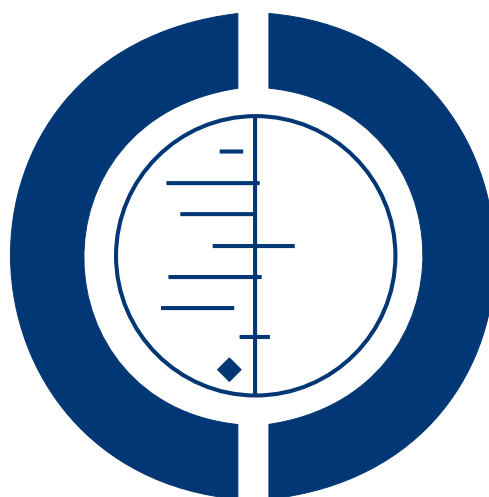


# Total disc replacement for chronic back pain in the presence of disc degeneration (Review)

Jacobs W, Van der Gaag NA, Tuschel A, de Kleuver M, Peul W, Verbout AJ, Oner FC



**THE COCHRANE  
COLLABORATION®**

This is a reprint of a Cochrane review, prepared and maintained by The Cochrane Collaboration and published in *The Cochrane Library* 2013, Issue 2

<http://www.thecochranelibrary.com>



---

Total disc replacement for chronic back pain in the presence of disc degeneration (Review)  
Copyright © 2013 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

## TABLE OF CONTENTS

HEADER . . . . .	1
ABSTRACT . . . . .	1
PLAIN LANGUAGE SUMMARY . . . . .	2
SUMMARY OF FINDINGS FOR THE MAIN COMPARISON . . . . .	3
BACKGROUND . . . . .	8
OBJECTIVES . . . . .	9
METHODS . . . . .	9
RESULTS . . . . .	12
Figure 1. . . . .	13
Figure 2. . . . .	15
Figure 3. . . . .	18
Figure 4. . . . .	19
Figure 5. . . . .	19
DISCUSSION . . . . .	21
AUTHORS' CONCLUSIONS . . . . .	24
ACKNOWLEDGEMENTS . . . . .	25
REFERENCES . . . . .	25
CHARACTERISTICS OF STUDIES . . . . .	30
DATA AND ANALYSES . . . . .	48
Analysis 1.1. Comparison 1 Disc replacement versus fusion, Outcome 1 Back pain at 6 Months. . . . .	49
Analysis 1.2. Comparison 1 Disc replacement versus fusion, Outcome 2 Back pain at 24 Months. . . . .	50
Analysis 1.3. Comparison 1 Disc replacement versus fusion, Outcome 3 Improvement in back pain at 24 months. . . . .	50
Analysis 1.4. Comparison 1 Disc replacement versus fusion, Outcome 4 Leg pain at 24 Months. . . . .	51
Analysis 1.5. Comparison 1 Disc replacement versus fusion, Outcome 5 Improvement in leg pain at 24 Months. . . . .	52
Analysis 1.6. Comparison 1 Disc replacement versus fusion, Outcome 6 Pain at 6 months. . . . .	52
Analysis 1.7. Comparison 1 Disc replacement versus fusion, Outcome 7 Pain at 24 months. . . . .	53
Analysis 1.8. Comparison 1 Disc replacement versus fusion, Outcome 8 Improvement in pain at 24 months. . . . .	53
Analysis 1.9. Comparison 1 Disc replacement versus fusion, Outcome 9 Overall improvement. . . . .	54
Analysis 1.10. Comparison 1 Disc replacement versus fusion, Outcome 10 Patient satisfaction. . . . .	55
Analysis 1.11. Comparison 1 Disc replacement versus fusion, Outcome 11 Oswestry at 6 Months. . . . .	56
Analysis 1.12. Comparison 1 Disc replacement versus fusion, Outcome 12 Oswestry at 24 Months. . . . .	56
Analysis 1.13. Comparison 1 Disc replacement versus fusion, Outcome 13 Improvement in Oswestry score at 24 months. . . . .	57
Analysis 1.14. Comparison 1 Disc replacement versus fusion, Outcome 14 Percentage of patients improved on Oswestry at 24 months. . . . .	58
Analysis 1.15. Comparison 1 Disc replacement versus fusion, Outcome 15 Improvement in working status at 24 months. . . . .	59
Analysis 1.16. Comparison 1 Disc replacement versus fusion, Outcome 16 Implant motion. . . . .	60
Analysis 1.17. Comparison 1 Disc replacement versus fusion, Outcome 17 Blood loss. . . . .	60
Analysis 1.18. Comparison 1 Disc replacement versus fusion, Outcome 18 Reoperations. . . . .	61
Analysis 1.19. Comparison 1 Disc replacement versus fusion, Outcome 19 Adjacent segment degeneration. . . . .	62
Analysis 1.20. Comparison 1 Disc replacement versus fusion, Outcome 20 Facet joint degeneration. . . . .	62
Analysis 1.21. Comparison 1 Disc replacement versus fusion, Outcome 21 Radiographic loosening (radiolucency). . . . .	63
Analysis 1.22. Comparison 1 Disc replacement versus fusion, Outcome 22 Subsidence. . . . .	63
Analysis 2.1. Comparison 2 Disc arthroplasty versus rehabilitation, Outcome 1 Back Pain at 12 months. . . . .	64
Analysis 2.2. Comparison 2 Disc arthroplasty versus rehabilitation, Outcome 2 Back Pain at 24 months. . . . .	65
Analysis 2.3. Comparison 2 Disc arthroplasty versus rehabilitation, Outcome 3 Oswestry at 12 months. . . . .	65
Analysis 2.4. Comparison 2 Disc arthroplasty versus rehabilitation, Outcome 4 Oswestry at 24 months. . . . .	66
Analysis 2.5. Comparison 2 Disc arthroplasty versus rehabilitation, Outcome 5 Patient satisfaction. . . . .	66
Analysis 2.6. Comparison 2 Disc arthroplasty versus rehabilitation, Outcome 6 Improvement in working status. . . . .	67
Analysis 2.7. Comparison 2 Disc arthroplasty versus rehabilitation, Outcome 7 Reoperations. . . . .	67
ADDITIONAL TABLES . . . . .	68
APPENDICES . . . . .	69

WHAT'S NEW . . . . .	77
CONTRIBUTIONS OF AUTHORS . . . . .	77
DECLARATIONS OF INTEREST . . . . .	78
SOURCES OF SUPPORT . . . . .	78
DIFFERENCES BETWEEN PROTOCOL AND REVIEW . . . . .	78
INDEX TERMS . . . . .	78

[Intervention Review]

# Total disc replacement for chronic back pain in the presence of disc degeneration

Wilco Jacobs<sup>1</sup>, Niels A Van der Gaag<sup>1</sup>, Alexander Tuschel<sup>2</sup>, Marinus de Kleuver<sup>3,4</sup>, Wilco Peul<sup>1</sup>, AJ Verbout<sup>5</sup>, F Cumhur Oner<sup>5</sup>

<sup>1</sup>Department of Neurosurgery, Leiden University Medical Center, Leiden, Netherlands. <sup>2</sup>Orthopaedic Hospital Vienna Speising, Vienna, Austria. <sup>3</sup>Sint Maartenskliniek, Nijmegen, Netherlands. <sup>4</sup>Department of Orthopedic Surgery, Vrije Universiteit (VUMC), Amsterdam, Netherlands. <sup>5</sup>Department of Orthopedics, University Medical Center Utrecht, Utrecht, Netherlands

Contact address: Wilco Jacobs, Department of Neurosurgery, Leiden University Medical Center, PO Box 9600, Leiden, 2300 RC, Netherlands. [w.c.h.jacobs@lumc.nl](mailto:w.c.h.jacobs@lumc.nl). [wch.jacobs@gmail.com](mailto:wch.jacobs@gmail.com).

**Editorial group:** Cochrane Back Group.

**Publication status and date:** Edited (no change to conclusions), published in Issue 2, 2013.

**Review content assessed as up-to-date:** 6 March 2012.

**Citation:** Jacobs W, Van der Gaag NA, Tuschel A, de Kleuver M, Peul W, Verbout AJ, Oner FC. Total disc replacement for chronic back pain in the presence of disc degeneration. *Cochrane Database of Systematic Reviews* 2012, Issue 9. Art. No.: CD008326. DOI: 10.1002/14651858.CD008326.pub2.

Copyright © 2013 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

## ABSTRACT

### Background

In the search for better surgical treatment of chronic low-back pain (LBP) in the presence of disc degeneration, total disc replacement has received increasing attention in recent years. A possible advantage of total disc replacement compared with fusion is maintained mobility at the operated level, which has been suggested to reduce the chance of adjacent segment degeneration.

### Objectives

The aim of this systematic review was to assess the effect of total disc replacement for chronic low-back pain in the presence of lumbar disc degeneration compared with other treatment options in terms of patient-centred improvement, motion preservation and adjacent segment degeneration.

### Search methods

A comprehensive search in Cochrane Back Review Group (CBRG) trials register, CENTRAL, MEDLINE, EMBASE, BIOSIS, ISI, and the FDA register was conducted. We also checked the reference lists and performed citation tracking of included studies.

### Selection criteria

We included randomised controlled trials (RCTs) comparing total disc replacement with any other intervention for degenerative disc disease.

### Data collection and analysis

We assessed risk of bias per study using the criteria of the CBRG. Quality of evidence was graded according to the GRADE approach. Two review authors independently selected studies and assessed risk of bias of the studies. Results and upper bounds of confidence intervals were compared against predefined clinically relevant differences.

## Main results

We included 40 publications, describing seven unique RCT's. The follow-up of the studies was 24 months, with only one extended to five years. Five studies had a low risk of bias, although there is a risk of bias in the included studies due to sponsoring and absence of any kind of blinding. One study compared disc replacement against rehabilitation and found a statistically significant advantage in favour of surgery, which, however, did not reach the predefined threshold for clinical relevance. Six studies compared disc replacement against fusion and found that the mean improvement in VAS back pain was 5.2 mm (of 100 mm) higher (two studies, 676 patients; 95% confidence interval (CI) 0.18 to 10.26) with a low quality of evidence while from the same studies leg pain showed no difference. The improvement of Oswestry score at 24 months in the disc replacement group was 4.27 points more than in the fusion group (five studies; 1207 patients; 95% CI 1.85 to 6.68) with a low quality of evidence. Both upper bounds of the confidence intervals for VAS back pain and Oswestry score were below the predefined clinically relevant difference. Choice of control group (circumferential or anterior fusion) did not appear to result in different outcomes.

## Authors' conclusions

Although statistically significant, the differences between disc replacement and conventional fusion surgery for degenerative disc disease were not beyond the generally accepted clinically important differences with respect to short-term pain relief, disability and Quality of Life. Moreover, these analyses only represent a highly selected population. The primary goal of prevention of adjacent level disease and facet joint degeneration by using total disc replacement, as noted by the manufacturers and distributors, was not properly assessed and not a research question at all. Unfortunately, evidence from observational studies could not be used because of the high risk of bias, while these could have improved external validity assessment of complications in less selected patient groups. Non-randomised studies should however be very clear about patient selection and should incorporate independent, blinded outcome assessment, which was not the case in the excluded studies. Therefore, because we believe that harm and complications may occur after years, we believe that the spine surgery community should be prudent about adopting this technology on a large scale, despite the fact that total disc replacement seems to be effective in treating low-back pain in selected patients, and in the short term is at least equivalent to fusion surgery.

## PLAIN LANGUAGE SUMMARY

### Total disc replacement for chronic low-back pain

Artificial joints have had a remarkable impact in reducing pain and improving function in the knee and hip. But what about joint replacement in the spine? This systematic review set out to determine how total disc replacement compares with other treatments for chronic low-back pain in randomised controlled trials.

The disc is a strong but flexible structure that cushions and separates the bony vertebrae of the spine. Disc degeneration is a nearly universal feature of the aging process. Though there are many theories about the causes of low-back pain, chronic symptoms are often attributed to disc degeneration. So when patients do not improve after nonsurgical care, they sometimes consider surgery to remove a degenerated disc.

The traditional surgical approach in this area is spinal fusion, which involves removing all or part of a degenerated disc and joining the vertebrae above and below it. Previous reviews suggest that fusion surgery can lead to moderate pain relief and modest gains in function. It appears to be superior to traditional physical therapy-but no better than an intensive rehabilitation program.

An alternative surgical approach is total disc replacement, which involves removing the disc and replacing it with an artificial implant made out of metal and plastic.

We identified seven randomised trials-involving a total of 1474 patients. Only one study compared total disc replacement with nonsurgical treatment, suggesting that surgery resulted in slightly better outcomes than intensive rehabilitation. But this did not translate into a *clinically significant* advantage that would make a major difference in patients' lives.

Six randomised trials compared disc replacement with spinal fusion surgery. Most of these studies had a high potential for bias, raising the possibility that they might not have provided a fair test of the treatments under study. These trials found that patients who underwent total disc replacement had slightly better outcomes in terms of back pain and function than those who had fusion surgery. But again the differences did not appear clinically significant.

The review could not find evidence of any other benefits of total disc replacement, and the studies provided no insights on the long-term risks associated with it. Given the gaps in the evidence, the review concluded that the spine surgery community should be prudent about adopting this technology on a large scale.

## SUMMARY OF FINDINGS FOR THE MAIN COMPARISON *[Explanation]*

Disc arthroplasty versus fusion for chronic low-back pain							
Patient or population: patients with chronic low-back pain Intervention: Disc arthroplasty versus fusion							
Outcomes	Comparisons		Relative effect (95% CI)	No of Participants (studies)	Quality of the evidence (GRADE)	Comments	
	Assumed risk / average estimate in control group	Corresponding values in interventional group (95% CI)					
	Fusion	Disc arthroplasty					
<b>Improvement in Back Pain</b> VAS 100 mm scale Follow-up: mean 24 months	The mean improvement in back pain at 24 months in the control groups was <b>42.5 mm</b>	The mean improvement in back pain at 24 months in the intervention groups was <b>5.22 mm better</b> (10.26 to 0.18)		676 (2 studies)	⊕⊕○○ <b>Low<sup>1</sup></b>	This difference is not clinically relevant as it is below the predefined clinically relevant difference of 15 mm	
<b>Patient satisfaction</b> Questionnaire Follow-up: mean 24 months	In the fusion group, 213 of 309 ( <b>69 of 100</b> ) patients were satisfied.	In the disc replacement group this was 529 of 649 ( <b>82 of 100</b> ) patients	<b>OR 1.93</b> (1.36 to 2.76)	958 (4 studies)	⊕⊕⊕○ <b>moderate<sup>2</sup></b>		
<b>Improvement in function</b> Oswestry score. Scale from: -100 to 100. Follow-up: mean 24 months	The mean improvement in Oswestry score at 24 months in the control groups was <b>29.2 points</b>	The mean improvement in Oswestry score at 24 months in the intervention groups was <b>4.27 better (lower)</b> (6.68 lower to 1.65 lower)		1207 (5 studies)	⊕⊕⊕○ <b>moderate<sup>3</sup></b>	This difference is not clinically relevant as it is below the predefined clinically relevant difference of 10 points.	

<p><b>Percentage of patients improved on function</b> Oswestry score Follow-up: mean 24 months</p>	<p>In the fusion group, 233 of 407 (<b>57 of 100</b>) patients improved on the Oswestry score</p>	<p><b>OR 1.45</b> (1.06 to 1.98)</p>	<p>1244 (5 studies)</p>	<p>⊕⊕○○ <b>low</b><sup>4</sup></p>
<p><b>Disability</b> Roland Morris</p>				<p><b>No evidence</b> Not measured</p>
<p><b>Physical and mental health</b> SF-36, SF-12</p>		<p>863 (2 studies)</p>	<p>⊕○○○ <b>Very low</b></p>	<p>Report of SF-36 differs per study. One study reports percentage improvements from baseline, the other reports average values. Percentage improvements from baseline were not significantly different (79.2% versus 70%), but average improvements were significant (17.0 versus 14.3; p=0.009)</p>
<p><b>Economic status</b> Euro-QoL; EQ-5D</p>		<p>152 (1 study)</p>	<p>⊕○○○ <b>Very low</b></p>	<p>EQ-5D used in only one study, no difference was found (0.67 +/- 0.33 for disc replacement versus 0.69 +/- 0.25 for fusion)</p>
<p><b>Daily activity</b></p>				<p><b>No evidence</b> Not measured</p>
<p><b>Return to Work</b></p>		<p>1188 (4 studies)</p>	<p>⊕○○○ <b>Very low</b><sup>5</sup></p>	<p>Not analyzed due to very low quality evidence</p>



**Adverse Effects**

⊕○○○  
**Very low**

or

**No evidence**  
(Various  
outcomes)

Thromboembolic complications (2 studies). 2 venous thromboses in the disc replacement group and none in the fusion group. 1 cardiovascular event in the disc replacement group and none in the fusion group  
Blood loss (5 studies): Mean difference -37 ml (-185 to 111) favouring disc replacement, not statistically different  
Reoperations (5 studies) : OR 0.80 (0.51 to 1.24) There were 63 of 810 (7.8%) re-operations in the total disc replacement group and 35 of 384 (9.1%) in the fusion group  
Neurological complications (1 study): no statistical difference  
Adjacent segment degeneration (1 study): six of 72 cases of fusion and only one of 80 cases of total disc replacement with adjacent segment problems  
Facet joint degeneration (1 study): no statistical difference

		Radiographic loosening (0 studies). Subsidence (0 studies). Polyethylene wear (0 studies).
--	--	--------------------------------------------------------------------------------------------------

CI: Confidence interval; **OR**: Odds ratio; **VAS**: Visual analogue scale

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.

<sup>1</sup> Suspected publication bias and imprecision

<sup>2</sup> Two studies found a significant difference; two other could not find a significant difference. Three studies favoured disc replacement and one favoured fusion.

<sup>3</sup> The quality of evidence had to be decreased because of inconsistency of the results as one found significant effects and four found no significant effects.

<sup>4</sup> The level of evidence had to be decreased because of inconsistency; two studies found a significant difference, three found no significant difference with one an estimate on the opposite side. The level of evidence was decreased because of imprecision; the CI is large

<sup>5</sup> The level of evidence had to be decreased because of inconsistency, possible reporting bias and imprecision.

## BACKGROUND

In the search for better surgical treatment of chronic low-back pain, total disc replacement has received increasing attention in recent years, especially at meetings of experts in this field and in non-peer-reviewed publications. In 2003, our group performed a systematic review on total disc replacement (de Kleuver 2003), but at that time, there were no randomised controlled trials. Since we published our findings, two more reviews with meta-analyses have been published by Yajun 2010 including five randomised studies and Van den Eerenbeemt 2010 including three randomised trials and 16 cohort studies. In addition, three studies were identified in the review on spondylosis by Gibson and Waddell (Gibson 2005; Gibson 2005a). Recently, additional trials have come to our attention, creating the necessity to update the previous review with this new evidence. In this review, we will convey the rationale for total disc replacement, present the results of a systematic review of the literature and attempt to summarise the findings on the current status of total disc replacement devices.

### Description of the condition

Chronic low-back pain is believed to have a multifactorial origin of somatic pain sources, accompanied by psychogenic and social aspects. A degenerating intervertebral disc is frequently incriminated as the most important somatic pain source in the lumbar spine (Coppes 1997; Kuslich 1991; Nachemson 1996). Degeneration or repetitive straining of the facet joints, two supporting joints, one at each side, at the back of the vertebral body, also has frequently been depicted as an important contributing source of pain (Ashton 1992; Cavanaugh 1996). The only available conventional surgical treatment to prevent further deterioration is surgically fusing the two vertebral bodies, i.e. spinal fusion, spondylosis or arthrodesis. However, the clinical outcomes of this procedure are variable, leading to continuing controversy about which patients might benefit from the procedure. The dispute is further strengthened by the difficulty to properly diagnose the disc as the source of pain. One prospective randomised trial showed significantly better clinical outcomes of fusion in comparison to non-operative treatment (Fritzell 2001). A large literature survey comprising 5600 patients (Boos 1997) reported a "satisfactory outcome" in 65% to 87% of the patients after a fusion, with fusions of only one motion segment having better results compared with fusions of multiple segments. However, as in other fields of orthopaedic surgery, fusion is generally not considered the optimal solution to treat degenerative disease, due to the increased stresses and subsequent degeneration at the adjacent joints.

### Description of the intervention

The development of artificial joint prostheses has almost obviated the need for fusion of the painfully and degenerated major peripheral joints. Similarly, in the lumbosacral spine, research has

concentrated on the development of mobile intervertebral artificial joints to retain motion and thereby possibly avoid the disadvantages of fusion. These disc prostheses are positioned anteriorly between the vertebral bodies, the two posterior facet joints are generally not replaced or removed. The first disc arthroplasty model, based on a steel ball, was described by Fernstrom 1966, but quickly abandoned due to poor outcome.

There are currently several replacements for the entire intervertebral disc on the market, two of which, the Charité and Prodisc L, are approved by the FDA. The Charité (DePuy Spine, Johnson&Johnson (DePuy 2009)) has been available since 1987 and was previously named Link SB Charité, developed by Schelnack and Büttner-Janzen (Buttner-Janzen 1987; David 2007; Lemaire 2005; Putzier 2006; Ross 2007). The Prodisc L (Synthes Spine (Kafer 2008; Synthes Spine 2009)), previously known as the ProDisc II (Bertagnoli 2005; Chung 2006; Delamarter 2005; Siepe 2006; Tropiano 2005) has been available since 1990. Other devices that are in the investigational stage are the Maverick (Medtronic, Sofamor Danek (Erkan 2009; Le Huec 2005)), the Flexicore (Stryker Spine), Activ-L (Aesculap (Ha 2009)), Nubac (Pioneer Surgical Technology (Alpizar-Aguirre 2008; Pioneer 2009a)), and Mobidisc (LDR Medical, Cedex 9 (LDR 2009)). Some devices appear to have been discontinued through market acquisitions, such as the Acroflex (AcroMed, purchased by DepPuy (Fraser 2004)).

The devices differ in mobility and other design aspects. Charité has the most extensive (least constrained) movement possibilities, ProDisc, Mobidisc and Maverick have more constraints and Flexicore has the least movement. Charité, Pro-disc and Mobidisc devices have low friction sliding surfaces, somewhat like a ball and socket joint, consisting of metal endplates with a polyethylene spacer. They are marketed as being developed from the successful total hip and knee arthroplasty materials. The Maverick and Flexicore are metal on metal implants. The metal endplates are meant to fuse to the vertebral endplates, much as the cup attaches to the pelvic bone in uncemented total hip arthroplasty. Supplementary fixation is provided on the SB Charité III by six teeth on each endplate, and on the Prodisc and Mobidisc by a fin on each endplate, which is sunk into the adjacent vertebral bodies.

### How the intervention might work

A possible advantage of total disc replacement is maintained mobility at the operated level, thus preventing adjacent segment degeneration. Leivseth 2006 evaluated ProDisc II and found decreased mobility at the surgical level. SariAli 2006 found normal kinematics with SB Charité prostheses, but only for one-level procedures. Huang 2006 found increased adjacent segment mobility for cases with decreased mobility at the surgical level. In conclusion, in a literature review on adjacent segment degeneration, Harrop 2008 showed more adjacent segment degeneration after fusion procedures than after total disc replacement, but the studies supporting this were of low quality. Using a computer model,

Chen 2009 predicts an increased instability at the surgical level, but not at the adjacent levels. In yet another computer model, Zander 2009 predicts decreased mobility of the segment, but no influence on adjacent segments. In conclusion, most short-term studies confirm preservation of motion to some degree and there are indications for less radiological degeneration of the adjacent segments when disc prosthesis are compared with spinal fusion.

### Possible disadvantages of the intervention

At the operated level, Park 2008 found progression of degeneration in the facet joints with ProDisc II and Rousseau 2006 showed increased facet joint loading in human cadavers. Goel 2005 and Rohlmann 2009 both predict an increase in facet joint loading. This might be present, especially in constrained devices (Moumene 2007). Rundell 2008 predicts from another computer model that this increase can be limited by posterior placement of the disc replacement in the intervertebral space. In a computer-aided design (CAD) model, Chung 2009 found that when constrained devices are used, the loading on the facet joints increases.

As total disc replacement is a relatively new procedure with limited evidence, salvage options such as conversion to a fusion are important factors when deciding in favour of treatment with total disc replacement. This is of major concern in these patients as revisions through the anterior approach carry large risks. Punt 2008 shows that removal of the damaged disc, with anterior and posterior fusion gives better outcomes than posterior fusion alone in salvage procedures. Complications following removal have been identified (de Maat 2009), and indeed salvage procedures are being investigated (Cunningham 2009). In a literature review, Patel 2008 points to the need for proper patient selection for initial disc replacement to avoid revision.

### Why it is important to do this review

Replacing an intervertebral disc in patients with chronic low-back pain with a total disc replacement raises the following questions, which need to be answered before large-scale implementation can be accepted.

1. How do the clinical results compare to fusion, the only surgical “gold standard” available?
2. How do the clinical results compare to other treatment options?
3. What are the radiologic results with regards to loosening, subsidence (slippage) of the implant into the vertebral bodies, and wear?
4. Does the motion segment retain its mobility? And if it does, do these motions resemble a normal motion segment?
5. Can the total disc replacement reduce the incidence of adjacent segment degeneration compared to fusion?
6. What is the incidence of facet joint degeneration at the operated level?

7. How does the perioperative complication rate compare to fusion operations?

8. Is there an acceptable and safe salvage procedure in case of failure?

In an attempt to answer these questions we performed this systematic literature review. Question 1 and 2 were the primary goals, while the other questions were of secondary interest.

## OBJECTIVES

The aim of this review was to assess the effect of total disc replacement for lumbar degenerative disc disease compared with fusion or other treatment options.

## METHODS

### Criteria for considering studies for this review

#### Types of studies

Primary studies for the review were randomised controlled trials (RCTs).

#### Types of participants

We included studies that involved patients scheduled for surgery for chronic (lasting longer than 12 weeks) degenerative disc disease. We did not apply exclusion criteria for age or gender of the populations, or type, location or duration of symptoms.

#### Types of interventions

The interventions evaluated in the trials were total disc replacement compared with any other treatment for lumbar degenerative disc disease.

#### Types of outcome measures

The required outcome parameters in the studies were clinical or functional. The primary outcome variable of our review was pain. Below is an indication of the anticipated outcome measures, but we made no exclusions on the type of outcome measure. Where outcome measures were composite, we analysed item scores as far as appropriate and possible, but analyses were also performed at the composite level. The minimal required duration of follow-up was six months.

## Primary outcomes

Patient-centred outcomes were of primary interest in this review. Short-term follow-up was defined as immediately postoperative to five years, long-term follow-up was defined as five years and longer. The following outcomes were prospectively included:

- Symptoms
  - Pain
  - Pain medication usage
- Overall improvement
- Patient Satisfaction
- Back-specific functional status
  - Oswestry
  - Roland-Morris
- Well being, quality of life
  - SF-36
  - SF-12
  - Euro-QOL
  - Specific disability questionnaires
    - ◇ Daily activity
    - ◇ Return to work

## Secondary outcomes

- Motion segment mobility
- General complications
  - Death
  - Hospitalisation
- General perioperative complications
  - Thrombo-embolic complications
  - Bleeding
- Surgical re-intervention
- Treatment specific complications
  - Adjacent segment degeneration
  - Facet joint degeneration at the operated level
  - Donor site pain (if appropriate)
  - Radiologic findings:
    - ◇ Loosening (In total hip arthroplasty radiologic loosening is recognised as a precursor of clinical loosening)
    - ◇ Subsidence of the implant into the vertebral bodies
    - ◇ Polyethylene wear

## Search methods for identification of studies

### Electronic searches

The following databases were electronically searched up to 22 December 2011:

- the Cochrane Central Register of Controlled Trials (CENTRAL) (*The Cochrane Library*)

- CBRG trials register
- MEDLINE (from 1966)
- EMBASE (from 1980)
- BIOSIS (from 2004)

Search strings for MEDLINE, EMBASE and ISI are given in [Appendix 1](#) and are adapted for the additional databases. A sensitive search was performed because terminology in the field of total disc replacement is not yet standardised. During the search, no restriction was made with regard to language or date. The search was performed by the Cochrane Back Review Group (CBRG) Trials Search Co-ordinator (RC).

### Searching other resources

We searched the FDA register and we contacted experts in the field. References and citation tracking results of selected articles were included in the search. We also searched ClinicalTrials.gov up to 22 December 2011 for ongoing or completed trials.

## Data collection and analysis

### Selection of studies

Two review authors (WJ, AT) independently selected the articles from the list of identified references. Consensus was sought, but when no consensus could be reached, a third review author (MK) was consulted. If relevance could not be excluded on the basis of the abstract, we retrieved the complete article. The final decision on inclusion was made upon review of the full text. Studies were selected when the [Criteria for considering studies for this review](#) were met and the study was published in a peer-reviewed journal as a full article, excluding grey literature and conference proceedings.

### Data extraction and management

Data were extracted onto separate, pre-developed forms. From each study, we gathered basic information concerning authors (affiliation, sponsoring), methods (study design, sample size), patients (selection criteria and diagnoses, pain location, age, gender), treatments (implant models, constraints, materials, levels involved), control treatments (non-surgical versus surgical, fusion versus non-fusion, anterior column versus posterior or circumferential), and outcome variables with results. One review author (WJ) extracted the data.

### Assessment of risk of bias in included studies

Two review authors (WJ, NvdG) independently assessed the risk of bias of the selected articles. Again, consensus was sought, but if necessary, a third review author was consulted (MK). Risk of bias of RCTs was assessed with the 12 criteria recommended by

the Cochrane Back Review Group (Furlan 2009). Criteria and operationalisation are given in Appendix 2. The items were scored as 'low risk', 'high risk', or 'unclear risk'. We categorised studies as having a "low risk of bias" when at least six of the 12 criteria were met and the study had no serious methodological flaws. Examples of serious methodological flaws are invalidating extensive loss to follow-up (> 20% for the whole study) or invalid trial stop after interim analysis.

### Measures of treatment effect

For dichotomous outcomes, we calculated odds ratios (OR). For continuous outcomes, we calculated the mean difference (MD). For each outcome, a 95% confidence interval (95% CI) was computed. A random-effects meta-analysis was used for all analyses. Clinical relevance was assessed with the five questions as recommended by Furlan 2009 (see Appendix 3). We evaluated clinically important changes using the guideline given by Ostelo 2008. The minimal clinically important difference of the primary outcome measurements (i.e. functional disability, perceived recovery and leg pain) was defined as a 30% improvement from baseline. This corresponds to mean difference of 15 for the Visual Analogue scale (VAS) (0-100), 2 of the Numerical rating Scale (0-10), 5 of the Roland Disability Questionnaire (0-24), 10 of the Oswestry Disability Questionnaire (0-100) and 20 for the Quebec Back Pain Disability Questionnaire (0-100). In case of absence of a clinical relevant point estimate difference, we also evaluated whether the upper bound of the 95% CI was smaller than this difference. If available, the analyses were performed on *improvements* of outcomes as these account for possible baseline differences of the specified outcome between and within the studies. We evaluated the primary outcomes at the longest follow-up term available, in this case 24 months.

### Unit of analysis issues

The unit of analysis was the patient, regardless of the number of disc arthroplasties that were performed. If allowed by the presented data, subgroup analysis was planned for one- and two-level surgeries.

### Dealing with missing data

We contacted the principle investigators of included trials by email with a request for clarifications about items with an unclear or high risk of bias. If responses were received, the two review authors responsible for the 'Risk of bias' assessment again met to discuss the impact of the clarifications on the 'Risk of bias' assessment and made corrections accordingly. Missing clinical data in trials were accepted when this was less than 20%, otherwise, the trial was excluded from the specific analysis of the outcome parameters and follow-up time points where missing data exceeded this requirement. Missing information about parameter variability was

estimated from ranges if provided or estimated from comparable trials.

Secondary publications of those studies with single-centre results were not used for analysis as this can be regarded as incomplete data reporting. We excluded from the analysis further secondary publications from multiple centres where the included sample represented less than 80% of the initial randomised sample.

### Assessment of heterogeneity

Heterogeneity between studies was first assessed clinically and then statistically. Clinical heterogeneity was evaluated for study design (allocation concealment, outcome assessor blinding, patient blinding), patient characteristics (pain location, levels involved, age, gender), device design (constraint, material), and control intervention (non-surgical versus surgical, fusion versus non-fusion, anterior column versus posterior or circumferential) variability. When studies were judged to be clinically homogeneous, statistical homogeneity was also tested with a  $I^2$ -test. We used the definition from the *Cochrane Handbook for Systematic Reviews of Interventions* (Chapter 9.5.2) (Higgins 2011).

- 0% to 40%: might not be important
- 30% to 60%: May represent moderate heterogeneity
- 50% to 90%: May represent substantial heterogeneity
- 75% to 100%: Considerable heterogeneity

### Assessment of reporting biases

We planned to use a funnel plot to identify publication bias. Also, we investigated inconsistencies between the FDA trial registries and published reports of trials. In general, we expected studies comparing interventions for low-back pain to assess at least pain, Oswestry (or similar) scores and to evaluate mobility/fusion of the motion segment. If there were too many studies without these parameters, reporting bias was suspected, unless confirmed by a full protocol.

### Data synthesis

The results from individual studies were pooled when the studies were judged to be sufficiently homogeneous after the assessment of heterogeneity (clinical and statistical).

We evaluated the quality of evidence for all outcome parameters, regardless of quantitative analysis, using the GRADE approach (GRADE Working Group 2004) and GRADE Profiler software, most current version). In short, the quality of evidence was judged from the presence of the following criteria (adapted from Furlan 2009).

- No risk of bias: 75% of studies with a low risk of bias (six or more items met, no serious methodological flaws)
- No inconsistency: Studies have consistent findings
- No indirectness: Results are based on direct comparison
- No imprecision: Estimate of effect is sufficiently precise (resulting from more than one trial, confidence interval narrow

and conclusive, and more than 75% of studies contributing to analysis)

- No publication bias: Analysis is free of reporting or publication bias

Depending on how many criteria were met, we judged the quality of evidence to be 'High' (all criteria), 'Moderate' (all but one), 'Low' (all but two) or 'Very Low' (all but three). Important outcomes for which there were no trials were considered to have 'no evidence'. An outcome with only one trial was automatically low quality and if it also had a high risk of bias, it dropped to very low quality. The clinical relevance of the review results was assessed with the five questions given in [Appendix 3](#). The results of this assessment were used to help in the discussion of the final results and conclusions.

### **Subgroup analysis and investigation of heterogeneity**

When heterogeneity existed, we performed post-hoc subgroup analyses and sensitivity analyses to explore the reason for heterogeneity. Analyses were performed for study design, (allocation concealment, outcome assessor blinding, patient blinding), patient characteristics (pain location, levels involved, age, gender) and implant design (constraint, material).

### **Sensitivity analysis**

Sensitivity analyses were planned to assess the effect of risk of bias (high or low) on outcomes (primary, secondary and complications).

## **RESULTS**

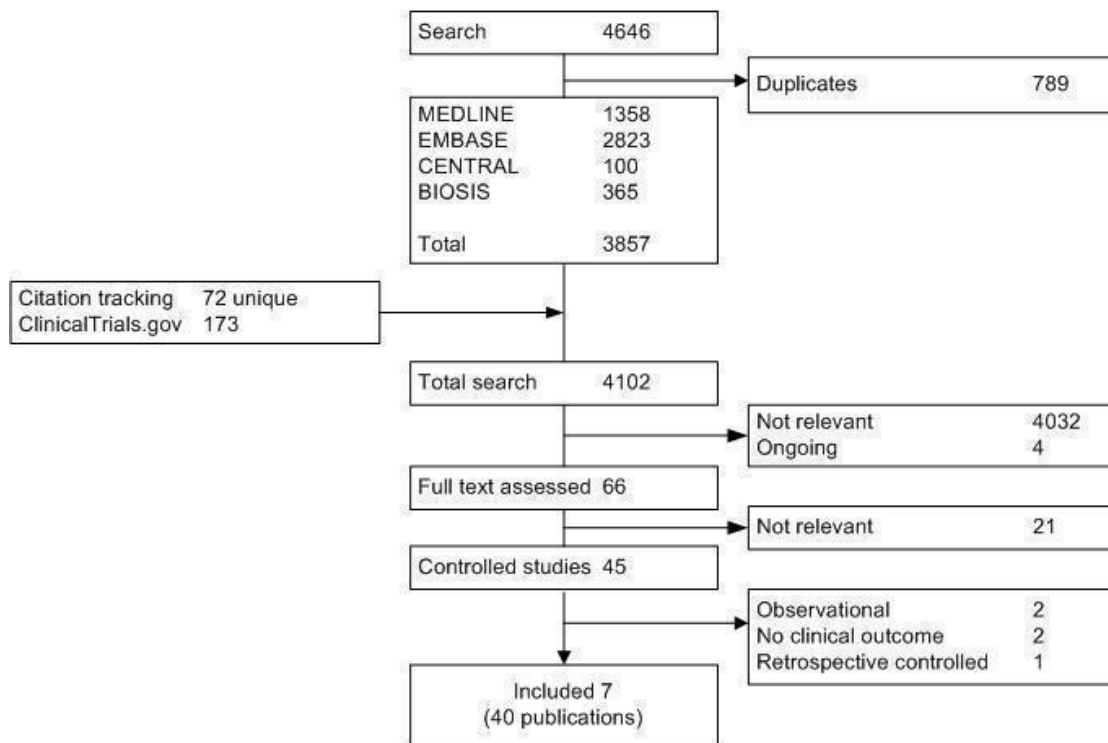
### **Description of studies**

See: [Characteristics of included studies](#); [Characteristics of excluded studies](#); [Characteristics of ongoing studies](#).

### **Results of the search**

The search in the databases resulted in 3857 references (MEDLINE 1358, EMBASE 2823, BIOSIS 365; CENTRAL 100, minus 789 duplicates). Kappa scores for the selection process were between 0.4 and 0.7 (moderate). We found 40 publications meeting our inclusion criteria, describing seven unique studies (see [Figure 1](#)). Four FDA approved Investigational Device Exemption studies were reported in one, two, 12 and 15 publications. Searching ClinicalTrials.gov resulted in four ongoing trials listed in the Ongoing studies section, but no additional completed trials. Citation tracking was performed on January 9, 2012 and of 195 cites, 70 were unique additional references but no new studies were found. For characteristics of studies see [Characteristics of included studies](#), [Characteristics of excluded studies](#) and [Table 1](#).

Figure 1.



## Included studies

### Methods

Three studies were non-inferiority trials (Blumenthal 2005; Gornet 2011; Zigler 2007) and one study was a superiority trial (Hellum 2011). In two studies, design and hypothesis were not described regarding superiority or non-inferiority, but these studies tested for superiority (Berg 2009; Moreno 2008). In one study, design and hypothesis was also not described regarding superiority or non-inferiority, but for this study with only single-centre, preliminary results, the clinical outcomes were not statistically tested (Sasso 2008).

### Participants

All studies included adult patients with back or leg pain or both in the presence of Degenerative Disc Disease (DDD) unresponsive to conservative treatment for at least three or six months. All studies included patients with one affected level, except for Berg 2009 which had 51% one-level procedures, and Hellum 2011 which included an unclear amount of one and two affected levels.

Zigler 2007 separately reported one- and two-level procedures. Five studies (Berg 2009; Blumenthal 2005; Gornet 2011; Sasso 2008; Zigler 2007) mentioned facet joint arthritis/degeneration as an exclusion criterion.

### Interventions

Four types of total disc replacement were used in the studies, Charite (Charite III, De Puy Spine, Raynham, MA, USA), Prodisc-L (Prodisc L or Prodisc II, Synthes Spine, West Chester, PA, USA), Maveric (MAVERICK Disc, Medtronic, Memphis, TN, USA), and Flexicore (Flexicore Intervertebral Disc, Stryker Spine, Allendale, NJ, USA).

Six studies (Berg 2009; Blumenthal 2005; Gornet 2011; Moreno 2008; Sasso 2008; Zigler 2007) compared total disc replacement with fusion. One study, Hellum 2011, compared Prodisc against a rehabilitation protocol with cognitive treatment and physical therapy.

### Outcomes

All studies used the Oswestry Disability Index (ODI) and a VAS pain score, specified as VAS back pain in three studies and completed with a VAS leg pain in one study. SF-36 was used in three



studies. Follow-up was up to 24 months for all studies with one study with a report of five-year results ([Blumenthal 2005](#)).

### **Excluded studies**

In the published version of the protocol for this review, we anticipated including non-randomised studies. The search and selection strategy as well as the 'Risk of bias' assessment were thus also aimed at evaluating observational studies. In addition to two observational concurrently controlled studies ([Cakir 2006](#); [Schroven 2006](#)), we excluded two studies that only looked at kinematics and did not report clinical outcome ([SariAli 2006](#); [Tournier 2007](#)), and one study that was a retrospective study with a comparison between two arthroplasty devices ([Shim 2007](#)). [Cakir 2006](#) et al specifically defined different selection criteria for the two inter-

ventions and this consequently resulted in two different groups with associated distinct age categories. Details of these studies are listed in [Characteristics of excluded studies](#).

### **Risk of bias in included studies**

We contacted all (seven) principle investigators of the trials for clarifications about unclear or high 'Risk of bias' items. Four responded and their clarifications together with the 'Risk of bias' assessment adjustments are shown in the 'Risk of bias' tables in [Characteristics of included studies](#) (denoted: "investigator response:"). The clarifications resulted in an upgrade of four items with an "unclear" to "low risk of bias". These four studies had in total 21 items with "high" or "unclear" risk of bias. The final risk of bias for the studies is given in [Figure 2](#).

Figure 2.

	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding (performance bias and detection bias): All outcomes - Participants?	Blinding (performance bias and detection bias): All outcomes - Care providers?	Blinding (performance bias and detection bias): All outcomes - Observers?	Incomplete outcome data (attrition bias): All outcomes - drop-outs?	Incomplete outcome data (attrition bias): All outcomes - ITT analysis?	Selective reporting (reporting bias)	Characteristics similar at baseline?	Co-interventions avoided or similar?	Compliance acceptable?	Timing of outcome assessments similar?
Berg 2009	+	+	-	-	-	+	+	?	-	+	+	+
Blumenthal 2005	+	+	-	-	-	+	+	?	+	+	+	+
Gornet 2011	+	+	?	-	?	+	+	?	+	+	+	+
Hellum 2011	+	?	-	-	-	+	+	?	-	+	+	+
Moreno 2008	?	?	-	-	-	+	+	?	?	+	+	-
Sasso 2008	?	?	?	-	?	-	?	?	?	+	+	+
Zigler 2007	+	-	-	-	?	+	+	?	+	+	+	+

## Allocation

Five of the included studies (Berg 2009; Blumenthal 2005; Gornet 2011; Hellum 2011; Zigler 2007) used adequate randomisation techniques, generated by computer programs or scrambled in a box.

Two studies used sequentially numbered, sealed envelopes provided by the sponsor to attempt allocation concealment (Blumenthal 2005; Gornet 2011). In one study, these were opened after informed consent, and consecutively; only one case was excluded from the intervention (Gornet 2011). In the other study, these were opened the day before surgery, and presumably no patients dropped out or crossed over before the intervention (Blumenthal 2005). One study only mentions sealed envelope “with no possibility to see through” opened after informed consent, and apparently no patients crossed over or dropped out (Berg 2009). One study had a co-ordinating secretary not involved in the treatment who could access the allocation on the internet, presumably at all times, but allocation was disclosed after randomisation (Hellum 2011). For this study it was unclear how long before intervention the allocation was disclosed, but 16 (of 173) participants did not proceed to the allocated intervention. The other studies did not mention allocation concealment (Moreno 2008; Sasso 2008; Zigler 2007).

## Blinding

Except for outcome assessor blinding in one study (Hellum 2011), blinding of any party is probably not used in any of the RCTs although description of blinding was insufficient. Even then, for the study by Hellum, the primary outcome measures were patient-reported outcomes and the patients were not blinded. Care provider blinding is obviously impossible in trials on surgical techniques and implants. However, patient blinding and especially outcome assessor blinding is feasible, but was not used in the included studies.

Regarding observer blinding, most of the (primary) outcomes were VAS pain and Oswestry scores. As these are self-report questionnaires, lack of assessor blinding has less influence on the risk of bias than lack of patient blinding.

## Incomplete outcome data

In general, the large FDA-IDE trials have acceptable loss to follow-up between 10% and 12% at two years. However, one trial represents only one centre in a multicentre trial (Sasso 2008), and, as the author acknowledges, these results should be read carefully until the complete patient sample has been reported upon. One study (Hellum 2011) reported a drop-out of 19.6% (13 and 21 patients in the two groups), but of these, 26 patients returned a

questionnaire with the primary outcomes. The remaining studies report loss to follow-up between 0% and 4%. Of note is the available data for the primary, composite, outcome parameter for the study of Gornet 2011, although only reported in supplementary tables, the loss to follow-up appears to be up to 22% for the total disc replacement group and 40% for the fusion group at 24 months follow-up.

For only one study (Blumenthal 2005) there is a report of five-year follow-up. There is a high percentage loss to follow-up in this report and thus a high risk of bias for the five-year follow-up results. It was decided not to include these results as the percentage loss to follow-up exceeded 20%.

## Selective reporting

From the FDA IDE trials we received no protocols from the FDA on the basis of the Freedom Of Information Act, but the application is still pending. From the three non-FDA trials we received no protocols from the authors. Hellum 2011 was registered on clinicaltrials.gov, but after two years of inclusion. Moreover, we assessed the descriptions in the trial registries valuable for prevention of publication bias, but for the purpose of this review, insufficient in assessing reporting bias on detailed items, such as what were exactly the primary and secondary outcome parameters and at what time points, and which analyses were anticipated.

## Other potential sources of bias

### Sponsoring

Three studies were sponsored solely by commercial parties, being the manufacturers of the total disc replacement devices (Gornet 2011: Medtronic; Sasso 2008: Stryker; Blumenthal 2005: DePuy). For one study (Zigler 2007), funding and conflict of interest statements were inconsistent across publications where the primary publication mentioned no funding for the work at hand but authors had received benefits from a commercial party related to the subject. However, secondary publications for this study do mention either funding, received benefits or both. One study was funded by non-commercial parties (Hellum 2011: the South Eastern Norway Regional Health Authority and EXTRA funds from the Norwegian Foundation for Health and Rehabilitation, through the Norwegian Back Pain Association). One study did not receive financial support (Berg 2009). One study did not disclose information about funding or conflict of interest (Moreno 2008). A secondary publication from Berg 2009 received financial support for health economy calculations (DePuySpine, Synthes and Medtronic).

### Baseline characteristics

For the comparison between total disc replacement and fusion, baseline characteristics were fairly similar. One trial ([Berg 2009](#)) has baseline differences for VAS leg pain, when evaluating the treatment difference in change scores, the significance of the outcome at two years diminished. Also, the change scores for back pain were not significant in contrast to the uncorrected scores at two years. Their primary outcome, a global assessment of back pain, has no baseline value, but probably relies on the baseline pain levels and therefore carries the risk of bias. Three other studies ([Blumenthal 2005](#); [Gornet 2011](#); [Zigler 2007](#)) reported change scores, but no difference with the uncorrected scores were found.

[Moreno 2008](#) has possibly differences in baseline for both pain and Oswestry score, but this study did not perform statistical testing, and no change scores are presented. [Sasso 2008](#) did not report baseline characteristics. The risk of bias from unbalanced baseline characteristics from these studies is believed to be high. However, the overall impact of these studies on the results of this review is limited, as [Sasso 2008](#) was not used for the analyses because it represents only one centre of a multicentre trial and [Moreno 2008](#) has only a sample size of 32 patients. Further, both were high risk of bias studies and were excluded from the sensitivity analyses.

For the only study with a comparison between total disc replacement and rehabilitation ([Hellum 2011](#)), baseline characteristics were significantly different for secondary outcome scores back pain and SF36 mental health score. The change from baseline scores was still significantly different.

### Co-interventions

The use of co-interventions was hard to evaluate. Many different rehabilitation protocols and concurrent interventions could be imagined. In general, as far as reported, the co-interventions were considered standard across interventions.

### Compliance

Compliance was generally not an issue in trials comparing different surgical techniques or implants. Only for the comparison between disc replacement and rehabilitation, cross-over was an issue, although cross-over from rehabilitation to surgery was within reasonable limits and an intention-to-treat analysis was performed in addition to a per protocol analysis.

### Timing of outcome assessments

The well-protocolised FDA-IDE studies have guaranteed assessment of outcomes within relatively standard and narrow time windows. Only one study ([Moreno 2008](#)) had a difference between the follow-up periods of the groups.

### Heterogeneity

In general, there appeared to be no heterogeneity between the different interventions on the  $I^2$  test. Only blood loss showed

considerable heterogeneity (94%). In addition, Oswestry at 24 months had an  $I^2$  of 39%, but we have not used this analysis in the results as we only used improvements from baseline.

## Effects of interventions

See: [Summary of findings for the main comparison Disc arthroplasty versus fusion for chronic low-back pain](#)

Because of the high risk of bias due to possible selective reporting of the single-centre results ( $n = 76$ ) of the FDA IDE study on the FlexiCore by [Sasso 2008](#), we decided not to include these preliminary results in this version of the review. The data being incomplete would have influenced the grading of the evidence as the comparison is not free of publication bias and no decision can be made about consistency of other studies with this study. The only report with five-year follow-up ([Blumenthal 2005](#)) was excluded from analysis because of extensive loss to follow-up (up to 43%). [Zigler 2007](#) separately reported one- and two-level procedures, but for this report we only used the one-level results.

### 1. Disc arthroplasty versus fusion

Five studies ([Berg 2009](#); [Blumenthal 2005](#); [Gornet 2011](#); [Moreno 2008](#); [Zigler 2007](#), one level) with 1301 patients were found that compared total disc replacement ( $n = 865$ ) with fusion ( $n = 436$ ). Charite was compared by [Blumenthal 2005](#) ( $n=205$ ) against ALIF with BAK fusion cages ( $n = 99$ , Zimmer Spine, Minneapolis, MN) and by [Moreno 2008](#) ( $n=14$ ) against ALIF with KLA cage ( $n = 18$ , Scient'X, Guyancourt, France), iliac crest autograft and anterior plate. [Zigler 2007](#) compared one- and two-level procedures of Prodisc ( $n = 161$  and  $n = 165$ ) against femoral ring allograft with posterior pedicle screws and iliac crest autograft ( $n = 75$  and  $n = 72$ ). [Berg 2009](#) compared Charite, Prodisc or Maverick ( $n = 80$ ) with fusion with local or iliac crest bone graft, posterior pedicle screws, posterolateral or interbody fusion with Brantigan cage ( $n = 72$ , DePuy Spine, Raynham, MA).

Four studies had a low risk of bias ([Berg 2009](#); [Blumenthal 2005](#); [Gornet 2011](#); [Zigler 2007](#)) and one study had a high risk of bias ([Moreno 2008](#)).

### Improvement in pain

All studies reported significant differences in improvement in either leg or back pain or unspecified pain favouring total disc replacement at various short-term follow-up time points up to 12 months. "Improvement in pain" (not specified) was measured by [Blumenthal 2005](#) and [Zigler 2007](#) at six weeks and three, six and 12 months where [Blumenthal 2005](#) found a significant difference at all follow-up time points and [Zigler 2007](#) (one level) only after three months. [Berg 2009](#) measured "improvement in leg pain" at 12 and 24 months and found no differences. [Gornet 2011](#) measured "improvement in leg pain" at six weeks and three, six and 12 months and found a difference at 12 months. [Berg 2009](#) measured

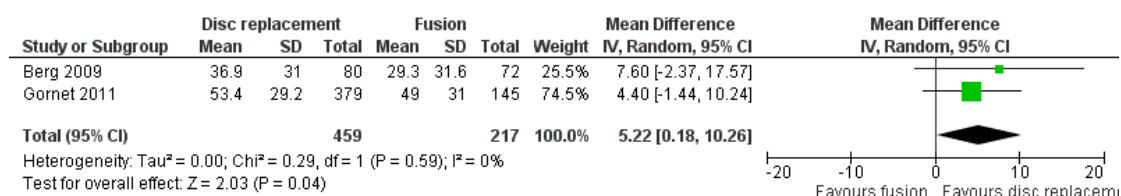
“improvement in back pain” at 12 and 24 months and found no differences. [Gornet 2011](#) measured “improvement in back pain” at six weeks and three, six and 12 months and found a difference at all follow-up time points.

At 24 months, all studies evaluated improvement of either leg or back pain or unspecified pain. Two studies ([Berg 2009](#); [Gornet 2011](#)) measured both “improvements in back pain” and “improvements in leg pain”, but only one significant difference was found for “improvement in back pain” favouring total disc replacement in one study ([Gornet 2011](#)). Three studies ([Blumenthal 2005](#);

[Moreno 2008](#); [Zigler 2007](#)) did not specify the type of pain (“pain” was assessed), but these studies did not provide an estimate of variability so results could not be pooled.

- There is low quality evidence from two low risk of bias studies ([Berg 2009](#); [Gornet 2011](#)) that the mean improvement in VAS back pain at 24 months in the disc replacement group was 5.2 mm (out of 100) higher (676 patients; 95% CI 0.2 to 10.3, see [Figure 3](#)) than in the fusion group. The quality of evidence had to be decreased because of risk of publication bias and imprecision.

**Figure 3. Forest plot of comparison: I Disc replacement versus fusion, outcome: I.3 Improvement in back pain at 24 months.**



- There is very low quality evidence from two studies ([Berg 2009](#); [Gornet 2011](#)) that the difference between disc replacement and fusion in mean improvement of VAS leg pain at 24 months was not statistically significant. The quality of evidence had to be decreased because of risk of publication bias, inconsistency and imprecision.

### Overall improvement

The definition of overall improvement differed between the studies. [Blumenthal 2005](#) defines clinical success as 25% improvement of Oswestry score, no device failure, no major complications and no neurological deterioration. [Zigler 2007](#) uses the FDA criteria needing an improvement in 10 endpoints, including Oswestry score (two analyses; 15% and 15 points), SF-36, device success, radiological success and neurological success. [Gornet 2011](#) defined a 15-point improvement also on a composite score including successful neurological status, disc height, no additional surgery and no serious device or surgical procedure-related adverse events. The composite variable appeared to suffer from the loss to follow-up of the compound scores as only 313/405 and 103/172 were analysed. [Moreno 2008](#) defines excellent result as a 60% gain in Oswestry

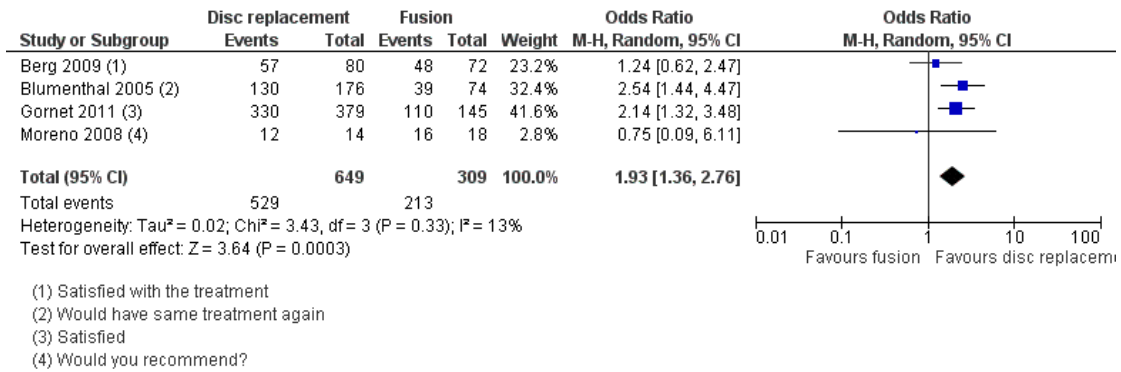
and 50% gain in VAS pain. [Berg 2009](#) did not define an overall outcome. Meta-analysis could not be performed.

### Patient satisfaction

All studies evaluated patient satisfaction, four studies (see below) used a dichotomised measure and could be pooled. [Zigler 2007](#) used a continuous VAS measure for patient satisfaction and found a significant difference of 8.6 mm favouring total disc replacement over fusion for one-level procedures and 8,8 mm for two-level procedures. The questions on satisfaction were “would you have the same treatment again”, “satisfied with results” and “would recommend the intervention”.

- There is moderate quality evidence from four studies ([Berg 2009](#); [Blumenthal 2005](#); [Gornet 2011](#); [Moreno 2008](#)) that patient satisfaction at 24 months is more prevalent in the total disc replacement group with an OR of 1.93 (958 patients; 95% CI 1.36 to 2.76; see [Figure 4](#)) than in the fusion group. The quality of evidence had to be decreased because of inconsistency, two non-sponsored studies found no effect and two sponsored studies found a statistically significant effect.

**Figure 4. Forest plot of comparison: I Disc arthroplasty versus fusion, outcome: I.10 Patient satisfaction.**



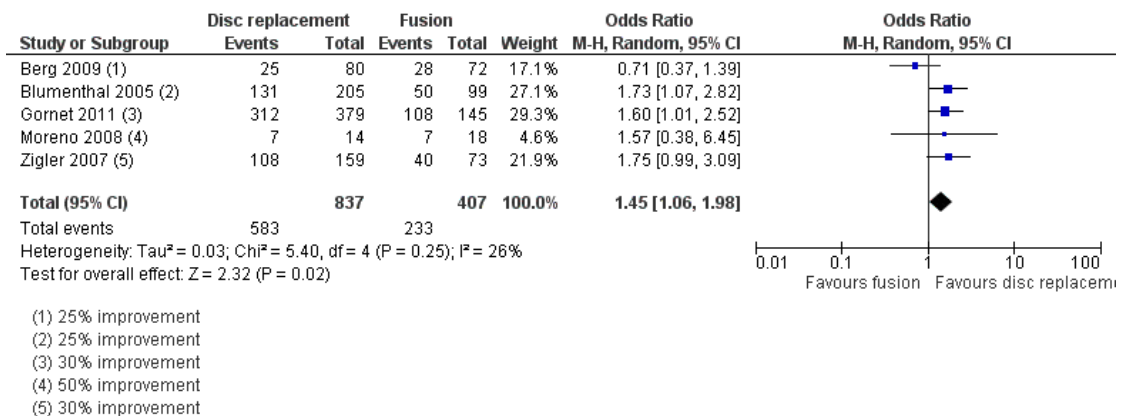
**Back-specific functional status**

All studies included an Oswestry score, other functional status scores were not used. All four low risk of bias studies reported a difference in improvement of the Oswestry score favouring total disc replacement over fusion at various follow-up time points. At six weeks, three months and six months differences were found by three studies (Blumenthal 2005; Gornet 2011; Zigler 2007). Additionally, Gornet 2011, Berg 2009 and Blumenthal 2005 found a difference at 12 months. Only one study (Gornet 2011), reported a difference in improvement of Oswestry score favouring total disc replacement at 24 months, while four other studies could not find a difference. The pooled analysis, however showed a statistically significant difference favouring disc replacement over fusion. The level for improvement on the Oswestry score varied between the studies from 25% to 50%.

- There is moderate quality evidence from five studies (Berg 2009; Blumenthal 2005; Gornet 2011; Moreno 2008; Zigler 2007) that the mean Improvement in Oswestry score at 24 months in the disc replacement group was 4.27 points (out of 100) better (1207 patients; 95% CI 1.85 to 6.68; see Figure 5) than in the fusion group. The quality of evidence had to be decreased because of inconsistency of the results as one found significant effects and four found no significant effects.

- There is low quality evidence from five studies (Berg 2009; Blumenthal 2005; Gornet 2011; Moreno 2008; Zigler 2007) that patients in the total disc replacement group are more likely to have been improved on the Oswestry scale at 24 months more than a predefined level (1244 patients; OR 1.45, 95% CI 1.06 to 1.98; see Figure 5) than in the fusion group. The quality of evidence had to be decreased because of inconsistency and imprecision.

**Figure 5. Forest plot of comparison: I Disc replacement versus fusion, outcome: I.14 Percentage of patients improved on Oswestry at 24 months.**



### Well being, quality of life

Only two low risk of bias studies (Gornet 2011 and Zigler 2007, two level) reported mean and standard deviation of SF-36 scores. Zigler 2007 (one level) only reports the SF-36 success rate. Blumenthal 2005 and Berg 2009 mention having used SF-36, but do not report results. The results from Gornet 2011 show that there are only very short term (1.5 and three months) differences in improvement in mental component score favouring total disc replacement. For the physical component score, the improvement was significant at every follow-up and favoured total disc replacement. Meta-analysis could not be performed. Because for both one-and two-level procedures, there is only one study, only a low quality of evidence can be appointed to this comparison.

Four studies reported working status at all follow-up measurements (Gornet 2011), at three, 12 and 24 months (Berg 2009), 12 and 24 months (Blumenthal 2005) and at 24 months (Zigler 2007, one level). Only Gornet 2011 reports time to return-to-work from a Kaplan-Meier analysis, but with no indication of variability. The preoperative working status varied between the studies, limiting the analysis to a change in working status between preoperative and 24 months.

- There is very low quality from four studies (Berg 2009; Blumenthal 2005; Gornet 2011; Zigler 2007, one level) that the difference in improvement working status at 24 months is not statistically significantly different between disc replacement and fusion. The level of evidence had to be decreased because of inconsistency, possible reporting bias and imprecision.

### Radiological and complications

- There is moderate quality evidence from four low risk of bias studies (Blumenthal 2005; Gornet 2011; Zigler 2007, one level; Berg 2009) that total disc replacement leads to a range of motion that is more comparable to the preoperative status. All four studies found a range of motion of total disc replacement up to two years comparable to the preoperative status, while the range of motion for fusion was nearly zero. The quality of evidence had to be decreased due to inconsistent reporting with the impossibility for meta-analysis. There was, however, a dramatic effect allowing moderate quality of evidence.

- Only two studies mentioned thromboembolic complications. Blumenthal 2005 mentioned two cases of venous thromboses in the disc replacement group and none in the fusion group. Gornet 2011 reports one cardiovascular event in the disc replacement group and none in the fusion group. No meta-

analyses could be performed because the definition of “cardiovascular event” in Gornet 2011 was not given.

- There is very low quality evidence from five studies (Berg 2009; Blumenthal 2005; Gornet 2011; Moreno 2008; Zigler 2007, one level) that there is conflicting evidence on the amount of perioperative blood loss. The level of evidence had to be decreased from high to very low because of imprecision, inconsistency and reporting bias.

- There were 63 of 810 (7.8%) re-operations in the total disc replacement group and 35 of 384 (9.1%) in the fusion group. There is very low quality evidence from five studies (Berg 2009; Blumenthal 2005; Gornet 2011; Moreno 2008; Zigler 2007, one level) that the difference in re-operations up to 24 months was not statistically significant. The quality of evidence had to be downgraded because of imprecision, inconsistency and possible reporting bias.

- Only one secondary publication of a low risk of bias study (Blumenthal 2005) reported neurological complications and found no difference between the two groups.

- There is very low quality evidence from one low risk of bias study (Berg 2009) that the difference in adjacent segment degeneration at 24 months was not statistically different. This one study only marginally reported adjacent segment degeneration mentioning six of 72 cases of fusion and only one of 80 cases of total disc replacement with adjacent segment problems. The level of evidence is very low because there was only one trial and had to be decreased because of imprecision.

- There is very low quality of evidence from one low risk of bias study (Berg 2009) that the occurrence of facet joint degeneration is not statistically significantly different. The quality of evidence had to be downgraded because of imprecision.

- Radiographic loosening was not adequately reported in any of the studies. Zigler 2007 mentions one case of radiolucency in the fusion group, but no criteria are defined.

- Subsidence was not adequately reported and/or defined in any of the studies. Blumenthal 2005 reports average subsidence in mm, but does not report numbers of patients above a predefined margin. Gornet 2011 and Zigler 2007 (one level) report number of patients with subsidence, but both studies do not report the criteria for when subsidence was scored as present. Also, both studies have a remarkable difference in subsidence (< 1% versus > 5%). Meta-analysis could not be performed.

- Polyethylene wear was not assessed in any of the studies.

### Cost-effectiveness

One study reported cost-effectiveness in a secondary publication

(Berg 2009). No difference in societal cost could be found while from a healthcare perspective fusion was more costly. Since clinical benefit was only small and because of the wide spread in possible cost per quality adjusted life year gained, cost-effectiveness could not be shown superior for either intervention.

## 2. Disc arthroplasty versus rehabilitation

One low risk of bias study (Hellum 2011) comparing disc replacement (n = 86) with rehabilitation (n = 87) found significant difference in improvement on the primary outcome, the Oswestry scale, but not beyond the clinically relevant difference of 10 points. For back pain and Oswestry score, mixed model analysis (all follow-ups) and univariate analysis (at baseline and one and two years) were presented, the last were used for analysis. No improvement scores were reported for these outcome parameters, but “Treatment effect” could be used for analysis of difference in improvement, so analyses below are from the authors’ report.

### Improvement in pain

- There is very low quality evidence from one low risk of bias study (Hellum 2011) that the improvement of back pain at 12 months in the total disc replacement group is 14.0 mm higher (152 patients; 95% CI 5.0 to 23.0 mm) than in the rehabilitation group. Quality of evidence had to be decreased because there was only one trial with an imprecise estimate.
- There is very low quality evidence from one low risk of bias study (Hellum 2011) that the improvement of back pain at 24 months in the total disc replacement group is 12.3 mm higher (152 patients; 95% CI 3.1 to 21.3 mm) than in the rehabilitation group. Quality of evidence had to be decreased because there was only one trial with an imprecise estimate.

### Overall improvement

Overall improvement was not assessed or calculated in this study.

### Patient satisfaction

- There is low quality evidence from one low risk of bias study (Hellum 2011) that patients are more likely to be satisfied with the outcome at two years in the total disc replacement group (OR 2.65 95% CI 1.42 to 4.96) than in the rehabilitation group. Quality of evidence had to be decreased because there was only one trial.

### Back-specific functional status

- There is low quality evidence from one low risk of bias study (Hellum 2011) that the improvement of Oswestry score at 12 months in the total disc replacement group is 8.90 points higher (172 patients; 95% CI 4.77 to 13.03 points) than in the

rehabilitation group. Quality of evidence had to be decreased because there was only one trial.

- There is low quality evidence from one low risk of bias study (Hellum 2011) that the improvement of Oswestry score at 24 months in the total disc replacement group is 6.90 points higher (172 patients; 95% CI 2.23 to 11.57 points) than in the rehabilitation group. Quality of evidence had to be decreased because there was only one trial.

### Well being, quality of life

- There is low quality evidence from one low risk of bias study (Hellum 2011) that there is no statistically significant difference in improvement of working status.

### Radiological and complications

The one study evaluating total disc replacement with rehabilitation did not measure any radiological parameters, thus implant motion, adjacent segment degeneration, facet joint degeneration, loosening, subsidence and polyethylene wear could not be analysed.

- Thromboembolic complications were reported for two patients with total disc arthroplasty.
- Blood loss of more than 1500 cc was reported for four patients.
- There is low quality evidence from one low risk of bias study (Hellum 2011) that there is no statistically significant difference in subsequent operation rate.

## DISCUSSION

### Summary of main results

#### Effectiveness

Most studies show clinically important changes (Ostelo 2008) from baseline for both interventions, being total disc replacement on one hand and fusion or rehabilitation on the other. However, the differences on the primary outcome parameters, including their 95% confidence intervals, were smaller than these predefined changes or the quality of evidence was of low or very low quality.

#### How do the clinical results compare to fusion, the only surgical “gold standard” available?

The higher improvement in back pain for total disc replacement (5.2 of 100 mm; 95% CI 0.2 to 10.3) did not exceed the predefined clinically relevant difference (15 mm) with a low quality of evidence. For other outcomes, moderate quality of evidence was



only found for difference in improvement on the Oswestry score at 24 months (4.3 points; 95% CI 1.85 to 6.68) in favour of disc replacement, which also did not exceed the predefined clinically relevant difference of 10 points. Patients reported higher satisfaction with total disc replacement and the disc replacement devices appeared to succeed in maintaining more natural motion characteristics. Overall, the clinical differences were considered as being relevant in only one study. From none of the included studies could we conclude that the likely clinical benefits, if present, could be worth the possible potential harms (Table 2).

#### **How do the clinical results compare to other treatment options?**

Only one study is available for the comparison between total disc replacement with rehabilitation. The average effect for improvement in back pain at 24 months (12.3 mm) is below the predefined clinical relevant difference of 15 mm although there is a chance (95% CI 3.1 to 21.3) that the clinical outcome is superior for total disc replacement. This conclusion could also apply on the improvement in Oswestry score (8.4 points; 95% CI 3.6 to 13.2). However, evidence for this comparison was considered of low or very low quality, because it was based on only one low risk of bias trial.

#### **Safety**

#### **What are the radiologic results with regards to loosening, subsidence (slippage) of the implant into the vertebral bodies, and polyethylene wear?**

Radiological results are poorly reported across the studies. Very few data are available and where reported, definitions are not provided.

#### **Does the motion segment retain its mobility? And if it does, do these motions resemble a normal motion segment?**

There is limited amount of evidence that motion segment mobility is preserved through disc replacement. Although it can be argued that clinical studies evaluating the effectiveness of disc arthroplasty should always include the assessment of mobility, motion assessment was only presented in three of the studies (Blumenthal 2005; Gornet 2011; Zigler 2007). However, the data presentation pertaining to motion was poor limiting the possibility for meta-analysis. These studies found that total disc replacement showed range of motion comparable to preoperative range of motion, while fusion surgery lead to (nearly) absence of motion as would be expected.

We have to comment that our search and selection process was not aimed at finding studies that exclusively valued the mobility of total disc replacement without assessment of clinical outcome, thus the included randomised studies could be not representative.

However, two excluded studies (SariAli 2006; Tournier 2007) and one secondary publication (Zigler 2007) confirmed our findings. A comparable mobility with healthy volunteers was found by SariAli 2006 in patients with disc arthroplasty at one level, but with more than one level, the mobility was abnormally increased. The secondary study for Zigler 2007 also finds physiological motion profiles for flexion-extension for operative and proximal levels.

#### **Can the total disc replacement reduce the incidence of adjacent segment degeneration compared to fusion?**

We did not find conclusive evidence of a beneficial effect of total disc replacement versus fusion on adjacent segment degeneration at relevant long-term follow-up. Only one study assessed adjacent segment degeneration (Berg 2009), but could not make a precise estimate of the effect due to a limited sample size and low incidence of adjacent segment degeneration at the short-term follow-up of two years. Ultimately, as the population receiving total disc replacement is relatively young, adjacent segment degeneration needs to be assessed at the long term. For example, Putzier 2006 found 17% adjacent segment degeneration after 17.3 years in 53 patients in a cohort of the Charite disc replacement.

#### **What is the incidence of facet joint degeneration at the operated level?**

Most studies excluded patients with facet joint degeneration. On the development of facet joint degeneration after implantation of a disc replacement, there is insufficient evidence. Only one study reports facet joint degeneration (Berg 2009) with an imprecise estimate due to very low incidence at two-year follow-up. The assessment of facet joint degeneration appears relevant as the flexion/extension appears to be comparable to normal, but especially if the axis of rotation is altered the motion might conflict with the facet joints.

#### **How does the perioperative complication rate compare to fusion operations?**

Perioperative complications are not adequately and consistently reported between the studies. Blood loss showed great heterogeneity and thrombo-embolic complications were only reported in one study. From these studies it is not clear how the perioperative complication rate of total disc replacement compares with fusion procedures.

#### **“Is there an acceptable and safe salvage procedure in case of failure?”**

Gornet 2011 mentions two removals of disc arthroplasty due to epidural abscess and allergic reaction. Berg 2009 mentions fusion

of four disc arthroplasty procedures close to the 24 month follow-up. It was difficult to link the failures of the device to salvage procedures and outcome. Definition of "failure" varies and it was difficult to classify failures as "device related". Outcomes of possible salvage cases were not reported in one study (Gornet 2011). Posterior decompression through laminectomy might very well have been necessary for pain due to stenosis of the foramen with extension. One secondary study reports on the revisability of the total disc replacement from their trial (Blumenthal 2005) and concluded that one third could be revised to a new total disc replacement and that two thirds of patients could be revised to a fusion procedure. Recent studies on revisability (de Maat 2009) show that revision is possible, but not without (vascular) complications and with highly variable outcome (Punt 2008).

### Overall completeness and applicability of evidence

Although clearly presented, the extensive selection criteria in the FDA-IDE trials (Blumenthal 2005; Gornet 2011; Sasso 2008; Zigler 2007) result in a very select patient sample regarding height loss, age and presence of any comorbidity. This limits the external validity of the results of these trials, and precludes extrapolation to the application of total disc replacement in all patients with degenerative disc disease.

The surgical techniques were fairly well described, although co-interventions such as the rehabilitation protocol used lack detailed description. There were few comparable control groups between the studies. Only the control groups of Berg 2009 and Zigler 2007, which both use circumferential fusion, are sufficiently comparable. Although both Blumenthal 2005 and Gornet 2011 use a stand-alone cage, Gornet 2011 fills the cage with a bone substitute whereas Blumenthal 2005 uses iliac crest autograft. The BAK cages used by Blumenthal 2005 have been criticized for resulting in large numbers of patients with unacceptable outcome. Since superiority of any technique in lumbar fusion for degenerative disc disease has not yet been established, these results were pooled when statistical heterogeneity was absent. However, future evidence about relative effectiveness of these interventions could warrant subgroup analyses. A systematic assessment of the evidence on this topic is thus clearly needed.

The most relevant outcome parameters, pain, disability with Oswestry score and return to work are used, and although statistically significant differences were found, no clinically relevant differences were found when these were compared with the clinically important differences as defined by Ostelo 2008. However, these guidelines were defined as a meaningful change from baseline for an intervention on an individual basis, whereas the clinically relevant difference in effect between two groups might be different. Although we agree that differences at the group level might be somewhat smaller than improvements at the individual level, relevant differences at the group level between surgical interventions

for low-back pain have not been published in the literature. For this review we had to rely on the best information available on clinically relevant differences. Furthermore, the studies only reported the percentage of the patients that improved more than a predefined level for the Oswestry score. On top of this, the predefined level varied between 25%, 30% and 50%. Therefore, the interpretation of the results was also supported with the results of the clinical relevance score from Table 2. The Oswestry score as used by Zigler 2007 was an adapted version of the official validated version. The adaptation probably raised the baseline levels, allowing more room for improvement. Their estimate of improvement on the Oswestry score can thus be regarded as an overestimate.

### Quality of the evidence

Although the risk of bias of five studies could be considered low, the most important items, sequence generation, allocation concealment, blinding, and comparable baseline values were not met in most studies.

### Potential biases in the review process

#### Publication bias

Sasso 2008 only represents a sub analysis from a multicentre study. This study is included but results were not used for meta-analytic purposes. The data being incomplete would have influenced the grading of the evidence as the comparison is not free of publication bias and no decision can be made about consistency of other studies with this study.

Evidence from the current observational studies could not be used because of the high risk of bias. We had felt that non-randomised studies could have contributed to the total evidence as there are many concerns in randomised surgical studies with blinding and randomisation. Further, non-randomised studies could have additional value for external validity of the evidence and for assessing the risk of complications in a wider range of patients than those usually included in strict randomised trials. Non-randomised studies should however, be very clear about patient selection and should incorporate independent, blinded outcome assessment. This was certainly not the case in those considered for inclusion in this review and, therefore, these studies were excluded.

#### Sensitivity analyses

There were only two high risk of bias studies (Moreno 2008; Sasso 2008). Sasso 2008 was excluded from all analyses. Excluding the small study by Moreno 2008 from the analysis did not influence the quality of evidence and had only very limited effect on the estimate for improvement in Oswestry score at 24 months.

Subgroup analyses for patients with one- and two-level procedures could not be performed because these results were not presented separately in two of the three studies including two-level procedures (Berg 2009; Hellum 2011). Zigler 2007 separately reported one- and two-level procedures and the results appear to be different. Berg 2009 states no difference in outcome between one- and two-level procedures, but no data are provided. Excluding Berg 2009 from the comparison against fusion did not influence the quality of evidence and had only a limited effect on the estimate for improvement in Oswestry score at 24 months. For improvement in back and leg pain, quality of evidence would be reduced, because only one study remained for this comparison.

### Agreements and disagreements with other studies or reviews

Our previous systematic review (de Kleuver 2003), included only case series and concluded that disc replacement seemed to be associated with a high rate of re-operations. On that evidence, disc replacement was considered to be an experimental procedure, only to be applied in strict clinical trials. Two early reviews (Freeman 2006; Randolph 2006) published after the first clinical results of comparisons against fusion concluded that there was no evidence that disc replacement reduced adjacent segment degeneration and that more independent, long-term prospective studies were needed to systematically evaluate success and failure of this innovative technique. Two recent systematic reviews included three (Van den Eerenbeemt 2010) and four (Yajun 2010) trials. Van den Eerenbeemt 2010 raises questions about the external validity of the trials as carefully selected patients were included. Further, the relevance of the size of the effect is small, as was found in our review. The conclusions of Yajun 2010 lack reliability because of duplicate inclusion of the same study (using a secondary reference (results of one centre) next to the primary (multicentre) reference). Both reviews suggest that the use of BAK cages is no longer indicated as a comparator giving more relevance to the comparisons with circumferential fusion as in the Prodisc-L study. In our review we could not find distinct differences between the different approaches for fusion. In line with the previous reviews, we conclude that still more studies and longer-term follow-up assessments are needed to assess the effectiveness and safety of this device.

## AUTHORS' CONCLUSIONS

### Implications for practice

Disc replacement compared with conventional fusion surgery for degenerative disc disease appears to result in clinically irrelevant superiority with respect to pain relief, Oswestry and Quality of Life in a selected population, so far only for the short term. Currently

available trials did not assess adjacent level disease and facet joint degeneration properly, which is a shortcoming as this is the reason the product was manufactured. Therefore, because we believe that harm and complications may occur after years, we believe that the spine surgery community should be prudent about adopting this technology on a large scale, despite the fact that total disc replacement seems to be effective in treating low-back pain in selected patients, and in the short term is at least equivalent to fusion surgery. Otherwise, if equivalence can be assumed along the proposed methodology, use of these expensive implants should remain limited including informed consent and mandatory quality assessment in registries similar to the recommendations for metal-on-metal (MoM) hip prostheses (Sedrakyan 2011).

### Implications for research

Despite the publication of several recent studies on the effectiveness of total disc replacement, there are still additional trials needed, with less conflict of interest. Since only one trial compared total disc replacement with conservative treatment, this design needs to be replicated. Current ongoing studies all compare different brands of total disc replacement, thus these will not provide this additional evidence. Most important are long-term follow-up studies as differences in adjacent segment degeneration will only then be identified. Furthermore, prosthesis-related complications such as wear and loosening are likely to develop after longer implantation periods. In order to better evaluate clinically relevant group differences, research and consensus are needed on the clinical relevant differences between (surgical) interventions for low-back pain (Dworkin 2009).

### Implications for those planning a total disc replacement trial

Choose current established interventions as control group, i.e. circumferential fusion or conservative treatment. Assessment of mobility at index level with facet joint degeneration as well as mobility at adjacent levels with adjacent segment degeneration needs to be performed to assess safety. Provide supplementary tables with scores at all follow-up time points with indication of variability, as well as changes from baseline with indication of variability. Also, provide the reader with sample sizes and numbers of patients followed up for each outcome parameter at each follow-up time point.

### Implications for other reviewers

For a better understanding of the effectiveness of total disc replacement, future reviews should be able to make more efficient use of the collected data in the studies, published as well as unpublished, especially from the FDA IDE trials. Individual patient data meta-analysis can provide a more powerful analysis and offers more possibilities for subgroup analyses. There were four out of seven responses to the author queries pertaining methodological

information, which makes patient data sharing promising. This step could improve the quality assessment of included studies, but the responses need to be published and regarded peer-reviewed information, in addition to the original publication. This review needs to be updated on a relatively short notice as new information from the ongoing trials will become available in the near future. The ongoing studies will specifically evaluate the relative effectiveness of disc replacements amongst each other. Both the studies on Activ-L versus Prodisc or Charite (*Aesculap Impl syst 2007*), and Kineflex versus Charite (*SpinalMotion 2006*) have passed the anticipated primary outcome evaluation. Studies on the Freedom disc versus an unknown control disc replacement device (*Axiomed Spine Corp 2008*) and on the Nubac nucleus replacement versus Prodisc (*Pioneer 2009b*) are anticipated to end respectively in 2012 and 2014.

## ACKNOWLEDGEMENTS

The authors are grateful for being supported by the Cochrane Back Review Group, in particular Victoria Pennick and Teresa Marin for prompt and continuous support and feedback and Rachel Couban for (repeated) search strategies and results. Special thanks to Rachel Couban for her efforts with FDA requests. Thanks to Moyo Kruyt for clinical interpretation and draft review.

The authors also thank the principle investigators of the included studies, dr J Zigler of the Texas Back Institute (Plano, Texas, US), dr S Berg of the Stockholm Spine Center (Stockholm, Sweden), dr C Hellum of the Department of Orthopedics of the Oslo University Hospital and Oslo University (Oslo, Norway), and dr M Gornet of The Orthopedic Center of St. Louis (St. Louis, Missouri, US) for providing additional information about unclear risk of bias items.

## REFERENCES

### References to studies included in this review

#### Berg 2009 *{published data only}*

- Berg S. On total disc replacement. *Acta Orthopædica* 2011; **82**(343):1–29.
- Berg S, Fritzell P, Tropp H. Sex life and sexual function in men and women before and after total disc replacement compared with posterior lumbar fusion. *Spine Journal: Official Journal of the North American Spine Society* 2009; **9**(12):987–94.
- Berg S, Tropp H. Results from a randomized controlled study between total disc replacement and fusion compared with results from a spine register. *SAS Journal* 2010; **4**(3): 68–74.
- Berg S, Tropp HT, Leivseth G. Disc height and motion patterns in the lumbar spine in patients operated with total disc replacement or fusion for discogenic back pain. Results from a randomized controlled trial. *Spine Journal: Official Journal of the North American Spine Society* 2011; **11**(11): 991–8.
- \* Berg S, Tullberg T, Branth B, Olerud C, Tropp H. Total disc replacement compared to lumbar fusion: a randomised controlled trial with 2-year follow-up. *European Spine Journal* 2009; **18**(10):1512–9.
- Fritzell P, Berg S, Borgstrom F, Tullberg T, Tropp H. Cost effectiveness of disc prosthesis versus lumbar fusion in patients with chronic low back pain: randomized controlled trial with 2-year follow-up. *European Spine Journal* 2011; **20**(1432-0932 (Electronic), 0940-6719 (Linking), 7): 1001–11.

#### Blumenthal 2005 *{published data only}*

- \* Blumenthal S, McAfee PC, Guyer RD, Hochschuler SH, Geisler FH, Holt RT, et al. A prospective, randomized, multicenter Food and Drug Administration investigational device exemptions study of lumbar total disc replacement with the CHARITE artificial disc versus lumbar fusion -

- Part I: Evaluation of clinical outcomes. *Spine* 2005; **30**(14): 1565–75.
- Geisler FH, Blumenthal SL, Guyer RD, McAfee PC, Regan JJ, Johnson JB, et al. Neurological complications of lumbar artificial disc replacement and comparison of clinical results with those related to lumbar arthrodesis in the literature: results of a multicenter, prospective, randomized investigational device exemption study of Charite intervertebral disc. Invited submission from the Joint Section Meeting on Disorders of the Spine and Peripheral Nerves, March 2004. *Journal of Neurosurgery Spine* 2004; **1**(2):143–54.
- Geisler FH, Guyer RD, Blumenthal SL, McAfee PC, Cappuccino A, Bitan F, et al. Effect of previous surgery on clinical outcome following 1-level lumbar arthroplasty. *Journal of Neurosurgery Spine* 2008; **8**(2):108–14.
- Geisler FH, Guyer RD, Blumenthal SL, McAfee PC, Cappuccino A, Bitan F, et al. Patient selection for lumbar arthroplasty and arthrodesis: the effect of revision surgery in a controlled, multicenter, randomized study. *Journal of Neurosurgery Spine* 2008; **8**(1):13–6.
- Geisler FH, McAfee PC, Banco RJ, Blumenthal SL, Guyer RD, Holt RT, et al. Prospective, randomized, multicenter FDA IDE study of CHARITE artificial disc versus lumbar fusion: effect at 5-year follow-up of prior surgery and prior discectomy on clinical outcomes following lumbar arthroplasty. *SAS Journal* 2009; **3**(1) (pp 17-25), 2009. Date of Publication: March 2009.):25.
- Guyer RD, McAfee PC, Banco RJ, Bitan FD, Cappuccino A, Geisler FH, et al. Prospective, randomized, multicenter Food and Drug Administration investigational device exemption study of lumbar total disc replacement with the CHARITE artificial disc versus lumbar fusion: five-year follow-up. *Spine Journal: Official Journal of the North*

- American Spine Society* 2009;**9**(5):374–86.
- Guyard RD, McAfee PC, Hochschuler SH, Blumenthal SL, Fedder IL, Ohnmeiss DD, et al. Prospective randomized study of the Charite artificial disc: data from two investigational centers. *Spine Journal: Official Journal of the North American Spine Society* 2004;**4**(6 Suppl):252S–9S.
- McAfee PC. Total disc replacement. *Operative Techniques in Orthopaedics* 2003;**13**(3):214–21.
- McAfee PC, Cunningham B, Holsapple G, Adams K, Blumenthal S, Guyard RD, et al. A prospective, randomized, multicenter Food and Drug Administration investigational device exemption study of lumbar total disc replacement with the CHARITE; artificial disc versus lumbar fusion - Part II: Evaluation of radiographic outcomes and correlation of surgical technique accuracy with clinical outcomes. *Spine* 2005;**30**(14):1576–83.
- McAfee PC, Fedder IL, Saiedy S, Shucosky EM, Cunningham BW. Experimental design of total disc replacement-experience with a prospective randomized study of the SB Charite. *Spine* 2003;**28**(20):S153–62.
- McAfee PC, Fedder IL, Saiedy S, Shucosky EM, Cunningham BW. SB Charite disc replacement: report of 60 prospective randomized cases in a US center. *Journal of Spinal Disorders & Techniques* 2003;**16**(4):424–33.
- McAfee PC, Geisler FH, Saiedy SS, Moore SV, Regan JJ, Guyard RD, et al. Revisability of the CHARITE artificial disc replacement: analysis of 688 patients enrolled in the U.S. IDE study of the CHARITE Artificial Disc. *Spine* 2006;**31**(11):1217–26.
- Gornet 2011** {published data only}
- \* Gornet MF, Burkus JK, Dryer RF, Pelozo JH. Lumbar disc arthroplasty with MAVERICK disc versus stand-alone interbody fusion: a prospective, randomized, controlled, multicenter investigational device exemption trial. *Spine (Phila Pa 1976.)* 2011;**36**(25):E1600–11.
- Hellum 2011** {published data only}
- Berg L, Gjertsen O, Hellum C, Neckelmann G, Johnsen LG, Eide GE, Espeland A. Reliability of change in lumbar MRI findings over time in patients with and without disc prosthesis-comparing two different image evaluation methods. *Skeletal Radiology* 2012;**Epub**:March 20.
- Hellum C, Johnsen LG, Gjertsen O, Berg L, Neckelmann G, Grundnes O, Rossvoll I, Skouen JS, Brox JI, Storheim K, The Norwegian Spine Study Group. Predictors of outcome after surgery with disc prosthesis and rehabilitation in patients with chronic low back pain and degenerative disc: 2-year follow-up. *European Spine Journal* 2012;**21**(4):681–690.
- \* Hellum C, Johnsen LG, Storheim K, Nygaard OP, Brox JI, Rossvoll I, Ro M, Sandvik L, Grundnes O. Surgery with disc prosthesis versus rehabilitation in patients with low back pain and degenerative disc: two year follow-up of randomised study. *British Medical Journal* 2011;**342**:d2786.
- Moreno 2008** {published data only}
- Moreno P, Boulot J. [Comparative study of short-term results between total artificial disc prosthesis and anterior lumbar interbody fusion]. [French]. *Revue de Chirurgie Orthopedique et Reparatrice de l Appareil Moteur* 2008;**94**(3):282–8.
- Sasso 2008** {published data only}
- \* Sasso RC, Foulk DM, Hahn M. Prospective, randomized trial of metal-on-metal artificial lumbar disc replacement: initial results for treatment of discogenic pain. *Spine* 2008;**33**(2):123–31.
- Tepper G, Wolf SB, Feldman L, Spector M. Artificial lumbar disc replacement vs circumferential fusion for the treatment of discogenic pain: A prospective randomized study Eurospine 2006. 8th Annual Meeting of the European Spine Society, 25-28 October 2006, Istanbul, Turkey - abstracts sp# 35. *European Spine Journal*. 2006; Vol. 15, issue Suppl 4:S502.
- Zigler 2007** {published data only}
- Auerbach JD, Jones KJ, Milby AH, Anakwenze OA, Balderston RA. Segmental contribution toward total lumbar range of motion in disc replacement and fusions: A comparison of operative and adjacent levels. *Spine* 2009;**34**(23):2510–7.
- Auerbach JD, Wills BP, McIntosh TC, Balderston RA. Evaluation of spinal kinematics following lumbar total disc replacement and circumferential fusion using in vivo fluoroscopy. *Spine* 2007;**32**(5):527–36.
- Auerbach JD, Wills BPD, McIntosh TC, Balderston RA. Lumbar disc replacement versus fusion for single-level degenerative disc disease: Two-year results from a randomized prospective study. *Seminars in Spine Surgery* 2005;**17**(4):310–8.
- Bendo JA, Quirno M, Errico T, Spivak JM, Goldstein J. A comparison of two retroperitoneal surgical approaches for total disc replacement of the lumbar spine. *Spine* 2008;**33**(2):205–9.
- Buttacavoli FA, Delamarter RB, Kanim LEA. Cost comparison of patients with 3-level artificial total lumbar disc replacements versus 360degree fusion at 3 contiguous lumbar vertebral levels: An analysis of compassionate use at 1 site of the US investigational device exemption clinical trial. *SAS Journal* 2010;**4**(4) : (pp 107-114), 2010. Date of Publication: December 2010.:114.
- Delamarter R, Zigler JE, Balderston RA, Cammisa FP, Goldstein JA, Spivak JM. Prospective, randomized, multicenter Food and Drug Administration investigational device exemption study of the ProDisc-L total disc replacement compared with circumferential arthrodesis for the treatment of two-level lumbar degenerative disc disease: results at twenty-four months. *Journal of Bone and Joint Surgery. American volume* 2011; Vol. 93, issue 8:705–15.
- Delamarter RB, Bae HW, Pradhan BB. Clinical results of ProDisc-II lumbar total disc replacement: Report from the United States clinical trial. *Orthopedic Clinics of North America* 2005;**36**(3 SPEC. ISS.):301–13.
- Delamarter RB, Fribourg DM, Kanim LE, Bae H. ProDisc artificial total lumbar disc replacement: introduction and early results from the United States clinical trial. *Spine* 2003;**28**(20):S167–75.
- Hannibal M, Thomas DJ, Low J, Hsu KY, Zucherman J.

ProDisc-L total disc replacement: a comparison of 1-level versus 2-level arthroplasty patients with a minimum 2-year follow-up. *Spine* 2007;**32**(21):2322–6.

Leahy M, Zigler JE, Ohnmeiss DD, Rashbaum RF, Sachs BL. Comparison of results of total disc replacement in postdiscectomy patients versus patients with no previous lumbar surgery. *Spine* 2008;**33**(15):1690–3.

Levin DA, Bendo JA, Quirno M, Errico T, Goldstein J, Spivak J. Comparative charge analysis of one- and two-level lumbar total disc replacement versus circumferential lumbar fusion. *Spine* 2007;**32**(25):2905–9.

Ohnmeiss DD, Bodemer W, Zigler JE. Effect of adverse events on low back surgery outcome: twenty-four-month follow-up results from a Food And Drug Administration investigational device exemption trial. *Spine* 2010;**35**(7): 835–8.

\* Zigler J, Delamarter R, Spivak JM, Linovitz RJ, Danielson GO III, Haider TT, et al. Results of the prospective, randomized, multicenter Food and Drug Administration investigational device exemption study of the ProDisc-L total disc replacement versus circumferential fusion for the treatment of 1-level degenerative disc disease. *Spine* 2007; **32**(11):1155–62.

Zigler JE. Lumbar spine arthroplasty using the ProDisc II. *Spine Journal: Official Journal of the North American Spine Society* 2004;**4**(6 Suppl):260S–7S.

Zigler JE, Burd TA, Vialle EN, Sachs BL, Rashbaum RF, Ohnmeiss DD. Lumbar spine arthroplasty: early results using the ProDisc II: a prospective randomized trial of arthroplasty versus fusion. *Journal of Spinal Disorders & Techniques* 2003;**16**(4):352–61.

## References to studies excluded from this review

**Cakir 2006** {*published data only*}

Cakir B, Richter M, Huch K, Puhl W, Schmidt R. Dynamic stabilization of the lumbar spine. *Orthopedics* 2006;**29**(8): 716–22.

**SariAli 2006** {*published data only*}

SariAli el-H, Lemaire JP, Pascal-Moussellard H, Carrier H, Skalli W. In vivo study of the kinematics in axial rotation of the lumbar spine after total intervertebral disc replacement: long-term results: a 10-14 years follow up evaluation. *European Spine Journal* 2006;**15**(10):1501–10.

**Schroven 2006** {*published data only*}

Schroven I, Dorofey D. Intervertebral prosthesis versus anterior lumbar interbody fusion: One-year results of a prospective non-randomised study. *Acta Orthopaedica Belgica* 2006;**72**(1):83–6.

**Shim 2007** {*published data only*}

Shim CS, Lee SH, Shin HD, Kang HS, Choi WC, Jung B, et al. CHARITE versus ProDisc: a comparative study of a minimum 3-year follow-up. *Spine* 2007;**32**(9):1012–8.

**Tournier 2007** {*published data only*}

Tournier C, Aunoble S, Le Huec JC, Lemaire JP, Tropiano P, Lafage V, et al. Total disc replacement: consequences for

sagittal balance and lumbar spine movement. *European Spine Journal* 2007;**16**(3):411–21.

## References to ongoing studies

**Aesculap Impl syst 2007** {*unpublished data only*}

Yue JJ, Mo FF. Clinical study to evaluate the safety and effectiveness of the Aesculap Activ-L artificial disc in the treatment of degenerative disc disease. *BMC Surgery* 2010; **10**:14.

**Axiomed Spine Corp 2008** {*unpublished data only*}

An IDE Study of the Freedom Lumbar Disc in the Treatment of Lumbar Degenerative Disc Disease. Ongoing study September 2008.

**Pioneer 2009b** {*unpublished data only*}

A Prospective, Multi-center, Randomized, Controlled Clinical Trial Evaluating the Safety and Effectiveness of NUBAC Disc Arthroplasty. Ongoing study February 2009.

**SpinalMotion 2006** {*unpublished data only*}

Clinical Study Protocol for the Investigation of the Kineflex Spinal System - a Pivotal Study in Continued Access Stage. Ongoing study January 2005.

## Additional references

**Alpizar-Aguirre 2008**

Alpizar-Aguirre A, Mireles-Cano JN, Rosales-Olivares M, Miramontes-Martinez V, Reyes-Sanchez A. Clinical and radiological follow-up of Nubac disc prosthesis. Preliminary report [Evaluacion clinica y radiologica de la protesis de nucleo discal Nubac. Informe preliminar]. *Cirugia y cirujanos* 2008;**76**(4):311–5.

**Ashton 1992**

Ashton IK, Ashton BA, Gibson SJ, Polak JM, Jaffray DC, Eisenstein SM. Morphological basis for back pain: the demonstration of nerve fibers and neuropeptides in the lumbar facet joint capsule but not in ligamentum flavum. *Journal of Orthopedic Research* 1992;**10**(1):72–8.

**Bertagnoli 2005**

Bertagnoli R, Yue JJ, Shah RV, Nanieva R, Pfeiffer F, Fenk-Mayer A, et al. The treatment of disabling single-level lumbar discogenic low back pain with total disc replacement utilizing the Prodisc prosthesis: a prospective study with 2-year minimum follow-up. *Spine* 2005;**30**(19):2230–6.

**Boos 1997**

Boos N, Webb JK. Pedicle screw fixation in spinal disorders: a European view. *European Spine Journal* 1997;**6**(1):2–18.

**Boutron 2005**

Boutron I, Moher D, Tugwell P, Giraudeau B, Poiraudou S, Nizard R, et al. A checklist to evaluate a report of a non pharmacological trial (CLEAR NPT) was developed using consensus. *Journal of Clinical Epidemiology* 2005;**58**: 1233–40.

**Buttner-Jan 1987**

Buttner-Jan K, Schellnack K, Zippel H. An alternative treatment strategy in lumbar intervertebral disk damage using an SB Charité modular type intervertebral disk

- endoprosthesis [Eine alternative Behandlungsstrategie beim lumbalen Bandscheibenschaden mit der Bandscheibenendoprothese Modultyp SB Charité]. *Zeitschrift für Orthopädie und ihre Grenzgebiete* 1987;**125**(1):1–6.
- Cavanaugh 1996**  
Cavanaugh JM, Ozaktay AC, Yamashita HT, King AI. Lumbar facet pain: biomechanics, neuroanatomy and neurophysiology. *Journal of Biomechanics* 1996;**29**(9): 1117–29.
- Chen 2009**  
Chen SH, Zhong ZC, Chen CS, Chen WJ, Hung C. Biomechanical comparison between lumbar disc replacement and fusion. *Medical Engineering & Physics* 2009;**31**(2):244–53.
- Chung 2006**  
Chung SS, Lee CS, Kang CS. Lumbar total disc replacement using ProDisc II: a prospective study with a 2-year minimum follow-up. *Journal of Spinal Disorders and Techniques* 2006;**19**(6):411–5.
- Chung 2009**  
Chung SK, Kim YE, Wang KC. Biomechanical effect of constraint in lumbar total disc replacement: a study with finite element analysis. *Spine* 2009;**34**(12):1281–6.
- Coppes 1997**  
Coppes MH, Marani E, Thomeer RT, Groen GJ. Innervation of "painful" lumbar discs. *Spine* 1997;**22**(20): 2342–9.
- Cunningham 2009**  
Cunningham BW, Hu N, Beatson HJ, Serhan H, Seftor JC, McAfee PC. Revision strategies for single- and two-level total disc replacement procedures: a biomechanical perspective. *Spine Journal: Official Journal of the North American Spine Society* 2009;**Epub**:ahead of print.
- David 2007**  
David T. Long-term results of one-level lumbar arthroplasty: minimum 10-year follow-up of the Charité artificial disc in 106 patients. *Spine* 2007;**32**(6):661–6.
- de Maat 2009**  
de Maat GH, Punt IM, Van Rhijn LW, Schurink GW, Van Ooij A. Removal of the Charité lumbar artificial disc prosthesis: surgical technique. *Journal of Spinal Disorders and Techniques* 2009;**22**(5):334–9.
- Delamarter 2005**  
Delamarter RB, Bae HW, Pradhan BB. Clinical results of ProDisc-II lumbar total disc replacement: report from the United States clinical trial. *The Orthopedic Clinics of North America* 2005;**36**(3):301–13.
- DePuy 2009**  
Charité Artificial disk. <http://www.depuyortho.com/products/spinearthroplasty/charite.asp> (accessed 31 July 2009).
- Dworkin 2009**  
Dworkin RH, Turk DC, McDermott MP, Peirce-Sandner S, Burke LB, Cowan P, Farrar JT, Hertz S, Raja SN, Rappaport BA, Rauschkolb C, Sampaio C. Interpreting the clinical importance of group differences in chronic pain clinical trials: IMMPACT recommendations. *Pain* 2009;**146**(3): 238–244.
- Erkan 2009**  
Erkan S, Rivera Y, Wu C, Mehdob AA, Transfeldt EE. Biomechanical comparison of a two-level Maverick disc replacement with a hybrid one-level disc replacement and one-level anterior lumbar interbody fusion. *Spine J* 2009. Epub 2009 May 28;**9**(10):830–5.
- Fernstrom 1966**  
Fernstrom U. Arthroplasty with intercorporeal endoprosthesis in herniated disc and in painful disc. *Acta Chirurgica Scandinavica* 1966;**Suppl**(357):154–9.
- Fraser 2004**  
Fraser RD, Ross ER, Lowery GL, Freeman BJ, Dolan M. AcroFlex design and results. *Spine Journal: Official Journal of the North American Spine Society* 2004;**4**(6 Suppl): 245S–51S.
- Freeman 2006**  
Freeman BJ, Davenport J. Total disc replacement in the lumbar spine: a systematic review of the literature. [Review] [44 refs]. *European Spine Journal* 2006;**15** Suppl 3:**S439–47**, 2006 Aug.:47.
- Fritzell 2001**  
Fritzell P, Hagg O, Wessberg P, Nordwall A. 2001 Volvo Award Winner in Clinical Studies: Lumbar fusion versus nonsurgical treatment for chronic low back pain: a multicenter randomized controlled trial from the Swedish Lumbar Spine Study Group. *Spine* 2001;**26**(23):2521–32.
- Furlan 2009**  
Furlan A, Pennick V, Bombardier C, Van Tulder M, Editorial Board, Cochrane Back Review Group. 2009 updated method guidelines for systematic reviews in the Cochrane Back Review Group. *Spine* 2009; Vol. 34, issue 18:1929–41.
- Gibson 2005**  
Gibson J, Waddell G. Surgery for degenerative lumbar spondylosis. *Cochrane Database of Systematic Reviews* 2005, Issue 4. [DOI: 10.1002/14651858.CD001352.pub3]
- Gibson 2005a**  
Gibson JN, Waddell G. Surgery for degenerative lumbar spondylosis: updated Cochrane Review. *Spine* 2005;**30** (20):2312–20.
- Goel 2005**  
Goel VK, Grauer JN, Patel TC, Biyani A, Sairyo K, Vishnubhotla S, et al. Effects of Charité artificial disc on the implanted and adjacent spinal segments mechanics using a hybrid testing protocol. *Spine* 2005;**30**(24):2755–64.
- GRADE Working Group 2004**  
GRADE Working Group. Grading quality of evidence and strength of recommendations. *BMJ* 2004;**328**:1490–4.
- Ha 2009**  
Ha SK, Kim SH, Kim DH, Park JY, Lim DJ, Lee SK. Biomechanical study of lumbar spinal arthroplasty with a semi-constrained artificial disc (activ L) in the human

- cadaveric spine. *Journal of the Korean Neurosurgical Society* 2009;**45**(3):169–75.
- Harrop 2008**  
Harrop JS, Youssef JA, Maltenfort M, Vorwald P, Jabbour P, Bono CM, et al. Lumbar adjacent segment degeneration and disease after arthrodesis and total disc replacement. *Spine* 2008;**33**(15):1701–7.
- Higgins 2011**  
Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from [www.cochrane-handbook.org](http://www.cochrane-handbook.org).
- Huang 2006**  
Huang RC, Tropiano P, Marnay T, Girardi FP, Lim MR, Cammisa FP Jr. Range of motion and adjacent level degeneration after lumbar total disc replacement. *Spine Journal: Official Journal of the North American Spine Society* 2006;**6**(3):242–7.
- Kafer 2008**  
Kafer W, Clessienne CB, Daxle M, Kocak T, Reichel H, Cakir B. Posterior component impingement after lumbar total disc replacement: a radiographic analysis of 66 ProDisc-L prostheses in 56 patients. *Spine* 2008;**33**(22):2444–9.
- Kuslich 1991**  
Kuslich SD, Ulstrom CL, Michael CJ. The tissue origin of low back pain and sciatica: a report of pain response to tissue stimulation during operations on the lumbar spine using local anesthesia. *The Orthopedic Clinics of North America* 1991;**22**(2):181–7.
- LDR 2009**  
Mobidisc. <http://www.ldrmedical.com/mobidisc.php?Langue=en> (accessed 31 July 2009).
- Le Huec 2005**  
Le Huec JC, Mathews H, Basso Y, Aunoble S, Hoste D, Bley B, et al. Clinical results of Maverick lumbar total disc replacement: two-year prospective follow-up. *The Orthopedic Clinics of North America* 2005;**36**(3):315–22.
- Leivseth 2006**  
Leivseth G, Braaten S, Frobin W, Brinckmann P. Mobility of lumbar segments instrumented with a ProDisc II prosthesis: a two-year follow-up study. *Spine* 2006;**31**(15):1726–33.
- Lemaire 2005**  
Lemaire JP, Carrier H, SariAli e-H, Skalli W, Lavaste F. Clinical and radiological outcomes with the Charité artificial disc: a 10-year minimum follow-up. *Journal of Spinal Disorders and Techniques* 2005;**18**(4):353–9.
- Moumene 2007**  
Moumene M, Geisler FH. Comparison of biomechanical function at ideal and varied surgical placement for two lumbar artificial disc implant designs: mobile-core versus fixed-core. *Spine* 2007;**32**(17):1840–51.
- Nachemson 1996**  
Nachemson A, Zdeblick TA, O'Brien JP. Lumbar disc disease with discogenic pain. What surgical treatment is most effective?. *Spine* 1996;**21**(15):1835–8.
- Ostelo 2008**  
Ostelo RW, Deyo RA, Stratford P, Waddell G, Croft P, Von Korf M, et al. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. *Spine* 2008;**33**(1):90–4.
- Park 2008**  
Park CK, Ryu KS, Jee WH. Degenerative changes of discs and facet joints in lumbar total disc replacement using ProDisc II: minimum two-year follow-up. *Spine* 2008;**33**(16):1755–61.
- Patel 2008**  
Patel AA, Brodke DS, Pimenta L, Bono CM, Hilibrand AS, Harrop JS, et al. Revision strategies in lumbar total disc replacement. *Spine* 2008;**33**(11):1276–83.
- Pioneer 2009a**  
Nubac. <http://www.pioneersurgical.com/products/spine/acp/nubac.asp?menu=2,5,0,22> (Accessed 31 July 2009).
- Punt 2008**  
Punt IM, Visser VM, Van Rhijn LW, Kurtz SM, Antonis J, Schurink GW, et al. Complications and re-operations of the SB Charité lumbar disc prosthesis: experience in 75 patients. *European Spine Journal* 2008;**17**(1):36–43.
- Putzier 2006**  
Putzier M, Funk JF, Schneider SV, Gross C, Tohtz SW, Khodadadyan-Klostermann C, et al. Charité total disc replacement—clinical and radiographic results after an average follow-up of 17 years. *European Spine Journal* 2006;**15**(2):183–95.
- Randolph 2006**  
Randolph GB, Scioscia TN, Wang JC. Lumbar total disc arthroplasty: state of the data. *Seminars in Spine Surgery* 2006;**18**(2):61–71.
- Rohlmann 2009**  
Rohlmann A, Mann A, Zander T, Bergmann G. Effect of an artificial disc on lumbar spine biomechanics: a probabilistic finite element study. *European Spine Journal* 2009;**18**(1):89–97.
- Ross 2007**  
Ross R, Mirza AH, Norris HE, Khatri M. Survival and clinical outcome of SB Charité III disc replacement for back pain. *Journal of Bone & Joint Surgery* 2007;**89**(6):785–9.
- Rousseau 2006**  
Rousseau MA, Bradford DS, Bertagnoli R, Hu SS, Lotz JC. Disc arthroplasty design influences intervertebral kinematics and facet forces. *Spine Journal: Official Journal of the North American Spine Society* 2006;**6**(3):258–66.



**Rundell 2008**

Rundell SA, Auerbach JD, Balderston RA, Kurtz SM. Total disc replacement positioning affects facet contact forces and vertebral body strains. *Spine* 2008;**33**(23):2510–7.

**Sedrakyan 2011**

Sedrakyan A, Normand SL, Dabic S, Jacobs S, Graves S, Marinac-Dabic D. Comparative assessment of implantable hip devices with different bearing surfaces: systematic appraisal of evidence. *British Journal of Medicine* 2011; **343**:d7434.

**Siepe 2006**

Siepe CJ, Mayer HM, Wiechert K, Korge A. Clinical results of total lumbar disc replacement with ProDisc II: three-year results for different indications. *Spine* 2006;**31**(17): 1923–32.

**Synthes Spine 2009**

Synthes Spine. <http://us.synthesprodisc.com/EN-US/DiscReplacement/ProDiscLumbar/> (Accessed 31 July 2009).

**Tropiano 2005**

Tropiano P, Huang RC, Girardi FP, Cammisa FP Jr, Marnay T. Lumbar total disc replacement. Seven to eleven-year follow-up. *Journal of Bone and Joint Surgery (Am)* 2005;**87**(3):490–6.

**Van den Eerenbeemt 2010**

Van den Eerenbeemt KD, Ostelo RW, Van Royen BJ, Peul WC, Van Tulder MW. Total disc replacement surgery for

symptomatic degenerative lumbar disc disease: a systematic review of the literature. *European Spine Journal* 2010; **19**(1432-0932 (Electronic), 0940-6719 (Linking), 8): 1262–80.

**Van Tulder 2003**

Van Tulder M, Furlan A, Bombardier C, Bouter L, Editorial Board Cochrane Back Review Group. Updated method guidelines for systemic reviews in the Cochrane Collaboration Back Review Group. *Spine* 2003;**28**(12): 1290–9.

**Yajun 2010**

Yajun W, Yue Z, Xiuxin H, Cui C. A meta-analysis of artificial total disc replacement versus fusion for lumbar degenerative disc disease. *European Spine Journal* 2010; **19**(1432-0932 (Electronic), 0940-6719 (Linking), 8): 1250–61.

**Zander 2009**

Zander T, Rohlmann A, Bergmann G. Influence of different artificial disc kinematics on spine biomechanics. *Clinical Biomechanics* 2009;**24**(2):135–42.

**References to other published versions of this review****de Kleuver 2003**

de Kleuver M, Oner FC, Jacobs WCH:. Total disc replacement for chronic low back pain: Background and a systematic review of the literature. *European Spine Journal* 2003;**12**(2):108–16.

\* Indicates the major publication for the study

## CHARACTERISTICS OF STUDIES

### Characteristics of included studies [author-defined order]

#### Berg 2009

Methods	RCT, method unclear, allocation concealed with sealed envelopes
Participants	LBP > 1 year, LBP > Leg pain, ODI > 30 or back pain VAS > 50 DDD, 1 or 2 level, MRI degeneration 20-55 years > 3 mo conservative treatment, no previous lumbar fusion, previous decompression No: spinal stenosis, spondylolysis, spondylolisthesis, facet joint arthritis, major deformity, osteoporosis, previous spinal infection, spinal tumour, compromised vertebral body
Interventions	Total disc replacement: Chrite, Prodisc or Maverick Control: Fusion with local or iliac crest bone graft, Pedicle screws, posterolateral or interbody fusion with Brantigan cage
Outcomes	VAS leg, VAS back, SF 36, ODI, EQ5D, work status @ preoperative, 1 and 2 years
Notes	Total disc replacement superior to fusion at one year, equal at 2 years Design and hypothesis not described regarding superiority/non-inferiority, tested for superiority

#### *Risk of bias*

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Not described. Investigator response: "The randomization was with pre-ordered sealed envelope technique, that means that before the study 200 envelopes, <b>with no possibility to look through them</b> were prepared. In 100 a folded paper with the line "fusion" was written, in the other 100 the folded paper had the line "TDR". All envelopes were then sealed. <b>The envelopes were then carefully scrambled</b> and the box with all the envelopes were delivered the our patient coordinators office, and never again came close to any of the health staff. When a patient that had entered the study was scheduled for surgery and admitted to the hospital, one of the coordinators draw one envelope, and that decided what type of surgery that was provided. There was also a stratification between the three different prostheses. If a patient was randomized for

		TDR, there were another two different boxes, one for one-level cases and one for two-level cases. Each of these boxes contained 90 pre-ordered sealed envelopes, 30 of the respective three brands of prostheses used for the study. The high number of envelopes in these “stratification-boxes” was due to the fact that we from the start did not know how the proportion of one- and two-level cases would actually be. <b>(upgraded)</b>
Allocation concealment (selection bias)	Low risk	Sealed envelopes
Blinding (performance bias and detection bias) All outcomes - Participants?	High risk	Not mentioned. Investigator response: “The patients were not blinded to the type of treatment that they were randomized to. That means that either the evening before surgery or at the latest the morning of the day of surgery, they were informed by their attending surgeon the result of the randomization. Since the two compared treatments had totally different approaches, it was not possible to blind the patients.”
Blinding (performance bias and detection bias) All outcomes - Care providers?	High risk	Not possible. Investigator response: “The care provider was not blinded, the same reason as above.”
Blinding (performance bias and detection bias) All outcomes - Observers?	High risk	Not mentioned. Investigator response: “Outcome was only assessed by the patients. The patients had questionnaires sent home to fill out, and the questionnaires were then sent the central registration office of the Swedish National Spine registry (SweSpine). The questionnaires thus never came into the hands of any of the care providers.”
Incomplete outcome data (attrition bias) All outcomes - drop-outs?	Low risk	
Incomplete outcome data (attrition bias) All outcomes - ITT analysis?	Low risk	
Selective reporting (reporting bias)	Unclear risk	No protocol was found
Characteristics similar at baseline?	High risk	VAS pain was significantly different at baseline. Investigator response: The two groups were at all baseline data (VAS for back-

**Berg 2009** (Continued)

		pain, Oswestry, SF-36, EQ5D, time of sick-leave, earlier surgery, work load, other sickness etc.) very similar. The only difference was at VAS for leg pain. This difference we have understood as a random significance, and since outcome regarding leg-pain was not the main focus of the study, less important
Co-interventions avoided or similar?	Low risk	
Compliance acceptable?	Low risk	
Timing of outcome assessments similar?	Low risk	

**Blumenthal 2005**

Methods	Multicentre RCT, computer-generated (SAS), 2:1, block size 6. Allocation concealment with sealed envelopes
Participants	Back and or leg pain ODI > 30, VAS > 40 DDD, single level, L4-S1, discography confirmed, 18-60 years, no morbid obesity > 6 months conservative treatment, no previous fusion TH1-S1, no chronic steroid use, no use of bone growth stimulator no: nerve root compression, noncontained HNP, spondylosis, spondylolisthesis > 3mm, scoliosis > 8°, positive straight leg test, spinal tumour, osteoporosis, osteopenia, metabolic bone disease, infection, psychosocial disorder, metal allergy, arachnoiditis, autoimmune disorder
Interventions	Total disc replacement: Charite Control: BAK (2 cages) with iliac crest autograft
Outcomes	VAS back pain, ODI, SF-36, neurological status, patient satisfaction @ preoperative, 6 weeks, 3, 6, 12 and 24 months ROM, disk height, neuroforaminal height, segmental translation @ preoperative, 3, 6, 12 and 24 months
Notes	Total disc replacement at least equivalent to fusion (ALIF) Non-inferiority trial

**Risk of bias**

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	

**Blumenthal 2005** (Continued)

Allocation concealment (selection bias)	Low risk	
Blinding (performance bias and detection bias) All outcomes - Participants?	High risk	Non-blinded study
Blinding (performance bias and detection bias) All outcomes - Care providers?	High risk	Not possible
Blinding (performance bias and detection bias) All outcomes - Observers?	High risk	nNon-blinded study
Incomplete outcome data (attrition bias) All outcomes - drop-outs?	Low risk	
Incomplete outcome data (attrition bias) All outcomes - ITT analysis?	Low risk	
Selective reporting (reporting bias)	Unclear risk	No protocol was found
Characteristics similar at baseline?	Low risk	
Co-interventions avoided or similar?	Low risk	
Compliance acceptable?	Low risk	
Timing of outcome assessments similar?	Low risk	

**Gornet 2011**

Methods	Allocation: Randomized Endpoint Classification: Safety/Efficacy Study Intervention Model: Parallel Assignment Masking: Open Label Primary Purpose: Treatment
Participants	DDD confirmed by plain radiography, CT or MRI Back pain w/wo leg pain L4-S1 Single level Oswestry > 30 Back pain > 20 > 6 month conservative treatment 18-70 years

Interventions	Device: MAVERICK Disc Device: Anterior fusion with LT-CAGE® Lumbar Tapered Fusion Device and INFUSE® Bone Graft (Fusion)	
Outcomes	Primary Outcome Measures: Overall Success = Oswestry Disability Index success; Maintenance/improvement in neurological status; Disc height success; No serious AE classified as implant or implant/surgical procedure associated; No additional surgical procedure classified as failure [ Time Frame: 24 Months ] [ Designated as safety issue: Yes ] Secondary Outcome Measures: SF-36; Back Pain Status; Leg Pain Status; Patient Satisfaction; Patient Global Perceived Effect; Other Measurements- Radiographic; Return to Work; Doctor's Perception of Results [ Time Frame: 24 Months ] [ Designated as safety issue: Yes ]	
Notes	Non-inferiority trial	
<b><i>Risk of bias</i></b>		
<b>Bias</b>	<b>Authors' judgement</b>	<b>Support for judgement</b>
Random sequence generation (selection bias)	Low risk	Not described, 2:1 fixed block size but not described how blocks were filled. Author response: "The sponsor (Medtronic) centrally ran the randomization schedule stratified by site with <b>SAS Proc Plan</b> . The randomization was in a 2:1 ratio and with a fixed block size of 6. The following is what described in the protocol: "Both the investigator and the patient will be blinded to the randomization prior to informed consent. Upon ensuring that the informed consent form has been properly completed, the investigator or designee will open the envelope that corresponds to the patient's assigned clinical trial number to determine if the patient will be randomized into the investigational or control group. The patient and surgeon will not be blinded following the opening of the treatment envelope." <b>(upgraded)</b>
Allocation concealment (selection bias)	Low risk	
Blinding (performance bias and detection bias) All outcomes - Participants?	Unclear risk	Not described
Blinding (performance bias and detection bias)	High risk	Not possible

**Gornet 2011** (Continued)

All outcomes - Care providers?		
Blinding (performance bias and detection bias) All outcomes - Observers?	Unclear risk	Not described
Incomplete outcome data (attrition bias) All outcomes - drop-outs?	Low risk	
Incomplete outcome data (attrition bias) All outcomes - ITT analysis?	Low risk	Only one patient crossed over, per protocol analysis was performed
Selective reporting (reporting bias)	Unclear risk	No protocol was found
Characteristics similar at baseline?	Low risk	
Co-interventions avoided or similar?	Low risk	
Compliance acceptable?	Low risk	
Timing of outcome assessments similar?	Low risk	

**Hellum 2011**

Methods	Allocation: Randomized Control: Active Control Endpoint Classification: Safety/Efficacy Study Intervention Model: Parallel Assignment Masking: Single Blind Primary Purpose: Treatment
Participants	Chronic Low-back pain Degenerative Disc Disease confirmed by imaging
Interventions	Device: Total Disc Prosthesis Behavioral: Multidisciplinary rehabilitation program
Outcomes	Primary Outcome Measures: Oswestry Disability Index Cost-effectiveness (full economic analysis) Secondary Outcome Measures: Side effects Time out of work Fear - avoidance beliefs Self - efficacy for pain Hopkins symptoms check list Drug consumption Life satisfaction (EQ 5D and SF - 36)
Notes	Superiority trial

<i>Risk of bias</i>		
<b>Bias</b>	<b>Authors' judgement</b>	<b>Support for judgement</b>
Random sequence generation (selection bias)	Low risk	Not described: "They were randomized in Blocks", but unclear how the blocks were filled. Investigator response: "The size of the blocks at each center varied from 2 to 8." ( <b>upgraded</b> )
Allocation concealment (selection bias)	Unclear risk	Uncertainty about period between disclosure and intervention. Sixteen participants did not start allocated intervention. Investigator response: " <b>Group allocation was generated by a web-site.</b> A coordinating secretary, without information about the patient, conducted the web-based randomisation after the patient were found eligible and had signed written informed consent. Thereafter, the coordinating secretary informed the patient about group allocation. All patients should receive their treatment before 3 months after disclosure/randomization. Reasons for patients not receiving allocated intervention are listed in the flow-chart fig. 1 in the paper."
Blinding (performance bias and detection bias) All outcomes - Participants?	High risk	Not possible
Blinding (performance bias and detection bias) All outcomes - Care providers?	High risk	Not possible
Blinding (performance bias and detection bias) All outcomes - Observers?	High risk	Only 2 years Prolo score and back performance scale blinded assessment was performed. Not described for all other outcomes. Investigator response: "At two year follow-up (the studies main endpoint), <b>all data (also questionnaires / main outcome (ODI)) were collected by blinded evaluators.</b> Data at all other follow-ups were collected by coordinating secretaries not involved in the trial.  All data were punched into SPSS by persons without knowledge about the trial (university students from non-medical disciplines)



**Hellum 2011** (Continued)

		“
Incomplete outcome data (attrition bias) All outcomes - drop-outs?	Low risk	
Incomplete outcome data (attrition bias) All outcomes - ITT analysis?	Low risk	
Selective reporting (reporting bias)	Unclear risk	No protocol was found. Investigator response: "We published our protocol www.clinicaltrial.gov NCT 00394732. We also sent British Medical Journal our protocol shortly after we submitted the manuscript to the BMJ. Other publications were not performed."
Characteristics similar at baseline?	High risk	Pain scores (primary outcome) and SF36 scores different between groups. Investigator response: "In our original protocol, we did not prespecify that differences considered to be by chance (due to the randomization process) should be adjusted for. Therefore British Medical Journal did not want to publish adjusted values. When adjusting for these differences, the difference in outcome between groups were larger in favour of surgery."
Co-interventions avoided or similar?	Low risk	
Compliance acceptable?	Low risk	
Timing of outcome assessments similar?	Low risk	

**Moreno 2008**

Methods	RCT, method unclear
Participants	Chronic LBP DDD, single level, L4-S1, no instability, Modic 1 or 2 on MRI < 55 years > 6 months conservative treatment, previous discectomy or chemonucleolysis allowed no spondylolisthesis, no HNP, no stenosis
Interventions	Total disc replacement: Charite III Control: ALIF with KLA cage, iliac crest autograft and anterior plate

**Moreno 2008** (Continued)

Outcomes	VAS, ODI @ preoperative, 6 months and latest (23 or 26 months) Return to work @ 6 months Satisfaction @ latest	
Notes	Conclusions: total disc replacement is possible alternative to fusion in severe DDD in young participant. More rapid and superior results Design and hypothesis not described regarding superiority/non-inferiority; tested for superiority	
<b><i>Risk of bias</i></b>		
<b>Bias</b>	<b>Authors' judgement</b>	<b>Support for judgement</b>
Random sequence generation (selection bias)	Unclear risk	Not described
Allocation concealment (selection bias)	Unclear risk	Not described
Blinding (performance bias and detection bias) All outcomes - Participants?	High risk	Allocation was clear because of insurance issues
Blinding (performance bias and detection bias) All outcomes - Care providers?	High risk	Not possible in surgery trials
Blinding (performance bias and detection bias) All outcomes - Observers?	High risk	Not described
Incomplete outcome data (attrition bias) All outcomes - drop-outs?	Low risk	
Incomplete outcome data (attrition bias) All outcomes - ITT analysis?	Low risk	
Selective reporting (reporting bias)	Unclear risk	No protocol was found
Characteristics similar at baseline?	Unclear risk	Sufficient information not given
Co-interventions avoided or similar?	Low risk	
Compliance acceptable?	Low risk	
Timing of outcome assessments similar?	High risk	Follow-up differed between groups (23 vs 26 months)

**Sasso 2008**

Methods	Multicentre RCT, method unclear
Participants	Discogenic pain, Axial pain > radicular pain DDD, single level, L1-S1, MRI/CT/FEX instability (translational >2 mm, rotational >5°, height loss >2 mm) 18-60, skeletally mature, BMI <40 Conservative treatment failed, no previous decompression or microdiscectomy No: spondylolisthesis, stenosis, Scoliosis > 10 degrees, metabolic bone disease, RA, HIV, Malignancies
Interventions	Total disc replacement: Flexicore Control: Circumferential fusion with femoral ring allograft, pedicle screws and iliac crest autograft
Outcomes	ODI, VAS back pain @ Preop, 6 weeks, 3, 6, 12 and 24 months Blood loss, operation time and length of stay ROM on flexion extension x-rays @ preoperative and 6 weeks
Notes	This study reports on a sub sample (2 sites) Conclusion: Flexicore compares very favourably to circumferential fusion for lumbar DDD Design and hypothesis not described regarding superiority/non-inferiority; clinical outcome not tested

***Risk of bias***

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Not described
Allocation concealment (selection bias)	Unclear risk	Not described
Blinding (performance bias and detection bias) All outcomes - Participants?	Unclear risk	Not described
Blinding (performance bias and detection bias) All outcomes - Care providers?	High risk	Not possible
Blinding (performance bias and detection bias) All outcomes - Observers?	Unclear risk	Not described
Incomplete outcome data (attrition bias) All outcomes - drop-outs?	High risk	Subgroup of larger trial

Incomplete outcome data (attrition bias) All outcomes - ITT analysis?	Unclear risk	Not mentioned
Selective reporting (reporting bias)	Unclear risk	No protocol was found
Characteristics similar at baseline?	Unclear risk	Insufficient information given
Co-interventions avoided or similar?	Low risk	
Compliance acceptable?	Low risk	
Timing of outcome assessments similar?	Low risk	

**Zigler 2007**

Methods	Multicentre RCT, method unclear, allocation concealment unclear
Participants	Back and/or leg (radicular) pain, ODI > 40, DDD, single level, L3-S1, CT/MRI/diskography/FEX (instability > 3 mm translation, > 5° angulation; disc height decrease > 2 mm; scarring/thickening annulus; HNP; vacuum phenomenon) 18-60 years > 6 months conservative therapy, no previous fusion No: vertebral endplates < 34.5x27 mm, metal allergy, spondylolisthesis, spondylosis, osteoporosis, Paget, osteomalacia, metabolic bone disease, deg spondylolisthesis > 1, BMI > 40, AIDS, HIV, active malignancy
Interventions	Total disc replacement: Prodisc-L Control: Circumferential fusion with femoral ring allograft with pedicle screws and iliac crest autograft
Outcomes	VAS, ODI, SF-36, Physical exam, neurological exam @ preoperative, postoperative, 6 weeks, 3, 6, 12, 18 and 24 months
Notes	Conclusion: Prodisc-L is safe and efficacious. In selected patients better than circumferential fusion Non-inferiority trial

**Risk of bias**

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Not described - Investigator response: "Randomization was performed external to both the sponsor and the investigator sites by an independent third party, in overlapping blocks of six patients at each site. The

		<p>randomization schedule was not available to either the sponsor or the investigator at any time during the study. Patient demographics showed so significant differences between the study group populations. The schedule was generated externally, so I have no idea what the technical aspects were. <b>I would assume it was computer generated because blocks of six were used for each site, with each site maintaining an internal 2:1 randomization. Thus, even if the last two patients had randomized to fusion, there was no predictor that the next patient would get an ADR because the fusion patients might have been at the end of a six-block, and next patient might be at the beginning of the following block. This is a perfect job for a computer-generated algorithm.</b> (upgraded)</p>
<p>Allocation concealment (selection bias)</p>	<p>High risk</p>	<p>Sponsor provided allocation after consent. Investigator response: "Allocation was made available to the investigator and sponsor only after the patient was enrolled. Allocation to the investigator was necessary for surgical planning, as the investigational surgery required two hours of Operating Room time, and the Control arm procedure required 4-5 hours of OR time, utilizing different instruments and implants. There were no withdrawals due to surgeon displeasure. There were approximately 30 additional patients who were enrolled but not treated, the majority due to insurance denials for surgery, completion of enrolment by the time their insurance was approved, and a small number due to improvement in symptoms by the time insurance was approved. Since they received no surgery, and were not followed any longer, there is no postop data to include."</p>
<p>Blinding (performance bias and detection bias) All outcomes - Participants?</p>	<p>High risk</p>	<p>This was a non blinded study. Investigator response: "Patients were blinded through the enrolment and preoperative processes, and were not told which group they were allocated into until they awoke from surgery. No patient was informed prior to surgery,</p>

		and there were consequently no patient withdrawals among the 292 subjects due to displeasure with their allocation.“
Blinding (performance bias and detection bias) All outcomes - Care providers?	High risk	Not possible. Investigator response: “Care providers were not blinded.“
Blinding (performance bias and detection bias) All outcomes - Observers?	Unclear risk	Not described. Investigator response: “The study results depended upon a complex endpoint of 10 separate success criteria. Investigator input was limited to only one category, Neurologic Status (consisting of recording the motor, sensory, and reflex exams as well as a Straight Leg Raise test). The remaining nine success criteria were either patient generated (VAS pain, VAS satisfaction, ODI, SF-36) or recorded by independent radiologists based upon 5 radiographic success criteria determined by their off-site review of digitized films.“
Incomplete outcome data (attrition bias) All outcomes - drop-outs?	Low risk	Investigator response: “Five patients were treated off-protocol (for example, a two-level fusion decision made by the Operating Surgeon at the time of surgery). Their data was followed for safety, but was excluded from the efficacy calculation.“
Incomplete outcome data (attrition bias) All outcomes - ITT analysis?	Low risk	
Selective reporting (reporting bias)	Unclear risk	No protocol was found. Investigator response: “This study had a 98% follow-up rate at the two-year postop reporting threshold for the FDA. As primary author, I was satisfied that this extraordinarily high follow-up rate ensured that all available data was included in the analysis from all sites, and that there was no significant opportunity for reporting bias.“
Characteristics similar at baseline?	Low risk	
Co-interventions avoided or similar?	Low risk	
Compliance acceptable?	Low risk	

**Zigler 2007** (Continued)

Timing of outcome assessments similar?	Low risk
----------------------------------------	----------

AE: adverse effects  
BMI: body mass index  
CT: computed tomography  
DDD: degenerative disc disease  
EQD: Euroqol  
HNP: herniated nucleus pulposus  
LBP: low-back pain  
mo: month  
MRI: magnetic resonance imaging  
ODI: Oswestry Disability Index  
RA: rheumatoid arthritis  
RCT: randomised controlled trial  
ROM: range of motion  
SF: short form  
VAS: visual analogue scale  
vs versus  
w/wo: with/without

**Characteristics of excluded studies** [ordered by study ID]

Study	Reason for exclusion
Cakir 2006	Non-randomized; specifically defined different selection criteria for the two interventions and this consequently resulted in two different groups with associated distinct age categories
SariAli 2006	No clinical data.
Schroven 2006	Non-randomized.
Shim 2007	Retrospective; comparison between disc arthroplasties.
Tournier 2007	No clinical data.

## Characteristics of ongoing studies *[ordered by study ID]*

### Aesculap Impl syst 2007

Trial name or title	Clinical Study to Evaluate the Safety and Effectiveness of the Aesculap Activ-L Artificial Disc in the Treatment of Degenerative Disc Disease
Methods	Allocation: Randomized Control: Active Control Endpoint Classification: Safety/Efficacy Study Intervention Model: Parallel Assignment Masking: Single Blind (Subject) Primary Purpose: Treatment
Participants	Degenerative Disc Disease
Interventions	Device: Activ-L Artificial Disc Device: ProDisc Total Disc Replacement or Charite Artificial Disc
Outcomes	Primary Outcome Measures: Improvement of at least 15 points in the Oswestry Disability Index score at 24 months compared to baseline [ Time Frame: 24 months ] Secondary Outcome Measures: Back Pain, measured at rest using a visual analogue scale (VAS); improvement of 20 mm or more on a 100 mm VAS scale at 24 months compared to baseline [ Time Frame: 24 months ]
Starting date	January 2007
Contact information	
Notes	

### Axiomed Spine Corp 2008

Trial name or title	An IDE Study of the Freedom Lumbar Disc in the Treatment of Lumbar Degenerative Disc Disease
Methods	Allocation: Randomized Endpoint Classification: Safety/Efficacy Study Intervention Model: Parallel Assignment Masking: Open Label Primary Purpose: Treatment
Participants	Degenerative Disc Disease (DDD)
Interventions	Device: FLD Device: Control
Outcomes	Primary Outcome Measures: Overall success will be determined by a composite of measures regarding subject self-assessment of function (disability), low-back pain, neurological function and device implantation status. [ Time Frame: 24 Months ] [ Designated as safety issue: Yes ] Secondary Outcome Measures: The improvement of subject self-assessment of function (disability), low-back pain, patient satisfaction, neurological function and device implantation status at the 24 month follow-up



**Axiomed Spine Corp 2008** (Continued)

	compared to baseline. [ Time Frame: 24 Months ] [ Designated as safety issue: Yes ]
Starting date	September 2008
Contact information	
Notes	

**Pioneer 2009b**

Trial name or title	A Prospective, Multi-center, Randomized, Controlled Clinical Trial Evaluating the Safety and Effectiveness of NUBAC Disc Arthroplasty
Methods	Allocation: Randomized Control: Active Control Endpoint Classification: Safety/Efficacy Study Intervention Model: Single Group Assignment Masking: Single Blind (Subject) Primary Purpose: Treatment
Participants	Degenerative Disc Disease
Interventions	Device: NUBAC Device: Prodisc
Outcomes	Primary Outcome Measures: Improved patient function [ Time Frame: 6 weeks, 3, 6, 12, 24 months, annually thereafter ]
Starting date	February 2009
Contact information	
Notes	

**SpinalMotion 2006**

Trial name or title	Clinical Study Protocol for the Investigation of the Kineflex Spinal System - a Pivotal Study in Continued Access Stage
Methods	Allocation: Randomized Control: Active Control Endpoint Classification: Safety/Efficacy Study Intervention Model: Parallel Assignment Masking: Single Blind (Subject) Primary Purpose: Treatment
Participants	Degenerative Disc Disease

**SpinalMotion 2006** (Continued)

Interventions	Device: Lumbar Artificial Disc Device: Charite Artificial Disc
Outcomes	Primary Outcome Measures: Improvement in Oswestry Low-Back Pain Disability Score at 24 months compared with baseline; no revision removal, supplemental fixation or device-related re-operations and no major adverse event as defined by the study protocol [ Time Frame: 24 months ] [ Designated as safety issue: Yes ] Secondary Outcome Measures: Maintenance or improvement in neurologic status [ Time Frame: 24 months ] [ Designated as safety issue: Yes ] Pain improvement [ Time Frame: 24 months ] Significant disc height increase [ Time Frame: 24 months ] No displacement or migration of the device [ Time Frame: 24 months ] Time to return to work [ Time Frame: 24 months ] Time to recovery [ Time Frame: 24 months ] Preservation of at least 4 degrees of motion in flexion/extension [ Time Frame: 24 months ]
Starting date	January 2005
Contact information	
Notes	

## DATA AND ANALYSES

### Comparison 1. Disc replacement versus fusion

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Back pain at 6 Months	2	706	Mean Difference (IV, Random, 95% CI)	-6.44 [-10.55, -2.34]
2 Back pain at 24 Months	2	676	Mean Difference (IV, Random, 95% CI)	-5.12 [-9.60, -0.64]
3 Improvement in back pain at 24 months	2	676	Mean Difference (IV, Random, 95% CI)	5.22 [0.18, 10.26]
4 Leg pain at 24 Months	1	524	Mean Difference (IV, Random, 95% CI)	-3.60 [-8.84, 1.64]
5 Improvement in leg pain at 24 Months	2	676	Mean Difference (IV, Random, 95% CI)	0.56 [-4.54, 5.65]
6 Pain at 6 months	1	275	Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
7 Pain at 24 months	2	500	Mean Difference (IV, Random, 95% CI)	-6.16 [-11.98, -0.35]
8 Improvement in pain at 24 months	3	532	Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
9 Overall improvement	4		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
10 Patient satisfaction	4	958	Odds Ratio (M-H, Random, 95% CI)	1.93 [1.36, 2.76]
11 Oswestry at 6 Months	2	309	Mean Difference (IV, Random, 95% CI)	-9.90 [-14.31, -5.49]
12 Oswestry at 24 Months	5	1207	Mean Difference (IV, Random, 95% CI)	-6.31 [-9.68, -2.94]
13 Improvement in Oswestry score at 24 months	5	1207	Mean Difference (IV, Random, 95% CI)	4.27 [1.85, 6.68]
14 Percentage of patients improved on Oswestry at 24 months	5	1244	Odds Ratio (M-H, Random, 95% CI)	1.45 [1.06, 1.98]
15 Improvement in working status at 24 months	4	1188	Odds Ratio (M-H, Random, 95% CI)	1.02 [0.66, 1.59]
16 Implant motion	2	653	Mean Difference (IV, Random, 95% CI)	8.3 [7.02, 9.58]
17 Blood loss	5	1301	Mean Difference (IV, Random, 95% CI)	-37.22 [-185.06, 110.62]
18 Reoperations	5	1194	Odds Ratio (M-H, Random, 95% CI)	0.80 [0.51, 1.24]
19 Adjacent segment degeneration	1	152	Odds Ratio (M-H, Random, 95% CI)	0.14 [0.02, 1.19]
20 Facet joint degeneration	1	152	Odds Ratio (M-H, Random, 95% CI)	12.65 [0.70, 228.67]
21 Radiographic loosening (radiolucency)	1	236	Odds Ratio (M-H, Random, 95% CI)	0.15 [0.01, 3.82]
22 Subsidence	3	912	Odds Ratio (M-H, Random, 95% CI)	0.38 [0.18, 0.79]

### Comparison 2. Disc arthroplasty versus rehabilitation

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Back Pain at 12 months	1	172	Mean Difference (IV, Random, 95% CI)	-17.6 [-26.12, -9.08]
2 Back Pain at 24 months	1	172	Mean Difference (IV, Random, 95% CI)	-14.30 [-22.89, -5.71]
3 Oswestry at 12 months	1	172	Mean Difference (IV, Random, 95% CI)	-8.90 [-13.03, -4.77]

Total disc replacement for chronic back pain in the presence of disc degeneration (Review)  
Copyright © 2013 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

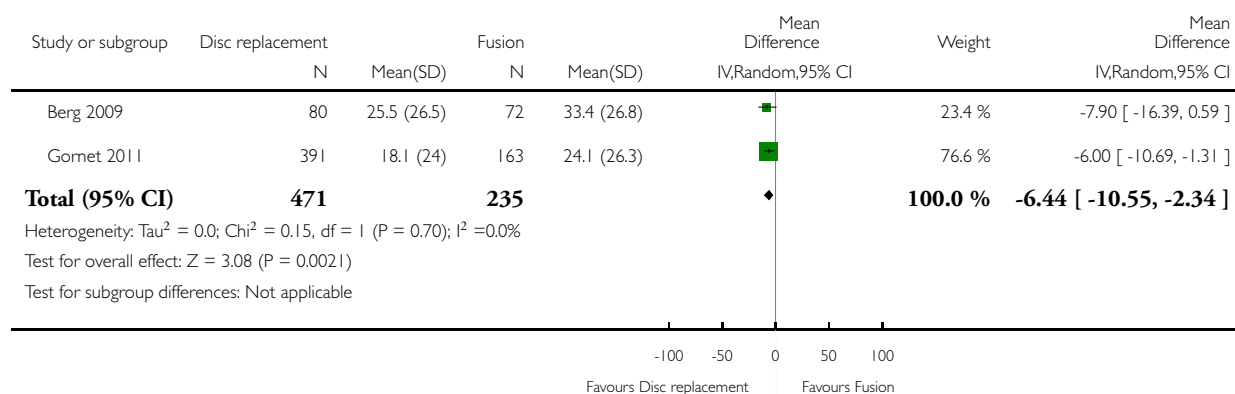
4 Oswestry at 24 months	1	172	Mean Difference (IV, Random, 95% CI)	-6.90 [-11.57, -2.23]
5 Patient satisfaction	1	172	Odds Ratio (M-H, Random, 95% CI)	2.65 [1.42, 4.96]
6 Improvement in working status	1	172	Risk Ratio (M-H, Random, 95% CI)	0.92 [0.78, 1.07]
7 Reoperations	1	157	Odds Ratio (M-H, Random, 95% CI)	1.04 [0.29, 3.75]

### Analysis 1.1. Comparison 1 Disc replacement versus fusion, Outcome 1 Back pain at 6 Months.

Review: Total disc replacement for chronic back pain in the presence of disc degeneration

Comparison: 1 Disc replacement versus fusion

Outcome: 1 Back pain at 6 Months

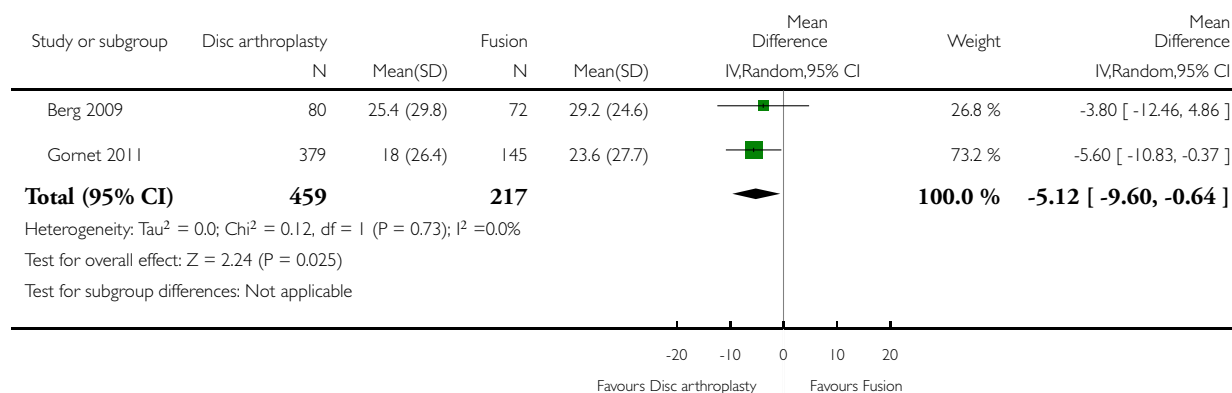


### Analysis 1.2. Comparison 1 Disc replacement versus fusion, Outcome 2 Back pain at 24 Months.

Review: Total disc replacement for chronic back pain in the presence of disc degeneration

Comparison: 1 Disc replacement versus fusion

Outcome: 2 Back pain at 24 Months

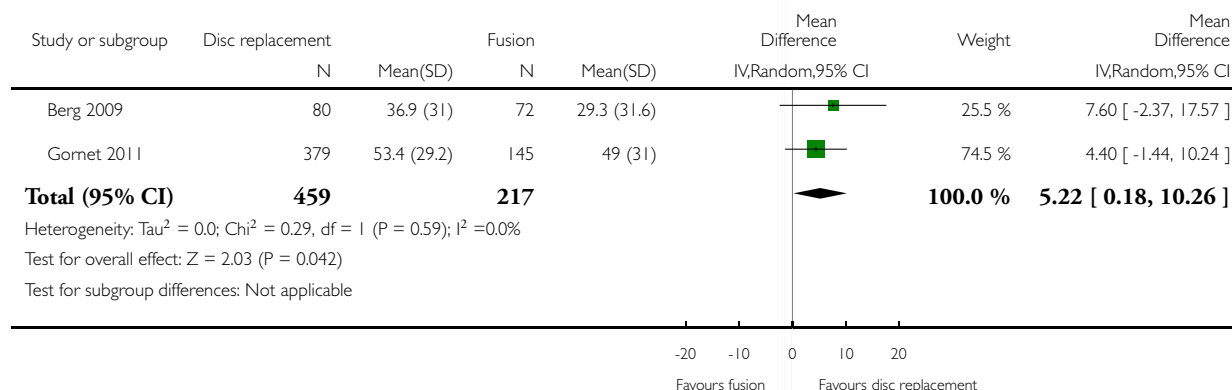


### Analysis 1.3. Comparison 1 Disc replacement versus fusion, Outcome 3 Improvement in back pain at 24 months.

Review: Total disc replacement for chronic back pain in the presence of disc degeneration

Comparison: 1 Disc replacement versus fusion

Outcome: 3 Improvement in back pain at 24 months

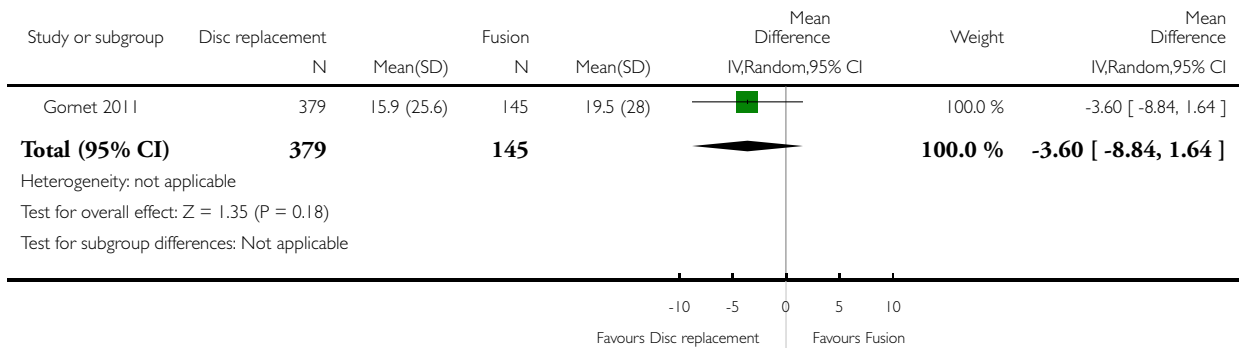


### Analysis 1.4. Comparison 1 Disc replacement versus fusion, Outcome 4 Leg pain at 24 Months.

Review: Total disc replacement for chronic back pain in the presence of disc degeneration

Comparison: 1 Disc replacement versus fusion

Outcome: 4 Leg pain at 24 Months

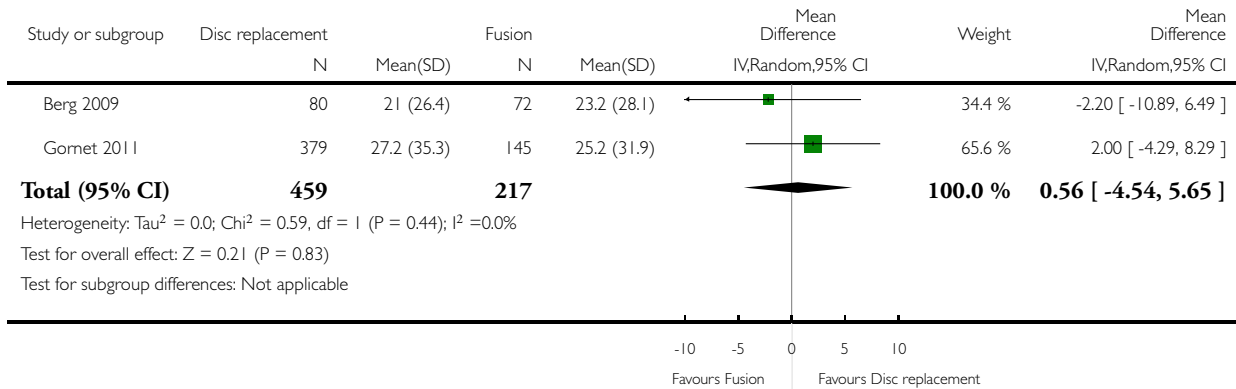


### Analysis 1.5. Comparison 1 Disc replacement versus fusion, Outcome 5 Improvement in leg pain at 24 Months.

Review: Total disc replacement for chronic back pain in the presence of disc degeneration

Comparison: 1 Disc replacement versus fusion

Outcome: 5 Improvement in leg pain at 24 Months

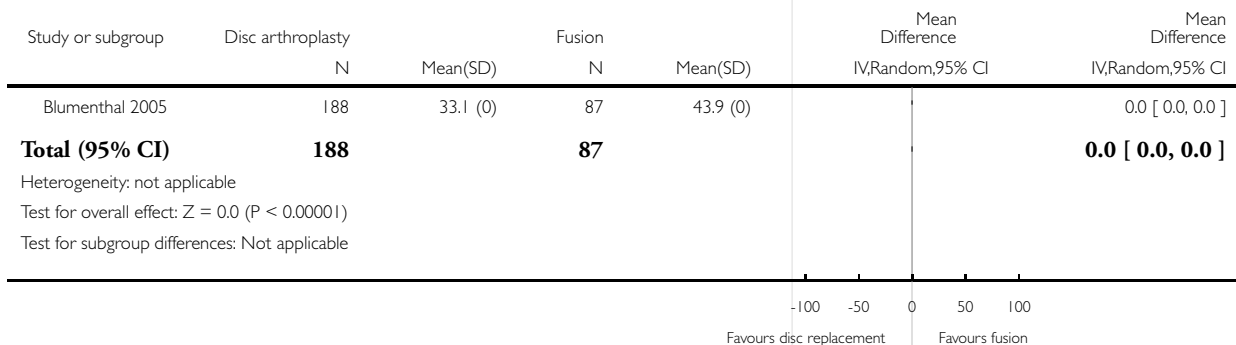


### Analysis 1.6. Comparison 1 Disc replacement versus fusion, Outcome 6 Pain at 6 months.

Review: Total disc replacement for chronic back pain in the presence of disc degeneration

Comparison: 1 Disc replacement versus fusion

Outcome: 6 Pain at 6 months

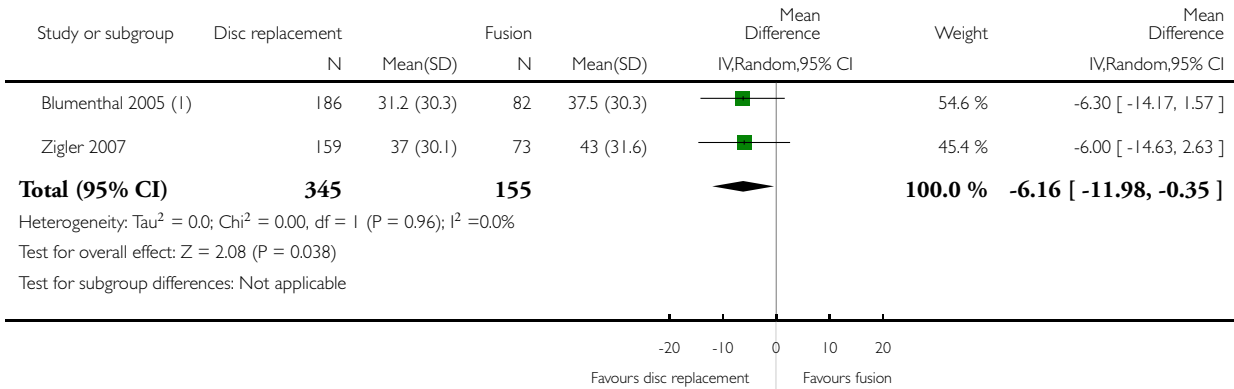


### Analysis 1.7. Comparison 1 Disc replacement versus fusion, Outcome 7 Pain at 24 months.

Review: Total disc replacement for chronic back pain in the presence of disc degeneration

Comparison: 1 Disc replacement versus fusion

Outcome: 7 Pain at 24 months



(1) SD Blumenthal imputed from Zigler

### Analysis 1.8. Comparison 1 Disc replacement versus fusion, Outcome 8 Improvement in pain at 24 months.

Review: Total disc replacement for chronic back pain in the presence of disc degeneration

Comparison: 1 Disc replacement versus fusion

Outcome: 8 Improvement in pain at 24 months



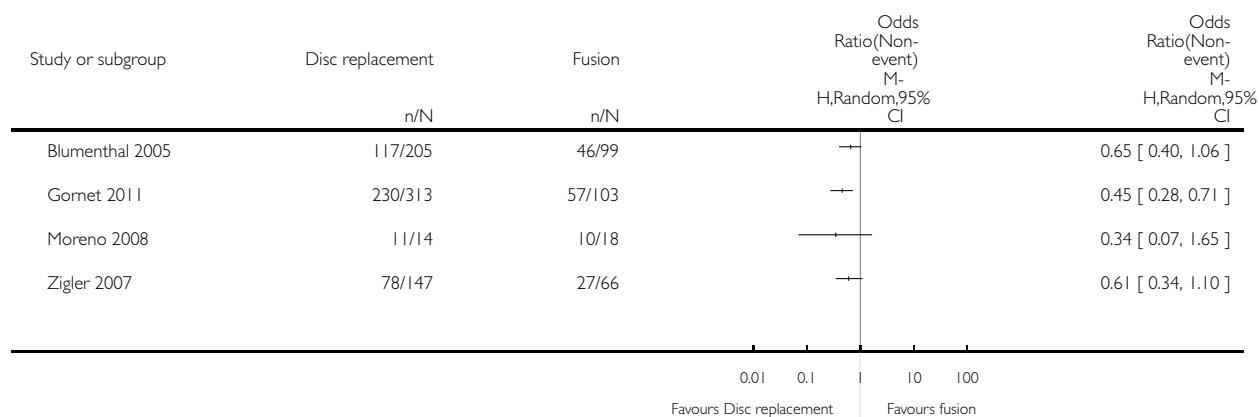


### Analysis 1.9. Comparison 1 Disc replacement versus fusion, Outcome 9 Overall improvement.

Review: Total disc replacement for chronic back pain in the presence of disc degeneration

Comparison: 1 Disc replacement versus fusion

Outcome: 9 Overall improvement

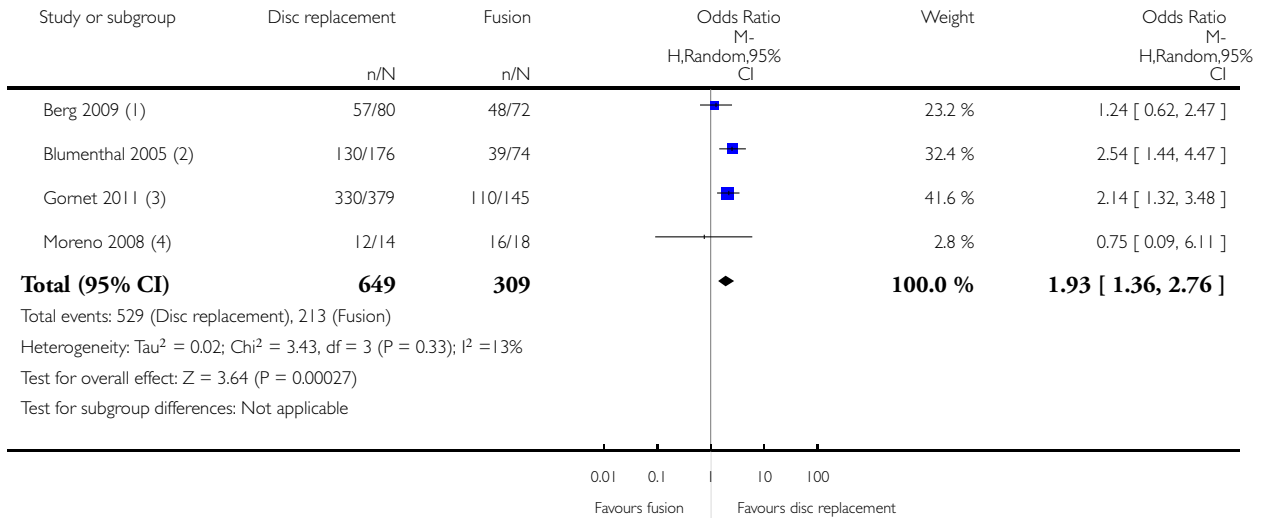


### Analysis 1.10. Comparison 1 Disc replacement versus fusion, Outcome 10 Patient satisfaction.

Review: Total disc replacement for chronic back pain in the presence of disc degeneration

Comparison: 1 Disc replacement versus fusion

Outcome: 10 Patient satisfaction



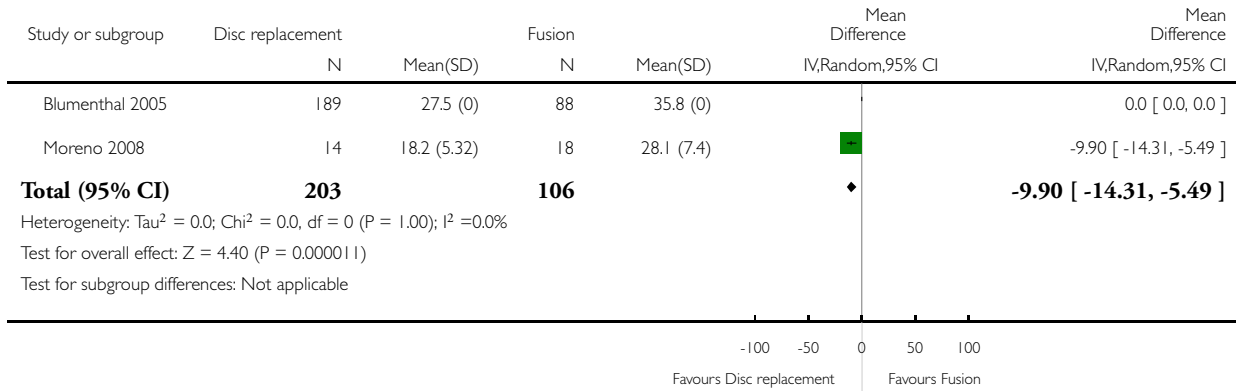
- (1) Satisfied with the treatment
- (2) Would have same treatment again
- (3) Satisfied
- (4) Would you recommend?

### Analysis 1.11. Comparison 1 Disc replacement versus fusion, Outcome 11 Oswestry at 6 Months.

Review: Total disc replacement for chronic back pain in the presence of disc degeneration

Comparison: 1 Disc replacement versus fusion

Outcome: 11 Oswestry at 6 Months

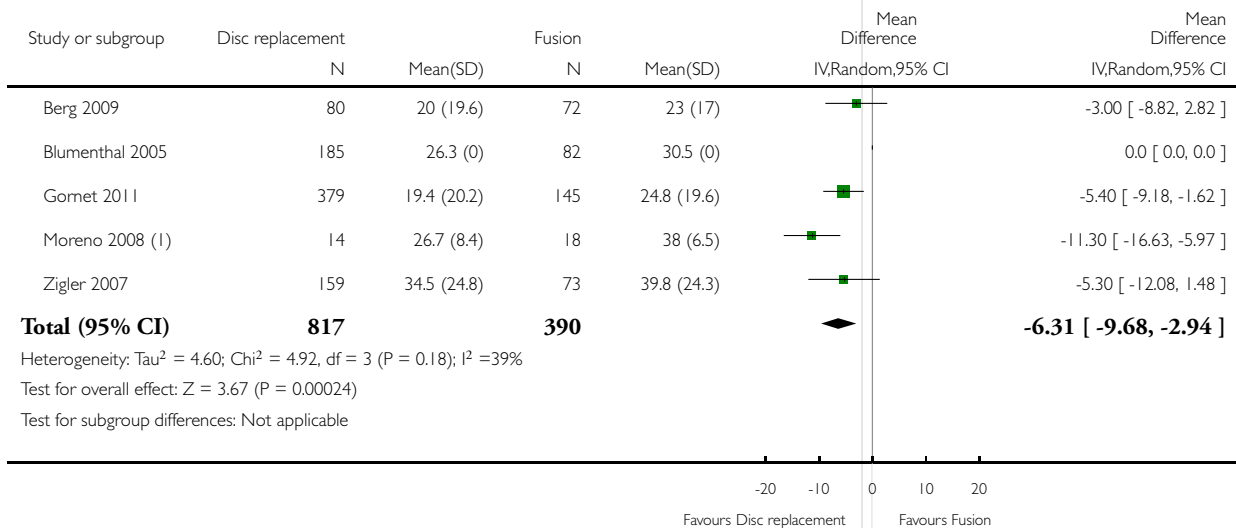


### Analysis 1.12. Comparison 1 Disc replacement versus fusion, Outcome 12 Oswestry at 24 Months.

Review: Total disc replacement for chronic back pain in the presence of disc degeneration

Comparison: 1 Disc replacement versus fusion

Outcome: 12 Oswestry at 24 Months



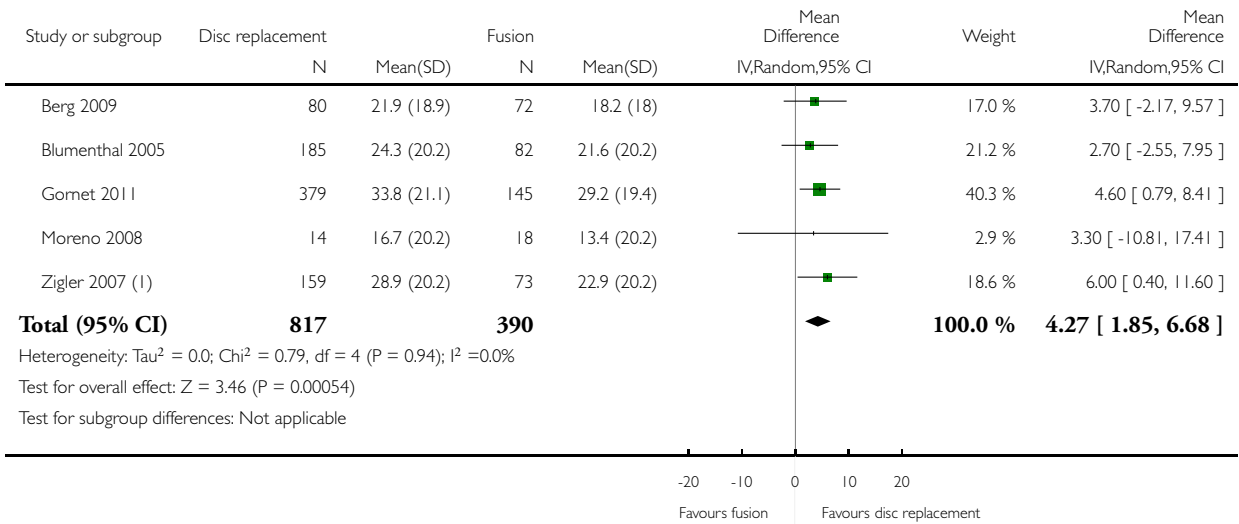
(I) Scores converted from a 60 point scale

**Analysis 1.13. Comparison 1 Disc replacement versus fusion, Outcome 13 Improvement in Oswestry score at 24 months.**

Review: Total disc replacement for chronic back pain in the presence of disc degeneration

Comparison: 1 Disc replacement versus fusion

Outcome: 13 Improvement in Oswestry score at 24 months



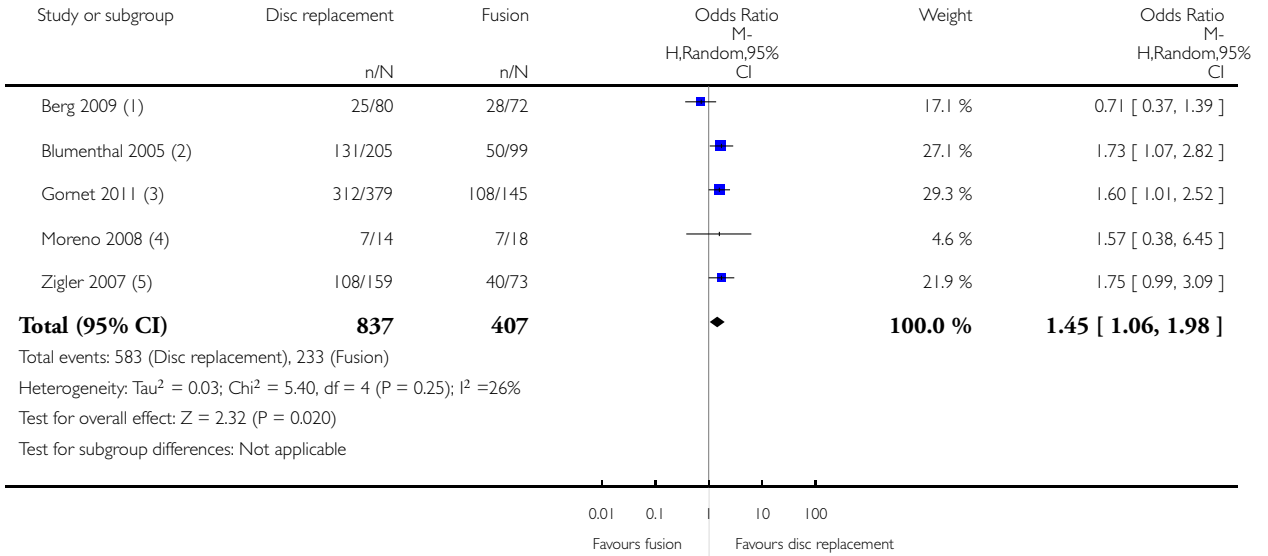
(I) Statistical significance differs from publication, probably due to multiple testing correction in the paper.

**Analysis 1.14. Comparison 1 Disc replacement versus fusion, Outcome 14 Percentage of patients improved on Oswestry at 24 months.**

Review: Total disc replacement for chronic back pain in the presence of disc degeneration

Comparison: 1 Disc replacement versus fusion

Outcome: 14 Percentage of patients improved on Oswestry at 24 months



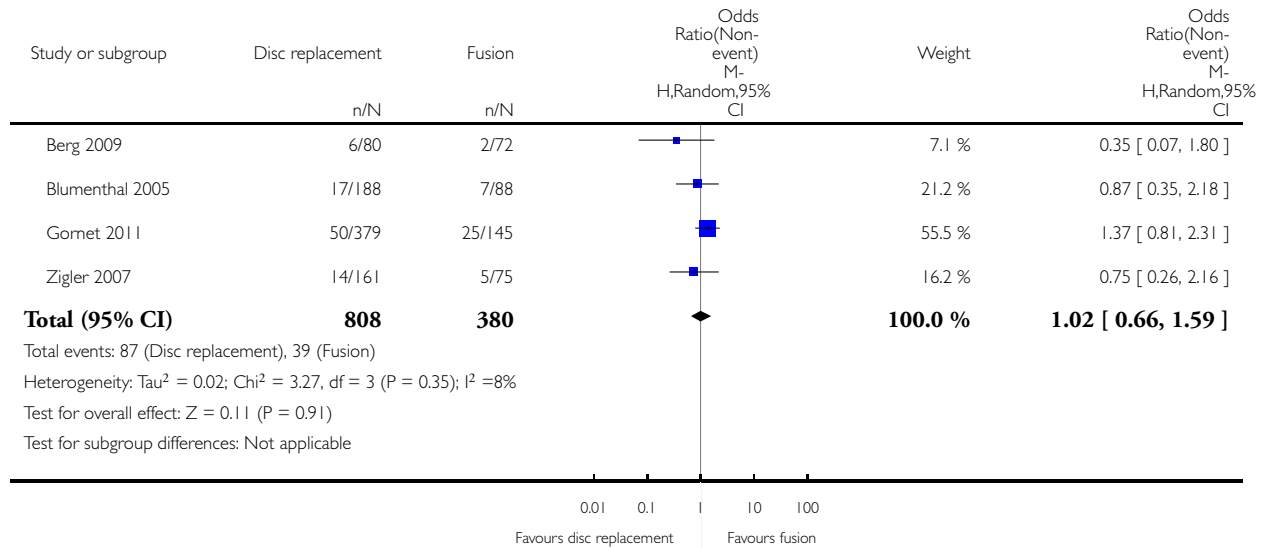
- (1) 25% improvement
- (2) 25% improvement
- (3) 30% improvement
- (4) 50% improvement
- (5) 30% improvement

**Analysis 1.15. Comparison 1 Disc replacement versus fusion, Outcome 15 Improvement in working status at 24 months.**

Review: Total disc replacement for chronic back pain in the presence of disc degeneration

Comparison: 1 Disc replacement versus fusion

Outcome: 15 Improvement in working status at 24 months



### Analysis 1.16. Comparison 1 Disc replacement versus fusion, Outcome 16 Implant motion.

Review: Total disc replacement for chronic back pain in the presence of disc degeneration

Comparison: 1 Disc replacement versus fusion

Outcome: 16 Implant motion

Study or subgroup	Disc replacement		Fusion		Mean Difference IV,Random,95% CI	Mean Difference IV,Random,95% CI
	N	Mean(SD)	N	Mean(SD)		
Berg 2009	68	9 (4.7)	61	0.7 (2.5)		8.30 [ 7.02, 9.58 ]
Gomet 2011 (1)	379	9.5 (0)	145	0.6 (0)		0.0 [ 0.0, 0.0 ]
<b>Total (95% CI)</b>	<b>447</b>		<b>206</b>			<b>8.30 [ 7.02, 9.58 ]</b>

Heterogeneity: Tau<sup>2</sup> = 0.0; Chi<sup>2</sup> = 0.0, df = 0 (P = 1.00); I<sup>2</sup> = 0.0%  
 Test for overall effect: Z = 12.70 (P < 0.00001)  
 Test for subgroup differences: Not applicable

-10 -5 0 5 10  
 Favours fusion Favours disc replacement

(1) at 24 months

### Analysis 1.17. Comparison 1 Disc replacement versus fusion, Outcome 17 Blood loss.

Review: Total disc replacement for chronic back pain in the presence of disc degeneration

Comparison: 1 Disc replacement versus fusion

Outcome: 17 Blood loss

Study or subgroup	Disc replacement		Fusion		Mean Difference IV,Random,95% CI	Weight	Mean Difference IV,Random,95% CI
	N	Mean(SD)	N	Mean(SD)			
Berg 2009 (1)	80	450 (447)	72	400 (404)		18.9 %	50.00 [ -85.29, 185.29 ]
Blumenthal 2005	205	205 (211.7)	99	208.9 (283.9)		21.6 %	-3.90 [ -66.89, 59.09 ]
Gomet 2011	405	240.7 (301.1)	172	95.2 (106.9)		22.2 %	145.50 [ 112.11, 178.89 ]
Moreno 2008 (2)	14	150 (188)	18	300 (317)		17.0 %	-150.00 [ -326.48, 26.48 ]
Zigler 2007	161	204 (231.3)	75	465 (440)		20.2 %	-261.00 [ -366.80, -155.20 ]
<b>Total (95% CI)</b>	<b>865</b>		<b>436</b>			<b>100.0 %</b>	<b>-37.22 [ -185.06, 110.62 ]</b>

Heterogeneity: Tau<sup>2</sup> = 25289.12; Chi<sup>2</sup> = 68.21, df = 4 (P < 0.00001); I<sup>2</sup> = 94%  
 Test for overall effect: Z = 0.49 (P = 0.62)  
 Test for subgroup differences: Not applicable

-200 -100 0 100 200  
 Favours disc replacement Favours fusion

(1) SD estimated from trendline found from mean and SD in the three other studies

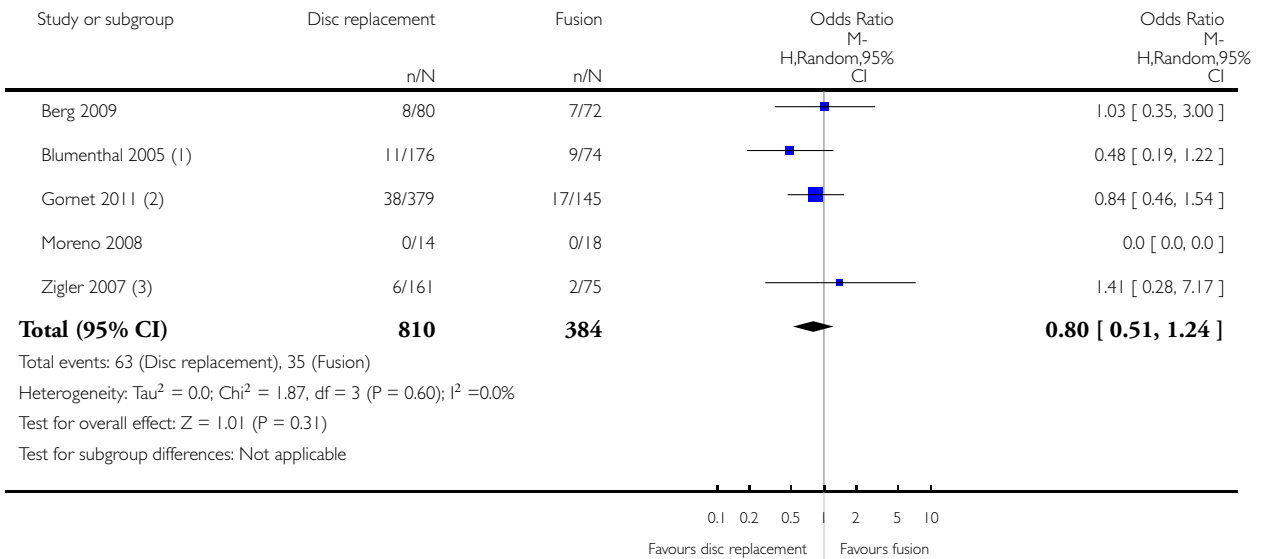
(2) SD estimated from trendline found from mean and SD in the three other studies

### Analysis 1.18. Comparison 1 Disc replacement versus fusion, Outcome 18 Reoperations.

Review: Total disc replacement for chronic back pain in the presence of disc degeneration

Comparison: 1 Disc replacement versus fusion

Outcome: 18 Reoperations



(1) Revision, reoperation, removals up to 24 months; N excluding loss to follow up and deaths

(2) Revisions (0 VS 0), removals (2 VS 0), supplementary fixations (15 vs 14) and reoperations (22 vs 3) up to 24 months

(3) Up to 24 months; excluding screw removals (2)

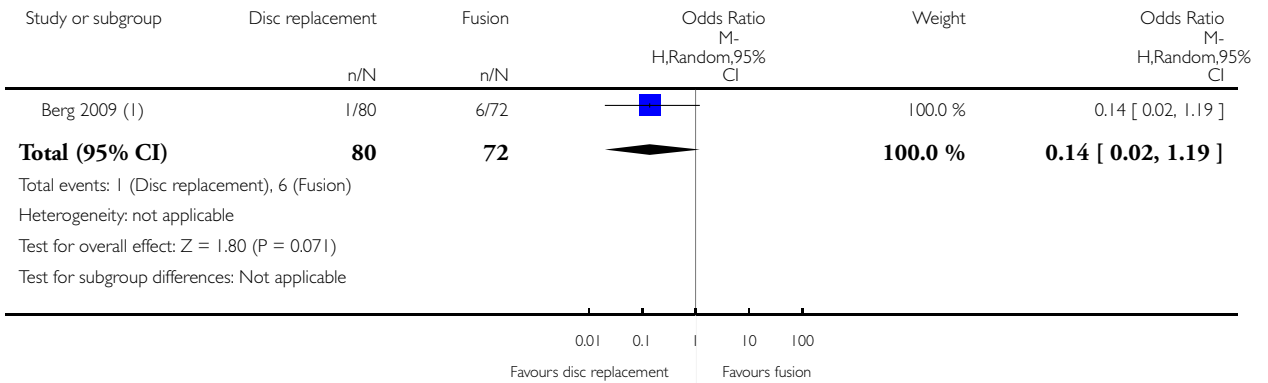


**Analysis 1.19. Comparison 1 Disc replacement versus fusion, Outcome 19 Adjacent segment degeneration.**

Review: Total disc replacement for chronic back pain in the presence of disc degeneration

Comparison: 1 Disc replacement versus fusion

Outcome: 19 Adjacent segment degeneration



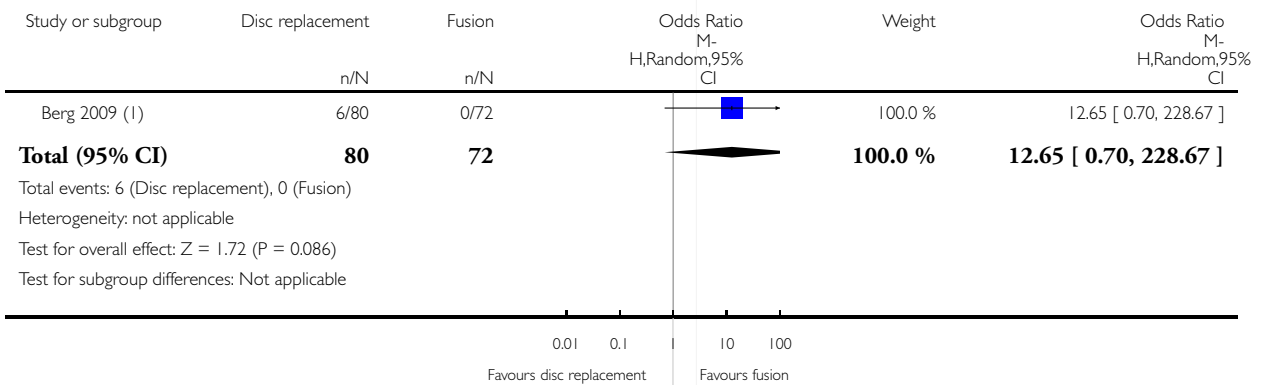
(1) up to 24 months

**Analysis 1.20. Comparison 1 Disc replacement versus fusion, Outcome 20 Facet joint degeneration.**

Review: Total disc replacement for chronic back pain in the presence of disc degeneration

Comparison: 1 Disc replacement versus fusion

Outcome: 20 Facet joint degeneration



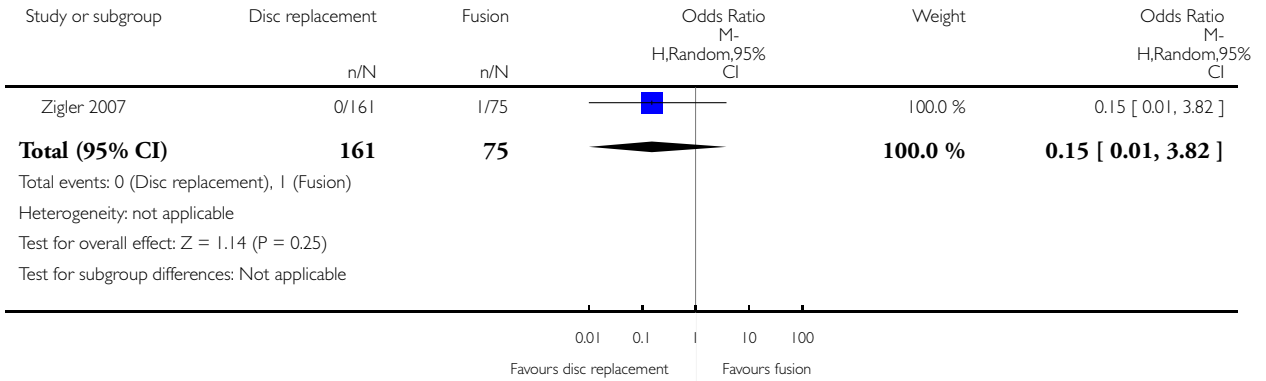
(1) up to 24 months

### Analysis 1.21. Comparison 1 Disc replacement versus fusion, Outcome 21 Radiographic loosening (radiolucency).

Review: Total disc replacement for chronic back pain in the presence of disc degeneration

Comparison: 1 Disc replacement versus fusion

Outcome: 21 Radiographic loosening (radiolucency)

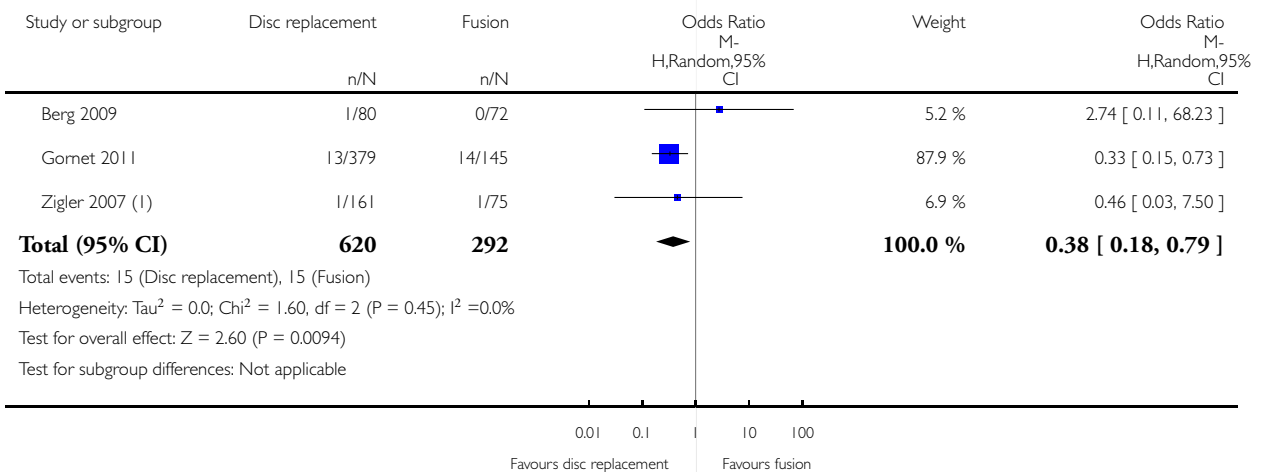


### Analysis 1.22. Comparison 1 Disc replacement versus fusion, Outcome 22 Subsidence.

Review: Total disc replacement for chronic back pain in the presence of disc degeneration

Comparison: 1 Disc replacement versus fusion

Outcome: 22 Subsidence



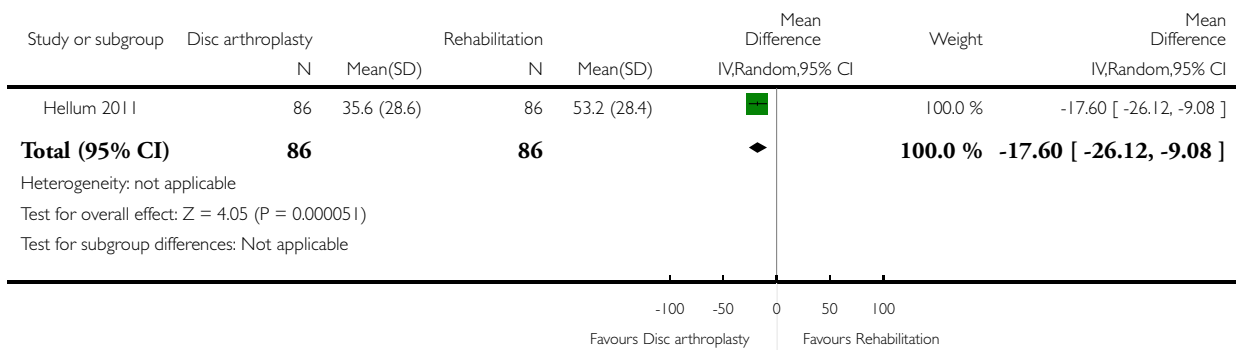
(1) At 24 Months, not clinically significant

### Analysis 2.1. Comparison 2 Disc arthroplasty versus rehabilitation, Outcome 1 Back Pain at 12 months.

Review: Total disc replacement for chronic back pain in the presence of disc degeneration

Comparison: 2 Disc arthroplasty versus rehabilitation

Outcome: 1 Back Pain at 12 months

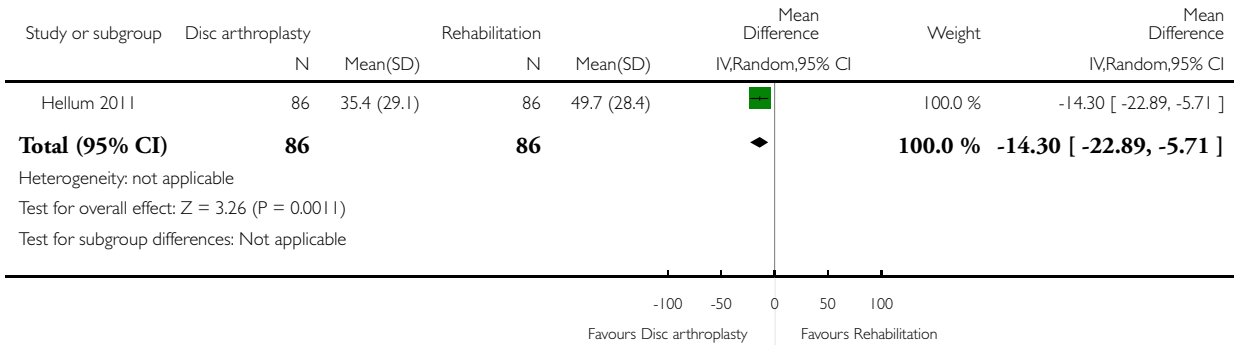


### Analysis 2.2. Comparison 2 Disc arthroplasty versus rehabilitation, Outcome 2 Back Pain at 24 months.

Review: Total disc replacement for chronic back pain in the presence of disc degeneration

Comparison: 2 Disc arthroplasty versus rehabilitation

Outcome: 2 Back Pain at 24 months

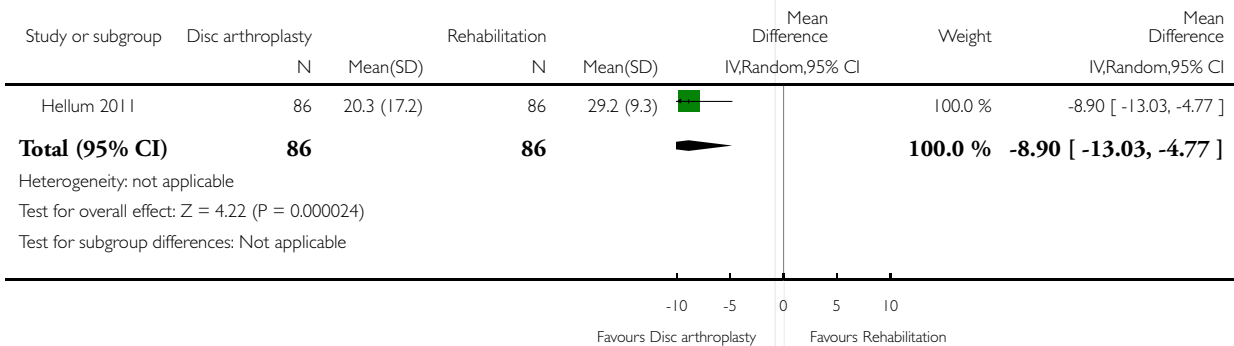


### Analysis 2.3. Comparison 2 Disc arthroplasty versus rehabilitation, Outcome 3 Oswestry at 12 months.

Review: Total disc replacement for chronic back pain in the presence of disc degeneration

Comparison: 2 Disc arthroplasty versus rehabilitation

Outcome: 3 Oswestry at 12 months

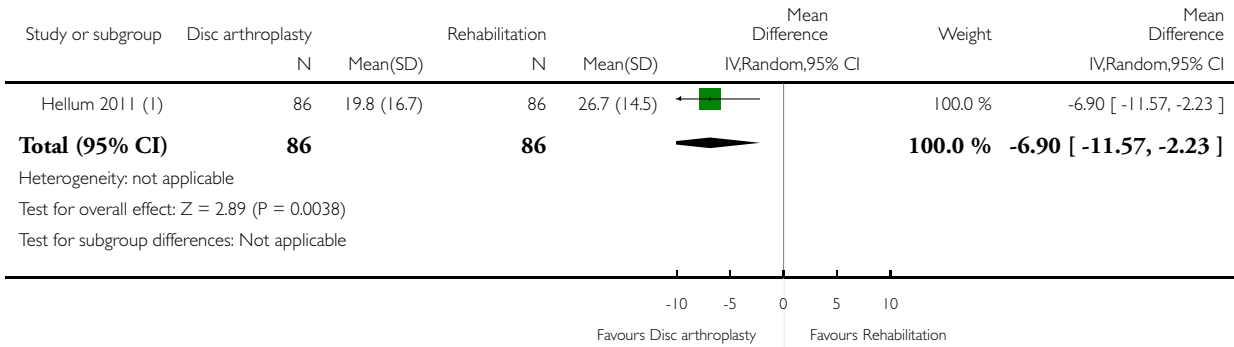


### Analysis 2.4. Comparison 2 Disc arthroplasty versus rehabilitation, Outcome 4 Oswestry at 24 months.

Review: Total disc replacement for chronic back pain in the presence of disc degeneration

Comparison: 2 Disc arthroplasty versus rehabilitation

Outcome: 4 Oswestry at 24 months



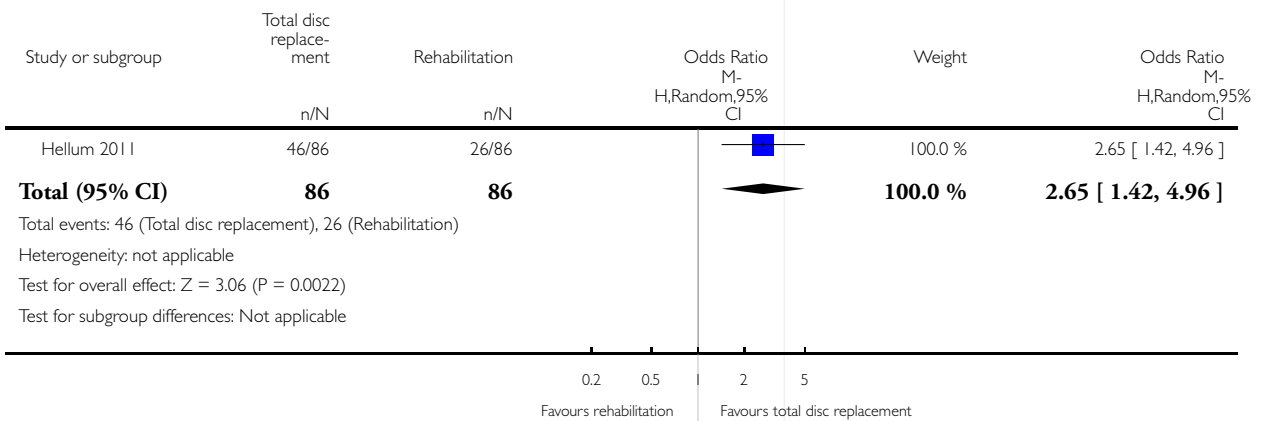
(1) ITT analysis

### Analysis 2.5. Comparison 2 Disc arthroplasty versus rehabilitation, Outcome 5 Patient satisfaction.

Review: Total disc replacement for chronic back pain in the presence of disc degeneration

Comparison: 2 Disc arthroplasty versus rehabilitation

Outcome: 5 Patient satisfaction

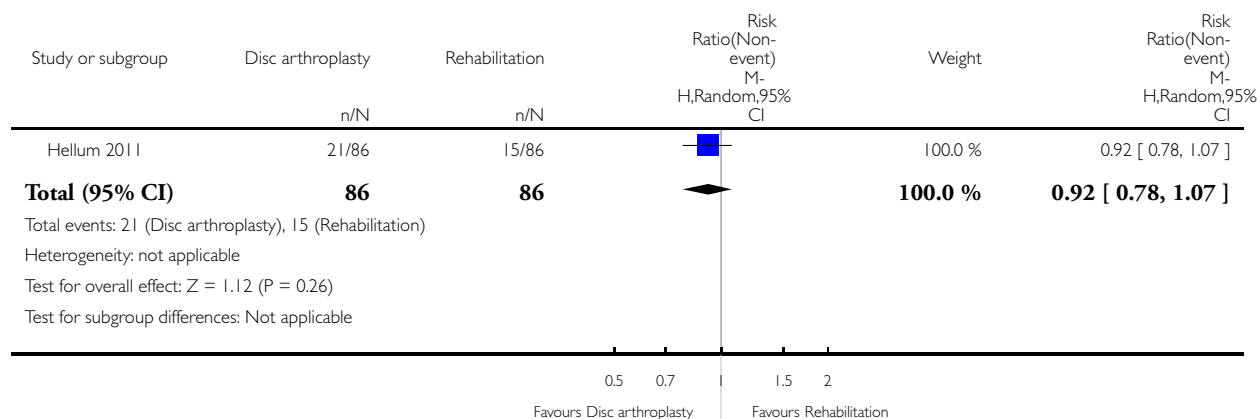


### Analysis 2.6. Comparison 2 Disc arthroplasty versus rehabilitation, Outcome 6 Improvement in working status.

Review: Total disc replacement for chronic back pain in the presence of disc degeneration

Comparison: 2 Disc arthroplasty versus rehabilitation

Outcome: 6 Improvement in working status

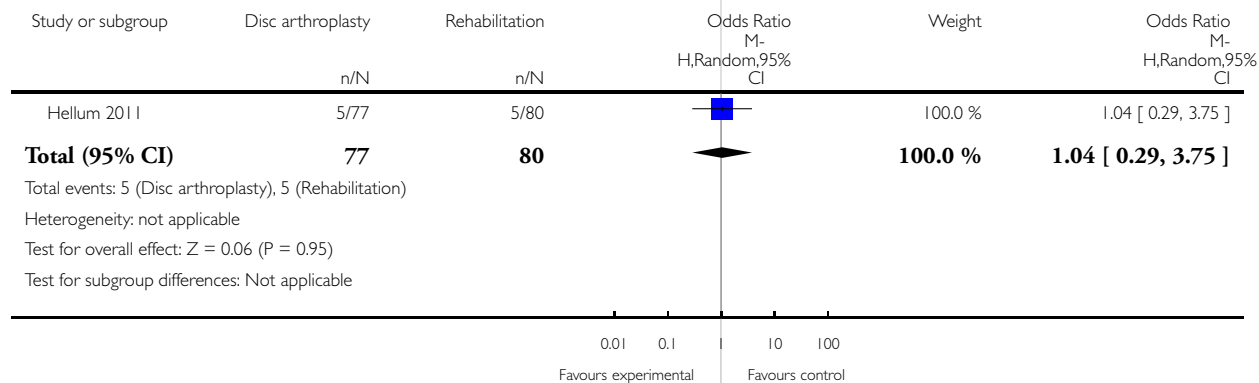


### Analysis 2.7. Comparison 2 Disc arthroplasty versus rehabilitation, Outcome 7 Reoperations.

Review: Total disc replacement for chronic back pain in the presence of disc degeneration

Comparison: 2 Disc arthroplasty versus rehabilitation

Outcome: 7 Reoperations



## ADDITIONAL TABLES

Table 1. Characteristics of (excluded) observational studies

Study_____	Characteristic____	Remark_____
Cakir 2006	Selection bias	Included spinal claudication patients specifically in the control group, but not in the experimental group. Resulting age difference large
	Population	> 12 months LBP, > 6 months conservative therapy, no spondylolisthesis, no spondylosis, no antero- or retrolisthesis, no facet joint arthritis total disc replacement: DDD, One level Control: Spinal Claudication
	Interventions	Total disc replacement: Prodisc Control: Dynesis (Zimmer, Winterthur, Switzerland; pedicle screws, spacers and rods) with decompression
	Outcomes	ODI, SF-36 @ preop and latest (15.3 and 21.8 months)
	Remarks	No difference between treatments discussed
Schroven 2006	Population	DDD or lumbar spondylosis, CT/MRI confirmed 18-60 years > 6 months of conservative therapy No baseline characteristics given
	Interventions	total disc replacement: Prodisc Control: ALIF (unknown cage with iliac crest autograft) Allocation to groups not described
	Outcomes	ODI @ preoperative and 6, 12 months
	Remarks	Prodisc better ODI improvement than ALIF, also better in length of stay, blood loss and operation time
Shim 2007	Design	Retrospective comparison
	Interventions	Two disc replacement devices.

CT: computed tomography

DDD: degenerative disc disease

LBP: low-back pain

MRI: magnetic resonance imaging

ODI: Oswestry Disability Index

**Table 2. Clinical relevance**

Item	Berg 2009	Blumenthal 2005	Gornet 2011	Hellum 2011	Moreno 2008	Sasso 2008	Zigler 2007
1. Are the patients described in detail so that you can decide whether they are comparable to those that you see in your practice?	Yes	Yes	Yes	Yes	No	No	Yes
2. Are the interventions and treatment settings described well enough so that you can provide the same for your patients?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
3. Were all clinically relevant outcomes measured and reported?	Yes	Yes	Yes	Yes	No	Yes	Yes
4. Is the size of the effect clinically important?	No	Yes	No	No	No	Unsure	No
5. Are the likely treatment benefits worth the potential harms?	No	Unsure	No	No	No	Unsure	No



## APPENDICES

### Appendix I. Search strategies

#### MEDLINE

*NB: MEDLINE strategy uses Cochrane Search Filter for RCTS plus BMJ Clinical evidence Cohort & case series filter*

1. randomized controlled trial.pt.
2. controlled clinical trial.pt.
3. randomized.ab.
4. placebo.ab,ti.
5. drug therapy.fs.
6. randomly.ab,ti.
7. trial.ab,ti.
8. groups.ab,ti.
9. or/1-8
10. (animals not (humans and animals)).sh.
11. 9 not 10
12. Comparative Study/
13. exp Evaluation Studies/
14. exp Follow-Up Studies/
15. exp Prospective Studies/
16. exp Cross-Over Studies/
17. exp Epidemiologic Studies/
18. exp Case-Control Studies/
19. exp Cohort Studies/
20. exp Cross-Sectional Studies/
21. (cohort adj (study or studies)).mp
22. cohort analy\$.mp
23. (follow up adj (study or studies)).mp.
24. (observational adj (study or studies)).mp.
25. longitudinal.mp.
26. retrospective.mp.
27. cross sectional.mp.
28. control\$.mp.
29. prospective\$.mp.
30. volunteer.mp.
31. or/12-30
32. 31 not 10
33. 32 or 11
34. exp Cohort Studies/
35. cohort\$.tw.
36. controlled clinical trial.pt.
37. Epidemiologic Methods/
38. limit 37 to yr=1966-1989
39. exp Case-Control Studies/
40. (case\$ and control\$).tw.
41. (case\$ and series).tw.
42. or/34-36,38-41
43. 42 not 10
44. 33 or 43
45. exp Spine/
46. exp Spinal Diseases/
47. exp Back Pain/

48. exp Intervertebral Disk/
49. intervertebral disc.mp.
50. exp Intervertebral Disk Displacement/
51. exp Discitis/
52. dorsalgia.ti,ab.
53. backache.ti,ab.
54. spondylosis.ti,ab.
55. lumbago.ti,ab.
56. (lumbar adj pain).ti,ab.
57. sciatic neuropathy/
58. sciatica.ti,ab.
59. (disc adj degeneration).mp.
60. (spine or spinal).mp.
61. vertebra\$.mp.
62. (disc or discs or disk or disks).mp.
63. or/45-62
64. "Prostheses and Implants"/
65. Implants, Experimental/
66. Prosthesis Implantation/
67. Arthroplasty, Replacement/
68. exp Arthroplasty/ or arthroplasty.mp.
69. or/64-68
70. 62 and 69
71. disc replacement.mp.
72. kinematic.mp.
73. active L.mp.
74. charite.mp.
75. prodisc.mp.
76. kineflex.mp.
77. maverick.mp.
78. mobidisc.mp.
79. flexicore.mp.
80. nubac.mp.
81. acroflex.mp.
82. artificial.ti,ab.
83. flexible.ti,ab.
84. mobile.ti,ab.
85. kinematic.ti,ab.
86. prestige.mp.
87. (pro adj disc).mp.
88. porous coated motion.mp.
89. or/71-88
90. 62 and 89
91. 70 or 90
92. 44 and 63 and 91

**EMBASE**

1. Clinical Article/
2. exp Clinical Study/
3. Clinical Trial/
4. Controlled Study/
5. Randomized Controlled Trial/
6. Major Clinical Study/
7. Double Blind Procedure/

8. Multicenter Study/
9. Single Blind Procedure/
10. Phase 3 Clinical Trial/
11. Phase 4 Clinical Trial/
12. crossover procedure/
13. placebo/
14. or/1-13
15. allocat\$.mp.
16. assign\$.mp.
17. blind\$.mp.
18. (clinic\$ adj25 (study or trial)).mp.
19. compar\$.mp.
20. control\$.mp.
21. cross?over.mp.
22. factorial\$.mp.
23. follow?up.mp.
24. placebo\$.mp.
25. prospectiv\$.mp.
26. random\$.mp.
27. ((singl\$ or doubl\$ or trebl\$ or tripl\$) adj25 (blind\$ or mask\$)).mp.
28. trial.mp.
29. (versus or vs).mp.
30. or/15-29
31. 14 and 30
32. human/
33. Nonhuman/
34. exp ANIMAL/
35. Animal Experiment/
36. 33 or 34 or 35
37. 32 not 36
38. 31 not 36
39. 37 and 38
40. exp SPINE/
41. exp Spine Disease/
42. exp Backache/
43. exp Intervertebral Disk/
44. intervertebral disc.mp.
45. exp Intervertebral Disk Hernia/
46. exp Diskitis/
47. exp LOW BACK PAIN/
48. (lumbar adj pain).mp.
49. exp ISCHIALGIA/
50. spondylosis.mp.
51. lumbago.mp.
52. sciatica.mp.
53. back pain.mp.
54. dorsalgia.mp.
55. discitis.mp.
56. (disc adj degeneration).mp.
57. (disk adj degeneration).mp.
58. vertebra\$.mp.
59. (disc or discs or disk or disks).mp.
60. exp Intervertebral Disk Degeneration/

- 61. or/40-60
- 62. prestige.mp.
- 63. porous coated motion.mp.
- 64. prodisc.mp.
- 65. exp PROSTHESIS/
- 66. exp IMPLANT/
- 67. exp Implantation/
- 68. exp ARTHROPLASTY/
- 69. exp Intervertebral Discectomy/
- 70. kinematic.mp. or exp KINEMATICS/
- 71. Active L.mp.
- 72. charite.mp.
- 73. kineflex.mp.
- 74. maverick.mp.
- 75. mobidisc.mp.
- 76. flexicore.mp.
- 77. nubac.mp.
- 78. acroflex.mp.
- 79. artificial.ti,ab.
- 80. flexible.ti,ab.
- 81. kinematic.ti,ab.
- 82. or/62-81
- 83. 59 and 82
- 84. 39 and 61 and 83

**BIOSIS**

*NB: strategy uses the RCT filter developed by Lisa Tjosvold Sept. 2, 2003 for the Child Health Field.*

- #24 #23 AND #16
- #23 #22 OR #21 OR #20 OR #19 OR #18 OR #17
- #22 TS=(disc OR disk) AND TS=(replacement OR kinematic OR artificial OR flexible OR mobile OR Active L OR charite OR Maverick)
- #21 Topic=(flexicore)
- #20 Topic=(Arthroplasty)
- #19 Topic=(porous coated motion)
- #18 Topic=(prodisc)
- #17 Topic=(prestige)
- #16 #15 AND #1
- #15 #14 OR #13 OR #12 OR #11 OR #10 OR #9 OR #8 OR #7 OR #6 OR #5 OR #4 OR #3 OR #2
- #14 Topic=(lumbar vertebra\*)
- #13 Topic=(disc) OR Topic=(discs) OR Topic=(disk) OR Topic=(disks)
- #12 Topic=(sciatica)
- #11 Topic=(lumbago)
- #10 Topic=(spondylosis)
- #9 Topic=(backache)
- #8 Topic=(dorsalgia)
- #7 Topic=(dis\*itis)
- #6 TS=intervertebral dis\*
- #5 Topic=(“low back pain“)
- #4 Topic=(“back pain“)
- #3 Topic=(back pain)
- #2 Topic=(spine) OR Topic=(spinal)
- #1 TS= clinical trial\* OR TS=research design OR TS=comparative stud\* OR TS=evaluation stud\* OR TS=controlled trial\* OR TS=follow-up stud\* OR TS=prospective stud\* OR TS=random\* OR TS=placebo\* OR TS=(single blind\*) OR TS=(double blind\*)

**CENTRAL**

- #1 MeSH descriptor Back Pain explode all trees
- #2 dorsalgia
- #3 backache
- #4 MeSH descriptor Low Back Pain explode all trees
- #5 (lumbar next pain) or (coccyx) or (coccydynia) or (sciatica) or (spondylosis)
- #6 MeSH descriptor Sciatica explode all trees
- #7 MeSH descriptor Spine explode all trees
- #8 MeSH descriptor Spinal Diseases explode all trees
- #9 (lumbago) or (discitis) or (disc near degeneration) or (disc near prolapse) or (disc near herniation)
- #10 spinal fusion
- #11 facet near joints
- #12 MeSH descriptor Intervertebral Disk explode all trees
- #13 postlaminectomy
- #14 arachnoiditis
- #15 failed near back
- #16 MeSH descriptor Cauda Equina explode all trees
- #17 lumbar near vertebra\*
- #18 spinal near stenosis
- #19 slipped near (disc\* or disk\*)
- #20 degenerat\* near (disc\* or disk\*)
- #21 stenosis near (spine or root or spinal)
- #22 displace\* near (disc\* or disk\*)
- #23 prolap\* near (disc\* or disk\*)
- #24 (#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20 OR #21 OR #22 OR #23)
- #25 MeSH descriptor Prostheses and Implants, this term only
- #26 MeSH descriptor Prosthesis Implantation, this term only
- #27 MeSH descriptor Implants, Experimental, this term only
- #28 MeSH descriptor Arthroplasty explode all trees
- #29 MeSH descriptor Arthroplasty, Replacement, this term only
- #30 arthroplasty
- #31 disc replacement
- #32 prestige
- #33 prodisc
- #34 porous coated motion
- #35 kinematic NEAR disc
- #36 kinematic
- #37 (#25 OR #26 OR #27 OR #28 OR #29 OR #30 OR #31 OR #32 OR #33 OR #34 OR #35 OR #36)
- #38 disc or discs or disk or disks
- #39 (#37 AND #38)
- #40 (#24 AND #39)

## Appendix 2. Criteria and operationalisation for 'Risk of bias' assessment

### Random sequence generation (selection bias)

#### Selection bias (biased allocation to interventions) due to inadequate generation of a randomised sequence

There is a low risk of selection bias if the investigators describe a random component in the sequence generation process such as: referring to a random number table, using a computer random number generator, coin tossing, shuffling cards or envelopes, throwing dice,

drawing of lots, minimisation (minimisation may be implemented without a random element, and this is considered to be equivalent to being random).

There is a high risk of selection bias if the investigators describe a non-random component in the sequence generation process, such as: sequence generated by odd or even date of birth, date (or day) of admission, hospital or clinic record number; or allocation by judgment of the clinician, preference of the participant, results of a laboratory test or a series of tests, or availability of the intervention.

## **Allocation concealment (selection bias)**

### **Selection bias (biased allocation to interventions) due to inadequate concealment of allocations prior to assignment**

There is a low risk of selection bias if the participants and investigators enrolling participants could not foresee assignment because one of the following, or an equivalent method, was used to conceal allocation: central allocation (including telephone, web-based and pharmacy-controlled randomisation); sequentially numbered drug containers of identical appearance; or sequentially numbered, opaque, sealed envelopes.

There is a high risk of bias if participants or investigators enrolling participants could possibly foresee assignments and thus introduce selection bias, such as allocation based on: using an open random allocation schedule (e.g. a list of random numbers); assignment envelopes were used without appropriate safeguards (e.g. if envelopes were unsealed or non-opaque or not sequentially numbered); alternation or rotation; date of birth; case record number; or other explicitly unconcealed procedures.

## **Blinding of participants**

### **Performance bias due to knowledge of the allocated interventions by participants during the study**

There is a low risk of performance bias if blinding of participants was ensured and it was unlikely that the blinding could have been broken; or if there was no blinding or incomplete blinding, but the review authors judge that the outcome is not likely to be influenced by lack of blinding.

## **Blinding of personnel/care providers (performance bias)**

### **Performance bias due to knowledge of the allocated interventions by personnel/care providers during the study**

There is a low risk of performance bias if blinding of personnel was ensured and it was unlikely that the blinding could have been broken; or if there was no blinding or incomplete blinding, but the review authors judge that the outcome is not likely to be influenced by lack of blinding.

## **Blinding of outcome assessor (detection bias)**

### **Detection bias due to knowledge of the allocated interventions by outcome assessors**

There is low risk of detection bias if the blinding of the outcome assessment was ensured and it was unlikely that the blinding could have been broken; or if there was no blinding or incomplete blinding, but the review authors judge that the outcome is not likely to be influenced by lack of blinding, or:

- for patient-reported outcomes in which the patient was the outcome assessor (e.g. pain, disability): there is a low risk of bias for outcome assessors if there is a low risk of bias for participant blinding (Boutron 2005); for outcome criteria that are clinical or therapeutic events that will be determined by the interaction between patients and care providers (e.g. co-interventions, length of hospitalisation, treatment failure), in which the care provider is the outcome assessor: there is a low risk of bias for outcome assessors if there is a low risk of bias for care providers (Boutron 2005); for outcome criteria that are assessed from data from medical forms: there is a low risk of bias if the treatment or adverse effects of the treatment could not be noticed in the extracted data (Boutron 2005).

## **Incomplete outcome data (attrition bias)**

### **Attrition bias due to amount, nature or handling of incomplete outcome data**

There is a low risk of attrition bias if there were no missing outcome data; reasons for missing outcome data were unlikely to be related to the true outcome (for survival data, censoring unlikely to be introducing bias); missing outcome data were balanced in numbers, with similar reasons for missing data across groups; for dichotomous outcome data, the proportion of missing outcomes compared with the observed event risk was not enough to have a clinically relevant impact on the intervention effect estimate; for continuous outcome data, the plausible effect size (difference in means or standardised difference in means) among missing outcomes was not enough to have a clinically relevant impact on observed effect size, or missing data were imputed using appropriate methods (if drop-outs are very large, imputation using even 'acceptable' methods may still suggest a high risk of bias) (Van Tulder 2003). The percentage of withdrawals and drop-outs should not exceed 20% for short-term follow-up and 30% for long-term follow-up and should not lead to substantial bias (these percentages are commonly used but arbitrary, not supported by literature) (Van Tulder 2003).

## **Selective reporting (reporting bias)**

### **Reporting bias due to selective outcome reporting**

There is low risk of reporting bias if the study protocol is available and all of the study's pre-specified (primary and secondary) outcomes that are of interest in the review have been reported in the pre-specified way, or if the study protocol is not available but it is clear that the published reports include all expected outcomes, including those that were pre-specified (convincing text of this nature may be uncommon).

There is a high risk of reporting bias if not all of the study's pre-specified primary outcomes have been reported; one or more primary outcomes is reported using measurements, analysis methods or subsets of the data (e.g. subscales) that were not pre-specified; one or more reported primary outcomes were not pre-specified (unless clear justification for their reporting is provided, such as an unexpected adverse effect); one or more outcomes of interest in the review are reported incompletely so that they cannot be entered in a meta-analysis; the study report fails to include results for a key outcome that would be expected to have been reported for such a study.

## **Group similarity at baseline (selection bias)**

Bias due to dissimilarity at baseline for the most important prognostic indicators.

There is low risk of bias if groups are similar at baseline for demographic factors, value of main outcome measure(s), and important prognostic factors (examples in the field of back and neck pain are duration and severity of complaints, vocational status, percentage of patients with neurological symptoms) (Van Tulder 2003).

## **Co-interventions (performance bias)**

### **Bias because co-interventions were different across groups**

There is low risk of bias if there were no co-interventions or they were similar between the index and control groups (Van Tulder 2003).

## **Compliance (performance bias)**

### **Bias due to inappropriate compliance with interventions across groups**

There is low risk of bias if compliance with the interventions was acceptable, based on the reported intensity/dosage, duration, number and frequency for both the index and control intervention(s). For single-session interventions (e.g. surgery), this item is irrelevant (Van Tulder 2003).

### **Intention-to-treat-analysis**

There is low risk of bias if all randomized patients were reported/analysed in the group to which they were allocated by randomisation.

### **Timing of outcome assessments (detection bias)**

#### **Bias because important outcomes were not measured at the same time across groups**

There is low risk of bias if all important outcome assessments for all intervention groups were measured at the same time ([Van Tulder 2003](#)).

### **Other bias**

#### **Bias due to problems not covered elsewhere in the table**

There is a low risk of bias if the study appears to be free of other sources of bias not addressed elsewhere (e.g. study funding).

## **Appendix 3. Assessment of Clinical Relevance**

1. Are the patients described in detail so that you can decide whether they are comparable to those that you see in your practice?
2. Are the interventions and treatment settings described well enough so that you can provide the same for your patients?
3. Were all clinically relevant outcomes measured and reported?
4. Is the size of the effect clinically important?
5. Are the likely treatment benefits worth the potential harms?

## **WHAT'S NEW**

Last assessed as up-to-date: 6 March 2012.

Date	Event	Description
18 December 2012	Amended	Amended information for sponsoring of one study (Berg 2009).

## **CONTRIBUTIONS OF AUTHORS**

Wilco Jacobs: Protocol development and preparation, Selection, Risk of bias assessment, RevMan data entry, Manuscript preparation, Overall Coordination

Niels van der Gaag: Risk of bias assessment, Draft review

Alexander Tuschel: Protocol review, Selection

Marinus de Kleuver: Protocol development, Clinical interpretation, Draft review

Cumhur Oner: Clinical interpretation

Wilco Peul: Clinical interpretation, Draft review

Ab Verbout: Clinical interpretation



## DECLARATIONS OF INTEREST

None declared.

## SOURCES OF SUPPORT

### Internal sources

- Sint Maartenskliniek, Netherlands.

Provided allocated time to develop the protocol for the review

- Leiden University Medical Center, Netherlands.

Provided allocated time to perform the review

### External sources

- No sources of support supplied

## DIFFERENCES BETWEEN PROTOCOL AND REVIEW

It was anticipated during the conception of the protocol that there were fewer than five randomised studies available. This assumption was based on a recent non-Cochrane review ([Van den Eerenbeemt 2010](#)) that included only three studies. According to the BRG guidelines, with fewer than five studies, non-randomised studies can be included in Cochrane reviews. Our initial methodology was thus aimed to include non-randomised studies. However, the search resulted in seven RCT's and only 2 non-randomised, prospective concurrent controlled studies. The risk of bias in the two small observational studies ([Cakir 2006](#); [Schroven 2006](#)) was high. [Cakir 2006](#) et al specifically defined different selection criteria for the two interventions and this consequently resulted in two different groups with associated distinct age categories. The risk of bias in the observational studies was considered too high to be used alongside the large randomised controlled trials, or even in addition to these trials. It was therefore decided to omit the poor quality studies from the analysis. A sensitivity analysis had already been planned between the randomised and non-randomised studies, and this analysis was now used to assess the effect of omitting these studies. For readability, we have omitted the methodology and results associated with the aim to include observational studies.

Handling of grey literature and conference proceedings was not described in the protocol. These were excluded and the criterion is added to the methods section. We decided to exclude conference proceedings because reliable data extraction and risk of bias assessment from an abstracts is not feasible.

The secondary goals as mentioned in the protocol stem from our previous review. That review deliberately included observational data on total disc replacement alone for support of some of the secondary questions. Because of the comparative effectiveness and clinical outcome requirements of this Cochrane version of the review we were unable to validly provide an answer to one secondary question "What could be considered the indication for total disc replacement of a vertebral motion segment?". This question requires extensive epidemiological and prognostic study designs registering extensive patient characteristics.

## **INDEX TERMS**

### **Medical Subject Headings (MeSH)**

Chronic Pain [etiology; \*surgery]; Intervertebral Disc Degeneration [complications; \*surgery]; Low Back Pain [etiology; \*surgery]; Lumbosacral Region; Randomized Controlled Trials as Topic; Spinal Fusion [methods]; Total Disc Replacement [adverse effects; \*methods]

### **MeSH check words**

Humans