Cervical Disc Arthroplasty

Abstract

This technology overview addressed four questions that compared the difference in outcomes between patients undergoing cervical disc arthroplasty with patients undergoing anterior cervical diskectomy fusion. Most studies did not either report or conduct the appropriate statistical analyses to examine predictive characteristics in patients with successful clinical outcomes. Most studies were inconclusive or unreliable regarding clinical outcomes and revision and/or complication rates in patients who present with neck and/or arm pain. No significant difference in the length of hospital stay was reported; however, two studies included in the overview reported that patients treated with cervical disc arthroplasty returned to work in significantly fewer days (range, 14 to 16 days) than did patients treated with anterior cervical diskectomy fusion.

The prospect of relieving radicular arm and neck pain while maintaining spine segmental motion, thereby eliminating adjacent-segment degeneration, may be possible with the advent of the artificial disc. Increasing numbers of artificial discs are becoming available for use, and the body of data with longer follow-up is growing. However, does the available evidence tell us whether artificial disc replacement is as good as, or superior to, anterior cervical fusion in relieving neck and arm pain when these discs are used to address clinical scenarios similar to those managed by anterior cervical fusion?

Evidence-based medicine uses a three-legged stool approach to reach appropriate clinical decisions: (1) the best available evidence is incorporated with (2) the physician’s experience and (3) patient values to select the best available treatment recommendation for an individual patient. Often, while incorporating new technology into clinical practice, physician experience is limited; also, exuberant marketing and unrealistic expectations regarding a new technique can unduly influence patient values. Given the potential inherent weaknesses of these two legs of the evidence-based medicine triad, the best available evidence becomes that much more important in clinical decision making early during incorporation of new technology into a clinician’s practice.

This technology overview examines the best available evidence on cervical artificial disc replacement compared with the standard of care of anterior cervical fusion and plating.

Methods Overview

This report was developed using the methods of a systematic review.1 We began by having a panel of physicians frame four key questions; we then developed rules (inclusion criteria) for determining which information to include. (The full list of criteria appears in Appendix I in the complete technology overview,
which includes all appendices, tables, and figures, available at http://www.aaos.org/research/overviews/cervicaldisc.pdf). Articles were included only when they met the a priori criteria. Finally, we conducted comprehensive literature searches (Appendix II) to ensure that the data we considered are not biased in favor of any particular point of view. Thereafter, we evaluated the quality of the relevant studies (including their methods of analysis), considered and compared their results, and summarized this information. The program TechDig 2.0 (Ronald B. Jones, Mundelein, IL) was used to estimate means and variances from studies presenting data only in graphic form. Also, we calculated the variance of the arcsine difference to confirm statistical significance \( P < 0.05 \) and converted one-sided probability values to two-sided values to consistently compare the results of all of the studies included in this overview.

### Included Articles

Our search identified 2,054 citations that were potentially relevant to this overview and that could potentially meet our inclusion criteria. Of these, seven studies\(^3\)\(^-\)\(^9\) were included to address the key questions (Table 1 in the complete overview online). Six of these studies compared the outcomes of patients treated with single-level cervical disc arthroplasty with those of patients treated with single-level fusion with adjunct augmentation. One study compared the outcomes of patients treated with cervical disc arthroplasty (CDA) at multiple levels with those of patients treated with anterior cervical disc fusion (ACDF) at multiple levels.\(^6\)

Most of the studies in this overview included patients with a herniated cervical disc and/or cervical degenerative disc disease \((n = 566)\).\(^3\)\(^-\)\(^8\) These studies did not include patients with moderate or severe or “marked” spondylosis. Two studies included patients with spondylosis and neck or arm pain (radicular) and/or functional/neurologic deficits \((n = 297)\) but excluded patients with severe spondylosis or ankylosing spondylitis (chronic spondylosis).\(^3\)\(^,\)\(^9\) One study does not report whether patients with spondylosis were included.\(^5\)

### Device Properties

The studies included in this overview reported that patients were treated with a metal-on-polymer or metal-on-metal artificial cervical disc; patients treated with diskectomy and fusion received an anterior cervical plate with varying adjunct augmentation, with or without a cage (Table 2). Information regarding the size of the artificial discs used to treat patients in these studies was not reported in detail. Any information the authors reported regarding disc size was given in general terms. Specifically, the authors reported the sizes available for a specific type of the artificial disc but did not disclose the number of patients who received any given size. Nor was disc size compiled for specific groups of patients identified; therefore, disc size information was not useful for this overview.

### Device Recall Information

The US FDA has issued recalls associated with two of the devices used to treat patients with CDA or ACDF reported by four of the studies included in this overview. Please see Tables 3 and 4 for further information regarding the recalls for these devices.
We evaluated quality on a per outcome basis rather than a per study basis because quality is not necessarily the same for all outcomes and all follow-up times reported in a study. For example, a study might report results immediately after patients received a given treatment and after some period of time has passed. Often, nearly all enrolled patients contribute data at early follow-up times; however, at much later follow-up times, only a few patients may contribute data. In this scenario, one would have more confidence in the earlier data than in the later data. Our assigning a higher quality score to the earlier results reflects this difference in confidence.

We assessed the quality of treatment studies using a two-step process. First, we assigned a level of evidence to all results reported in a study based solely on the design of that study. Accordingly, all data presented in randomized controlled trials were initially categorized as level I evidence; all results presented in nonrandomized controlled trials and other prospective comparative studies were initially categorized as level II. We next assessed each outcome at each reported time point using a quality questionnaire and, when quality standards were not met, downgraded the level of evidence (for this outcome at this time point) by one level.

### Outcomes Considered

We preferentially included patient-oriented outcomes over surrogate outcomes. This was partly because patient-oriented outcomes are important to patients and indicate, without the need for extrapolation, whether an intervention is effective. Patient-oriented outcomes include pain, quality of life, ability to perform activities of daily living, and revision surgery. However, surrogate outcomes substitute for a clinical event of true importance. Common surrogate outcomes include laboratory tests, biomarkers, range of motion, and radiographic findings. Unlike the use of patient-oriented outcomes, use of surrogate outcomes can be misleading and can even make harmful treatments look beneficial. We found patient-oriented evidence for every question.

### Minimal Clinically Important Improvement

Wherever possible, we considered the effects of treatments in terms of the minimal clinically important improvement (MCII), in addition to whether the effects were statistically significant. The MCII is the smallest clinical change that is important to patients, in recognition of the fact that some treatment-induced, statistically significant improvements are too small to matter to patients. The values we used for MCIs are derived from published studies investigating the visual analog scale (VAS) and the Neck Disability Index (NDI). The associated descriptive terms in this technology overview, and the conditions for using each of these terms, are outlined in Table 6.

### Power

To assess the power of an outcome to detect a statistically significant difference, we determined whether the number of patients in a study was sufficient to detect a small, medium, or large effect, while assuming an $\alpha$ of 0.05 as the significance level, 80% power, and Cohen’s definitions of small, medium, and large effects (small, $d = 0.2$; medium, $d = 0.5$; and large, $d = 0.8$). When a study with a nonsignificant difference was unable to detect a medium or large effect, it was categorized as low power. Studies able to detect medium effects or studies with statistically significant differences were categorized as high power. Six of the seven studies included in this overview were categorized as having high power, and one study was categorized a low powered study.

### Question 1

What patient characteristics predict successful outcomes in patients who undergo CDA compared with those in patients who undergo anterior cervical diskectomy and fusion?

The outcomes of interest for this question included the following: previous surgeries per patient, all demographics available, age, sex, smoking status, workers’ compensation status, narcotic use, opioid use, analgesic use, use of transcutaneous electrical nerve stimulation (TENS) unit, and any ongoing pain management when evaluated in a study. Most of the studies considered for this question did not report or conduct the appropriate statistical analyses (eg, regression, multiple regression) to examine predictive patient characteristics with patients considered to have successful clinical patient-oriented outcomes.

At 24 months, the authors of one level II study with 147 patients reported no statistically significant difference in the percentage of successful patients treated with CDA compared with successful patients treated with ACDF in regard to the continuation of the use of strong narcotics and muscle relaxants (Table 1 [Table 7 in the online technology overview]). These results are inconclusive about which patient charac-
teristics predict successful outcomes in patients treated with CDA compared with patients treated with ACDF.

The results of the included study that addressed this question (Appendix V) reported unreliable level II evidence for the outcomes\(^3\) (Appendix III). The results are unreliable because they were reported as a composite measure. Composite outcome measures, such as “overall success” as reported in this included study,\(^3\) are unreliable because each individual outcome might not equally influence or contribute to the overall significance of the estimated effects of the given treatment; hence, less important outcomes can be more influential than more serious outcomes (ie, death and/or serious adverse events). Studies suggest the results of the individual outcome measures, along with the results of the composite outcome measures, should be examined to ensure a comprehensive examination of the effects of a given treatment.\(^18\)\(^-\)\(^20\)

Other studies considered for this question did not report or conduct the appropriate statistical analyses or did not compare patients treated with CDA to patients treated with ACDF; therefore, these studies were not included (Appendix V, Table 1).

**Question 2**

Do patients with herniated cervical disc who present with arm pain with or without neck pain, and who are treated with a CDA, have equal or better clinical outcomes than patients treated with ACDF?

To address this question, we included five studies that examined five outcomes\(^3\)\(^-\)\(^7\) (Tables 26 through 30). All of the data were level II except for three of the outcomes reported by Mummaneni et al.\(^5\) The results reported by Mummaneni et al\(^5\) at 24 months were level III data because patient follow-up at this duration was <80% (see Quality of the Literature, above). Based on this flaw, the outcomes data reported at 24 months were not included to address this question; the NDI, neurologic success, and VAS pain results at earlier follow-up durations were included to address this question.

**Neck Disability Index Scores**

Four studies reported NDI scores of patients treated with CDA compared with those of patients treated with ACDF at various follow-up durations\(^5\)\(^-\)\(^6\) (Table 8, Figure 1). One study reported NDI results of patients treated with CDA at multiple levels compared with those of patients treated with ACDF at multiple levels.\(^6\) Patients with lower NDI scores are considered to have less disability when performing activities of daily living compared with patients with higher NDI scores. Three of the four studies reported that, at earlier follow-up durations (1.5 to 3 months), patients treated with CDA had statistically significantly lower NDI scores than did patients treated with ACDF, but the differences are not considered to be clinically important. Results at longer follow-up durations are inconclusive in that one of the four studies reported statistically significant differences in favor of patients treated with CDA at 6 months, and one study reported no statistically significant results; two of the four studies reported statistically significant results in favor of CDA at 12 months and at 24 months; and two of the three studies reported statistically significant results in favor of patients treated with CDA (Table 8).

**Table 1**

<table>
<thead>
<tr>
<th>Study</th>
<th>LOE</th>
<th>Treatment Group</th>
<th>N(^a)(^b)</th>
<th>Duration</th>
<th>Medication Use(^c)(^d)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Murrey et al(^3)</td>
<td>III</td>
<td>CDA</td>
<td>75</td>
<td>24 mo</td>
<td>10%</td>
<td>P = 0.1</td>
</tr>
<tr>
<td>—</td>
<td>III</td>
<td>ACDF</td>
<td>72</td>
<td>24 mo</td>
<td>20.8%</td>
<td>P = 0.1</td>
</tr>
</tbody>
</table>

ACDF = anterior cervical disc fusion, CDA = cervical disc arthroplasty, LOE = level of evidence

\(^a\) Number of patients considered an overall success

\(^b\) Overall success defined by the authors as the percentage of patients with Neck Disability Index success (≥15-point improvement/reduction from baseline), maintenance or improvement in neurologic status (measured by motor function, sensory function, and tendon function; all three conditions had to be satisfied to be considered a success), no serious implant-related adverse event or adverse event related to the implant procedure, or secondary surgical procedure.

\(^c\) Medication use includes the use of strong narcotics and muscle relaxants.

\(^d\) Strong narcotics defined as Schedule II drugs with high abuse and high dependency risk

\(^e\) The authors reported the results of one-tailed tests. We converted the values reported by the authors to two-tailed values.

\(^f\) American Academy of Orthopaedic Surgeons study authors test of arcsine difference

634 Journal of the American Academy of Orthopaedic Surgeons
Neck Disability Index Success Rate

Three studies reported NDI success rates as the percentage of patients with 15-point improvement/reduction from baseline.3,5 Two studies reported that, at 3 months, patients treated with CDA had statistically significantly greater NDI success rates than did patients treated with ACDF3,5 (Table 9). Three studies reported no statistically significant differences in NDI success rates of patients at later follow-up durations (6 to 24 months). One study reported that, at 24 months, the NDI success rates of patients treated with CDA were statistically significantly noninferior (margin of inferiority, δ = 0.10) compared with the success rate of patients treated with ACDF (Table 9).4

Neurologic Success Rate

Three studies reported unreliable (see Question 1, above, regarding composite outcomes) and inconclusive results of neurologic success rates, defined as the maintenance or improvement in neurologic status from baseline as measured by motor function, sensory function, and tendon function.3,5 All three conditions had to be satisfied in order for a patient’s outcome to be considered a success. Details were not reported regarding how motor, sensory, and tendon function were measured.

All three studies reported no statistically significant differences in neurologic success rates at earlier follow-up durations (Table 10). One study reported that, at 12 months, patients treated with CDA had statistically significantly greater neurologic success rates than did patients treated with ACDF (Table 10). One study reported that the neurologic success rates of patients treated with CDA were statistically significantly noninferior to those of patients treated with ACDF at 24 months.

Neck Pain (VAS)

The results reported by five level II studies are inconclusive.3,7 One study reported neck pain results that are incomparable to the results reported by the other studies included to address this question.3 Four of the five studies reported no statistically significant differences in neck pain at earlier follow-up durations (1 to 6 months); one study reported that patients treated with CDA had statistically significantly less neck pain than did patients treated with ACDF (Tables 11 and 12, Figure 2). At later follow-up durations (12 to 36 months), two of the five studies reported that patients treated with CDA had statistically significantly less neck pain than did patients treated with ACDF; however, in one of these two studies, patients were treated with either CDA or ACDF at multiple levels of the cervical spine. One study reported that, at 3 months, patients treated with CDA had statistically significantly less neck pain intensity than did patients treated with ACDF; no statistically significant differences in neck pain scores were reported by these authors at all other follow-up durations3 (Table 12).

Arm Pain (VAS)

Five studies reported the results of arm pain scores of patients up to 36 months following treatment3,7 (Tables 13 and 14, Figure 3). Four studies reported no statistically significant differences in the arm pain scores of patients at 36 months. One study reported that patients treated with CDA at multiple levels had statistically significantly less arm pain than did patients treated with ACDF at multiple levels.6

Medical Outcomes Study 36-Item Short Form

Three studies reported the Medical Outcomes Study 36-Item Short Form physical component summary (PCS) scores,4,6 and two studies reported the mental component summary (MCS) scores, up to 24 months following treatment. One of the three studies reported that patients treated with CDA had statistically significantly greater improvements in PCS and MCS scores up to 12 months following treatment, but the difference was not statistically significant at 24 months4 (Tables 12 and 15). One of three studies reported no statistically significant differences in PCS and MCS scores at all follow-up durations.5 Two studies categorized as having high power reported conflicting results at 6 months;4,5 one study reported statistically significant results, whereas the second study reported no statistically significant results. One study reported that, at 12 and 24 months, patients treated with CDA at multiple levels had statistically significantly higher PCS scores than did patients treated with ACDF at multiple levels6 (Tables 15 and 16, Figures 4 and 5).

Return to Work

Two studies reported no statistically significant difference in the percentage of patients who returned to work at 24 months.3,4 One study reported similar results for patients returning to heavy work3 (Table 17). (See Question 4, below, for the results of patients’ length of time to return to work.)

Question 3

Do patients with herniated cervical disc who present with arm pain, with or without neck pain, and who are treated with a CDA, have equal or better revision rates and/or complication rates than those treated with ACDF?

To address this question, we included four level II studies that...
reported secondary surgical procedures, adverse events, and complications of patients treated with CDA at a single level or ACDF at a single level,\(^3,6,8\) and one study that reported the complications of patients treated with either CDA or ACDF at multiple levels\(^8\) (Tables 31 and 32). One study reported unreliable results due to the use of a composite measure (see Question 1, above, regarding composite outcomes).\(^3\)

**Secondary Surgical Procedures**

Three level II studies reported inconclusive and incomparable results of secondary surgical procedures of patients treated with CDA or ACDF at 24 months.\(^3,6,8\) Secondary surgical procedures included revisions, supplemental fixation, implant removal, and reoperations (Table 18).

One of the three studies reported unreliable results of the device success rate of patients treated with CDA compared with those of patients treated with ACDF\(^3\) (Table 18) (see Question 1, above, regarding composite measures). One of the three studies reported no statistically significant differences in the overall reoperation rates of patients treated with CDA compared with those of patients treated with ACDF\(^3\) (Table 18). This study also reported that statistically significantly fewer patients treated with CDA required reoperations at any level of the cervical spine. The authors reported that the difference was statistically significant, but AAOS calculations cannot confirm this (Table 18). One of the three studies reported that no secondary surgical procedures occurred in patients treated with CDA at multiple levels compared with patients treated with ACDF at multiple levels.\(^6\)

**Adverse Events**

The results reported by four level II studies regarding the number of adverse events of patients treated with CDA compared with the adverse events of patients treated with ACDF are inconclusive.\(^3,5,6,8\) One study excluded complications or any adverse events “not meaningful” to the treatment and that had no effect on the results of patients\(^8\) (eg, postoperative face-lift surgery, being hit with a golf ball). Two studies reported the severity of adverse events based on the World Health Organization (WHO) severity scale. See Table 19 for information and description of each grade.

One of the four studies reported that, at 24 months, there was no statistically significant difference in the number of adverse events that occurred in patients treated with CDA compared with patients treated with ACDF\(^3\) (Table 20). One of the four studies reported that, at 36 months, patients treated with CDA had statistically significantly fewer serious adverse events than did patients treated with ACDF\(^5\) but AAOS calculations cannot confirm this (Table 20).

One study reported that, within the perioperative period, patients treated with CDA had statistically significantly more surgically related adverse events or acute neurologic adverse events (grades 1 through 4) than did patients treated with ACDF.\(^3\) The authors reported the difference as statistically significant, but AAOS calculations cannot confirm this (Tables 21 and 22). One of the four studies reported no statistically significant differences in the number of patients with adverse events within the perioperative period\(^1\) (Table 21). One of the four studies reported that one patient treated with CDA at multiple levels had a deep vein thrombosis and that one patient treated with ACDF at multiple levels had dysphagia.\(^6\)

### Question 4

For patients, is CDA or ACDF more economical, as defined by hospital length of stay and length of time to return to work?

Four level II studies were included to address this question\(^1,5,9\) (Tables 30 and 31). Three level II studies reported the average length of hospital stay,\(^5,9\) and two studies reported the length of time to return to work for patients treated with CDA compared with patients treated with ACDF.\(^4,5\) The studies included to address both questions did not report patient-level data that would allow for appropriate correlation analyses; therefore, no association between adverse events, length of hospital stay, and time to return to work can be provided.

Three studies reported no significant difference in the length of hospital stay for patients treated with CDA compared with patients treated with ACDF\(^3,5,9\) (Table 23). Two studies reported that patients treated with CDA returned to work in significantly fewer days (range, 14 to 16 days) than did patients treated with ACDF.\(^4,5\) (Table 24).

### References

3. Murrey D, Janssen M, Delamarter R, et al: Results of the prospective, randomized, controlled multicenter Food and Drug Administration investigational device exemption study of the ProDisc-C total disc replacement versus anterior discectomy and fusion for the treatment...


