Cervical and Lumbar Spinal Arthroplasty: Clinical Review


*AJNR Am J Neuroradiol* 2012, 33 (9) 1631-1641
doi: https://doi.org/10.3174/ajnr.A2758
http://www.ajnr.org/content/33/9/1631
Cervical and Lumbar Spinal Arthroplasty: Clinical Review

SUMMARY: In contrast to cervical and lumbar fusion procedures, the principal aim of disk arthroplasty is to recapitulate the normal kinematics and biomechanics of the spinal segment affected. Following decompression of the neural elements, disk arthroplasty allows restoration of disk height and maintenance of spinal alignment. Based on clinical observations and biomechanical testing, the anticipated advantage of arthroplasty over standard arthrodesis techniques has been a proposed reduction in the development of symptomatic ALD. In this review of cervical and lumbar disk arthroplasty, we highlight the clinical results and experience with standard fusion techniques, incidence of ALD in the population of patients with surgical fusion, and indications for arthroplasty, as well as the biomechanical and clinical outcomes following arthroplasty. In addition, we introduce the devices currently available and provide a critical appraisal of the clinical evidence regarding arthroplasty procedures.

ABBRVATIONS: ACDF = anterior cervical disectomy and fusion; ALD = adjacent-level degenerative disease; AOR = axis of rotation; BAK = Bagby and Kuslich; DDD = degenerative disk disease; HO = heterotopic ossification; NDI = neck disability index; ODI = Ovesey Disability Index; RCT = randomized controlled trial; ROM = range of motion; SF-36 = Short Form 36; UHMWPE = ultrahigh molecular weight polyethylene; VAS = Visual Analog Scale

Anterior cervical fusion techniques were first introduced and popularized by Cloward,1 Bailey and Badgley,2 Robinson,3 Smith and Robinson,4 and others in the late 1950s and early 1960s. Today, anterior cervical disectomy with interbody fusion and plating is the predominant technique used in the treatment of symptomatic disk herniations, spondylosis, segmental instability, selected traumas, malalignment, and, more controversially, axial neck pain. The nuances of patient selection, surgical technique, and choices of interbody device and instrumentation are beyond the scope of the current review. When one broadly considers anterior cervical arthrodesis procedures, rates of symptomatic improvement and radiographic fusion are extremely favorable for single- and multilevel constructs for the range of indications mentioned above.5,6

Although clinical experience with the first cervical arthroplasty prototype was initially reported in 1966 by Fernström7 and expanded by Reitz and Joubert,8 this alternative to fusion fell into early disfavor due to hardware-related complications and postoperative adjacent-level hypermobility.8,10 With time, however, motion-preservation strategies were reconsidered largely due to clinical and radiographic observations of progressive degenerative disease at levels immediately adjacent to surgically fused segments (ALD). The distinction between purely radiographic and symptomatic (myelopathy, radiculopathy, or myeloradiculopathy) ALD is critical to our understanding of this process. Unfortunately, few studies have been adequately designed or powered to allow definitive conclusions about the pathophysiology, natural history, or incidence of ALD.11 In 1996, Wu et al12 reported a statistically higher rate of radiographic degenerative changes, including anterior osteophytes and disk herniations, in 68 patients treated with 1- or 2-level anterior cervical fusions compared with asymptomatic matched controls at 3-year follow-up. Although such results seem to implicate surgical arthrodesis in the development of ALD, interpretation must allow the limitations of such a matched case-control study.

Radiographic evidence alone of ALD has been reported to occur in as many as 92% of patients at 5-year follow-up.13 Perhaps the most compelling data are derived from larger retrospective series. In 1980, Lunsford et al14 reported a series of 253 patients treated for disk herniations with at least 1-year follow-up. A 6.7% reoperation rate for newly symptomatic levels was reported, though adjacent-level pathology per se was not specifically tabulated. Hilibrand et al15 calculated a 2.9% annual risk of symptomatic ALD from their review of 374 surgically treated patients. Development of ALD, however, appeared complex and likely related to preoperative imaging findings at adjacent levels as well as to the location of the level affected.15 Most interesting, Goffin et al13 were able to demonstrate a similar rate of ALD in younger patients with trauma compared with older patients with degeneration following anterior cervical arthrodesis. The primary difficulty among all such studies has been controlling for the natural rate of degenerative changes in the aging spine, particularly among a subgroup of symptomatic patients who have already demonstrated surgically significant degeneration.

Cadaveric testing has provided similarly compelling, though complex and controversial, evidence for the biomechanical factors involved in the development of ALD following arthrodesis. Illustrating the complexity of the issue, studies have reported increased intradiskal pressures and segmental motion that are dependent on both the location of the adjacent level (superior versus inferior to the fused segment) and the type of loading used (flexion versus extension).16 Studies additionally reveal that the degree of lordosis achieved during fusion significantly alters adjacent-level range of motion,17 and that in some cases, segmental mobility may be maximally
increased at levels distant from the index level. Cadaveric studies have also supported the finding that arthrodesis generally results in increased adjacent-level intradiskal pressures and statistically significant changes in adjacent-level motion, compared with arthroplasty under the same testing paradigms. In their review, Bartolomei et al reported that critics of cadaveric studies have noted the technique’s limitations, including the following: the inability to simulate true bony fusion, the ability to investigate only immediate rather than long-term functional changes, the failure to account for the in vivo effects of stabilizing paraspinal musculature, and the inability to assess the clinical relevance of such induced changes.

Total disk arthroplasty cannot uniformly supplant arthrodesis but is an alternative for a subset of patients who would otherwise require anterior fusion. Candidates are adult patients who present after a period of failed conservative therapy with radiculopathy or myelopathy due to subaxial compressive pathology. Exclusion criteria derived from clinical trials are in some cases device-specific but generally include the following: ≥3 surgical levels, spondylolisthesis, cervical instability as documented with dynamic imaging, severe loss of disk height, adjacent prior fusion, pregnancy, axial neck pain alone without neurologic symptoms, local infection, local tumor or metastasis, cervical trauma, osteoporosis, chronic steroid use, and notable systemic disease (eg, rheumatoid arthritis, insulin-dependent diabetes mellitus, malignancy, and so forth). In relative contrast to the lumbar spine, the goal of cervical arthroplasty is to preserve rather than to restore normal spinal kinematics. As a result, patients with existing ankylosing disease or fused segments are excluded. Reconstitution of the full mobile spinal segment similarly requires functional facet articulations, and patients with significant degeneration or fusion elsewhere within the trijoint complex are, therefore, also excluded (Fig 1).

In 2008, Auerbach et al retrospectively reviewed 167 consecutive patients undergoing cervical spine procedures and reported that 43% of patients would be candidates for total disk replacement based on published recommendations. That number would have increased to 47% if adjacent-level disease

Fig 1. Ideal candidate for cervical arthroplasty. A, Lateral radiograph depicting minimal facet arthropathy and degenerative disease. B and C, Extension (B) and flexion (C) radiographs depicting normal segmental motion at the index level and throughout the cervical spine. D, Sagittal T2-weighted MR image depicting single-level degenerative disk disease, endplate changes, relative preservation of disk height, and posterior disk bulge without marked osteophyte formation.
after prior fusion was removed as a contraindication. These figures, however, represent significant overestimates based on our clinical experience. There are currently a few published studies with small sample sizes that have reported cervical disk replacement after prior adjacent fusion, and preliminary results have compared favorably with those of fusion.27–30

Total disk arthroplasty devices may be classified according to modular versus nonmodular design, device endplate treatment, fixation properties, articular surface composition, articular design (uniarticular, biarticular, nonarticular), as well as kinematics.31–33 Currently available implants and selected devices under trial in the United States are presented in the Table. Methods of endplate fixation by order of increasing pull-out strength include serrated teeth, >1-mm toothed ridges or keels, and vertebral body screws. All modalities were tested as many as 59% and 76% of patients at 1-35 and 3-year follow-up, respectively.36 Although some deterioration in the range of motion with time is not atypical or necessarily unexpected, HO-related rates of clinically relevant hypomobility or bony union across the device have been incompletely evaluated. Early exuberant bone incorporation is typically addressed by a short course of postoperative nonsteroidal anti-inflammatory medications administered to all patients.

Articular surfaces include varying combinations of metal (eg, titanium, stainless steel, cobalt, and chromium), polymer composite, and ceramic. Sekhon et al37 found that devices composed of cobalt-chrome-molybdenum alloys obscured visualization of index and adjacent levels on MR imaging compared with titanium alloy or composite alternatives. This effect, however, was dependent on magnet strength in another study by Antosh et al.38 Although well-described in reference to large-joint orthopedic prostheses39 and less commonly after lumbar arthroplasty,40 wear-related failure, debris accumulation,41 and aseptic loosening have been uncommonly encountered following cervical arthroplasty. Biomaterials and prosthesis design are key to the long-term viability of any disk implant, and the risk of device fatigue may be more negligible in the cervical region due to lower stresses, loads, and mobility compared with the lumbar region.32–34,42 The long-term clinical effects of cervical prosthesis osteointegration and wear, however, have been incompletely evaluated.43

Total artificial disk implants are categorized kinematically as constrained, semiconstrained, and unconstrained on the basis of their ability to restrict or permit segmental motion relative to normal physiologic parameters. Bearing material, position, joint configuration (eg, uniarticular ball and socket, uniarticular ball and trough, biarticular ball and socket, or biarticular saddle configuration), and geometry (eg, spheric, toroidal, or ovate surface) of the articulation create a unique kinematic profile for each device. Constrained implants provide superior stability but require greater precision during surgical placement to adequately match the device’s fixed axis of rotation to the spinal level affected. Unconstrained implants rely on axial ligaments, muscle, and tensioning support to provide stability, but their unfixed axis of rotation may distribute loads away from the bone-device interface. Optimal device kinetics have not been established on the basis of the available short-term clinical and biomechanical studies. Device selection is highly individualized, and successful reproduction of physiologic kinematics must consider multiple factors, including facet loading, disk height, device design, and position of the implant within the interspace (Fig 2).34,44,45

The Prestige-ST (Medtronic, Memphis, Tennessee), ProDisc-C (Synthes Spine, West Chester, Pennsylvania), and Bryan artificial cervical disks (Medtronic) were approved for implantation in the United States between 2007 and 2009 following the publication of FDA Investigational Device Exemp-

### FDA-approved cervical and lumbar arthroplasty devices

<table>
<thead>
<tr>
<th>Device</th>
<th>Application</th>
<th>Design</th>
<th>Biomaterials</th>
<th>Endplate</th>
<th>Fixation</th>
<th>Kinematics</th>
<th>FDA IDE Approval</th>
<th>Manufacturer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prestige ST</td>
<td>Cervical</td>
<td>Uniarticular ball and trough</td>
<td>Metal-on-metal articulation, stainless steel</td>
<td>Roughened surface</td>
<td>Vertebral body screws</td>
<td>Unconstrained</td>
<td>July 2007</td>
<td>Medtronic</td>
</tr>
<tr>
<td>Bryan</td>
<td>Cervical</td>
<td>Biarticular</td>
<td>Titanium alloy shells with polyurethane nucleus, saline lubricant</td>
<td>Applied porous coating</td>
<td>Milled, press-fit</td>
<td>Unconstrained</td>
<td>May 2009</td>
<td>Medtronic</td>
</tr>
<tr>
<td>ProDisc-C</td>
<td>Cervical</td>
<td>Uniarticular ball and socket</td>
<td>CCM endplate with UHMWE inlay, metal-on-polyurethane articulation</td>
<td>Roughened titanium</td>
<td>Central keel</td>
<td>Semi-constrained</td>
<td>December 2007</td>
<td>Synthes Spine</td>
</tr>
<tr>
<td>CHARITE</td>
<td>Lumbar</td>
<td>Biarticular ball and socket</td>
<td>CC endplates with UHMWE sliding core</td>
<td>Titanium and calcium phosphate plasma spray</td>
<td>6 Fixation teeth at cranial/ caudal endplates</td>
<td>Unconstrained</td>
<td>October 2004</td>
<td>Depuy Spine</td>
</tr>
<tr>
<td>ProDisc-L</td>
<td>Lumbar</td>
<td>Uniarticular ball and socket</td>
<td>CCM endplates with UHMWE insert</td>
<td>Titanium plasma spray</td>
<td>Large central keel, 2 lateral spikes</td>
<td>Semi-constrained</td>
<td>August 2008</td>
<td>Synthes Spine</td>
</tr>
</tbody>
</table>

Note:—CCM indicates cobalt-chrome-molybdenum, CC, cobalt-chrome alloy; IDE, Investigational Device Exemption.
tion trials (Table).46-48 Adult patients were enrolled on the basis of single-level radicular symptoms referable to degenerative disk disease and were randomized to receive control ACDF versus investigational arthroplasty device implantation. These studies were designed and powered as prospective multicenter nonblinded 2-year noninferiority efficacy and safety trials. In 2007, Mummaneni et al46 reported on 541 patients treated at 32 centers. They found statistically significant higher rates of neurologic success ($P < .005$) and lower rates of adjacent-level surgeries (1.1% versus 3.4%, $P = .0492$) with Prestige ST (Medtronic) placement compared with ACDF at 24-month follow-up. Re-examination of the data by Botelho et al,49 however, by using number-needed-to-treat analysis, failed to show statistical significance for adjacent-segment disease necessitating surgery (Fig 3). In addition, at 5-year follow-up on the same cohort of patients, rates for surgery at adjacent levels were lower for the investigational group but not statistically significant (4.9% versus 2.9%, $P = .376$).50 The implant was noted to maintain motion at the index level, averaging more than 6.5° at last follow-up.

Murrey et al48 reported statistically significant results favoring arthroplasty with the ProDisc-C device in terms of narcotic use and the need for secondary surgery at last follow-up. In this study, only 1 patient with fusion underwent secondary surgery for ALD. Most interesting, only 84.4% of patients in the investigational arm met criteria for motion preservation at the index level at 2-year follow-up. In a Bryan cervical disk trial reported by Heller et al,47 statistically significant differences favoring arthroplasty were noted for NDI scores, time to return to work, and “overall success” (defined as ≥15-point increase in the NDI score, stable or improved neurologic status, no need for a repeat operation, and absence of a surgical or implant-related adverse event). An analysis of secondary outcome measures also confirmed noninferiority and, in some instances, indicated a nonsignificant trend favoring arthroplasty. Patients with myelopathy were included in this trial.
and despite unanticipated crossovers from the investigational to control arms, the superiority of arthroplasty in terms of “overall success” remained robust in intent-to-treat analysis. In terms of artificial disk long-term mobility, the authors noted that 7%–8% of investigational patients displayed <2° of motion at each follow-up time point, but no single patient remained consistently below that threshold at all time points to indicate reliable hypomobility. Last, radiographic or symptomatic adjacent-level disease was not reported.

Despite the findings of many of these published studies, definitive conclusions regarding the utility of arthroplasty are limited due to the following factors affecting each trial: non-inferiority study design, short-term follow-up, operator bias due to the impossibility of blinding, patient and assessor bias due to unblinded design, failure to specifically address ALD as a primary outcome measure, and the use of survey data. In fact, in 2010 a meta-analysis of randomized controlled trials concluded that cervical disk prostheses were unjustified compared with fusion.51

The devices introduced above, new investigational implants, and additional secondary end points have been studied through subset analyses, subsequent follow-up studies, or within smaller trials.52-57 Clinical studies specifically addressing ALD as a primary outcome, however, remain sparse. Anderson et al58 reported that World Health Organization grade 3 or 4 events as well as repeat operations were more common for the arthrodesis group in a review of the FDA Investigational Device Exemption Bryan disk evaluation data. However, the difference in the number of adjacent-level repeat operations between the 2 groups did not reach significance. Jawahar et al59 examined the incidence of ALD among 93 patients randomized to 1- or 2-level ACDF versus arthroplasty, pooling their cohort from 3 different device FDA Investigational Device Exemption clinical trials. They reported no statistically significant difference in the development of ALD (15% controls, 18% investigational arms) at mean follow-up of 36.4 months. ALD was defined as clinical and radiologic evidence of adjacent level degeneration requiring “active intervention.” Unfortunately, because the results involve data pooled from multiple independent device trials, with varied surgical indications and a high incidence of ALD in both arms, there remains significant room for additional interpretation. Last, as mentioned previously, the recently published 5-year follow-up data for 271 of the patients enrolled in the original Prestige-ST FDA Investigational Device Exemption trial also failed to demonstrate a statistically significant difference for the rates of surgeries performed for ALD (2.9% versus 4.9%, arthroplasty versus arthrodesis).50

Summary

Critical review of the literature suggests that total cervical disk arthroplasty is a comparatively safe and effective alternative to anterior fusion for appropriately selected patients in the short term. With a reduction in the development of symptomatic ALD as the primary goal of arthroplasty, however, there is no compelling long-term level 1 clinical evidence to suggest its superiority for routine use over fusion. Moreover, the short- and long-term effects of motion preservation on the posterior elements (eg, ligamentum flavum, facet joints) have not been adequately studied. Marketed arthroplasty implants have a broad range of biomechanical and kinematic properties, and the long-term clinical consequences of these alternative designs are also incompletely understood. In the opinion of the authors, evidence showing clear superiority for cervical disk replacement over fusion will be difficult to generate due to the excellent results currently achieved with ACDF; the biomechanical complexity of the trijoint complex; and the requirements for rigorous clinical study designs with longer-term follow-up, larger study populations, and standardized means of evaluating competing devices.

Lumbar Arthroplasty

Low-back pain is a major health problem in Western countries, the major causes of which are thought to be DDD and facet arthropathy.60,61 It has been hypothesized that through disk dehydration, annular tears, and loss of disk height, DDD can result in abnormal motion of the involved segment and biomechanical instability causing pain. Similarly, chronic
facial stress leads to hypertrophy, osteophyte formation, distortion of innervating elements, pathologic motion within the facet capsule, and pain.62,63 While conservative treatment modalities such as physical therapy, massage, and oral medication regimens are initially used, incomplete pain control often leads patients and health care providers to consider surgical intervention. Lumbar fusion or arthrodesis has been considered the criterion standard surgical treatment for DDD and facet arthropathy.64 This can be accomplished via multiple surgical corridors (eg, posterior, posterolateral, lateral retroperitoneal, anterior retroperitoneal) and may involve fusion across any or all of the 3 lumbar columns via screw/rod constructs, interbody devices, clamp/plating systems, and posterolateral autograft/allograft supplementation. The theoretic means by which lumbar arthrodesis achieves pain relief is the elimination of motion at the fused segment, regardless of which column is serving as the pain generator.

The long-term results of lumbar spinal fusion have been mixed.62,65,66 While selected patients experience decreased pain and disability versus conservative treatment,65 not all patients achieved significant pain relief, and a proportion of patients developed pain caused by further degeneration at levels not treated during the initial operation (ie, ALD). Furthermore, infectious complications are common, seen in 10%–40% of patients.67 Despite the lack of convincing prospective evidence supporting spinal fusion for pain relief68,69 the number of fusion procedures performed is continually increasing, with an estimated 77% increase between 1996 and 2001.70

As an alternative surgical procedure, total lumbar disk replacement or lumbar arthroplasty was developed as a means of relieving pain while restoring and maintaining segmental load transfer, sagittal balance, and the spinal segment motion.71-73 This can be accomplished via multiple surgical corridors (eg, posterior, posterolateral, lateral retroperitoneal, anterior retroperitoneal) and may involve fusion across any or all of the 3 lumbar columns via screw/rod constructs, interbody devices, clamp/plating systems, and posterolateral autograft/allograft supplementation. The theoretic means by which lumbar arthrodesis achieves pain relief is the elimination of motion at the fused segment, regardless of which column is serving as the pain generator.

The long-term results of lumbar spinal fusion have been mixed.62,65,66 While selected patients experience decreased pain and disability versus conservative treatment,65 not all patients achieved significant pain relief, and a proportion of patients developed pain caused by further degeneration at levels not treated during the initial operation (ie, ALD). Furthermore, infectious complications are common, seen in 10%–40% of patients.67 Despite the lack of convincing prospective evidence supporting spinal fusion for pain relief68,69 the number of fusion procedures performed is continually increasing, with an estimated 77% increase between 1996 and 2001.70

As an alternative surgical procedure, total lumbar disk replacement or lumbar arthroplasty was developed as a means of relieving pain while restoring and maintaining segmental load transfer, sagittal balance, and the spinal segment motion.71-73 It was additionally hypothesized that the use of these devices would decrease the incidence of fusion-induced degeneration at adjacent segments, further improving clinical outcomes. The concept of lumbar disk replacement as an alternative to fusion initially gained momentum following the success of total knee and hip arthroplasty.74,75 Since the first described total disk replacement in the late 1950s,7 multiple disk replacement prostheses have been designed for use in the lumbar spine. These devices, however, carry significant cost and have a variable track record for both safety and clinical outcomes.76 In this section of the review, we will examine the evidence supporting the efficacy of these implants as a durable treatment for chronic low back pain attributed to degenerative disease in the lumbar spine.

Degeneration of the intervertebral disk is an inevitable consequence of aging, bipedal ambulation, and upright posture. It is widely accepted that disk degeneration is caused, at least in part, by the gradual deterioration of tissues subject to these constant physiologic stresses.76 Additionally, it has been well established via in vitro modeling that immobilization of a spinal motion segment further increases adjacent segment stress, as manifested by increased intradiskal pressures and an angular range of motion.77-80 Several studies, however, have identified multiple potential independent risk factors for the development of ALD, including age, postmenopausal status in women, diagnosis of regional lumbar stenosis, osteoporosis, and postfusion sagittal and coronal malalignment (eg, transition syndrome).81-84 Multiple longitudinal studies of patients receiving lumbar fusion (see below) have also failed to identify elevated rates of ALD. This failure of in vitro models to predict the clinical consequences of fusion is generally thought to be due to the inability of the pure moment loads used in laboratory models to adequately represent the more complex spinomuscular loading schemes observed in vivo.85

In 1978, Frymoyer et al86 reported their experience in a group of 207 patients (n = 143, fusion-group) followed for at least 10 years (mean, 13 years) after lumbar disk surgery. While radiographic signs of adjacent segment degeneration were more common in the fused group, there were no statistically significant clinical differences between the groups at last follow-up. Indeed, fewer (30%) of the fusion patients required further surgery compared with patients treated without fusion (37%). Nonetheless, selection bias and inherent limitations with case-control matching prevent definitive interpretation of these data. Seitsalo et al87 studied a group of 227 patients treated for isthmic spondylolisthesis (mean age, 13.8 years). One hundred forty-five patients were treated with surgical fusion, and 82 were followed conservatively. Patients were followed for a mean of 16 years. These authors found that the incidence of ALD was not influenced by the presence or absence of a fusion. Furthermore, when such changes were noted, there were no statistically significant correlations between the number of degenerative disks, the severity of ALD, or the subjective symptoms of low back pain.87 These results are particularly compelling when considering the study cohort of young otherwise healthy patients without global degenerative disease.

More recently, Kumar et al88 studied a group of 28 patients who had been treated with lumbar fusion 30 years previously and compared them with age- and sex-matched controls who had undergone lumbar microdiskectomy with or without fusion during the same time period. They, too, found that though the fusion group had a higher incidence of radiographic changes at adjacent segments, functional outcomes as measured by SF-36 and ODI were statistically no different in the 2 groups.88 While each of the aforementioned series had extensive clinical follow-up, they represent a mixture of retrospective, case-control, and prospective cohort designs. None of the studies were specifically powered to make statistically significant conclusions regarding the natural history and/or risk of ALD in both nonfused and fused populations.

The debate over the clinical relevance of fusion-induced ALD calls into question one of the theoretic indications for lumbar arthroplasty over fusion. In fact, while arthroplasty is designed to treat low back pain caused by degenerative disease, several differences between fusion and arthroplasty should be clarified to understand the specific indications for the latter. First, fusion across a motion segment should eliminate pain deriving from any spinal pain generator, including the disk space, facets, and associated structures. Given that arthroplasty only addresses the disk space and does not eliminate motion, the procedure is not designed to eliminate pain from sources outside the disk (eg, facet arthropathy).76 Second, arthroplasty is not designed to stabilize the spine; therefore, patients with translational deformity, particularly spondylolysis, are not good candidates for the procedure.89,90 Third, because arthroplasty is designed to maintain if not restore physiologic motion at the disk space, patients with minimal-
to-no segmental motion secondary to degenerative or pathologic (eg, diffuse idiopathic skeletal hyperostosis, ankylosing spondylitis) segmental autofusion are not appropriate candidates for the procedure. As Bertagnoli and Kumar described, the ideal patient for lumbar disk arthroplasty has refractory low back pain, a single level of disk disease, >4 mm of retained disk space height in the index level, no evidence of facet arthropathy, intact posterior elements, and no neurologic deficit. In a recent review, Simons noted that such patients are rare—fewer than 7% of those receiving surgical intervention for degenerative low back pain would be potential candidates for arthroplasty based on the Charité Artificial Disk (DePuy Spine; Raynham, Massachusetts; see below) exclusion criteria. Ultimately, in seeking to replicate or augment the functioning of the normal spinal elements, a lumbar arthroplasty device must take into consideration both the quantity and quality of motion that occurs across the replaced joint.

The simplest but most relevant parameter for evaluating the biomechanical effectiveness of an arthroplasty device is the physiologic ROM, defined as the amount of motion possible across the joint at a prechosen nondestructive load. ROM can be evaluated in terms of translation or rotation about any axis. Angular ROM is more pertinent for rotational motion (eg, flexion/extension), whereas linear ROM is more pertinent for translational motion (eg, axial translation during loading). Although challenging to precisely define in the clinical setting, the effectiveness of the arthroplasty device is measured by comparing appropriate ROM before and after arthroplasty with the normal range for a given lumbar level. (Replicating the preoperative ROM for the level of interest is not advantageous given that the level is already abnormal.) For the arthroplasty to be effective, postarthroplasty ROM should be at least proportionally equivalent to normal.

The AOR is next in importance to the ROM as a parameter for evaluating lumbar arthroplasty. The AOR is the line in space about which rotation occurs during motion of the spine. In purely linear motions (eg, compression, anteroposterior translation), the AOR is at infinity. During more common bending and twisting motions, the AOR lies in or near the disk space. Most important, the AOR is not a fixed point. Rather, the path or “centrode” of the AOR must be evaluated over a complete movement. This complexity can hamper precise evaluation in vivo, where 2D AOR approximations are taken from x-ray data. In vitro, the use of optical markers allows for precise (3D) observations, though loads applied in such a setting are a simplification of true in vivo loading. A spine in which clinically successful arthroplasty has been applied may have exactly the same ROM as the preoperative condition but may have shifted the AOR to a physiologic location. Ultimately, preservation and/or restoration of both normal ROM and AOR are required to declare an arthroplasty device effective.

“Coupling” refers to secondary motion that occurs in addition to the primary expected motion at a joint. The best known pattern of coupling in the lumbar spine is the coincident lateral bending that occurs during axial rotation, as a consequence of the sloped facet joints.

For example, during primary axial rotation to the left, the upper lumbar spine bends laterally toward the left and the lower lumbar spine bends laterally toward the right as coupled motions. Coupling is an important biomechanical parameter because it indicates the 3D quality of spine motion. Lumbar arthroplasty should maintain the normal coupling pattern of the spine to minimize bony tissue stress and resultant facet hypertrophy and osteophyte formation. Not surprisingly, coupling properties explain why the AOR is not perpendicular to the plane of primary motion.

Most important, lumbar arthroplasty devices are designed to mimic the biomechanics of an intact motion segment but not recapitulate the biomechanics of the natural disk. Attempts to recapitulate the natural disk by using a flexible elastomeric prosthesis, such as the AcroFlex Lumbar Disk (DePuy Spine), were met with early mechanical failure, very poor clinical outcomes, and removal from the market. The alternative approach has been to change the nature of the disk from a deformable cushion to a sliding rotational joint. The sliding joint arthroplasty has greater ROM than a flexible core and relies more on native tissues to limit ROM, reducing the mechanical requirements of the arthroplasty.

While multiple different lumbar prostheses have been available and approved for use in the European market (with many subsequently withdrawn), Investigational Device Exemption trials in the United States have led to FDA approval for the Charité (III) Artificial Disk (DePuy Spine) and multiple iterations of the ProDisc design (successive generations include ProDisc I, ProDisc II, and ProDisc-L; Synthes Spine, Paoli, Pennsylvania) (Table). Both are sliding joint arthroplasty devices and are placed following complete discectomy via anterior retroperitoneal approaches to the lumbar spine. The Charité was designed in the early 1980s and is currently approved for single-level DDD from L4 to S1. It is composed of cobalt chromium endplates and an UHMWPE mobile sliding core. The ProDisc was designed in the late 1980s and is similar in composition of cobalt chromium molybdenum endplates with a UHMWPE core. The ProDisc device, however, was developed as a semi-constrained ball-and-socket articular design and fixes the AOR of the motion segment regardless of loading technique.

By contrast, the mobile core of the unconstrained Charité prosthesis makes it possible for the AOR to shift anteriorly during extension and posteriorly during flexion. This pattern more successfully recreates the AOR observed in the normal native spine at L4–5 and L5–S1 in vitro. The better ability of the Charité to allow the joint to find its natural AOR has been described as its most significant biomechanical advantage over the ProDisc. As might be anticipated from its semi-constrained design, the advantage of the ProDisc over the Charité appears to be its improved ability to limit anteroposterior translational ROM in response to shear loads, thereby theoretically providing greater stability. Of particular concern in both arthroplasty devices, however, is the unconstrained nature of axial rotation. In this mode, the arthroplasty relies heavily on the remaining annulus and posterior elements to resist motion, providing only a guiding effect to ensure that the facets interact squarely. Fortunately, in vitro modeling has shown the facets to be potent resistors of axial rotation.

As of 2010, two RCTs have been conducted specifically to address the safety and efficacy of lumbar arthroplasty versus lumbar fusion in patients with symptomatic lumbar DDD. While 16 prospective comparative cohort studies were conducted...
ducted in the same period, the strength of their conclusions regarding efficacy (by design) are inferior to that of an RCT. As such, we discuss their results only when addressing device safety and complications.

The Charité trial, which was designed as a noninferiority trial, randomized 304 patients to either arthroplasty with the Charité III disk \((n = 205)\) or anterior interbody fusion with the BAK (Zimmer Spine, Minneapolis, Minnesota) cage \((n = 99)\) with follow-ups of 2 and 5 years. Inclusion criteria were single-level symptomatic DDD at L4-S1, back and/or leg pain without radiculopathy, VAS \(\geq 40\), ODI \(\geq 30\), and failure after \(\geq 6\) months of conservative treatment. Exclusion criteria were significant and included previous lumbar fusion or fracture, osteoporosis, facet joint arthrosis, collapsed disk space, spinal stenosis, spondylolisthesis of \(\geq 3\) mm, and scoliotic deformity of \(\geq 11^\circ\), among others. The primary outcomes were pain (VAS), functional impairment (ODI), and overall clinical success (defined by using 4 criteria: \(\geq 25\%\) improvement in ODI, no device failure, no major complication, and no neurologic deterioration).

As a secondary outcome, patient satisfaction was measured. The comparative improvements in pain scores \((-40.6 \text{ versus } -34.1, \text{ arthroplasty versus fusion})\) and functional impairment \((-24.3 \text{ versus } -21.6, \text{ arthroplasty versus fusion})\) were not statistically different at 2-year follow-up. These findings were recapitulated at 5-year follow-up. Composite clinical success percentages revealed that the Charité group was noninferior to the lumbar fusion group both at 2-year (57.1 versus 46.5\%, \(P < .0001\)) and 5-year (57.8 versus 51.2\%, \(P < .04\)) follow-ups. Patient satisfaction scores were significantly better in the Charité group (73.7\%) at 2-year follow-up compared with the control group (53.1\%, \(P < .002\)). Five-year follow-up satisfaction scores were broadly in line with 2-year results, though these data were drawn from only 57\% of the originally randomized population and were thought to be highly biased. Radiographic analysis by McAfee et al\(^7\) demonstrated maintenance of flexion/extension ROM in the Charité group with a mean ROM of 7.5\° versus a baseline value of 6.6\°.

Complications can be separated into those related to the surgical approach (eg, vascular injury, nerve root injury, retrograde ejaculation), prosthesis/fusion failure (eg, subsidence, osteolysis, migration, implant fracture, endplate fracture, pseudoarthrosis), donor-site complications, and miscellaneous (eg, infection, pain) (Fig 4). Blumenthal et al\(^9\) described overall complication rates in the Charité trial as 29.1\% for arthroplasty and 50.2\% for fusion at 2-year follow-up, though the FDA report on the trial noted overall adverse event
percentages of ~75% for both groups.95 This discrepancy is most likely a by-product of a more exhaustive definition of “adverse events” by the FDA. Geisler et al102 examined neurologic complications (eg, dysesthesia, pain, index-level motor deficit) and found no difference (16.6% for arthroplasty versus 17.2% for fusion, P > .3) between the 2 groups. Device failures necessitating repeat operations have been reported between 5.4% and 6.3% for arthroplasty and 9.1% and 10.1% for fusion at 2-year follow-up.90,103

The ProDisc trial, also designed as a noninferiority trial, randomized 236 patients to either arthroplasty with the ProDisc-L device (n = 161) or to lumbar circumferential fusion (anterior interbody fusion with femoral ring allograft and post-terolateral fusion with autologous iliac crest bone graft and pedicle screws) (n = 75).104 Outcomes were reported with 2-year follow-up. Inclusion and exclusion criteria were similar to those of the Charité trial. Clinical success was defined by using a combination of 4 clinical and 6 radiographic outcomes as required by the FDA (ie, ODI increase ≥15%, SF-36 improvement, no repeat operation to revise ProDisc or fusion, no neurologic injury, no device migration, no subsidence, no loss of disk height of ≥3 mm, among others). Pain (VAS) and functional impairment were additional primary outcomes. Although the clinical success (as defined above) rate was reported as significantly better in the ProDisc (54.3%) than in the fusion group (40.8%) (P = .044), again demonstrating noninferiority of the arthroplasty device, it is unclear even from the 2007 Zigler et al104 publication what statistical testing was applied to derive this calculation. As a consequence, this is considered a highly biased result.64

There were no significant differences with respect to mean functional impairment change (~28.9 versus ~22.9%, arthroplasty versus fusion) and pain score change (~39 versus ~32, arthroplasty versus fusion). The overall complication rates reported by Zigler et al104 were similar between the 2 groups: 7.3% and 6.3% for arthroplasty and fusion, respectively. Similar to the Charité trial, the FDA report on the Pro-Disc trial noted adverse event rates of ~85% for both groups (“FDA Approval ProDisc,” 200696). Repeat operation rates were statistically no different when comparing the arthroplasty group (3.7%) and controls (5.4%).104

Numerous concerns regarding design and outcome methodology in both the Charité and ProDisc trials have been described. To begin, the selection of the BAK fusion in the control group for the Charité trial has been criticized by multiple authors.76,105,106 While the BAK cages were the only FDA-approved interbody devices available at the time the study was designed and most closely resembled lumbar disk arthroplasty in terms of approach-related morbidity, the greatest success observed with BAK cages for interbody fusion had been in patients with collapsed disk spaces.107 Unfortunately, patients with disk space heights of <4 mm were excluded from the Charité trial, and this decision caused bias against the control group. In fact, the results obtained in the fusion group were very poor (clinical success rate, 46.5%) compared with other contemporary series of anterior lumbar interbody fusion in properly selected patients (equivalent clinical success rate, 85%-95%).107,108

Second, both the Charité and ProDisc trials have been cited for their liberal ODI increase requirement as a component of clinical success (Charité, 25%; ProDisc, 15%). Recent consensus suggests that a 30% ODI increase defines clinically relevant improvement for conservative interventions and that this benchmark should be further elevated for more investigational and/or costly procedures.109 Re-stratification of clinical success rates in both trials based on a 30%-35% ODI benchmark has been suggested as a means of clarifying the clinical relevance of the data.64 Third, the means by which pain scores were incorporated as an outcome measure has been challenged. Resnick and Watters76 noted that neither trial incorporated pain relief or opioid use into the definition of clinical success, yet 64% of those judged to have achieved such success in the Charité trial were using narcotic pain medications 24 months following surgery.80 Furthermore, in the ProDisc trial, given that ODI and VAS scores do not account for pain location, there is high likelihood of bias against the control group given the increased pain related to the harvest of the iliac crest autograft and the combined anteroposterior fusion.64 Last, the 2-year and 5-year maximum follow-up periods for the 2 trials are likely inadequate, particularly given the relatively young age (mean, ~39 years) of the trial populations as well as the unknown in vivo life span of a lumbar arthroplasty device.85

Summary

At the present time, there is insufficient evidence to support the superiority or routine use of lumbar arthroplasty for symptomatic lumbar degenerative disk disease, even in the highly selected populations that meet exclusion criteria for the placement of these devices. Improved outcomes related to pain and functional status versus fusion have not been reliably reported in the short or long term. Most important, the true incidence of ALD attributable directly to fusion in the lumbar spine remains unclear, and significant reductions in the development of ALD following fusion versus arthroplasty have not been demonstrated. The authors of this review again speculate that evidence demonstrating the superiority of lumbar arthroplasty over fusion is unlikely to be generated. The necessary stringency of study design, study population size, the development of further device modifications, and the need for longitudinal follow-up may no longer be possible in the current era of device cost containment.

Disclosures: Steven W. Chang—Research Support (including provision of equipment or materials): Medtronic; Details: institutional funding received for biomechanical studies of lumbar stability after posterolateral versus lateral fusion constructs; Speaker Bureau: Stryker Spine; Details: educational consultant for physician seminars and corporate educational series; Other Financial Relationships: Johnson & Johnson; Details: 2010 resident educational course speaker fees (Boston, Massachusetts).

References


