

# A meta-analysis of artificial total disc replacement versus fusion for lumbar degenerative disc disease

Wu Yajun · Zhu Yue · Han Xiuxin ·  
Cui Cui

Received: 27 October 2009/Revised: 14 March 2010/Accepted: 21 March 2010/Published online: 4 April 2010  
© Springer-Verlag 2010

**Abstract** Lumbar fusion has been developed for several decades and became the standard surgical treatment for symptomatic lumbar degenerative disc disease (DDD). Artificial total disc replacement (TDR), as an alternative for spinal arthrodesis, is becoming more commonly employed treating lumbar DDD. It is still uncertain whether TDR is more effective and safer than lumbar fusion. To systematically compare the effectiveness and safety of TDR to that of the fusion for the treatment of lumbar DDD, we performed a meta-analysis. Cochrane review methods were used to analyze all relevant randomized controlled trials published up to July 2009. Five relevant randomized controlled trials involving 837 patients were identified. Patients in TDR group have slightly better functioning and less back or leg pain without clinical significance, and significantly higher satisfaction status in TDR group compared with lumbar fusion group at the 2-year follow-up. But these outcomes are highly influenced by the study with BAK cage interbody fusion, the function/pain and patient satisfaction status are no longer significantly different between two groups after excluding this study. At 5 years, these outcomes are not significantly different between comparing groups. The complication and reoperation rate of two groups are similar both at 2 and at 5 years. In conclusion, TDR does not show significant superiority for the treatment of lumbar DDD compared with fusion. The benefits of motion preservation and the long-term complications are still unable to be concluded. More high-quality RCTs with long-term follow-up are needed.

**Keywords** Meta-analysis · Randomized controlled trial · Spinal fusion · Total disc replacement · Lumbar degenerative disc disease

## Introduction

Degenerative disc disease (DDD) is a major cause for chronic low-back pain with lumbar segmental instability in which surgical intervention is required when failed of conservative treatment. Spinal fusion for DDD is the most common accepted treatment in effort to eliminate abnormal motion and instability at the symptomatic degenerated levels, and thereby reduce or eliminate low-back pain [1, 4, 7, 17]. Artificial total disc replacement (TDR), as an alternative to spinal arthrodesis, is increasingly applied for surgically treating lumbar DDD [30, 32, 42]. By performing lumbar TDR, it is postulated that the patient's normal intervertebral segment motion is restored and maintained while the adjacent level is prevented from non-physiologic loading and thus the pain is relieved [43]. Previous studies that compared the clinical effects of TDR to fusion for treating lumbar DDD are constituted of ambiguous results [10, 21, 35, 47]. Therefore, it is still uncertain whether TDR is more effective and safer than fusion. The objective of this study is to systematically compare the effectiveness and safety of TDR to fusion for the treatment of lumbar DDD.

## Materials and methods

### Criteria for selected trials

All randomized controlled clinical trials (RCTs) comparing the TDR to fusion for the treatment of lumbar DDD were

W. Yajun · Z. Yue (✉) · H. Xiuxin · C. Cui  
Department of Orthopaedics, First Affiliated Hospital,  
China Medical University, 155 Nan Jing Bei Street,  
Shenyang 110001, Liaoning, China  
e-mail: zhuyuedr@163.com

identified in this study. Patients older than 18 years of age with lumbar systematic DDD were included in the study. The interventions in this study included various types of TDR and fusion in the lumbar spine. The outcomes obtained in this study were labeled as the primary outcome and the secondary outcome. The primary outcome includes: (1) the improvement in pain measured by a validate pain scale, (2) the improvement of movement and functioning measured by a disability scale, (3) patient satisfaction with the treatment, and (4) the complications. The secondary outcome includes: (1) clinical success rate, (2) operative level range of motion (ROM) measured on the flexion/extension films, (3) the operation time and the blood loss, (4) employment rate, and (5) the reoperation rate.

#### Search methods for identification of studies

Updating to July 2009, the relevant RCTs in all languages were identified through computer and other research methods. The sources of computer searching include PubMed, The Cochrane Central Register of Controlled Trials, Ovid MEDLINE, and EMBASE. Other searching methods include hand searching of Spine, European Spine Journal and Journal of Bone and Joint Surgery abstracts from 1990, and communications with international experts. Key words that have been used for researching are degenerative disc disease, low-back pain, lumbar fusion, artificial total disc replacement, and randomized controlled trial.

#### Data collection and analysis

##### *Selection of studies*

Titles and abstracts identified from the database are checked by two reviewers (Wu Y and Han X). The full text of all the possibly relevant studies are assessed independently by the two reviewers. The reviewers decide which trials fit the inclusion criteria. Any disagreements were resolved by discussion between two reviewers, when necessary, further discussion with another independent expert.

##### *Data extraction and management*

The data were extracted from included reports independently by two reviewers (Wu Y and Han X), and further discussions would be needed to deal with the disagreements. The data extracted include the following categories: the participant characteristics, the number of participants, and the loss to follow-up; study characteristics; the intervention details; the primary and the secondary outcomes; odds ratios (OR) or mean difference (MD), and 95% confidence limits (95% CI) of the comparisons.

##### *Assessment of risk of bias*

The risk of bias was assessed by two reviewers (Wu Y and Han X) independently. In a subsequent meeting, the reviewers tried to reach consensus on each criterion that they initially disagreed on. The criteria list from Koes et al. [28] is used for assessment. Different weights are attached to each criterion, respectively. The maximum score is 100 points. The higher score indicates lower risk of bias. The study which scored more than 50 points would be considered as the good study according to the criteria [28].

##### *Assessment of clinical relevance*

Two reviewers independently assessed the clinical relevance of the included studies referring to the five questions recommended by the Cochrane Back Review Group [46]. Positive (+) would be recorded for the question if the clinical relevance item is met, negative (−) for the irrelevance, and unclear (?) if the data is inadequate for answering the question. A 20% of improvement in the pain score [37] and a 25% of improvement in the functioning score [29] are considered to be clinically important.

##### *Assessment of heterogeneity*

The clinical homogeneity was evaluated by contrasting the RCTs while taking into account the following considerations: the participant characteristics including age, sex, clinical manifestation, pain, and function status baseline; surgical technique including the type of artificial lumbar disc and fusion method; follow-up period, and measurement method. The chi-squared test was performed to identify the heterogeneity of clinical outcomes, which can be interpreted as the proportion of observed variation in measured outcomes caused by heterogeneity besides rather than random variation [25, 26]. The statistical pooling would not be performed for measured outcomes when heterogeneity is statistically significant ( $P < 0.05$ ).

##### *Measures of treatment effect*

Attempts were made to statistically pool the data of homogeneous studies in order to obtain the primary and the secondary outcomes. The results were expressed in terms of odds ratio (OR) and a 95% confidence interval (95% CI) for dichotomous outcomes, and in terms of mean difference (MD) and 95% CI for continuous outcomes. When the same continuous outcomes are measured in different scales, standardized mean difference (SMD) and 95% CI are calculated. If in some studies outcomes are shown as dichotomous data while in the other studies expressed as continuous data, odds ratios would be re-pressed as standardized mean

difference to allow dichotomous and continuous data to be pooled together [27]. We performed a sensitivity analysis for the measured effects omitting the study which may largely influence the clinical results. Collected data were checked and entered into the computer by the two reviewers (Wu Y and Han X). RevMan software (version 5.0) was used for data analysis. A random-effects model was used in this meta-analysis [11]. A rating system with five levels of evidence taken from the Cochrane Back Review Group was used to evaluate the level of evidence [46].

## Results

### Description of studies

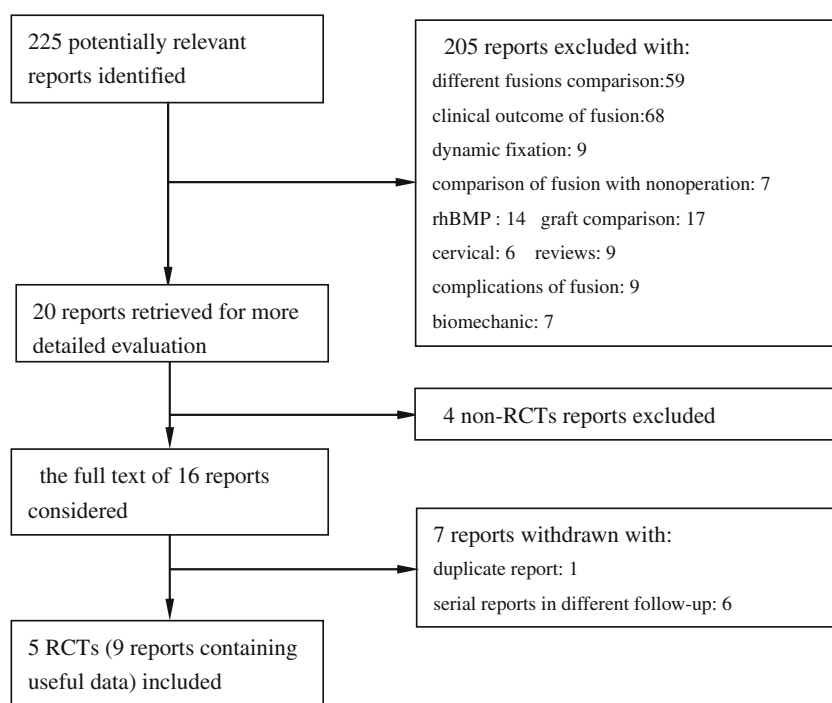
The process of identifying relevant studies is summarized in Fig. 1. From the selected databases, 696 references were obtained. By screening the titles and abstracts, 471 references were excluded due to the irrelevance to this topic. In 225 potentially relevant references, 209 references were omitted according to conditions listed in Fig. 1. The remaining 16 reports were taken for a comprehensive evaluation. These reports were based upon five independent continuous clinical randomized control trials, reporting for different follow-up periods or containing separated results. Nine reports from the five RCTs with the relevant information are eventually included involving 837 patients [2, 3, 9, 18, 20, 34, 36, 41, 48].

The characteristics of five included studies are summarized in Table 1. In all included studies, the adult patients with symptomatic lumbar DDD were recruited with sample size ranging from 67 to 304 patients. All the included studies have definite inclusion/exclusion criteria. Blumenthal et al. [3] performed the CHARITÉ artificial disc (DePuy Spine, Raynham, MA) replacement compared with anterior lumbar interbody fusion (ALIF) with BAK cages. In three studies, the ProDisc-L (Synthes Spine, West Chester, PA) [9, 48] and FlexiCore (Stryker Spine, Allendale, NJ) [41] artificial disc were employed compared with circumferential fusion. Berg et al. [2] performed artificial disc replacement with one of three following devices: CHARITÉ, ProDisc-L, or Maverick (Medtronic, Memphis, TE, USA) compared with the posterolateral fusion (PLF) with autologous bone graft or posterior interbody fusion (PLIF) with two carbon fiber cages. The clinical outcomes, surgical data, and complications were analyzed in 2- or 5-year follow-up period.

### Risk of bias in included studies

The assessment of bias risk of the included studies is presented in Table 2. Ten items (10/90, 11.1%) with inconsistent scores were further discussed to reach the consensus. Four studies with 770 patients participated have scored more than 50 points for the 18 questions [2, 3, 9, 48]. One study involving 67 patients has only scored 44 [41]. The most prevalent methodologic shortcomings are the

**Fig. 1** Study selection process. The flow-chart showed the selection of randomized controlled trials for meta-analysis



**Table 1** Characteristics of the included studies

Study	Blumenthal et al. [3]	Delamarter et al. [9]	Zigler et al. [48]	Sasso et al. [41]	Berg et al. [2]
Methods	R: fixed blocking method C: sealed envelope B: blind patients until after surgery L: Exp 44/205, Ctl 33/99 (at 2 years), total 100/233 (at 5 years)	R: fixed blocking method C: unclear B: blind patients until after surgery L: Exp 0/56, Ctl 0/22 (at 2 years)	R: fixed blocking method C: no use B: blind patients until after surgery L: Exp 2/161, Ctl 2/75 (at 2 years)	R: unclear C: unclear B: no use L: Exp 34/45, Ctl 15/22 (at 2 years)	R: randomization with unclear method C: sealed envelope B: no use L: Exp 0/80, Ctl 0/72 (at 2 years)
Participants	304 Patients; 157 males, 147 females; mean age 39.6 (19–60); single level DDD at L4–L5 or L5–S1; Oswestry score $\geq 30$ , VAS score $\geq 40$ (of 100); Back and/or leg pain with no nerve root compression; Failed $\geq 6$ months of appropriate nonoperative care; good compliance	78 Patients; 43 males, 35 females; mean age 41 (19–59); one or two level DDD between L3–S1; Oswestry score $>20/50$ ( $>40\%$ ); Back and/or leg pain; Failed $\geq 6$ months of conservative treatment; good compliance	236 Patients; 116 males, 120 females; mean age 39.2; Single level DDD at L3–S1; Oswestry score $\geq 40$ (20/50); Back and/or leg (radicular) pain; Failed $\geq 6$ months of conservative treatment; good compliance	67 Patients; 33 males, 34 females; mean age 37.7; single level DDD at L4–L5 or L5–S1; Oswestry score $\geq 40$ , VAS score $\geq 40$ (of 100); Back and/or leg pain; Failed $\geq 6$ months of conservative treatment; good compliance	152 Patients; 62 males, 90 females; mean age 40; one or two level DDD between L3 and S1; Oswestry score $\geq 30$ , VAS score $\geq 50$ (of 100); Back with/without leg pain $>1$ year; Failed $>3$ months of conservative treatment; good compliance, open-minded to treatment options
Interventions	Exp: TDR with the CHARITÉ artificial disc Ctl: ALIF with BAK cages packed with iliac crest autograft	Exp: TDR with ProDisc-L artificial disc Ctl: ALIF with femoral ring allograft plus instrumented PLF with autogenous iliac crest bone graft	Exp: TDR with ProDisc-L artificial disc Ctl: ALIF with femoral ring allograft plus instrumented PLF with autogenous iliac crest bone graft	Exp: TDR with Flexi-Core artificial disc Ctl: ALIF with femoral ring allograft plus instrumented PLF with autogenous iliac crest bone graft	Exp: TDR with CHARITÉ, or ProDisc-L, or Maverick artificial disc Ctl: instrumented PLF with autologous bone graft, or instrumented PLIF with two carbon fiber cages and bone graft
Outcomes	Surgical data; clinical outcomes: ODI, VAS score, patient satisfaction; clinical success rate; work status; radiographic outcome: ROM, disc space height; complications; reoperation (in 2-year follow-up)	Surgical data; clinical outcomes: ODI, VAS score, patient satisfaction; radiographic outcome: ROM (in 1–1.5-year follow-up)	Surgical data; clinical outcomes: ODI, VAS, SF-36 score, patient satisfaction; device, neurologic, and overall success; work status; radiographic outcome: ROM; complications; reoperation (in 2-year follow-up)	Surgical data; clinical outcomes: ODI, VAS score; radiographic outcome: ROM; complications; reoperation (in 2-year follow-up)	Surgical data; clinical outcomes: ODI, VAS, EQ5D, SF-36 score; patient satisfaction; work status; complications; reoperation (in 2-year follow-up)

R randomization, C concealment of allocation, B blinding, L losses to follow-up, Exp experimental group, Ctl control group, DDD degenerative disc disease, TDR total disc replacement, ALIF anterior lumbar interbody fusion, PLF posterolateral fusion, PLIF posterior lumbar interbody fusion

limited size of the study population, and a lack of reference concerning the blinding of outcome assessor. A fixed blocking method of randomization with six assignments per block was described in three studies [3, 9, 48]. Sealed envelop technique for allocation concealment was applied

in two studies [3, 48]. In three studies, the participants remained blinded until the operation was finished [3, 9, 48]. All of the participants in the five studies had performed the follow-up for at least 2 years and a follow-up rate of more than 89% was obtained in four of these studies [2, 3, 9, 48].

**Table 2** Risk of bias assessment of included studies

Criterion	Weight	Blumenthal et al. [3]	Delamarter et al. [9]	Zigler et al. [48]	Sasso et al. [41]	Berg et al. [2]
A. Homogeneity	2	2	2	2	2	2
B. Comparability of relevant baseline characteristics	5	3	3	3	3	3
C. Randomization procedure adequate	4	4	2	4		2
D. Drop-outs described for each study group separately	3	3	3	3	3	3
E. <20% loss to follow-up	2	2	2	2		2
<10% loss to follow-up	2	2		2		2
F. >50 subjects in the smallest group	8	8		8		8
>100 subjects in the smallest group	9					
G. Interventions included in protocol and described	10	10	10	10	10	10
H. Pragmatic study	5	5	5	5	5	5
I. Co-interventions avoided	5	5	5	5	5	5
J. Placebo-controlled	5					
K. Patients blinded	5	5	5	5		
L. Outcome measures relevant	10	10	10	10	6	8
M. Blinded outcome assessments	10					
N. Follow-up period adequate	5	5	5	5	5	5
O. Intention-to-treat analysis	5					
P. Frequencies of most important outcomes presented for each treatment group	5	5	5	5	5	5
Total score	100	67	59	69	44	60

Nevertheless, in one study only 18 out of the 67 patients have completed the 2-year follow-up [41]. The extremely low follow-up rate is the reason why outcomes of Sasso's study have not been statistically pooled into meta-analysis. None of the included studies encompassed the information of intention-to-treat (ITT) analysis.

#### Clinical relevance

The results of clinical relevance are presented in Table 3. There was one disagreement on 20 items (5%) between the two reviewers (Wu Y and Han X). Consensus was then reached on all scorings after discussion. The patient details and intervention procedures were explicitly recoded in all included studies in effort to allow researchers to replicate the treatment in clinical practice. In one study trial, the relevant outcomes, such as complications were not reported [9]. An improvement more than 20% in pain scores and improvement more than 25% in functioning were accomplished in all five studies. The consistent outcomes of all included studies suggested that the treatment benefits were likely worth the potential harms.

#### Heterogeneity

According to the data of Table 1, the participant groups from five studies had similar demographic characteristics,

and comparable pain and functioning status baseline. In the above studies, four different artificial discs (CHARITÉ, ProDisc-L, Maverick, and FlexiCore) were employed. Circumferential fusion was performed in three studies [9, 41, 48], ALIF in Blumenthal's study [3, 29], and instrumented PLF or PLIF in Berg's study [2]. Therefore, the surgical data were not pooled together. In the included studies, most outcomes were measured by the same method and at similar follow-up points. Pooling analysis of clinical success rate and range of motion was not performed since the definitions were inconsistent among studies. In random-effects meta-analysis, heterogeneity was observed in patient satisfaction ( $I^2 = 37%$ ,  $P = 0.20$ ), proportion of patients choosing the same treatment again ( $I^2 = 64%$ ,  $P = 0.06$ ), proportion of patients returning to full-time/part-time work ( $I^2 = 29%$ ,  $P = 0.24$ ), and the duration of hospitalization ( $I^2 = 80%$ ,  $P = 0.002$ ). The outcomes regarding patient functioning, painfulness, complication, and reoperation rate were consistent ( $I^2 = 0%$ ).

#### Meta-analysis results

At 2-year follow-up, the patient functioning ability measured by ODI in the TDR group was better than that of the fusion group (MD  $-4.06$ ; 95% CI  $[-7.28, -0.84]$ ;  $P = 0.01$ ) with statistical significance, but the mean difference of 4 Oswestry points was not clinically relevant.

**Table 3** Clinical relevance

	Blumenthal et al. [3]	Delamarter et al. [9]	Zigler et al. [48]	Sasso et al. [41]	Berg et al. [2]
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?	+	+	+	+	+

The VAS score of painfulness for TDR group was less than that of the fusion group (MD  $-4.75$ ; 95% CI  $[-9.14, -0.35]$ ;  $P = 0.03$ ), but the mean difference of 5 points was also not clinically significant. Patient satisfaction status was significantly better in TDR group than fusion group (SMD 0.29; 95% CI  $[0.05, 0.53]$ ;  $P = 0.02$ ). A greater proportion of patients in TDR group was willing to choose the same operation again (OR 2.86; 95% CI  $[1.41, 5.77]$ ;  $P = 0.003$ ), with 77.3% in TDR group and 58.2% in fusion group. There was no significant difference in complications (OR 0.80; 95% CI  $[0.50, 1.28]$ ;  $P = 0.36$ ), the proportion of patients who returned to full-time/part-time work (OR 1.21; 95% CI  $[0.76, 1.91]$ ;  $P = 0.43$ ), and reoperation rate (OR 0.73; 95% CI  $[0.40, 1.33]$ ;  $P = 0.31$ ) between TDR and the control group. In sensitivity analysis, the Blumenthal's study [3] was found to be highly influential. After excluding this study, there was no longer significant difference in the ODI score (MD  $-3.92$ ; 95% CI  $[-7.92, 0.08]$ ;  $P = 0.05$ ), VAS score (MD  $-4.19$ ; 95% CI  $[-9.72, 1.33]$ ;  $P = 0.14$ ), patient satisfaction (SMD 0.16; 95% CI  $[-0.08, 0.40]$ ;  $P = 0.19$ ), and proportion of patients who would choose the same treatment again (OR 4.13; 95% CI  $[0.77, 22.15]$ ;  $P = 0.10$ ). The outcomes of the meta-analysis are enumerated by Figs. 2 and 3.

### Qualitative results

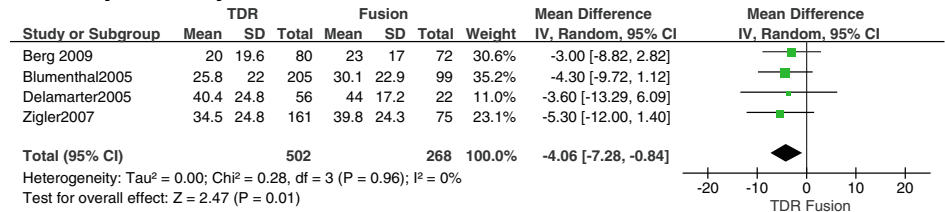
In all included studies, patients in TDR and fusion groups have demonstrated significant improvement in ODI and VAS scores compared with preoperative scores at all follow-up time points. Blumenthal et al. [3] reported that the overall clinical success was achieved in 57.1% of patients in the TDR group and 46.5% in the control group ( $P < 0.0001$ ) with the fulfillment of the following four criteria:  $\geq 25\%$  improvement in ODI score at 24 months compared with the preoperative score, no device failure, no major complications, and no neurological deterioration. Zigler et al. [48] reported the overall success rate of 53.4% in TDR group and 40.8% in control group using FDA criteria ( $P = 0.0438$ ). McAfee et al. [34] reported an operative level ROM of

113.6% of preoperative status in TDR group at 24 months ( $7.4 \pm 5.28$ ), and the mean ROM in fusion group has dropped to  $1.1 \pm 0.87$  (with 91.9% patient gaining successful lumbar fusion). In Delamarter's study [23], the result showed significantly better motility at L4–L5 for disc replacement patients ( $10.5^\circ$  at 12 months) comparing with the fusion patients ( $P < 0.05$ ), but at L5–S1 the differences between two groups was not statistically significant at 6-month point. Zigler et al. [48] reported that operative level ROM averaged  $7.7^\circ$  was maintained within a normal functional range in 93.7% of the lumbar disc replacement patients. The flexion–extension ROM of the FlexiCore recipients gained  $3.8^\circ$  at 6-week follow-up after surgery [41]. For operative time (OT) and blood loss (BL) there were significant differences between TDR group and circumferential fusion group in Zigler's report [48] (mean OT 121 vs. 229 min; mean BL 204 vs. 465 ml) and Sasso's report [41] (mean OT 82 vs. 179 min; mean BL 97 vs. 179 ml), but no significant differences between TDR group and ALIF group (or PLF/PLIF) in Blumenthal's [3] (mean OT 110.8 vs. 114 min; mean BL 212.1 vs. 204.3 ml) and Berg's report [2] (mean BL 560 vs. 444 ml). Significant differences were observed in patients in duration of hospitalization in TDR and fusion group from the studies reported by Blumenthal et al. [3] (mean 3.7 vs. 4.2 days,  $P = 0.0039$ ), Zigler et al. [48] (mean 3.5 vs. 4.4 days,  $P = 0.0001$ ), Sasso et al. [41] (mean 2 vs. 3 days,  $P < 0.005$ ), and Berg et al. [2] (mean 4.4 vs. 5.9,  $P < 0.00001$ ).

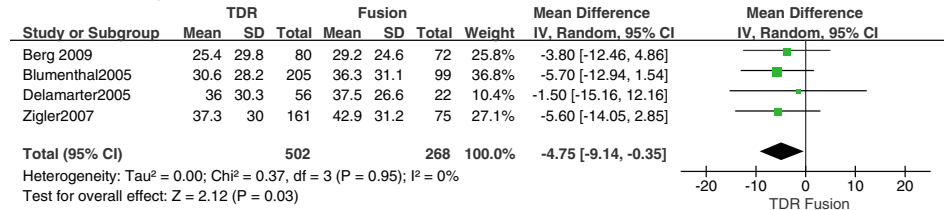
There is strong evidence (5 trials, 837 patients) that patients in both groups demonstrated significant improvement as measured by ODI and VAS scores compared with preoperative scores at all follow-up time points. There is conflicting evidence (4 trials [9, 34, 41, 48], 685 patients) that the operative level ROM of disc replacement patients maintained within a normal functional range different from that of fusion patients in 6 to 24 months period of time. There is moderate evidence (4 trials [2, 3, 41, 48], 759 patients) that the duration of hospitalization is significantly shorter with patients of TDR group than those of the fusion group.

**Fig. 2** Pooled results of artificial total disc replacement versus lumbar fusion. The 2-year results of TDR versus fusion for the treatment of lumbar DDD were shown, concerning function status, pain, patient satisfaction, complications, work status, and reoperation rate

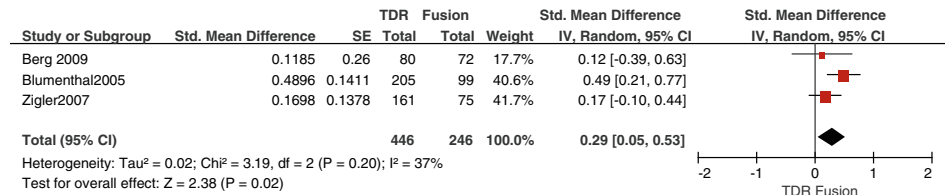
**Oswestry Disability Index**



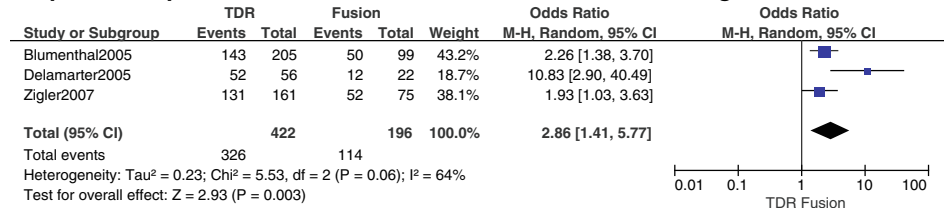
**Visual Analog Scale in Pain**



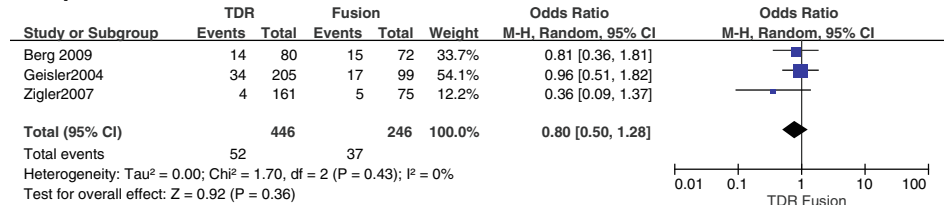
**Patient satisfaction**



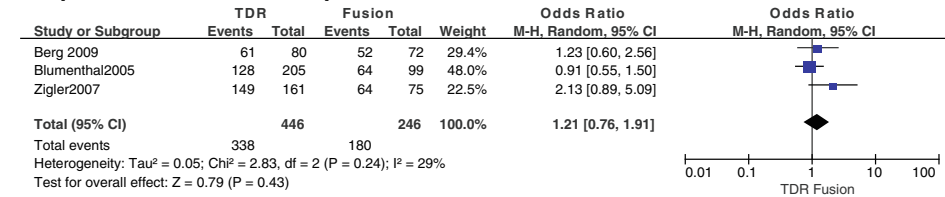
**Proportion of patients who would choose the same treatment again**



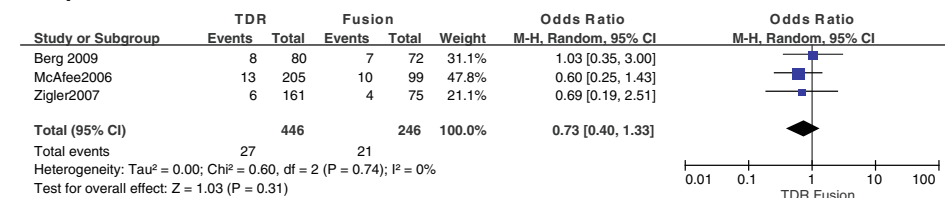
**Complications**



**Proportion of full-time and part-time work**

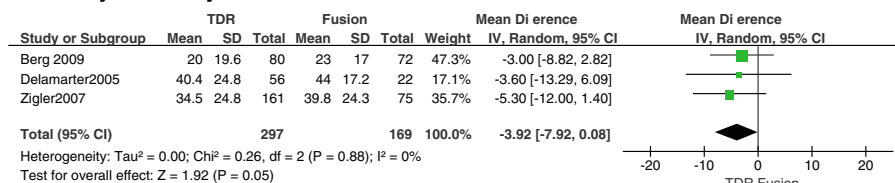


**Reoperation rate**

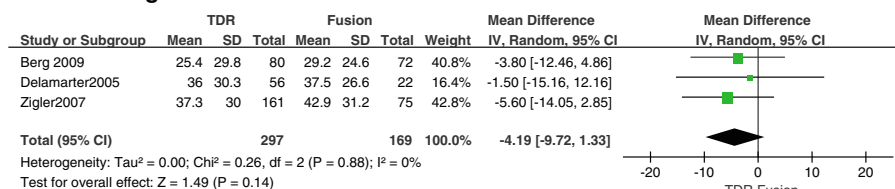


**Fig. 3** Pooled results of artificial total disc replacement versus lumbar fusion after excluding the study with stand-alone cage interbody fusion

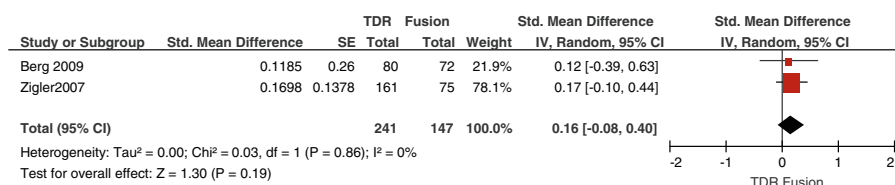
**Oswestry Disability Index**



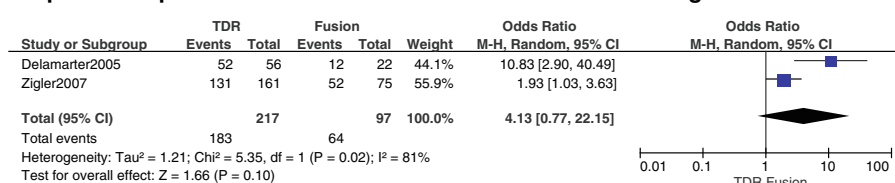
**Visual Analog Scale in Pain**



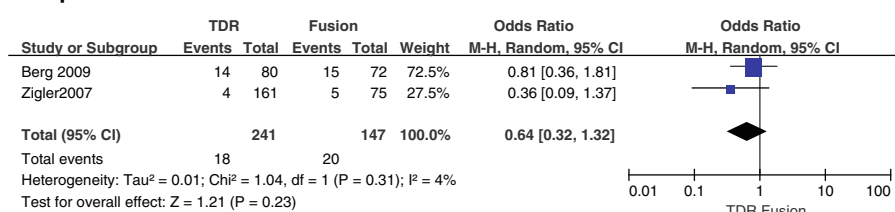
**Patient satisfaction**



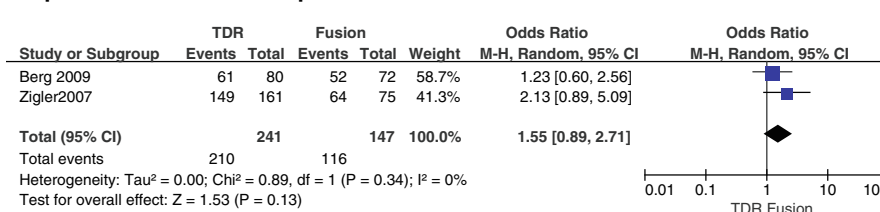
**Proportion of patients who would choose the same treatment again**



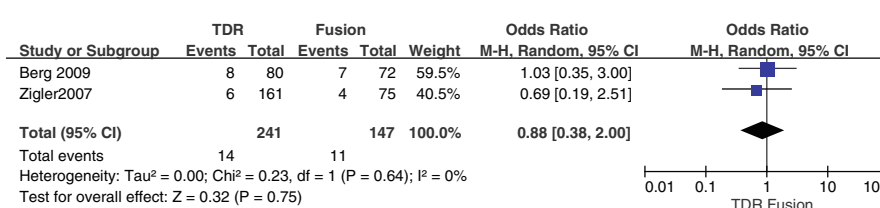
**Complications**



**Proportion of full-time and part-time work**



**Reoperation rate**





## Five-year results

Guyer et al. [29] reported the 5-year clinical follow-up results on large part of the patients participated in the previous 2-year randomized controlled trial performed by Blumenthal et al. [3]. As much as 133 randomized patients (90 CHARITE patients vs. 43 fusion patients) were available at 5 years with a follow-up rate of 57% (133/233). Overall success rate was 57.8% in the CHARITE' group versus a 51.2% rate in the fusion group by using Blumenthal's definition ( $P = 0.0359$ ) [3]. There was no statistical difference between the two groups in terms of ODI scores, VAS score, SF-36 PCS scores, and patient satisfaction at the 5-year follow-up point. A total of 65.6% in the CHARITE group versus 46.5% of patients in the fusion group have been employed full-time ( $P = 0.0403$ ). Long-term disability was recorded 8.0% of CHARITE patients statistically different from 20.9% of fusion patients ( $P = 0.0441$ ). There was no significant difference in complication (TDR 22.2% vs. fusion 32.6%,  $P = 0.20$ ) and reoperation rates (TDR 7.7% vs. fusion 16.3%,  $P = 0.14$ ) between the two groups. Additional surgeries for adjacent-level disease were performed in one (1.1%) CHARITE' patient and two (4.7%) fusion patients. The mean ROM at the index level was  $6.0^\circ$  for CHARITE' patients and  $1.0^\circ$  for fusion patients. The preoperative adjacent-level ROM was not statistically different from postoperative ROM at the 2 or 5-year point, for both CHARITE' and fusion patients, regardless of the implanted level. For the assessment of longitudinal ossification, 17 (18.9%) CHARITE cases showed lack of motion and a rating  $\geq 3$  on the longitudinal classification system using the  $5^\circ$  cut-off FDA guideline.

## Discussion

In this meta-analysis, five RCTs comparing total disc replacement to spinal fusion for degenerative lumbar disc disease are identified. It reveals that TDR results in a slightly better function and less back or leg pain in the lumbar DDD patients without clinical significance and a significantly better patient satisfaction when compared with lumbar fusion at 2-year follow-up point. In sensitivity analysis, Blumenthal's study [3] was found to be highly influential. After excluding this study, there is no longer significant difference in the ODI and VAS score, and patient's satisfaction. At 5-year point, there is no significant difference in the function and pain status, and patient satisfaction between TDR and fusion group. The complication and reoperation rates of the two groups are similar when measured respectively at 2- and 5-year points. There is strong evidence supporting that patients in both groups

demonstrated significant improvement in function and pain status compared with preoperative status at all follow-up points up to 5 years.

Lumbar fusion has been developed for several decades as the standard surgical treatment for symptomatic DDD, and various methods for achieving successful arthrodesis have been suggested. The spinal fusion is applied to eliminate the segment motion and to treat instability at the symptomatic degenerated levels and thereby could reduce or eliminate lower back pain [1]. Brantigan et al. [5] reported the 10-year results of circumferential fusion for the treatment of lumbar DDD. The high rate of clinical success (87.8%), fusion success (96.7%), and patient satisfaction (93.9%) was achieved at 10 years. Spinal fusion could alter the original biomechanics of the spine, where the loss of motion at the fused levels is compensated by increasing motion at the adjacent unfused segments [45], and a significant amount of additional force is placed on the facet joints at the adjacent unfused levels [6, 31]. As a result, the degeneration at adjacent segment may be accelerated, which is known as the adjacent segment disease (ASD) [39]. It was reported in Brantigan's study that adjacent segment degeneration occurred in 61% of patients, but was clinically significant only in 20% at 10 years after lumbar fusion [5]. Total disc replacement has been employed in an attempt to avoid disadvantages of the fusion surgery, such as adjacent segment degeneration. Harrop et al. [24] performed a system review of published incidence of radiographic adjacent segment degeneration (ASDeg) and symptomatic adjacent segment disease (ASDis) after arthrodesis or total disc replacement. The study suggested a correlation between fusion and the development of ASDeg and a stronger correlation between fusion and ASDis compared to arthroplasty, but the data support only a class C recommendation (lowest tier) for the use of arthroplasty to reduce ASDis and disc degeneration compared to arthrodesis. Despite the controversy surrounding surgical fusion of the painful degenerative functional spinal unit, without a better alternative it has de facto become the "gold standard" procedure for conservative-resistant cases [16]. So the total disc replacement, as a newly developed method for the treatment of lumbar DDD, should be examined by the "golden standard" of spinal fusion in order to measure the validity and guide for clinical practice.

The purpose of this study is to systematically compare the effectiveness and the safety of TDR to fusion for the treatment of lumbar DDD. Freeman et al. [14] have prior reviewed TDR in the lumbar spine in 2006, and subsequently Gibson et al. [19] reviewed surgery for degenerative lumbar spondylosis in 2007. Due to the lack of relevant RCTs, the statistically pooling results of TDR versus fusion were not stated in both of their reviews. In

our study, five RCTs which compare TDR to the spinal fusion are included to evaluate a total of 837 patients with lumbar DDD. When the 2-year data from four high-quality included studies were pooled, we found that the patients with TDR had a slightly better function and back or leg pain status, greater patient satisfaction, similar occurrence of complications and reoperation rate, and employment status. But we found Blumenthal's study [3] to be highly influential to the overall results in sensitivity analysis, which reported good function result and significantly higher patient satisfaction rate in TDR group compared with BAK cage interbody fusion. Because stand-alone interbody cages have limitations in biomechanical properties such as inadequate stabilization and subsidence [13, 38, 40], ALIF with BAK cage may overestimate the effect of compared TDR. After omitting Blumenthal's study, there is no longer significant difference in function and pain status and patient satisfaction between TDR and fusion group. Qualitative analysis reveals that there is strong evidence for significantly improved functioning (ODI score) and painfulness (VAS score) status in both groups compared with preoperative status. It is demonstrated that both treatments are efficacious for lumbar DDD. The 2-year results suggest that TDR has similar effectiveness and safety comparable to lumbar fusion for the treatment of lumbar DDD. However, the 2-year follow-up is too short to conclusively assess the complications and the long-term effects of adjacent-level degeneration. Guyer's study [20] provided an opportunity for the long-term assessment. At 5-year follow-up, no statistical difference is observed between TDR and the fusion groups. In addition, TDR (CHARITE') patients have shown a statistically greater rate of part-time/full-time employment and a lower rate of long-term disability than BAK fusion patients. There is no significant difference in complication and reoperation rate between two groups. Additional surgery for adjacent-level disease was operated in cases of one CHARITE' patient and two BAK fusion patients. However, these numbers are too low to draw statistical conclusion. The 5-year results are consistent with 2-year results which show the similarity of effectiveness and safety between TDR and lumbar fusion. This result is also supported by other long-term TDR researches studied beforehand [8, 32]. However, from this meta-analysis still the benefits of motion preservation and protecting adjacent levels remain unproved. Moreover, only one study with 5-year follow-up is not enough for assessment of long-term complications. In every included study, the patients were selected carefully by means of employing the similar inclusion and exclusion criteria. The mean VAS score at baseline for both groups from all studies was  $>60$  (of 100) whereas the mean ODI score at baseline was  $>40$  (of 100). These criteria did in fact result in a patient sample with significant

painfulness and disability which could contribute to good clinical results. The benefit of both interventions may not be repeated when performing them onto every DDD patient. Meticulous patient selection is essential to obtain a good clinical result [41, 48].

Meta-analysis is a statistical analysis of data collected from several different researches and surveys on the same problem, pooling outcomes in order to arrive at a more unbiased and scientific conclusion [12, 22]. Ideally, each of the studies included in meta-analysis should contain large numbers of cases and have a similar validated design. To avoid outcomes distorted by language bias, we made efforts to look for studies in all languages. Finally, only five published RCTs on lumbar TDR versus fusion were analyzed. Four studies had good methodological qualities (score  $> 50$ ), one study only gained 44 score which implies a higher risk of bias. The most prevalent methodological shortcomings appeared to be the small size of populations and the insufficiency regarding the outcome assessor blinding to intervention. The low number of included studies limited our assessment of potential publication bias by the funnel plot and unpublished researches with negative results can not be identified. Therefore, publication bias may exist, which could result in the overestimation of the effectiveness of interventions. In included studies, the interventions were inconsistent. Different procedures of lumbar fusion and different types of artificial disc may modify the comparing effect between the interventions, although no artificial disc is shown to be superior or inferior to the others [2]. Fusion method could result in different operative data and radiographic results, even if there is no significant difference in clinical and function results [15, 23, 33, 44]. Due to these limitations, the combined results of this meta-analysis should be cautiously accepted. In addition, the benefits of motion preservation and protecting adjacent levels, long-term complications and surgical revisions still remain unproved from the existing data. More independent high-quality RCTs with long-term outcomes and cost/effectiveness analyses are needed.

## Conclusions

Compared with lumbar fusion, total artificial disc replacement results in a slightly better functioning and back or leg pain status without clinical significance, and a significantly greater patient satisfaction at the two-year follow-up point. However, the study that used stand-alone cage interbody fusion as the control is highly influential to overall results. After omitting this study, there is no longer significant difference in function and pain status and patient satisfaction between TDR and fusion group. At five years, these outcomes are not significantly different

between comparing groups. Complication and reoperation rate are similar between the two groups at 2 and 5 years, respectively. From the existing outcomes, the total artificial disc replacement does not show significant superiority for the treatment of lumbar DDD compared with fusion. To assess the benefits of motion preservation and the long-term complications, more high-quality RCTs with long-term outcomes are needed.

**Conflict of interest statement** None.

## References

- An H, Boden SD, Kang J et al (2003) Summary statement: emerging techniques for treatment of degenerative lumbar disc disease. *Spine* 28:S24–S25
- Berg S, Tullberg T, Branth B et al (2009) Total disc replacement compared to lumbar fusion: a randomised controlled trial with 2-year follow-up. *Eur Spine J* 18:1512–1519
- Blumenthal S, McAfee PC, Guyer RD et al (2005) 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* 30:1565–1575
- Booth DC, Bridwell KH, Eisenberg BA et al (1999) Minimum 5-year results of degenerative spondylolisthesis treated with decompression and instrumented posterior fusion. *Spine* 24:1721–1727
- Brantigan JW, Neidre A, Toohey JS (2004) The Lumbar I/F Cage for posterior lumbar interbody fusion with the variable screw placement system: 10-year results of a Food and Drug Administration clinical trial. *Spine J* 4:681–688
- Chow DH, Luk KD, Evans JH et al (1996) Effects of short anterior lumbar interbody fusion on biomechanics of neighboring unfused segments. *Spine* 21:549–555
- Csecsei GI, Klekner AP, Dobai J et al (2000) Posterior interbody fusion using laminectomy bone and transpedicular screw fixation in the treatment of lumbar spondylolisthesis. *Surg Neurol* 53:2–7
- David T (2007) Long-term results of one-level lumbar arthroplasty: minimum 10-year follow-up of the CHARITE artificial disc in 106 patients. *Spine* 32:661–666
- Delamarter RB, Bae HW, Pradhan BB (2005) Clinical results of ProDisc-II lumbar total disc replacement: report from the United States clinical trial. *Orthop Clin North Am* 36:301–313
- Delamarter RB, Fribourg DM, Kanim LE et al (2003) ProDisc artificial total lumbar disc replacement: introduction and early results from the United States clinical trial. *Spine* 28:S167–S175
- DerSimonian R, Laird N (1986) Meta-analysis in clinical trials. *Control Clin Trials* 7:177–188
- Dickman CA, Yahiro MA, Lu HTC et al (1994) Surgical treatment alternatives for fixation of unstable fractures of the thoracic and lumbar spine. A meta-analysis. *Spine* 19:S2266–S2273
- Dimar JR 2nd, Beck DJ, Glassman SD et al (2001) Posterior lumbar interbody cages do not augment segmental biomechanical stability. *Am J Orthop (Belle Mead NJ)* 30:636–639
- Freeman BJ, Davenport J (2006) Total disc replacement in the lumbar spine: a systematic review of the literature. *Eur Spine J* 15:S439–S447
- Fritzell P, Hägg O, Wessberg P et al (2001) Volvo award winner in clinical studies: lumbar fusion versus non-surgical treatment for chronic low back pain. A multi-centre randomised controlled trial from the Swedish lumbar spine study group. *Spine* 26:2521–2534
- Fritzell P, Hägg O, Wessberg P et al (2002) Chronic low back pain and fusion: a comparison of three surgical techniques: a prospective multicenter randomized study from the Swedish Lumbar spine study group. *Spine* 27:1131–1141
- Fujiya M, Saita M, Kaneda K et al (1990) Clinical study on stability of combined distraction and compression rod instrumentation with posterolateral fusion for unstable degenerative spondylolisthesis. *Spine* 15:1216–1222
- Geisler FH, Blumenthal SL, Guyer RD et al (2004) 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 Charité intervertebral disc. *J Neurosurg Spine* 1:143–154
- Gibson JN, Waddell G (2005) Surgery for degenerative lumbar spondylosis. *Cochrane Database Syst Rev* 4:CD001352
- Guyer RD, McAfee PC, Banco RJ et al (2009) 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 J* 9:374–386
- Guyer RD, McAfee PC, Hochschuler SH et al (2004) Prospective randomized study of the Charite artificial disc: data from two investigational centers. *Spine J* 4:S252–S259
- Haheer TR, Merola A, Zpinick RI et al (1995) Meta-analysis of surgical outcome in adolescent idiopathic scoliosis. A 35 year English literature review of 11, 000 patients. *Spine* 20:1575–1584
- Han X, Zhu Y, Cui C et al (2009) A meta-analysis of circumferential fusion versus instrumented posterolateral fusion in the lumbar spine. *Spine* 34:E618–E625
- Harrop JS, Youssef JA, Maltenfort M et al (2008) Lumbar adjacent segment degeneration and disease after arthrodesis and total disc arthroplasty. *Spine* 33:1701–1707
- Higgins JP, Thompson SG (2002) Quantifying heterogeneity in a meta-analysis. *Stat Med* 21:1539–1558
- Higgins JP, Thompson SG, Deeks JJ et al (2003) Measuring inconsistency in meta analyses. *BMJ* 327:557–560
- Higgins JPT, Green S (2008) *Cochrane handbook for systematic reviews of interventions* version 5.0.1 [updated September 2008]. The Cochrane Collaboration
- Koes BW, Scholten RJ, Mens JM et al (1995) Efficacy of epidural steroid injections for low-back pain and sciatica: a systematic review of randomized clinical trials. *Pain* 63:279–288
- Lauridsen HH, Hartvigsen J, Manniche C et al (2006) Responsiveness and minimal clinically important difference for pain and disability instruments in low back pain patients. *BMC Musculoskelet Disord* 25:82
- Le Huec JC, Mathews H, Basso Y et al (2005) Clinical results of Maverick lumbar total disc replacement: two-year prospective follow-up. *Orthop Clin North Am* 36:315–322
- Lee CK, Langrana NA (1984) Lumbosacral spinal fusion: a biomechanical study. *Spine* 9:574–581
- Lemaire JP, Carrier H, Sariali el-H et al (2005) Clinical and radiological outcomes with the Charité artificial disc: a 10-year minimum follow-up. *J Spinal Disord Tech* 18:353–359
- Madan SS, Boeree NR (2003) Comparison of instrumented anterior interbody fusion with instrumented circumferential lumbar fusion. *Eur Spine J* 12:567–575
- McAfee PC, Cunningham B, Holsapple G et al (2005) 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* 30:1576–1583
35. McAfee PC, Fedder IL, Saiedy S et al (2003) SB Charité disc replacement: report of 60 prospective randomized cases in a US center. *J Spinal Disord Tech* 16:424–433
  36. McAfee PC, Geisler FH, Saiedy SS et al (2006) Revisability of the CHARITE artificial disc replacement: analysis of 688 patients enrolled in the U.S. IDE study of the CHARITE Artificial Disc. *Spine* 31:1217–1226
  37. Ostelo RW, de Vet HC (2005) Clinically important outcomes in low back pain. *Best Pract Res Clin Rheumatol* 19:593–607
  38. Oxland TR, Hoffer Z, Nydegger T et al (2000) A comparative biomechanical investigation of anterior lumbar interbody cages: central and bilateral approaches. *J Bone Joint Surg Am* 823:383–393
  39. Park P, Garton HJ, Gala VC et al (2004) Adjacent segment disease after lumbar or lumbosacral fusion: review of the literature. *Spine* 29:1938–1944
  40. Pellisé F, Puig O, Rivas A et al (2002) Low fusion rate after L5–S1 laparoscopic anterior lumbar interbody fusion using twin stand-alone carbon fiber cages. *Spine* 27:1665–1669
  41. Sasso RC, Foulk DM, Hahn M (2008) Prospective, randomized trial of metal-on-metal artificial lumbar disc replacement: initial results for treatment of discogenic pain. *Spine* 33:123–131
  42. Siepe CJ, Mayer HM, Wiechert K et al (2006) Clinical results of total lumbar disc replacement with ProDisc II: three-year results for different indications. *Spine* 31:1923–1932
  43. Szpalski M, Gunzburg R, Mayer M (2002) Spine arthroplasty: a historical review. *Eur Spine J* 11:S65–S84
  44. Tiusanen H, Schlenzka D, Seitsalo S et al (1996) Results of a trial of anterior or circumferential lumbar fusion in the treatment of severe isthmic spondylolisthesis in young patients. *J Pediatr Orthop B* 5:190–194
  45. Tokes IA, Wilder DG, Frymoyer JW et al (1981) 1980 Volvo award in clinical sciences. Assessment of patients with low-back pain by biplanar radiographic measurement of intervertebral motion. *Spine* 6:233–240
  46. Van Tulder M, Furlan A, Bombardier C et al (2003) Updated method guidelines for systematic reviews in the Cochrane collaboration back review group. *Spine* 28:1290–1299
  47. Zigler JE, Burd TA, Vialle EN et al (2003) Lumbar spine arthroplasty: early results using the ProDisc II: a prospective randomized trial of arthroplasty versus fusion. *J Spinal Disord Tech* 16:352–361
  48. Zigler J, Delamarter R, Spivak JM et al (2007) 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* 32:1155–1162