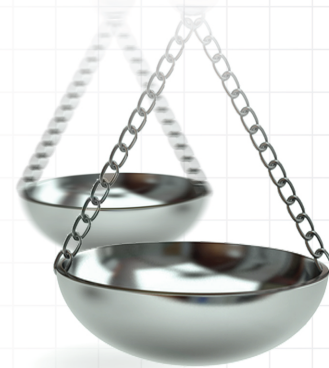




Comparison of therapies in lumbar degenerative disc disease: a network meta-analysis of randomized controlled trials

Journal of **Comparative Effectiveness Research**

Jack Zigler¹, Nicole Ferko², Chris Cameron² & Leena Patel^{*,2}

¹Texas Back Institute, 6020 West Parker Road #200, Plano, TX 75093, USA

²Cornerstone Research Group, 204–3228 South Service Rd., Burlington ON, Canada

* Author for correspondence: Tel.: +1 905 637 6231, ext. 222; Fax: +1 906 637 5014; lpatel@cornerstone-research.com

Aim: To compare the efficacy and safety of total disc replacement, lumbar fusion, and conservative care in the treatment of single-level lumbar degenerative disc disease (DDD). **Materials & methods:** A network meta-analysis was conducted to determine the relative impact of lumbar DDD therapies on Oswestry Disability Index (ODI) success, back pain score, patient satisfaction, employment status, and reoperation. Odds ratios or mean differences and 95% credible intervals were reported. **Results:** Six studies were included (1417 participants). Overall, the activL total disc replacement device had the most favorable results for ODI success, back pain, and patient satisfaction. Results for employment status and reoperation were similar across therapies. **Conclusion:** activL substantially improves ODI success, back pain, and patient satisfaction compared with other therapies for single-level lumbar DDD.

First draft submitted: 20 June 2017; Accepted for publication: 25 August 2017; Published online: 20 September 2017

Keywords: conservative care • lumbar arthrodesis • lumbar arthroplasty • lumbar degenerative disc disease • lumbar fusion • network meta-analysis • total disc replacement

Degenerative disc disease (DDD) in the lumbar spine is a leading cause of pain and disability in adults and is the reason for over 90% of spinal surgeries performed [1,2]. Treatment for patients with symptomatic lumbar DDD always begins with rehabilitation and pain management [3]. For patients with functionally disabling discogenic low back pain that does not improve after conservative care, surgical options such as lumbar spine fusion and arthroplasty, which is the surgical reconstruction or replacement of a joint, may be considered. Unlike lumbar fusion, which is broadly indicated for several spinal disorders [4], arthroplasty is generally only indicated in patients with discogenic low back pain due to DDD between the fourth and fifth lumbar vertebrae (L4/L5) or fifth lumbar and first sacral vertebrae (L5/S1) that has been unresponsive to conservative care [5,6].

Surgical approaches to the treatment of discogenic low back pain are associated with risks of procedure- and device-related complications and additional stress on other spinal segments, potentially leading to repeat surgeries or adjacent segment disease [7–10]. Since its commercial introduction over 10 years ago, arthroplasty continues to be an effective alternative to current therapies. In patients with lumbar DDD, evidence from randomized controlled trials (RCTs) and meta-analyses demonstrate greater success with arthroplasty than with lumbar fusion or conservative care and lower risk of adjacent segment disorder with arthroplasty than with lumbar fusion [11–21]. Advancements in the design and technology of total disc replacement (TDR) devices may further improve effectiveness and safety outcomes in arthroplasty [22]. First-generation TDR devices such as Charité and ProDisc-L offered a degree of motion preservation to protect against the increased mechanical stress imposed by lumbar fusion. The latest-generation device, the activL[®] Artificial Disc (Aesculap, Tuttlingen, Germany), advances motion-preserving technology to more closely align with the natural motion of the healthy human spine, thereby potentially further reducing wear on facet joints and adjacent segments [22].

With the evolution of treatments for lumbar DDD and proven benefits associated with arthroplasty over lumbar fusion, the design of recent RCTs has so far precluded comparison of newer TDR devices such as activL (comparators

consisted of two other TDR devices) with fusion or conservative care [23,24]. Thus, there is need for a comprehensive analysis that allows comparison across all relevant treatments for lumbar DDD. A network meta-analysis (NMA) enables the indirect comparison of two treatments that have not been directly compared in clinical trials but have a direct comparator in common. Therefore, the objective of this study was to conduct an NMA of RCTs to evaluate the efficacy and safety of the activL Artificial Disc compared with currently available alternative therapies, including other TDR devices for arthroplasty, lumbar fusion approaches and conservative care modalities, in patients with single-level lumbar DDD.

Materials & methods

A Bayesian statistical model was used to conduct the NMAs. Reporting of the analysis follows the Preferred Reporting Items of Systematic Reviews and Meta-analyses (PRISMA) Extension Statement for Reporting of Systematic Reviews Incorporating Network Meta-analyses of Health Care Interventions [25].

Data sources & search strategy

A comprehensive strategy was developed to search the PubMed and Cochrane Central Register of Controlled Trials (CENTRAL) databases to identify relevant RCTs published until December 2015 (Supplementary Appendix 1). Bibliographic and grey literature searches were also conducted. Only English language articles were reviewed.

Eligibility criteria & study selection

RCTs that included patients with discogenic low back pain due to single-level lumbar DDD, who were unresponsive to conservative therapy, were considered if they compared a TDR device (Charité, ProDisc-L, Maverick, Kineflex-L, Flexicore, activL) with other TDR devices, fusion (anterior, posterior, or circumferential) or conservative care (rehabilitation, exercise). Studies must have reported at least one of the following outcomes at 2-year follow-up: Oswestry Disability Index (ODI) success (≥ 15 -point improvement), mean back pain score (Visual Analog Scale [VAS] or Numeric Rating Scale [NRS]), patient satisfaction, employment status, reoperations (device failures resulting in removal, revision, reoperation or supplemental fixation) and device-related serious adverse events (SAEs). Studies were excluded if they had $>50\%$ loss-to-follow-up, were single-arm or observational studies, reported outcomes for a subset of the population captured in the full study (i.e., double reporting) or presented data for treatment arms in an aggregated form. Studies were selected by two independent reviewers, with differences resolved by discussion resulting in consensus or by a third party.

Data extraction

Data on study and patient baseline characteristics, treatments, loss to follow-up and key outcomes pertinent to this analysis were extracted by a single reviewer and checked for accuracy by a second reviewer; disagreements were resolved by consensus or by a third party.

Some data were extracted by digitizing data (TechDig v2.0 digitizing software) and mean differences were calculated from baseline and follow-up data provided. For continuous outcomes, a normal distribution was assumed. No data imputation was required.

Quality assessment

The Cochrane Collaboration's tool for assessing risk of bias was used to evaluate study quality. The tool assesses the risks of selection bias (sequence generation, allocation concealment), performance bias (blinding of participants and personnel), detection bias (blinding of outcome assessment), attrition bias (incomplete outcome data), reporting bias (selective outcome reporting) and other bias [26]. Two independent reviewers scored studies for low, unclear or high risk of bias, with discrepancies resolved by a third party.

Data synthesis & analysis

An NMA permits comparison of treatments across a network of trials, where different treatment strategies are employed within the same or very similar patient population. The NMA approach allows any two treatments within the network of evidence to be compared, even when a direct comparison from a trial is not available. Each treatment was considered as an individual node in the network, with networks constructed using NetMetaXL (Figure 1) [27] and analyses performed using WinBUGS Software (version 1.4.3, MRC Biostatistics Unit, Cambridge, UK). For outcomes such as device-related SAEs where a network could not be constructed because of insufficient

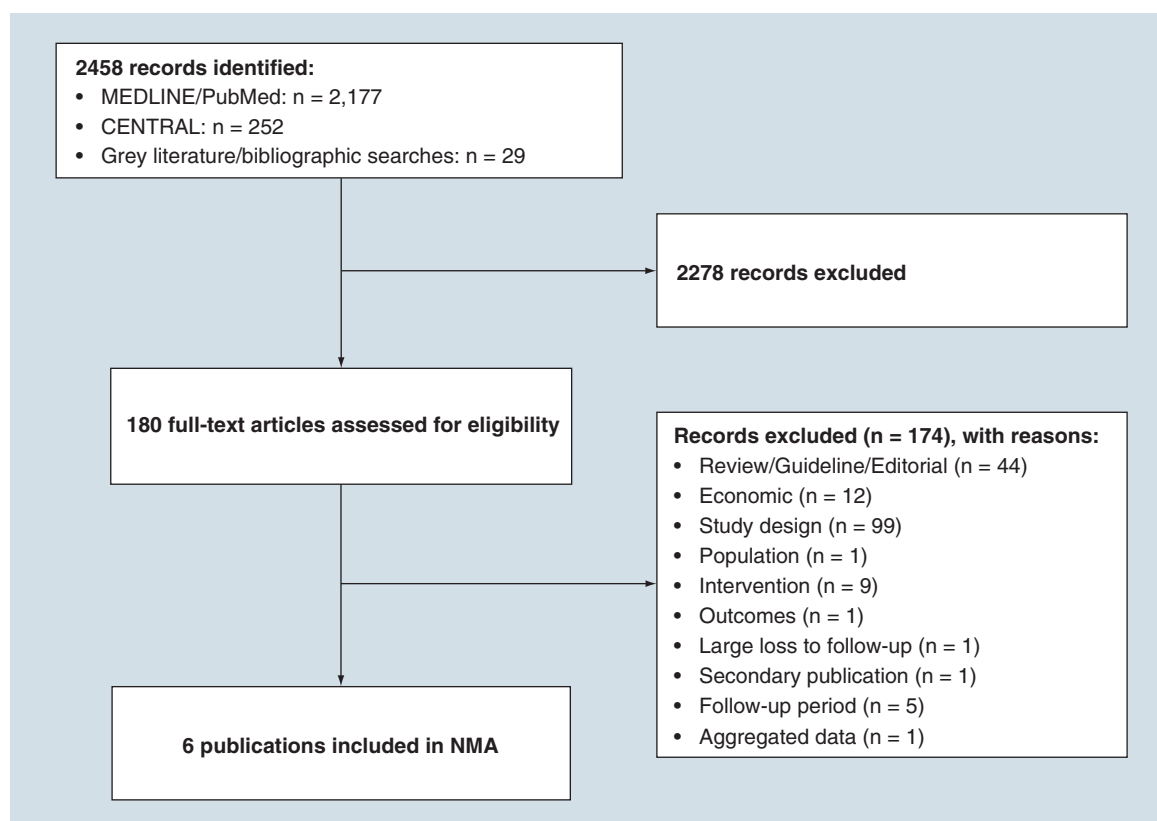


Figure 1. PRISMA flowchart.
NMA: Network meta-analysis.

data, an NMA was not conducted. Analyses were conducted using Bayesian fixed effects models that were fit to the data with a binomial likelihood. The models and priors used were consistent with recommendations made by the NICE Decision Support Unit [28]. Results of pairwise comparisons were reported as odds ratios (ORs) and mean differences (MDs). ORs were summarized using the median and 95% credible interval (CrI). ORs less than or greater than 1 favor one of the compared lumbar DDD therapies over the other, whereas ORs equal to 1 indicate equivalency. Mean differences were summarized using the mean and 95% CrI; MDs less than or greater than 0 favor one of the compared lumbar DDD treatments over the other, whereas MDs equal to 0 indicate equivalency. In brief, CrIs can be interpreted as the Bayesian equivalent to confidence intervals. Credible intervals excluding 1 for ORs or 0 for MDs indicate statistical significance. The probability that each treatment was the most efficacious was calculated by counting the number of times each treatment had the highest OR or MD in the model [29]. A single numerical summary of the probability that a treatment is best, second best and so on, known as the surface under the cumulative ranking curve area (SUCRA), was also calculated for each treatment. More detailed information regarding the statistical methods is included in Supplementary Appendix 3.

Because an NMA requires that studies are sufficiently similar in order to pool results, available study and patient characteristics were assessed to ensure similarity and to investigate the potential effect of clinical heterogeneity on ORs or MDs and 95% CrIs. Another key assumption behind NMA is that the analyzed network is consistent; that is, there is no conflict between direct and indirect evidence within a closed loop in the network, where the closed loop is comprised of different studies [30]. Evidence networks in our NMA consisted of single-study connections and contained a single closed loop based on data from 1 trial; therefore, we were unable to perform an inconsistency analysis [30]. Additionally, the results from our NMA were qualitatively compared with the results from direct comparisons generated from traditional meta-analyses.

Table 1. Summary of study characteristics and baseline patient demographics.						
Study (year)	Blumenthal (2005)	Gornet (2011)	Zigler (2007)	Garcia (2015)	Guyer (2014)	Hellum (2011)
Comparison	TDR vs fusion	TDR vs fusion	TDR vs fusion	TDR vs TDR	TDR vs TDR	TDR vs CC
Treatment	Charité	Maverick	ProDisc-L	activL	Kineflex-L	ProDisc-L
Comparator #1	ALIF	ALIF	Circumferential fusion	Charité	Charité	Rehabilitation
Comparator #2	–	–	–	ProDisc-L	–	–
n	304	577	292	324	394 [§]	173
Blinding	OL	OL	SB [†]	SB	SB [‡]	–
Analysis population	ITT	–	ITT	ITT	–	ITT
Lost to follow-up at 2 years (%)	9%	9%	1.80%	9%	–	20%
Mean age (years)	39.6	40	39	39	40	41
Prior spinal surgery	34%	28%	34%	25%	27%	28%
Mean BMI (kg/m ²)	26	–	27	27	27	26
Number of vertebral levels affected	1	1	1	1	1	1 & 2
L4/5 level treated (%)	30%	24%	32%	30%	24%	22% [¶]
L5/S1 level treated (%)	69%	75%	65%	70%	76%	46% [¶]
Work status (% working)	53%	60%	–	–	–	27%
Mean baseline ODI score	51%	54%	63%	58%	60%	42%
Mean baseline back pain score [#]	72	72	76	79	80	69
[#] Back pain score assessed using VAS (in mm) in Blumenthal 2005, Zigler 2007, Garcia 2015 and Hellum 2011; NRS used in Gornet 2011. [†] Patients were unblinded after the surgery. [§] Represents total number of patients randomized to either treatment group and does not include nonrandomized patients in the treatment groups. [‡] Patients were blinded until after surgery. [¶] Represents surgery group only. Not reported for nonsurgical group. BMI: Body mass index; CC: Conservative care; ITT: Intention to treat; L4/L5: Fourth and Fifth Lumbar Vertebral Segment; L5/S1: Fifth Lumbar and First Sacral Vertebral Segment; NRS: Numeric Rating Scale; ODI: Oswestry Disability Index; OL: Open label; SB: Single blind; TDR: Total disc replacement; VAS: Visual Analog Scale.						

Sensitivity analyses

Sensitivity analyses were conducted for the following: variations in efficacy of anterior lumbar interbody fusion (ALIF) instrumentation [31], where efficacy was adjusted using a simulated trial of 100 patients to consider assumptions regarding the difference between using recombinant human bone morphogenetic protein 2 (rhBMP-2) and iliac crest bone graft in ALIF (i.e., 10% increase in efficacy in favor of rhBMP-2) [14,17]; outcome definitions, where studies that used different definitions or measures for outcomes of interest were excluded [15,17]; follow-up rate, where studies with a ≥20% loss to follow-up were excluded [12]; exclusion of studies that included patients with 1 or 2 affected spinal segments [12]; and exclusion of studies with a high risk of bias [12,17].

Results

Study inclusion & baseline characteristics

From the literature search, 2458 potentially relevant records were identified, of which 2278 records were excluded during title and abstract review. Of the 180 eligible, full-text articles assessed for inclusion, 6 publications were included in the analysis [12,14,15,17,23,24] (Figure 1). Three trials compared a TDR device with fusion [14,15,17], two compared a TDR device with another TDR device [23,24] and one compared a TDR device with conservative care [12] (Table 1). Only one study included three treatment arms [23]. Charité and ProDisc-L were included in three trials each (each as an investigational device in one and as a control device in two others), whereas Maverick, Kineflex-L and activL were included in one trial each. ALIF was included as a comparator in two trials, circumferential fusion in one trial and rehabilitation in one trial. Five studies, which were corporately sponsored, were regulated under the US FDA Investigational Device Exemption (IDE) program; one study was not involved in the FDA’s IDE program and was not corporately sponsored.

Sample sizes for included trials ranged from 173 to 577 participants (Table 1). Follow-up for all studies was 2 years; 1.8–20% loss to follow-up was observed at 2 years. Only one study included patients with lumbar DDD at 1 and 2 levels [12]. Most patients were treated at the L5/S1 level. In all studies, patients were unresponsive or had insufficient improvement to non-surgical therapy. At baseline, the ODI scores were 42–63% and mean back pain scores were 69–80 across studies.

Risk of bias assessment

The risk of bias was similar across studies (Supplementary Appendix 2). Methods for randomized sequence generation were reported in all studies: all studies used block randomization, with three studies reporting use of block sizes of six [14,15,17]. Among patients randomized to the control arm in one study, surgeons were given the choice of implanting ProDisc-L or Charité [23]. Central allocation was reported as the method used in five studies [12,14,15,17,24], with sites notified of allocation using sealed envelopes in three studies [14,17,24], by telephone in one study [15] and through website access in one study [12]. Surgeons and/or staff were not blinded for preparatory purposes in three studies [14,15,23], were blinded until informed consent was received in one study [17] and were blinded until shortly after randomization in one study [12]. Participants were blinded to their randomization group until after surgery in four studies [14,15,23,24]. Intention-to-treat (ITT) analyses were conducted in five studies [12,14,15,23,24]. Analyses in one study were not conducted by ITT, however, this study was deemed to have low risk for selection bias because analyses were approached conservatively: the one study participant who was randomized to receive the investigational device received the control treatment and had successful outcomes postoperatively; therefore, this patient was included within the control group [17].

ODI success

All studies were included in the evidence network for ODI success, with all eight treatment comparators considered (Figure 2A). Outcomes data from the individual studies are presented in Supplementary Appendix 4A. When compared with all treatment comparators, activL was associated with the highest ORs of ODI success (Supplementary Appendix 5A). The ORs of achieving ODI success were significantly in favor of activL when compared with circumferential fusion (OR: 2.58; 95% CrI: 1.13–5.83), ALIF (OR: 2.57; 95% CrI: 1.08–6.05) and rehabilitation (OR: 3.87; 95% CrI: 1.64–9.13) (Figure 3A, Supplementary Appendix 5A). Overall, activL had the highest probability of being the best treatment among all therapies (72%; Figure 4) and had the highest overall SUCRA (94%; Supplementary Appendix 6).

The OR estimates for ODI success from direct comparisons aligned well with the estimates obtained from the NMA (Supplementary Appendix 8). Model fit statistics indicated reasonable model fit.

Back pain score

The analysis for back pain score included all studies and comparators (Figure 2B). Outcomes data from the individual studies are presented in Supplementary Appendix 4B. When compared with all treatment comparators, activL showed the greatest mean difference from baseline in back pain score (Supplementary Appendix 5B). The mean change for back pain score, from baseline to 2 years, was significantly greater for activL than for Charité (MD -10.42; 95% CrI: -20.07, -0.82), Kineflex-L (MD -11.60; 95% CrI: -22.98, -0.33), and ALIF (MD -16.84; 95% CrI: -29.22, -4.39) (Figure 3B, Supplementary Appendix 5B). Overall, activL had the highest probability of being the best treatment among all therapies (probability best, 61%; Figure 4) and had the highest SUCRA (91%; Supplementary Appendix 6).

The MD estimates for back pain score from direct comparisons aligned well with the estimates obtained from the NMA (Supplementary Appendix 8). Model fit statistics indicated reasonable model fit.

Patient satisfaction

For patient satisfaction, the evidence network included five studies and seven comparators (Figure 2D). Outcomes data from the individual studies are presented in Supplementary Appendix 4A. Overall, activL had the highest ORs for patient satisfaction from all comparisons (Supplementary Appendix 5C). The ORs for patient satisfaction significantly favored activL compared with rehabilitation (OR 3.30; 95% CrI: 1.39–7.84) and ALIF (OR 3.75; 95% CrI: 1.56–8.77) (Figure 3C, Supplementary Appendix 5C). Overall, activL had the highest probability of being the best treatment among all therapies (probability best, 46%; Figure 4) and had the highest SUCRA (86%; Supplementary Appendix 6).

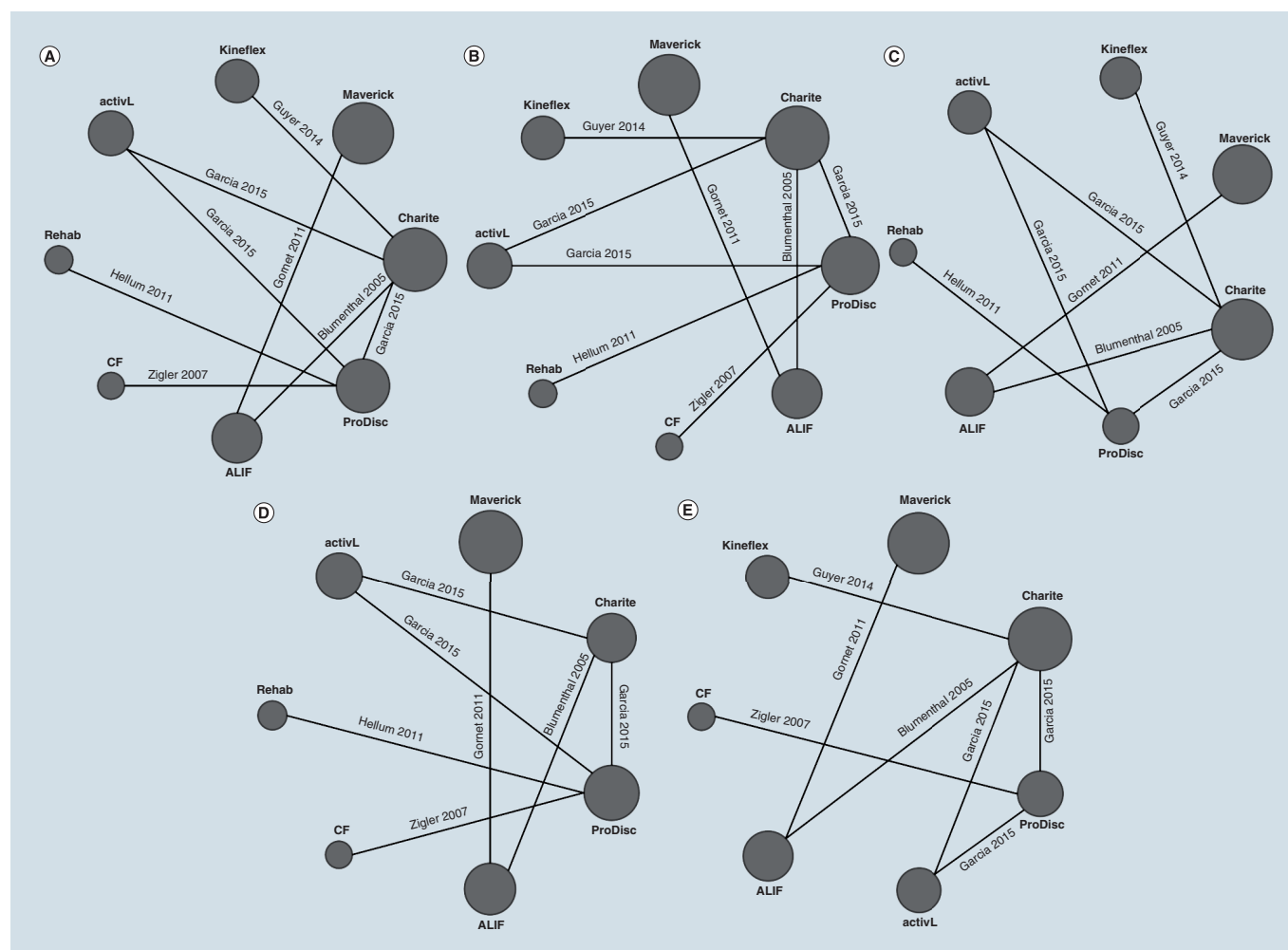


Figure 2. Evidence networks for (A) ODI success, (B) back pain score, (C) patient satisfaction, (D) employment status and (E) reoperation. In the evidence networks, the width of the lines for each connection is proportional to the number of randomized controlled trials comparing each pair of treatments. The size of each treatment node is proportional to the number of randomized participants (sample size).

ALIF: Anterior lumbar interbody fusion; CF: Circumferential fusion; Rehab: Rehabilitation.

The OR estimates for patient satisfaction from direct comparisons aligned well with the estimates obtained from the NMA (Supplementary Appendix 8). Model fit statistics indicated reasonable model fit.

Employment status

Four studies were included in the evidence network for employment status and seven comparators were considered (Figure 2D). Outcomes data from the individual studies are presented in Supplementary Appendix 4A. When compared with patients who received other treatments, activL patients had the highest ORs of being employed (Supplementary Appendix 5D). Overall, there were no statistically significant differences between the comparators (Figure 3D; Supplementary Appendix 5D).

The OR estimates for employment status from direct comparisons aligned well with the estimates obtained from the NMA (Supplementary Appendix 8). Model fit statistics indicated reasonable model fit.

Reoperation

The evidence network for reoperation, defined as device failures requiring revision, removal, reoperation, or supplemental fixation, included all studies but one [12] (Figure 2E); seven treatment comparators were considered. Outcomes data from the individual studies are presented in Supplementary Appendix 4A. Overall, ORs for reop-

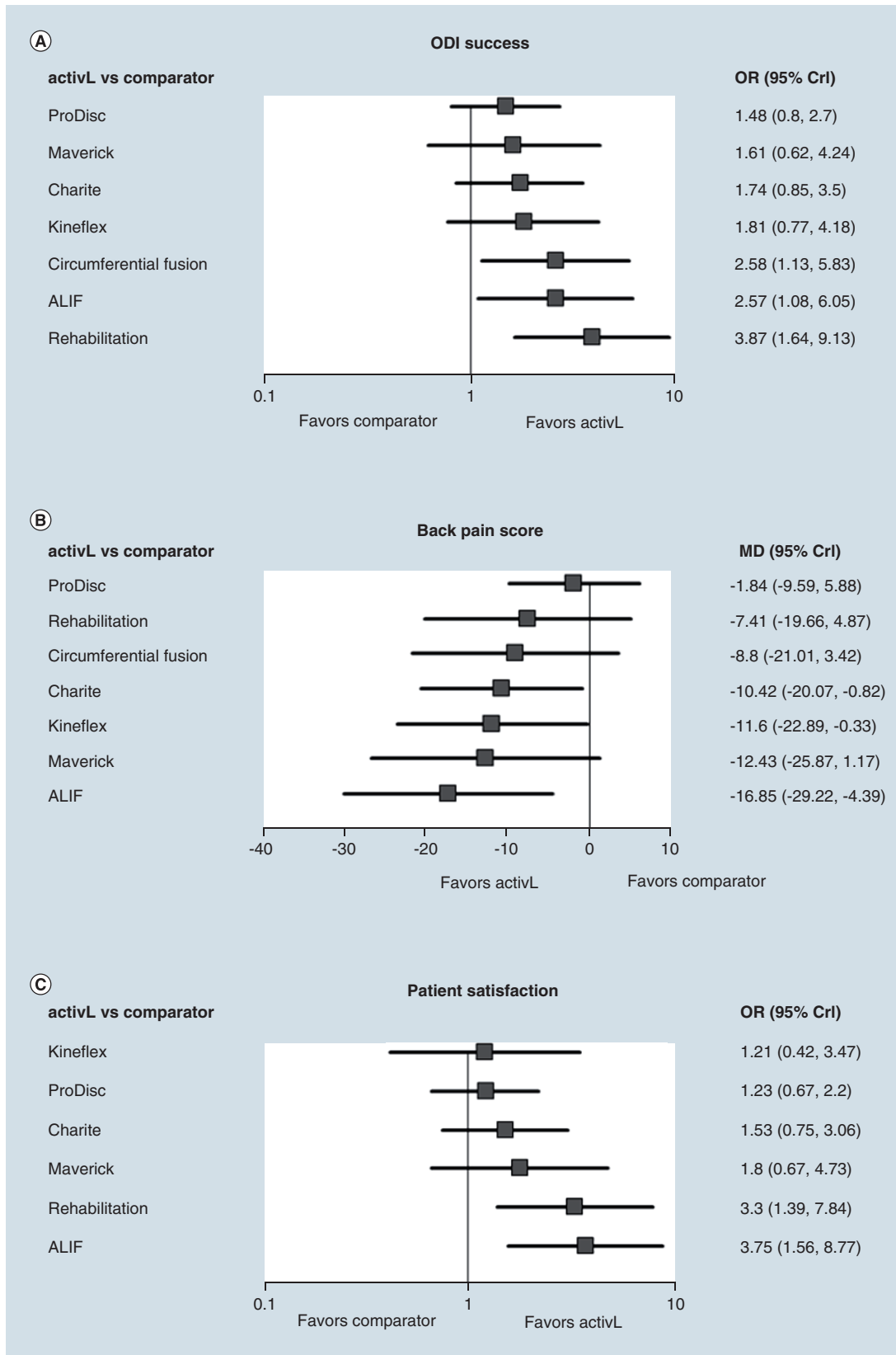


Figure 3. Forest plots presenting effect estimates and 95% credible intervals of activL versus comparator for (A) ODI success, (B) back pain score, (C) patient satisfaction, (D) employment status and (E) reoperation. Comparisons of activL versus comparator are ordered in relative ranking of success for activL. For ODI success, back pain score, patient satisfaction and employment status, ORs > 1 favor activL, whereas for reoperations, ORs < 1 favor activL. ALIF: Anterior lumbar interbody fusion; CrI: Credible interval; MD: Mean difference; ODI: Oswestry Disability Index; OR: Odds ratio.

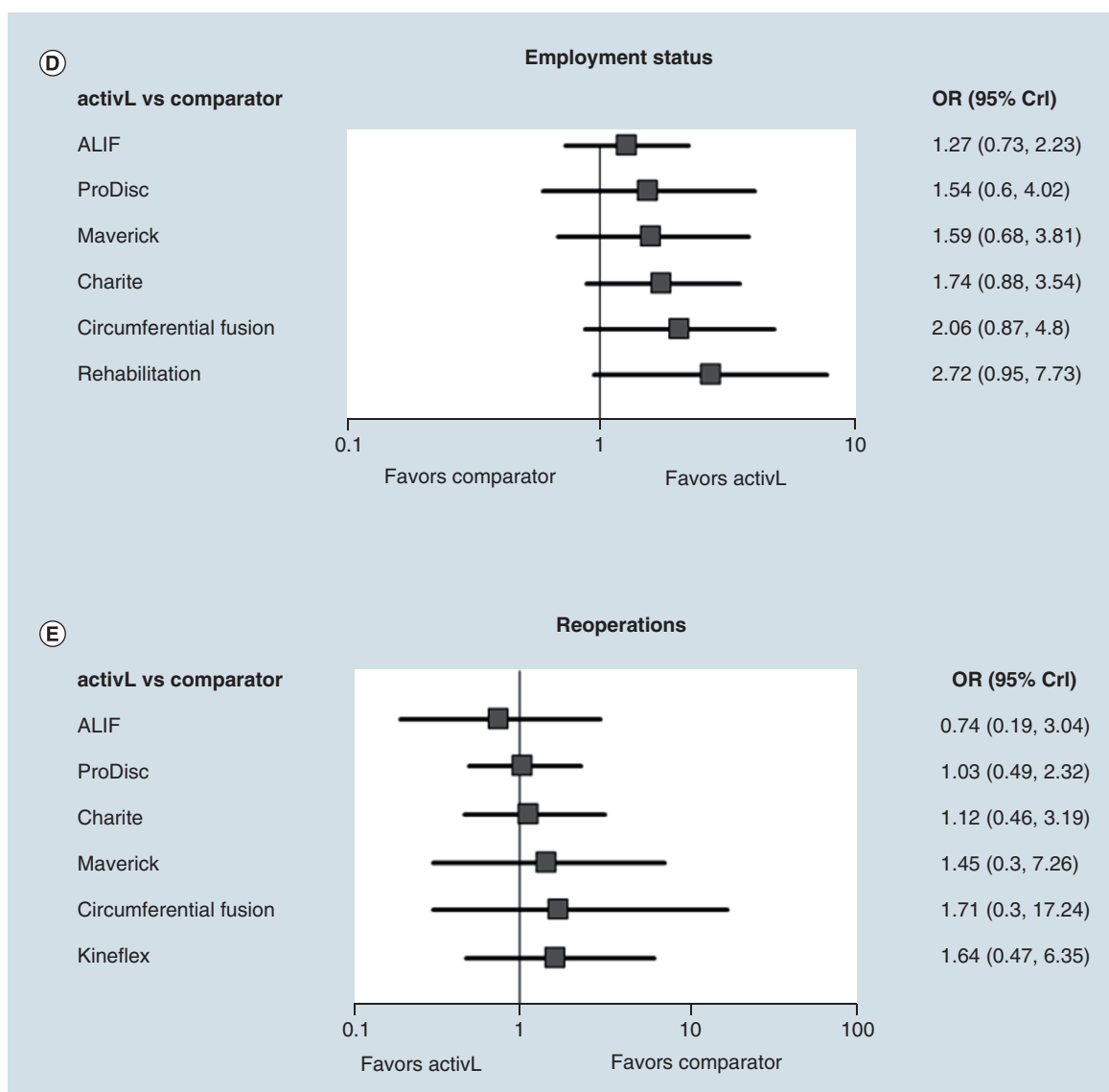


Figure 3. Forest plots presenting effect estimates and 95% credible intervals of activL versus comparator for (A) ODI success, (B) back pain score, (C) patient satisfaction, (D) employment status and (E) reoperation (cont.). Comparisons of activL versus comparator are ordered in relative ranking of success for activL. For ODI success, back pain score, patient satisfaction and employment status, ORs > 1 favor activL, whereas for reoperations, ORs < 1 favor activL. ALIF: Anterior lumbar interbody fusion; CrI: Credible interval; MD: Mean difference; ODI: Oswestry Disability Index; OR: Odds ratio.

eration were in favor of Kineflex-L (Supplementary Appendix 5E). However, there were no statistically significant differences between comparators.

The OR estimates for reoperations from direct comparisons aligned well with the estimates obtained from the NMA (Supplementary Appendix 8). Model fit statistics indicated reasonable model fit.

Sensitivity analyses

Results of the primary analysis were robust for the variables tested in the sensitivity analyses (Supplementary Appendix 9). For ODI success, when adjusted for differences in efficacy associated with ALIF based on the technique used, results for ODI success were slightly sensitive to the assumption of better efficacy with rhBMP-2 than with iliac crest bone graft. Results remained in favor of activL when the analysis excluded studies using a modified version of the ODI questionnaire, studies with $\geq 20\%$ loss to follow-up, studies that included 1- and



Figure 4. Probability of being the best treatment (%) across outcomes. Bayesian network meta-analysis allows estimates for the probability that a treatment is best, second best and so on, for a particular outcome. The figure presents results for the probability that a treatment is best for the given outcome, where the darker green indicates the highest probability of being the best treatment for a given outcome. ALIF: Anterior lumbar interbody fusion; ODI: Oswestry Disability Index.

2-level surgeries and studies with a high risk of bias. Results for back pain score were sensitive to studies with a high loss to follow-up and studies that included 1- and 2-level surgeries, with MDs significantly in favor of activL. Back pain score results were robust to studies using the NRS pain scale and those with a high risk of bias. Patient satisfaction results were robust to excluding studies with $\geq 20\%$ loss to follow-up and studies with a high risk of bias. Results from the primary analysis for employment status were robust to exclusion of studies with $\geq 20\%$ loss to follow-up and those that included 1- and 2-level surgeries. Excluding studies with a high risk of bias resulted in numerically higher ORs for employment status that continued to favor activL; however, statistical significance was not reached. Results for reoperations were slightly sensitive to the use of the FDA's criteria for reoperation success in the Kineflex-L IDE study and to the exclusion of studies with a high risk of bias.

Discussion

Several treatment options are available for the treatment of symptomatic, single-level lumbar DDD, including conservative care, fusion and arthroplasty. Evidence from several studies shows that arthroplasty is superior to fusion and conservative care [18–20,32–34]. The activL Artificial Disc is the latest evolution of TDR devices for arthroplasty that further advances mechanical motion-preservation in the lumbar spine. Clinical trial data show that activL is noninferior to Charité and ProDisc-L in terms of efficacy and safety, with range of motion outcomes being superior, but does not directly evaluate the impact of activL compared with fusion and conservative care [23].

Thus, the current study sought to conduct a comprehensive NMA of data from all available RCTs of lumbar TDR devices to evaluate the efficacy and safety of activL compared with not only other devices, but now also with lumbar fusion and conservative care.

Overall results from this analysis demonstrate that activL is associated with improved benefits in efficacy and with similar benefits in safety compared with other treatments for single-level lumbar DDD. Benefits include having the most favorable effect estimates for ODI success, back pain score and patient satisfaction compared with all other treatments. Additionally, activL is associated with the highest probability of being the best treatment not only among all surgical options for ODI success, back pain score and patient satisfaction, but also for all nonsurgical options. The results from the current analysis, given similar study designs and patient populations, are unlikely to arise from technical differences between studies; rather, its findings are more likely due to the advanced design attributes and improved outcomes associated with the activL Artificial Disc. As reported in this NMA, activL has shown better outcomes, albeit not statistically significant, than ProDisc-L in 2-year patient outcomes for ODI improvement, back pain score, patient satisfaction and employment status. Charité, the other approved TDR comparator, is no longer commercially available.

Currently, there is no consensus on what constitutes a meaningful improvement [35]. Several different thresholds have been used to interpret the importance of ODI changes, however, meaningful score changes are likely different for each patient and a single cut-off value may not appropriately capture change for all [35]. One study defined a 10-point minimum clinically important difference between treatment groups, which was not achieved despite results being statistically significant [12]. Alternatively, the FDA recommends the use of a 15-point minimum improvement for each patient; thus, FDA IDE studies have assessed the proportion of patients who have achieved this threshold [14,15,17,23,24]. Because of the lack of agreement on a single threshold and difficulties in interpreting clinical meaningfulness the current analysis did not assess the minimum clinically important difference for ODI, and instead chose to adopt the FDA's recommendation for interpreting ODI success.

This analysis includes the most recent evidence from all available RCTs evaluating TDR devices at 2-year follow-up, including the activL FDA IDE trial [23]. All studies included had relatively low loss-to-follow-up. Interestingly, the highest loss-to-follow-up occurred in the study comparing ProDisc-L with rehabilitation, with greater loss-to-follow-up observed in the rehabilitation group (24%) than in the TDR group (15%) [12]. When this particular study was omitted in the sensitivity analysis for back pain score, results were significantly in favor of activL across all treatment comparisons in contrast to those from the primary analysis. Sensitivity analysis results for ODI success, patient satisfaction and employment status results remained in favor of activL.

Meta-analyses of therapies for lumbar DDD are prevalent in the literature and are consistent with our findings. Most of these focused on comparing TDR with fusion and results demonstrated that TDR is safe and effective for the treatment of lumbar DDD. Across several outcomes such as disability, pain and patient satisfaction, improvements were statistically significantly in favor of TDR compared with fusion [18–20,32–34]. Furthermore, TDR was demonstrated to have either similar or improved safety compared with fusion, as typically measured by total complications, reoperation rates and adjacent segment degeneration [18–20,32–34,36,37]. When compared with conservative care, TDR showed greater improvements in disability, pain and quality of life [32]. Typically, meta-analyses of therapies for lumbar DDD included studies of single-level lumbar DDD with a 2-year follow-up period; only one of these studies compared TDR with conservative care. The latest TDR devices such as activL have not been included in published meta-analyses thus far.

One limitation of this study was the use of different outcome measures among included studies. In particular, a modified version of the ODI questionnaire was used in one study [15] and the NRS was used to assess pain in one study [17]. Results of sensitivity analyses removing these studies were comparable to the primary analyses of ODI success and back pain score, respectively. In another study, results for a different definition for reoperation were also reported [24], and inclusion of these data had little impact on the primary analysis results. Furthermore, there were limited data to inform analyses; evidence networks consisted of single-study connections between the multiple treatments. Data for specific outcomes were further limited due to differences in reporting; for instance, device-related SAEs were reported in only a small number of studies that lacked a common comparator by which to conduct indirect comparisons. Another limiting factor is that the results of this study are relevant for a 2-year follow-up period; however, evidence from RCTs and recently pooled analyses on long-term outcomes support these results, showing that the benefits with TDR devices over fusion are maintained over 5 years [ZIGLER ET AL., UNPUBLISHED DATA] [38–42].

The lack of direct comparative data between alternative treatments creates variations across health technology assessments and challenges for decision-makers. Network meta-analyses have not been previously published in the field of lumbar DDD, but have the potential to significantly aid in the clinical management and decision-making around numerous potential therapies. NMA, incorporating relevant clinical data while maintaining rigorous and credible methods, provides a feasible alternative to direct comparisons of all therapeutic options through RCTs.

Our study allows for the indirect comparison of activL with all relevant treatment comparators, including competitive TDR devices, lumbar fusion, and conservative care, such that decision-makers can make informed and relevant decisions in this continuously growing therapeutic area. Enhanced outcomes with TDR have been shown to improve the economic burden associated with lumbar DDD. Health economic studies using 2-year data clearly demonstrate that reduced operating time and length of hospital stay with TDR compared with fusion translate into lower direct costs [43–47]. Further, evaluation of cost–effectiveness using RCT data indicate that TDR may be cost-effective compared with conservative care at 2-year follow-up [48].

Future research using different materials to improve axial load transmission, wear and interface with the native spine should only further improve outcomes. Even at this relatively early stage in the evolution of lumbar arthroplasty, clinical results are very encouraging. Using the broader view afforded by this NMA allows us to clearly see that in appropriately selected and screened patients, arthroplasty offers better results than fusion or continued conservative care.

Conclusion

This is the first highly rigorous NMA that compares TDR devices with each other and with other forms of standard care, including conservative care, for the treatment of single-level lumbar DDD at 2-year follow-up. The analysis includes the most recent evidence from all available RCTs evaluating lumbar TDR devices, including data from the FDA IDE trial of the most recently approved device, activL.

This NMA clearly demonstrates that, compared with other TDR devices, surgical fusion approaches, and even conservative care, the activL Artificial Disc substantially improves ODI success, back pain score and patient satisfaction without significantly impacting reoperations in patients with single-level lumbar DDD. When choosing arthroplasty, comparison of outcomes shows that patients who received activL did better than those who received ProDisc-L, currently its only competitor in the USA.

Acknowledgements

The authors would like to thank David Banko of B Braun (PA, USA) and Katie Kleinschuster of Aesculap (PA, USA) for constructive discussions on study design.

Financial & competing interests disclosure

This study was sponsored by Aesculap Implant Systems, LLC. J Zigler received no compensation for work on this manuscript; he has received consultancy fees from Aesculap and DePuy Synthes outside the submitted work. Cornerstone Research Group, Inc., was contracted both to conduct the analysis and develop the manuscript. N Ferko, C Cameron and L Patel are employees of Cornerstone Research Group, Inc. Cornerstone Research Group, Inc., receives consultancy fees from major pharmaceutical and device companies, including Aesculap. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

References

Papers of special note have been highlighted as: • of interest; •• of considerable interest

- 1 Cihangiroglu M, Yildirim H, Bozgeyik Z *et al.* Observer variability based on the strength of MR scanners in the assessment of lumbar degenerative disc disease. *Eur. J. Radiol.* 51(3), 202–208 (2004).
- 2 An HS, Anderson PA, Haughton VM *et al.* Introduction: disc degeneration: summary. *Spine (Phila Pa 1976)* 29(23), 2677–2678 (2004).
- 3 Qaseem A, Wilt TJ, Mclean RM, Forciea MA. Noninvasive treatments for acute, subacute, and chronic low back pain: a clinical practice guideline from the American College of Physicians. *Ann. Intern. Med.* 166(7), 514–530 (2017).
- 4 North American Spine Society. NASS coverage policy recommendations: lumbar fusion. May 2014 (2014).
- 5 U.S. Food and Drug Administration (FDA). Summary of safety and effectiveness data, activL artificial disc (2015). www.accessdata.fda.gov/cdrh_docs/pdf12/P120024b.pdf

Summary points

- Network meta-analyses allow pooling of data from multiple studies and indirect comparison of treatments that have not been compared in direct head-to-head trials.
- This is the first highly rigorous network meta-analysis that compares total disc replacement (TDR) devices, lumbar fusion and conservative care for the treatment of single-level lumbar degenerative disc disease (DDD) at 2-year follow-up.

Methods

- Randomized controlled trials comparing a TDR device with another TDR device, fusion or conservative care for single-level lumbar DDD were included. Studies reported at least one of the following outcomes: Oswestry Disability Index (ODI) success, back pain score, patient satisfaction, employment status, need for reoperation and device-related serious adverse events.
- Network meta-analyses were conducted using a fixed-effects model that reported odds ratios (ORs) or mean differences (MDs) and 95% credible intervals (95% CrIs). Other measures reported were the probability of being the best treatment and the surface under the cumulative ranking curve area.

Results

- activL had the highest ORs for achieving ODI success overall; statistically significant results were observed when activL was compared with circumferential fusion (OR 2.58; 95% CrI: 1.13–5.83), anterior lumbar interbody fusion (ALIF; OR 2.57; 95% CrI: 1.08–6.05), and rehabilitation (OR 3.87; 95% CrI: 1.64–9.13).
- activL had the greatest MDs for back pain score overall, with significantly greater reductions than Charité (MD -10.42; 95% CrI: -20.07, -0.82), Kineflex-L (MD -11.60; 95% CrI: -22.98, -0.33) and ALIF (MD -16.84; 95% CrI: -29.22, -4.39).
- Results for patient satisfaction were highest for activL; ORs significantly favored activL over rehabilitation (OR 3.30; 95% CrI: 1.39–7.84) and ALIF (OR 3.75; 95% CrI: 1.56–8.77).
- Results for employment status and reoperations were similar across comparators.

Conclusion

- Compared with other TDR devices, surgical fusion approaches and even conservative care, the activL Artificial Disc substantially improves ODI success, back pain score and patient satisfaction without significantly impacting reoperations in patients with single-level lumbar DDD.
- When choosing arthroplasty, comparison of outcomes shows that patients who received activL did better than those who received ProDisc-L, currently its only competitor in the USA.

- U.S. Food and Drug Administration (FDA). Summary of safety and effectiveness data, PRODISC-L total disc replacement (2006). www.accessdata.fda.gov/cdrh_docs/pdf5/P050010b.pdf
- Gibson JN, Waddell G. Surgery for degenerative lumbar spondylosis. *Cochrane Database Syst. Rev.* 3, CD001352 (2005).
- Pan A, Hai Y, Yang J, Zhou L, Chen X, Guo H. Adjacent segment degeneration after lumbar spinal fusion compared with motion-preservation procedures: a meta-analysis. *Eur. Spine J.* 25(5), 1522–1532 (2016).
- Van Den Eerenbeemt KD, Ostelo RW, Van Royen BJ, Peul WC, Van Tulder MW. Total disc replacement surgery for symptomatic degenerative lumbar disc disease: a systematic review of the literature. *Eur. Spine J.* 19(8), 1262–1280 (2010).
- Kalakoti P, Missios S, Maiti T *et al.* Inpatient outcomes and postoperative complications after primary versus revision lumbar spinal fusion surgeries for degenerative lumbar disc disease: a national (nationwide) inpatient sample analysis, 2002–2011. *World Neurosurg.* 85, 114–124 (2016).
- Hellum C, Berg L, Gjertsen O *et al.* Adjacent level degeneration and facet arthropathy after disc prosthesis surgery or rehabilitation in patients with chronic low back pain and degenerative disc: second report of a randomized study. *Spine (Phila Pa 1976)* 37(25), 2063–2073 (2012).
- Hellum C, Johnsen LG, Storheim K *et al.* Surgery with disc prosthesis versus rehabilitation in patients with low back pain and degenerative disc: two year follow-up of randomised study. *BMJ* 342, d2786 (2011).
- Johnsen LG, Brinckmann P, Hellum C, Rossvoll I, Leivseth G. Segmental mobility, disc height and patient-reported outcomes after surgery for degenerative disc disease: a prospective randomised trial comparing disc replacement and multidisciplinary rehabilitation. *Bone Joint J.* 95-B(1), 81–89 (2013).
- Blumenthal S, McAfee PC, Guyer RD *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)* 30(14), 1565–1575; discussion E1387–E1591 (2005).
- Zigler J, Delamarter R, Spivak JM *et al.* Results of the prospective, randomized, multicenter Food and Drug Administration investigational device exemption study of the ProDisc-L total disc replacement versus circumferential fusion for the treatment of 1-level degenerative disc disease. *Spine (Phila Pa 1976)* 32(11), 1155–1162; discussion 1163 (2007).

- 16 Berg S, Tullberg T, Branth B, Olerud C, Tropp H. Total disc replacement compared with lumbar fusion: a randomised controlled trial with 2-year follow-up. *Eur. Spine J.* 18(10), 1512–1519 (2009).
- 17 Gornet MF, Burkus JK, Dryer RF, Pelozo JH. Lumbar disc arthroplasty with Maverick disc versus stand-alone interbody fusion: a prospective, randomized, controlled, multicenter investigational device exemption trial. *Spine (Phila Pa 1976)* 36(25), E1600–E1611 (2011).
- 18 Noshchenko A, Hoffecker L, Lindley EM, Burger EL, Cain CM, Patel VV. Long-term treatment effects of lumbar arthrodeses in degenerative disc disease: a systematic review with meta analysis. *J. Spinal Disord. Tech.* 28(9), E493–E521 (2014).
- 19 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.* 134(2), 149–158 (2014).
- 20 Nie H, Chen G, Wang X, Zeng J. Comparison of total disc replacement with lumbar fusion: a meta-analysis of randomized controlled trials. *J. Coll. Physicians Surg. Pak.* 25(1), 60–67 (2015).
- **A recent meta-analysis comparing total disc replacement (TDR) with lumbar fusion that reports significantly better clinical and safety outcomes, including Oswestry Disability Index success, patient satisfaction and complication rate, with TDR than with lumbar fusion at 2-year follow-up.**
- 21 Jacobs WC, Van Der Gaag NA, Kruijt MC *et al.* Total disc replacement for chronic discogenic low back pain: a Cochrane review. *Spine (Phila Pa 1976)* 38(1), 24–36 (2013).
- 22 Yue JJ, Garcia R Jr., Miller LE. The activL(RR) artificial disc: a next-generation motion-preserving implant for chronic lumbar discogenic pain. *Med. Devices. (Auckl.)* 9, 75–84 (2016).
- 23 Garcia R Jr, Yue JJ, Blumenthal S *et al.* Lumbar total disc replacement for discogenic low back pain: two-year outcomes of the activL multicenter randomized controlled IDE clinical trial. *Spine (Phila Pa 1976)* 40(24), 1873–1881 (2015).
- **FDA IDE trial showing that activL was noninferior to ProDisc-L and Charite ($p < 0.001$) for the primary composite end point of treatment success and was superior for the primary composite end point in a protocol-defined analysis ($p = 0.02$) at 2-year follow-up.**
- 24 Guyer RD, Pettine K, Roh JS *et al.* Comparison of 2 lumbar total disc replacements: results of a prospective, randomized, controlled, multicenter Food and Drug Administration trial with 24 month follow-up. *Spine (Phila Pa 1976)* 39(12), 925–931 (2014).
- 25 Hutton B, Salanti G, Caldwell DM *et al.* The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. *Ann. Intern. Med.* 162(11), 777–784 (2015).
- **Recommended PRISMA guidance on reporting network meta-analyses of therapeutic interventions.**
- 26 Higgins JP, Altman DG, Gotzsche PC *et al.* The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 343, d5928 (2011).
- 27 Brown S, Hutton B, Clifford T *et al.* A Microsoft-Excel-based tool for running and critically appraising network meta-analyses—an overview and application of NetMetaXL. *Syst. Rev.* 3, 110 (2014).
- **Canadian Agency for Drugs and Technology in Health supported Microsoft Excel-based tool for conducting a WinBugs-based Bayesian network meta-analysis.**
- 28 Dias S, Welton NJ, Sutton AJ, Ades AE. NICE DSU Technical Support Document 2: a generalised linear modelling framework for pairwise and network meta-analysis of randomised controlled trials (2011). <http://scharr.dept.shef.ac.uk/nicedsu/wp-content/uploads/sites/7/2017/05/TSD2-General-meta-analysis-corrected-2Sep2016v2.pdf>
- 29 Dias S, Welton NJ, Sutton AJ, Caldwell DM, Lu G, Ades AE. Evidence synthesis for decision making 4: inconsistency in networks of evidence based on randomized controlled trials. *Med. Decis. Making* 33(5), 641–656 (2013).
- 30 Dias S, Welton NJ, Sutton AJ, Caldwell DM, Lu G, Ades AE. NICE DSU Technical Support Document 4: inconsistency in networks of evidence based on randomised controlled trials. (2011). <http://scharr.dept.shef.ac.uk/nicedsu/wp-content/uploads/sites/7/2016/03/TSD4-Inconsistency.final..15April2014.pdf>
- 31 Fu R, Selph S, Mcdonagh M *et al.* Effectiveness and harms of recombinant human bone morphogenetic protein-2 in spine fusion: a systematic review and meta-analysis. *Ann. Intern. Med.* 158(12), 890–902 (2013).
- 32 Jacobs W, Van Der Gaag NA, Tuschel A *et al.* Total disc replacement for chronic back pain in the presence of disc degeneration. *Cochrane Database Syst. Rev.* (9), CD008326 (2012).
- 33 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.* 37(7), 1315–1325 (2013).
- 34 Yajun W, Yue Z, Xiuxin H, Cui C. A meta-analysis of artificial total disc replacement versus fusion for lumbar degenerative disc disease. *Eur. Spine J.* 19(8), 1250–1261 (2010).
- 35 Schwind J, Learman K, O'halloran B, Showalter C, Cook C. Different minimally important clinical difference (MCID) scores lead to different clinical prediction rules for the Oswestry disability index for the same sample of patients. *J. Man. Manip. Ther.* 21(2), 71–78 (2013).
- 36 Ren C, Song Y, Liu L, Xue Y. Adjacent segment degeneration and disease after lumbar fusion compared with motion-preserving procedures: a meta-analysis. *Eur. J. Orthop. Surg. Traumatol.* 24(Suppl. 1), S245–S253 (2014).

- **A recent meta-analysis highlighting the significantly lower incidence of adjacent segment degeneration (ASD) and reoperations with TDR than with lumbar fusion (ASD: 12.2 vs 33.0%, $p < 0.001$; reoperations: 1.0 vs 7.8%, $p < 0.001$).**
- 37 Wang JC, Arnold PM, Hermsmeyer JT, Norvell DC. Do lumbar motion preserving devices reduce the risk of adjacent segment pathology compared with fusion surgery? A systematic review. *Spine (Phila Pa 1976)* 37(22 Suppl.), S133–S143 (2012).
- 38 Hiratzka J, Rastegar F, Contag AG, Norvell DC, Anderson PA, Hart RA. Adverse event recording and reporting in clinical trials comparing lumbar disk replacement with lumbar fusion: a systematic review. *Global Spine J.* 5(6), 486–495 (2015).
- 39 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* 17(6), 493–501 (2012).
- 40 Guyer RD, McAfee PC, Banco RJ *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.* 9(5), 374–386 (2009).
- 41 Gornet M, Dryer R, Pelozo J, Schranck F. Lumbar disc arthroplasty vs anterior lumbar interbody fusion: five-year outcomes for patients in the Maverick degree disc IDE study. *Spine J.* 10(9 Suppl. 1), S64 (2010).
- 42 Guyer RD, Pettine K, Roh JS *et al.* Five-year follow-up of a prospective, randomized trial comparing two lumbar total disc replacements. *Spine (Phila Pa 1976)* 41(1), 3–8 (2016).
- 43 Fritzell P, Berg S, Borgstrom F, Tullberg T, Tropp H. Cost effectiveness of disc prosthesis versus lumbar fusion in patients with chronic low back pain: randomized controlled trial with 2-year follow-up. *Eur. Spine J.* 20(7), 1001–1011 (2011).
- 44 Kurtz SM, Lau E, Iannuzzi A *et al.* National revision burden for lumbar total disc replacement in the United States: epidemiologic and economic perspectives. *Spine (Phila Pa 1976)* 35(6), 690–696 (2010).
- 45 Patel VV, Estes S, Lindley EM, Burger E. Lumbar spinal fusion versus anterior lumbar disc replacement: the financial implications. *J. Spinal Disord. Tech.* 21(7), 473–476 (2008).
- **Retrospective costing study showing that TDR patients had shorter hospital stays (2.6 vs 2.9–5.0 days) and shorter operating room times (222 vs 274–439) than fusion patients, leading to lower total hospital costs with TDR (\$27,972) than with fusion (\$32,167–\$44,633).**
- 46 Guyer RD, Tromanhauser SG, Regan JJ. An economic model of one-level lumbar arthroplasty versus fusion. *Spine J.* 7(5), 558–562 (2007).
- **An economic study reporting substantially lower costs with TDR than with fusion, from 12.0 to 36.5% lower index hospitalization costs to 82.8–99.0% lower total costs, including index hospitalization and subsequent costs up to 2-year follow-up.**
- 47 Parkinson B, Goodall S, Thavaneswaran P. Cost-effectiveness of lumbar artificial intervertebral disc replacement: driven by the choice of comparator. *ANZ J. Surg.* 83(9), 669–675 (2013).
- 48 Johnsen LG, Hellum C, Storheim K *et al.* Cost-effectiveness of total disc replacement versus multidisciplinary rehabilitation in patients with chronic low back pain: a Norwegian multicenter RCT. *Spine (Phila Pa 1976)* 39(1), 23–32 (2014).