ORIGINAL ARTICLE

Optimizing success with lumbar disc arthroplasty

Matthew F. Gornet · Francine Schranck ·
Nicholas D. Wharton · Douglas P. Beall ·
Elizabeth Jones · Mark E. Myers · John A. Hipp

Abstract

Purpose This study aimed at determining the variables that may prove useful in predicting clinical outcomes following lumbar disc arthroplasty.

Methods Pre- and post-operative imaging assessments were obtained for 99 single-level lumbar disc arthroplasty patients from a prospective IDE study. The assessments and patient demographics were tested to identify variables that were significantly associated with clinical outcomes.

Results Clinical outcome data were available for 85 % of patients at the 5-year follow-up. Numerous assessments made from the pre-operative imaging were found to have statistically significant associations with clinical outcomes at 2 and 5 years. The most notable factors were related to the amount of degeneration at the index level, with patients achieving better outcome scores at 5 years if they have higher grades of degeneration preoperatively.

Conclusions Several variables may prove effective at optimizing clinical outcomes including a preoperative disc height <8 mm, Modic type 2 changes adjacent to the target disc, a low amount of lordosis present at the treatment level, low levels of fatty replacement of the paraspinal musculature, a prominent amount of facet joint or disc degeneration, and the presence of flat or convex vertebral endplates. There were also post-operative findings associated with better patient outcomes including a larger percent of the endplate covered with the implant, larger implant heights, greater increases in disc space heights, and a larger increase in index level lumbar lordosis. These variables could be explored in other clinical studies to facilitate meta-analyses that could identify effective strategies to optimize clinical outcomes with lumbar disc arthroplasty.

Keywords Lumbar disc arthroplasty · Predicting outcomes · Disc height · Disc degeneration · Modic changes · Lordosis · Endplates

M. F. Gornet (✉)
The Orthopedic Center of St. Louis, St. Louis, MO, USA
e-mail: mfgspine@gmail.com

F. Schranck
SPIRITT Research, 12977 North Forty Road, Suite 200,
St. Louis, MO 63141, USA
e-mail: fschranck@spirittresearch.com

N. D. Wharton · J. A. Hipp
Medical Metrics, Houston, TX, USA

D. P. Beall
Department of Radiology, Clinical Radiology of Oklahoma,
Oklahoma City, OK, USA

E. Jones
Houston Physician's Hospital, Webster, TX, USA

M. E. Myers
Center for Diagnostic Imaging, St. Louis Park, MN, USA

Introduction

Previous research has provided evidence for potential benefits of lumbar disc replacements, although clinical adoption of the technology is currently limited [1]. Uncertainty about guidelines for patient selection may be partially responsible for the limited adoption of this technology. Surgeons must consider many patient characteristics when offering treatment options, and clinical research data are essential to determining the optimal treatment for each patient. If disc arthroplasty is chosen, they must select

the specific device design, size of the disc arthroplasty device, the amount of lordosis, and other factors.

The predictive efficacy of baseline patient demographics, standardized clinical assessment scores, and psychological characteristics has been examined in several studies [2–5]. The importance of evidence-based medicine in the field of disc arthroplasty has been addressed, and the inability to draw definitive conclusions from existing data has been recognized [5]. There is currently only limited evidence to support decisions about using preoperative imaging for patient selection and limited evidence explaining the individual technical details of implanting disc arthroplasties. It may be possible to overcome this limitation and to optimize the benefits of motion preserving technologies for the lumbar spine through a better understanding of patient selection and operative techniques. Meta-analyses of the literature can be of great benefit in assessing the evidence for disc arthroplasty, but require consistent reporting of data between studies. Consistent reporting is also difficult given the paucity of data indicating the most important variables to select for analysis.

With a goal of identifying strategies to optimize patient benefits and identify variables that might be reported in future studies, comprehensive clinical and imaging data for 99 patients from a single site were analyzed to identify the factors associated with clinical outcomes.

Materials and methods

The data for the current study are for patients from a single site that was part of an ongoing Investigational Device Exemption (IDE) study [6]. The MAVERICK (Medtronic, Memphis, TN) lumbar disc replacement is an Investigational Device and is limited by Federal or United States law to investigational use. The inclusion and exclusion criteria are described in detail in a prior publication [6]. Briefly, patients were 18- to 70-year-old, with degenerative disc disease and associated back pain, a Preoperative Oswestry Disability Index (ODI) ≥30, and had failed 6 months of conservative treatment. Patients were followed for up to 7 years. Clinical outcomes were recorded using numerical pain rating scales evaluating frequency and duration of back and leg pain, the ODI (as implemented in the Lumbar Spine Questionnaire Copyright 2000 AAOS/NASS/SRS/CSRS/ORA/ASIA/COSS), and the SF-36 for general health. Data were collected pre-operatively and at 6 weeks, 3 months, 6 months, 1 year, 2 years, and continuing for a maximum of 7 years. Neutral–lateral, flexion–extension, and anterior–posterior radiographs were also obtained at each time point.

The current single-site study is based on data for 99 patients where a MAVERICK lumbar disc replacement

Table 1 Quantitative measurements collected and analyzed for associations with preoperative and postoperative clinical outcome scores

Variable	Details	Radiographs
Disc height	Measured at the anterior and posterior edges of the disc space. The average of the anterior and posterior disc heights was also measured	Neutral–lateral
Disc angle	Angle between endplates	Neutral–lateral
Global lordosis	Measured as the angle between the superior endplate of L1 and the superior endplate of S1	Neutral–lateral
Angular motion	Relative sagittal plane intervertebral rotation between flexion and extension	Flexion–extension
Translation	Sagittal plane translational motion between vertebrae. The translation per degree of rotation was also calculated	Flexion–extension

All measurements were collected at the treatment level. All measurements were made at preoperative, 2 years, and 5 years. All measurements were made using previously validated [7] computer-assisted technology (QMA®, Medical Metrics, Inc., Houston, TX). The change in these parameters was also calculated at postoperative time-points relative to preoperative

was implanted at a single level (38 at L4–L5 and 61 at L5–S1). Clinical outcomes data and radiographic assessments from preoperative, 2- and 5-year time points were analyzed. Multiple measurements were obtained from the radiographs using previously validated software (Quantitative Motion Analysis—QMA®, Medical Metrics, Inc.) [7]. The measurements are summarized in Table 1.

In addition to the X-rays, CT and MRI exams were obtained preoperatively and used to qualitatively assess the variables described in Table 2. The qualitative assessments were made by two radiologists with a third radiologist to adjudicate disagreements. All of the radiologists had extensive experience in systematic assessment of lumbar disc replacements in IDE studies, and all worked independent of each other and independent of the enrolling site. All three radiologists received a comprehensive training program prior to performing any assessments. The training program included multiple examples of each grade and addressed potential sources of reader disagreement. Image atlases based on original publications, and exception handling rules were also used by the readers for the assessments of adjacent level degeneration [8, 9].

Logistic regressions, analysis of variance, and Kruskal–Wallis equality of means tests were used to identify variables that may help to predict whether patients had a good clinical outcome (Stata Ver 11, College Station, TX). Two definitions of a "good" clinical outcome were used: (1) a 15-point improvement in the ODI score relative to the

Table 2 Summary of qualitative assessments collected and analyzed

Parameter	Description	κ Reader 1	κ Reader 2	References
Fatty replacement	Amount of fat interspersed within muscles posterior to the spine at the L4–L5 level	0.64	0.71	[25]
Modic change	Changes in MRI signal characteristics in the marrow adjacent to the vertebral endplates, relative to normal marrow signal for the patient disc at the treatment level	0.79	0.84	[26]
Index level disc degeneration from MRI	Changes in the disc height and MRI signal characteristics in the intervertebral disc at the treatment level relative to normal adjacent discs	0.91	0.92	[9]
Adjacent level disc degeneration	Changes in disc height, extent of osteophytes and endplate sclerosis at the intervertebral discs adjacent to the treatment level	0.80	0.90	[8]
Facet degeneration	Amount of facet degeneration at the treatment level	0.55	0.77	[27]
Endplate morphology	Morphology of the endplates at the treatment level	0.67	0.51	[28]

The assessments were collected from imaging obtained preoperatively and two and five years postoperatively. The level of agreement above chance, calculated using κ statistics (Stata Ver. 11, Stata Corp., College Station, TX), between each of the two primary readers and the final tie-broken result that was analyzed for predicting outcomes, is provided in the first column. Since a multi-grade difference between readers is far more significant than a single grade difference, quadratic weighted κ was used for most assessments, except Modic and endplate morphology where unweighted κ is provided since a single-grade difference is equally as important as a multi-grade difference
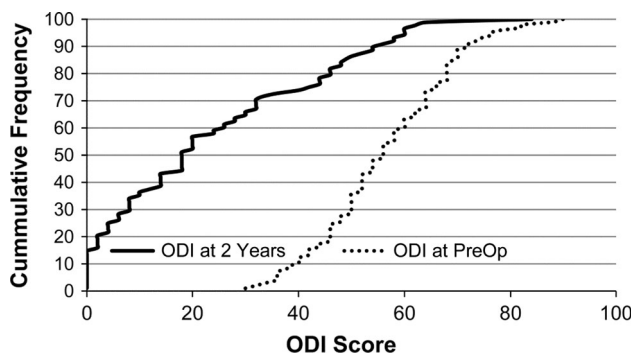

Fig. 1 The cumulative frequency of ODI scores shows the shift in ODI scores that resulted from disc arthroplasty surgery. Note that no patient had an ODI score <30 prior to surgery, whereas 65 % had an ODI <30 2-years after surgery
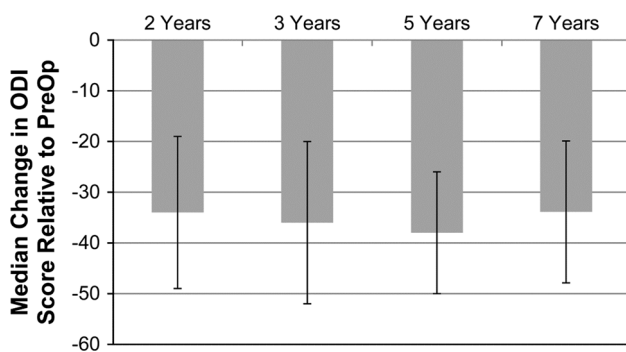

Fig. 2 Median change in the ODI score relative to PreOp. The *error bars* show the inter-quartile range. The data support that the clinical improvement was sustained out to 7 years

Preoperative ODI score (commonly used in clinical studies) and (2) a substantial clinical benefit (SCB), identified as a post-operative ODI score that was <31.3 [10].

Results

Summary of clinical outcomes

Clinical outcomes were obtained for 90 % of patients at 2 years, and 85 % at 5 years. The preoperative ODI scores were normally distributed and averaged 56.5 (SD 12.7). In contrast, the ODI scores were strongly distributed toward the low end of the ODI scale after surgery. The shift in the cumulative frequency of ODI scores also illustrates the effect of surgery on these patients (Fig. 1). A 15-point improvement in the ODI score relative to the preoperative ODI was observed in 76.1 % of patients at 2 years, and 84.3 % at 5 years. An SCB, defined as an ODI <31.3, was achieved in 65.9 % of patients at 2 years and 78.3 % at 5 years. There was no difference between postoperative time-points with respect to the median change in ODI score (Fig. 2, $P = 0.6$ Kruskal–Wallis test), suggesting a stable clinical benefit over time. The preoperative ODI score was not associated with whether the patients achieved a 15-point improvement in the ODI score at 2 years ($P = 0.69$) or at 5 years ($P = 0.13$). However, patients were less likely to achieve an SCB at 2 years if the preoperative ODI score was high (odds ratio = 0.94, $P = 0.001$). Preoperative ODI was not associated with SCB at 5 years. Siepe et al. [3] also reported an association between preoperative and postoperative outcome scores, as did Deutsch [11], who suggested a very high preoperative ODI may suggest a psychological overlay.

Reliability of quantitative and qualitative imaging assessments

The accuracy and reliability of the quantitative assessments (collected using QMA®, Medical Metrics, Inc., Houston,

TX) have been previously described [7, 12, 13]. Briefly, the accuracy of intervertebral rotation measurements at levels implanted with a disc arthroplasty was determined by comparison to optoelectronic motion measurements methods [14]. The limits of agreement (Bland-Altman test) between QMA® and optoelectronic measurements was ±1.0°. The reproducibility, calculated as the adjusted percent agreement between two observers, was reported by Pearson et al. [12] as >0.8 for intervertebral rotation, translation, and disc height change.

The reliability of the qualitative assessments was determined using kappa statistics by analyzing the assessments of each of the two primary readers compared to the adjudicated results. The levels of agreement were consistent with or better than has been described in a prior study [15]. The reliability results are included in Table 2.

Predictive efficacy of quantitative measurements

To determine the *preoperative* imaging assessments that were associated with the *preoperative* clinical outcome scores, every available demographic and preoperative radiographic imaging parameter was tested for an association with preoperative ODI, with the goal of identifying potential subgroups of patients that might be considered outliers, and to help identify potential sources of the patient's symptoms. Only a few parameters had statistically significant associations with the preoperative ODI score. All significant associations were weak, and although significant, they explained only a small amount of the variation between patients in preoperative ODI scores. Note that these associations are for all 99 patients. It can be hypothesized that there is a proportion of patients where individual parameters were very important to that patient, whereas in other patients, the parameter was not important, yielding the appearance that the parameter is only modestly important. This hypothesis is difficult to validate and would require a larger sample size to test for subpopulations that do and do not benefit.

The preoperative ODI was higher for patients with a small (low lordosis) disc angle at the treated level ($P = 0.04$, $R^2 = 0.042$). This was not simply due to the propensity for patients with low back pain to stand with reduced lordosis, since there was no significant association with the L1–S1 global lordosis angle (measured from the neutral-lateral standing X-ray, $P = 0.99$). These results were consistent with a study by Siepe et al. [16], in which there was no association between the preoperative disc height and the preoperative ODI score ($P = 0.6$). There was a weak ($R^2 = 0.04$) but significant ($P = 0.033$) association between preoperative sagittal plane angular range-of-motion at the treatment level and
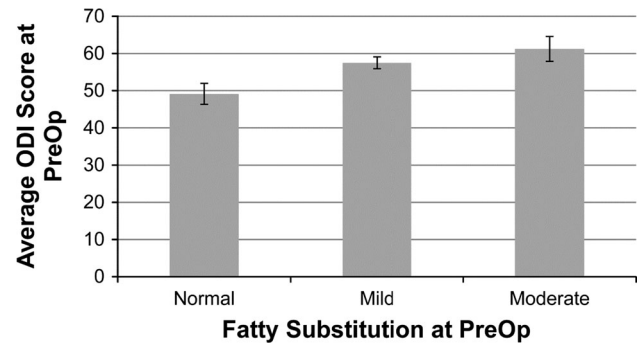

Fig. 3 The average ODI score at PreOp was significantly ($P = 0.024$) associated with the amount of fatty substitution in the muscles posterior to the L4–L5 level. The *error bars* show the standard error

the preoperative ODI score. Patients with a higher ROM had a lower ODI score. Higher levels of intervertebral translation per degree of rotation in the sagittal plane during flexion-to-extension were associated with a higher preoperative ODI score ($P = 0.019$, $R^2 = 0.056$). This higher level of translation per degree of rotation is an indicator of instability [17]. The preoperative ODI score was highest for patients with the greatest amount of fatty replacement in the posterior muscles ($P = 0.024$; Fig. 3). Preoperative Modic changes, Pfirrmann grade of disc degeneration, or the amount of facet degeneration were not generally associated with the preoperative ODI score ($P > 0.42$).

Preoperative parameters associated with postoperative outcome scores

There was no evidence that the 2-year ODI scores depended on whether the patient was male or female ($P = 0.1$), or whether they smoked ($P = 0.44$). Worker's compensation patients had significantly less improvement in the ODI score at 2 and 5 years ($P = 0.0009$ at 2 years, $P = 0.013$ at 5 years; Fig. 4). Younger patients were more likely to have achieved a 15-point improvement at 2 years ($P = 0.047$), but the strength of the association was weak (odds ratio = 0.94). This relationship did not exist at 5 years ($P = 0.99$).

Several imaging parameters assessed preoperatively were predictive of whether the patient achieved a 15-point improvement in the ODI or whether they had an ODI <31.3 at follow-up, and are summarized below. A variable that was nearly predictive included intervertebral rotation, which was just out of the range of being associated with an ODI <31.3 at follow-up ($P = 0.06$ at 2 years). The trend was for more motion in patients with good outcomes. This finding was similar to the results reported by Siepe et al. [16].
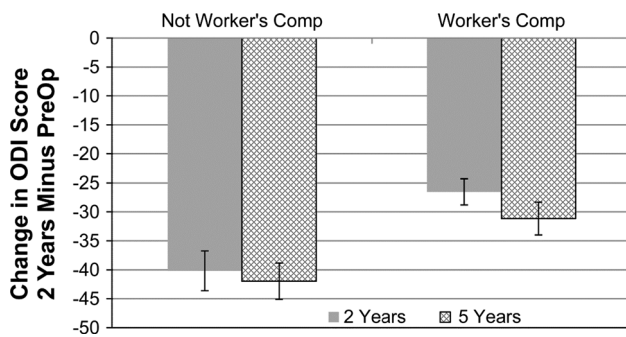
Fig. 4 The average changes in non-Worker's Compensation patients and Worker's Compensation patients at 2 and 5 years. Worker's compensation patients had less improvement in the ODI scores at 2 years ($P = 0.0009$) and at 5 years ($P = 0.013$). The *error bars* provide the standard errors

Preoperative Pfirrmann grade and disc height at the index level

Patients with a low Pfirrmann [9] grade of intervertebral disc degeneration (mildly degenerated disc) had a significantly higher ODI score at 5 years than patients with more degenerated discs ($P = 0.04$, one-way ANOVA). Patients with the most degenerated disc had the greatest improvement in ODI scores at 5 years postoperative ($P = 0.018$, one-way ANOVA; Fig. 5). An association was also found between the preoperative disc height and whether the patient achieved a 15-point improvement in the ODI ($P = 0.033$ at 2 years, $P = 0.027$ at 5 years), with those patients achieving good outcomes having thinner discs preoperatively. The preoperative grade (Kellgren–Lawrence) of disc degeneration at the adjacent levels was not a significant predictor of clinical outcomes ($P = 0.88$ at 2 years).

Preoperative vertebral marrow changes (Modic)

At 5 years before surgery, patients who had Modic 2 changes in the marrow adjacent to the endplates preoperatively had lower ODI scores than patients with no Modic changes or type 1 Modic changes ($P = 0.037$, Kruskal–Wallis test for equality of distributions, $P = 0.043$, one-way ANOVA). The median ODI at 5 years was 24 in patients with no preoperative Modic changes, 21 for patients with type 1 changes, and 14 for patients with type 2 changes.

Preoperative facet degeneration

At 2 years postoperative, patients who had the most severe facet degeneration at preoperatively also had the greatest improvement in their ODI score ($P = 0.025$, one-way ANOVA; Fig. 6).
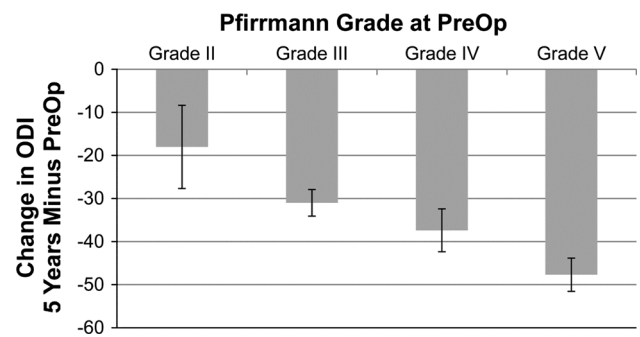

Fig. 5 The average change in ODI scores at 5 years PostOp was significantly ($P = 0.018$) greater for patients who had the most degenerated discs at PreOp. The *error bars* show the standard error
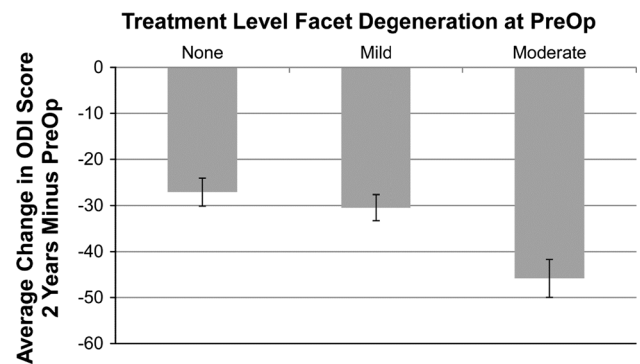

Fig. 6 The average change in the ODI score, measured at 24 months after surgery relative to before surgery, grouped by the amount of facet degeneration at PreOp at the treatment level. These data show greater improvement in patients who had more severe degenerative changes of the facet joints. The *error bars* show the standard error

Endplate width

Patients who did not achieve an improvement in ODI of ≥15 at 2 years had slightly larger (in the AP direction) endplate width (31.3 ± 3.13 versus 29.5 ± 3.0 mm, $P = 0.024$, one-way ANOVA). Endplate measurements were also larger on average in patients who did not have an ODI <31.3 at 2 years ($P = 0.018$, one-way ANOVA).

Endplate morphology

The change in the ODI score at 2 years after surgery was significantly associated with the preoperative morphology of the vertebral endplates at the treatment level ($P = 0.04$ one-way ANOVA; Fig. 7). Patients with flat or convex endplates tended to do better than patients with hooked or concave endplates.

Percent of overall lordosis at the treatment level

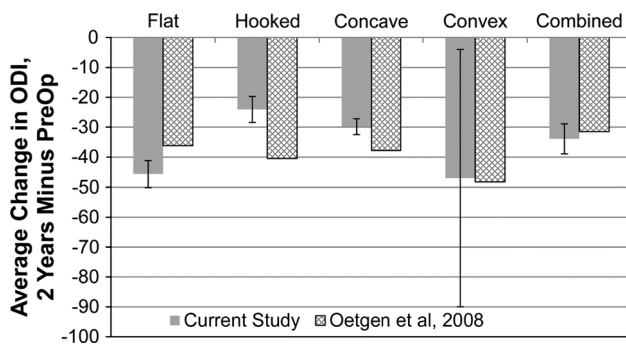The average overall lordosis was slightly greater at 5 years compared to preoperative lordosis (60.8 versus 56.7,

Fig. 7 The average change in the ODI score at 2 years relative to before surgery, grouped by the morphology of the vertebral endplates at the treatment level. These data support greater improvement in patients with flat or convex endplates than in those patients with hooked or concave endplates. Similar data from a study by Oetgen et al. [24] are also included. The *error bars* show the standard errors

Fig. 8 The average percent of the vertebral endplate covered by the implant was significantly greater ($P = 0.024$) in patients who had and ODI score <31.3 at 2 years. The *error bars* show the standard error

$P = 0.02$). Those patients where a larger percent of the overall (L1–S1) lumbar lordosis was found at the treatment level preoperatively were less likely to have an ODI <31.3 at 5 years ($P = 0.044$, logistic regression), and were less likely to have at least a 15-point improvement in the ODI score at 5 years ($P = 0.0024$, logistic regression). Patients who did not achieve at least a 15-point improvement in the ODI score had on average 32 ± 8.1 % of the overall lordosis at the treatment level preoperatively compared with 24.9 ± 10.9 % in patients who did improve ($P = 0.018$, one-way ANOVA).

Fatty replacement

Patients who did not achieve at least a 15-point improvement in the ODI score at 2 years tended to have more fatty replacement in the muscles posterior to the L4–L5 level preoperatively (average fatty replacement score 1.25 ± 0.55 versus 0.98 ± 0.58, $P = 0.075$ one-way ANOVA).

Variables that can be controlled at surgery and may influence clinical outcomes

There are multiple variables that are at least partially under the control of the surgeon at the time of surgery and may influence clinical outcomes.

Percent of endplate covered by the implant

The percent of the vertebral endplate that was covered by the implant, as measured in the sagittal plane from neutral–lateral radiographs, had a significant association with whether the patient had an ODI score <31.3 at 2 years ($P = 0.024$; Fig. 8). The association was also evident in the 5-year ODI data ($P = 0.014$). Patients with a larger percent covered were more likely to have an ODI <31.3 at
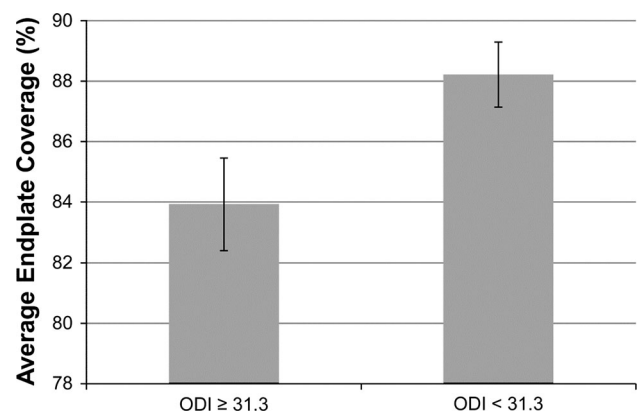
2 years (odds ratio 1.07, $P = 0.028$). The percent of the endplate covered varied from 66 to 109 % (average 87 %).

Disc distraction

The patients who achieved a 15-point improvement in the ODI score at 2 years tended to have a greater increase in disc space at the treatment level (measured as the average of the disc heights at the anterior- and posterior-most extent of the disc space), measured at 6 weeks relative to preoperative ($P = 0.019$; Fig. 9). The difference is <2 mm, but the data suggest that the modest extra distraction may benefit the patient.

Implant height relative to vertebral size

The implant height that was implanted into each patient, measured as a percent of the vertebral endplate width, was significantly greater in those patients who had achieved at least a 15-point improvement in ODI at 2 years ($P = 0.022$, one-way ANOVA; Fig. 10).

Treatment level lordosis The patients who achieved a 15-point improvement in the ODI score tended to have a greater increase in lordosis at the treatment level, measured at 6 weeks relative to preoperative ($P = 0.049$; Fig. 11). Patients with greater improvement in treatment level lordosis were also significantly ($P = 0.044$) more likely to have an ODI <31.3 at 2 years.

Discussion

The improvement criteria in this study included a 15-point improvement in the ODI and a postoperative ODI score <31.3. These criteria were adopted based on prior studies
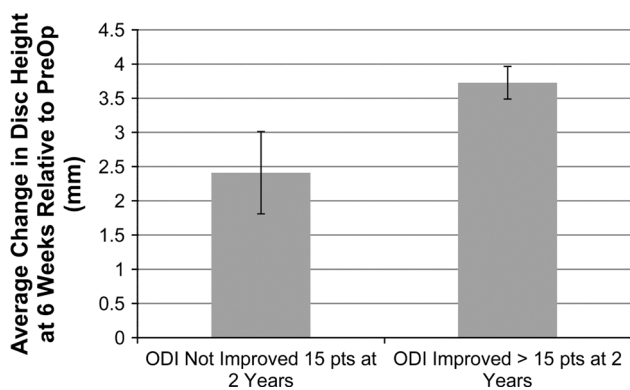
Fig. 9 Average disc distraction achieved by implantation of the disc arthroplasty, measured as the change in disc height (average of the anterior- and posterior disc space height) 6 weeks after surgery relative to before surgery, grouped by whether the patient achieved at least a 15-point improvement in the ODI score at 24 months. The *error bars* show the standard error

Fig. 11 The average change in the treatment level lordosis achieved at surgery, measured as the lordosis at 6 weeks minus lordosis at PreOp, grouped by whether the patient's ODI score had improved at 24 months by at least 15 points, relative to preoperative ODI. Patients with a greater change in the amount of lordosis at the treated level improved more than those patients with less of an increase in lumbar lordosis. The *error bars* show the standard error
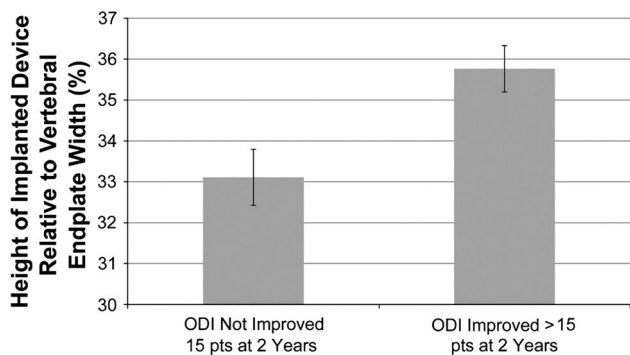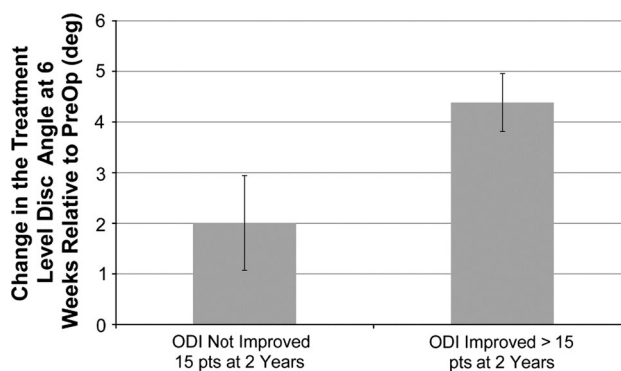

Fig. 10 The height of the implanted disc arthroplasty, expressed as a percentage of the vertebral endplate width, grouped by whether the patient had achieved at least a 15-point improvement in ODI at 24 months. Greater implant height was associated with greater patient improvement. The *error bars* show the standard errors

that used the 15-point ODI improvement threshold and a prior validation study of what defined a SCB as an ODI <31.3.

The clinical outcomes following lumbar disc arthroplasty in the current study are similar to clinical outcomes reported in prior studies of lumbar disc arthroplasties. At 2 years, 76.1 % of patients had achieved a 15-point improvement in the ODI score. In the Depuy Charite disc IDE study, 64 % achieved this improvement at 2 years [18]; in the Prodisc-L IDE, 67.8 % achieved this improvement at 2 years [19], and 82.2 % of patients achieved this improvement in the multi-site Maverick IDE [6].

Data from the current study point to several radiographic assessments that might be considered in subsequent registry studies to test specific hypotheses that might lead to

treatment optimization guidelines for lumbar disc arthroplasty. These parameters, the associated hypotheses, and any comparable reference data from prior publications include (and are summarized in Table 3):

The amount of preoperative disc degeneration at the treatment level

Hypothesis: patients do better if they have grade 3 or 4 disc degeneration preoperatively, measured using the Pfirrmann grading system [9]. Siepe et al. [20] found an association between the histological grade of disc degeneration at Preoperative and the ODI score several years later in their study of the Prodisc II disc arthroplasty. The results from the current study are also consistent with the findings of Siepe et al. that the severity of disc degeneration was not associated with the preoperative ODI score.

Preoperative disc height at the treatment level

Hypothesis: patients will do better if the disc height is <8 mm. The preoperative disc height as measured from neutral–lateral radiographs as the average of the anterior and posterior disc heights was also a significant predictor of 24- and 60-month outcomes. When pre and postoperative disc height and Pfirrmann grade were included in logistic regression analysis, only preoperative disc height was significant. Disc height was under 8 mm in only 11 % of the patients that did not have an ODI <31.3 at 5 years. However, 54 % of patients who did have an ODI <31.3 at 5 years also had a preoperative disc height >8 mm. Thus, although a significant predictor of outcomes, the sensitivity as a single parameter is suboptimal. Siepe et al. [16] observed that patients with more advanced narrowing

Table 3 Summary of the pre-operative and intra/post-operative radiographic parameters that may be significant in clinical trials of lumbar disc arthroplasty

Parameter	Potential treatment level criteria
Pre-operative	
Disc degeneration	Outcomes better if disc is Pfirrmann grade 3 or 4
Disc height	Outcomes better if disc height is <8 mm
Facet degeneration	Outcomes better in patients with degenerated facets
Modic changes	Outcomes better if Modic 2 changes
Lordosis	Outcomes worse when >32 % of overall lordosis is at the treatment level
Fatty replacement in muscle	Outcomes worse with high level of fat in muscles
Endplate morphology	Outcomes worse with hooked or concave endplates
Intra/Post-operative	
Endplate covered by implant	Outcomes worse when <85 % of endplate is covered
Increased disc height	Outcomes better with ≥3 mm increase in disc height
Increased lordosis	Outcomes better if lordosis improved by at least 3°

Additional studies, designed specifically to test these parameters, would be required to verify or refute the significance observed in the current study

reported greater postoperative satisfaction after disc arthroplasty.

Preoperative facet degeneration

Hypothesis: patients with more severe facet degeneration will experience greater improvements in the postoperative ODI score. A study by Le Huec et al. [21] would support that the null hypothesis. Our data support that the greater the facet degeneration, the greater the patient improvement when measured at 2 years. Clearly, additional data are needed on this issue. Observations by Zweig et al. [22] would support that patients with posterior element changes that create or will create radicular symptoms are a subset that will not benefit from disc arthroplasty.

Preoperative Modic changes at the treatment level

Hypothesis: patients with type 2 changes will do somewhat better. Hellum et al. [4] also found this association.

Preoperative lordosis at the treatment level relative to overall lordosis in the lumbar spine

Hypothesis: Patients do worse when preoperatively, >32 % of the overall lordosis is found at the treatment level. The

importance of lordosis was also observed by Le Huec et al. [23].

Preoperative fatty replacement

Hypothesis: Patients do worse when higher levels of fatty replacement are observed preoperatively. This was also observed by Le Huec et al. [21].

Preoperative endplate morphology at: Hypothesis: patients with hooked or concave endplates will do worse than patients with flat or convex endplates. It is likely that this association will depend on the interaction between implant design and endplate morphology.

Percent of endplate covered by the implant

Hypothesis: patients do worse when at least 85 % of the vertebral endplates are not covered by the implant.

The increase in disc height achieved at surgery

Hypothesis: patients do better with at least 3 mm increase in disc height. The implanted device height, normalized to the endplate width, may be related and a threshold level of 34 % may be an appropriate target.

Increase in treatment level lordosis achieved at surgery

Hypothesis: patients do better if lordosis at the treatment level is improved by at least 3° at surgery.

Overall, there are many factors that may contribute toward an optimal outcome in patients undergoing lumbar intervertebral disc arthroplasty surgery including a preoperative disc height <8 mm, Modic type 2 changes adjacent to the target disc, <32 % of the overall lordosis present at the treatment level, a low level of fatty replacement of the paraspinal musculature, a prominent amount of facet joint degeneration, a more advanced degree of intervertebral disc degeneration, and the presence of flat or convex vertebral endplates. There were also post-operative findings that were associated with a better patient outcome including a larger percent of the endplate covered with the implant, larger implant heights, greater postoperative increases in disc space heights, and a larger change in the amount of postoperative lumbar lordosis.

Although data from a single-site study are not sufficient to validate patient selection criteria, the results of the current study provide evidence in support of variables that should be considered in future studies. Consistent reporting of variables between studies can facilitate more meaningful meta-analyses. The statistically significant associations imaging variables and clinical outcomes that were found in the current study were generally modest. The relative

predictive value of each individual parameter or the efficacy of predictive models that include multiple parameters can be studied in large registry studies. Registry studies may prove a more powerful source of data to validate predictive equations, since inclusion and exclusion criteria can be expected to be less rigorous than in IDE studies. Large studies will enable testing for possible associations and thereby validation of guidelines that may subsequently help to avoid application of lumbar disc replacements in patients least likely to benefit. Large enough studies would enable validation of equations to provide a probability of success individualized to each patient.

References

1. Awe OO, Maltenfort MG, Prasad S, Harrop JS, Ratliff JK (2011) Impact of total disc arthroplasty on the surgical management of lumbar degenerative disc disease: analysis of the Nationwide Inpatient Sample from 2000 to 2008. Surg Neurol Int 2:139–144

2. Guyer RD, Siddiqui S, Zigler JE, Ohnmeiss DD, Blumenthal SL, Sachs BL, Hochschuler SH, Rashbaum RF (2008) Lumbar spinal arthroplasty: analysis of one center's twenty best and twenty worst clinical outcomes. Spine 33(23):2566–2569

3. Siepe CJ, Tepass A, Hitzl W, Meschede P, Beisse R, Korge A, Mayer HM (2009) Dynamics of improvement following total lumbar disc replacement: is the outcome predictable? Spine 34(23):2579

4. Hellum C, Johnsen LG, Gjertsen Ø, Berg L, Neckelmann G, Grundnes O, Rossvoll I, Skouen JS, Brox JI, Storheim K (2012) Predictors of outcome after surgery with disc prosthesis and rehabilitation in patients with chronic low back pain and degenerative disc: 2-year follow-up. Eur Spine J 21(4):1–10

5. Zindrick MR, Tzermiadianos MN, Voronov LI, Lorenz M, Hadjipavlou A (2008) An evidence-based medicine approach in determining factors that may affect outcome in lumbar total disc replacement. Spine 33(11):1262

6. Gornet MF, Burkus JK, Dryer RF, Peloza JH (2011) Lumbar disc arthroplasty with MAVERICK (TM) disc versus stand-alone interbody fusion: a prospective, randomized, controlled, multicenter investigational device exemption trial. Spine 36:E1600–E1611

7. Zhao KD, Yang C, Zhao C, Stans AA, An KN (2005) Assessment of noninvasive intervertebral motion measurements in the lumbar spine. J Biomech 38(9):1943–1946

8. Kellgren J, Lawrence J (1963) Atlas of standard radiographs: the epidemiology of chronic rheumatism, vol 2. Blackwell, Oxford, pp 46–57

9. Pfirrmann CW, Metzdorf A, Zanetti M, Hodler J, Boos N (2001) Magnetic resonance classification of lumbar intervertebral disc degeneration. Spine 26(17):1873–1878

10. Glassman SD, Copay AG, Berven SH, Polly DW, Subach BR, Carreon LY (2008) Defining substantial clinical benefit following lumbar spine arthrodesis. J Bone Joint Surg 90(9):1839

11. Deutsch H (2010) The predictive value of the baseline Oswestry Disability Index in lumbar disc arthroplasty. Neurosurg Focus 28(6):7

12. Pearson AM, Spratt KF, Genuario J, McGough W, Kosman K, Lurie J, Sengupta DK (2011) Precision of lumbar intervertebral measurements: does a computer-assisted technique improve reliability? Spine 36(7):572–580

13. Auerbach J, Namdari S, Milby A, White A, Reddy S, Lonner B, Balderston R (2008) The parallax effect in the evaluation of range of motion in lumbar total disc replacement. SAS J 2(4):184–188

14. Patwardhan AG, Havey RM, Wharton ND, Tsitsopoulos PP, Newman P, Carandang G, Voronov LI (2012) Asymmetric motion distribution between components of a mobile-core lumbar disc prosthesis: an explanation of unequal wear distribution in explanted CHARITE polyethylene cores. J Bone Joint Surg Am 94(9):846–854

15. Berg L, Gjertsen O, Hellum C, Neckelmann G, Johnsen LG, Eide GE, Espeland A (2012) Reliability of change in lumbar MRI findings over time in patients with and without disc prosthesis—comparing two different image evaluation methods. Skelet Radiol 41(12):1547–1557

16. Siepe CJ, Hitzl W, Meschede P, Sharma AK, Khattab MF, Mayer MH (2009) Interdependence between disc space height, range of motion and clinical outcome in total lumbar disc replacement. Spine 34(9):904–916

17. Frobin W, Brinckmann P, Leivseth G, Biggemann M, Reikeras O (1996) Precision measurement of segmental motion from flexion–extension radiographs of the lumbar spine. Clin Biomech (Bristol, Avon) 11(8):457–465

18. FDA (2004) PMA P040006: Summary of safety and effectiveness, CHARITE Artificial Disc

19. FDA (2006) PMA P050010: Summary of safety and efficacy, PRODISC-L total disc replacement

20. Siepe CJ, Heider F, Haas E, Hitzl W, Szeimies U, Stäbler A, Weiler C, Nerlich AG, Mayer MH (2012) Influence of lumbar intervertebral disc degeneration on the outcome of total lumbar disc replacement: a prospective clinical, histological, X-ray and MRI investigation. Eur Spine J 21(11):1–13

21. Le Huec JC, Basso Y, Aunoble S, Friesem T, Bruno MB (2005) Influence of facet and posterior muscle degeneration on clinical results of lumbar total disc replacement: two-year follow-up. J Spinal Disord Tech 18(3):219–223

22. Zweig T, Hemmeler C, Aghayev E, Melloh M, Etter C, Röder C (2011) Influence of preoperative nucleus pulposus status and radiculopathy on outcomes in mono-segmental lumbar total disc replacement: results from a nationwide registry. BMC Musculoskelet Disord 12(1):1–9

23. Le Huec J, Basso Y, Mathews H, Mehbod A, Aunoble S, Friesem T, Zdeblick T (2005) The effect of single-level, total disc arthroplasty on sagittal balance parameters: a prospective study. Eur Spine J 14(5):480–486

24. Oetgen M, Yue J, Jaramillo-de la Torre J, Bertagnoli R (2008) Does vertebral endplate morphology influence outcomes in lumbar total disc arthroplasty? part II: clinical and radiographic results as evaluated utilizing the vertebral endplate Yue–Bertagnoli (VEYBR) classification. SAS J 2(2):101–106

25. Kader DF, Wardlaw D, Smith FW (2000) Correlation between the MRI changes in the lumbar multifidus muscles and leg pain. Clin Radiol 55(2):145–149

26. Modic MT, Steinberg PM, Ross JS, Masaryk TJ, Carter JR (1988) Degenerative disk disease: assessment of changes in vertebral body marrow with MR imaging. Radiology 166(1 Pt 1):193–199

27. Weishaupt D, Zanetti M, Boos N, Hodler J (1999) MR imaging and CT in osteoarthritis of the lumbar facet joints. Skelet Radiol 28(4):215–219

28. Yue J, Oetgen M, Jaramillo-de la Torre J, Bertagnoli R (2008) Does vertebral endplate morphology influence outcomes in lumbar disc arthroplasty? Part I: an initial assessment of a novel classification system of lumbar endplate morphology. SAS J 2(1):16–22