

## CLINICAL CASE SERIES

OPEN

# Comparison of Mid- to Long-term Follow-up of Patient-reported Outcomes Measures After Single-level Lumbar Total Disc Arthroplasty, Multi-level Lumbar Total Disc Arthroplasty, and the Lumbar Hybrid Procedure for the Treatment of Degenerative Disc Disease

Matthew Scott-Young, FRACS,<sup>a,b</sup> So Mang (Simon) Lee, BExsc,<sup>a</sup> David Nielsen, FRACS,<sup>c</sup> Evelyne Rathbone, CStat,<sup>b</sup> Matthew Rackham, FRACS,<sup>a</sup> and Wayne Hing, PhD<sup>b</sup>

**Study Design.** Prospective cohort study.

**Objective.** The aim of this article is to compare the mid- to long-term patient-reported outcome measures (PROMs) between single-level total disc arthroplasty (TDA), multi-level TDA, and hybrid constructs (combination of TDA and anterior lumbar interbody fusion [ALIF] across multiple levels) for symptomatic degenerative disc disease (DDD).

**Summary of Background Data.** The treatment of single-level DDD is well documented using TDA. However, there is still a paucity of published evidence regarding long-term outcomes on multi-level TDA and hybrid constructs for the treatment of multi-level DDD, as well as lack of long-term comparisons regarding treatment of single-level DDD and multi-level DDD.

**Methods.** A total of 950 patients underwent surgery for single-level or multi-level DDD between July 1998 and February 2012 with single-level TDA ( $n=211$ ), multi-level TDA ( $n=122$ ), or hybrid

construct ( $n=617$ ). Visual Analog Score for the back (VAS-B) and leg (VAS-L) were recorded, along with the Oswestry Disability Index (ODI) and Roland Morris Disability Questionnaire (RMDQ).

**Results.** All PROMs in all groups showed statistically and clinically significant improvements ( $P<0.005$ ) in pain and function that is well above the corresponding minimum clinically important difference (MCID) and exceeds literature thresholds for substantial clinical benefit (SCB). Unadjusted analyses show that there were no statistically significant differences in the change scores between the surgery groups for VAS back and leg pain, and RMDQ up to 8 years' follow-up. Adjusted analyses showed the ODI improvement score for the single group was 2.2 points better (95% confidence interval [CI]: 0.6–3.9,  $P=0.009$ ) than in the hybrid group. The RMDQ change score was better in the hybrid group than in the multi-level group by 1.1 points (95% CI: 0.4–1.9,  $P=0.003$ ) at 6 months and a further 0.4 point at 2 years (95% CI: 0.1–0.8,  $P=0.011$ ).

**Conclusion.** In the setting of meticulous preoperative evaluation in establishing a precision diagnosis, clinically and statistically equivalent results can be achieved when treating symptomatic DDD through single-level TDA, multi-level TDA, and hybrid constructs. These results are sustained at mid- to long-term follow-up.

**Key words:** arthroplasty, artificial disc, back pain, comparison, degenerative disc disease, long-term results, lumbar spine, motion preservation, total disc replacement.

**Level of Evidence:** 3

**Spine 2022;47:377–386**

From the <sup>a</sup>Gold Coast Spine, Gold Coast, Queensland, Australia; <sup>b</sup>Faculty of Health Science & Medicine, Bond University, Gold Coast, Queensland, Australia; and <sup>c</sup>Department of Orthopedic Surgery, Princess Alexandra Hospital, Brisbane, Queensland, Australia.

Acknowledgment date: June 11, 2021. First revision date: August 12, 2021. Second revision date: August 23, 2021. Acceptance date: August 31, 2021.

The device(s)/drug(s) is/are FDA-approved or approved by corresponding national agency for this indication.

No funds were received in support of this work.

Relevant financial activities outside the submitted work: consultancy, royalties.

This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

Address correspondence and reprint requests to Matthew Scott-Young, FRACS, 27 Garden Street, Southport, Gold Coast, QLD 4215, Australia; E-mail: swalter@goldcoastspine.com.au

DOI: 10.1097/BRS.0000000000004253

Spine

The prevalence of low back pain worldwide ranges from 1.4% to 20% with a lifetime incidence ranging from 50% to 80%.<sup>1,2</sup> Low back pain caused by discogenic pain from lumbar degenerative disc disease

(DDD) is a specific diagnosis that can be treated operatively or nonoperatively.<sup>3</sup> The diagnosis is made from a combination of a clinical history, physical examination, radiological investigations, such as magnetic resonance imaging (MRI), and discriminating provocative discography with post discography computed tomography (CT) scans.<sup>4,5</sup> Other authors have also found electrophysiological studies,<sup>6</sup> MR spectroscopy,<sup>7</sup> and SPECT scanning<sup>8</sup> adjunctive in supporting a diagnosis. DDD can be present at single or multiple levels, and a variety of reconstructive options are available.

Options to treat single-level DDD operatively include total disc arthroplasty (TDA) or anterior lumbar interbody fusion (ALIF). TDA acts as a dynamic motion segment stabilizer, and has been shown to reduce back pain, improve functional outcomes, and reduce clinical and radiological adjacent segment pathology (CASP/RASP).<sup>9</sup> ALIF acts as a static motion segment stabilizer with similar improvement in clinical outcomes.<sup>10,11</sup> The pathogenesis of CASP/RASP remains unknown but is probably multifactorial. Changes may be related to the biomechanical effects after ALIF that alter the biomechanics at the adjacent levels, leading to increased intradiscal pressure and, therefore, increased load and mobility resulting in accelerated degeneration.<sup>12,13</sup>

Although there is ample evidence to suggest single-level TDA is efficacious compared to arthrodesis, there is less long-term evidence to support the use of multi-level TDA in multi-level DDD.<sup>14–17</sup> Some authors found no significant difference between single-level TDA and multi-level TDA but lacked long-term data.<sup>18</sup> Another surgical option is to treat multi-level DDD with multi-level ALIF, although this can result in higher rates of CASP/RASP and increased nonunion rates.<sup>19</sup> An alternative to multi-level TDA and multi-level ALIF is the hybrid procedure, which combines the motion preservation of TDA and stability from ALIF.<sup>20,21</sup> Other surgical preferences for treating DDD include transforaminal or posterior lumbar interbody fusion, and lateral based techniques. There is considerable evidence on the benefits of TDA<sup>10,15</sup> ALIF<sup>22,23</sup> and hybrid<sup>9,21,24,25</sup> constructs in treating symptomatic DDD with comparison studies showing similar clinical outcomes.<sup>11,17,23,26–28</sup> However, there are few long-term studies comparing these procedures with each other.

Given the lack of long-term data and comparator studies, the aim of this prospective cohort study was to compare the mid- to long-term patient-reported outcome measures (PROMs) between single-level TDA, multi-level TDA, and hybrid constructs for the indication of symptomatic DDD. It is hypothesized that patients who are carefully selected and matched with the appropriate treatment technology will achieve similar clinical outcomes over the mid- to long-term. The clinical decision to offer multilevel LTDA or a hybrid procedure was based on quality of bone, the number of levels, facet arthropathy, the location and type of radiculopathy, and the relationship of the pelvic incidence to actual and theoretical lumbar lordosis. In addition, before 2006 in Australia

LTDA had informal funding arrangements and could be performed at multiple levels using ALIF reimbursement Medicare item numbers. Subsequently, an application for a TDR specific Medicare item number was made and only single level TDA was supported by Medicare. Presently, multiple-level TDA are only available on approval by third party payers or to self-funded patients. Therefore, hybrid procedures became the main treatment option for those patients with symptomatic multi-level DDD. This precipitated the emerging concept of level-specific matching the technology with the pathology.

## MATERIALS AND METHODS

This was a prospective cohort study of 950 patients who were operated for single-level or multi-level DDD between July 1998 and February 2012 with single-level TDA ( $n = 211$ ), multi-level TDA ( $n = 122$ ), or a hybrid construct ( $n = 617$ ). This study was approved by the Bond University Human Research Ethics Committee (0000015881).

All participants suffered chronic low back pain ( $\geq 12$  months) that was unresponsive to active nonoperative treatment, including physical therapy. A diagnosis of single-level or multi-level discogenic low back pain, with or without radicular pain, was established through clinical history, examination, and diagnostic imaging, which included a combination of standing lumbar radiographs, MRI, and provocative discography, with post-discography fine cut CT scan. Because of the high sensitivity and specificity of MRI, it remains an excellent tool for assessing disc morphology, but should be used in conjunction with discography when planning surgical treatment.<sup>29</sup> Discography was used in all patients. The general principles outlined by the International Association for the Study of Pain (IASP)<sup>30</sup> were followed when utilizing discography as an investigative tool. Patients whose discographic results that were nonconcordant were not offered surgery. Electrophysiological studies (needle electromyography and nerve conduction studies) were performed to confirm the presence or absence of radiculopathy, myopathies, peripheral neuropathies, and degenerative neurological conditions. In patients with complex vascular anatomy, a CT angiogram was obtained. Surgery was offered to patients with a precision diagnosis of discogenic pain who had exhausted nonoperative modalities, and where their condition was having significant impact upon their social, recreational, and employment activities.

Surgery was performed via a midline rectus split with a left- or right-sided retroperitoneal approach. Different types of prosthesis were used through the study for each group. Levels undergoing ALIF in the hybrid construct group received PEEK cages, either with integrated cage and screw systems or with a cage and plate with screws combination. Posterior instrumentation was not used. Recombinant human bone morphogenic protein-2 (rhBMP-2), INFUSE Bone Graft (Medtronic Inc, Memphis, TN) was used in all ALIFs. There was no evidence of pseudoarthrosis. At 6 months post-op, this was verified by flexion/extension films and a fine cut three-dimensional reconstructed

**TABLE 1. Characteristics of Surgery Patients (N = 950)**

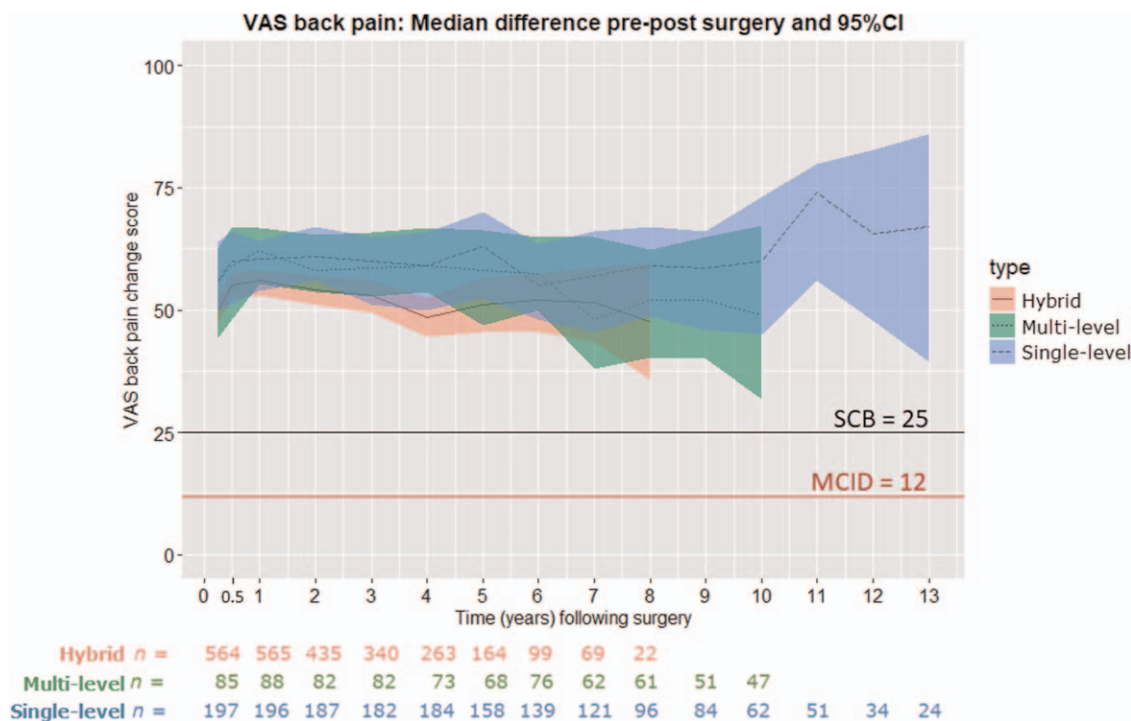
Characteristic	Hybrid Procedure, N = 617	Single-level, N = 211	Multi-level, N = 122
Follow-up, y, median (IQR)	3 (2–5)	8 (6–11)	9 (6–10)
Minimum	2	4	2
Maximum	8	13	10
Sex, n (%)			
Female	238 (38.6)	93 (44.1)	45 (36.9)
Male	379 (61.4)	118 (55.9)	77 (63.1)
Age at time of surgery*, mean (SD)	47.7 (10.8)	42.2 (11.1)	42.0 (8.2)
Levels operated, n (%)	2 level–500 (81.0)	1 level–211 (100)	2 level–120 (98.4)
	3 level–116 (18.8)		3 level–2 (1.6)
	4 level–1 (0.2)		
Baseline pain score, median (IQR)	74.0 (60.0–86.0)	75.0 (55.0–89.0)	80.0 (65.5–91.0)
VAS-Bpain <sup>†</sup>	51.0 (14.0–80.0)	60.0 (17.0–83.0)	54.5 (19.3–81.0)
VAS-L pain			
Baseline disability score, median (IQR)			
ODI <sup>‡</sup>	44.0 (34.0–54.0)	46.0 (34.0–56.0)	48.0 (34.5–60.0)
RMDQ	16.0 (13.0–19.0)	16.0 (12.0–19.0)	18.0 (13.0–20.0)

*IQR indicates Interquartile range; ODI, Oswestry Disability Index; RMDQ, Roland-Morris Disability Questionnaire; SD, standard deviation; VAS-B, Visual Analogue Scale Back; VAS-L, Visual Analogue Scale Leg.*

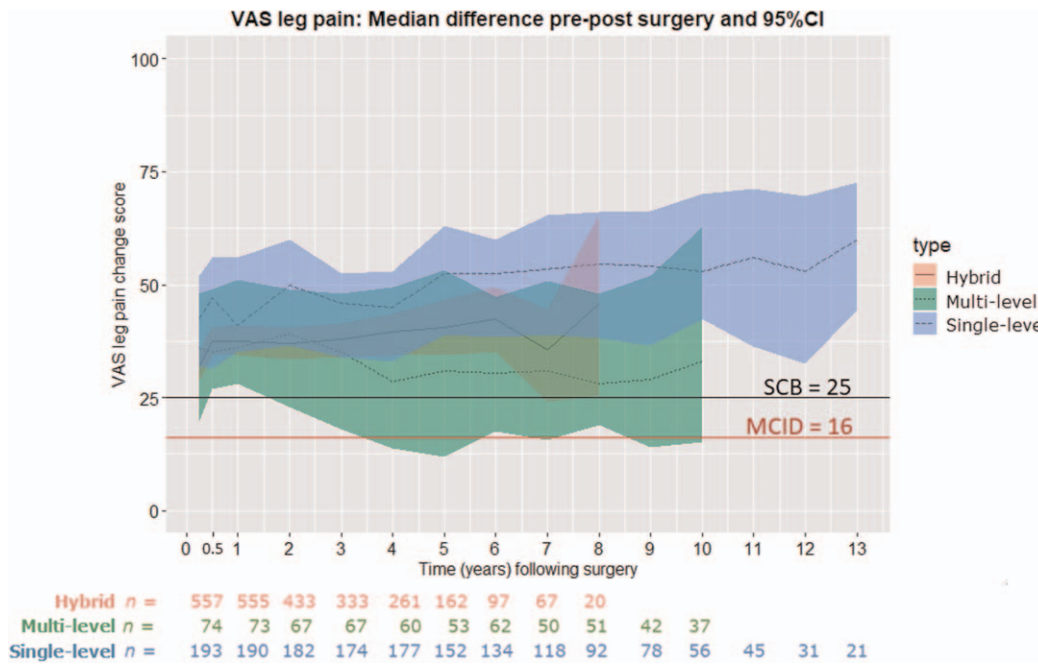
\*Age at time of surgery was significantly different between the hybrid group and each of the other two groups ( $P < 0.001$ ).

<sup>†</sup>Baseline VAS-B pain in the multi-level group was higher compared with the hybrid ( $P = 0.02$ ) and the single-level ( $P = 0.03$ ) groups.

<sup>‡</sup>Baseline ODI was higher in the multi-level group compared with the hybrid group ( $P = 0.04$ ).



**Figure 1.** Median improvement in VAS-B pain score after surgery was statistically ( $P < 0.001$ ) and clinically significant ( $>MCID 12$ ,  $>SCB 25$ ) for each of three surgery procedures. There were no statistically significant differences between surgery types. CI indicates confidence interval; MCID, minimum clinically important difference; SCB, substantial clinical benefit; VAS-B indicates Visual Analogue Scale Back.



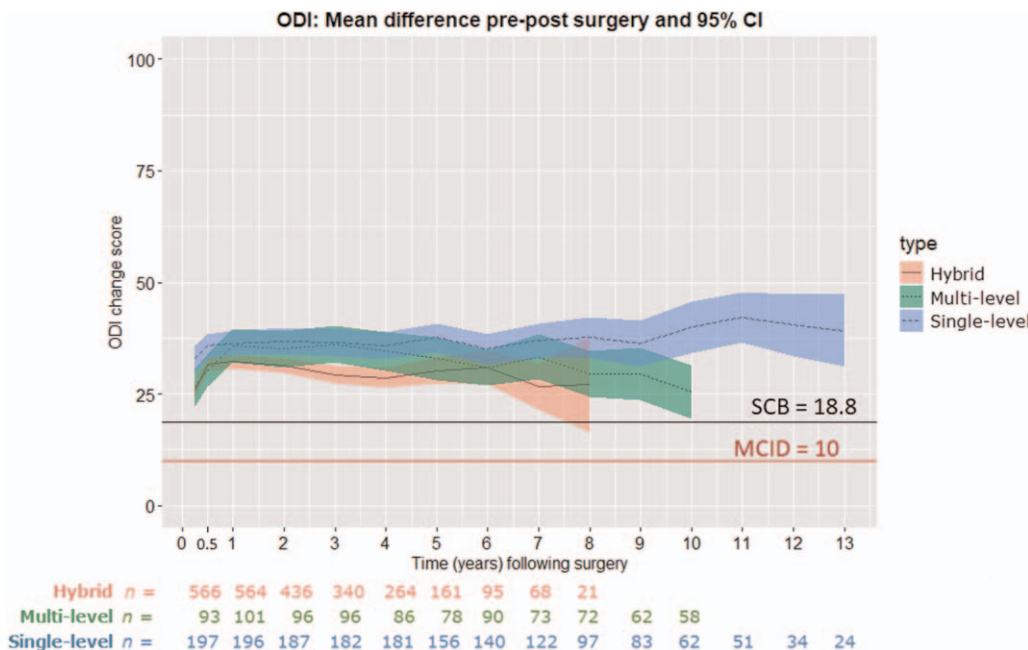
**Figure 2.** Median improvement in VAS-L pain score after surgery was statistically ( $P < 0.001$ ) and clinically significant ( $>MCID 16$ ,  $>SCB 25$ ) for each of three surgery procedures. There were no statistically significant differences between surgery types. CI indicates confidence interval; MCID, minimum clinically important difference; SCB, substantial clinical benefit; VAS-L indicates Visual Analogue Scale Leg.

computed tomography scan. The variation in specific prostheses was because of availability, design evolution, and surgeon preference at the time of surgery.

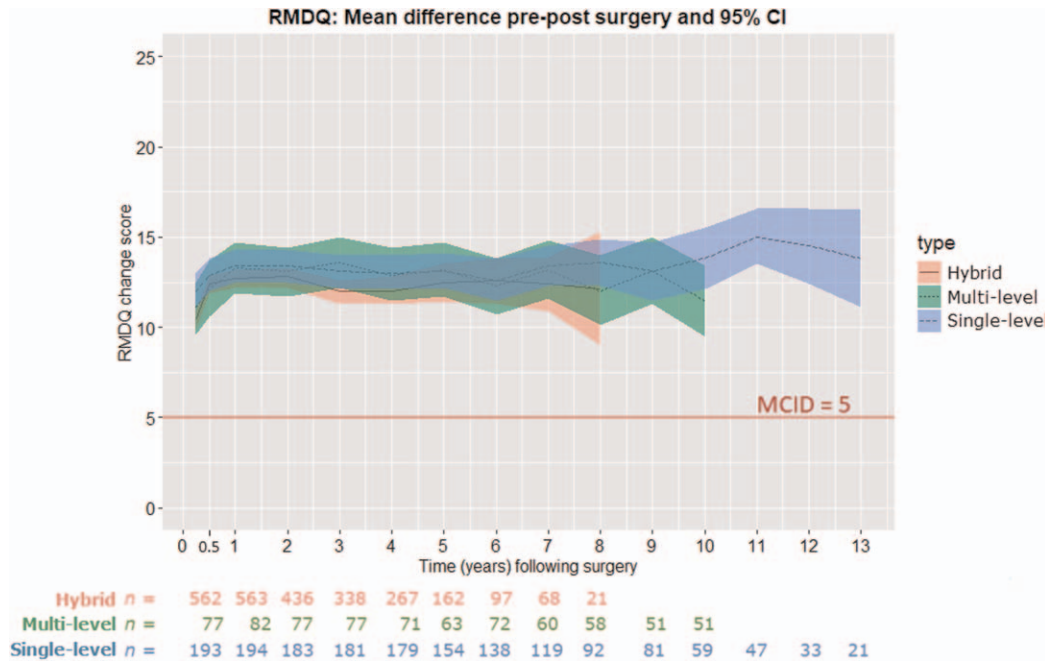
Contraindications to surgery included active infection, tumors, significant scoliosis ( $>20^\circ$ ), severe atherosclerosis

and pregnancy. Obesity and involvement in worker’s compensation or other litigation were regarded as relative contraindications.

PROMs were obtained preoperatively, then postoperatively at 3, 6, and 12 months, and annually thereafter.



**Figure 3.** Mean improvement in ODI score after surgery was statistically ( $P < 0.001$ ) and clinically significant ( $>MCID 10$ ,  $>SCB 18.8$ ) for each of three surgery procedures. The single-level procedure was statistically better than the hybrid procedure by 7.2 points at 3 months, 3 and 4 years ( $P < 0.001$ ). CI, confidence interval; MCID, minimum clinically important difference; ODI indicates Oswestry Disability Index; SCB, substantial clinical benefit.



**Figure 4.** Mean improvement in RMDQ disability score after surgery was statistically ( $P < 0.001$ ) and clinically significant ( $>MCID 5$ ) for each of three surgery procedures. There were no statistically significant differences between surgery types. CI indicates confidence interval; MCID, minimum clinically important difference; RMDQ indicates Roland Morris Disability Questionnaire.

Collected PROMs included Visual Analog Score for back (VAS-B) and leg (VAS-L) were recorded on a 0 to 100 scale, Oswestry Disability Index (ODI), and Roland Morris Disability Questionnaire (RMDQ). Patient satisfaction was

assessed with a four-scale written questionnaire (excellent, good, satisfactory, and poor). Radiographic assessment was performed to confirm movement, alignment, and lack of complication of the construct, for example, subsidence,

**TABLE 2. Results of the Linear Mixed Regression Modelling of Pain Change Scores Over 8 Years**

Parameter	Beta Coefficient	95% CI	P
<b>VAS-B pain</b>			
Intercept	56.5	(54.8, 58.2)	<0.001*
Baseline score (centered on 75)	0.8	(0.80, 0.90)	<0.001*
Age (centered on 45)	-0.1	(-0.21, 0.01)	0.08
Year 2	0.5	(-1.1, 2.1)	0.56
Year 4	-3.1	(-5.0, -1.1)	0.002*
Year 8	-7.4	(-11.0, -3.8)	<0.001*
Single-level	3.2	(0.4, 6.1)	0.03
Multi-level	-0.8	(-4.6, 3.0)	0.69
<b>VAS-L pain</b>			
Intercept	42.6	(41.0, 44.1)	<0.001*
Baseline score (centered on 54)	0.9	(0.86, 0.93)	<0.001*
Age (centered on 45)	-0.2	(-0.3, -0.1)	<0.001*
Year 2	0.6	(-1.0, 2.2)	0.45
Year 4	-0.3	(-2.1, 1.5)	0.71
Year 8	-1.4	(-4.3, 1.4)	0.32
Single-level	-0.1	(-2.7, 2.4)	0.91
Multi-level	-2.0	(-5.6, 1.6)	0.28

\*Statistically significant  $P < 0.0125$ .

The intercept coefficient represents the mean improvement at 6 months for patients in the hybrid group, with baseline pain score (VAS-B: 75; VAS-L: 54) and age 45. A negative coefficient signifies a decrease in improvement.

CI indicates confidence interval; VAS-B indicates Visual Analogue Scale Back; VAS-L, Visual Analogue Scale Leg.

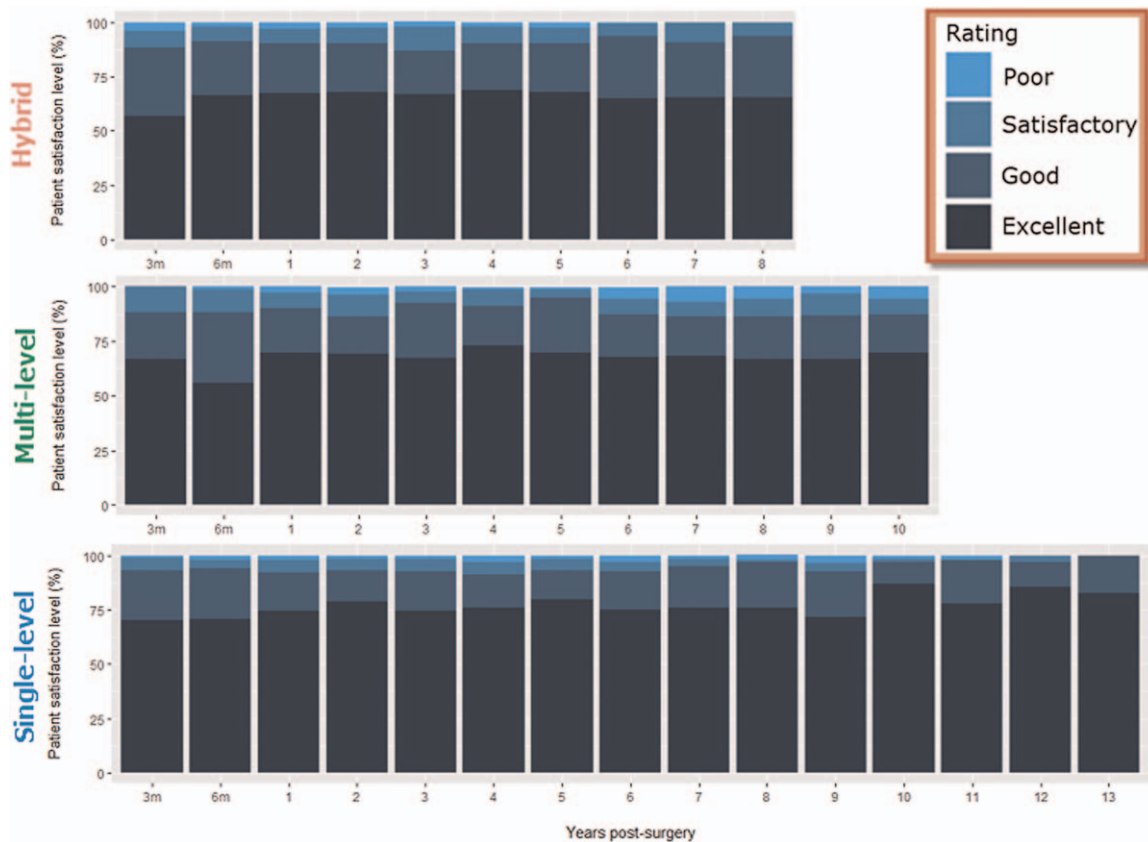
**TABLE 3. Results of the Linear Mixed Regression Modelling of Disability Change Scores Over 8 Years**

Parameter	Beta Coefficient	95% CI	P
<b>ODI</b>			
Intercept	31.3	(30.2, 32.3)	<0.001*
Baseline score (centered on 44)	0.8	(0.75, 0.84)	<0.001*
Age (centered on 45)	-0.2	(-0.23, -0.16)	<0.001*
Year 2	0.9	(-0.1, 1.8)	0.07
Year 4	-0.1	(-1.3, 1.0)	0.80
Year 8	-1.6	(-3.5, 0.4)	0.12
Single-level	2.2	(0.6, 3.9)	0.009*
Multi-level	-2.6	(-4.7, -0.5)	0.02
<b>RMDQ</b>			
Intercept	12.6	(12.3, 13.0)	<0.001*
Baseline score (centered on 16)	0.9	(0.81, 0.90)	<0.001*
Age (centered on 45)	-0.1	(-0.07, -0.03)	<0.001*
Year 2	0.4	(0.1, 0.8)	0.011*
Year 4	0.2	(-0.2, 0.6)	0.31
Year 8	0.1	(-0.5, 0.7)	0.82
Single-level	0.6	(0.1, 1.2)	0.03
Multi-level	-1.1	(-1.9, -0.4)	0.003*

\*Statistically significant  $P < 0.0125$ .

The intercept coefficient represents the mean improvement at 6 months for patients in the hybrid group, with baseline disability score (ODI: 44; RMDQ: 16) and age 45. A negative coefficient signifies a decrease in improvement.

CI indicates confidence interval; ODI indicates Oswestry Disability Index; RMDQ, Roland Morris Disability Questionnaire.



**Figure 5.** Patient Satisfaction Levels Were Consistently High Over the Years For all 3 Surgery Procedures.

**TABLE 4. The Mean Percentage of Surgery Patients Who Achieved MCID and SCB in the First 8 Years Post-surgery**

Surgery type	VAS-B		VAS-L		ODI		RMDQ
	% MCID	% SCB	% MCID	% SCB	% MCID	% SCB	% MCIDe
Hybrid	89.2	80.0	66.5	59.7	86.2	73.0	86.6
Single	85.5	75.4	70.6	64.3	90.3	79.0	89.2
Multiple	92.0	82.7	65.5	58.8	86.9	73.1	88.2

MCID indicates minimum clinically important difference; ODI, Oswestry Disability Index; RMDQ, Roland Morris Disability Questionnaire; SCB, substantial clinical benefit; VAS-B indicates visual analogue scale back; VAS-L, visual analogue scale leg.

heterotopic ossification, and adjacent segment degeneration. Plain radiographs, including erect flexion-extension views, were obtained at 3, 6, and 12 months. Cases that received ALIF, a fine-cut CT scan was obtained at 6 months to verify radiographic union.

### Statistical Analyses

Data were analyzed by an independent university affiliated research team with R statistical software version 3.5.0. Each cohort was compared for VAS-B, VAS-L,<sup>31</sup> ODI<sup>32</sup> and RMDQ disability<sup>33,34</sup> before surgery, at baseline and post-operatively at 3 months, 6 months, 12 months, and yearly thereafter. Change-from-baseline scores for each of these four measures were used as main outcomes. Baseline characteristics were compared using  $\chi^2$  tests for categorical variables and one-way analysis of variance (ANOVA) or Kruskal-Wallis tests depending on the distribution of the continuous variables. The change in scores for the surgery groups were compared using one-way ANOVA or Kruskal-Wallis tests at each time-point, with Bonferroni method used to correct for multiple comparisons. Adjusted analyses were carried out using linear mixed models to assess the effect of surgery type on the outcomes, while controlling for baseline score and age, at specific time points: 6 months, 2 years, 4 years, and 8 years post-surgery. Interactions between time and surgery type were assessed but were not used in the final models as they were not significant. Significance for the adjusted analyses was set at 0.0125 *a priori*.

Graphical representations of median changes in VAS and mean differences in ODI and RMDQ were plotted along with 95% confidence interval (CI) and the corresponding minimum clinically important difference (MCID) and substantial clinical benefit (SCB) for each outcome.<sup>35,36</sup>

### RESULTS

A total of 950 patients underwent one of three procedures: hybrid ( $n = 617$ ), single-level TDA ( $n = 211$ ) or multi-level TDA ( $n = 122$ ). The median follow-up times, with interquartile ranges (IQR), were 3 years (IQR 2–5) for hybrid, 8 years (IQR 6–11) for single, and 9 years (IQR 6–10) for multi-level. In the single-level cohort, 13.2% of patients were lost to follow-up at an average of 58.3 months post-operatively. The majority of these were lost despite 77%

rating their satisfaction as good to excellent at the last point of follow-up. In the multi-level TDA, 19.67% were lost to follow-up at an average of 79.7 months postoperatively. The majority were lost to follow-up despite 58% rating their satisfaction as good to excellent. In the hybrid cohort, 25% of patients were lost to follow-up, the majority of which were lost at the 12- to 24-month stage. The majority were lost to follow-up despite 82.8% rating their satisfaction as good to excellent at final follow-up. The demographic and preoperative patient characteristics are presented in Table 1. All groups were similar for surgical indications (DDD) and surgical approach. Demographics were comparable, except for age where the hybrid group had significantly older patients ( $P < 0.001$ ). Baseline PROMs were comparable, except in the multi-level group where VAS-B (median 80, IQR 65.5–91.0) was statistically higher than the single-level group (median 75.0, IQR 55.0–89.0,  $P = 0.03$ ) and the hybrid group (74.0, 60.0–86.0,  $P = 0.02$ ). The ODI baseline score was also higher in the multi-level group (median 48.0, IQR 34.5–60.0) than the hybrid group (median 44.0, IQR 34.0–54.0,  $P = 0.04$ ). The revision/reoperation rate was for single level 2.8%, multiple level 5.4%, and hybrid 3.7%. The index level and adjacent level revision and reoperation rate were equally divided. All index levels were at the TDA level. There were no reoperations for pseudoarthrosis and all fusions were standalone constructs.

The indications for index level revision spine surgery of the total disc replacements included facet arthropathy, subsidence, and migration of the implant. Most patients underwent RSS by a posterior spinal fusion/decompression for FA and/or neural compression. The median time to revision was 35 months (IQR 9–51), suggesting an element of kinematic mismatch resulting in facet degeneration and time for the pathology to develop and become symptomatic. Most (85%) were at L5-S1 with unconstrained prosthesis.

Figures 1–4 are graphical representations of the change in scores for VAS-B and VAS-L and the ODI and RMDQ outcome measures over time. The relevant MCID and SCB for each outcome measure are included. All groups showed a statistical improvement from base line in pain and function that was well above the corresponding MCID. Unadjusted analyses show that there were no statistically significant differences in the change scores between the surgery groups for VAS-B, VAS-L, and RMDQ. The ODI

improvement in the single-level group compared with the hybrid group reached statistical significance with an average of 7.2 points at 3 months, 3 years, and 4 years post-surgery ( $P < 0.001$ ). However, these differences were not clinically significant.

Table 2 and Table 3 display the PROMs summary statistics for the adjusted analyses using linear mixed models to assess the effect of surgery type on the outcomes, while controlling for baseline score and age. These results show that, after adjusting for baseline score and age, there were no significant differences in change scores between surgery types for VAS-B and VAS-L pain. The ODI improvement for the single-level group was 7.2 points better (95% CI: 0.6–3.9,  $P = 0.009$ ) than in the hybrid group. The RMDQ change was better in the hybrid group than in the multi-level group by 1.1 points (95% CI: 0.4–1.9,  $P = 0.003$ ) at 6 months and a further 0.4 points at 2 years (95% CI: 0.1–0.8,  $P = 0.011$ ).

Figure 5 compares patient satisfaction between the three surgical groups, with consistently high rates of good to excellent patient satisfaction reported and sustained over time.

Table 4 shows the mean percentage of surgery patients who achieved MCID and SCB in the first 8 years post-surgery.

## DISCUSSION

The aim of this prospective cohort study was to compare the mid- to long-term PROMs between single-level TDA, multi-level TDA, and hybrid constructs for the treatment of symptomatic DDD when appropriate methods of diagnosis, patient selection, and technique are followed. Clinically relevant and statistically significant improvements in all PROMs were achieved from baseline measures at all time points postoperatively in all three groups. Comparing the groups showed no statistically significant differences in the changes in score between the groups for VAS-B and VAS-L through unadjusted analyses and linear mixed modeling. However, statistically significant difference in improvements in ODI was found in the single-level group compared to the hybrid group in the early postoperative period ( $\leq 4$  years post-op). Past this time-point, the data clustered together showing similar improvements in outcomes independent of group or whether the diagnosis was single- or multi-level DDD. The results of this cohort study demonstrate that single-level TDA, multi-level TDA, and hybrid constructs are all effective in treating symptomatic DDD, with no clinical difference in PROMs between the groups up to 8 years follow-up.

All three groups had similar demographics, except for age in the hybrid group. The hybrid group was significantly older ( $P < 0.001$ ). This can be explained by the higher incidence of facet joint osteoarthritis in older patients (a contraindication for TDA) and age-related degenerative processes that leads to multi-level pathology.<sup>20</sup> The premise behind the hybrid construct is to address multi-level DDD by matching pathology with

technology, whereby levels with advanced degeneration, facet arthropathy, or segment instability are treated with ALIF while TDA is reserved for levels without significant segmental instability or facet arthropathy.<sup>37</sup> Despite the difference in age between the hybrid group and single- and multi-level TDA, the clinical outcomes for VAS-B, VAS-L, and RMDQ were similar.

Long-term studies for single-level TDA show favorable outcomes compared to fusion surgery, with comparable outcomes for back and leg pain and reduction of CASP/RASP.<sup>38,39</sup> However, with multi-level DDD, there were concerns regarding rates of failure in multi-level TDA because of the difficulty in positioning and balancing of multiple TDAs, particularly in the setting of unconstrained and semi-constrained prostheses, as well as concerns regarding facet arthritis and segmental instability.<sup>40–42</sup> Other considerations for multi-level TDA include spinopelvic parameters, operative level, bone density, and presence of other comorbidities.<sup>43</sup> Zigler *et al* compared single-level TDA to two-level TDA and found that two-level TDA is just as effective as one-level in treating symptomatic DDD in appropriately selected patients.<sup>44</sup> Schätz *et al* also found adequate initial safety and effectiveness at the two year time point for both single and multi-level TDA but cautioned that a larger sample size and extended follow-up were necessary.<sup>27</sup> Yue *et al* presented 5-year data comparing single-level and two-level arthroplasty.<sup>45</sup> Similarly, they did not find statistical differences in all clinical and radiological outcomes, but also recommended longer follow-up to confirm.

The findings previously reported are consistent with the results from our study, where both the single- and multi-level groups showed statistically significant improvements in all PROMs, with no statistically significant difference between the groups up to 8 years postoperatively. This study highlights the sustained improvements and minimal difference in mid- to long-term data on the efficacy of single- and multi-level TDA, when matching the technology to pathology. The authors recognize the importance of coronal and sagittal balance, as well as use of constrained and unconstrained TDA, which is highlighted in a previous study.<sup>43</sup>

Only the single-level TDA group showed significantly better ODI improvement over the hybrid group at 3 months, 3 years, and 4 years post-surgery ( $P < 0.001$ ). The difference was not clinically significant and, after 4 years post-surgery, the difference became insignificant. The authors believe this relates to the greater complexity in pathology and surgery in the hybrid group compared to the single-level group. The hybrid group was older and the surgery involves two or more levels, which results in longer recovery time. However, given the PROMs eventually clustered together over time, the authors suggest that clinical improvements continue as the patient gains strength and confidence, until a sustained steady state is reached. To our knowledge, this is the first cohort study comparing the mid- to long-term PROMs of single-level TDA with hybrid constructs.

There are a limited number of published comparative studies that compare hybrid constructs with either multi-



level TDA or multi-level ALIF.<sup>20,21</sup> Andrieu *et al* found no statistical difference in pain relief ( $-3.9$  points *vs.*  $-3.5$  points for VAS) or reduction in ODI ( $-29.5\%$  *vs.*  $-27.0\%$ ) between two-level TDA and two-level hybrid constructs, respectively, after 2 years of follow-up.<sup>20</sup> Similarly, in this study, the PROMs comparing the hybrid group and multi-level TDA were not statistically significant. However, after a linear mixed model was applied, the RMDQ in the hybrid group was better than the multi-level TDA group at 6 months by 1.1 points (95% CI: 0.4–1.9,  $P=0.003$ ) and 1.5 point at 2 years (95% CI: 0.1–0.8,  $P=0.011$ ). The authors cannot account for the initial difference in the early postoperative period in RMDQ but suggest this may relate to improved sagittal and coronal correction made possible with the ALIF component of hybrid procedures. Use of hyper lordotic ALIF cages enables greater sagittal and coronal realignment and disc height restoration than TDA prostheses alone.

Chen *et al* compared the hybrid construct against multi-level ALIF and found the hybrid construct was a valid and viable alternative, with at least equal if not better clinical outcomes in terms of survivorship, back pain, and disability scores.<sup>21</sup> This present study similarly demonstrates equivalent improvement in PROMs in all groups.

Limitations of this study include all cases being performed by a single surgeon at a single institution, which affects the generalizability of the results.<sup>46,47</sup> Another limitation is the lack of a control group in the context of an RCT protocol and is inherent in observational studies.

Finally, there is the question of the ideal PROMs to use in a spinal surgery study. Some PROMs scoring systems lack sensitivity in measurement of subtleties of function to accurately define utility. Furthermore, PROMs do not evaluate important economic outcomes, which are increasingly being recognized as relevant to healthcare payers and the sustainability of a treatment. A systematic review by Ishaque *et al* suggests PROMs are an effective intervention in evidence based medicine, through improving communication and decision-making process about the potential benefits of surgery.<sup>48</sup> Finally, there is also the issue of reporting bias inherent in PROMs. Despite these limitations, the value of PROMs in comparing the results of treatment modalities is well established.

## CONCLUSION

This study shows clinically and statistically significant reduction in back and leg pain, as well as self-rated disability outcomes in the single-level TDA, multi-level TDA, and hybrid construct groups, which was sustained for at least 8 years. There was no clinically significant difference in PROMs between all the groups up to 8 years postoperatively. This suggests that single-level TDA, multi-level TDA, and hybrid constructs are all viable and suitable options for patients suffering chronic back and leg pain, provided a precision diagnosis has been made and it is matched by an appropriate surgical treatment. To our knowledge, this represents the largest cohort and longest follow-up

comparing PROMs of single-, multi-level TDA, and hybrid constructs in the literature.

## ➤ Key Points

- ❑ Single-level TDA, multi-level TDA, and hybrid constructs are viable options for treating symptomatic DDD.
- ❑ Statistically and clinically significant benefits were achieved with all surgery groups, which were clustered and sustained up to 8 years post-surgery.
- ❑ There is no clinically significant difference in PROMs between single-level TDA, multi-level TDA, and hybrid constructs.

## References

1. Andersson GB. Epidemiological features of chronic low-back pain. *Lancet (London, England)* 1999;354:581–5.
2. Fatoye F, Gebrye T, Odeyemi I. Real-world incidence and prevalence of low back pain using routinely collected data. *Rheumatol Int* 2019;39:619–26.
3. Bogduk N. Management of chronic low back pain. *Med J Aust* 2004;180:79–83.
4. Sehgal N, Fortin JD. Review article internal disc disruption and low back pain. *Management* 2000;3:143–57.
5. Lee YC, Zotti MGT, Osti OL. Operative management of lumbar degenerative disc disease. *Asian Spine J* 2016;10:801–19.
6. Cho SC, Ferrante MA, Levin KH, et al. Utility of electrodiagnostic testing in evaluating patients with lumbosacral radiculopathy: an evidence-based review. *Muscle Nerve* 2010;42:276–82.
7. Zuo J, Joseph GB, Li X, et al. In vivo intervertebral disc characterization using magnetic resonance spectroscopy and T1ρ imaging: association with discography and Oswestry Disability Index and Short Form-36 Health Survey. *Spine (Phila Pa 1976)* 2012;37:214–221.
8. Malham GM, Parker RM, Ballok ZE, et al. Bone scans are reliable for the identification of lumbar disk and facet pathology. *Glob spine J* 2015;5:23–30.
9. Scott-Young M, McEntee L, Schram B, et al. Concurrent use of lumbar total disc arthroplasty and anterior lumbar interbody fusion. *Spine (Phila Pa 1976)* 2018;43:E75–81.
10. Van Den Eerenbeemt KD, Ostelo RW, Van Royen BJ, et al. Total disc replacement surgery for symptomatic degenerative lumbar disc disease: a systematic review of the literature. *Eur Spine J* 2010;19:1262–80.
11. Wei J, Song Y, Sun L, et al. Comparison of artificial total disc replacement versus fusion for lumbar degenerative disc disease: a meta-analysis of randomized controlled trials. *Int Orthop* 2013;37:1315–25.
12. Zhang C, Berven SH, Fortin M, et al. Adjacent segment degeneration versus disease after lumbar spine fusion for degenerative pathology: a systematic review with meta-analysis of the literature. *Clin spine Surg* 2016;29:21–9.
13. Epstein NE. Adjacent level disease following lumbar spine surgery: A review. *Surg Neurol Int* 2015;6 (suppl 24):S591–9.
14. Blumenthal S, McAfee PC, Guyer RD, et al. A prospective, randomized, multicenter Food and Drug Administration Investigational Device Exemptions study of lumbar total disc replacement with the CHARITETM artificial disc versus lumbar fusion—part I: evaluation of clinical outcomes. *Spine (Phila Pa 1976)* 2005;30:1565–1575.
15. Zigler J, Delamarter R, Spivak JM, et al. Results of the prospective, randomized, multicenter food and drug administration investigational device exemption study of the ProDisc®-L total disc

- replacement versus circumferential fusion for the treatment of 1-level degenerative disc disease. *Spine (Phila Pa 1976)* 2007;32:1155–1162.
16. Gornet MF, Dryer RF, Pelozo JH, et al. Anterior lumbar interbody fusion: five-year outcomes for patients in the maverick® disc ide study. *Spine J* 2010;10:S64; doi:10.1016/j.spinee.2010.07.174.
  17. Zigler J, Ferko N, Cameron C, et al. Comparison of therapies in lumbar degenerative disc disease: a network meta-analysis of randomized controlled trials. *J Comp Eff Res* 2018;7:233–46.
  18. Hannibal M, Thomas DJ, Low J, et al. ProDisc-L total disc replacement: a comparison of 1-level versus 2-level arthroplasty patients with a minimum 2-year follow-up. *Spine (Phila Pa 1976)* 2007;32:2322–6.
  19. Rölinghoff M, Schlüter-Brust K, Groos D, et al. Mid-range outcomes in 64 consecutive cases of multilevel fusion for degenerative diseases of the lumbar spine. *Orthop Rev (Pavia)* 2010;2:e3.
  20. Andrieu K, Allain J, Longis P-M, et al. Comparison between total disc replacement and hybrid construct at two lumbar levels with minimum follow-up of two years. *Orthop Traumatol Surg Res* 2017;103:39–43.
  21. Chen B, Akpolat YT, Williams P, et al. Survivorship and clinical outcomes after multi-level anterior lumbar reconstruction with stand-alone anterior lumbar interbody fusion or hybrid construct. *J Clin Neurosci* 2016;28:7–11.
  22. Giang G, Mobbs R, Phan S, et al. Evaluating outcomes of stand-alone anterior lumbar interbody fusion: a systematic review. *World Neurosurg* 2017;104:259–71.
  23. Shah M, Kolb B, Yilmaz E, et al. Comparison of lumbar laminectomy alone, lumbar laminectomy and fusion, stand-alone anterior lumbar interbody fusion, and stand-alone lateral lumbar interbody fusion for treatment of lumbar spinal stenosis: a review of the literature. *Cureus* 2019;11:e5691.
  24. Aunoble S, Meyrat R, Al Sawad Y, et al. Hybrid construct for two levels disc disease in lumbar spine. *Eur Spine J* 2010;19:290–6.
  25. Hoff EK, Strube P, Pumberger M, et al. ALIF and total disc replacement versus 2-level circumferential fusion with TLIF: a prospective, randomized, clinical and radiological trial. *Eur Spine J* 2016;25:1558–66.
  26. Nie H, Chen G, Wang X, et al. Comparison of Total Disc Replacement with lumbar fusion: a meta-analysis of randomized controlled trials. *J Coll Physicians Surg Pak* 2015;25:60–7.
  27. Schätz C, Ritter-Lang K, Gössel L, et al. Comparison of single-level and multiple-level outcomes of total disc arthroplasty: 24-month results. *Int J spine Surg* 2015;9:14.
  28. Zigler J, Gornet MF, Ferko N, et al. Comparison of lumbar total disc replacement with surgical spinal fusion for the treatment of single-level degenerative disc disease: a meta-analysis of 5-year outcomes from randomized controlled trials. *Glob spine J* 2018;8:413–423.
  29. Manchikanti L, Soin A, Benyamin RM, et al. An update of the systematic appraisal of the accuracy and utility of discography in chronic spinal pain. *Pain Physician* 2018;21:91–110.
  30. Spine Intervention Society. Safety Practices for Interventional Pain Procedures. 2nd ed San Fr CA International Spine Interv Soc 2013–Updat 2019 retrieved from [https://cdn.ymaws.com/www.spineintervention.org/resource/resmgr/patient\\_safety/safety\\_practices/safetypractices\\_module5\\_prov.pdf](https://cdn.ymaws.com/www.spineintervention.org/resource/resmgr/patient_safety/safety_practices/safetypractices_module5_prov.pdf), accessed 10 Sept 2. 2019:1–5. [https://www.spineintervention.org/global\\_engine/download.aspx?fileid=2D659E35-0E4C-4337-A277-EF4B7E59E364](https://www.spineintervention.org/global_engine/download.aspx?fileid=2D659E35-0E4C-4337-A277-EF4B7E59E364).
  31. Hjerstad MJ, Fayers PM, Haugen DF, et al. Studies comparing numerical rating scales, verbal rating scales, and visual analogue scales for assessment of pain intensity in adults: a systematic literature review. *J Pain Symptom Manage* 2011;41:1073–93.
  32. Fairbank JCT, Pynsent PB. The Oswestry disability index. *Spine (Phila Pa 1976)* 2000;25:2940–52.
  33. Stratford PW, Binkley JM, Riddle DL, et al. Sensitivity to change of the Roland-Morris Back Pain Questionnaire: Part 1. *Phys Ther* 1998;78:1186–96.
  34. Roland M, Fairbank J. The Roland-Morris disability questionnaire and the Oswestry disability questionnaire. *Spine (Phila Pa 1976)* 2000;25:3115–24.
  35. Copay AG, Glassman SD, Subach BR, et al. Minimum clinically important difference in lumbar spine surgery patients: a choice of methods using the Oswestry Disability Index, Medical Outcomes Study questionnaire Short Form 36, and Pain Scales. *Spine J* 2008;8:968–74.
  36. Glassman SD, Copay AG, Berven SH, et al. Defining substantial clinical benefit following lumbar spine arthrodesis. *J Bone Jt Surg - Ser A* 2008;90:1839–47.
  37. Rasouli A, Delamarter RB, Kanim LEA, et al. 158 The hybrid ADR: long-term analysis of sagittal motion and clinical outcome after concurrent ADR+ALIF (minimum 2-year follow-up) spine j 2009;9:83S.
  38. Othman YA, Verma R, Qureshi SA. Artificial disc replacement in spine surgery. *Ann Transl Med* 2019;7 (suppl 5):S170.
  39. Lu Sbao, Hai Y, Kong C, et al. An 11-year minimum follow-up of the Charite III lumbar disc replacement for the treatment of symptomatic degenerative disc disease. *Eur Spine J* 2015;24:2056–64.
  40. Ching AC, Birkenmaier C, Hart RA. Short segment coronal plane deformity after two-level lumbar total disc replacement. *Spine (Phila Pa 1976)* 2010;35:44–50.
  41. van Ooij A, Oner F, Verbout A. Complications of artificial disc replacement: a report of 27 patients with the SB Charite disc. *J Spinal Disord Tech* 2003;16:369–83.
  42. Siepe CJ, Mayer HM, Heinz-Leisenheimer M, et al. Total lumbar disc replacement: different results for different levels. *Spine (Phila Pa 1976)* 2007;32:782–90.
  43. Scott-Young M, McEntee L, Zotti M, et al. Patient reported outcome measures following multi-level lumbar total disc arthroplasty for the treatment of multi-level degenerative disc disease. *Spine (Phila Pa 1976)* 2019;45:18–25.
  44. Zigler JE, Ohnmeiss DD. Comparison of 2-level versus 1-level total disc replacement: results from a prospective FDA-regulated trial. *SAS J* 2008;2:140–4.
  45. Yue J, Zhang K, Bai HX, et al. A comparison of patients who have undergone 1-Level versus 2-Level ProDisc arthroplasty: a prospective study with minimum of 5-year follow-up. *Spine (Phila Pa 1976)* 2013;38:1194–8.
  46. Regan JJ, McAfee PC, Blumenthal SL, et al. Evaluation of surgical volume and the early experience with lumbar total disc replacement as part of the investigational device exemption study of the Charité Artificial Disc. *Spine (Phila Pa 1976)* 2006;31:2270–6.
  47. Birkmeyer JD, Stukel TA, Siewers AE, et al. Surgeon volume and operative mortality in the United States. *N Engl J Med* 2003;349:2117–2127.
  48. Ishaque S, Karnon J, Chen G, et al. A systematic review of randomised controlled trials evaluating the use of patient-reported outcome measures (PROMs). *Qual Life Res* 2019;28:567–92.