

# Comparison of artificial total disc replacement versus fusion for lumbar degenerative disc disease: a meta-analysis of randomized controlled trials

Jiangbo Wei · Yueming Song · Lin Sun · Chaoliang Lv

Received: 8 March 2013 / Accepted: 21 March 2013 / Published online: 4 May 2013  
© The Author(s) 2013. This article is published with open access at Springerlink.com

## Abstract

**Purpose** To compare the efficacy and safety of TDR to that of the fusion for the treatment of lumbar degenerative disc disease (LDDD).

**Methods** Randomized controlled trials comparing TDR with any other intervention for LDDD were acquired by a comprehensive search in PubMedCentral, MEDLINE, EMBASE, BIOSIS, ClinicalTrials.gov, and the FDA trials register. Methodologic quality was assessed and relevant data were retrieved, and appropriate meta-analysis was performed. Two review authors independently selected studies, extracted data, and assessed risk of bias. Results and upper bounds of confidence intervals were compared with predefined clinically relevant differences.

**Results** Six relevant randomized controlled trials (RCTs) involving 1,603 patients were identified and reported two year follow-up results. Patients in TDR group compared with lumbar fusion group demonstrated significant improvements in ODI, VAS scores and complication rates at the two year follow-up. Meanwhile, except for operating time in anterior group, intra-operative blood loss, operating time in posterior group, and reoperation rate were without clinical significance between the two groups. In addition, the range of motion (ROM) was maintained within normal ranges after TDR.

**Conclusions** The results showed the TDR has significant safety and efficacy comparable to lumbar fusion at two year follow-up. Although superiority compared to fusion could not be proved, by comparing clinical symptoms relieved, motion preserved, and the low reoperation rate during long-term follow-up on TDR, TDR was considered safe and

effective. Therefore, the authors suggest adopting TDR on a large scale; with failure of TDR, interbody fusion would be performed.

## Introduction

Lumbar degenerative disc disease (LDDD) is a major cause of chronic low-back pain with lumbar segmental instability, in which surgical intervention is required when conservative treatment fails. Lumbar fusion has been developed for several decades as the gold standard for treatment of symptomatic LDDD. A mechanically stable fusion of the involved lumbar segments should reduce pain [1, 2]. However, this method is not perfect, not only because of complications, but also because of increasing stress on the adjacent segments may cause new instability and pain [3–5]. A variety of surgical fusion techniques have been performed during the last century; radiographically confirmed fusion rates have exceeded 96.7 % [6]. However, all of this has not translated into an improvement of successful clinical outcomes [7, 8]. This seems to indicate that fusion was not the parameter which determines the clinical success. In recent years, the idea of the nonfusion treatment has been gradually accepted by spine surgeons and the patients, and used in clinical practice widely.

Artificial total disc replacement (TDR) as an alternative to spinal fusion has been increasingly applied for surgically treating LDDD [9, 10]. It was postulated that the patient's normal intervertebral segment motion might be restored and maintained while the adjacent level was prevented from nonphysiologic loading and thus the pain was relieved [11–13]. Previous studies that compared the clinical effects of TDR to fusion for treating LDDD provided ambiguous results [14–16]. Therefore, it was still uncertain whether TDR was more effective and safer than fusion. The objective

J. Wei · Y. Song (✉) · L. Sun · C. Lv  
Department of Orthopedics, West China Hospital,  
Sichuan University, 37# Guoxue Rd,  
Chengdu 610041, People's Republic of China  
e-mail: sym\_cd@yahoo.com.cn

of this study was to systematically compare the efficacy and safety of TDR to fusion for the treatment of LDDD.

## Materials and methods

### Study selection

All randomized controlled trials comparing TDR with fusion for the treatment of LDDD were identified. We searched databases including PubMedCentral, MEDLINE (from 1966), EMBASE(from 1980), BIOSIS (from 2004), ClinicalTrials.gov, and FDA trials register. The following sources were searched up to 30 January 2013. The search strategy consisted of a combination of keywords concerning the technical procedure (*lumbar degenerative disc disease, lumbar disc replacement, artificial total disc replacement, lumbar arthroplasty, lumbar fusion, prosthesis, implantation, and randomized controlled trial*) and keywords regarding the anatomical features and pathology (*lumbar vertebrae*). These keywords were used as MESH headings and free text words. In addition, a search was performed using the specific names of the prostheses. In this systematic review and meta-analysis, all of relevant review articles and other potentially eligible studies were randomized controlled trials (RCTs), the search was limited to studies published in English. Studies for the review were randomized controlled trials, published in a peer-reviewed journal as a full article, excluding grey literature and conference proceedings.

Two authors (W.J.B. and S.Y.M.) checked all titles and abstracts from the databases independently which met our search terms and reviewed full publications, excepted from this, an authors (L.S.) as a referee independently selected the articles from the list of identified references whenever necessary. The reference section of all primary studies was inspected for additional references, and only those reporting the results of a randomized controlled trial (RCTs) were included in this analysis. This review was conducted under the suggested QUORUM guideline standards [17]. If studies did not report the actual number or the standard deviation, but rather presented the data only in graph format, the authors were contacted. Most authors responded but were not able to provide additional clarification because of personal circumstances or because the data presented were preliminary and not available for scientific research. The main characteristics of the studies included in the meta-analysis are presented in Table 1.

### Data extraction

Two authors (W.J.B. and S.Y.M.) independently extracted relevant data from the included studies regarding design, age, gender, type of disc prosthesis, type of fusion

intervention, and follow-up period. The outcomes pooled in this analysis include visual analogue scale (VAS), Oswestry disability index (ODI), intra-operative blood loss, operating time, proportion of full-time and part-time work, complication, reoperation rate, and the range of motion (ROM).

### Assessment of heterogeneity

The clinical homogeneity related to characteristics of the participants, including age, sex, clinical manifestation, pain, function status baseline, and so on. Surgical technique homogeneity compared 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, which was directly calculated from the  $Q$  statistic and describes the percentage of variation across the studies that was due to heterogeneity rather than change.  $I^2$  ranges from 0 to 100 %, with 0 indicating the absence of any heterogeneity. When  $I^2$  was <50 %, low heterogeneity was assumed, and the effect was thought to be due to change. Conversely, when  $I^2$ >50 %, heterogeneity was thought to exist and the effect was random.

### Assessment of risk of bias

Inadequate methodology in randomized controlled trials leads to the risk of intervention effects overestimation [18], so assessment of risk of bias in included studies was necessary. The assessment of bias risk of the included studies is presented in Table 2. Controlled trials were assessed using a criteria list recommended by the Cochrane Back Review Group [19]. The following criteria were scored yes, no, or unsure: criteria 12 was scored not applicable because we consider compliance no relevant for surgical interventions. If studies met at least 6 of the 12 items, it was considered low.

### Statistical analysis

Attempts were made to statistically pool the data of homogeneous studies in order to obtain the primary and the secondary outcomes. According to the recommendations of the Cochrane Collaboration, the meta-analyses were using RevMan software (version5.0) provided by the organization. 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 were measured in different scales, standardized mean difference (SMD) and 95 % CI were calculated. If in some studies outcomes were shown as dichotomous data while in the other studies expressed as continuous data, odds

**Table 1** Main characteristics of studies included in the meta-analysis

	Patients, n		Mean Age, y		Male, %		Type of disk	Surgical approaches (fusion)	Follow-up, y
	TDR	Fusion	TDR	Fusion	TDR	Fusion			
Berg et al. [20]	80	72	40.2	38.5	40	42	CHARITE or ProDisc-L, or Maverick	posterior	2
Blumenthal et al. [21]	205	99	39.6	39.6	55.1	44.4	CHARITE	anterior	2
Delamarter et al. [22]	56	22	39.7	44.2	57	45	ProDisc-L	posterior	2
Zigler et al. [23]	161	75	38.7	40.4	50.9	45.3	ProDisc-L	posterior	2
Delamarter et al. [24]	174	82	41.8	41.8	57.6	54.2	ProDisc-L	posterior	2
Gornet et al. [25]	405	172	39.9	40.2	50	50.6	ProDisc-L	anterior	2

ratios would be repressed as standardized mean difference to allow dichotomous and continuous data to be pooled together. We performed a sensitivity analysis for the measured effects omitting the study which may largely influence the clinical results. If when  $I^2 > 50\%$ , only the randomized effect model was accurate.

## Results

The process of identifying relevant studies is summarized in Table 1. From the selected databases, 382 references were obtained. By screening the titles and abstracts, 366 references were excluded due to the irrelevance to this topic. In 16 potentially relevant references, 10 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 six independent continuous clinical randomized fusion trials, reporting for different follow-up periods or containing separated results. The six RCTs with the relevant information were eventually included involving 1,603 patients [20–25]. All of these studies reported two year follow-up results.

### Description

The characteristics of six included studies are summarized in Table 1. In all included studies, the patients with symptomatic LDDD were recruited with sample size ranging from 78 to 577 patients. Berg et al. [20]. performed artificial disc replacement with one of three following devices: CHARITE<sup>®</sup>, ProDisc-L, or Maverick (Medtronic, Memphis, TN) compared with the posterolateral fusion (PLF) with autologous bone graft or posterior interbody fusion (PLIF) with two carbon fiber cages. Blumenthal et al. [21]. performed the CHARITE<sup>®</sup> artificial disc (DePuy Spine, Raynham, MA) replacement compared with anterior lumbar interbody fusion (ALIF) with BAK cages. In two studies, the ProDisc-L (Synthes Spine, West Chester, PA) [22, 23], artificial disc were employed compared with circumferential fusion. Delamarter et al. [24]. reported the prosthesis

consists of two end plates manufactured from cobalt-chromium-molybdenum (CoCrMo) alloy and a convex ultra-high molecular weight polyethylene (UHMWPE) insert. Fixation was provided by a central keel and titanium plasma spray coating on each end plate. In addition, Gornet et al. [25]. performed the MAVERICK Disc (Medtronic, Memphis, TN) replacement compared with anterior interbody fusion with rhBMP-2 on an absorbable collagen sponge (INFUSE Bone Graft, Medtronic) and tapered fusion cages (LT-CAGE Lumbar Tapered Fusion Device, Medtronic).

### Risk of bias

The outcomes are presented in Table 3. A fixed blocking method of randomization with six assignments per block was described in five studies [20–24]. Sealed envelop technique for allocation concealment was applied in three studies [20, 22, 23]. In three studies, the participants remained blinded until the operation was finished. All of the participants in the six studies had performed the follow-up for at two years and a follow-up rate of more than 89 % was obtained in four of these studies [20–23]. None of the included studies encompassed the information of intention-to-treat (ITT) analysis. Overall, there trials were classified as more high methodological quality.

### Visual analogue scale. (VAS)

The scales measuring the duration and intensity of pain have scores ranging from 0 to 10, with a lower score representing a better condition. The composite back or/and leg pain score was derived by multiplying the intensity and duration scores. By this method, the composite score could range from 0 to 100. At two years, VAS pain scores recorded for back or/and leg pain indicated statistically significant improvement from preoperative levels regardless of treatment. Only five trials reported the continuous outcome measures in the form of mean SD, and so they were included in the meta-analysis [20–24]. They enrolled 1,603 patients, with 1,081 patients assigned to the TDR group and 522 patients

**Table 2** Criteria for risk of bias assessment

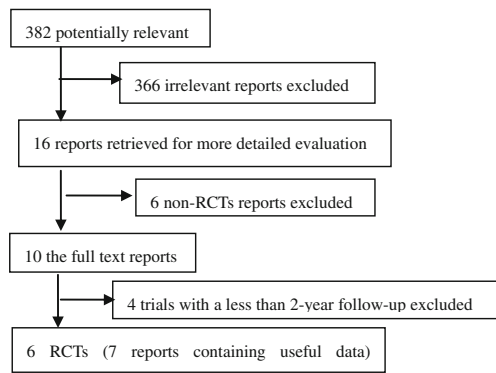
Question	Criteria	Judgment
1 Was the method of randomization adequate?	A random (unpredictable) assignment sequence. Adequate methods including computer-generated random sequence, coin toss, rolling a dice, drawing of ballots with the study group labels from a dark bag, preordered sealed envelopes, and sequentially ordered vials.	Yes/No/Unsure
2 Was the treatment allocation concealed?	Adequate if the allocation of patients involved a central independent unit, on-site locked computer, identically appearing numbered drug bottles or containers prepared by an independent pharmacist or investigator, or sealed envelopes.	Yes/No/Unsure
3 Was the patient blinded to the intervention?	The index and fusion groups are indistinguishable for the patients.	Yes/No/Unsure
4 Was the care provider blinded to the intervention?	The index and fusion groups are indistinguishable for the care providers.	Yes/No/Unsure
5 Was the outcome assessor blinded to the intervention?	For patient-reported outcomes with adequately blinded patients for outcome criteria that suppose a contact between participants and outcome assessors: the blinding procedure is adequate if patients are blinded, and the treatment or adverse effects of the treatment cannot be noticed during examination for outcome criteria that do not suppose a contact with participants: the blinding procedure is adequate if the treatment or adverse effects of the treatment cannot be noticed during the assessment for outcome criteria that are clinical or therapeutic events that will be determined by the interaction between patients and care providers, in which the care provider is the outcome assessor: the report needs to be free of selective outcome reporting.	Yes/No/Unsure
6 Was the dropout rate described and acceptable?	The number of participants who were included in the study but did not complete the observation period or were not included in the analysis are described and reasons are given and are <20 % or short-term and <30 % for long-term follow-up.	Yes/No/Unsure
7 Were all randomized participants analyses in the group to which they were allocated?	All randomized patients are reported/analyzed in the group they were allocated to by randomization for the most important moments of effect measurement (minus missing values) irrespective of noncompliance and cointerventions.	Yes/No/Unsure
8 Are reports of the study free of suggestion of selective outcome reporting?	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.	Yes/No/Unsure
9 Were the groups similar at baseline regarding the most important prognostic indicators?	The groups have to be similar at baseline regarding demographic factors, duration, and severity of complaints, percentage of patients with neurological symptoms, and value of main outcome measure(s).	Yes/No/Unsure
10 Were co-interventions avoided or similar?	There were no cointerventions or they were similar between the index and fusion groups.	Yes/No/Unsure
11 Was compliance acceptable in all groups?	The compliance with the interventions based on the reported intensity, duration, number, and frequency of sessions for both the index intervention and fusion intervention.	Yes/No/Unsure
12 Was the timing of the outcome assessment similar in all groups?	Timing of outcome assessment was identical for all intervention groups and for all important outcome assessments.	Yes/No/Unsure

assigned to the fusion group. As for the outcome of VAS scores, the test for heterogeneity demonstrated that no significant heterogeneity existed across the five studies ( $P=0.80$ ;  $I^2=0\%$ ), and the fixed model was performed. Compared with patients treated with fusion, patients treated with artificial TDR showed a significant decrease (SMD,  $-3.18$ ; CI,  $-5.74$  to  $-0.63$ ;  $P=0.01$ ) (Figs. 2 and 3). At 2-year follow-up, the patient functioning ability measured by

VAS in the TDR group was better than that of the fusion group with statistical significance.

Oswestry disability index(ODI)

The ODI was a validated questionnaire that assesses a patient's disability during activities of daily living. ODI success was defined as 15 % improvement from baseline. Regardless of



**Fig. 1** Flowchart showing the identification, inclusion, and exclusion of the randomized controlled trials (RCTs)

treatment, all patients showed statistically significant improvement in ODI scores at two year compared with baseline. Six trials reported the continuous outcome measures as mean standard deviation(SD), and they were included in the meta-analysis. Six trials enrolled 1,603 patients, with 1,081 patients assigned to the TDR group and 522 patients assigned to the fusion group. The test for heterogeneity demonstrated significant heterogeneity existed across the six studies ( $P=0.88$ ;  $I^2=0\%$ ), and the random model was performed. The outcome shown low heterogeneity. Overall, a two year follow-up, the patient functioning ability measured by ODI in the TDR group (SMD,  $-5.13$ ; CI,  $-7.35$  to  $-2.90$ ;  $P<0.0001$ ) was better than that of the fusion group with statistical significance.

**Intraoperative blood loss and operating time**

Both the intra-operative blood loss and operating time were reported in five trials [20, 21, 23–25]. Five trials enrolled

1,525 patients with 1,025 patients assigned to the TDR group and the other 500 patients assigned to the fusion group. The test for heterogeneity of two groups demonstrated existed significant heterogeneity( $I^2=95\%$ ;  $I^2=98\%$ ; respectively). The greater heterogeneity came from surgery approaches different in fusion group, therefore this trails divided into two subgroup (e.g., posterior and anterior group) separately based on surgery approaches different when meta-analysis, and the randomized effect model was performed. Overall, patients treated with TDR showed no significant difference no matter which methods was used compared to patients treated with fusion in the intra-operative blood loss (SMD,  $-92.39$ ; CI,  $-309.05$ – $124.27$ ;  $P=0.40$  or SMD, $70.53$ ; CI,  $-75.87$ – $216.94$ ;  $P=0.35$  respectively) (Figs. 4 and 5). Meantime, in operating time, there were significant difference in anterior group (SMD,  $-81.16$ ; CI,  $-143.60$ – $18.71$ ;  $P=0.01$ )(Fig. 6), while in posterior group, that between TDR and fusion were no difference(SMD, $12.49$ ; CI,  $-13.85$ – $38.83$ ;  $P=0.35$ ) (Fig. 7).

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

There were five trails [20, 21, 23–25] reports proportion of full-time and part-time work. Its enrolled 1,525 patients, with 1,025 patients assigned to the TDR group and the other 500 patients assigned to the fusion group. The rate was 72.7 % (745 of 1025) in the TDR group and 70.2 % (351 of 500) in the fusion group. The test for heterogeneity demonstrated that no significant heterogeneity existed across the five studies ( $P=0.62$ ;  $I^2=0\%$ ), and the fixed model was performed. Patients treated with TDR showed a significant decrease (OR 1.14; 95 % CI [0.89–1.44];  $P=0.30$ ) compared to patients treated with fusion (Fig. 6).

**Table 3** Risk of bias assessment of included studies

Comparison	Reference	Berg et al. [20]	Blumenthal et al. [21]	Delamarter et al. [22]	Zigler et al. [23]	Delamarter et al. [24]	Gornet et al. [25]
1	Randomization	Y	Y	Y	Y	Y	Y
2	Allocation Concealment	Y	Y	Y	N	Y	N
3	Patient Blinding	Y	N	Y	N	Y	Y
4	Surgeon Blinding	Y	N	N	N	N	Y
5	Outcome Blinding	N	N	N	N	N	Y
6	Dropouts	Y	Y	Y	Y	Y	Y
7	Intention to Treat	N	N	N	N	N	N
8	Selective Reporting	U	U	U	U	U	U
9	Baseline	Y	Y	Y	Y	Y	Y
10	Cointerventions	Y	Y	Y	Y	Y	Y
11	Compliance	Y	Y	Y	Y	Y	Y
12	Outcome Timing	Y	Y	Y	Y	Y	Y
Risk of bias		9/12 (low)	7/12 (low)	8/12(low)	6/12 (low)	8/12(low)	9/12 (low)

U unsure, Y yes, N no

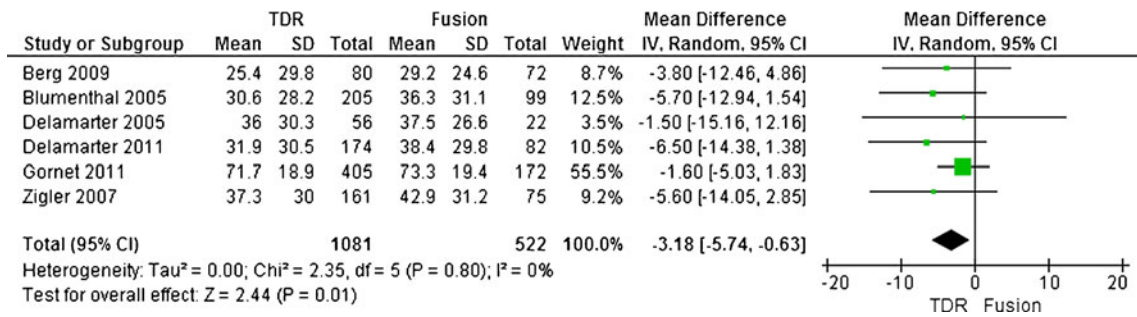


Fig. 2 Results of the meta-analysis for the visual analogue scale (VAS) for TDR and fusion groups at 2-year

Overall, the patient functioning ability measured by the proportion of full-time and part-time work in the TDR group was equivalent to the fusion group follow-up two years.

Complication

Complication including device failures necessitating reoperation, revision, removal occurred, major vessel injury, neurologic damage, nerve root injury, death and so on in the study. The complication rate was reported in five trials [21–25]. Five trials enrolled 1,525 patients, with 1,025 patients assigned to the TDR group and the other 500 patients assigned to the fusion group. The complication rate was 5.8 % (59 of 1025) in the TDR group and 10.8 % (54 of 500) in the fusion group. The test for heterogeneity of two studies demonstrated existed significant heterogeneity (P=0.03; I<sup>2</sup>=63 %), and the randomized effect model was performed. Patients treated with TDR showed no significant decrease in the overall reoperation rate (OR 0.57; 95 % CI [0.38–0.84]; P=0.005) compared with patients treated with fusion (Fig. 7). Overall, a two year follow-up, the patient functioning ability measured by complication in the TDR group was better than that of the fusion group with statistical significance.

Overall reoperation rate

Secondary surgical procedures were defined as any revision, removal, or reoperation of the implant or supplemental

fixation. The overall reoperation rate was reported in five trials [20, 22–25]. Five trials enrolled 1,525 patients, with 1,025 patients assigned to the TDR group and 500 patients assigned to the fusion group. The overall reoperation rate was 5.2 % (53 of 1025) in the TDR group and 6 % (30 of 500) in the fusion group. The test for heterogeneity of two studies demonstrated existed significant heterogeneity (P=0.07; I<sup>2</sup>=51 %), and the randomized effect model was performed. Patients treated with TDR showed no different in the overall reoperation rate (OR, 0.91; CI, 0.57 to 1.46; P=0.71) compared with patients treated with fusion (Figs. 8, 9 and 10).

The range of motion (ROM)

Angular motion in the sagittal plane was measured by comparing lateral flexion and extension radiographs. Because lacked of the consistent criterion to ROM of lumbar operation, the five RCTs reports lumbar ROM after treatment of two groups only made common statistical descriptive analysis. By the FDA definition, failures would include a patient with 7° ROM before surgery at the index level that maintains 7° at 24 months or a hypermobile segment returned to a normal functional ROM, and successes would include improvement from 1° ROM before surgery to a nonfunctional 2° ROM after surgery. Delamarter et al. [22] followed up patients at 6 months, and found that difference of operative segments in L4-5 were statistically significant while operative segments in L5-S1 was not. Berg et al. [20]

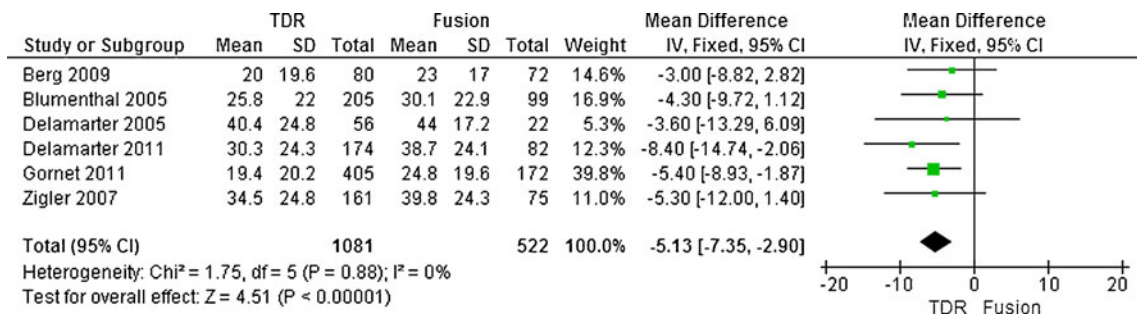
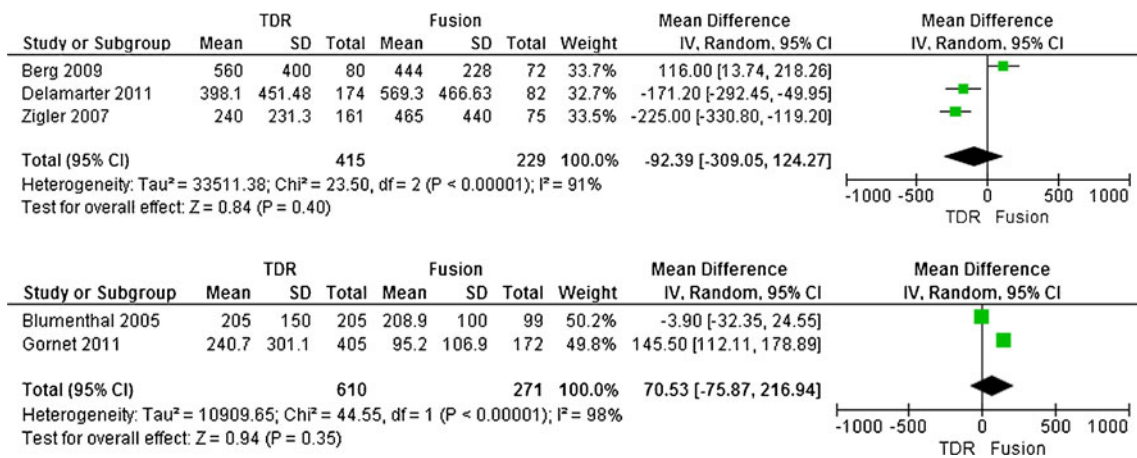


Fig. 3 Results of the meta-analysis for the Oswestry disability index (ODI) for TDR and fusion groups at 2-year



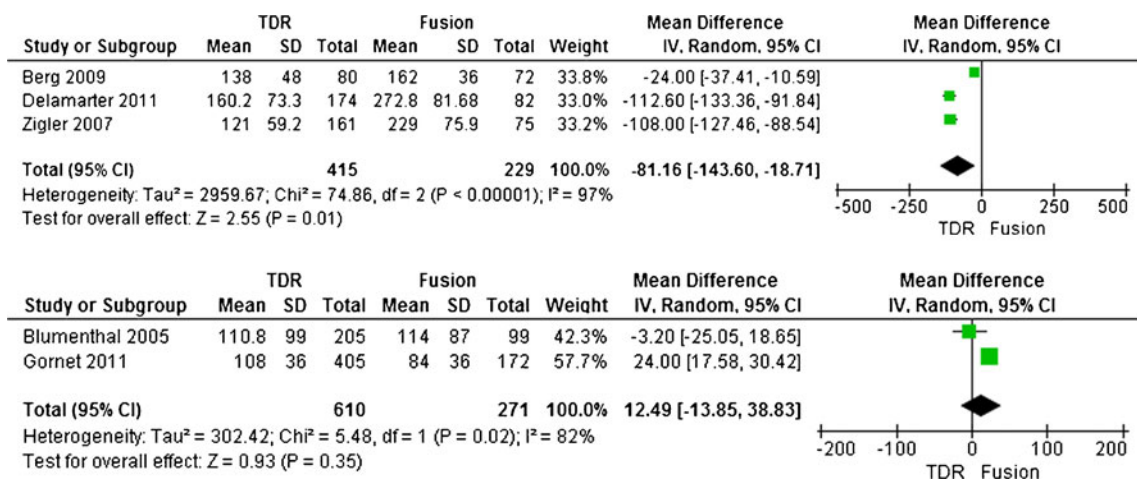
**Figs. 4, 5** Results of the meta-analysis for intraoperative blood loss for TDR and fusion subgroups at 2-year

reports the ROM of TDR better than the fusion group at operation level of L4-5 or/and L5-S1, the different of index mean were not statistically significant compared to pre- and postoperative after two years, in other words, TDR retained the normal ranges at operation level. Zigler et al. [23] found that ROM was maintained within a normal functional range in 93.7 % of TDR patients at two year follow-up (Delamarter et al. [24]) reported the ROM in the TDR group averaged  $7.8^{\circ} \pm 5.3^{\circ}$  at the level of the superior disc and  $6.2^{\circ} \pm 4.1^{\circ}$  at the level of the inferior disc at two year. Gornet et al. [25] found at the TDR group, Compared to the mean angular motion value was  $7.0^{\circ}$  before surgery, motion varied at 12 and two year were  $9.4^{\circ}$  and  $9.5^{\circ}$  respectively, demonstrating not only maintenance but also increase in segmental motion after TDR. Overall, the range of motion (ROM) was maintained within normal ranges after TDR, and demonstrated the benefits of motion preservation for the deceleration of adjacent level degeneration.

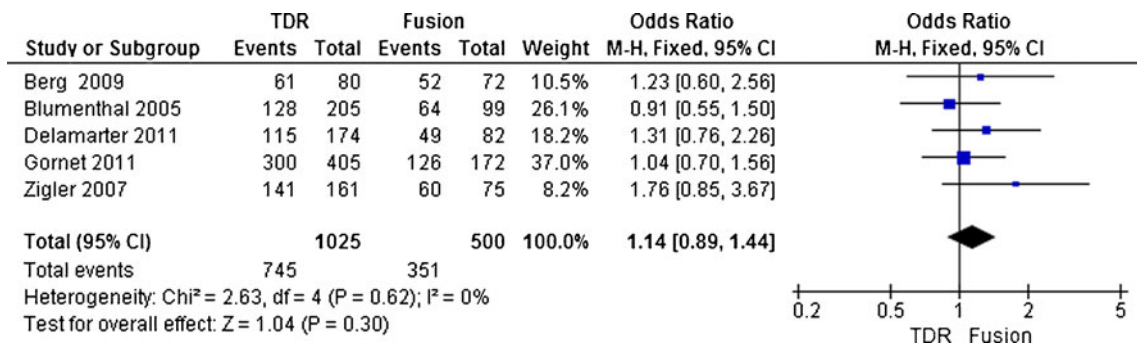
**Discussion**

Safety

There were five trials [20, 22–25] assessing the safety from four facets, including intra-operative blood loss, operative time, surgical complications and the reoperation rate. In Berg et al. [20], both operation time and hospital stay were shorter in the TDR group than in the fusion group, but complications and reoperations were similar in both groups. Blumenthal et al. [21] reported there was no difference between the two groups with respect to operative time, blood loss and complications, but the hospital stay was significantly shorter in the TDR group. Moreover, he believed its safety and efficacy would significantly increase when performed by a skilled surgeon. On the contrary, Zigler et al. [23] showed that the TDR group was statistically significantly lower with regard to operative time,



**Figs. 6, 7** Results of the meta-analysis for operating time for TDR and fusion subgroups at 2-year



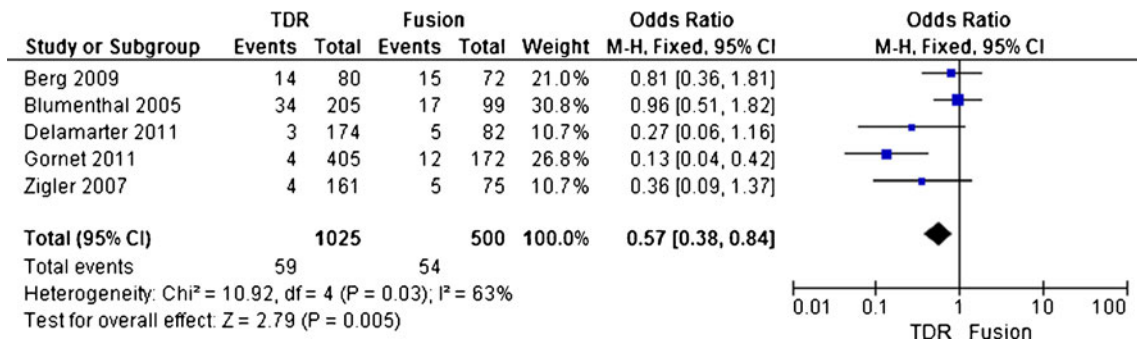
**Fig. 8** Results of the meta-analysis for proportion of full-time and part-time work for TDR and fusion groups at 2-year

estimated blood loss, and length of hospital stay. In addition, retrograde ejaculation was reported in two patients (1.2 %) in the TDR group, and three patients developed deep venous thrombosis after surgery (two TDR and one fusion). Gornet et al. [25] found that the mean operative time was approximately 24 min longer, and a higher blood loss for the TDR group, which were a statistically significant differences, but the average lengths of hospital stay for TDR and fusion patients were similar, and overall adverse event rates showed no statistical difference. Delamarter et al. [24] reported intraoperative data showing that operative times, estimated blood loss, and length of hospital stay were significantly decreased in TDR group. Complications included one dural tear in TDR group (one of 165) and three dural tears in the fusion group (three of 72). One of the two patients in TDR group sustained an iliac artery tear, whereas the other patient in TDR group and the two patients in fusion group had excessive oozing from the decompression, decorticated bone, and graft sites. Postoperatively, deep venous thrombosis was reported in two (1.2 %) of 165 patients in TDR group and two (2.8 %) of 72 in fusion group. According to meta-analysis results, the intraoperative blood loss and operating time in the posterior group showed no significant difference compared to fusion group; apart from this, operating time in the anterior group and complications are significantly reduced compared to the fusion group. Thus, the authors draw a conclusion that safety

in TDR group is better than fusion group at two years follow-up.

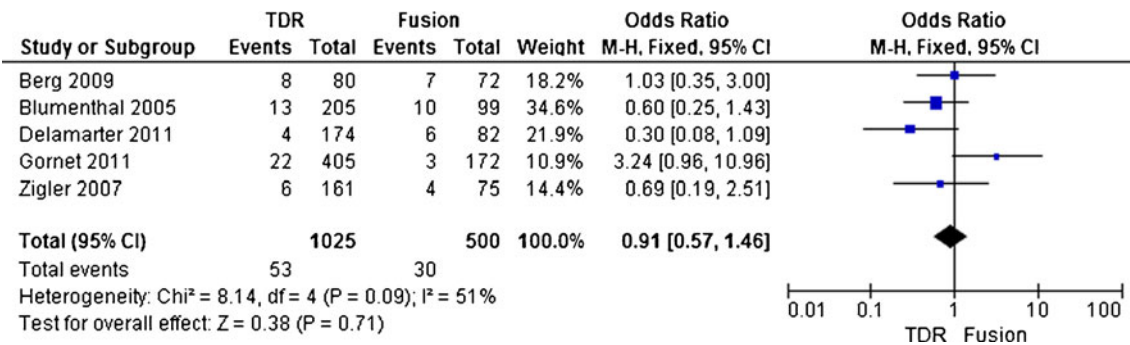
**Efficacy**

Most of the included studies showed clinically relevant improvement from VAS, ODI, proportion of full-time and part-time work, and ROM. In the study of Berg et al. [20], there were no differences in ODI success; and they were satisfied with their treatment and full or part-time between groups at one and two years. He also believed that strictly choosing surgical indications could improve the efficacy of the TDR. Blumenthal et al. [21] reported that both the improvement of ODI and VAS and employment rate had statistically significant differences favoring the TDR patients compared to the fusion group at 24 months. Overall clinical success was significantly higher in TDR group compared with fusion group. In Zigler et al. [23], except for SF-36, improvement in ODI, VAS patient satisfaction, and recreation status had increased, showing a statistically significant difference favoring the TDR patients compared to the fusion group at 24 months. Using the FDA definition, 53.4 % of TDR and 40.8 % of fusion patients were successful, with a statistically significant difference favoring the TDR group. Gornet et al. [25] observed that 24 months follow-up ODI, a mean improvement of low back pain and leg pain, and SF-36 scores after surgery were significantly



**Fig. 9** Results of the meta-analysis for complication for TDR and fusion groups at 2-year





**Fig. 10** Results of the meta-analysis for reoperation rate for TDR and fusion groups at 2-year. Abbreviations: TDR, Artificial total disc replacement; CI, Confidence interval; M-H, Mantel-Haenszel; SD, standard deviation

higher than the values for the fusion group. In addition, the percentage of working participants in the TDR group was slightly higher than that of the fusion group. Overall, a significantly higher percentage of TDR patients showed satisfactory outcomes than fusion patients. Delamarter et al. [24] reported that the mean improvement of ODI from baseline, the SF-36 score, and VAS score for patient satisfaction showed a significant difference in favor of the TDR group at all follow-up time points. The percentage of patients participating in work and recreation status had increased in both groups at 24 months, but there was no significant difference between the TDR group and the fusion group. Wilco C. H. Jacobs et al. [26] conclude that TDR seems to be effective in treating low back pain in selected patients, and in the short term was at least equivalent to fusion surgery. Berg et al. [20], believed that a strict grasp the indications for surgery could enhance effectiveness. In addition, among VAS, ODI, proportion of full-time and part-time work, meta-analysis results showed statistically significant improvement compared to fusion group at 24 months. Therefore, it could be said that the efficacy in the TDR group better than the fusion group.

#### Deficiency

The trials showed TDR had significant safety and efficacy compare to fusion at short-term follow-up, but there was a lack of long-term follow-up, high-quality RCT articles, so evaluation of the efficacy and safety of TDR comparison of fusion for the treatment of LDDD was not sufficient. The current literature reported only long-term follow-up of TDR. Both medium-term follow-up (from five to ten years) [27–30] and long-term follow-up (>10 years) [31, 32] showed significant improvement postoperatively in clinical, and in radiological the motion preserved in surgical segment, lower rate of adjacent segment degeneration. No special complications were noted. In addition, the authors believe that long-term follow-up should focus on potential failure of TDR and adjacent segment degenerative.

The potential failures of disc replacement include early wear, malposition, and prostheses loosening, which are dependent upon numerous factors related to implant design, surgical technique, and patient-specific factors. This needed to be revised. Followed up at two years, Blumenthal et al. [21] and Berg et al. [20] reported four (5 %) and five (2.4 %) patients, respectively, requiring prostheses revision, but mentioned no surgery method. David [31] reported a mean follow-up time of 13.2 years, with eight (7.5 %) patients requiring posterior fusion. In Putzier et al. [32], 6 (11 %) of patients needed reoperation during a follow-up of 17 years. The authors believed an organized approach reduced operative time, minimizes risks, decreases stress, and increases the success rate. At early failure of surgery TDR prostheses are removed and revision to another artificial disc can be considered in order to account for renovated segments without destroying them. Conversely, once renovated sections of lumbar bone structure have been destroyed, lumbar fusion must be performed, including posterolateral fusion, 360° fusion with cages or allograft bone. On the choice of surgical approaches, Pimenta et al. [33] believed anterior revision approaches were associated with significant risks due to scarring and adhesions resulting from the primary procedure, making mobilization of the vessels very difficult, especially at the L4-5 bifurcation; thus primary revision of a failed TDR could be planned as a posterior fusion.

With increased age, lumbar disc degeneration result in lumbar disc space collapse and corresponding stiffening of spinal segments. It is a normal physiological process. There are individual differences in this process; thought to be related to the altered mechanics or loss of motion from the fusion, adjacent segment degeneration was a controversial problem that occurred after a spinal fusion. Motion preservation TDR theoretically may decrease or prevent adjacent segment disease from occurring, but it was difficult to judge postoperative adjacent disc degeneration, whether from physiological processes or TDR reasons. In some literature, follow-up of TDR postoperative more than ten years reported adjacent segment degenerative disease incidence of less than 2.8 %, although longer follow-up was not found.

Despite the lack of comparability, these long-term follow-up results were encouraging.

The literature holds a positive attitude on the clinical and radiological in TDR long-term follow-up. Despite lacking of fusion group comparison, TDR was found to be at least as safe and effective as a treatment of LDDD.

### Limitation

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 [34, 35]. The purpose of this study is to systematically compare the efficacy and the safety of TDR to fusion for the treatment of LDDD. In this meta-analysis, although we identified six RCTs, and five of six trails were regarded as high-quality, the results still have limited application because there exist some problems. First of all was the methodologic quality. In these trials, two studies [23, 25] had no allocation concealment. Some scholars [36] believe that allocation concealment, rather than a perfect test and a non-hidden distribution plan or distribution plan to hide imperfections test, often exaggerated treatment effect of 30–41 %. It may produce measurement bias that patient blinded in two trails [21, 23] and surgeon blinded in three trails [21–23] to the intervention. In addition to this, the literature of Zigler et al. [23] used allocation concealment, patient blinded and surgeon blinded at the same time, and thus was a relatively high-bias risk trial. Overall, their trials were classified as of higher methodological quality. Secondly, results are affected by heterogeneity caused by random sampling. For example, the results of the intra-operative blood loss and operating time presented significant heterogeneity. Third, these literature are adopted to different models of artificial disc and different interventions, and thus there may implement bias. Fourth, most of the retrieved documents were English; there may be language bias. Results of these trials tend to report the superior efficacy of lumbar disc replacement, so the system studies have higher risk of publication bias.

### Conclusion

The results showed the TDR has significant safety and efficacy, comparable to lumbar fusion at two year follow-up. Although the superiority compared to fusion could not be proved, because clinical symptoms were relieved, motion preserved, and a low reoperation rate during long-term follow-up on TDR, TDR was considered safe and effective. Therefore, authors suggest adoption of TDR on a large scale; once there is failure of TDR, interbody fusion could be performed.

**Acknowledgments** Drs Wei, Song, Sun, and Lv have no relevant financial relationships to disclose.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

### References

1. An H, Boden SD, Kang J et al (2003) Summary statement: emerging techniques for treatment of degenerative lumbar disc disease. *Spine* 28:S24–S25
2. 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
3. Glaser J, Stanley M, Sayre H et al (2003) A 10-year follow-up evaluation of lumbar spine fusion with pedicle screw fixation. *Spine* 28(13):1390–1395
4. Chou WY, Hsu CJ, Chang WN, Wong CY et al (2002) Adjacent segment degeneration after lumbar spinal posterolateral fusion with instrumentation in elderly patients. *Arch Orthop Trauma Surg* 122:39–43
5. Okuda S, Iwasaki M, Miyauchi A et al (2004) Risk factors for adjacent segment degeneration after PLIF. *Spine* 29:1535–1540
6. 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
7. Resnick DK, Choudhri TF, Dailey AT et al (2005) Guidelines for the performance of fusion procedures for degenerative disease of the lumbar spine. Part 5: Correlation between radiographic and functional outcome. *J Neurosurg Spine* 2:658–661
8. Sengupta DK, Mulholland RC (2005) Fulcrum assisted soft stabilization system: A new concept in the surgical treatment of degenerative low back pain. *Spine* 30:1019–1029
9. Lemaire JP, Carrier H, Sariali el-H et al (2005) Clinical and radiological outcomes with the Charite' artificial disc: a 10-year minimum follow-up. *J Spinal Disord Tech* 18:353–359
10. 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
11. Szpalski M, Gunzburg R, Mayer M (2002) Spine arthroplasty: A historical review. *Eur Spine J* 11:S65–S84
12. Ingalhalikar AV, Reddy CG et al (2009) Effect of lumbar total disc arthroplasty on the segmental motion and intradiscal pressure at the adjacent level: An in vitro biomechanical study: presented at the 2008 joint spine section meeting laboratory investigation. *J Neurosurg Spine* 11(6):715–723
13. Weisskopf M, Ohnsorge JA, Martini F et al (2008) Influence of inlay height on motion characteristics of lumbar segments in TDR. *Orthop Unfall* 146(4):452–457
14. Guyer RD, McAfee PC, Hochschuler SH et al (2004) Prospective randomized study of the Charite artificial disc: Data from two TDR centers. *Spine J* 4:S252–S259
15. McAfee PC, Fedder IL, Saiedy S et al (2003) SB Charite' disc replacement: Report of 60 prospective randomized cases in a US center. *J Spinal Disord Tech* 16:424–433
16. 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
17. Shea B, Dube C, Moher D (2006) Assessing the quality of reports of systematic reviews: The Quorum statement compared to other

- tools. In: Egger M, Smith GD, Altman DG (eds) *Systematic reviews in health care: meta-analysis in context*. BMJ Publishing Group, London, pp 122–139
18. Higgins JPT, S (2011) *Green “Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. Available from [www.cochrane-handbook.org](http://www.cochrane-handbook.org)*
  19. Furlan AD, Pennick V, Bombardier C et al (2009) 2009 updated method guidelines for systematic reviews in the Cochrane Back Review Group. *Spine (Phila Pa 1976)* 34:1929–1941
  20. Berg S, Tullberg T, Branth B et al (2009) TDR compared to lumbar fusion: a randomised controlled trial with 2-year follow-up. *Eur Spine J* 18:1512–1519
  21. Blumenthal S, McAfee PC, Guyer RD et al (2005) A prospective, randomized, multicenter Food and Drug Administration TDR device exemption study of lumbar TDR with the CHARITE artificial disc versus lumbar fusion: Part I: Evaluation of clinical outcomes. *Spine* 30:1565–1575
  22. Delamarter RB, Bae HW, Pradhan BB (2005) Clinical results of ProDisc-II lumbar TDR: report from the United States clinical trial. *Orthop Clin North Am* 36:301–313
  23. Zigler J, Delamarter R, Spivak JM et al (2007) Results of the prospective, randomized, multicenter Food and Drug Administration TDR device exemption study of the ProDisc-L TDR versus circumferential fusion for the treatment of 1-level degenerative disc disease. *Spine* 32:1155–1162
  24. Delamarter BR, Zigler JE, Balderston RA et al (2011) Prospective, randomized, multicenter Food and Drug Administration TDR device exemption study of the ProDisc-L TDR compared with circumferential fusion for the treatment of two-level lumbar degenerative disc disease. *J Bone Joint Surg Am* 93:705–715
  25. Gornet MF, Burkus JK, Dryer RF et al (2011) Lumbar disc arthroplasty with MAVERICK disc versus stand-alone interbody fusion. *Spine* 36(25):E1600–E1611
  26. Jacobs WCH, van der Gaag NA et al (2013) Total disc replacement for chronic discogenic low back pain: A Cochrane review. *Spine* 38(1):24–36
  27. Guyer RD, McAfee PC, Banco RJ et al (2009) Prospective, randomized, multicenter Food and Drug Administration TDR device exemption study of lumbar TDR with the CHARITE artificial disc versus lumbar fusion: Five-year follow-up. *Spine J* 9:374–386
  28. Park C-K, Ryu K-S, Lee K-Y et al (2012) Clinical outcome of lumbar TDR using ProDisc-L in degenerative disc disease. *Spine* 37(8):672–677
  29. Katsimihis M, Bailey CS, Issa K et al (2010) Prospective clinical and radiographic results of CHARITÉ III artificial total disc arthroplasty at 2- to 7-year follow-up: A Canadian experience. *Can J Surg* 53(6):408–4145
  30. Zigler JE, Glenn J, Delamarter RB (2012) Five-year adjacent-level degenerative changes in patients with single-level disease treated using lumbar TDR with ProDisc-L versus circumferential fusion. *J Neurosurg Spine* 10:19
  31. 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
  32. Putzier M, Funk JF, Schneider SV et al (2006) Charité total disc replacement—clinical and radiographical results after an average follow-up of 17 years. *Eur Spine J* 15(2):183–195
  33. Pimenta L, Díaz RC, Guerrero LG (2006) Charité lumbar artificial disc retrieval: Use of a lateral minimally invasive technique. Technical note. *J Neurosurg Spine* 5(6):556–561
  34. 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
  35. Haer 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
  36. Schulz KF, Chalmers I, Hayes RJ et al (1995) Empirical evidence of bias. Dimensions of methodological quality associated with estimates of effects in controlled trials. *JAMA* 273(5):408–412