in virtually all outcomes in one trial [14-16, 57]. Follow-up in all these trials was very long-term at 2, 4, 6 or 10 years. Thus overall with a strict interpretation there was conflicting evidence between conservative treatment and surgery. However in the last trial by 4 years decompression surgery had occurred in 54% of those randomised to conservative treatment, compared to 66% in those randomised to surgery [16]. Given this degree of non-adherence to treatment protocol it is doubtful if the intention-to-treat results, given in table 2, can be taken at face value. In the as-treated analysis, which however included both randomised and observational cohorts, there were significantly better outcomes for those receiving surgery (p<0.001 in all primary and secondary outcomes) [16]. Given this it is probably truer to say there was moderate evidence that surgery is better than conservative treatment. There is moderate quality evidence (5 RCTS; n = 664) that surgery produced significantly better outcomes than conservative treatment in function, pain, and walking short and long-term.

Epidural injections produced no statistically significant differences compared to physical therapy, saline, saline and anaesthetic or anaesthetic injection at long-term follow-up in 6 high-quality trials [42, 44, 46, 49, 52, 59]; there were a
few minor significant differences short-term in one trial [44]. A bilateral transforaminal injection was more effective than an interlaminar steroid injection [48]. Percutaneous adhesiolysis and decompression surgery were more effective than an epidural steroid injection [50, 64, 65]. There is strong evidence (6 RCTS; $n = 239$) that epidural injections were no more effective than active controls.

In one study calcitonin was better than placebo short-term, but there were no differences long-term [43]. However, 4 studies (2 high quality) found no significant differences between calcitonin and placebo or paracetamol [34, 53, 56, 58]. In the one positive study patients were their own controls and reporting was not optimal [43]. Thus there is moderate evidence (5 RCTs; $n = 222$) that calcitonin was no more effective than placebo or paracetamol. All further studies have been considered individually as treatments could not be grouped in any meaningful way. Prostaglandin was more effective at high rather than low doses in one high quality study [47], but not in another high quality study [60], and more effective than an NSAID, but only at short-term in a low quality study [51]. Methylcobalamin, a vitamin B12, in addition to physical therapy and NSAID was better than placebo in one low quality study, but only in one outcome [61]. Gabapentin in addition to physical therapy and NSAID was more effective than physical therapy and NSAID alone, but follow-up was only to 4 months, and the study was low quality [63].

Physical therapy interventions were less effective than decompression surgery in 3 studies [10, 12, 13], but equal to decompression surgery in one
study [14-16], but see earlier discussion about the non-adherence to randomisation protocol. Exercise was more effective than a control group, but results were only very short-term; the addition of active ultrasound made no difference in one high quality study [45]. Body-weight sustained treadmill training was no more effective than a cycling programme, though both improved slightly in one high quality study [55]. However when used in conjunction with manual therapy and flexion exercises was more effective than flexion exercises, treadmill walking and ultrasound, though only short-term in another high quality study [62]. There were long-term differences favouring the body-weight sustained treadmill and manual therapy group and improvements in both groups in the numeric pain rating scale long-term, but none of these were significant. Use of a walking stick did not produce any improvement in one high quality study [41].

Discussion

In a systematic review of the literature into surgical versus non-surgical care and into different forms of non-surgical care for patients with spinal stenosis, 31 articles and 27 separate studies were found. Eighteen articles were deemed to be high quality.

Decompression surgery was better than non-surgical care in 4 out of 5 comparisons, but 3 of those 4 were not considered high quality, and one had no statistical analysis. In the study in which there were no significant differences between surgical and non-surgical care this was by intention-to-treat analysis [14-16, 57]; but nearly as many of those randomised to
conservative care received surgery as in the surgery group, and an as-treated analysis showed significant improvements in those treated surgically. Therefore we considered it justified to suggest that there was moderate evidence to support surgery over conservative care.

During the preparation of our manuscript another systematic review that compared surgery with conservative treatment was published, which considered the same RCTs, but also conducted a meta-analysis [38]. Their conclusions were similar to ours, with surgery showing better results for pain, disability and quality of life both short and long-term. As with our conclusions, the intention-to-treat analysis found minimal differences with an effect size of 0.35 \( (p=0.35) \); whereas the as-treated analysis found an effect size of 3.27 \( (p=0.001) \) [38].

There was strong evidence that epidural injections were no more effective than saline or anaesthetic in 5 out of 6 studies. There was moderate evidence that calcitonin was not effective in 5 studies. There was no consistent evidence to support the use of prostaglandin, B12, and gabapentin in 5 studies. There was no consistent long-term evidence to support any particular exercise programme. Except for the surgical studies the majority of RCTs that we included had a follow-up of less than year. Given the chronic nature of the problem, and the likely placebo effect of any treatment provided, there is limited validity for only short-term outcomes. However a programme of manual therapy, body-weight supported treadmill walking, and flexion and strengthening exercises was more effective than treadmill walking, flexion
exercises, and sub-therapeutic ultrasound [62], though another trial did not support the use of body-weight supported treadmill walking compared to cycling [55]. It maybe the manual therapy component that was the important element of this intervention, but as this was individualised the detail of what was done is not clear.

These generally negative conclusions are in line with other guidelines and reviews. Two guidelines [32] concluded that there was insufficient evidence to firmly recommend any intervention [32, 66]. Two previous reviews were equally negative about the use of calcitonin for use in spinal stenosis [33, 34], though one review suggested it may have transient benefit [37]. Another 2 reviews on chiropractic and manual therapy failed to find any supporting evidence [35, 36].

We used the PEDro scale to assess quality of the studies, which was developed from the Jadad and Delphi criteria, and has demonstrated moderate to substantial levels of reliability between raters [39]. It has been used in a number of previous systematic reviews to evaluate study quality in different musculoskeletal problems [67-69], with the latter 2 studies using the same criterion to judge quality. The PEDro scale has been shown to be a more comprehensive measure of methodological quality than the Jadad score in the stroke rehabilitation literature [70], and been demonstrated to be a valid measure of the methodological quality of clinical trials [71]. The criteria used are in fact very similar to other contemporary methods for judging quality criteria, such as the recent Cochrane Back Review Group criteria [72].
Despite which the use of slightly different criteria can lead to differences in judgements about quality, for instance between this and another review [38].

This large systematic review attempted to include all conservative treatment options offered by a range of healthcare providers. We conducted filtering of the references, PEDro scoring, and data extraction independently and settled disagreements by consensus. We were able to include non-English language studies, and thus avoided the risk of language bias. As with all reviews we could not avoid the risk of publication bias, but as this suggests that negative studies are less likely to be published, this would only re-enforce our conclusions.

The clinical implications of our findings are that decompression surgery is probably more effective than conservative care. Although we have included surgical treatment when the control group was some kind of non-surgical care it should be recognised that the population included in these surgical trials might a priori differ from the populations in the conservative treatment trials. The fact that they are likely to be worse at intake makes the success of surgical management and the relative failure of conservative management more likely. The decision to be treated with surgery or conservatively must lie with the clearly-informed patient, and will depend on symptom severity and functional impairment.

The other aspects of this debate that should be considered carefully are the costs and risks involved in surgical interventions. In the USA the cost of over
122,000 lumbar fusions for degenerative conditions was estimated at $4.8 billion in 2001 with every sign that numbers are still growing, especially in patients over 60 [8, 9]. Furthermore complication rates of up to 35% have been reported, and re-operation rates may be as high as 23% [11, 17, 18, 19, 20, 21].

The clinical implications of our findings are that it is difficult to firmly recommend any particular conservative therapy, though there is some evidence to support body-weight supported treadmill walking and manual therapy, this was not consistent.

We conducted a systematic review of the literature that compared different non-surgical treatments and surgical and non-surgical treatment for patients with spinal stenosis. We included 27 separate studies, 18 of which we considered high quality against the PEDro criteria. Surgery appeared to be more effective than non-surgical care when data was analysed in an as-treated way, but not when considered by an intention-to-treat analysis. From the evidence we reviewed steroid epidural injections and calcitonins were ineffective. There was no consistent evidence for the effectiveness of any other non-surgical treatments, including exercise and manual therapy.

Acknowledgements

Thanks to Filiz Can, Sinikka Kilpikoski, Georg Supp, and Hiroshi Takasaki for help in reviewing non-English papers.
Conflict of interest statement

The authors report no conflict of interest.

Funding

No funding was received for this work.

References


42. Cuckler JM, Bernini PA, Wiesel SW, Booth RE, Rothman RH, Pickens GT. The use of epidural steroids in the treatment of lumbar radicular


55. Pua YH, Cai CC, Lim KC. Treadmill walking with body weight support is no more effective than cycling when added to an exercise program


