Washington State Health Care Authority

# Artificial Disc Replacement for Cervical and Lumbar Degenerative Disc Disease

## Assessing Signals for Update

*July 18, 2024*

Health Technology Assessment Program (HTA)

**Washington State Health Care Authority**

PO Box 42712
Olympia, WA 98504-2712
(360) 725-5126
www.hca.wa.gov/hta
shtap@hca.wa.gov

**Prepared by:**
RTI International–University of North Carolina Evidence-based Practice Center
Research Triangle Park, NC 27709 www.rti.org



This report is based on research conducted by the RTI–University of North Carolina Evidence-based Practice Center through a contract between RTI International and the State of Washington Health Care Authority (HCA). The findings and conclusions in this document are those of the authors, who are responsible for its contents. The findings and conclusions do not represent the views of the Washington HCA and no statement in this report should be construed as an official position of Washington HCA.

The information in this report is intended to help the State of Washington Health Care Authority make decisions. This report is not intended to be a substitute for the application of clinical judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical reference and in conjunction with all other pertinent information (i.e., in the context of available resources and circumstances presented by individual patients).


None of the individuals involved in producing this report reported any financial or nonfinancial conflicts of interest regarding the topic presented in this report.

**Acknowledgments**

The following individuals contributed to this report:
*Lead Investigator:* Shannon Kugley, MLIS
*Co-Investigator:* Leila Kahwati, MD MPH

# Contents

## Abbreviations

**ACDF** Anterior cervical discectomy and fusion

**ADR** Artificial disc replacement

**AHRQ** Agency for Healthcare Research and Quality

**CADR** Cervical artificial disc replacement

**CI** Confidence interval

**DDD** Degenerative disc disease

**FDA** United States Food and Drug Administration

**HTA** Health technology assessment

**HTCC** State of Washington Health Technology Clinical Committee

**IDE** Investigational device exemption

**LADR** Lumbar artificial disc replacement

**LBP** Low back pain

**LCD** Local Coverage Determination

**NA** Not applicable

**NCD** Medicare National Coverage Determination

**NDI** Neck Disability Index

**NR** Not reported

**ODI** Oswestry Disability Index

**RCT** Randomized controlled trial

**RR** Risk ratio

**SR** Systematic review

**SF-36** Short-Form 36 questionnaire

**VAS** Visual Analogue Scale

# Executive Summary

## Background

The State of Washington's Health Technology Assessment Program published a health technology assessment (HTA) titled "Artificial disc replacement – Re-review Final evidence report" in December 2016[1] to assess the evidence for efficacy, harms, differential effects and costs of artificial disc replacement compared with other therapies including non-operative therapy, spinal fusion, and other surgery. Findings indicated that cervical artificial disc replacement (CADR) may be superior to anterior cervical discectomy and spinal fusion (ACDF) for safety and efficacy based on low to moderate quality evidence.[1] Based on these findings, the State of Washington Health Technology Clinical Committee (HTCC) issued a *covered benefit with conditions* determination for CADR.[2] The 2016 HTA reported that lumbar artificial disc replacement (LADR) was comparable to single level anterior lumbar interbody fusion or circumferential lumbar fusion in the treatment of symptomatic DDD at 24 and 60 months based on low quality evidence, but noted that the efficacy of the comparator treatment, lumbar fusion, is uncertain leading to an HTCC non-coverage determination for LADR.[2]

## Methods

To determine whether there is a need for an update to the 2016 HTA and reconsideration of the 2017 HTCC coverage decision, we conducted a signal search using a modified Ottawa approach,[3,4] relying on recent systematic reviews (SRs) to detect a signal for an update. We searched MEDLINE for relevant studies published in or after 2016 that addressed the research questions and study selection criteria used in the original HTA report. The search retrieved numerous relevant systematic reviews (SRs); therefore, we narrowed our assessment to SRs published in the last 5 years. We selected 5 SRs (3 for CADR[5-7] and 2 for LADR[8,9]) that were the most comprehensive and recent for determining a signal for update.

## Results

### *CADR*

We reviewed findings from 3 recent SRs for effectiveness and harms of 1- or 2-level CADR. [5-7] All included studies comparing CADR with ACDF or disc replacement with an alternative cervical artificial disc. We relied mainly on findings from an Agency for Healthcare Research and Quality (AHRQ)-commissioned SR published in 2023[7] because the review authors evaluated outcomes and harms for 1- and 2- level disc disease separately and by short, intermediate, and long-term follow-up, making a direct comparison with the 2016 HTA possible. The other 2 reviews provided supporting information: 1 SR assessed the statistical fragility of the evidence base,[6] and the other focused on mid- and long-term outcomes.[5] The AHRQ review concluded that 1- and 2-level CADR is comparable to ACDF for pain and function, based on moderate strength of evidence for both outcomes with similar or fewer harms.[7] The SR designed to assess statistical fragility[6] reported no consensus regarding superiority of CADR or ACDF for pain, disability, or adjacent segment disease (ASD, a complication that occurs from fusion procedures), but did conclude that the evidence has fair to moderate statistical robustness and is

not vulnerable to statistical fragility, meaning that inclusion of future studies is unlikely to change the findings.[6] The third review that focused on longer term outcomes reported no differences between CADR and ACDF for arm pain, neck pain, and harms, but CADR was superior to ACDF for neck disability index, quality of life, with fewer reoperations and less ASD.[5]

### *LADR*

We reviewed the findings from 2 SRs for effectiveness and harms of 1- or 2-level LADR.[8,9] One included RCTs and cohort studies with at least 5 years follow-up but limited reporting to changes from pre- to postoperative in the LADR study arm only.[9] Pain and disability outcomes improved significantly from baseline among individuals treated LADR.[9] However, as this review included studies without a comparison group, it is difficult to draw clear conclusion on the benefit or harms of treatment. This review also included an RCT evaluating LADR versus multidisciplinary rehabilitation with data at 8 years follow-up that is new since the 2016 HTA.[10] In intention-to-treat analyses, disability scores improved more in the LADR groups compared with the rehabilitation group at 2 years[11] and the improvement persisted at 8-year follow-up[10] but the treatment effects fell short of a clinically meaningful difference. The other review included RCTs that compared LADR with lumbar fusion.[8] Based on results of 10 RCTs, the review reported that LADR is superior to lumbar fusion for disability, pain, quality of life, patient satisfaction, overall success, reoperation rate, and complications.[8]

## Conclusions

Our assessment of the evidence for CADR suggests that that cervical disc replacement remains at least as effective as ACDF, the current standard of care, with similar or fewer harms. Although 2 cervical disc replacement devices have been approved since the 2016 HTA, we conclude there is no signal suggesting the need for an update.

Several new studies comparing LADR to lumbar fusion suggests that disc replacement may be more effective than fusion with similar or fewer harms. In comparison to multidisciplinary rehabilitation, which may be a more appropriate comparator, new long-term follow-up evidence from an RCT with some methodological limitations suggests LADR may be superior, but the differences may not be clinically meaningful. No new lumbar disc replacement devices have been approved since the 2016 HTA. Thus, we conclude there is no signal suggesting the need for an update as long as fusion is not considered a relevant comparator.

# 1. Introduction

Low back and neck pain due to degenerative disc disease (DDD) is a frequent cause of disability in adults in the United States, accounting for a large proportion of health care expenditures to evaluate and treat the condition.[12] Because aging is the primary risk factor for development of DDD, as the US population ages, the incidence of DDD is expected to increase.

First-line treatment is usually nonsurgical, such as physical therapy, epidural steroid injections, and medications. However, an estimated 10% to 20% of people with lumbar DDD and up to 30% with cervical DDD are unresponsive to nonsurgical treatment. Surgery may be considered when adequate trials (e.g., at least 6 months) of nonoperative treatments have failed to relieve symptoms attributed to DDD or to prevent progression of nerve damage from radiculopathy or myelopathy. For lumbar DDD that has been non-responsive to conservative treatment, fusion (i.e., arthrodesis) is sometimes used for chronic lumbar pain without radiculopathy, myelopathy, or disc prolapse. The goal of spinal fusion is to remove the disc and fuse the vertebrae, thereby limiting the motion at the symptomatic segment. Spinal fusion may promote degeneration of the vertebrae above or below the fusion site (adjacent segment disease, ASD); however, uncertainties remain regarding the extent to which this occurs.[13-15] Guidelines recommend intensive multidisciplinary rehabilitation and appropriate patient selection as an integral part of decision making particularly for lumbar fusion.[16,17] For cervical DDD, which often involves radiculopathy, the current standard of care is anterior cervical discectomy and spinal fusion (ACDF), the goal of which is nerve decompression and restoration of spinal alignment and stability.[18]

A surgical alternative to fusion is artificial disc replacement (ADR).[19] Disc prostheses mimic the decompressive and supportive properties of intervertebral discs and preserve motion at the index level, thereby improving pain and function as well as decreasing stress on adjacent segment structures and theoretically the risk of ASD. Lumbar disc arthroplasty, also known as lumbar total disc replacement or lumbar artificial disc replacement (LADR) is a surgical procedure on the lumbar spine that involves complete removal of the damaged or diseased lumbar intervertebral disc and implantation of an artificial disc[20,21] while cervical disc arthroplasty, or cervical artificial disc replacement (CADR), involves removal and replacement of the cervical intervertebral disc.[22] These procedures may be done as an alternative to lumbar or cervical spinal fusion and are intended to reduce pain, preserve motion at the site of surgery, and restore intervertebral disc height.[21]

### *Cervical Disc Replacement*

The U.S. FDA issued the first premarket approval (PMA) for a cervical disc replacement device, the Prestige ST, in 2007. Since then, numerous disc replacement designs (Product Code “mjo” in FDA’s PMA Database) have been developed and have received FDA approval for 1- and 2-level disease (9 devices currently listed in Appendix A).[23] Two devices for cervical disc replacement have been newly approved since the 2016 HTA: Simplify (NuVasive) and M6-C (Spinal

Kinetics). The FDA issued an original PMA for Simplify on 9/18/2020[24] and for M6-C on 2/6/2019.[25] There is no Center for Medicare & Medicaid Services (CMS) National Coverage Determination for cervical disc replacement; however, 1 Medicare Administrative Contractor (MAC) has issued a Local Coverage Determination (LCD) for cervical disc replacement with FDA-approved devices in skeletally mature patients who meet certain clinical conditions, effective 10/28/2019.[26]

***Lumbar Disc Replacement***

The Activ-L lumbar disc replacement system (Aesculap Implant Systems, Center Valley, Pennsylvania)[27] and ProDisc-L (Centinel Spine, West Chester, Pennsylvania)[28] are the only FDA-approved lumbar artificial disc products available in the U.S (**Appendix B**). For lumbar disc replacement in patients ***older than age 60*** years, there is a CMS National Determination for non-coverage.[29] There is no national coverage determination for patients aged ***60 years and younger***. For beneficiaries aged 60 years and younger, CMS leaves the coverage determination to local MACs. One local MAC (Palmetto GBA) has a LCD specifying non-coverage for LADR in beneficiaries 60 years of age and younger effective 6/17/2021.[30]

## 1.1 Policy Context

The Health Technology Clinical Committee (HTCC) considered the evidence for lumbar and cervical disc replacement from the 2016 HTA.[1] For CADR, the committee issued a *covered benefit with conditions* determination and for LADR they issues a *not-covered benefit* determination.

***HTCC coverage determination for CADR***

CADR is a covered with conditions consistent with the criteria identified in the reimbursement determination (March 17, 2017).[2] The conditions for coverage include that patients must meet FDA approved indications for use and not have any contraindications. FDA approval is device specific but includes:

- Skeletally mature patients
- Disc replacement following 1- or 2-level discectomy for intractable symptomatic radiculopathy or myelopathy confirmed by patient findings and imaging.
- Patients must have advanced imaging and clinical evidence of corresponding nerve root or spinal cord compression and have failed or be inappropriate for non-operative care.
- For 2-level procedures, objective evidence of radiculopathy, myelopathy or spinal cord compression at 2 consecutive levels is required.

The rationale for the committee's decision was as follows:

> "The committee reviewed and discussed the available studies of cervical artificial disc replacement. Details of study design, inclusion criteria and other factors affecting study quality were discussed. A majority of committee members found the evidence sufficient to determine that cervical artificial discs replacements were at least equivalent for safety and effectiveness compared to alternatives for some conditions, and unproven for cost-effectiveness. A majority of the committee voted to cover with conditions, cervical artificial disc replacement."

### *HTCC coverage determination for LADR*

LADR is not a covered benefit.[2] The rationale for the committee's decision was as follows:

> "The committee reviewed and discussed the available studies of lumbar artificial disc replacement. Details of study design, inclusion criteria and other factors affecting study quality were discussed. A majority of committee members found the evidence sufficient to determine that lumbar artificial discs replacements were unproven for safety and unproven for effectiveness compared to alternatives for some conditions, and unproven for cost effectiveness. A majority of the committee voted to not cover lumbar artificial disc replacement."

## 1.2 Scope and Key Questions of the 2016 HTA

The key questions guiding the previous HTA and signal searches on this topic are listed below.

**Key Question 1 (KQ1):** What is the evidence of efficacy and effectiveness of artificial disc replacement (ADR) compared with other therapies including non-operative therapy, spinal fusion, and other surgery?

**Key Question 2 (KQ2):** What is the evidence related to the ADR safety profile including device failure and reoperation for ADR?

**Key Question 3 (KQ3):** What is the evidence of differential efficacy or safety amongst special populations including but not limited to the elderly and workers compensation populations?

**Key Question 4 (KQ4):** What are the cost implications and cost effectiveness for ADR?

The full inclusion and exclusion criteria is in **Appendix C** and are summarized as follows:

### *Population*

*Lumbar:* Patients undergoing primary LADR for DDD without neurological compromise and who have not had prior spine surgery at the instrumented level.

*Cervical:* Patients undergoing primary CADR for DDD resulting in radiculopathy or myelopathy and who have not had prior surgery at the instrumented level.

### *Intervention*

LADR or CADR with commercially available device (defined as FDA-approved devices or unapproved devices in Phase III trials with ≥ 1 year of follow-up data in a peer-reviewed journal).

### *Comparators*

Non-operative treatment, spinal fusion, other spine surgery. Comparator interventions that employ a device not FDA-approved for use in the US will be excluded.

### *Outcomes*

- Physical function/disability (overall clinical success, validated disability indices)
- Pain/pain reduction
- Device failure (reoperation at the index level – to include revision, reoperation, or removal)
- Complications (e.g., migration, subsidence, neurologic injury as well as infection, vascular damage, heterotopic ossification, others)

### *Key Findings From 2016 HTA*

The 2016 HTA reported outcomes stratified by duration of follow-up based on commonly reported time points: 24 months, 48 to 60 months, and 84 months. Results were reported separately for CADR and LADR and for 1- and 2-level arthroplasty. The report included 24 randomized controlled trials (RCTs), 14 non-randomized comparative studies and 9 economic evaluations. A detailed summary of findings from the 2016 HTA are in Appendix D.

For CADR, a total of 19 RCTs (in 49 publications), 9 non-randomized comparative studies, and 6 economic evaluations were included. The 2016 HTA reported that CADR appears to be superior to ACDF for overall success at 24 months and at 48 to 60 months for 1- and 2-level replacement (GRADE: moderate quality of evidence) and up to 84 months for 1-level (GRADE: low quality of evidence). Evidence of overall success at 84 months for 2-level replacement was insufficient.

For LADR, a total of 5 RCTs (in 11 publications), 5 non-randomized comparative studies, and 3 economic evaluations were included. The 2016 HTA reported that LADR was comparable to single level anterior lumbar interbody fusion or circumferential lumbar fusion in the treatment of symptomatic DDD at 24 and 60 months (GRADE: low quality evidence).[1] There was low quality evidence indicating a similar rate of secondary surgery, major complications, and adverse events. Authors noted that sample sizes may have precluded detection of such events and their frequency may be underestimated. One of the 2 investigational device exemption (IDE) trials for LADR included in the 2016 HTA report, reported no serious adverse events (SAEs) at 60 months; the other IDE trial reported that serious life-threatening events were more common in the LADR group (mean 0.58 per patient) compared with fusion (mean 0.38 per patient). Two level LADR was also similar to fusion for efficacy (GRADE: low quality evidence) and similar or superior to fusion for safety (GRADE: low quality evidence).

[1] Although results suggest that 24-month outcomes for LADR are similar to lumbar fusion, for non-inferiority trials the assumption is that reference treatment must have an established efficacy or that it is in widespread use. For the lumbar spine, the efficacy of the comparator treatment, lumbar fusion, for DDD remains uncertain, especially when it is compared with nonoperative care. Given what is known about lumbar fusion as a comparator and having evidence that only compares LADR with lumbar fusion limits the ability to fully answer the efficacy/effectiveness question for LADR.

## 1.3 Objectives

The primary aim of this signal search was to determine whether or not there is new evidence on the efficacy, safety, or cost-effectiveness of cervical or lumbar ADR that will change the conclusions of the most recent health technology assessment (HTA) or HTCC coverage decision.

# 2. Methods

We used a modified Ottawa approach[3,4] to determine whether a signal for an update was present, relying primarily on recent systematic reviews (SRs).

## 2.1. Literature Search

We searched MEDLINE® (via PubMed) for relevant English-language studies between January 1, 2016, and May 21, 2024, allowing an overlap of 6 months with the previous search. The search strategy is described in **Appendix E**. We limited the search to SRs and RCTs using database study design filters. In addition to PubMed, we searched AHRQ, CADTH, NICE, Cochrane, and ICER websites for reviews, guidelines, and consensus statements published in 2016 or later.

## 2.2. Study Selection

We used the same inclusion and exclusion criteria from the 2016 HTA (**Appendix C**). However, we also looked for evidence on the 2 new devices that were approved by the FDA for CADR since the prior 2016 HTA.

Because the search retrieved numerous SRs, we did not assess RCTs further. Instead, we sought SRs that included studies that met inclusion and exclusion criteria from the original 2016 HTA. We included SRs with broader inclusion and exclusion criteria if findings were reported separately for eligible studies. For example, we only abstracted and reported findings on comparative effectiveness and harms of CADR versus a comparator from a review that included evidence for other surgical and non-surgical interventions for cervical DDD.

We prioritized the most recently published SRs for abstraction and those with the largest number of included primary studies. Note, that when SRs have similar inclusion and exclusion criteria, newer SRs are likely to include the same primary research studies as older reviews, reducing the utility of also abstracting data from older SRs. We did not include studies that reported only on effects among subgroups (KQ3) or cost (KQ4). Because we relied on SRs to detect a signal, some of the abstracted outcomes include data from studies evaluating arthroplasty using artificial discs that are not FDA-approved or are not available in the U.S.

## 2.3. Data Abstraction and Signal Assessment

One reviewer evaluated titles and abstracts retrieved by our search; that same reviewer assessed full text systematic review articles to determine if they met selection criteria and reported relevant findings. We abstracted relevant data and stopped abstracting additional reviews once

we had enough information to determine whether a signal was present. We also prioritized abstraction of high impact SRs (e.g., Agency for Healthcare Research and Quality). We abstracted the condition, intervention, comparator, included studies (i.e., number of studies, study designs, and publication years), the outcomes (from the list of eligible outcomes), and review findings. We abstracted the name(s) of the artificial disc evaluated when reported by the review authors. We abstracted effect size and confidence intervals or numerical differences and variance when reported.

Because CADR is currently a covered benefit with conditions, we determined our signal assessment for CADR primarily based on identifying evidence suggesting new harms (KQ2). Because LADR is not currently a covered benefit, we determined our signal assessment for LADR based on identifying new evidence indicating effectiveness (KQ 1) with emphasis on comparisons of LADR to treatment other than lumbar fusion. This is because lumbar fusion for uncomplicated DDD is not a covered benefit and fusion is not considered an appropriate comparator for evaluating effectiveness of LADR.

# 3. Results

## 3.1. Search Yield and Overview of Studies

Our search in PubMed that was limited to RCTs or SRs retrieved 156 records. We prioritized 30 SRs (24 cervical, 5 lumbar, 1 both) published in or after 2019 for signal assessment. We selected 5 SRs,[5-9] 3 for cervical[5-7] and 2 for lumbar[8,9] arthroplasty that were the most comprehensive and current for data abstraction.

## 3.2. Study Characteristics

Search dates used in the 5 SRs spanned from 1980 to 2023 with publications dates of primary studies from 2005 to 2023. Three of the reviews limited inclusion to RCTs only.[5,6,8] Most assessed the methodological quality of the included studies and 2 reviews assessed the strength of evidence for at least some of the outcomes reported.[5,7] One reported overall direction of effect and number of studies,[6] 3 reported effect sizes,[5,7,8] and 1 included pre- and post-surgery change in outcomes.[9] One review was commissioned by AHRQ and was sponsored by a professional organization. The other reviews were conducted by individual authors affiliated with academic institutions located in the U.S.,[5-7] Australia,[9] and China.[8] Below we provide a summary of the SRs. Detailed information on the reviews is reported separately in **Appendix F** for CADR (Table F-1) and LADR (Table F-2).

### *CADR*

The most recently published review (Ortiz-Babilonia et al, 2024)[6] evaluated 25 studies of CADR versus ACDF published between 2007 and 2021. Outcomes assessed included overall pain, neck disability index (NDI), radicular arm pain, disability, modified Japanese Orthopedic Association (mJOA) score, and ASD (overall, superior-level, and inferior-level). The review authors did not calculate effect sizes as the objective of the review was to assess statistical fragility. Authors reported the number of studies that favored CADR or ACDF by outcome.

The review by Selph and colleagues (2023)[31] was commissioned by the AHRQ as part of the Effective Healthcare Program. The topic was nominated by the Congress of Neurological Surgeons (CNS) to inform guidelines on surgical management of DDD. The review included studies published between 1980 and 2023 that evaluated surgical interventions[2] compared with non-surgical interventions[3] or other surgical interventions for symptomatic cervical DDD (e.g., DDD with pain, radiculopathy, myelopathy). For the comparison of CADR versus ACDF, the review included 22 RCTs, 2 multicenter FDA IDE trials of newer cervical arthroplasty devices (M6-C and Simplify) versus historic ACDF controls, 8 non-IDE non-randomized studies, 7 database/registry studies, and 1 post-hoc analysis of an FDA IDE trial. Notably, the review included evidence from a 7-year follow up of an IDE RCT of Secure-C and a follow-up of an a non-randomized study of an intervention (NRSI) of MOBI-C,[32]contributing data to long-term outcomes and harms that were not available when the 2016 HTA was published.[33] The review assessed outcomes by 1-level, 2-level or multi-level disc replacement. Authors assessed effectiveness and harms for the presence of at least a small clinically meaningful difference. That is, a statistically significant findings with a less than small clinical effect were assessed as no difference for strength of evidence (SOE) assessments. The AHRQ-commissioned review included 1 RCT of CADR with an unnamed device versus ACDF that reported 3 years followup[34] and 6 RCTs[32,33,35-38] reporting long-term outcomes that were not in the 2016 HTA.

Finally, a review by Kim et al. (2023)[5] evaluated CADR compared with ACDF for symptomatic cervical DDD. The review included 14 RCTs published between 2014 and 2023 and conducted meta-analyses of pain, quality of life, ASD, SAEs, and reoperation outcomes.

### *LADR*

One of the 2 included SRs of LADR (Wen et al) focused on long-term outcomes (minimum of 5 years follow-up) following LADR.[9] This review included 22 studies (7 RCTs, 15 cohort studies) published between 2012 and 2021, of which, 10 were published in or after 2017. The focus of this review was on reporting pre- to post-operative changes among the study group receiving LADR and differences in these changes based on type of artificial disc. Among the included studies was an article reporting on 8-year follow-up of a study comparing LADR with multidisciplinary rehabilitation for chronic low back pain and localized degenerative changes in the lumbar intervertebral discs.[10] The comparators evaluated in the rest of the studies were not reported.

A second review (Bai et al) published in 2019 limited inclusion to RCTs, assessed risk of bias, conducted a meta-analysis, and assessed strength of evidence.[8] Outcomes included visual analog scale (VAS) for pain, Oswestry disability index (ODI), reoperation and overall success. The comparator in these studies was lumbar fusion. One of the 10 RCTs, a 5-year follow-up of

[2] Eligible surgical interventions included: discectomy, disc replacement, fusion up to T2, cervical arthroplasty, laminectomy, laminoplasty, corpectomy, cervical hybrid surgery, foraminotomy, ACDF cage vs. ACDF cage + plate
[3] Non-surgical comparators included: heat, exercise, acupuncture, drugs, radiofrequency ablation, steroid injections, Botox for neck pain, psychological strategies [e.g., cognitive behavioral therapy], occupational therapy, multidisciplinary rehabilitation).

reoperation rates following 2-level lumbar disk replacement with the ProDisc-L,[39] was published after the 2016 HTA.

## 3.3. Signal Findings

### *CADR*

For assessment of a signal for CADR, we examined findings from the 3 recent SRs that compared CADR with another surgical intervention, usually, ACDF or an alternative cervical artificial disc.[5-7] We did not find any reviews that included studies evaluating CADR compared with conservative treatment. We focus on the findings from the comprehensive AHRQ-commissioned SR[7], which systematically evaluated outcomes and harms for 1- and 2- level disc disease independently and by short, intermediate, and long-term, making a direct comparison with the 2016 HTA possible (**Appendix G, Tables G1-G5**). We include findings from the other 2 reviews[5,6] as supporting information, primarily because they did not separate results for 1-level and 2-level CADR to allow for direct comparison to the prior 2016 HTA and because they synthesized fewer outcomes.

As compared to the 2016 HTA, the AHRQ review includes several new studies across multiple outcomes. This review concludes that 1- and 2-level CADR is comparable to ACDF for pain and function, based on moderate strength of evidence for both outcomes.[7] We note that this review found several statistically significant differences that favored CADR compared with ACDF but review authors only considered CADR more effective if the difference observed was equal to or greater than a clinically meaningful threshold. Considering this, we believe the AHRQ review's findings related to effectiveness align with those of the 2016 HTA report.

For harms (KQ2), the AHRQ review reported that 1-level CADR was superior (large effect) to ACDF for reoperation (high SOE) and was superior (small effect) to ACDF for SAEs (low SOE).[7] Neurological events were similar for 1-level CADR and ACDF (low strength of evidence).[7] Study findings for pain and function were similar in patients with 2-level cervical arthroplasty or ACDF. Reoperation was more likely following 2-level CADR versus ACDF; however, the findings are based on low strength of evidence and the authors noted that the indication for reoperation was not consistently described and the potential impact on re-operation at index level for plate removal to treat ASD is unknown.[7] The likelihood of an adverse event was slightly lower at 24 with months and not different at 120 months based on low strength of evidence for 2-level CADR versus ACDF.[7] There was insufficient evidence on neurological events for 2-level arthroplasty Evidence was sparse for this comparison beyond 2 levels.[7]

The review by Ortiz-Babilonia et al. (2024)[6] was designed to assess statistical fragility and did not report effect sizes. Authors concluded that there is no consensus regarding superiority of CADR or ACDF for pain, disability, or ASD because most of the included studies found no significant difference between interventions. The authors did, however, report that the evidence appears to have fair to moderate statistical robustness and is not vulnerable to statistical fragility, which suggests that the inclusion of future studies is unlikely to change the findings.[6]

The 2023 meta-analysis of data from 14 RCTs by Kim et al. reported that CADR was superior to ACDF for neck disability index, quality of life, reoperation and ASD.[5] There were no differences between CADR and ACDF for arm pain, neck pain, and AEs.[5]

### *LADR*

Unlike the 2016 HTA, the 2 SRs of LADR did not report outcomes separately for 1- and 2-level procedures; nonetheless, findings were not remarkably dissimilar, thus we included outcomes for 1- and 2-level LADR versus comparators (fusion unless otherwise specified) in a single table for effectiveness (**Appendix**

**Table G-6**) and for harms (**Appendix**

**Table G-7**). Based on these 2 reviews, there have been no new comparative studies evaluating the effectiveness or safety of LADR versus conservative treatment. However, we did identify one new article reporting long-term follow-up from an RCT evaluating LADR versus multidisciplinary rehabilitation that was included in the 2016 HTA (**Appendix**

**Table G-8**).[10]

The review by Bai et al. 2019 reported that LADR is superior to lumbar fusion for disability (ODI), pain (VAS), quality of life (SF-36), patient satisfaction, overall success, reoperation rate, and complications based on the evidence from 10 RCTs.[8] Authors assessed all outcomes as moderate strength of evidence except for complications (low SOE).[8] Only 1 of the included articles was published after 2015 (Radcliff, 2018) and it only reported on reoperation rates.[39]

The review by Wen et al. (2024) included studies with a minimum follow-up of 5 years to examine efficacy and safety of LADR.[9] While several of these studies are newly published since the 2016 HTA, the new studies are either head-to-head comparisons of alternative disc devices or single arm studies reporting outcomes without any comparator. The review reported clinical outcomes from 22 studies including 7 RCTs.[9] The review authors found that pain (VAS) and disability (ODI) outcomes improved significantly from baseline to post-operatively among individuals treated with LADR (mean improvement in VAS pain 44.7, SD 10.7; mean improvement in ODI 30.8, SD 9.4).[10] However, as this review included studies with and without a comparison group and only reported pre/post changes, it is difficult to draw clear conclusion on the benefit of treatment in relationship to other procedures or conservative treatment.

One of the included reviews included long term data on ProDisc II (no longer available as has been superseded by newer versions) from an RCT of LADR versus multidisciplinary rehabilitation.[10] Outcomes are summarized in **Appendix D,**

**Table G-8**. The 2-year outcomes from this RCT[11] were included in the 2016 HTA. Outcomes at 8 years from this study were reported in a 2017 publication[10] that was not included in the prior HTA. In intention-to-treat analyses, disability scores improved more in the LADR groups compared with the rehabilitation group at 2 years (mean improvement in ODI from baseline: 20.9 points and 12.4 points, respectively; difference between groups -8.4 favoring surgery; 95% CI, -13.2 to -3.6[11]) and the improvement persisted at 8-year follow-up (mean difference -6.1 favoring surgery; 95% CI -11.0 to -1.2[10]). We note that a clinically meaningful difference in the ODI is generally considered about 10 points; thus, the treatment effects observed fall short of a clinically meaningful difference. Nearly a quarter of participants randomized to multidisciplinary rehabilitation crossed over to surgery (LADR, fusion, or discectomy), some participants randomized to LADR also received multidisciplinary rehabilitation, and several participants in both groups received additional types of lumbar surgery (e.g., fusion, discectomy). Thus, the treatment effect observed in the intention-to-treat analyses is likely biased towards a null effect.

# 4. Discussion and Conclusions

The following artificial discs for the cervical spine were represented by the primary studies included in the 3 reviews of cervical arthroplasty: ActivC, Bryan, Discover, Kineflex-C, Mobi-C, PCM, Prestige LP, ProDisc-C, and Secure-C. Of these, ActiveC, Bryan, Discover, Kineflex-C, PCM, and Prestige are either not FDA-approved or are no longer available commercially in the U.S. One review[7] included reports from 2 completed prospective, multicenter FDA IDE trials that evaluated newer cervical arthroplasty devices (M6-C and Simplify discs) compared with historic ACDF controls. These devices received an FDA PMA after the HTA 2016 was completed. Several new studies are available from this signal search comparing CADR to ACDF and the evidence suggests CADR is at least as effective as ACDF, with similar or fewer harms. Thus, despite new evidence and new devices, the conclusions of the prior 2016 HTA are stable, and we did not identify any signal for an update.

The following artificial discs for the lumbar spine were represented by the primary studies included in the 3 reviews of lumbar arthroplasty: AcroFlex, ActivL, Charité, Flexicore, Kineflex, M6L, Maverick, ProDisc-L, and XL-TDR. Of these, only ActivL and ProDisc-L are FDA approved and commercially available in the U.S. One previously included RCT has new longer-term follow-up data assessing the effectiveness of LADR compared to conservative treatment; although LADR appears more effective, the difference does not exceed current thresholds for a clinically meaningful difference at 2 or 8 years of follow-up. Other new studies are available suggesting that LADR may be more effective than fusion, with fewer complications. However, if fusion is not considered a relevant comparator, then this new evidence would not constitute a signal for an update.

## 4.1 Limitations

This signal search has several limitations. First, we searched a single electronic database (PubMed); therefore, we may have missed relevant SRs or studies published in journals not indexed in PubMed. Second, we conducted a limited data abstraction and assessment of the evidence reported in the most recent SRs; we did not conduct risk-of-bias assessments of the reviews we identified or of the primary studies included in those reviews. We also did not perform GRADE certainty of evidence assessments. Third, the included SRs may have used broader inclusion and exclusion criteria than the 2016 HTA. Not all the included reviews specified the artificial disc evaluated in the included studies and most of the reviews did not restrict eligibility to studies of FDA-approved, commercially available devices.

## 4.2 Conclusions

Two new cervical disc replacement devices have been approved since the previously published 2016 HTA. Evidence that has accumulated since the 2016 HTA suggests that CADR remains at least as effective as ACDF, the current standard of care, with similar or fewer harms. Thus, we conclude there is no signal for an update concerning artificial cervical disc replacement.

No new LADR devices have been approved since the previous 2016 HTA. Several new studies comparing lumbar disc replacement to lumbar fusion suggests that disc replacement may be more effective than fusion with similar or fewer harms. In comparison to multidisciplinary rehabilitation, a more appropriate comparator, new evidence from long-term follow-up suggests the difference between lumbar disc replacement and multidisciplinary rehabilitation may not be clinically meaningful; however, this study has some limitations. Overall, we conclude there is no signal for an update concerning artificial lumbar disc replacement.

# 5. References

# Appendixes

## Appendix A. FDA-approved Cervical Artificial Discs

| Year | Name | Manufacturer | Level |
|---|---|---|---|
| 2007 | Prestige® Cervical Disc | Medtronic Sofamor Danek, Memphis, TN | 1-level |
| 2007 | ProDisc™-C Total Disc Replacement | Centinel Spine, West Chester, PA | 1 level |
| 2009 | Bryan® Cervical Disc System | Medtronic Sofamor Danek, Memphis, TN | 1-level |
| 2012 | SECURE-C® Artificial Cervical Disc | Globus Medical, Audubon, PA | 1-level |
| 2012 | PCM® Cervical Disc | NuVasive, San Diego, CA | 1-level |
| 2013 | Mobi-C® Cervical Disc Prosthesis | Highridge Medical, Westminster, Co | 2-level |
| 2014 | PRESTIGE® LP Cervical Disc | Medtronic Sofamor Danek, Memphis, TN | 1-level |
| 2019 | M6-C Cervical Disc[4] | Orthofix Spinal Kinetics, Sunnyvale, CA | 1-level |
| 2021 | Simplify® Cervical Artificial Disc[1] | NuVasive, San Diego, CA | 2-level |

U.S. Food and Drug Administration (FDA). Prestige® cervical disc system. Summary of Safety and Effectiveness Data. PMA P060018. Rockville, MD: FDA; July 16, 2007. Accessed 6/17/2024. https://www.accessdata.fda.gov/cdrh_docs/pdf6/p060018c.pdf

U.S. Food and Drug Administration (FDA). ProDisc™-C Total Disc Replacement. Summary of Safety and Effectiveness Data. PMA No. 070001. Rockville, MD: FDA; December 17, 2007. Accessed 6/17/2024. https://www.accessdata.fda.gov/cdrh_docs/pdf7/p070001b.pdf

U.S. Food and Drug Administration (FDA). BRYAN® cervical disc. Summary of Safety and Effectiveness Data. PMA P060023. Rockville, MD: FDA; May 12, 2009. Accessed 6/17/2024. https://www.accessdata.fda.gov/cdrh_docs/pdf6/P060023b.pdf

U.S. Food and Drug Administration (FDA). SECURE-C® cervical artificial disc. Summary of Safety and Effectiveness Data. PMA P100003. Rockville, MD: FDA; September 28, 2012. 6/17/2024. https://www.accessdata.fda.gov/cdrh_docs/pdf10/p100003b.pdf

U.S. Food and Drug Administration (FDA). PCM® Cervical Disc. Summary of Safety and Effectiveness Data. PPMA P100012. Rockville, MD: FDA; October 26, 2012. Accessed 6/17/2024. https://www.accessdata.fda.gov/cdrh_docs/pdf10/P100012B.pdf

U.S. Food and Drug Administration (FDA). Mobi-C® Cervical Disc Prosthesis (two-level) - P110009. Silver Spring, MD: FDA; August 23, 2013. Accessed 6/17/2024. https://www.accessdata.fda.gov/cdrh_docs/pdf11/P110009B.pdf

U.S. Food and Drug Administration (FDA). Prestige PMA P090029. Silver Spring, MD: FDA; July 24, 2014. Accessed 6/24/2024. https://www.accessdata.fda.gov/cdrh_docs/pdf9/P090029A.pdf

[4] Approved since the 2016 HTA for artificial disc replacement

U.S. Food and Drug Administration (FDA). M6-C™ Artificial Cervical Disc. PMA P170036. Silver Spring, MD: FDA; February 6, 2019. Accessed 6/17/2024. https://www.accessdata.fda.gov/cdrh_docs/pdf17/P170036A.pdf

U.S. Food and Drug Administration (FDA). Simplify® Cervical Artificial Disc. PMA P200022/S003. Silver Spring, MD: FDA; April 1, 2021. Accessed 6/17/2024. https://www.accessdata.fda.gov/cdrh_docs/pdf20/P200022S003A.pdf

## Appendix B. FDA-approved and Available Lumbar Artificial Discs

| Year | Name | Manufacturer | Level |
|---|---|---|---|
| 2006 (1-level); 2020 (2-level) | ProDisc-L Lumbar Disc | Depuy Spine, Raynham, MA | 2-level |
| 2015 | ActivL Artificial Disc | Aesculap Implant Systems; Center Valley, PA | 1-level |

**Note:** The ActivL artificial disc the ProDisc-L are the only commercially available FDA-approved devices for lumbar disc arthroplasty. Other devices (i.e., Maverick™, FlexiCore®, KineFlex-L™, and AcroFlex) completed premarket approval studies but withdrew before FDA consideration or are not sold commercially in the United States. The Charité disc had FDA approval, but is no longer sold in the U.S.

U.S. Food & Drug Administration (FDA). Summary of Safety and Effectiveness Data (SSED): CHARITE Artificial Disc. http://www.accessdata.fda.gov/cdrh_docs/pdf4/P040006B.pdf2004

U.S. Food & Drug Administration (FDA). Summary of Safety and Effectiveness Data (SSED): PRODISC-L Total Disc Replacement. http://www.accessdata.fda.gov/cdrh_docs/pdf5/P050010B.pdf2006

U.S. Food & Drug Administration (FDA). Summary of Safety and Effectiveness Data (SSED): activL Artificial Disc (activL). http://www.accessdata.fda.gov/cdrh_docs/pdf12/p120024b.pdf2015

# Appendix C. PICOTS

| Study Component | Inclusion | Exclusion |
|---|---|---|
| Population | Lumbar Artificial Disc Replacement (LADR)<br>Patients undergoing primary lumbar artificial disc replacement for degenerative disc disease (DDD)<br>Cervical Artificial Disc Replacement (CADR)<br>Patients undergoing primary CADR for DDD resulting in radiculopathy or myelopathy | Neurological compromise<br>Prior spine surgery at the instrumented level<br>Patients with contraindications to receive LADR or CADR<br>ADR in the thoracic spine |
| Intervention | Commercially available device (defined as FDA-approved devices or unapproved devices in Phase III trials with ≥ 1 year of follow-up data in a peer-reviewed journal) | Device not approved by the FDA or without at least one year of follow-up data from a Phase III trial<br>Disc nucleus replacement |
| Comparator | Non-operative treatment<br>Spinal fusion<br>Other spine surgery | No comparator<br>Non-FDA approved device |
| Outcomes | Physical function/disability<br>Pain/pain reduction.<br>Device failure including reoperation at the index level, revision, reoperation or removal<br>Complications (e.g., migration, subsidence, neurologic injury as well as infection, vascular damage, heterotopic ossification)<br>Quality of life<br>Incidence of adjacent segment disease | range of motion<br>alignment |
| Publication Type | Studies published in English<br>Studies published in a peer-reviewed journal<br>Publicly available FDA reports<br>LADR: Summary of Safety and Effectiveness Data (SSED), In-depth Statistical Review, In-depth Clinical Review<br>CADR: SSED, Executive Summary of FDA panel meeting | Abstracts, editorials, letters<br>Duplicate publications of the same study which do not report on different outcomes<br>Single reports from multicenter trials<br>White papers<br>Narrative reviews<br>Articles identified as preliminary reports when results are published in later versions |
| Study Design | KQ1 – KQ4<br>RCT<br>Comparative study with concurrent control group (N≥50 for L-ADR; N≥100 for CADR)<br>Systematic review<br>Meta-analysis<br>KQ3<br>RCTs which stratify on patient or other characteristics and formally evaluate statistical interaction (effect modification)<br>KQ4<br>Formal economic analyses (e.g., cost-utility study)<br>In the absence of formal economic analyses, cost data reported in other systematic reviews or technology assessments | In vitro/ kinematic studies<br>Narrative reviews<br>Cross-sectional studies |

Abbreviations: ACDF: anterior cervical discectomy and fusion; ADR: artificial disc replacement; CADR: cervical artificial disc replacement; DDD: degenerative disc disease; FDA: Food and Drug Administration; KQ: key question; N: number; RCT: randomized controlled trial; SSED: Summary of Safety and Effectiveness Data

# Appendix D. 2016 HTA Strength of Evidence Summary Tables for CADR and LADR

## Table D-1. Strength of Evidence Summary: CADR vs. ACDF (1-level) Efficacy Results

| Outcome | Follow-up | RCTs | N* | Reasons for Downgrading | Conclusion* | Quality |
|---|---|---|---|---|---|---|
| **Overall success†** | 24 mos. | 5 RCTs (Prestige ST, ProDisc-C, Bryan, SECURE-C, & PCM IDE trials) | N= 1681 | Risk of bias[1] (-1) | Pooled RD 9.5% (95% CI 5.3%, 13.7%)<br>Conclusion: CADR was superior to ACDF in terms of the percentage of patients who achieved overall success at 24 months. | ⊕⊕⊕○ MODERATE |
| | 48-60 mos. | 3 RCTs (Mobi-C, Prestige ST, & Bryan IDE trials) | N= 933 | Risk of bias[1] (-1) | Pooled RD 9.6% (95% CI 3.9%, 15.3%)<br>Conclusion: CADR was superior to ACDF in terms of the percentage of patients who achieved overall success at 48 to 60 months. | ⊕⊕⊕○ MODERATE |
| | 84 mos. | 1 RCT (Prestige ST IDE trial) | N= 933 | Risk of bias[1] (-1)<br>Imprecision[3] (-1) | RD 11.8% (95% CI 2.0%, 20.1%)<br>Conclusion: CADR was superior to ACDF in terms of the percentage of patients who achieved overall success at 84 months. | ⊕⊕○○ LOW |
| **NDI success** (≥15-point improvement) | 24 mos. | 5 RCTs (Prestige ST, ProDisc-C, Bryan, SECURE-C, & PCM IDE trials) | N= 1640 | Risk of bias[1] (-1) | Pooled RD 4.3% (95% CI 0.6%, 8.1%)<br>Conclusion: Slightly more CADR than ACDF patients achieved NDI success (≥15-point improvement from baseline) at 24 months. | ⊕⊕⊕○ MODERATE |
| | 48-60 mos. | 3 RCTs (Mobi-C, Prestige ST, & Bryan IDE trials) | N= 933 | Risk of bias[1] (-1)<br>Imprecision[3] (-1) | Pooled RD 5.8% (95% CI -1.8%, 13.3%)<br>Conclusion: CADR and ACDF appear to be comparable; no significant difference between groups. | ⊕⊕○○ LOW |
| | 84 mos. | 1 RCT (Prestige ST IDE trial) | N= 395 | Risk of bias[1] (-1)<br>Imprecision[3] (-1) | RD 3.2% (95% CI -4.5%, 10.8%)<br>Conclusion: CADR and ACDF appear to be comparable; no significant difference between groups. | ⊕⊕○○ LOW |
| **NDI scores** (0-100) | 24 mos. | 9 RCTs (Prestige ST, ProDisc-C, Mobi-C, Bryan, PCM, & SECURE-C IDE trials; Karabag 2014; Zhang 2012; Zhang 2014) | N=2183 | Risk of bias[1] (-1) | WMD 1.11 (95% CI -0.06, 2.27)<br>Conclusion: CADR may be comparable to ACDF in terms of mean NDI scores at 24 months; the difference between groups was not significant. | ⊕⊕⊕○ MODERATE |

| Outcome | Follow-up | RCTs | N* | Reasons for Downgrading | Conclusion* | Quality |
|---|---|---|---|---|---|---|
| | 48-60 mos. | 6 RCTs (ProDisc-C, Mobi-C, Bryan, Prestige ST, & PCM IDE trials; Zhang 2014) | N= 1443 | Risk of bias[1] (-1) | WMD 4.21 (95% CI 1.67, 6.75) Conclusion: CADR patients had slightly higher NDI scores than did ACDF patients at 48 to 60 months, although the difference between groups is probably not clinically meaningful. Additionally, this effect appears to stem largely from three moderately high risk of bias trials, as the two moderately low risk of bias trials together suggest equivalence. | ⊕⊕⊕◯ MODERATE |
| | 84 mos. | 2 RCTs (ProDisc-C & Prestige ST IDE trials) | N= 544 | Risk of bias[1] (-1) Imprecision[4] (-1) | WMD 4.41 (95% CI 0.68, 8.14) Conclusion: CADR conferred a slight benefit over ACDF in mean NDI scores, although the difference between groups is probably not clinically meaningful. Additionally, this effect appears to stem largely from the moderately high risk of bias trial, as the moderately low risk of bias trial found no difference between groups. | ⊕⊕◯◯ LOW |
| **Neurological success** (maintenance/ improvement of motor function, sensory function, and deep tendon reflexes) | 24 mos. | 6 RCTs (Mobi-C, ProDisc-C, Prestige ST, Bryan, PCM, & SECURE-C IDE trials) | N= 1882 | Risk of bias[1] (-1) | Pooled RD 3.2% (95% CI 0.8%, 5.7%) Conclusion: CADR may be slightly better than ACDF in terms of neurological success at 24 months. | ⊕⊕⊕◯ MODERATE |
| | 48-60 mos. | 4 RCTs (ProDisc-C, Bryan, Prestige ST, & PCM IDE trials) | N= 1147 | Risk of bias[1] (-1) | Pooled RD 4.0% (95% CI 0.5%, 7.5%) Conclusion: CADR may be slightly better than ACDF in terms of neurological success at 48 to 60 months. | ⊕⊕⊕◯ MODERATE |
| | 84 mos. | 2 RCTs(ProDisc-C & Prestige ST IDE trials) | N= 531 | Risk of bias[1] (-1) Imprecision[3] (-1) | Pooled RD 4.5% (95% CI -4.9%, 13.8%) Conclusion: CADR and ACDF appear to be comparable; no significant difference between groups. | ⊕⊕◯◯ LOW |

| Outcome | Follow-up | RCTs | N* | Reasons for Downgrading | Conclusion* | Quality |
|---|---|---|---|---|---|---|
| **Arm pain success** (≥20-point VAS improvement) | 24 mos. | 2 RCTs (SECURE-C & PCM IDE trials) | N= 578 | Risk of bias[1] (-1) Imprecision[3] (-1) | Conclusion: Two trials each found no difference between groups in the percentage of patients who achieved arm pain success at 24 months:‡ • SECURE-C trial: RD 4.7% (95% CI -7.9%, 17.4%) (left arm); RD -2.5% (95% CI -15.1%, 10.1%) (right arm) • PCM trial: RD 3.8% (95% CI -5.2%, 12.8%) (worst arm) | ⊕⊕◯◯ LOW |
| | 60 mos. | 1 RCT (PCM IDE trial) | N= 288 | Risk of bias[1] (-1) Imprecision[3] (-1) | RD 9.5% (95% CI -0.4%, 19.5%) Conclusion: CADR and ACDF appear to be comparable, no significant difference between groups. | ⊕⊕◯◯ LOW |
| | 84 mos. | No trials | | | No data reported. | ⊕◯◯◯ INSUFFICIENT |
| **Arm pain VAS scores** (0-100) | 24 mos. | 7 RCTs (Prestige ST, ProDisc-C, Mobi-C, Bryan, PCM, & SECURE-C IDE trials; Zhang 2012) | N= 2015 | Risk of bias[1] (-1) | WMD 1.60 (95% CI 0.51, 2.70) Conclusion: Arm pain VAS scores were slightly better with CADR versus ACDF; however, the difference between groups is probably not clinically meaningful. Two additional trials (Rozankovic 2016 (N=101), Nabhan 2007 (N=39)), reached similar conclusions but were not included in the pooled analysis.§ | ⊕⊕⊕◯ MODERATE |
| | 48-60 mos. | 5 RCTs (ProDisc-C, Mobi-C, Bryan, Prestige ST, & PCM IDE trials) | N= 1332 | Risk of bias[1] (-1) | WMD 3.82 (95% CI 1.15, 6.48) Conclusion: Arm pain VAS scores may be slightly better with C-ADR versus ACDF; however, the difference between groups is probably not clinically meaningful. | ⊕⊕⊕◯ MODERATE |
| | 84 mos. | 2 RCTs (ProDisc-C & Prestige ST IDE trials) | N= 543 | Risk of bias[1] (-1) Imprecision[4] (-1) | WMD 2.21 (95% CI -2.08, 6.50) Conclusion: CADR and ACDF appear to be comparable; no significant difference between groups. | ⊕⊕◯◯ LOW |
| **Neck pain success** (≥20-point VAS improvement) | 24 mos. | 2 RCTs (SECURE-C & PCM IDE trials) | N= 578 | Risk of bias[1] (-1) Imprecision[3] (-1) | Pooled RD 3.6% (95% CI -6.1%, 13.4%) Conclusion: CADR and ACDF appear to be comparable; no significant difference between groups. | ⊕⊕◯◯ LOW |
| | 60 mos. | 1 RCT (PCM IDE trial) | N= 288 | Risk of bias[1] (-1) Imprecision[3] (-1) | -4.0% (95% CI -14.1%, 6.3% Conclusion: CADR and ACDF | ⊕⊕◯◯ LOW |

| Outcome | Follow-up | RCTs | N* | Reasons for Downgrading | Conclusion* | Quality |
|---|---|---|---|---|---|---|
| | | | | | appear to be comparable; no significant difference between groups. | |
| | 84 mos. | No trials | | | No data reported. | ⊕◯◯◯ INSUFFICIENT |
| **Neck pain VAS scores** (0-100) | 24 mos. | 3 RCTs (Prestige ST, ProDisc-C, Mobi-C IDE trials) | N= 905 | Risk of bias[1] (-1) | WMD 1.29 (95% CI -1.28, 3.86) Conclusion: CADR is as good as ACDF. For the three trials at moderately low risk of bias only, no difference was seen between groups. Six additional trials (Bryan, PCM, & SECURE-C IDE trials; Nabhan 2007; Rozankovic 2016; Zhang 2012) (N=1250) reported this outcome; however, the resulting pooled estimate, which favored CADR, had high statistical heterogeneity ($I^2$=80%) (WMD 5.11 (95% CI 2.55, 7.66)). | ⊕⊕⊕◯ MODERATE |
| | 48-60 mos. | 5 RCTs (ProDisc-C, Mobi-C, Bryan, Prestige ST, & PCM IDE trials) | N= 1331 | Risk of bias[1] (-1) | WMD 6.63 (95% CI 3.29, 9.97) Conclusion: CADR is as good as or slightly better than ACDF; C-ADR may confer a slight benefit over ACDF in mean NDI scores, although the difference between groups is most likely not clinically meaningful. | ⊕⊕⊕◯ MODERATE |
| | 84 mos. | 2 RCTs (ProDisc-C & Prestige ST IDE trials) | N= 543 | Risk of bias[1] (-1) Imprecision[4] (-1) | WMD 5.59 (95% CI 1.31, 9.86) Conclusion: CADR is as good as or slightly better than ACDF; C-ADR may confer a slight a slight benefit over ACDF in terms of mean neck pain VAS scores, although the difference between groups is probably not clinically meaningful. Additionally, this effect appears to stem largely from the moderately high risk of bias trial, as the moderately low risk of bias trial found no difference between groups. | ⊕⊕◯◯ LOW |

* Unless otherwise specified, conclusion and patient numbers were based on completer analysis, for which the effect estimate was more conservative than that of the ITT analysis.

† Overall clinical success included the following components:

- NDI score improvement ≥ 15 points (from baseline)
- Maintenance or improvement in neurological status
- No additional surgery from device failure (removal, revision, supplemental fixation)
- No device-related adverse events and/or major complications

- In addition, one trial required patients to achieve radiological success for motion (PCM trial); another stipulated no changes to the treatment plan made intraoperatively (SECURE-C trial)

‡ Results could not be pooled due to differences in data reporting between the trials.

§ Two trials were excluded from the pooled analysis because their mean differences were both considerably different from those reported by other trials and their inclusion led to high statistical heterogeneity

Reasons for downgrading:

1. Serious risk of bias: the majority of studies violated one or more of the criteria for good quality RCT (or cohort study) related to the outcome reported (see Appendix for details)
2. Inconsistency: differing estimates of effects across trials
3. Imprecise effect estimate for a dichotomous outcome: small sample size and/or confidence interval includes both negligible effect and appreciable benefit or harm with ADR
4. Imprecise effect estimate for a continuous outcome: wide (or unknown) confidence interval and/or small sample size

## Table D-2. Strength of Evidence Summary: CADR vs. ACDF (2-level) Efficacy Results

| Outcome | Follow-up | RCTs | N* | Reasons for Downgrading | Conclusion* | Quality |
|---|---|---|---|---|---|---|
| **Overall success†** | 24 mos. | 1 RCT (Mobi-C (2-level) ST IDE trial)) | N= 320 | Risk of bias[1] (-1) | RD 23.2% (95% CI 11.6%, 34.8%) Conclusion: CADR was superior to ACDF in terms of the percentage of patients who achieved overall success at 24 months. | ⊕⊕⊕◯ MODERATE |
| | 60 mos. | 1 RCT (Mobi-C (2-level) ST IDE trial)) | N= 297 | Risk of bias[1] (-1) | RD 29.6% (95% CI 18.1%, 41.2%) Conclusion: CADR was superior to ACDF in terms of the percentage of patients who achieved overall success at 60 months. | ⊕⊕⊕◯ MODERATE |
| | 84 mos. | No trials | | | No data reported. | ⊕◯◯◯ INSUFFICIENT |
| **NDI success‡** | 24 mos. | 1 RCT (Mobi-C (2-level) ST IDE trial)) | N= 320 | Risk of bias[1] (-1) | RD 16.7% (95% CI 5.7%, 27.7%) Conclusion: CADR was superior to ACDF in terms of the percentage of patients who achieved NDI success at 24 months. | ⊕⊕⊕◯ MODERATE |
| | 48 mos. | 1 RCT Mobi-C (2-level) ST IDE trial)) | N=285 | Risk of bias[1] (-1) | RD 26.6% (95% CI 14.6%, 38.6%) Conclusion: CADR was superior to ACDF in terms of the percentage of patients who achieved NDI success at 24 months. | ⊕⊕⊕◯ MODERATE |
| | 84 mos. | No trials | | | No data reported. | ⊕◯◯◯ INSUFFICIENT |
| **NDI scores** | 24 mos. | 2 RCTs (Mobi-C (2-level) ST IDE trial), Cheng 2009) | N= 353 | Risk of bias[1] (-1) Imprecision[4] (-1) | Conclusion: CADR may be slightly better than ACDF in terms of NDI scores; both trials reported significantly better scores following CADR: one moderately low risk of bias trial (Mobi-C, N=291) (MD -7.5 (95% | ⊕⊕◯◯ LOW |

| Outcome | Follow-up | RCTs | N* | Reasons for Downgrading | Conclusion* | Quality |
|---|---|---|---|---|---|---|
| | | | | | CI -12.0, -3.0)) and another moderately high risk of bias trial (Cheng 2009, N=62) (11 vs. 19, MD -8 (95% CI NR), p=0.02). Differences may not be clinically meaningful. | |
| | 60 mos. | 1 RCT (Mobi-C (2-level) ST IDE trial)) | N= 258 | Risk of bias[1] (-1) Imprecision[4] (-1) | MD -9.6 (95% CI -14.6, -4.6) Conclusion: NDI scores may be slightly better with CADR versus ACDF; however, differences may not be clinically meaningful. | ⊕⊕◯◯ LOW |
| | 84 mos. | No trials | | | No data reported. | ⊕◯◯◯ INSUFFICIENT |
| **Neurological success** (maintenance/ improvement of motor function, sensory function, and deep tendon reflexes) | 24 mos. | 1 RCT (Mobi-C (2-level) ST IDE trial)) | N= 320 | Risk of bias[1] (-1) Imprecision[3] (-1) | RD 1.6% (95% CI -4.2%, 7.5%) Conclusion: CADR and ACDF appear to be comparable; no significant difference between groups. | ⊕⊕◯◯ LOW |
| | 60 mos. | 1 RCT (Mobi-C (2-level) ST IDE trial)) | N= 297 | Risk of bias1 (-1) Imprecision3 (-1) | RD -2.4% (95% CI -8.7%, 4.0%) Conclusion: CADR and ACDF appear to be comparable; no significant difference between groups. | ⊕⊕◯◯ LOW |
| | 84 mos. | No trials | | | No data reported. | ⊕◯◯◯ INSUFFICIENT |
| **Arm or neck pain success** | Any | No trials | | | No data reported. | ⊕◯◯◯ INSUFFICIENT |
| **Arm pain VAS scores** (0-100) | 24 mos. | 2 RCTs (Mobi-C (2-level) ST IDE trial), Cheng 2009) | N= 353 | Risk of bias[1] (-1) Imprecision[4] (-1) | Conclusion: CADR is as good as or better than ACDF in terms of arm pain scores: while one moderately low risk of bias trial (Mobi-C, N=291) found no difference between groups (MD -4.3 (95% CI -9.5, 0.9)), another moderately high risk of bias trial (Cheng 2009, N=62) found better scores with CADR than with ACDF (14 vs. 27, MD -13 (95% CI NR), p=0.01). Differences may not be clinically meaningful. | ⊕⊕◯◯ LOW |
| | 48 mos. | 1 RCT (Mobi-C (2-level) ST IDE trial)) | N= 255 | Risk of bias[1] (-1) Imprecision[4] (-1) | MD in Δ scores: -3.0 (95% CI -11.6, 5.6) Conclusion: CADR and ACDF appear to be comparable; no significant difference between groups. | ⊕⊕◯◯ LOW |

| Outcome | Follow-up | RCTs | N* | Reasons for Downgrading | Conclusion* | Quality |
|---|---|---|---|---|---|---|
| | 84 mos. | No trials | | | No data reported. | ⊕◯◯◯ INSUFFICIENT |
| **Neck pain VAS scores** (0-100) | 24 mos. | 2 RCTs (Mobi-C (2-level) ST IDE trial), Cheng 2009) | N= 353 | Risk of bias[1] (-1) Imprecision[4] (-1) | Conclusion: CADR is as good as or slightly better than ACDF in terms of neck pain scores: while one moderately low risk of bias trial (Mobi-C, N=291) found no difference between groups (MD -3.9 (95% CI -10.1, 2.3)), another moderately high risk of bias trial (Cheng 2009, N=62) reported better scores with CADR than with ACDF (15 vs. 26, MD -11 (95% CI NR), p=0.01). Differences may not be clinically meaningful | ⊕⊕◯◯ LOW |
| | 48 mos. | 1 RCT (Mobi-C (2-level) ST IDE trial)) | N= 255 | Risk of bias[1] (-1) Imprecision[4] (-1) | MD in Δ scores: -5.0 (95% CI -13.3, 3.3) Conclusion: CADR and ACDF appear to be comparable; no significant difference between groups. | ⊕⊕◯◯ LOW |
| | 84 mos. | No trials | | | No data reported. | ⊕◯◯◯ INSUFFICIENT |

* Unless otherwise specified, conclusion and patient numbers were based on completer analysis, for which the effect estimate was more conservative than that of the ITT analysis.

† Overall clinical success required all of the following:

- NDI improvement of at least 15 points (out of 50) from baseline
- Maintenance or improvement in all components of neurological status
- No subsequent surgical intervention at the index level or levels;
- No potentially (possibly or probably) device-related adverse event;
- No Mobi-C intraoperative changes in treatment.

‡ NDI success was defined as postoperative ≥30-point improvement on the NDI if the baseline score was ≥60, or ≥50% improvement if the baseline score was <60.

Reasons for downgrading:

1. Serious risk of bias: the majority of studies violated one or more of the criteria for good quality RCT (or cohort study) related to the outcome reported (see Appendix for details)
2. Inconsistency: differing estimates of effects across trials
3. Imprecise effect estimate for a dichotomous outcome: small sample size and/or confidence interval includes both negligible effect and appreciable benefit or harm with ADR
4. Imprecise effect estimate for a continuous outcome: wide (or unknown) confidence interval and/or small sample size.

## Table D-3. Strength of Evidence Summary: CADR vs. ACDF (Mixed levels (1-, 2-, or 3-level)) Efficacy Results

| Outcome | Follow-up | RCTs | N* | Reasons for Downgrading | Conclusion* | Quality |
|---|---|---|---|---|---|---|
| **Overall, NDI, or neurological success** | Any | No trials | | | No data reported. | ⊕○○○ INSUFFICIENT |
| **NDI scores** | 24 mos. | 1 RCT (Skeppholm 2015) | N= 143 | Risk of bias[1] (-1) Imprecision[4] (-1) | MD -1.0 (95% CI -7.4, 5.4) Conclusion: CADR and ACDF appear to be comparable. No significant difference between groups in one trial of radiculopathy patients. | ⊕⊕○○ LOW |
| | 24-36 mos. | 1 RCT (Cheng 2011) | N= 81 | Risk of bias[1] (-1) Imprecision[4] (-1) | Conclusion: CADR is as good as or slightly better. One trial of myelopathy patients reported better scores with CADR than with ACDF at 24 months (13 vs. 16, MD -3 (95% CI NR), p=0.01) and 36 months (12 vs. 17, MD -5 (95% CI NR), p<0.01), although this difference is not likely to be clinically meaningful. | ⊕⊕○○ LOW |
| | 48-60 or 84 mos. | No trials | | | No data reported. | ⊕○○○ INSUFFICIENT |
| **Arm or neck pain success** | Any | No trials | | | No data reported. | ⊕○○○ INSUFFICIENT |
| **Arm pain VAS scores** (0-100) | 24 mos. | 1 RCT (Skeppholm 2015) | N= 143 | Risk of bias[1] (-1) Imprecision[4] (-1) | MD 0.4 (95% CI -7.7, 8.5) Conclusion: CADR and ACDF appear to be comparable. No significant difference between groups. | ⊕⊕○○ LOW |
| | 48-60, 84 mos. | No trials | | | No data reported. | ⊕○○○ INSUFFICIENT |
| **Arm pain VAS scores**(0-100) | 24 mos. | 1 RCT (Skeppholm 2015) | N= 143 | Risk of bias[1] (-1) Imprecision[4] (-1) | MD -1.2 (95% CI -9.9, 7.5) Conclusion: CADR and ACDF appear to be comparable. No significant difference between groups. | ⊕⊕○○ LOW |
| | 48-60, 84 mos. | No trials | | | No data reported. | ⊕○○○ INSUFFICIENT |

* Unless otherwise specified, conclusion and patient numbers were based on completer analysis, for which the effect estimate was more conservative than that of the ITT analysis.

Reasons for downgrading:

1. Serious risk of bias: the majority of studies violated one or more of the criteria for good quality RCT (or cohort study) related to the outcome reported (see Appendix for details)

2. Inconsistency: differing estimates of effects across trials
3. Imprecise effect estimate for a dichotomous outcome: small sample size and/or confidence interval includes both negligible effect and appreciable benefit or harm with ADR
4. Imprecise effect estimate for a continuous outcome: wide (or unknown) confidence interval and/or small sample size

## Table D-4. Strength of Evidence Summary: CADR vs. ACDF with a zero-profile device (2 non-contiguous levels) Efficacy Results

| Outcome | Follow-up | RCTs | N* | Reasons for Downgrading | Conclusion* | Quality |
|---|---|---|---|---|---|---|
| **Overall, NDI, or neurological success** | Any | No trials | | | No data reported. | ⊕◯◯◯ INSUFFICIENT |
| **NDI scores** | Mean 32.4 (24-46) mos. | 1 RCT (Qizhi 2016) | N= 30 | Risk of bias[1] (-1) Imprecision[4] (-1) | MD 0.3 (95% CI -0.4, 1.0) Conclusion: CADR and ACDF appear to be comparable. No significant difference between groups possibly due in part to small sample size. | ⊕⊕◯◯ LOW |
| **Arm or neck pain success or scores** | Any | No trials | | | No data reported. | ⊕◯◯◯ INSUFFICIENT |

* Unless otherwise specified, conclusion and patient numbers were based on completer analysis, for which the effect estimate was more conservative than that of the ITT analysis.

Reasons for downgrading:

1. Serious risk of bias: the majority of studies violated one or more of the criteria for good quality RCT (or cohort study) related to the outcome reported (see Appendix for details)
2. Inconsistency: differing estimates of effects across trials
3. Imprecise effect estimate for a dichotomous outcome: small sample size and/or confidence interval includes both negligible effect and appreciable benefit or harm with ADR
4. Imprecise effect estimate for a continuous outcome: wide (or unknown) confidence interval and/or small sample size

## Table D-5. Strength of Evidence Summary: CADR vs. ACDF (1-level) Safety Results

| Outcome | Follow-up | RCTs | N* | Reasons for Downgrading | Conclusion* | Quality |
|---|---|---|---|---|---|---|
| **Secondary surgery at the index level** | 24 mos. | 8 RCTs (Prestige ST, Mobi-C, ProDisc-C, Bryan, PCM, & SECURE-C IDE trials; Karabag 2014; Rozankovic 2016) | N= 2299 | Risk of bias[1] (-1) | CADR 2.9%, ACDF 6.2% Pooled RD 3.1% (95% CI 1.1%, 5.1%) Conclusion: Fewer patients in the CADR group underwent secondary surgery at the index level through 24 months compared with those in the ACDF group. | ⊕⊕⊕◯ MODERATE |

| Outcome | Follow-up | RCTs | N* | Reasons for Downgrading | Conclusion* | Quality |
|---|---|---|---|---|---|---|
| | 48-60 mos. | 4 RCTs (Mobi-C, ProDisc-C, Bryan, & PCM IDE trials) | N= 1335 | Risk of bias[1] (-1) Imprecision[3] (-1) | CADR 4.6%, ACDF 9.3% Pooled RD 4.8% (95% CI 0.8%, 8.8%) Conclusion: Fewer patients in the CADR group underwent secondary surgery at the index level through 48 or 60 months compared with those in the ACDF group. | ⊕⊕◯◯ LOW |
| | 84 mos. | 2 RCTs (ProDisc-C & Prestige ST IDE trials) | N= 750 | Risk of bias[1] (-1) Imprecision[3] (-1) | CADR 4.5%, ACDF 12.1% RD 7.5% (95% CI 3.6%, 11.4%) Conclusion: CADR was superior to ACDF in terms of the percentage of patients who underwent secondary surgery at the index level through 84 months. | ⊕⊕◯◯ LOW |
| **Serious/ major adverse events* (as classified by the trial)** | 24 mos. | 5 RCTs (Prestige ST, ProDisc-C, Bryan, SECURE-C, & PCM IDE trials) | N= 2388 | Risk of bias[1] (-1) Imprecision[3] (-1) | CADR 24.3%, ACDF 31.0% Pooled RD 6.8% (95% CI 2.0%, 11.6%) Conclusion: Slightly fewer C-ADR than ACDF patients had serious adverse events (as classified by the trial) through at 24 months. | ⊕⊕◯◯ LOW |
| | 24-48 mos. | 1 RCT (Bryan ST IDE trial) | N= 463 | Risk of bias[1] (-1) Imprecision[3] (-1) | CADR 17.4%, ACDF 17.1% RD -0.3% (95% CI -7.2%, 6.7%) Conclusion: No significant difference between groups. | ⊕⊕◯◯ LOW |
| | 0-48 mos. | 1 RCT (Mobi-C IDE trial) | N= 260 | Risk of bias[1] (-1) Imprecision[3] (-1) | CADR 10.1%, ACDF 9.9% RD -0.2% (95% CI -8.0%, 7.7%) Conclusion: No significant difference between groups. | ⊕⊕◯◯ LOW |
| | 24-84 mos. | 1 RCT (PCM ST IDE trial) | N= 404 | Risk of bias[1] (-1) Imprecision[3] (-1) | CADR 21.0%, ACDF 17.4% RD -3.7% (95% CI -11.3%, 4.0%) Conclusion: No significant difference between groups. | ⊕⊕◯◯ LOW |
| **Device-related adverse events† (as classified by the trial)** | 24 mos. | 6 RCTs (Prestige ST, ProDisc-C, Mobi-C, Bryan, PCM, & SECURE-C IDE trials) | N= 2167 | Risk of bias[1] (-1) | CADR 4.9%, ACDF 10.8% Pooled RD 5.0% (95% CI 2.7%, 7.4%) Conclusion: Device-related adverse events (as classified by the trial) were less common with CADR than ACDF through at 24 months. | ⊕⊕⊕◯ MODERATE |

| Outcome | Follow-up | RCTs | N* | Reasons for Downgrading | Conclusion* | Quality |
|---|---|---|---|---|---|---|
| | 60 mos. | 2 RCTs (Mobi-C & ProDisc-C IDE trials) | N= 469 | Risk of bias[1] (-1) | CADR 3.9%, ACDF 3.2% Pooled RD 0.4% (95% CI -3.4%, 4.3%) Conclusion: No significant difference between groups. | ⊕⊕⊕◯ MODERATE |
| | 84 mos. | 1 RCT (ProDisc-C IDE trial) | N= 209 | Risk of bias[1] (-1) Imprecision[3] (-1) | CADR 27.2%, ACDF 28.3% RD 1.1% (95% CI -11.0%, 13.3%) Conclusion: No significant difference between groups. | ⊕⊕◯◯ LOW |

* Defined as:

- Bryan IDE trial: Most serious adverse events were related to medical conditions and not to the procedure, implant, or cervical spine disease. Classified as WHO grade 3 or 4 (taken from Anderson 2008) (grade 3 events required medical treatment or may have had a long-term health effect; grade 4 events required an operation, were life threatening, permanent disability, or caused death).
- PCM IDE trial: any event that results in death, serious injury, permanent impairment; or that prolongs hospitalization or requires surgical intervention to prevent death or serious injury; classified by the Clinical Events Committee.
- Mobi-C IDE trial: any event that results in death, serious injury, permanent impairment; or that prolongs hospitalization or requires surgical intervention to prevent death or serious injury; or that was a congenital anomaly or birth defect; classified by the Clinical Events Committee.
- ProDisc-C IDE trial: "Severe or life-threatening adverse event": defined as any event requiring hospitalization or surgery (see SSED Table 18).
- Secure-C IDE trial: "Severe or life-threatening adverse event": a severe event was defined as any event that significantly limits the patient's ability to perform routine activities despite symptomatic therapy; a life-threatening event was defined as any event that required removal of the implant or put the patient at immediate risk of death (including death) (see SSED Table 19).

† Defined as:

- Prestige ST IDE trial: events included anatomical/technical difficulty, implant displacement/loosening, infection, neck and/or arm pain, neurological, non-union, pending non-union, and subsidence.
- Bryan IDE trial: events included malpositioned implant, neck and/or arm pain, non-union, other, pending non-union, spinal event, and trauma.
- Mobi-C IDE trial: events included spinal ligament ossification, neck pain, muscle spasms, radiculopathy, subsidence, medical device complication, misplaced screw coded as device complication.
- ProDisc-C IDE trial (0-24 months): events included dysphagia, superficial wound infection, musculoskeletal, neck pain, and index-level surgery.
- ProDisc-C IDE trial (0-84 months): adjacent-level degenerative disc disease or degenerative joint changes, cardiovascular, dysphagia, headache, musculoskeletal, musculoskeletal neck spasms, neurologic, numbness, ossification, other, back and lower extremity pain, incision site pain, neck pain, neck and other pain, neck and shoulder pain, neck and upper extremity pain, neck and upper extremity pain with numbness, surgery for device related events (index or other level), wound issues.Secure-C IDE trial: device-related adverse events were classified by the Clinical Events Committee and included those events that were linked to the device (revision, removal, reoperation, or supplemental fixation at the index level; fracture or mechanical failure of the device, pseudarthrosis, radiolucency around the device, migration, subsidence, loosening, etc. Neck and arm pain were excluded from this category of adverse events.
- Riina 2008: not defined

Reasons for downgrading:

1. Serious risk of bias: the majority of studies violated one or more of the criteria for good quality RCT (or cohort study)

related to the outcome reported (see Appendix for details)

2. Inconsistency: differing estimates of effects across trials
3. Imprecise effect estimate for a dichotomous outcome: small sample size and/or confidence interval includes both negligible effect and appreciable benefit or harm with ADR. If sample size is likely too small to detect rare outcomes, evidence may be downgraded twice.

## Table D-6. Strength of Evidence Summary: CADR vs. ACDF (2-level) Safety Results

| Outcome | Follow-up | RCTs | N* | Reasons for Downgrading | Conclusion* | Quality |
|---|---|---|---|---|---|---|
| **Secondary surgery at the index level** | 24 mos. | 1 RCT (Mobi-C (2-level) IDE trial) | N= 330 | Risk of bias[1] (-1) Imprecision[3] (-1) | CADR 3.1%, ACDF 11.4% RD -8.3% (95% CI -14.8%, -1.8%) Conclusion: Secondary surgery at the index level was performed in fewer CADR than ACDF patients through 24 months. | ⊕⊕◯◯ LOW |
| | 60 mos. | 1 RCT (Mobi-C (2-level) IDE trial) | N= 339 | Risk of bias[1] (-1) Imprecision[3] (-1) | CADR 4.7%, ACDF 12.4% RD -7.7% (95% CI -14.5%, -0.8%) Conclusion: Fewer patients in the CADR group underwent secondary surgery at the index level through 60 months compared with those in the ACDF group. | ⊕⊕◯◯ LOW |
| | 84 mos. | No trials | | | No data reported. | ⊕◯◯◯ INSUFFICIENT |
| **Serious/ major adverse events* (as classified by the trial)** | 24 mos. | 1 RCT (Mobi-C (2-level) IDE trial) | N= 330 | Risk of bias[1] (-1) Imprecision[3] (-1) | CADR 24.4%, ACDF 32.4% RD -7.9% (95% CI -18.5%, 2.6%) Conclusion: Device-related adverse events (as classified by the trial) were less common with CADR than ACDF through at 24 months. | ⊕⊕◯◯ LOW |
| | 48-60 mos. | No trials | | | No data reported. | ⊕◯◯◯ INSUFFICIENT |
| | 84 mos. | No trials | | | No data reported. | ⊕◯◯◯ INSUFFICIENT |
| **Device-related adverse events† (as classified by the trial)** | 24 mos. | 1 RCT (Mobi-C (2-level) IDE trial) | N= 330 | Risk of bias[1] (-1) Imprecision[3] (-1) | CADR 16.0%, ACDF 34.3% RD -18.3% (95% CI -28.6%, -8.0%) Conclusion: Device-related adverse events (as classified by the trial) were less common with CADR than ACDF through at 24 months. | ⊕⊕◯◯ LOW |
| | 48-60 mos. | No trials | | | No data reported. | ⊕◯◯◯ INSUFFICIENT |
| | 84 mos. | No trials | | | No data reported. | ⊕◯◯◯ INSUFFICIENT |

* Classified by the Clinical Events Committee as possibly or definitely related to the device, and included anatomy/technical difficulty, dysphagia/dysphonia, gastrointestinal, heterotopic ossification, malpositioned implant, neck and/or arm pain, neurological, non-union, other, other pain, respiratory, spinal disorder, trauma.

† Serious adverse events met one or more of the following criteria: 1) resulted in death; 2) was life-threatening (immediate risk of death); 3) required inpatient hospitalization or prolonged hospitalization; 4) resulted in persistent or significant disability or incapacity; 5) necessitated medical or surgical intervention to preclude permanent impairment of a body function or permanent damage to a body structure; or 6) was a congenital anomaly or birth defect. Reported events included: anatomy/technical difficulty, cancer, cardiovascular, death, dysphagia/dysphonia, gastrointestinal, infection (systemic or local), malpositioned implant, migration of implant, neck and/or arm pain, neurological, non-union, other, other pain, respiratory, spinal disorder, trauma, upper extremity nerve entrapment, urogenital, non-infectious wound issue (hematoma, CSF leakage).

Reasons for downgrading:

1. Serious risk of bias: the majority of studies violated one or more of the criteria for good quality RCT (or cohort study) related to the outcome reported (see Appendix for details)
2. Inconsistency: differing estimates of effects across trials
3. Imprecise effect estimate for a dichotomous outcome: small sample size and/or confidence interval includes both negligible effect and appreciable benefit or harm with ADR. If sample size is likely too small to detect rare outcomes, evidence may be downgraded twice.

## Table D-7. Strength of Evidence Summary: CADR vs. ACDF (Mixed level (1-, 2-, or 3-level) Safety Results

| Outcome | Follow-up | RCTs | N* | Reasons for Downgrading | Conclusion* | Quality |
|---|---|---|---|---|---|---|
| **Secondary surgery at the index level** | 24-36 mos. | 2 RCTs (Skeppholm 2015, Cheng 2011) | N= 234 | Risk of bias[1] (-1)<br>Imprecision[3] (-1) | 24 mos. (N=151):<br>CADR 6.2%, ACDF 1.4%<br>RD 4.7% (95% CI -1.2%, 10.7%)<br>36 mos. (N=83):<br>CADR 0%, ACDF 0%<br>RD 0% (95% CI not calculable)<br>Conclusion: No significant difference between groups. | ⊕⊕◯◯ LOW |
| | 48-60 mos. | No trials | | | No data reported. | ⊕◯◯◯ INSUFFICIENT |
| | 84 mos. | No trials | | | No data reported. | ⊕◯◯◯ INSUFFICIENT |
| **Serious/ major adverse events** | 24-36 mos. | 2 RCTs (Skeppholm 2015, Cheng 2011) | N= 234 | Risk of bias[1] (-1)<br>Imprecision[3] (-1) | Conclusion: No serious adverse events were reported by either trial. | ⊕⊕◯◯ LOW |
| | 48-60 mos. | No trials | | | No data reported. | ⊕◯◯◯ INSUFFICIENT |
| | 84 mos. | No trials | | | No data reported. | ⊕◯◯◯ INSUFFICIENT |
| **Device-related adverse events** | 24-36 mos. | 2 RCTs (Skeppholm 2015, Cheng 2011) | N= 234 | Risk of bias[1] (-1)<br>Imprecision[3] (-1) | Conclusion: No overall summary of device-related adverse events was reported by either trial. With the exception of dysphagia, which was less common in the CADR group than in the ACDF group | ⊕⊕◯◯ LOW |

| Outcome | Follow-up | RCTs | N* | Reasons for Downgrading | Conclusion* | Quality |
|---|---|---|---|---|---|---|
| | | | | | (Skeppholm: 11.8% vs. 19.9% through 24 months, p=0.31; Cheng 2011: 2.4% vs. 16.7% through 36 months, p<0.01), complications attributable to the device occurred similarly between groups, and occurred in relatively few patients (0-2.4% of the CADR group; 0% in the ACDF group) across both trials. | |
| | 48-60 mos. | No trials | | | No data reported. | ⊕◯◯◯ INSUFFICIENT |
| | 84 mos. | No trials | | | No data reported. | ⊕◯◯◯ INSUFFICIENT |

Reasons for downgrading:
1. Serious risk of bias: the majority of studies violated one or more of the criteria for good quality RCT (or cohort study) related to the outcome reported (see Appendix for details)
2. Inconsistency: differing estimates of effects across trials
3. Imprecise effect estimate for a dichotomous outcome: small sample size and/or confidence interval includes both negligible effect and appreciable benefit or harm with ADR. If sample size is likely too small to detect rare outcomes, evidence may be downgraded twice.

## Table D-8. Strength of Evidence Summary: CADR vs. ACDF with a zero-profile device (2 non-contiguous levels) Safety Results

| Outcome | Follow-up | RCTs | N* | Reasons for Downgrading | Conclusion* | Quality |
|---|---|---|---|---|---|---|
| **Secondary surgery at the index level** | Any | | | | No data reported. | ⊕◯◯◯ INSUFFICIENT |
| **Serious/ major adverse events** | Mean 32.4 (24-46) mos. | 1 RCT (Qizhi 2016) | N=30 | Risk of bias[1] (-1) Imprecision[3] (-2) | Conclusion: No serious adverse events were reported. | ⊕◯◯◯ INSUFFICIENT |
| **Device-related adverse events** | Mean 32.4 (24-46) mos. | 1 RCT (Qizhi 2016) | N=30 | Risk of bias[1] (-1) Imprecision[3] (-2) | Conclusion: All events that could be attributed to the device occurred similarly between groups, but no summary of device-related adverse events was reported. | ⊕◯◯◯ INSUFFICIENT |

Reasons for downgrading:
1. Serious risk of bias: the majority of studies violated one or more of the criteria for good quality RCT (or cohort study) related to the outcome reported (see Appendix for details)
2. Inconsistency: differing estimates of effects across trials
3. Imprecise effect estimate for a dichotomous outcome: small sample size and/or confidence interval includes both negligible effect and appreciable benefit or harm with ADR. If sample size is likely too small to detect rare outcomes, evidence may be downgraded twice.

## Table D-9. Strength of Evidence Summary: Differential Efficacy and Safety Results for CADR

| Outcome | Follow-up | RCTs | N* | Reasons for Downgrading | Conclusion | Quality |
|---|---|---|---|---|---|---|
| **Any** | Any | | | | No studies were identified which stratified on patient characteristics or evaluated effect modification. | ⨁◯◯◯ INSUFFICIENT |

## Table D-10. Strength of Evidence Summary: L-ADR vs. Fusion (1-level) Efficacy Results

| Outcome | Follow-up | RCTs | N* | Reasons for Downgrading | Conclusion* | Quality |
|---|---|---|---|---|---|---|
| **L-ADR vs. fusion (1-level)** | | | | | | |
| **Overall success†** | 24 mos. | 2 RCTs (Charité, ProDisc-L IDE trials) | N=484 | Risk of Bias[1] (-1)<br>Imprecision[3] (-1) | Pooled RD 7.9% (95% CI -1.7%, 17.4%),<br>Conclusion: L-ADR is comparable with single level anterior lumbar interbody fusion or circumferential fusion up to 24 months following surgery in terms of the proportion of patients achieving overall clinical success | ⊕⊕◯◯ LOW |
| | 60 mos. | 2 RCTs (Charité, ProDisc-L IDE trials) | N=319 | Risk of Bias[1] (-1)<br>Imprecision[3] (-1) | Pooled RD 7.1%, (95% CI -4.9%, 18.9%)<br>Conclusion: L-ADR is comparable with single level anterior lumbar interbody fusion or circumferential fusion up to 60 months following surgery in terms of the proportion of patients achieving overall clinical success | ⊕⊕◯◯ LOW |
| **ODI success** (≥15-point improvement) | 24 mos. | 2 RCTs (Charité, ProDisc-L IDE trials) | N=485 | Risk of Bias[1] (-1)<br>Imprecision[3] (-1) | Pooled RD 8.9% (95% CI -0.5%, 18.3%),<br>Conclusion: L-ADR is comparable with single level anterior lumbar interbody fusion or circumferential fusion up to 24 months following surgery in terms of the proportion of patients achieving ODI success | ⊕⊕◯◯ LOW |
| | 60 mos. | 2 RCTs (Charité, ProDisc-L IDE trials) | N=310 | Risk of Bias[1] (-1)<br>Imprecision[3] (-1) | Pooled RD 7.8%, (95% CI -3.6%, 19.2%)<br>Conclusion: L-ADR is comparable with single level anterior lumbar interbody fusion or circumferential fusion up to 60 months following surgery in terms of the proportion of patients achieving ODI success | ⊕⊕◯◯ LOW |
| **Neurological success‡** | 24 mos. | 2 RCTs (Charité, ProDisc-L IDE trials) | N=483 | Risk of Bias[1] (-1)<br>Inconsistency[2] (-1)<br>Imprecision[3] (-1) | Pooled RD 2.2%, (95% CI -12.6%, 17.1%)<br>Conclusion: L-ADR is comparable with single level anterior lumbar interbody fusion or circumferential fusion up to 24 months following surgery in terms of the proportion of patients achieving neurological | ⊕◯◯◯ INSUFFICIENT |

| Outcome | Follow-up | RCTs | N* | Reasons for Downgrading | Conclusion* | Quality |
|---|---|---|---|---|---|---|
| | | | | | success | |
| | 60 mos. | 2 RCTs (Charité, ProDisc-L IDE trials) | N=306 | Risk of Bias[1] (-1) Imprecision[3] (-1) | Pooled RD 0.2%, (95% CI -7.9%, 8.3%) Conclusion: L-ADR is comparable with single level anterior lumbar interbody fusion or circumferential fusion up to 60 months following surgery in terms of the proportion of patients achieving neurological success | ⊕⊕◯◯ LOW |
| **VAS Pain scores** (0-100) | 24 mos. | 2 RCTs (Charité, ProDisc-L IDE trials) | N=488 | Risk of Bias[1] (-1) Imprecision[4] (-1) | WMD 6.84, 95% CI 0.63, 12.32) Conclusion: L-ADR may be comparable to fusion with regard to pain relief at 24 months. Neither trial individually reported a significant difference between treatments. Based on pooled estimates, VAS pain at 24 months may be somewhat less following L-ADR compared with fusion (pooled mean difference however the difference is likely not clinically meaningful). | ⊕⊕◯◯ LOW |
| | 60 mos. | 2 RCTs (Charité, ProDisc-L IDE trials) | N=309 | Risk of Bias[1] (-1) Imprecision[4] (-1) | WMD MD 1.16, 95% CI -6.43, 8.74 Conclusion: L-ADR may be as good as fusion with regard to pain relief at 24 months. | ⊕⊕◯◯ LOW |

* Unless otherwise specified, conclusion and patient numbers were based on completer analysis, for which the effect estimate was more conservative than that of the ITT analysis.

† Overall clinical success: The FDA criterion of at least a 15-point improvement from baseline ODI scores was used for both RCTs to minimize heterogeneity in the meta-analysis. The definition of overall clinical success was similar in the two studies, but not identical. In the Prodisc-L trial (Zigeler 2007), success was defined more conservatively than the Charité (Blumenthal 2005) trial in that it required improvement in the SF-36 and radiological success as additional criteria. The addition of these parameters would make success more difficult to achieve resulting in a lower proportion of patients attaining overall clinical success, but not likely biasing the results between study groups. Therefore, these two studies were pooled;

‡ Neurological success was defined as no neurological change (i.e. defined as lack of neurological deterioration compared with preoperative status, at any point of time in the Charité trial and as neurological status improved or maintained (motor, sensory, reflex, straight leg raise) in the ProDisc-L trial.

Reasons for downgrading:

1. Serious risk of bias: the majority of studies violated one or more of the criteria for good quality RCT (or cohort study) related to the outcome reported (see Appendix for details)

2. Inconsistency: differing estimates of effects across trials
3. Imprecise effect estimate for a dichotomous outcome: small sample size and/or confidence interval includes both negligible effect and appreciable benefit or harm with ADR
4. Imprecise effect estimate for a continuous outcome: wide (or unknown) confidence interval and/or small sample size

## Table D-11. Strength of Evidence Summary: L-ADR vs. Fusion (2-level) Efficacy Results

| Outcome | Follow-up | RCTs | N* | Reasons for Downgrading | Conclusion* | Quality |
|---|---|---|---|---|---|---|
| **L-ADR vs. fusion (2-level)** | | | | | | |
| **Overall success†** | 24 mos. | 1 RCTs (Delamarter) | N=215 | Risk of Bias[1] (-1)<br>Imprecision[3] (-1) | RD 11.0% (95% CI -3.3%, 25.4%)<br>Conclusion: At 24 months, 2-level L-ADR is as good as fusion with regard to the proportion of patients achieving clinical success; no statistical difference was observed between treatments observed. | ⊕⊕◯◯ LOW |
| **ODI Scores** (0-100) | | | | Risk of Bias[1] (-1)<br>Imprecision[4] (-1) | MD -8.4 (95% CI -15.4, -1.4)<br>Conclusion: Two-level ADR may be as good as or slightly better than fusion with respect to function measured via ODI. Patients receiving 2-level L-ADR had significant improvement (lower) in ODI scores; It is not clear if this difference is clinically meaningful. Change from baseline for ADR was 52.4% ± 38.1% and for fusion was 40.9% ± 36.0%. | ⊕⊕◯◯ LOW |
| **Neurological success‡** | | | | Risk of Bias[1] (-1)<br>Imprecision[3] (-1) | RD 8.5% (95% CI -2.5%, 19.6%)<br>Conclusion: Two-level ADR may be as good as fusion by 24 months in terms of neurological success; no statistical difference was observed between treatments observed | ⊕⊕◯◯ LOW |
| **VAS Pain scores** (0-100) | | | | Risk of Bias[1] (-1)<br>Imprecision[4] (-1) | MD -6.5 (-15.7, 2.7)<br>Conclusion: Two-level ADR may be as good as fusion with regard to pain relief; no statistical difference was observed between treatments observed | ⊕⊕◯◯ LOW |

* Unless otherwise specified, conclusion and patient numbers were based on completer analysis, for which the effect estimate was more conservative than that of the ITT analysis.

† Overall clinical success: The FDA criterion of at least a 15-point improvement from baseline ODI scores was used, other components of the composite: 1) Improvement in SF-36 PCS compared with baseline; 2) Neurological status improved or maintained from baseline; 3) No secondary surgical procedures to remove or modify the total disc replacement implant or arthrodesis implant/site; 4) no subsidence >3 mm; 5) no migration >3 mm; 6) no radiolucency/loosening; 7) no loss of disc height >3 mm); and 8) for ADR, range of motion improved for maintained from baseline and for Fusion, no motion (<10° angulation, total for two levels combined) on flexion and extension radiographs.

‡ Neurological success was defined as neurological status improved or maintained (motor, sensory, reflex, straight leg raise).

Reasons for downgrading:

1. Serious risk of bias: the majority of studies violated one or more of the criteria for good quality RCT (or cohort study) related to the outcome reported (see Appendix for details)

2. Inconsistency: differing estimates of effects across trials
3. Imprecise effect estimate for a dichotomous outcome: small sample size and/or confidence interval includes both negligible effect and appreciable benefit or harm with ADR
4. Imprecise effect estimate for a continuous outcome: wide (or unknown) confidence interval and/or small sample size

## Table D-12. Strength of Evidence Summary: L-ADR vs. Fusion (1 or 2-level) Efficacy Results

| Outcome | Follow-up | RCTs | N* | Reasons for Downgrading | Conclusion* | Quality |
|---|---|---|---|---|---|---|
| **L-ADR vs. fusion (1- or 2-level)** | | | | | | |
| **Overall success†** | 24 mos. | 1 RCT (Berg/Skold) | N= 152 | Risk of Bias[1] (-1)<br>Imprecision[3] (-1) | RD 5.8% (95% CI -8.8%, 20.5%)<br>Conclusion: L-ADR is comparable to fusion with regard to the proportion of patients who reported being totally pain free or much better. | ⊕⊕◯◯ LOW |
| | 60 mos. | | N=151 | Risk of Bias[1] (-1)<br>Imprecision[3] (-1) | RD 4.9% (95% CI -9.7%, 19.5%)<br>Conclusion: L-ADR is comparable to fusion with regard to the proportion of patients who reported being totally pain free or much better. | ⊕⊕◯◯ LOW |
| **ODI success** (≥ 25% improvement) | 24 mos. | | N= 152 | Risk of Bias[1] (-1)<br>Imprecision[3] (-1) | RD 8.2% (95% CI -7.4%, 23.8%)<br>Conclusion: L-ADR is comparable to fusion with regard to the proportion of patients who achieved ODI success | ⊕⊕◯◯ LOW |
| | 60 mos. | | N=151 | Risk of Bias[1] (-1)<br>Imprecision[4] (-1) | RD 12.7% (95% CI -1.7%, 27.1%)<br>Conclusion: L-ADR is comparable to fusion with regard to the proportion of patients achieved ODI success. | ⊕⊕◯◯ LOW |
| **Back Pain VAS scores** (0-100) | 24 mos. | | N= 152 | Risk of Bias[1] (-1)<br>Imprecision[4] (-1) | MD -3.8 (95% CI -12.6, 5.0)<br>Conclusion: L-ADR is comparable to fusion with regard to back pain relief at 24 months. | ⊕⊕◯◯ LOW |
| | 60 mos. | | N=151 | Risk of Bias[1] (-1)<br>Imprecision[4] (-1) | MD -7.8 (-16.9, 1.3)<br>Conclusion: L-ADR is comparable to fusion with regard to back pain relief at 60 months | ⊕⊕◯◯ LOW |
| **Leg Pain VAS scores** (0-100) | 24 mos. | | N= 152 | Risk of Bias[1] (-1)<br>Imprecision[4] (-1) | MD -4.3 (-12.1, 3.5)<br>Conclusion: L-ADR is comparable to fusion with regard to leg pain relief at 24 months. | ⊕⊕◯◯ LOW |
| | 60 mos. | | N=151 | Risk of Bias[1] (-1)<br>Imprecision[4] (-1) | MD -6.3 (-14.0, 1.4) | ⊕⊕◯◯ LOW |

| Outcome | Follow-up | RCTs | N* | Reasons for Downgrading | Conclusion* | Quality |
|---|---|---|---|---|---|---|
| | | | | | Conclusion: L-ADR is comparable to fusion with regard to leg pain relief at 60 months | |

| | | | | | | |
|---|---|---|---|---|---|---|
| **SF-36 pain subscale** (0-100 [best]) | 60 mos. | | N= 151 | Risk of Bias[1] (-1)<br>Imprecision[4] (-1) | MD 10.8 (1.2, 20.4)<br>Conclusion: L-ADR is comparable to or slightly better than fusion with regard at 60 months; It is not clear that the difference in SF-36 pain scores is clinically meaningful. | ⊕⊕◯◯ LOW |

* Unless otherwise specified, conclusion and patient numbers were based on completer analysis, for which the effect estimate was more conservative than that of the ITT analysis. For this trial, authors report no loss to follow-up at 24 months; however it is not clear if there were randomized patients who did not receive the allotted treatment.

† Overall clinical success was defined differently in the Berg 2009 (totally pain free) and Skold 2013 (totally pain free OR much better) publications; The latter definition is used here as it is more conservative; Using the definition of "totally pain free" RDs at 24 months ( RD 22.2% , 95%CI 8.8, 35.7)and 60 months RD 22.0% (95% CI 8.5, 35.5)suggest L-ADR is better than fusion however substantial imprecision is noted and strength of evidence is low. Full detail is provided in the report.

Reasons for downgrading:

1. Serious risk of bias: the majority of studies violated one or more of the criteria for good quality RCT (or cohort study) related to the outcome reported (see Appendix for details)
2. Inconsistency: differing estimates of effects across trials
3. Imprecise effect estimate for a dichotomous outcome: small sample size and/or confidence interval includes both negligible effect and appreciable benefit or harm with ADR
4. Imprecise effect estimate for a continuous outcome: wide (or unknown) confidence interval and/or small sample size

## Table D-13. Strength of Evidence Summary: L-ADR vs. Multidisciplinary Rehabilitation Efficacy Results

| Outcome | Follow-up | RCTs | N* | Reasons for Downgrading | Conclusion* | Quality |
|---|---|---|---|---|---|---|
| **L-ADR vs. Multidisciplinary Rehabilitation** | | | | | | |
| **Overall success/ODI success†** (≥15-point improvement in ODI) | 24 mos. | 1 RCT (Hellum) | N=139 | Risk of Bias[1] (-1)<br>Imprecision[3] (-1) | RD 22.9% (95% CI 6.9%, 38.9%)<br>Conclusion: L-ADR appears to be superior to multidisciplinary rehabilitation; the proportions of L-ADR participants achieving clinical success based on ODI improvement of at least 15 points is significantly higher (57.3%) than the proportion in the rehabilitation group (34.4%). | ⊕⊕◯◯ LOW |
| **VAS Pain scores** (0-100) | | | | Risk of Bias[1] (-1)<br>Imprecision[4] (-1) | MD -14.3 (95% CI -23.0, -5.6)<br>Conclusion: Results for VAS pain scores for suggest that L-ADR may be associated with less pain at 24 months compared with multidisciplinary rehabilitation however, baseline low back pain scores were significantly worse in the rehabilitation group than in the surgery group | ⊕⊕◯◯ LOW |

* Unless otherwise specified, conclusion and patient numbers were based on completer analysis, for which the effect estimate was more conservative than that of the ITT analysis.

† Overall clinical success: The FDA criterion of at least a 15-point improvement from baseline ODI scores was used to define clinical success

Reasons for downgrading:

1. Serious risk of bias: the majority of studies violated one or more of the criteria for good quality RCT (or cohort study) related to the outcome reported (see Appendix for details)
2. Inconsistency: differing estimates of effects across trials
3. Imprecise effect estimate for a dichotomous outcome: small sample size and/or confidence interval includes both negligible effect and appreciable benefit or harm with ADR
4. Imprecise effect estimate for a continuous outcome: wide (or unknown) confidence interval and/or small sample size

## Table D-14. Strength of Evidence Summary: L-ADR vs. Fusion (1-level) Safety Results

| Outcome | Follow-up | RCTs | N* | Reasons for Downgrading | Conclusion* | Quality |
|---|---|---|---|---|---|---|
| **L-ADR vs. fusion (1-level)** | | | | | | |
| **Secondary Surgery at Index Level+** | 24 mos. | 2 RCTs (Charité, ProDisc-L IDE trials) | N=540 | Risk of Bias[1] (-1) Imprecision[3] (-1) | L-ADR 4.9%, Fusion 1.4% Pooled RD 2.3% (95% CI -2.1%, 6.6%) Conclusion: L-ADR is comparable with single level anterior lumbar interbody fusion or circumferential fusion up to 24 months following surgery in terms of the proportion of patients who had subsequent surgery at the index level. | ⊕⊕◯◯ LOW |
| | 24–60 mos. | 1 RCTs (ProDisc-L IDE trial) | N=236 | Risk of Bias[1] (-1) Imprecision[3] (-1) | L-ADR 6.6%, Fusion 3.7% RD 2.9% (95% CI -3.4%, 9.3%) Conclusion: L-ADR is comparable with single level anterior lumbar interbody fusion or circumferential fusion between 24 and 60 months following surgery in terms of the proportion of patients | ⊕⊕◯◯ LOW |
| | 60 mos. | 1 RCTs (ProDisc-L IDE trial) | N=236 | Risk of Bias[1] (-1) Imprecision[3] (-1) | L-ADR 12.0%, Fusion 8.1% RD 3.9% (95% CI -4.6%, 12.4% Conclusion: L-ADR is comparable with single level anterior lumbar interbody fusion or circumferential fusion up to 24 months following surgery in terms of the proportion of patients achieving ODI success | ⊕⊕◯◯ LOW |

| Outcome | Follow-up | RCTs | N* | Reasons for Downgrading | Conclusion* | Quality |
|---|---|---|---|---|---|---|
| **Major Adverse Events‡** | 24 mos. | 2 RCTs (Charité, ProDisc-L IDE trials) | N=540 | Risk of Bias[1] (-1)<br>Imprecision[3] (-2) | Frequency ≤ 1% of patients for both treatments across both trials.<br>Conclusion: Firm conclusions regarding the comparability of L-ADR and fusion regarding the frequency of major adverse events are not possible: sample sizes may be inadequate to detect rare events. It is possible that reported frequency of such events is underestimated. | ⊕◯◯◯<br>INSUFFICIENT |
| **Major‡, serious or life-threatening adverse event§** | 60 mos. | 2 RCTs (Charité, ProDisc-L IDE trials) | N=133 (Charité)<br>N=236 (ProDisc-L) | Risk of Bias[1] (-1)<br>Inconsistency (-1)<br>Imprecision[3] (-1) | Charité: No major adverse events were reported for L-ADR or fusion, however the small sample size and substantial loss to follow-up preclude drawing firm conclusion **<br>Prodisc- L: Serious or life-threatening event risks for L-ADR were 0.58 per patient, fusion 0.38 per patient, p = 0.036; They appear to be more common with L-ADR than with fusion.<br>Conclusion: Firm conclusions regarding the comparability L-ADR and fusion across these studies at 60 months is not possible. Differing definitions of what may constitute such events may impact the discrepancy across studies in addition to factors related to the population available for the Charite trial at 60 months. | ⊕◯◯◯<br>INSUFFICIENT |
| **Device-related adverse events (excluding secondary surgery at index level)** | 24 mos. | 2 RCTs (Charité, ProDisc-L IDE trials) | N=540 | Risk of Bias[1] (-1)<br>Imprecision[3] (-1) | L-ADR 11.5%, fusion 9.2%<br>Pooled RD -2.7% (95% CI -7.4 %, 1.9%)<br>Conclusion: L-ADR is comparable with single level anterior lumbar interbody fusion or circumferential fusion up to 60 months following surgery in terms of the proportion of patients | ⊕⊕◯◯<br>LOW |

| Outcome | Follow-up | RCTs | N* | Reasons for Downgrading | Conclusion* | Quality |
|---|---|---|---|---|---|---|
| **Any Adverse Event** | 24 mos. | 2 RCTs (Charité, ProDisc-L IDE trials) | N=540 | Risk of Bias[1] (-1)<br>Imprecision[4] (-1) | L-ADR 79.5%, fusion 84.5%<br>Pooled RD 6.2% (95% CI -0.7%, 13.0%)<br>Conclusion: L-ADR may be comparable to fusion with regard to experiencing any adverse event by 24 months. | ⊕⊕◯◯ LOW |
| | 60 mos. | 1 RCT (ProDisc-L IDE trials) | N=236 | Risk of Bias[1] (-1)<br>Imprecision[4] (-1) | L-ADR 5.1 per patient, fusion 5.4 per patient, p = 0.507<br>Conclusion: L-ADR may be comparable with fusion with regard frequency of any adverse event by 60 months. | ⊕⊕◯◯ LOW |

*Percentages were calculated based on the number of patients who received treatment (i.e., excludes those who dropped out after randomization but prior to undergoing surgery) unless otherwise noted.

†Secondary surgery at index level included revision, reoperation, device/hardware removal, supplemental fixation, hemilaminectomy and discectomy with decompression

‡Major adverse event defined as major vessel injury, neurological damage, nerve root injury, and death.

§Zigler 2012 does not provide detail regarding what constitutes a serious or life threatening event; unclear if these events were defined the same way as "major adverse events" for the ProDisc-L trial at 24 months.

**For the Charite IDE trial, of the 14 initial sites, 6 declined participation in the 60-month continuation study, which eliminated 64 randomized patients and only those with both 24 month and 60 month data were included.

Reasons for downgrading:

1. Serious risk of bias: the majority of studies violated one or more of the criteria for good quality RCT (or cohort study) related to the outcome reported (see Appendix for details)
2. Inconsistency: differing estimates of effects across trials
3. Imprecise effect estimate for a dichotomous outcome: small sample size and/or confidence interval includes both negligible effect and appreciable benefit or harm with ADR. If sample size is likely too small to detect rare outcomes, evidence may be downgraded twice.
4. Imprecise effect estimate for a continuous outcome: wide (or unknown) confidence interval and/or small sample size

## Table D-15. Strength of Evidence Summary: L-ADR vs. Fusion (2-level) Safety Results

| Outcome | Follow-up | RCTs | N* | Reasons for Downgrading | Conclusion* | Quality |
|---|---|---|---|---|---|---|
| **L-ADR vs. Fusion (2-level)** | | | | | | |
| **Secondary surgical procedure at index level(s)†** | 24 mos. | 1 RCTs (Delamarter) | N=237 | Risk of Bias[1] (-1)<br>Imprecision[3] (-1) | L-ADR 2.4%, fusion 8.3%<br>RD -5.9% (95% CI -12.7%, 0.09%)<br>Conclusion: At 24 months, additional surgery at the index level was less common following 2-level L-ADR vs. fusion. | ⊕⊕◯◯<br>LOW |
| **Major surgery-related complications‡** | | | | Risk of Bias[1] (-1)<br>Imprecision[4] (-1) | L-ADR 0.7%, fusion 4.9%<br>RD -6.7% (95% CI -14.0%, 0.6%)<br>Conclusion: Major surgery-related complications were less common with L- ADR compared with fusion, however there was no statistical difference between groups, perhaps partly due to sample size. | ⊕⊕◯◯<br>LOW |
| **Device related complications (Subsidence or migration)§** | 24 mos. | 1 RCTs (Delamarter) | N=237 | Risk of Bias[1] (-1)<br>Imprecision[3] (-2) | L-ADR 2.4%, Fusion 1.4%<br>RD 1.0% (-2.5%, 4.6%)<br>Conclusion: There was no statistical difference between groups; however, this may in part be a function of sample size. The frequency of device-related events may be underestimated. | ⊕◯◯◯<br>INSUFFICIENT |

* Percentages were calculated based on the number of patients who received treatment (i.e., excludes those who dropped out after randomization but prior to undergoing surgery) unless otherwise noted.

† Includes revision (1 ADR, 1 fusion), decompression (3 ADR, 1 fusion), and device/implant removal (0 ADR, 6 fusion). One fusion patients underwent implant removal, decompression and revision of the bone fusion sites due to pseudarthrosis at L5-S1; this patient is only counted once in the overall estimate.

‡ Included dural tear (1 ADR, 3 fusion; all successful repaired), blood loss >1500 mL (2 ADR, 2 fusion; 1 iliac artery tear in ADR group while all others had excessive oozing from the surgical site), and deep vein thrombosis (2 ADR, 2 fusion; all successfully treated).

§ Based on radiographic evaluation, implant subsidence of >3 mm for L-ADR patients (not clinically relevant) or migration and implant migration or subsidence of > 3mm was reported for fusion. There was one anterior migration of L-ADR which resulted in need for revision.

Reasons for downgrading:

1. Serious risk of bias: the majority of studies violated one or more of the criteria for good quality RCT (or cohort study) related to the outcome reported (see Appendix for details)
2. Inconsistency: differing estimates of effects across trials
3. Imprecise effect estimate for a dichotomous outcome: small sample size and/or confidence interval includes both negligible effect and appreciable benefit or harm with ADR. If sample size is likely too small to detect rare outcomes, evidence may be downgraded twice.
4. Imprecise effect estimate for a continuous outcome: wide (or unknown) confidence interval and/or small sample size.

## Table D-16. Strength of Evidence Summary: L-ADR vs. Fusion (1- or 2-level) Safety Results

| Outcome | Follow-up | RCTs | N* | Reasons for Downgrading | Conclusion* | Quality |
|---|---|---|---|---|---|---|
| **L-ADR vs. fusion (1- or 2-level)** | | | | | | |
| **Any Secondary Surgical Procedure at Index Level†** | 24 mos. | 1 RCT (Berg/Skold) | N= 152 | Risk of Bias[1] (-1)<br>Imprecision[3] (-1) | L-ADR 10.0%, fusion 30.6%<br>RD -20.6% (-33.1%, -8.1%)<br>Conclusion: L-ADR was associated with significantly fewer secondary surgeries compared with fusion up to 24 months; the majority were device related | ⊕⊕◯◯ LOW |
| | 60 mos. | | N=151 | Risk of Bias[1] (-1)<br>Imprecision[3] (-1) | L-ADR 17.5%, fusion 36.6%<br>RD -19.1% (-33.1%, -5.2%)<br>Conclusion: L-ADR was associated with significantly fewer secondary surgeries compared with fusion through 60 months; the majority was device related. | ⊕⊕◯◯ LOW |
| **Device-related reoperation†** | 24 mos. | | N= 152 | Risk of Bias[1] (-1)<br>Imprecision[3] (-1) | L-ADR 5.0%, fusion 27.8%<br>RD -22.8% (95% CI -34.2%, -11.4%)<br>Conclusion: L-ADR was associated fewer device-related surgeries compared with fusion up to 24 months; these are the only device-related adverse events that authors report. | ⊕⊕◯◯ LOW |
| | 60 mos. | | N=151 | Risk of Bias[1] (-1)<br>Imprecision[3] (-1) | L-ADR 11.3%, fusion 28.2%<br>RD -16.9% (95% CI -29.5%, -4.4%)<br>Conclusion: L-ADR was associated fewer device-related surgeries compared with fusion through 60 months; these are the only device-related adverse events that authors report. | ⊕⊕◯◯ LOW |
| **Total major complications§** | 60 mos. | | N= 152 | Risk of Bias[1] (-1)<br>Imprecision[3] (-1) | L-ADR 2.5%, fusion 8.3%<br>RD -5.8% (95% CI -13.1%, 1.4%)<br>Conclusion: Fewer major complications occurred following L-ADR compared with fusion; however statistical significance was not reached, possibly in part due to sample size. All events occurred within 24 months with no additional events reported through 60 months. | ⊕⊕◯◯ LOW |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Any (total) complication** | 60 mos. | | N= 152 | Risk of Bias[1] (-1)<br>Imprecision[4] (-1) | L-ADR 17.5%, fusion 20.8%<br>RD -3.3% (95% CI -15.9%, 9.2%)<br>Conclusion: L-ADR was comparable to fusion with regard to frequency of any complications through 24 months. All events occurred within 24 months with no additional events reported through 60 months. | ⊕⊕◯◯ LOW |

* Percentages were calculated based on the number of patients who received treatment (i.e., excludes those who dropped out after randomization but prior to undergoing surgery) unless otherwise noted.
† Based on authors' description: Subsequent device-related procedures included subsequent fusion (in the ADR group), pedicle screw extraction due to pain or irritation. Non-device related secondary procedures includes decompression, decompression + pedicle screw extraction, re-fusion, hematoma removal, hernia repair and repair of dural tear.
§ Major complications include deep infection (4 fusion), pseudarthrosis (2 fusion), nerve entrapment (1 ADR), and subsidence/reoperation (1 ADR).
Reasons for downgrading:

1. Serious risk of bias: the majority of studies violated one or more of the criteria for good quality RCT (or cohort study) related to the outcome reported (see Appendix for details)
2. Inconsistency: differing estimates of effects across trials
3. Imprecise effect estimate for a dichotomous outcome: small sample size and/or confidence interval includes both negligible effect and appreciable benefit or harm with ADR. If sample size is likely too small to detect rare outcomes, evidence may be downgraded twice.
4. Imprecise effect estimate for a continuous outcome: wide (or unknown) confidence interval and/or small sample size

## Table D-17. Strength of Evidence Summary: L-ADR vs. Multidisciplinary Rehabilitation Safety Results

| Outcome | Follow-up | RCTs | N* | Reasons for Downgrading | Conclusion* | Quality |
|---|---|---|---|---|---|---|
| **L-ADR vs. Multidisciplinary Rehabilitation** | | | | | | |
| **Secondary Surgery at Index Level†** | 24 mos. | 1 RCT (Hellum) | N=77 | Risk of Bias[1] (-1)<br>Imprecision[3] (-1) | L-ADR: 6.5% (5/77)<br>Conclusion: Secondary surgery risk is only applicable to the L-ADR group; conclusions regarding comparative safety are not possible | ⊕⊕◯◯ LOW |
| **Major complication resulting in impairment‡** | | | | Risk of Bias[1] (-1)<br>Imprecision[4] (-1) | L-ADR: 7.8% (6/77)<br>Conclusion: Conclusions regarding comparative safety are not possible. As defined in this study, major complications resulting in impairment are only applicable to those receiving L-ADR. | ⊕⊕◯◯ LOW |
| **Any complication§** | | | | Risk of Bias[1] (-1)<br>Imprecision[4] (-1) | L- ADR: 33.8% (26/77)<br>Conclusion: Over 1/3 of L-ADR recipient s experienced some type of complication. | ⊕⊕◯◯ LOW |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | Conclusions regarding comparative safety with respect to any complications as defined are not possible; authors do not provide information on any events in the rehabilitation group. | |

* ITT analyses are based on the baseline, as-treated population: Six patients (3 in each group) were excluded shortly after randomization and not accounted for in the studies analyses. Safety events were only reported for L-ADR, thus although the total study populations was 139, only 77 received ADR.
† Surgeries included fusion at level with disc prosthesis and level above (n=1); insertion of new polyethylene inlay (n=1); and partial resection of spinous process because of possible painful contact between adjacent levels (n=2)
‡ Includes: polyethylene inlay dislodgement requiring revision surgery, during which injury to the left common iliac artery led to compartment syndrome resulting in a lower leg amputation (n=1); arterial thrombosis of dorsalis pedis artery resulting in a slightly colder foot (n=1); retrograde ejaculation (n=1); sensory loss in the thigh (n=2); and new radicular pain (n=2); there were a total of 7 events in 6 patients.
§ Includes "major complications resulting in impairment" as well as perioperative and other surgery-related adverse events such as dural tear, blood loss >1500 mL, hematoma, infection, etc. Authors report the most frequent treatment-related events as blood loss >1500 mL; temporary sensory loss and temporary radicular pain occurring in 5.2% of LADR patients (4/77). It is not clear if patients could experience more than one complication.

Reasons for downgrading:

1. Serious risk of bias: the majority of studies violated one or more of the criteria for good quality RCT (or cohort study) related to the outcome reported (see Appendix for details)
2. Inconsistency: differing estimates of effects across trials
3. Imprecise effect estimate for a dichotomous outcome: small sample size and/or confidence interval includes both negligible effect and appreciable benefit or harm with ADR. If sample size is likely too small to detect rare outcomes, evidence may be downgraded twice.
4. Imprecise effect estimate for a continuous outcome: wide (or unknown) confidence interval and/or small sample size

## Table D-18. Strength of Evidence Summary: Differential Efficacy and Safety Results for L-ADR

| Outcome | Follow-up | RCTs | N* | Reasons for Downgrading | Conclusion | Quality |
|---|---|---|---|---|---|---|
| **L-ADR vs. Fusion or Multidisciplinary Rehabilitation** | | | | | | |
| Any | Any | | | | No studies were identified which stratified on patient characteristics or evaluated effect modification. | ⊕◯◯◯ INSUFFICIENT |

## Appendix E. Search Strategy

Search date: May 21, 2024

| # | Search Query | Retrieval |
|---|---|---|
| 1 | (((((("Intervertebral Disc Degeneration"[Mesh]) OR "Intervertebral Disc Displacement"[Mesh]) OR "Intervertebral Disc"[Mesh]) OR "Intervertebral disc disease" [Supplementary Concept]) AND (arthroplasty[tiab] OR replacement[tiab] OR artificial[tiab])) OR (((((artificial[TI] OR prosthetic*[TI] OR prosthes*[TI] OR total[TI] OR replacement*[TI] OR arthroplasty[TI]) AND (Disk[TI] OR Disc[TI] OR Intervertebral[TI] OR vertebr*[TI] OR spine[TI] OR spinal[TI])) AND (cervical[TI] OR neck[TI])) NOT (Medline[sb])) OR ((((Disk[TI] OR Disc[TI] OR Intervertebral[TI] OR vertebr*[TI] OR spine[TI] OR spinal[TI]) AND (artificial[TI] OR prosthetic*[TI] OR prosthes*[TI] OR total[TI] OR replacement*[TI] OR arthroplasty[TI] OR activl[TI] OR activ-l[TI] OR prodisc[TI])) AND ((lumbar[TI] OR "low-back"[TI] OR "low back"[TI] OR "lower back"[TI]))) NOT (Medline[sb]))))) OR (Total Disc Replacement[MeSH])) OR ("Bryan Cervical Disc" OR "M6-C Artificial Cervical Disc" OR "MOBI-C" OR "Prestige Cervical Disc" OR "ProDisc-C" OR "Secure-C" OR "Simplify Cervical Artificial Disc" OR "Prestige LP") | 2802 |
| 2 | (((((("Intervertebral Disc Degeneration"[Mesh]) OR "Intervertebral Disc Displacement"[Mesh]) OR "Intervertebral Disc"[Mesh]) OR "Intervertebral disc disease" [Supplementary Concept]) AND (arthroplasty[tiab] OR replacement[tiab] OR artificial[tiab])) OR (((((artificial[TI] OR prosthetic*[TI] OR prosthes*[TI] OR total[TI] OR replacement*[TI] OR arthroplasty[TI]) AND (Disk[TI] OR Disc[TI] OR Intervertebral[TI] OR vertebr*[TI] OR spine[TI] OR spinal[TI])) AND (cervical[TI] OR neck[TI])) NOT (Medline[sb])) OR ((((Disk[TI] OR Disc[TI] OR Intervertebral[TI] OR vertebr*[TI] OR spine[TI] OR spinal[TI]) AND (artificial[TI] OR prosthetic*[TI] OR prosthes*[TI] OR total[TI] OR replacement*[TI] OR arthroplasty[TI] OR activl[TI] OR activ-l[TI] OR prodisc[TI])) AND ((lumbar[TI] OR "low-back"[TI] OR "low back"[TI] OR "lower back"[TI]))) NOT (Medline[sb]))))) OR (Total Disc Replacement[MeSH])) OR ("Bryan Cervical Disc" OR "M6-C Artificial Cervical Disc" OR "MOBI-C" OR "Prestige Cervical Disc" OR "ProDisc-C" OR "Secure-C" OR "Simplify Cervical Artificial Disc" OR "Prestige LP"); *Limited to Publication dates: 2016 - 2024* | 1220 |
| 3 | (((((("Intervertebral Disc Degeneration"[Mesh]) OR "Intervertebral Disc Displacement"[Mesh]) OR "Intervertebral Disc"[Mesh]) OR "Intervertebral disc disease" [Supplementary Concept]) AND (arthroplasty[tiab] OR replacement[tiab] OR artificial[tiab])) OR (((((artificial[TI] OR prosthetic*[TI] OR prosthes*[TI] OR total[TI] OR replacement*[TI] OR arthroplasty[TI]) AND (Disk[TI] OR Disc[TI] OR Intervertebral[TI] OR vertebr*[TI] OR spine[TI] OR spinal[TI])) AND (cervical[TI] OR neck[TI])) NOT (Medline[sb])) OR ((((Disk[TI] OR Disc[TI] OR Intervertebral[TI] OR vertebr*[TI] OR spine[TI] OR spinal[TI]) AND (artificial[TI] OR prosthetic*[TI] OR prosthes*[TI] OR total[TI] OR replacement*[TI] OR arthroplasty[TI] OR activl[TI] OR activ-l[TI] OR prodisc[TI])) AND ((lumbar[TI] OR "low-back"[TI] OR "low back"[TI] OR "lower back"[TI]))) NOT (Medline[sb]))))) OR (Total Disc Replacement[MeSH])) OR ("Bryan Cervical Disc" OR "M6-C Artificial Cervical Disc" OR "MOBI-C" OR "Prestige Cervical Disc" OR "ProDisc-C" OR "Secure-C" OR "Simplify Cervical Artificial Disc" OR "Prestige LP") AND (2016:2024[pdat]); *Limited to Publication type: Meta-Analysis, Systematic Review* | 104 |
| 4 | (((((("Intervertebral Disc Degeneration"[Mesh]) OR "Intervertebral Disc Displacement"[Mesh]) OR "Intervertebral Disc"[Mesh]) OR "Intervertebral disc disease" [Supplementary Concept]) AND (arthroplasty[tiab] OR replacement[tiab] OR artificial[tiab])) OR (((((artificial[TI] OR prosthetic*[TI] OR prosthes*[TI] OR total[TI] OR replacement*[TI] OR arthroplasty[TI]) AND (Disk[TI] OR Disc[TI] OR Intervertebral[TI] OR vertebr*[TI] OR spine[TI] OR spinal[TI])) AND (cervical[TI] OR neck[TI])) NOT (Medline[sb])) OR ((((Disk[TI] OR Disc[TI] OR Intervertebral[TI] OR vertebr*[TI] OR spine[TI] OR spinal[TI]) AND (artificial[TI] OR prosthetic*[TI] OR | 52 |

| | | |
|---|---|---|
| | prosthes*[TI] OR total[TI] OR replacement*[TI] OR arthroplasty[TI] OR activl[TI] OR activ-l[TI] OR prodisc[TI])) AND ((lumbar[TI] OR "low-back"[TI] OR "low back"[TI] OR "lower back"[TI])) NOT (Medline[sb]))) OR (Total Disc Replacement[MeSH])) OR ("Bryan Cervical Disc" OR "M6-C Artificial Cervical Disc" OR "MOBI-C" OR "Prestige Cervical Disc" OR "ProDisc-C" OR "Secure-C" OR "Simplify Cervical Artificial Disc" OR "Prestige LP") AND (2016:2024[pdat]); *Limited to Publication type: Randomized controlled trial* | |
| 5 | #3 OR #4 | 156 |

**Abbreviations:** MeSH: Medical Subject Heading; pdat: publication date; sb: subset; TI: title: tiab: title/abstract

# Appendix F. Detailed Systematic Review Tables

## Table F-1. Summary of systematic reviews of CADR

| Citation<br>Search Dates | Condition<br>Intervention<br>Comparator | Studies Included in the Review | Outcomes | Results | Review Authors' Conclusions |
|---|---|---|---|---|---|
| Ortiz-Babilonia, 2024[6]<br>Search dates not reported | Condition<br>Degenerative cervical disc disease<br>Intervention<br>CADR<br>Comparator<br>Anterior cervical discectomy and fusion (ACDF) | 25 RCTs published between 2007 and 2021<br>Devices represented<br>*ActivC*<br>Bryan<br>*Discover*<br>Mobi-C<br>PCM<br>Prestige<br>ProDisc-C<br>Secure-C | Overall pain<br>Neck pain<br>Radicular arm pain<br>NDI<br>mJOA<br>ASD (overall, superior-level, and inferior-level) | Effect size/ numerical values not reported. As reported by SR authors: > indicates significantly better <, indicates significantly worse<br><br>**Overall pain (k=6)**<br>No difference (k=6)<br>**Neck pain (k=17)**<br>CADR>ACDF (k=1)<br>No difference (k=16)<br>**Radicular arm pain (k=16)**<br>CADR>ACDF (k=2)<br>No difference (k=14)<br>**Neck disability (k=21)**<br>CADR>ACDF (k=2)<br>CADR<ACDF (k=2)<br>No difference (k=17)<br>**mJOA (k=5)**<br>No difference (k=5)<br>**ASD overall (k=4)**<br>CADR>ACDF (k=2)<br>No difference (k=2)<br>**ASD, superior-level (k=4)**<br>CADR>ACDF (k=2)<br>No difference (k=2)<br>**ASD, inferior-level (k=5)**<br>CADR>ACDF (k=1)<br>No difference (k=4) | Most of the included studies found no significant difference between CADR and ACDF<br><br>Studies had fair to moderate statistical robustness and do not seem to suffer from statistical fragility. |

| Citation<br>Search Dates | Condition<br>Intervention<br>Comparator | Studies Included in the Review | Outcomes | Results | Review Authors' Conclusions |
|---|---|---|---|---|---|
| Kim, 2023[5]<br>2000 – June 2023 | Condition<br>Symptomatic cervical degenerative disc disease<br>Intervention<br>CADR<br>Comparator<br>ACDF | 14 RCTs (all assessed as low risk of bias) published between 2014 and 2023<br>Devices represented<br>*ActivC*<br>Bryan<br>*Discover*<br>*Kineflex-C*<br>Mobi-C<br>Prestige LP<br>ProDisc-C<br>Secure-C | Arm pain score<br>Neck pain score<br>NDI<br>Neurological success<br>QoL (SF-36)<br>ASD (radiographic)<br>Symptomatic adjacent-level disease<br>Secondary surgery<br>Adverse events | **Arm Pain (k=10)**<br>SMD =−0.12; 95% CI: −0.22 to 0.02; P= 0.22; I2= 20%)<br>**Neck Pain Score (k=10)**<br>SMD =−0.17; 95% CI: −0.39 to 0.06; P=0.14; I2=83%<br>**Neurological success (k=9)**<br>RR =0.95, 95% CI: 0.90–1.00, P= 0.04; I2=58%<br>**NDI (k=11)**[e]<br>SMD = −0.10; 95% CI: −0.19 to −0.01; P=0.03; I2=91%<br>**SF-36 PCS (k=7)**<br>SMD =0.15; 95% CI:0.02 to 0.27; P =0.02; I2=34%<br>**Radiographic ASD (k=7)**<br>RR = 0.58, 95% CI: 0.49 to 0.68, P <0.00001; I2=37%<br>**Symptomatic adjacent-level disease (k=11)**<br>RR = 0.44, 95% CI: 0.32 to 0.61, P < 0.00001, I2=16%<br>**Secondary surgical procedures (k=15)**<br>RR =0.49, 95% CI: 0.35–0.70; P<0.0001; I2=61%<br>**Adverse events (k=11)**<br>RR = 0.96; 95% CI: 0.86 to 1.08; P= 0.51; I2=55% | **CADR is better than ACDF** for NDI, QoL, ASD (radiographic and symptomatic) and reoperation<br>**CADR is similar to ACDF** for arm pain, neck pain, and adverse events |
| Selph 2023[7]<br>1980 - 2023<br>AHRQ Commissioned Review | Condition<br>Symptomatic cervical DDD | 22 RCTs (6 high risk of bias, 6 low risk of bias, and 10 moderate risk of bias) | 1-level CADR<br>Overall success<br>Pain | 1-level CADR vs. ACDF<br>**Overall success at short, intermediate, or long term** | **CADR is superior to ACDF** for reoperation (large effect; high SOE for 1-level; low |

[e] Due to high level of heterogeneity, authors assessed NDI at 60 months, 84 months, and more than 108 months and found no difference between CADR and ACDF

| Citation<br>Search Dates | Condition<br>Intervention<br>Comparator | Studies Included in the Review | Outcomes | Results | Review Authors' Conclusions |
|---|---|---|---|---|---|
| *Note: summarized only the studies that compared CADR vs. ACDF*[f] | Intervention<br>Cervical disc replacement, and others not relevant to this review<br>Comparator<br>One of the surgical interventions listed under "Intervention" above or non-surgical treatment (e.g., heat, exercise, acupuncture, drugs, radiofrequency ablation, steroid injections, Botox® for neck pain, psychological strategies [e.g., cognitive behavioral therapy], occupational therapy, multidisciplinary rehabilitation) | 2 prospective, multicenter FDA IDE trials of newer cervical arthroplasty devices (M6-C and Simplify) vs. historic ACDF controls<br>8 non-IDE NRSIs for evaluation of harms<br>7 database/registry studies<br>1 post-hoc analysis of an FDA IDE trial<br>Devices represented<br>*ActivC*<br>Bryan<br>*Discover*<br>*Kineflex-C*<br>Mobi-C<br>Prestige LP<br>ProDisc-C<br>Secure-C | Neurologic status<br>Function<br>Reoperation<br>SAEs<br>Mortality<br>2-level CADR<br>Pain<br>Neurologic status<br>Function<br>Reoperation<br>SAEs<br>Mortality<br>World Health Organization (WHO) Grade 3 or 4 (scale 0-4, 4 most serious) adverse events<br>Neurological events or deficits<br>Mortality | No difference at short and intermediate term<br>CADR slightly better than ACDF at long term<br>**Pain (neck or arm) at short, intermediate, or long term**<br>No difference (moderate SOE)<br>**Neurologic status**<br>No difference (Moderate SOE)<br>**General function at short, intermediate, or long term**<br>No difference (moderate SOE):<br>**Reoperation at the index level**<br>Substantially lower likelihood for CADR vs. ACDF (High SOE)<br>**SAE at short term**<br>Slightly lower likelihood for CADR vs. ACDF (Low SOE)<br>**SAE at >24 months**<br>No difference<br>**Neurological events or deficits at short, intermediate, or long term**<br>No difference (Low SOE)<br>**Mortality**<br>Inadequate evidence (SOE: Insufficient)<br>2-level CADR vs. ACDF<br>**Pain (neck or arm) at short, intermediate, or long term**<br>No difference (Moderate SOE):<br>**Neurologic function at short, intermediate, or long term** | SOE for 2-level) and SAEs (small effect; low SOE)<br>**CADR is comparable to ACDF** for pain (moderate SOE), function (moderate SOE), and neurological events (low SOE for 1-level)<br>**Insufficient evidence for CADR vs. ACDF** for neurological events following 2-level arthroplasty |

[f] 114 studies in 140 publications included for all KQs; 57 RCTs, 56 observational studies, 1 SR

| Citation<br>Search Dates | Condition<br>Intervention<br>Comparator | Studies Included in the Review | Outcomes | Results | Review Authors' Conclusions |
|---|---|---|---|---|---|
| | | | | No difference (Moderate SOE)<br>**General function at short, intermediate, or long term**<br>No difference (Moderate SOE)<br>**Reoperation at the index level**<br>Substantially lower likelihood for CADR vs. ACDF (Low SOE)<br>**SAE at 24 months**<br>Slightly lower likelihood for CADR vs. ACDF (Low SOE)<br>**World Health Organization (WHO) Grade 3 or 4 (scale 0-4, 4 most serious) adverse events 120 months**<br>No difference (Low SOE)<br>**Neurological events or deficits at short, intermediate, or long term**<br>Inadequate evidence (SOE: Insufficient)<br>**Mortality**<br>Inadequate evidence (SOE: Insufficient)<br><br><u>1-, 2-, or 3-level CADR vs. ACDF</u><br><br>**VAS Neck Pain at intermediate**<br>No difference (Low SOE)<br>**Neurologic function**<br>Inadequate evidence (SOE: Insufficient)<br>**General function**<br>Inadequate evidence (SOE: Insufficient)<br>**Harms**<br>Inadequate evidence (SOE: Insufficient) | |

**Abbreviations:** ACDF: anterior cervical discectomy and fusion; ASD: adjacent segment disease; CADR: cervical artificial disc replacement; CT: computed tomography; DDD: degenerative disc disease; KQ: Key Question; IDE: investigational device exemption; k: number of studies; LADR: lumbar artificial disc replacement; NDI: neck disability index; NRSI: nonrandomized studies of interventions; ODI: Oswestry Disability Index; QOL: quality of life; RCT: randomized controlled trial; RR: risk ratio; SF-36: short form health survey- 36 item; SAE: serious adverse event; SF-36: short form 36 item; SMD: standardized mean difference; SOE: strength of evidence; VAS: visual analogue scale;

## Table F-2. Summary of systematic reviews of LADR

| Citation<br>Search Dates | Condition Intervention<br>Comparator | Studies Included in the Review | Outcomes | Results | Review Authors' Conclusions |
|---|---|---|---|---|---|
| Wen, 2024[g]<br>2012 to 2022 | Condition<br>Lumbar DDD<br>Intervention<br>LADR<br>Comparator<br>Other lumbar artificial disc or no comparator (pre-post) | 22 studies (7 RCTs, 15 cohort studies) published between 2012 and 2021 (10 published in or after 2017)<br>Devices represented[g]<br>*AcroFlex*<br>ActivL<br>Charité<br>*Kineflex*<br>*Maverick*<br>ProDisc-L | Pain (VAS)<br>Disability (ODI)<br>Clinical success<br>Complication<br>Reoperation | **VAS, mean (SD)[h]**<br>Preoperative: 75.4 (3.5)<br>Postoperative: 24.7 (8.4)<br>**ODI, mean (SD)[4]**<br>Preoperative: 49.8 (5.1)<br>Postoperative: 19.4 (3.6)<br>**Clinical success rate, % (SD) (k=11)**<br>74.8 (6.3)<br>**Complication rate, % (SD) (k=17)**<br>18.5 (6.3)<br>**Complication rate at 5 years, % (SD)**<br>18.8 (6.8)<br>**Complication rate ≥10 years, % (SD)**<br>27.2 (5.3)<br>**Reoperation rate, % (SD) (k=17)**<br>13.6 (3.8)<br>**Reoperation rate at 5 years, % (SD)**<br>7.8 (2.8)<br>**Reoperation rate ≥10 years, % (SD)**<br>16.7 (9.6) | Pain and disability scores improved at follow-up compared to baseline<br><br>Charité, ProDisc-L and ActivL devices showed significant improvements in pain and disability scores from baseline<br><br>Clinical success and patient satisfaction were better with the Charité device than the ProDisc-L, but the differences were not significant<br><br>Clinical success rates were highest for the Active-L |

[g] Devices in italicized text are investigational/ not FDA approved
[h] Lower score is lower pain

| Citation<br>Search Dates | Condition Intervention<br>Comparator | Studies Included in the Review | Outcomes | Results | Review Authors'<br>Conclusions |
|---|---|---|---|---|---|
| Bai, 2019[8]<br>up to October 2018 | Condition<br>Symptomatic lumbar DDD<br>Intervention<br>LADR<br>Comparator<br>Fusion: anterior, posterior, or circumferential | 10[i] RCTs<br>Devices represented<br>*Charité*<br>*Flexicore*<br>*Maverick*<br>ProDisc-L | Overall success<br>Pain (VAS)<br>Function/disability (ODI)<br>QoL (SF-36)<br>Device failure<br>Complications | **Overall success (k=4)**<br>RR = 1.260 (95% CI: 1.118 to 1.420)<br>**VAS (k=5)**<br>SMD = -0.206 (95% CI: - 0.326 to - 0.085)<br>**ODI score (k=4)**<br>SMD = - 0.276 (95% CI: - 0.4 to - 0.152)<br>**ODI success (k=4)**<br>RR = 1.146 (95% CI: 1.056 to 1.243)<br>**SF-36 (k=4)**<br>SMD = 0.283 (95% CI: 0.157 to 0.409)<br>**Complications (k=6)**<br>RR = 0.437 (95% CI: 0.282 to 0.678)<br>**Reoperation (k=5)**<br>RR = 0.534 (95% CI: 0.288 to 0.992) | Disc replacement is superior to lumbar fusion for overall success, VAS, ODI, SF-36, reoperation rate, and complications |

**Abbreviations:** ACDF: anterior cervical discectomy and fusion; ASD: adjacent segment disease CT: computed tomography; DDD: degenerative disc disease; KQ: Key Question; IDE: investigational device exemption; k: number of studies; LADR: lumbar artificial disc replacement; mJOA: modified Japanese orthopedic association scale; NRSI: nonrandomized studies of interventions; ODI: Oswestry Disability Index; QOL: quality of life; RCT: randomized controlled trial; RR: risk ratio; SF-36: short form health survey- 36 item; SAE: serious adverse event; SF-36: short form 36 item; SMD: standardized mean difference; SOE: strength of evidence; VAS: visual analogue scale;

[i] Authors reported 14 RCTs but it is 14 publications in 10 RCTs

# Appendix G. Signal Search Findings

## Table G-1. Summary of findings for 1-level CADR (KQ1 Effectiveness)

| Outcome | Follow-up, months | 2016 HTA | Selph et al[7] |
|---|---|---|---|
| | | Direction of effect<br>SOE; Number of studies | Direction of effect<br>SOE; Number of studies |
| **Overall success** | 24 | CADR > ACDF<br>Moderate; 5 RCTs | No difference<br>Not assessed; 4 RCTs |
| | 48-60 | CADR > ACDF<br>Moderate; 3 RCTs | No difference<br>Not assessed; 6 RCTs |
| | 84 | CADR > ACDF<br>Low; 1 RCT | CADR > ACDF<br>Not assessed; 3 RCTs |
| **NDI success** | 24 | CADR > ACDF<br>Moderate; 5 RCTs | No difference<br>Moderate; 6 RCTs |
| | 48-60 | No difference<br>Low; 3 RCTs | No difference<br>Moderate; 6 RCTs |
| | 84 | No difference<br>Low; 1 RCT | No difference<br>Moderate; 4 RCTs |
| **NDI score** | 24 | No difference<br>Moderate; 9 RCTs | No difference<br>Moderate; 8 RCTs |
| | 48-60 | CADR > ACDF<br>Moderate; 6 RCTs | No difference<br>Moderate; 12 RCTs |
| | 84 | CADR > ACDF<br>Low; 2 RCTs | No difference<br>Moderate; 6 RCTs |
| **Neurological success** | 24 | C-ADR > ACDF<br>Moderate; 6 RCTs | No difference<br>Moderate; 5 RCTs |
| | 48-60 | CADR > ACDF<br>Moderate; 4 RCTs | No difference<br>Moderate; 6 RCTs |
| | 84 | No difference<br>Low; 2 RCTs | No difference<br>Moderate; 5 RCTs |
| **Arm pain success** | 24 | No difference<br>Low; 2 RCTs | No difference<br>Moderate; 2 RCTs |
| | 60 | No difference<br>Low; 1 RCT | No difference<br>Moderate; 4 RCTs |
| | 84 | Insufficient<br>No trials | No difference<br>Moderate; 1 RCTs |
| **Arm pain VAS score** | 24 | CADR > ACDF<br>Moderate; 7 RCTs | No difference<br>Moderate; 6 RCTs |
| | 48-60 | CADR > ACDF<br>Moderate; 5 RCTs | No difference<br>Moderate; 9 RCTs |
| | 84 | No difference<br>Low; 2 RCTs | No difference<br>Moderate; 5 RCTs |
| **Neck pain success** | 24 | No difference<br>Low; 2 RCTs | No difference<br>Moderate; 2 RCTs |
| | 60 | No difference<br>Low; 1 RCT | No difference<br>Moderate; 4 RCTs |
| | 84 | Insufficient | No difference |

| Outcome | Follow-up, months | 2016 HTA | Selph et al[7] |
|---|---|---|---|
| | | Direction of effect<br>SOE; Number of studies | Direction of effect<br>SOE; Number of studies |
| | | No trials | Moderate; 1 RCT |
| **Neck pain<br>VAS score** | 24 | CADR > ACDF<br>Moderate; 3 RCTs | No difference<br>Moderate; 8 RCTs |
| | 48-60 | CADR > ACDF<br>Moderate; 5 RCTs | No difference<br>Moderate; 11 RCTs |
| | 84 | CADR > ACDF<br>Low; 2 RCTs | No difference<br>Moderate; 5 RCTs |

* Maintenance/ improvement of motor function, sensory function, and deep tendon reflexes

Note: ">" symbol means the intervention is better than the listed comparator (more effective or fewer harms); "<" means the intervention is worse than the listed comparator (less effective or more harms).

**Abbreviations:** CADR: cervical artificial disc replacement; ACDF: anterior cervical discectomy and fusion; RCT: randomized controlled trial; VAS: visual analogue scale: SOE: strength of evidence; Neck disability index

## Table G-2. Summary of findings for 2-level CADR (KQ1 Effectiveness)

| Outcome | Follow-up, months | 2016 HTA | Selph et al[7] |
|---|---|---|---|
| | | Direction of effect<br>SOE; Number of studies | Direction of effect<br>SOE; Number of studies |
| **Overall success** | 24 | CADR > ACDF<br>Moderate; 1 RCT | CADR > ACDF<br>Not reported; 2 RCTs |
| | 48-60 | CADR > ACDF<br>Moderate; 1 RCT | CADR > ACDF<br>Not reported; 2 RCTs |
| | 84 | Insufficient | CADR > ACDF<br>Not reported; 1 RCT |
| **NDI success** | 24 | CADR > ACDF<br>Moderate; 1 RCT | No difference<br>Moderate; 2 RCTs |
| | 48-60 | CADR > ACDF<br>Moderate; 1 RCT | No difference<br>Moderate; 1 RCT |
| | 84 | Insufficient | No difference<br>Moderate; 2 RCTs |
| **NDI score** | 24 | CADR > ACDF<br>Low; 2 RCTs | CADR > ACDF<br>Moderate; 3 RCTs |
| | 48-60 | CADR > ACDF<br>Low; 1 RCT | CADR > ACDF<br>Moderate; 4 RCTs |
| | 84 | Insufficient | CADR > ACDF<br>Moderate; 3 RCTs |
| **Neurological success** | 24 | No difference<br>Low; 1 RCT | No difference<br>Moderate; 2 RCTs |
| | 48-60 | No difference<br>Low; 1 RCT | No difference<br>Moderate; 2 RCTs |
| | 84 | Insufficient | No difference<br>Moderate; 2 RCTs |
| **Arm pain success** | 24 | Insufficient | No difference<br>Moderate; 2 RCTs |
| | 60 | Insufficient | No difference<br>Moderate; 2 RCTs |
| | 84 | Insufficient | No difference<br>Moderate; 1 RCT |
| **Arm pain VAS score** | 24 | CADR > ACDF<br>Low; 2 RCTs | No difference<br>Moderate; 2 RCTs |
| | 48-60 | No difference<br>Low; 1 RCT | CADR > ACDF<br>Moderate; 3 RCTs |
| | 84 | Insufficient | No difference<br>Moderate; 2 RCTs |
| **Neck pain success** | 24 | Insufficient | No difference<br>Moderate; 2 RCTs |
| | 60 | Insufficient | No difference<br>Moderate; 2 RCTs |
| | 84 | Insufficient | No difference<br>Moderate; 1 RCT |

| Outcome | Follow-up, months | 2016 HTA | Selph et al[7] |
|---|---|---|---|
| | | Direction of effect<br>SOE; Number of studies | Direction of effect<br>SOE; Number of studies |
| **Neck pain VAS score** | 24 | CADR > ACDF<br>Low; 2 RCTs | No difference<br>Moderate; 3 RCTs |
| | 48-60 | No difference<br>Low; 1 RCT | CADR > ACDF<br>Moderate; 4 RCTs |
| | 84 | Insufficient | CADR > ACDF<br>Moderate; 3 RCTs |

* Maintenance/ improvement of motor function, sensory function, and deep tendon reflexes.

Note: ">" symbol means the intervention is better than the listed comparator (more effective or fewer harms); "<" means the intervention is worse than the listed comparator (less effective or more harms).

**Abbreviations:** CADR: cervical artificial disc replacement; ACDF: anterior cervical discectomy and fusion; RCT: randomized controlled trial; VAS: visual analogue scale: SOE: strength of evidence; Neck disability index

## Table G-3. Summary of findings for 1-level CADR (KQ2 Harms)

| Outcome | Follow-up, months | 2016 HTA | Selph et al[7] |
|---|---|---|---|
| | | Direction of effect<br>SOE; Number of studies | Direction of effect<br>SOE; Number of studies |
| **Secondary surgery at the index level** | 24 | CADR > ACDF<br>Moderate; 8 RCTs | CADR > ACDF<br>High; 9 RCTs |
| | 48-60 | CADR > ACDF<br>Low; 4 RCTs | CADR > ACDF<br>High; 7 RCTs |
| | >60 | NR | CADR > ACDF<br>High; 7 RCTs |
| | 84 | CADR > ACDF<br>Low; 2 RCTs | NR |
| **Serious/ major adverse events** | 24 | CADR > ACDF<br>Low; 5 RCTs | CADR > ACDF<br>Low; 5 RCTs |
| | 24-48 | No difference<br>Low; 1 RCT | No difference<br>Low; 2 RCTs |
| | 60 | NR | No difference<br>Low; 3 RCTs |
| | 84 | No difference<br>Low; 1 RCT | NR |
| **Device-related adverse events** | 24 | CADR > ACDF<br>Moderate; 6 RCTs | NR |
| | 60 | No difference<br>Moderate; 2 RCTs | NR |
| | 84 | No difference<br>Low; 1 RCT | NR |

Note: ">" symbol means the intervention is better than the listed comparator (more effective or fewer harms); "<" means the intervention is worse than the listed comparator (less effective or more harms).

**Abbreviations:** CADR: cervical artificial disc replacement; ACDF: anterior cervical discectomy and fusion; RCT: randomized controlled trial; VAS: visual analogue scale: SOE: strength of evidence; Neck disability index

## Table G-4. Summary of findings for 2-level CADR (KQ2 Harms)

| Outcome | Follow-up, months | 2016 HTA | Selph[7] |
|---|---|---|---|
| | | Direction of effect<br>SOE; Number of studies | Direction of effect<br>SOE; Number of studies |
| **Secondary surgery at the index level** | 24 | CADR > ACDF<br>Low; 1 RCT | CADR > ACDF<br>Low; 2 RCTs |
| | 60 | CADR > ACDF<br>Low; 1 RCT | CADR > ACDF<br>Low; 1 RCT |
| | >60 | NR | CADR > ACDF<br>Low; 2 RCTs |
| | 84 | No data | NR |
| **Serious/ major adverse events** | 24 | CADR > ACDF<br>Low; 1 RCT | CADR > ACDF<br>Low; 2 RCTs |
| | 60 | No data | NR |
| | 84 | No data | NR |
| | 120 | NR | CADR > ACDF<br>Low; 1 RCT |
| **Device- related adverse events** | 24 | CADR > ACDF<br>Low; 1 RCT | NR |
| | 60 | No data | NR |
| | 84 | No data | NR |

Note: ">" symbol means the intervention is better than the listed comparator (more effective or fewer harms); "<" means the intervention is worse than the listed comparator (less effective or more harms).

**Abbreviations:** CADR: cervical artificial disc replacement; ACDF: anterior cervical discectomy and fusion; RCT: randomized controlled trial; VAS: visual analogue scale: SOE: strength of evidence; Neck disability index

## Table G-5. Summary of findings for 1-, 2-, or 3-level CADR (KQ2 Harms)

| Outcome | Follow-up, months | 2016 HTA | Selph et al[7] |
|---|---|---|---|
| | | Direction of effect<br>SOE; Number of studies | Direction of effect<br>SOE; Number of studies |
| **Secondary surgery at the index level** | 24-36 | No difference<br>Low; 2 RCTs | CADR > ACDF<br>Insufficient; 2 RCTs |
| | 60 | No data | NR |
| | 84 | No data | NR |
| **Serious/ major adverse events** | 24-36 | No difference<br>Low; 2 RCTs | Not reported<br>Insufficient; 2 RCTs |
| | 60 | No data | NR |
| | 84 | No data | NR |
| **Device-related adverse events** | 24 | CADR > ACDF<br>Low; 2 RCTs | NR |
| | 60 | No data | NR |
| | 84 | No data | NR |

Note: ">" symbol means the intervention is better than the listed comparator (more effective or fewer harms); "<" means the intervention is worse than the listed comparator (less effective or more harms).

**Abbreviations:** CADR: cervical artificial disc replacement; ACDF: anterior cervical discectomy and fusion; RCT: randomized controlled trial; VAS: visual analogue scale: SOE: strength of evidence; Neck disability index

**Table G-6. Summary of findings for 1-level or 2-level LADR vs. fusion (KQ1 Effectiveness)**

| Outcome | Level | 2016 HTA | Bai et al[8] |
|---|---|---|---|
| | | Direction of effect<br>SOE; Number of studies | Direction of effect<br>SOE; Number of studies |
| **Overall success** | 1 | No difference<br>Low; 2 RCTs | NR |
| | 2 | No difference<br>Low; 1 RCT | NR |
| | 1 or 2 | No difference<br>Low; 1 RCT | LADR > Fusion<br>Moderate; 4 RCTs |
| **ODI success** | 1 | No difference<br>Low; 2 RCTs | NR |
| | 2 | NR | NR |
| | 1 or 2 | No difference<br>Low; 1 RCT | LADR > Fusion<br>4 RCTs; Moderate |
| **ODI score** | 1 | NR | NR |
| | 2 | LADR > Fusion<br>Low; 1 RCT | NR |
| | 1 or 2 | NR | LADR > Fusion<br>Moderate; 5 RCTs |
| **Neurological success** | 1 | No difference<br>Low; 2 RCTs | NR |
| | 2 | No difference<br>Low; 1 RCT | NR |
| | 1 or 2 | NR | LADR > Fusion<br>Moderate; 3 RCTs |
| **Pain VAS score** | 1 | No difference<br>Low; 2 RCTs | NR |
| | 2 | No difference<br>Low; 1 RCT | NR |
| | 1 or 2 | NR | LADR > Fusion<br>Moderate; 5 RCTs |
| **Back pain VAS score** | 1 | NR | NR |
| | 2 | NR | NR |
| | 1 or 2 | No difference<br>Low; 1 RCT | NR |
| **Leg pain VAS score** | 1 | NR | NR |
| | 2 | NR | NR |
| | 1 or 2 | No difference<br>Low; 1 RCT | NR |
| **SF-36 pain subscale** | 1 | NR | NR |
| | 2 | NR | NR |
| | 1 or 2 | LADR > Fusion<br>Low; 1 RCT | LADR > Fusion<br>Moderate; 5 RCTs |

Note: ">" symbol means the intervention is better than the listed comparator (more effective or fewer harms); "<" means the intervention is worse than the listed comparator (less effective or more harms).

**Abbreviations:** LADR: lumbar artificial disc replacement; RCT: randomized controlled trial; SF-36: Short Form-36 item; SOE: strength of evidence; VAS: visual analogue scale; vs: versus

## Table G-7. Summary of findings for 1-level or 2-level LADR vs. fusion (KQ2 Harms)

| Outcome | Level | 2016 HTA | Wen et al[9] | Bai et al[8] |
|---|---|---|---|---|
| | | Direction of effect<br>SOE; Number of studies | Incidence<br>Number of studies | Direction of effect<br>Number of studies |
| **Any adverse event/ complication** | 1 | No difference<br>Low; 2 RCTs | NR | NR |
| | 1 or 2 | No difference<br>Low; 1 RCT | LADR: 18.5% (6.3%)*<br>17 studies | LADR > Fusion<br>Low; 6 RCTs |
| **Reoperation (secondary surgical procedure at index level or device-related operation)** | 1 or 2 | LADR > Fusion<br>Low; 1 RCT | LADR: 13.6 (3.8%)†<br>17 studies | LADR > Fusion<br>Moderate; 5 RCTs |

* An implant or non-implant related complications. Only data for LDR group reported in SR; comparator data not reported.
† Secondary spinal surgery, related to the implant or any other unspecified indication. Only data for LDR group reported in SR; comparator data not reported.

Note: ">" symbol means the intervention is better than the listed comparator (more effective or fewer harms); "<" means the intervention is worse than the listed comparator (less effective or more harms).

**Abbreviations:** LADR: lumbar artificial disc replacement; RCT: randomized controlled trial; SD: standard deviation; vs: versus

## Table G-8. Summary of findings for 1-level or 2-level LADR vs. multidisciplinary rehabilitation (KQ1 Effectiveness)

| Outcome | 2016 HTA | Signal Search |
|---|---|---|
| | LADR Versus Multidisciplinary Rehabilitation Study at 24 months[11] | LADR Versus Multidisciplinary Rehabilitation Study at 96 months[10] |
| **ODI** | LADR > Multidisciplinary rehabilitation<br>Mean improvement: 8.4 points (95% CI: 3.6 to 13.2) | LADR > Multidisciplinary rehabilitation<br>Mean improvement<br>ITT analysis: 6.1 points (95% CI 1.2 to 11.0)<br>PP analysis: 8.1 points (95% CI 2.3 to 13.9) |
| **VAS Pain score** | LADR > Multidisciplinary rehabilitation | LADR > Multidisciplinary rehabilitation<br>ITT analysis: 9.9 (95% CI 0.6 to 19.2) |
| **Secondary surgery at index level** | LADR: 6.5% (5/77) | LADR: 12/86 had additional surgery<br>Multidisciplinary rehab: 21/87 crossed over to LADR, fusion, or discectomy) |
| **Major complications resulting in impairment** | LADR: 7.8% (6/77) | LADR: Only reported for all surgeries, not specific to LADR; will require author query to clarify/confirm data. |
| **Any complications** | LADR: 33.8% (26/77) | NR |

Note: ">" symbol means the intervention is better than the listed comparator (more effective or fewer harms); "<" means the intervention is worse than the listed comparator (less effective or more harms).

**Abbreviations:** LADR: lumbar artificial disc replacement; RCT: randomized controlled trial; SF-36: Short Form-36 item; SOE: strength of evidence; ODI Oswestry Disability Index; VAS: visual analogue scale; vs: versus

1. Spectrum Research. Artificial Disc Replacement Re-Review Final evidence Report. https://www.hca.wa.gov/assets/program/adr-rr-final-report-20161219.pdf. Published 2016.
2. Health Technology Clinical Committee (HTCC). Health Technology Clinical Committee Final Findings and Decision: Artificial disc replacement – Re-review. https://www.hca.wa.gov/assets/program/adr-rr-final-findings-decision-20170317.pdf. Published 2017.
3. Goossen K, Bieler D, Hess S, et al. An adapted 'Ottawa' method allowed assessing the need to update topic areas within clinical practice guidelines. *J Clin Epidemiol.* 2022;150:1-11. PMID: https://www.ncbi.nlm.nih.gov/pubmed/35710055. doi: 10.1016/j.jclinepi.2022.06.003
4. Shojania KG, Sampson M, Ansari MT, Ji J, Doucette S, Moher D. How quickly do systematic reviews go out of date? A survival analysis. *Ann Intern Med.* 2007;147(4):224-233. PMID: https://www.ncbi.nlm.nih.gov/pubmed/17638714. doi: 10.7326/0003-4819-147-4-200708210-00179
5. Kim LJY, Mazur MD, Dailey AT. Mid-term and Long-term Outcomes After Total Cervical Disk Arthroplasty Compared With Anterior Cervical Discectomy and Fusion: A Systematic Review and Meta-analysis of Randomized Controlled Trials. *Clin Spine Surg.* 2023;36(9):339-355. PMID: https://www.ncbi.nlm.nih.gov/pubmed/37735768. doi: 10.1097/bsd.0000000000001537
6. Ortiz-Babilonia CD, Gupta A, Cartagena-Reyes MA, et al. The Statistical Fragility of Trials Comparing Cervical Disc Arthroplasty and Anterior Cervical Discectomy and Fusion: A Meta Analysis. *Spine (Phila Pa 1976).* 2024;49(10):708-714. PMID: https://www.ncbi.nlm.nih.gov/pubmed/37368958. doi: 10.1097/brs.0000000000004756
7. Selph SS, Skelly AC, Jungbauer RM, et al. Cervical Degenerative Disease Treatment: A Systematic Review (Comparative Effectiveness Review, No. 266.) https://www.ncbi.nlm.nih.gov/books/NBK597699/. Published 2023. Updated Nov. 2023. Accessed 7/10/2024.
8. Bai DY, Liang L, Zhang BB, et al. Total disc replacement versus fusion for lumbar degenerative diseases - a meta-analysis of randomized controlled trials. *Medicine (Baltimore).* 2019;98(29):e16460. PMID: https://www.ncbi.nlm.nih.gov/pubmed/31335704. doi: 10.1097/md.0000000000016460
9. Wen DJ, Tavakoli J, Tipper JL. Lumbar Total Disc Replacements for Degenerative Disc Disease: A Systematic Review of Outcomes With a Minimum of 5 years Follow-Up. *Global Spine J.* 2024:21925682241228756. PMID: https://www.ncbi.nlm.nih.gov/pubmed/38263726. doi: 10.1177/21925682241228756
10. Furunes H, Storheim K, Brox JI, et al. Total disc replacement versus multidisciplinary rehabilitation in patients with chronic low back pain and degenerative discs: 8-year follow-up of a randomized controlled multicenter trial. *Spine J.* 2017;17(10):1480-1488. PMID: https://www.ncbi.nlm.nih.gov/pubmed/28583869. doi: 10.1016/j.spinee.2017.05.011
11. 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.* 2011;342:d2786. PMID: https://www.ncbi.nlm.nih.gov/pubmed/21596740. doi: 10.1136/bmj.d2786

12. Martin BI, Deyo RA, Mirza SK, et al. Expenditures and Health Status Among Adults With Back and Neck Problems. *JAMA.* 2008;299(6):656-664. doi: 10.1001/jama.299.6.656
13. Cho SK, Riew DK. Adjacent segment disease following cervical spine surgery. *JAAOS-Journal of the American Academy of Orthopaedic Surgeons.* 2013;21(1):3-11.
14. Levin DA, Hale JJ, Bendo JA. Adjacent segment degeneration following spinal fusion for degenerative disc disease. *Bulletin of the NYU hospital for joint diseases.* 2007;65(1):29-29.
15. Malveaux WMSC, Sharan AD. Adjacent segment disease after lumbar spinal fusion: a systematic review of the current literature. 2011.
16. Robinson J, Kothari MJ. Treatment and prognosis of cervical radiculopathy. *UpToDate, Waltham, MA.(Accessed on August 21, 2019.).* 2017.
17. Chou R, Atlas S. Subacute and chronic low back pain: Surgical treatment. *UptoDate, Waltham, MA.* 2024.
18. Hillard VH, Apfelbaum RI. Surgical management of cervical myelopathy: indications and techniques for multilevel cervical discectomy. *The Spine Journal.* 2006;6(6):S242-S251.
19. Othman YA, Verma R, Qureshi SA. Artificial disc replacement in spine surgery. *Annals of translational medicine.* 2019;7(Suppl 5).
20. Frelinghuysen P, Huang RC, Girardi FP, Cammisa FP, Jr. Lumbar total disc replacement part I: rationale, biomechanics, and implant types. *Orthop Clin North Am.* 2005;36(3):293-299. PMID: https://www.ncbi.nlm.nih.gov/pubmed/15950689. doi: 10.1016/j.ocl.2005.02.014
21. Abi-Hanna D, Kerferd J, Phan K, Rao P, Mobbs R. Lumbar Disk Arthroplasty for Degenerative Disk Disease: Literature Review. *World Neurosurg.* 2018;109:188-196. PMID: https://www.ncbi.nlm.nih.gov/pubmed/28987839. doi: 10.1016/j.wneu.2017.09.153
22. Acosta FL, Jr., Ames CP. Cervical disc arthroplasty: general introduction. *Neurosurg Clin N Am.* 2005;16(4):603-607, vi. PMID: https://www.ncbi.nlm.nih.gov/pubmed/16326283. doi: 10.1016/j.nec.2005.07.003
23. U.S. Food & Drug Administration. Premarket Approval (PMA). https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfPMA/pma.cfm. Published. Accessed May 8, 2024.
24. U.S. Food and Drug Administration (FDA). Simplify® Cervical Artificial Disc. PMA P200022/S003. . https://www.accessdata.fda.gov/cdrh_docs/pdf20/P200022S003A.pdf. Published. Updated April 1, 2021. Accessed 6/17/2024.
25. U.S. Food and Drug Administration (FDA). M6-C™ Artificial Cervical Disc. PMA P170036. . https://www.accessdata.fda.gov/cdrh_docs/pdf17/P170036A.pdf. Published. Updated February 6, 2019. Accessed 6/17/2024.
26. Centers for Medicare & Medicaid Services. Local Coverage Decision (LCD): Cervical Disc Replacement. https://www.cms.gov/medicare-coverage-database/view/lcd.aspx?lcdid=38033&ver=13&keyword=artificial%20disc%20replacement&keywordType=starts&areaId=all&docType=NCD,F&contractOption=all&sortBy=relevance&bc=1. Published. Accessed May 8, 2024.
27. Aesculap, Inc. Activ-L artificial disc. https://www.aesculapimplantsystems.com/en/healthcare-professionals/spine/activl-artificial-disc.html. Published. Accessed May 8, 2024.

28. Centinel Spine. PRODISC L Anterior Lumbar Total Disc Replacement. https://www.centinelspine.com/prodisc_l.php. Published. Accessed May 8, 2024.
29. Centers for Medicare & Medicaid Services. National Coverage Determination (NCD): Lumber Artificial Disc Replacement (LADR). https://www.cms.gov/medicare-coverage-database/view/ncd.aspx?ncdid=313. Published. Accessed May 8, 2024.
30. Centers for Medicare & Medicaid Services. Local Coverage Determination (LCD): Lumbar Artificial Disc Replacement (L37826). https://www.cms.gov/medicare-coverage-database/view/lcd.aspx?lcdId=37826. Published. Accessed May 8, 2024.
31. Hood C, Zamani R, Akrami M. Impact of heterotopic ossification following lumbar total disk replacement: a systematic review. *BMC Musculoskelet Disord.* 2022;23(1):382. PMID: https://www.ncbi.nlm.nih.gov/pubmed/35461244. doi: 10.1186/s12891-022-05322-9
32. Radcliff K, Davis RJ, Hisey MS, et al. Long-term Evaluation of Cervical Disc Arthroplasty with the Mobi-C© Cervical Disc: A Randomized, Prospective, Multicenter Clinical Trial with Seven-Year Follow-up. *Int J Spine Surg.* 2017;11(4):31. PMID: https://www.ncbi.nlm.nih.gov/pubmed/29372135. doi: 10.14444/4031
33. Vaccaro A, Beutler W, Peppelman W, et al. Long-Term Clinical Experience with Selectively Constrained SECURE-C Cervical Artificial Disc for 1-Level Cervical Disc Disease: Results from Seven-Year Follow-Up of a Prospective, Randomized, Controlled Investigational Device Exemption Clinical Trial. *Int J Spine Surg.* 2018;12(3):377-387. PMID: https://www.ncbi.nlm.nih.gov/pubmed/30276095. doi: 10.14444/5044
34. Chen X, Shi L, Yu X, Pang Q, Yang J. Comparative study of artificial cervical disc replacement and anterior cervical discectomy/fusion in the treatment of cervical spondylotic myelopathy. *Int J Clin Exp Med.* 2019;12(8):10597-10604.
35. Donk RD, Verbeek ALM, Verhagen WIM, Groenewoud H, Hosman AJF, Bartels R. What's the best surgical treatment for patients with cervical radiculopathy due to single-level degenerative disease? A randomized controlled trial. *PLoS One.* 2017;12(8):e0183603. PMID: https://www.ncbi.nlm.nih.gov/pubmed/28850600. doi: 10.1371/journal.pone.0183603
36. Gornet MF, Lanman TH, Burkus JK, et al. Two-level cervical disc arthroplasty versus anterior cervical discectomy and fusion: 10-year outcomes of a prospective, randomized investigational device exemption clinical trial. *J Neurosurg Spine.* 2019;31(4):508-518. PMID: https://www.ncbi.nlm.nih.gov/pubmed/31226684. doi: 10.3171/2019.4.Spine19157
37. Lavelle WF, Riew KD, Levi AD, Florman JE. Ten-year Outcomes of Cervical Disc Replacement With the BRYAN Cervical Disc: Results From a Prospective, Randomized, Controlled Clinical Trial. *Spine (Phila Pa 1976).* 2019;44(9):601-608. PMID: https://www.ncbi.nlm.nih.gov/pubmed/30325888. doi: 10.1097/BRS.0000000000002907
38. Loidolt T, Kurra S, Riew KD, Levi AD, Florman J, Lavelle WF. Comparison of adverse events between cervical disc arthroplasty and anterior cervical discectomy and fusion: a 10-year follow-up. *Spine J.* 2021;21(2):253-264. PMID: https://www.ncbi.nlm.nih.gov/pubmed/33080376. doi: 10.1016/j.spinee.2020.10.013
39. Radcliff K, Spivak J, Darden B, 2nd, Janssen M, Bernard T, Zigler J. Five-Year Reoperation Rates of 2-Level Lumbar Total Disk Replacement Versus Fusion: Results of a Prospective, Randomized Clinical Trial. *Clin Spine Surg.* 2018;31(1):37-42. PMID: https://www.ncbi.nlm.nih.gov/pubmed/28005616. doi: 10.1097/BSD.0000000000000476