RESEARCH ARTICLE

The short-term efficacy and safety of artificial total disc replacement for selected patients with lumbar degenerative disc disease compared with anterior lumbar interbody fusion: A systematic review and meta-analysis

Xiaoping Mu[1ᵒ], Jianxun Wei[1ᵒ], Jiancuo A.[2ᵒ], Zhuhai Li[1], Yufu Ou[1]*

1 Department of Orthopaedic, The People's Hospital of Guangxi Zhuang Autonomous Region, Nanning, China, 2 Department of Spinal surgery, Qinghai Red Cross Hospital, Xining, Qinghai, China

ᵒ These authors contributed equally to this work.
* Ou.YF1972@gmail.com

Abstract

Purpose

To systematically compare the efficacy and safety of lumbar total disc replacement (TDR) with the efficacy and safety of anterior lumbar interbody fusion (ALIF) for the treatment of lumbar degenerative disc disease (LDDD).

Methods

The electronic databases PubMed, Web of Science and the Cochrane Library were searched for the period from the establishment of the databases to March 2018. The peer-reviewed articles that investigate the safety and efficacy of TDR and ALIF were retrieved under the given search terms. Quality assessment must be done independently by two authors according to each item of criterion. The statistical analyses were performed using RevMan (version 5.3) and Stata (version 14.0). The random-effect model was carried out to pool the data. The I^2 statistic was used to evaluate heterogeneity. The sensitivity analysis was carried out to assess the robustness of the results of meta-analyses by omitting the articles one by one.

Results

Six studies (5 randomized controlled trials (RCT) and 1 observational study) involving 1093 patients were included in this meta-analysis. The risk of bias of the studies could be considered as low to moderate. Operative time (MD = 4.95; 95% CI -18.91–28.81; P = 0.68), intraoperative blood loss (MD = 4.95; 95% CI -18.91–28.81; P = 0.68), hospital stay (MD = -0.33; 95% CI, -0.67–0.01; P = 0.05), complications (RR = 0.96; 95% CI 0.91–1.02; P = 0.18) and re-operation rate (RR = 0.54; 95% CI 0.14–2.12; P = 0.38) were without significant clinical difference between groups. Patients in the TDR group had higher postoperative

satisfaction (RR = 1.19; 95% CI 1.07–1.32; P = 0.001) and, better improvements in ODI (MD = -10.99; 95% CI -21.50- -0.48; P = 0.04), VAS (MD = -10.56; 95% CI -19.99- -1.13; P = 0.03) and postoperative lumbar mobility than did patients in the ALIF group.

Conclusions

The results showed that TDR has significant superiority in term of reduced clinical symptoms, improved physical function and preserved range of motion for the treatment of LDDD compared to ALIF. TDR may be an ideal alternative for the selected patients with LDDD in the short-term. However, the results of this study cannot suggest the use of TDR instead of ALIF in lumbar spine treatment only in the light of short term results. More studies that are well-designed, that are of high-quality and that have larger samples are needed to further evaluate the efficacy and safety of TDR with at the long-term follow-up.

Level of evidence

Therapeutic Level 3

Introduction

Chronic low back pain (CLBP) is generally defined as lower back pain that persists for at least 3 months or 12 weeks [1]. More than 80% of all individuals experienced at least 1 episode of LBP at some point in their lifetime [2]. Obviously, the impact of LBP is substantial, not only on the individual, but also on communities and health systems [3,4]. Current research is still not clear about the causes of CLBP, but lumbar disc degenerative disease is considered to be closely related to CLBP.

For the patients seeking medical intervention for CLBP, lumbar operation will be considered when conservative treatments are ineffective for 6 months. ALIF, which was first reported by Capener in 1932, has developed into a mature and popular operative method for the treatment of LDDD. However, as an important part of many different fusion approaches, ALIF is not perfect. It not only has some complications similar to other fusion approaches, such as adjacent segment degeneration [5] and postoperative fusion pain, but is also associated with the risk of vascular, intestinal and nerve injury [6]. Although those complications associated with the approach can be decreased through improving techniques and continuous training, the restriction in the range of motion (ROM) of the fusion segment generally cannot be changed.

As an alternative to lumbar fusion, artificial lumbar total disc replacement (TDR), which is continuously applied in clinical practice, can not only eliminate the adverse effects on spinal stability after discectomy by restoring and preserving normal inter-segmental motion [7,8], but also avoid the non-physiological load of adjacent segments to slow down degeneration [9]. Biomechanical study has already demonstrated that TDR can maintain the three-dimensional motion and restore the biomechanical properties of lumbar vertebrae [10]. Relevant clinical studies have been reported, but it is not clear which one is better or worse in the efficacy of TDR and ALIF.

Scholars [11–14] have carried out several meta-analysis to survey the safety and efficacy of TDR for patients with LDDD compared with lumbar fusion. However, the application of different fusion approaches in various studies may lead to certain bias and reduce the level of

evidence after pooling data for the same outcome. ALIF and TDR have similar anatomical approaches, and there are several clinical studies comparing TDR and ALIF in the treatment of LDDD. Therefore, we believe that it is necessary to perform a detailed stratification of lumbar fusion in a new meta-analysis on comparison of TDR and ALIF.

Materials and methods

This systematic review and meta-analysis was performed according to the Preferred Reporting Items of systematic review and meta-analysis (PRISMA) items [15]. (S1 File)

Search strategy

Two authors (M.XP and W.JX) who have been educated in literature retrieval courses independently searched electronic databases including PubMed, Web of Science and the Cochrane Library for the period from the establishment of the databases to March 2018. The following search terms were used: ("total disc replacement" OR "lumbar disc arthroplasty" OR "artificial disc replacement") AND ("anterior lumbar interbody fusion" OR "anterior fusion" OR "ALIF"). Due to the limitations of the authors' own language, the studies written in other languages other than English and Chinese, were not considered for the inclusion in the present study. The list of references of the relevant review and the included studies were also further hand-checked one by one to identify studies that had not been retrieved in the preliminary database search.

Selection criteria

The eligibility criteria in the present study included the following: (1) study design: randomized control trials and observational studies; (2) case population: adult patients with LDDD underwent lumbar TDR or ALIF; (3) intervention methods: TDR (investigative group) versus ALIF (control group); and (4) outcome measures: study containing at least one of the desired evaluation indicators for this meta-analysis. Exclusion criteria were formulated as follows: (1) a study containing patients with the history of lumbar surgery prior to TDR/ALIF; (2) a case report, animal experiment, or biomechanical research paper; and (3) a study reporting patients who underwent combinations of both of the interventions (TDR or ALIF) and/or one of the interventions and other surgical procedures.

Data extraction

Two authors (M.XP and O.YF) independently extracted the data in accordance with the established criteria and filled in the standardized form immediately. The following information was extracted from every study: (1) study characteristics: authors, publication year, study design, and number of patients each group; (2) surgical data: operative time, estimated blood loss, and duration of hospitalization; (3) curative effect evaluation: VAS scores, ODI scores, and patient satisfaction; (4) radiological parameter: range of motion (ROM) of each lumbar segment; and (5) postoperative complications and re-operation: infections, approach-related events, and neurological events, among other complications.

Risk of bias assessment and quality of evidence

Three reviewers (M.XP, L.ZH and W.JX) independently used the bias risk assessment tool provided by Cochrane back review group [16] for RCTs and the Newcastle-Ottawa Quality Assessment Scale (NOQAS) [17] for cohort and case-controlled studies to evaluate the quality of each study. Each RCT was re-examined and "yes", "no" or "unsure" were answered

provided for in the following items: (1) random sequence generation, allocation concealment and baseline similarity for the important prognostic indicators (selection bias); (2) blinding of patients and care provider, co-interventions (performance bias); (3) acceptable drop-out rate and analysis of randomized participants (attrition bias); (4) blinding of outcome assessor and identical timing of outcome assessment (measurement bias); (5) suggestion of selective outcome reporting (reporting bias); and (6) other: other sources of potential bias. NOQAS was composed of nine items in the following three categories: selection of the study population, comparability among groups and outcome evaluation for cohort studies or exposure for case-controlled studies. Out of a total score of 9 points, studies with less than 5 points were considered as low quality studies; studies with 5 points or greater were rated as high quality studies. The Grading of Recommendations Assessment, Development and Evaluation (GRADE) system [18] was used by two reviewers (M.XP and O.YF) to assess the confidence in effect estimates. The quality of evidence was considered as high, moderate, low or very low involving the following domains: risk of bias, inconsistency, indirectness, imprecision, and publication bias. When disputes between the reviewers regarding results could not be resolved through their internal negotiation, new reviewers would be added.

Date analysis and interpretation

An open and free statistic software, RevMan 5.3 version and Stata 14.0, were downloaded and synthesized for all of available data in these relevant studies. Meta-analysis and forest plots were expected to construct. Mean difference (MD) was used for continuous outcomes with identical scales. Otherwise, we used standardized mean difference (SMD). The relative risks (RRs) were carried out to analysis for binary outcomes. the corresponding 95% confidence interval (95% CI) was provided for each outcome. The random-effect (RE) model was carried out to pool the data, as this model is more conservative and provides better estimates with wider confidence intervals than the fixed-effects model. The I^2 statistic was used to evaluate heterogeneity; values of 25%, 50%, and 75% were considered low, moderate, and high heterogeneity, respectively [19]. The presence of heterogeneity warrants examining the sources where we used covariates in a meta-regression analysis. In this analysis, the covariates were used were the following: sample size (<200 cases vs. >200 cases), study design (prospective vs. retrospective), and follow-up points (<24 months vs. >24 months). The publication bias was not evaluated, as a small meta-analysis normally under-powered to detect much bias and this analysis tends to lead to conclusions that are not justified. The sensitivity analysis was carried out to assess the robustness of the results of meta-analyses by omitting the articles one by one. Under the same evaluation index, an inappropriate data format from the included studieswas not pooled with this software. But, to avoid missing meaningful results and to minimize the bias of the report, these data were still displayed reported in the results section by descriptive text. A double-sided p value less than 0.05 was considered as a significant difference.

Results

Literature search and study characteristic

The flow chart based on the PRISMA Statement [20] is shown in Fig 1. A total of 1641 records were preliminarily identified under our search strategy. Of these, 564 duplicated articles were eliminated. After screening the titles and abstracts, 39 potentially eligible studies were required to undergo the full-text analysis. Finally, 6 studies [21–26] that met all of the selective criteria were included in the current analysis. It's worth pointing out that two [21,22] of the 6 articles were from a randomized, multicenter US Food and Drug Administration Investigational
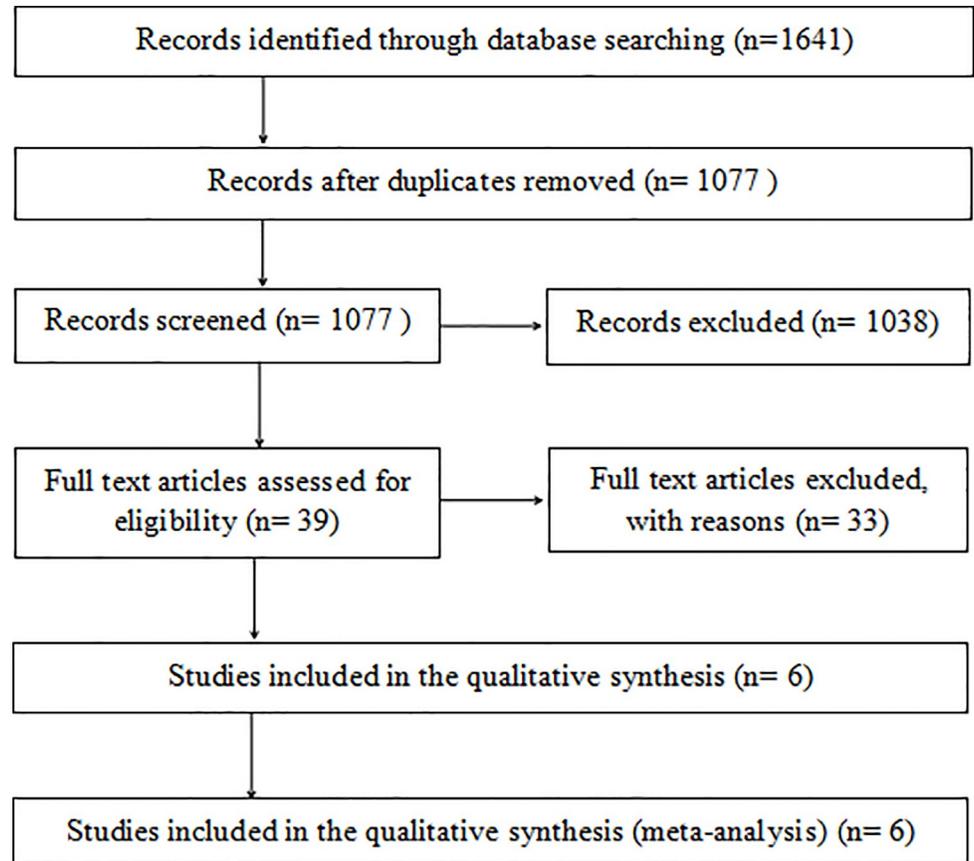
Fig 1. The flow chart of systematic review based on PRISMA statement.

https://doi.org/10.1371/journal.pone.0209660.g001

Device Exemption Study. Following in-depth discussion at the reviewers' meeting, the consistent conclusion was that we could not rule out either of these 2 articles, because the evaluation indicators used in each article were different from those in the other article. Also, each of those indicators was exactly what our meta-analysis needed.

Characteristics of the included studies are summarized in Table 1. These 6 articles were published from 2005 to 2017 and included a total of 1093 patients, with sample numbers ranging from 48 to 577. Of these studies, 5 studies were RCTs [21–25] and 1 article [26] was an observational study. The mean age of patients ranged from 39.6 to 48.4. The follow-up time in the investigative and control groups ranged from 12 months to 60 months.

Results of risk of bias assessment

According to the criteria of the Cochrane review group, the risk of bias of the included RCTs were rated as low to moderate, and the results are summarized in Table 2. Negative answers to the blinding could be found in almost all included studies. In fact, blinding of patients and care providers is often not feasible in the field of surgery. So, given the nature of the surgical research, we decided that we should not pay much more attention to the impact of blinding on the quality of the study. Moreover, only one observational study received a high quality score of 6 points (selection of cohorts: 2 scores, comparability of cohorts: 2 scores and assessment of outcome: 2 scores) based on NOQAS.

Table 1. Summary of characteristic of included studies.

Study	Design	No. of patients		Mean age		Gender		Follow up	Inclusion criteria	Outcome
		TDR	ALIF	TDR	ALIF	TDR	ALIF			
Blumenthal, 2005	RCT	205	99	39.6	39.6	M:113 F:92	M:44 F:55	24m	Symptomatic degenerative disc disease; single level at L4–L5 or L5-S1	Operative time, blood loss, hospital stay, patient satisfaction
Holt, 2007	RCT	205	99	39.6	39.6	M:113 F:92	M:44 F:55	24m	Symptomatic degenerative disc disease; single level at L4–L5 or L5-S1	Total adverse events, re-operation
Geisler, 2009	RCT	53	31	41.0	40.2	M:24 F:29	M:18 F:13	60m	Single level, symptomatic degenerative disc disease involvement L4–S1; TDR or ALIF no prior surgery	Operative time, blood loss, patient satisfaction,VAS, ODI
Gornet, 2011	RCT	405	172	39.9	40.2	M:205 F:200	M:86 F:86	24m	Degenerative disc disease; single-level, symptomatic involvement L4–S1	Operative time, blood loss, hospital stay, patient satisfaction,VAS, ODI, total adverse events, re-operation
Strube, 2016	RCT	23	25	47.3	48.4	M:10 F:13	M:10 F:15	12m	Degenerative disc disease; single level, symptomatic involvement L4–S1	ROM, VAS score, ODI score
Mattei, 2017	OS	30	50	40.5	45.3	M:15 F:15	M:25 F:25	12m	Degenerative disc disease; single-level symptomatic occurred at L4/L5 or L5/S1	Blood loss, VAS score, ODI score

OS = observational study; M = male; F = female; m = months; L = lumbar; S = sacral.

Surgical data (TDR *versus* ALIF)

Three RCTs [21,23,24] with sufficient data reported the operative time and duration of hospitalization of patients who were underwent TDR or ALIF. The pooled results shown that there was no significant difference in the operative time between TDR and ALIF groups [(Operative time: MD, 4.95; 95% CI, -18.91 to 28.81; P = 0.68; Fig 2); Duration of hospitality: MD, -0.33; 95% CI, -0.67 to 0.01; P = 0.05; Fig 3)]. Three RCTs [21,23,24] and one observational study [25] including a total of 1045 patients had the sufficient data on estimated blood loss. The synthetic results revealed that TDR did not show any difference from ALIF (MD, 60.77; 95% CI, -21.85 to 143.40; P = 0.15; Fig 4).

Radiographic outcome (TDR *versus* ALIF)

Range of motion (ROM), as a good evaluation indicator after fusion had been reported in two studies [24,25]. Meta-analysis could not be performed due to insufficient data. Gornet et al [23] reported that the mean preoperative segmental ROM was 7˚, and it increased at 12 and 24 months follow-up (9.4˚and 9.5˚, respectively). In the ALIF group, the mean ROM was less than 0.6˚at any time after operation. However, a significant difference in postoperative lumbar mobility was observed in patients undergoing TDR (median ROM = 11.4˚, interquartile range 4.95˚), compared with that of patients in the ALIF group (median ROM = 0.4˚, interquartile range 1.30˚) [25].

Clinical outcome (TDR *versus* ALIF)

In our meta-analysis, VAS, ODI score and postoperative patient satisfaction were considered as a category to assess the clinical curative efficacy. Each of the four clinical evaluation indicators was reported by three studies. The pooled results signified that there were significant differences in the VAS score at the final follow-up (MD, -10.56; 95% CI, -19.99 to -1.13; P = 0.03; Fig 5), in the ODI score at the final follow-up (MD, -10.99; 95% CI, -21.50 to -0.48; P = 0.04; Fig 6), and in the postoperative patient satisfaction (RR, 1.20; 95% CI, 1.11 to 1.30; P<0.00001; Fig 7).

Table 2. Risk of bias assessment of included RCTs.

Type of bias	Criteria	Answers				
		Blumenthal, 2005	Holt, 2007	Geisler, 2009	Gornet, 2011	Strube, 2016
Selection	Was the method of randomization adequate?	Yes	Yes	Unsure	Yes	Yes
	Was the treatment allocation concealed?	Yes	Yes	Unsure	Yes	Yes
	Were the groups similar at baseline regarding the most important prognostic indicators?	Yes	Yes	Yes	Yes	Yes
Performance	Was the patient blinded to the intervention?	No	No	No	No	Unsure
	Was the care provider blinded to the intervention?	No	No	No	No	Unsure
	Were co-interventions avoided or similar?	Yes	Yes	Yes	Yes	Yes
Attrition	Was the drop-out rate described and acceptable?	Yes	Unsure	Yes	Yes	Yes
	Were all randomized participants analyzed in the group to which they were allocated?	Yes	Yes	Yes	Yes	Yes
Measurement	Was the outcome assessor blinded to the intervention?	No	No	No	No	Unsure
	Was the timing of the outcome assessment similar in all groups?	Yes	Unsure	Yes	Yes	Yes
Reporting	Are reports of the study free of suggestion of selective outcome reporting?	No	No	Yes	Yes	Yes
Other	Are other sources of potential bias unlikely?	Unsure	Unsure	Unsure	Unsure	Unsure

Complications and second surgery (TDR *versus* ALIF)

Two RCTs [22,24] reported data on total adverse events and the second surgery. The pooled results reflected that there were no significant difference in the complications occurrence (RR, 0.96; 95% CI, 0.90 to 1.02; P = 0.22; Fig 8) or in the re-operation rates (RR, 0.54; 95% CI, 0.14 to 2.12; P = 0.38; Fig 9) between the two groups.

Results of the GREAD system assessment

Based on the GREAD system, confidence in the estimates were high for patient satisfaction, moderate for operative time, duration of hospitalization and total adverse events, and low for blood loss, VAS, ODI score and re-operation (Table 3).

Heterogeneity, sensitivity analysis

Meta-regression analyses were carried out in accordance with some covariates, including sample size, study design, and follow-up points; however, meta-regression outcomes did not detect the sources of heterogeneity (Table 4). A sensitivity analysis was performed by omitting one study (Blumenthal et al [21]). There was a significant difference between groups (MD = 0.42; 95% CI, 0.17–0.66)., which means that the results cannot be considered robust.

Study or Subgroup	TDR (operative time) Mean	SD	Total	ALIF (operative time) Mean	SD	Total	Weight	Mean Difference IV, Random, 95% CI
Blumenthal et al, 2005	110.8	47.7	205	114	67.9	99	34.7%	-3.20 [-18.08, 11.68]
Geisler et al, 2009	115.6	53.69	53	128.2	66.23	31	26.2%	-12.60 [-40.03, 14.83]
Gornet et al, 2011	108	36	405	84	36	172	39.0%	24.00 [17.58, 30.42]
Total (95% CI)			663			302	100.0%	4.95 [-18.91, 28.81]

Heterogeneity: Tau² = 369.08; Chi² = 15.92, df = 2 (P = 0.0003); I² = 87%
Test for overall effect: Z = 0.41 (P = 0.68)

Fig 2. Forest plot for operative time. df = degrees of freedom, and IV = Inverse Variance.

Study or Subgroup	TDR (hospital stay)			ALIF (hospital stay)			Weight	Mean Difference IV, Random, 95% CI
	Mean	SD	Total	Mean	SD	Total		
Blumenthal et al, 2005	3.7	1.18	205	4.2	1.99	99	30.3%	-0.50 [-0.92, -0.08]
Geisler et al, 2009	3.7	0.86	53	4.3	1.47	31	22.0%	-0.60 [-1.17, -0.03]
Gornet et al, 2011	2.2	1.3	405	2.3	1.1	172	47.7%	-0.10 [-0.31, 0.11]
Total (95% CI)			663			302	100.0%	-0.33 [-0.67, 0.01]

Heterogeneity: Tau² = 0.05; Chi² = 4.69, df = 2 (P = 0.10); I² = 57%
Test for overall effect: Z = 1.92 (P = 0.05)
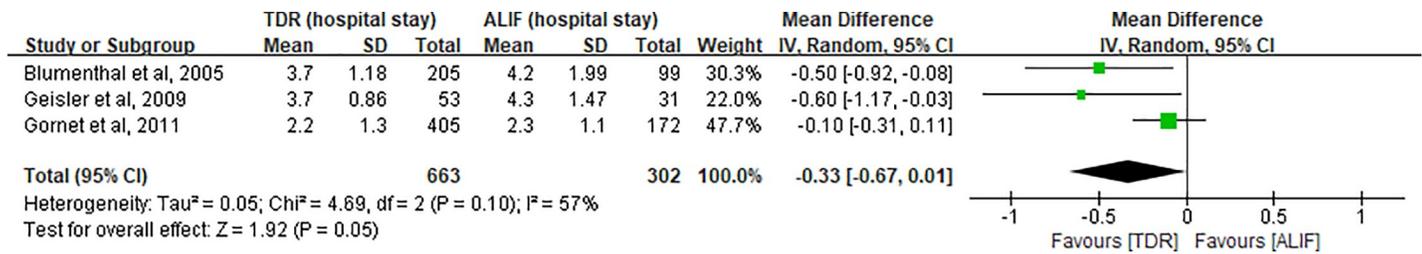
Fig 3. Forest plot for duration of hospitalization. df = degrees of freedom, and IV = Inverse Variance.

https://doi.org/10.1371/journal.pone.0209660.g003

Discussion

Main findings

Five RCTs and one observational study involving 1093 cases were included in this meta-analysis. In the present study, we found that patients who underwent TDR had similar surgical data (operative time, intraoperative blood loss and hospital stay) and similar risk of postoperative adverse events compared with patients who underwent ALIF in the short term. However, the statistical differences in the results of comparison of clinical efficacy showed that patients who underwent TDR could achievebetter symptom relief and that their physical function improved.

The results of the prior relevant meta-analysis

Several previous meta-analysis [11–13] comparing TDR to fusion for the treatment of lumbar disc degenerative disease have been published. A meta-analysis with 5 RCTs published in 2010 [11], reported that a significant statistical difference was only found in the postoperative patients' satisfaction rate between both groups at 2-year follow-up. Although TDR revealed slight advantages in better functioning and pain remission, the pooled effects of multiple studies were not statistically significant. Importantly, TDR did not show significant superiority in the clinical outcomes at 5-year follow-up. Subsequently, two meta-analysis that included 6 and 7 RCTs under the same thesis were published in different peer-review journals by Wei et al [12] and Rao et al [13], respectively. Wei et al [11] reported that, the safety and efficacy of TDR during the 2-year follow-up were significantly better than that of lumbar fusion, but it was still not considered that TDR is superior to lumbar fusion. However, TDR was demonstrated significant superiorities in improved physical function, reduced pain and shortened duration of hospitalization in the Rao and colleagues' manuscript [13].

Inconsistent findings prompted Ding et al [27] to conduct a systematic review of overlapping meta-analysis. After comparing 5 meta-analysis with the same topic, a paper by Jacobas
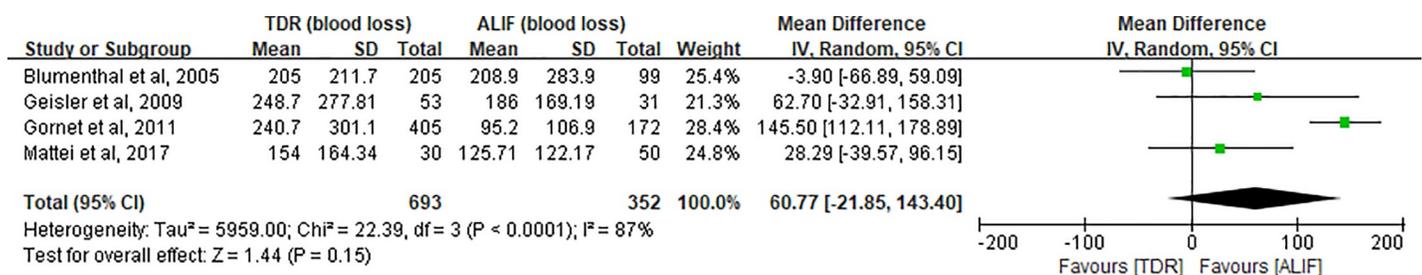
Study or Subgroup	TDR (blood loss)			ALIF (blood loss)			Weight	Mean Difference IV, Random, 95% CI
	Mean	SD	Total	Mean	SD	Total		
Blumenthal et al, 2005	205	211.7	205	208.9	283.9	99	25.4%	-3.90 [-66.89, 59.09]
Geisler et al, 2009	248.7	277.81	53	186	169.19	31	21.3%	62.70 [-32.91, 158.31]
Gornet et al, 2011	240.7	301.1	405	95.2	106.9	172	28.4%	145.50 [112.11, 178.89]
Mattei et al, 2017	154	164.34	30	125.71	122.17	50	24.8%	28.29 [-39.57, 96.15]
Total (95% CI)			693			352	100.0%	60.77 [-21.85, 143.40]

Heterogeneity: Tau² = 5959.00; Chi² = 22.39, df = 3 (P < 0.0001); I² = 87%
Test for overall effect: Z = 1.44 (P = 0.15)

Fig 4. Forest plot for estimated blood loss. df = degrees of freedom, and IV = Inverse Variance.

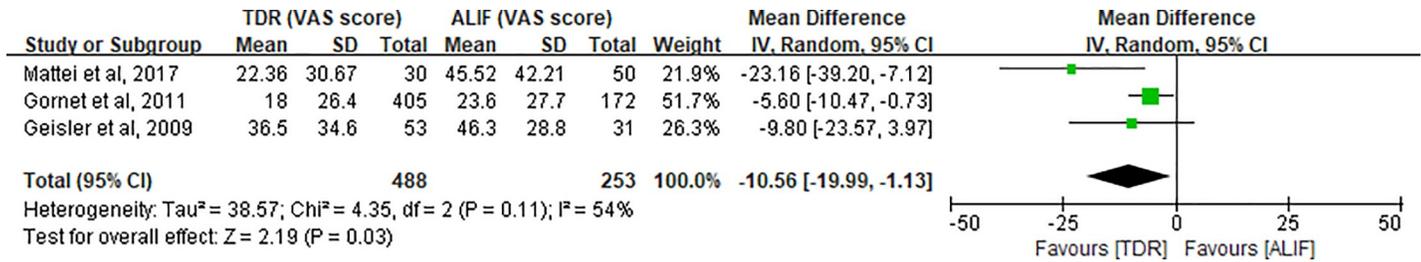https://doi.org/10.1371/journal.pone.0209660.g004

Fig 5. Forest plot for VAS score at the final follow-up. df = degrees of freedom, and IV = Inverse Variance.

et al [14] was rated to provide the best available evidence. They demonstrated that TDR was superior to lumbar fusion in patient satisfaction, ODI, VAS, pain, implant motion and subsidence [14]. Finally, the cautious conclusions that TDR was at least as safe and effective as lumbar fusion in the short term were presented in Ding and coleagues' study [27]. However, the common defect in these meta-analysis that affected their findings is that different approaches of lumbar fusion for LDDD would result in different outcomes [28]. Circumferential fusion was the most used fusion approach in these meta-analysis. So, the results were largely influenced by this approach. In order to minimize the impact of confounding factors on the level of evidence, it is necessary to conduct a meta-analysis on the comparison of TDR and a specific surgical approach, such as ALIF, PLIF.

The efficacy and safety of TDR

The efficacy of operation should be based primarily on patients' safety. In the present study, surgical time, intraoperative blood, complications and re-operation rate were used to evaluate operative safety. The results were consistent with other studies [29]. Currently, most researchers believe that the early clinical efficacy of TDR is positive [30,31] on the premise of grasping strictly the operative indication. A variety of evaluation indicators were comprehensively analyzed in our study and the conclusion could be obtained that the efficacy of TDR is indeed better than that of ALIF for LDDD. As a new surgical technique, there were few mid- or long-term follow-up studies. Guyer et al [32] reported no statistical differences were found in clinical outcomes between TDR and ALIF, but patients who underwent TDR reached a statistically greater rate of part- and full-time employment and a statistically lower rate of long-term disability. Also, Zigler et al [33] believed that patients in the TDR group had significantly better improvement on some scales and were more satisfied about avoiding the stiffness of fusion than patients in the lumbar fusion group during the 5-year follow-up. Similar results were reported in the more than 15-year follow-up retrospective study [34].
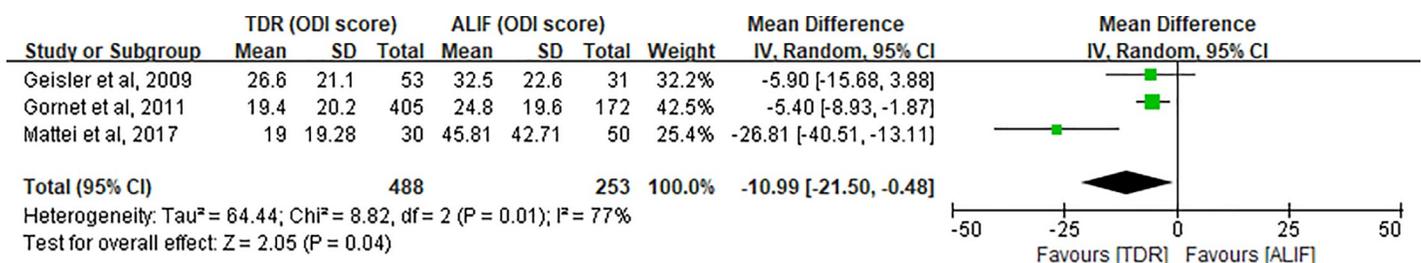

Fig 6. Forest plot for ODI score at the final follow-up. df = degrees of freedom, and IV = Inverse Variance.
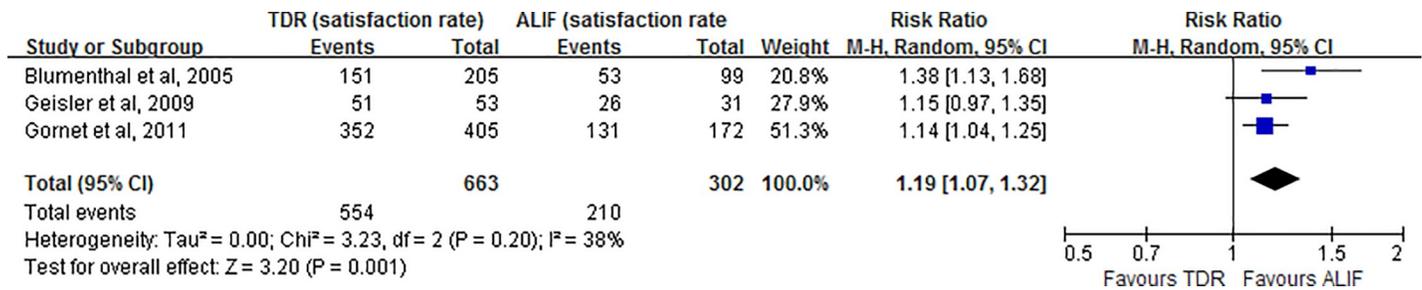
Study or Subgroup	TDR (satisfaction rate)		ALIF (satisfaction rate		Weight	Risk Ratio M-H, Random, 95% CI	Risk Ratio M-H, Random, 95% CI
	Events	Total	Events	Total			
Blumenthal et al, 2005	151	205	53	99	20.8%	1.38 [1.13, 1.68]	
Geisler et al, 2009	51	53	26	31	27.9%	1.15 [0.97, 1.35]	
Gornet et al, 2011	352	405	131	172	51.3%	1.14 [1.04, 1.25]	
Total (95% CI)		663		302	100.0%	1.19 [1.07, 1.32]	
Total events	554		210				

Heterogeneity: Tau² = 0.00; Chi² = 3.23, df = 2 (P = 0.20); I² = 38%
Test for overall effect: Z = 3.20 (P = 0.001)

0.5   0.7   1   1.5   2
Favours TDR   Favours ALIF

Fig 7. Forest plot for postoperative patient satisfaction. df = degrees of freedom, and M-H = Mantel-Haenszel.

https://doi.org/10.1371/journal.pone.0209660.g007

The retention of lumbar segmental motion is considered to be the greatest advantage of TDR and the original intention of designing the artificial disc device. However, since the included studies didn't provide sufficient data about ROM, we could not evaluate the studies by quantitative analysis. Descriptive qualitative analysis in our study showed that ROM after TDR was significantly better than ALIF. And whether it is with anyone of different fusion approaches or different artificial disc devices [35], the advantage of TDR in preserving postoperative ROM can be seen.

Strengths and limitations

To our knowledge, this is the first meta-analysis of published clinical studies comparing the safety and efficacy of TDR and ALIF in patients with LDDD. Further, in our study, the application of the confidence in effect estimates can provide more institutional guidance for clinical decision-maker according to the GREAD system. However, this methodology has not been found in the previous studies with similar theses. In addition, under the strict methodology, significant superiorities of TDR in improved physical function, reduced pain and preserved range of motion were reported compared with ALIF in this meta-analysis. It is our hope that these results may be used to inform the clinical management of lumbar DDD, a condition that substantially influences quality of life in older adults.

However, this study was restricted by several limitations that require the cautious interpretation of the results. First, our meta-analysis that mainly focused on to investigating the safety and efficacy of TDR and ALIF for patients with LDDD had already excluded the influence of other fusion approaches on the results. However, the types of artificial intervertebral discs in the TDR group could not be stratified because of the limitation of relevant studies, so there may be an implementation bias. Moreover, although RCTs were the predominant studies included in this meta-analysis, blinding was not applied in these included studies, which may result in measurement bias. In addition, publication bias that can also affect the evidence level
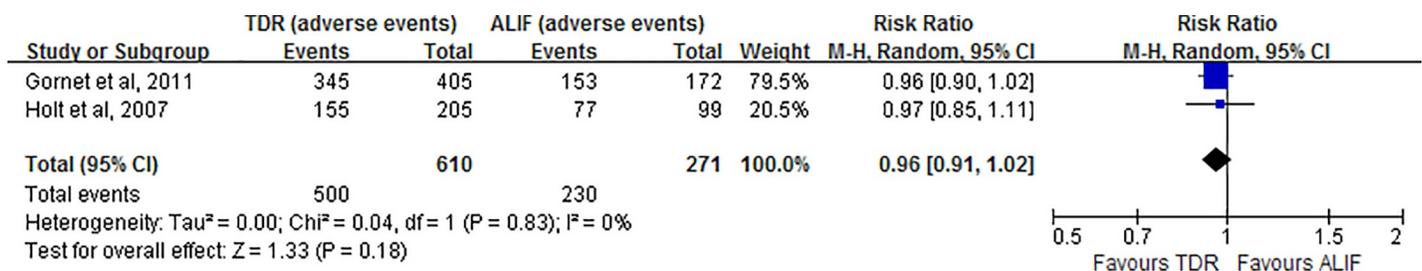

Study or Subgroup	TDR (adverse events)		ALIF (adverse events)		Weight	Risk Ratio M-H, Random, 95% CI	Risk Ratio M-H, Random, 95% CI
	Events	Total	Events	Total			
Gornet et al, 2011	345	405	153	172	79.5%	0.96 [0.90, 1.02]	
Holt et al, 2007	155	205	77	99	20.5%	0.97 [0.85, 1.11]	
Total (95% CI)		610		271	100.0%	0.96 [0.91, 1.02]	
Total events	500		230				

Heterogeneity: Tau² = 0.00; Chi² = 0.04, df = 1 (P = 0.83); I² = 0%
Test for overall effect: Z = 1.33 (P = 0.18)

0.5   0.7   1   1.5   2
Favours TDR   Favours ALIF

Fig 8. Forest plot for total adverse events. df = degrees of freedom, and M-H = Mantel-Haenszel.
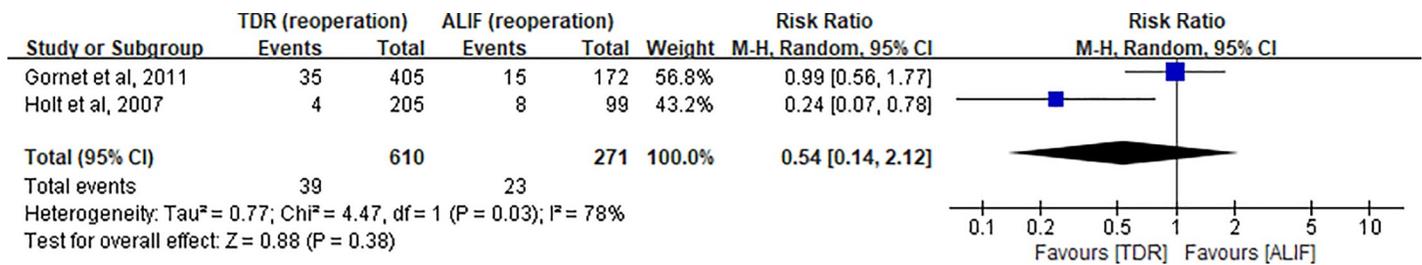
https://doi.org/10.1371/journal.pone.0209660.g008

Fig 9. Forest plot for the number of re-operation. df = degrees of freedom, and M-H = Mantel-Haenszel.

of conclusions couldn't be well detected in the small meta-analysis with less than 10 included studies, because only publicly published studies were included in this meta-analysis and studies that reported the contrary finding may be difficult to publish. Also Additionally, language bias was found in our study. Finally, the result of the sensitivity analysis forced us to cautiously interpret the results of the study only.

Given that significant heterogeneity was present in the current study, we used the meta-regression analysis to test the factors that may be the sources of heterogeneity. Unfortunately, these specific covariates have proven to be not the factors influencing the safety and effectiveness of TDR. The cause for this finding has yet to be determined.

Conclusions

In summary, this meta-analysis based on the current available studies shows that the efficacy of TDR is superior to that of ALIF in the short term. TDR may be an ideal alternative for

Table 3. Quality of evidence for outcome based on GREAD system.

Outcomes	No. of patients and trials	Study design (No.)	Risk of Bias	Inconsistency	Indirectness	Imprecision	Publication bias	Estimate of Effect (95% CI)	Confidence in Effect Estimates (GRADE)
Patient satisfaction	764 (3)	RCT (3)	No Serious	Not Serious	Not Serious	Not Serious	None detected	1.20 (1.11–1.30)	High ⊕⊕⊕⊕
Operative time	965 (3)	RCT (3)	No Serious	Not Serious	Not Serious	Serious	None detected	4.95 (-18.91–28.81)	MODERATE ⊕⊕⊕O
Duration of hospitalization	965 (3)	RCT (3)	No Serious	Not Serious	Not Serious	Serious	None detected	-0.33 (-0.67–0.01)	MODERATE ⊕⊕⊕O
Total adverse events	881 (2)	RCT (2)	Serious*	Not Serious	Not Serious	Not Serious	None detected	0.96 (0.90–1.02)	MODERATE ⊕⊕⊕O
Blood loss	1045 (4)	RCT (3) CC (1)	No Serious	Serious†	Not Serious	Serious	None detected	60.77 (-21.85–143.4)	LOW ⊕⊕OO
VAS score	741 (3)	RCT (2) CC (1)	No Serious	Serious†	Not Serious	Serious	None detected	-10.56 (-19.99—-1.13)	LOW ⊕⊕OO
ODI score	741 (3)	RCT (2) CC (1)	No Serious	Serious†	Not Serious	Serious	None detected	-10.99 (-21.50—-0.48)	LOW ⊕⊕OO
Reoperation	881 (2)	RCT (3)	Serious*	Not Serious	Not Serious	Serious	None detected	0.54 (0.14–2.12)	LOW ⊕⊕OO

*Selective outcome reporting among included studies.

†Different pooled estimates of effect between RCT and case-controlled study.

Wide 95% CIs.

CC = case-controlled study; CI = confidence interval.

Table 4. Meta-regression analysis of potential sources of heterogeneity.

Factors	Coefficient	Standard error	P-value	95% confidence intervals
Sample size	0.544	0.606	0.435	-1.384, 2.472
Study design	1.094	0.362	0.802	-0.382, 3.135
Follow-up points	0.091	0.331	0.802	-0.962, 1.143

selected patients with LDDD in the short term. However, the results of this study cannot suggest the use of TDR above the use of ALIF for lumbar spinal treatments only on the basis of short term results. Furthermore, we think that this study still has a certain clinical significance, although the limitations of this meta-analysis require us to be cautious about the present conclusions. Multicenter, well-designed, high-quality, large sample and long-term follow-up studies are needed to further evaluate the short- and long-term safety and efficacy of TDR comparison of ALIF or other fusion approaches in the treatment of LDDD.

Supporting information

S1 File. PRISMA 2009. Checklist.
(DOC)

S2 File. Search strategy.
(DOCX)

Acknowledgments

Author Contributions

Conceptualization: Xiaoping Mu, Jiancuo A., Yufu Ou.

Data curation: Xiaoping Mu, Jianxun Wei, Zhuhai Li.

Formal analysis: Jianxun Wei, Jiancuo A., Yufu Ou.

Methodology: Xiaoping Mu, Jianxun Wei, Jiancuo A., Zhuhai Li, Yufu Ou.

Project administration: Yufu Ou.

Resources: Yufu Ou.

Software: Xiaoping Mu, Zhuhai Li.

Supervision: Yufu Ou.

Writing – original draft: Xiaoping Mu, Jianxun Wei.

Writing – review & editing: Jiancuo A., Yufu Ou.

References

1. Airaksinen O, Brox JI, Cedraschi C, Hildebrandt J, Klaber-Moffett J, Kovacs F, et al. Chapter 4. European guidelines for the management of chronic nonspecific low back pain. Eur Spine J. 2006; 15(Suppl 2): S192–300.

2. Cassidy JD, Carroll LJ, Côté P. The Saskatchewan Health and Back Pain Survey: The Prevalence of low back Pain and Related Disability in Saskatchewan Adults. Spine (Phila Pa 1976). 1998; 23 (17):1860–7.

3. Deyo RA, Weinstein JN. Low back pain. N Engl J Med. 2001; 344: 363–70. https://doi.org/10.1056/NEJM200102013440508 PMID: 11172169

4. Golob AL, Wipf JE. Low Back Pain. Med Clin North Am.2014; 98: 405–28. https://doi.org/10.1016/j.mcna.2014.01.003 PMID: 24758954

5. Glaser J, Stanley M, Sayre H, Woody J, Found E, Spratt K. A 10-year follow-up evaluation of lumbar spine fusion with pedicle screw fixation. Spine (Phila Pa 1976). 2003; 28(13): 1390–5.

6. Snyder LA, Kalb S, Kakarla UK, Porter RW, Kaibara T, Dickman CA, et al. Effects of anterior plating on clinical outcomes of anterior lumbar interbody fusion. Clin Spine Surg. 2016; 29(7):300–4. https://doi.org/10.1097/BSD.0b013e31827f5642 PMID: 23222098

7. Booth KC, Bridwell KH, Eisenberg BA, Baldus CR, Lenke LG. Minimum 5-year results of degenerative spondylolisthesis treated with decompression and instrumented posterior fusion. Spine (Phila Pa 1976). 1991; 24(16):1721–7.

8. Ingalhalikar AV, Reddy CG, Lim TH, Torner JC, Hitchon PW. 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. 2009; 11(6):715–23. https://doi.org/10.3171/2009.7.SPINE094 PMID: 19951025

9. Siepe CJ, Mayer HM, Wiechert K, Korge A. Clinical results of total lumbar disc replacement with Pro-Disc II: three-year results for different indications. Spine (Phila Pa 1976). 2006; 31(17):1923–32.

10. Dooris AP, Goel VK, Grosland NM, Gilbertson LG, Wilder DG. Load-sharing between anterior and posterior elements in a lumbar motion segment implanted with an artificial disc. Spine (Phila Pa 1976). 2001; 26(6): E122–9.

11. Wu Y, Zhu Y, Han X. A meta-analysis of artificial total disc replacement versus fusion for lumbar degenerative disc disease. Eur Spine J. 2010; 19(8):1250–61. https://doi.org/10.1007/s00586-010-1394-x PMID: 20364392

12. Wei J, Song Y, Sun L, Lv C. Comparison of artificial total disc replacement versus fusion for lumbar degenerative disc disease: a meta-analysis of randomized controlled trials. Int Orthop. 2013; 37 (7):1315–25. https://doi.org/10.1007/s00264-013-1883-8 PMID: 23645003

13. Rao MJ, Cao SS. Artificial total disc replacement versus fusion for lumbar degenerative disc disease: a meta-analysis of randomized controlled trials. Arch Orthop Trauma Surg. 2014; 134(2):149–58. https://doi.org/10.1007/s00402-013-1905-4 PMID: 24323061

14. Jacobs W, Van der Gaag NA, Tuschel A, de Kleuver M, Peul W, Verbout AJ, et al. Total disc replacement for chronic back pain in the presence of disc degeneration. Cochrane Database Syst Rev. 2012; 9:CD008326.

15. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. Ann Intern Med 2009; 151(4):W-65–W-94.

16. Furlan AD, Malmivaara A, Chou R, Maher CG, Deyo RA, Schoene M, et al. 2015 updated method guideline for systematic reviews in the Cochrane back and neck group. Spine (Phila Pa 1976). 2015; 40:1660–73.

17. Wells GA, Shea B, O'Connell D. The Newcastle-Ottawa Scale (NOS) for assessing the quality of non-randomised studies in Meta-analyses. Ottawa: Ottawa Hospital Research Institute. 2011.

18. Guyatt G, Oxman AD, Sultan S, Brozek J, Glasziou P, Alonso-Coello P, et al. GRADE guidelines: 11. Making an overall rating of confidence in effect estimates for a single outcome and for all outcomes. J Clin Epidemiol. 2013; 66(2):151–7. https://doi.org/10.1016/j.jclinepi.2012.01.006 PMID: 22542023

19. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003; 327(7414):557–60. https://doi.org/10.1136/bmj.327.7414.557 PMID: 12958120

20. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol. 2009; 62(10):1006–12. https://doi.org/10.1016/j.jclinepi.2009.06.005 PMID: 19631508

21. Blumenthal S, McAfee PC, Guyer RD, Hochschuler SH, Geisler FH, Holt RT, et al. A prospective, randomized, multicenter Food and Drug Administration investigational device exemptions study of lumbar

total disc replacement with the CHARITE artificial disc versus lumbar fusion: part I: evaluation of clinical outcomes. Spine (Phila Pa 1976). 2005; 30(14):1565–75; discussion E387-391.

22. Holt RT, Majd ME, Isaza JE, Blumenthal SL, McAfee PC, Guyer RD, et al. Complications of lumbar artificial disc replacement compared to fusion: results from the prospective, randomized, multicenter US Food and Drug Administration Investigational Device exemption study of the Charité artificial disc. SAS J. 2007; 1(1): 20–7. https://doi.org/10.1016/SASJ-2006-0004-RR PMID: 25802575

23. Geisler FH, McAfee PC, Banco RJ, Blumenthal SL, Guyer RD, Holt RT, et al. Prospective, randomized, multicenter FDA IDE Study of CHARITÉ artificial disc versus lumbar fusion: effect at 5-year follow-up of prior surgery and prior discectomy on clinical outcomes following lumbar arthroplasty. SAS J. 2009; 3 (1):17–25. https://doi.org/10.1016/SASJ-2008-0019-RR PMID: 25802625

24. Gornet MF, Burkus JK, Dryer RF, Peloza JH. Lumbar disc arthroplasty with Maverick disc versus stand-alone interbody fusion: a prospective, randomized, controlled, multicenter investigational device exemption trial. Spine (Phila Pa 1976). 2011; 36(25):E1600–11.

25. Strube P, Putzier M, Streitparth F, Hoff EK, Hartwig T. Postoperative posterior lumbar muscle changes and their relationship to segmental motion preservation or restriction: a randomized prospective study. J Neurosurg Spine. 2016; 24(1):25–31. https://doi.org/10.3171/2015.3.SPINE14997 PMID: 26360146

26. Mattei TA, Beer J, Teles AR, Rehman AA, Aldag J, Dinh D. Clinical outcomes of total disc replacement versus anterior lumbar interbody fusion for surgical treatment of lumbar degenerative disc disease. Global Spine J. 2017; 7(5): 452–9. https://doi.org/10.1177/2192568217712714 PMID: 28811990

27. Ding F, Jia Z, Zhao Z, Xie L, Gao X, Ma D, et al. Total disc replacement versus fusion for lumbar degenerative disc disease: a systematic review of overlapping meta-analyses. Eur Spine J. 2017; 26(3):806–5. https://doi.org/10.1007/s00586-016-4714-y PMID: 27448810

28. Liao G, Deng F, Yin X, Lin J. Comment on Wei et al.: Comparison of artificial total disc replacement versus fusion for lumbar degenerative disc disease: a meta-analysis of randomized controlled trials. Int Orthop. 2013; 37(10):2107–8. https://doi.org/10.1007/s00264-013-1991-5 PMID: 23873172

29. Li ZY, Han X, Ma S, Tian W. Effect of total lumbar disc replacement on the treatment of discogenic low lumbar pain: preliminary outcomes. Chin Med J (Engl). 2013; 126(8):1504–8. PMID: 23595385

30. McAfee PC, Cunningham B, Holsapple G, Adams K, Blumenthal S, Guyer RD, et al. A prospective, randomized, multicenter Food and Drug Administration investigational device exemption study of lumbar total disc replacement with the CHARITE artificial disc versus lumbar fusion: part II: evaluation of radiographic outcomes and correlation of surgical technique accuracy with clinical outcomes. Spine (Phila Pa 1976). 2005; 30(14):1576–83; discussion E388-390.

31. Sasani M, Oktenoğlu T, Tuncay K, Canbulat N, Carilli S, Ozer FA. Total disc replacement in the treatment of lumbar discogenic pain with disc herniation: a prospective clinical study. Turk Neurosurg. 2009; 19(2):127–34. PMID: 19431121

32. Guyer RD, McAfee PC, Banco RJ, Bitan FD, Cappuccino A, Geisler FH, et al. Prospective, randomized, multicenter Food and Drug Administration investigational device exemption study of lumbar total disc replacement with the CHARITE artificial disc versus lumbar fusion: five-year follow-up. Spine J. 2009; 9 (5):374–386. https://doi.org/10.1016/j.spinee.2008.08.007 PMID: 18805066

33. Zigler JE, Delamarter RB. Five-year results of the prospective, randomized, multicenter, Food and Drug Administration investigational device exemption study of the ProDisc-L total disc replacement versus circumferential arthrodesis for the treatment of single-level degenerative disc disease. J Neurosurg Spine. 2012; 17(6):493–501. https://doi.org/10.3171/2012.9.SPINE11498 PMID: 23082846

34. Lu S, Sun S, Kong C, Sun W, Hu H, Wang Q, et al. Long-term clinical results following Charite III lumbar total disc replacement. Spine J. 2017. pii: S1529–9430(17)30969-5.

35. Guyer RD, Pettine K, Roh JS, Dimmig TA, Coric D, McAfee PC, et al. Five-year follow-up of a prospective, randomized trial comparing two lumbar total disc replacements. Spine (Phila Pa 1976). 2016; 41 (1):3–8.