Peer Review File
Article Information: https://dx.doi.org/10.21037/jss-22-36

Reviewer A

Comment 1: Thank you for sharing your experience with the scientific community. This is a very nice series of 60 patients that offers tangible lessons to others about patient selection, complications, and anticipated outcomes. I understand the drop-off of patient for long-term follow up, but would have appreciated a 2y time-point as well, if possible. Additionally, a simple graph or table depicting the time-point for last follow up (4.3y mean is good, but 1-10y is too wide a spectrum) - how many patients made it past 5y, etc.?
Reply 1: We have added in a 2-yr timepoint. A total of 37 patients were available for final follow-up, with 36/37 at 2-years and 13/37 past 5 years. Please see Figure 2 for the bar graph depicting the time-points at last follow-up.
Changes in the text: *Abstract – Results*: Thirty-seven (62%) were followed for a mean of 4.3 (1-10) years with 36/37 reviewed at a minimum of 2-years and 13/37 followed for over 5-years.
: *Manuscript - Results*: Most patients (n=42, 70%) had a hybrid procedure, 16 (27%) underwent standalone LTDR, and 2 (3%) a three-level procedure. Twenty-three (38%) patients were lost to follow-up. Thirty-seven (62%) were followed for a mean of 4.3 (1-10) years with 36/37 reviewed at a minimum of 2-years and 13/37 followed for over 5-years. (*Table 1, Figure 2*).

Comment 2: I find table 2 and 3 repetitive.
Reply 2: Thank you, we have combined the old Tables 2 and 3 into a new Table 2.
Changes in the text: See Table section with new Table 2.

Comment 3: Why only do radiographic final follow up - adjacent segment degeneration is best evaluated by MRI. 8/37 figure would be quite different.
Reply 3: Radiographic follow-up was undertaken with flexion-extension x-rays to assess position and motion of the M6-L prosthesis. Adjacent segment degeneration is a radiographic not clinical presentation. We had no symptomatic adjacent segment disease that would have warranted an MRI for consideration of further surgical intervention. We have added that adjacent segment degeneration is best evaluated by MRI in the Limitations of our study with a reference.
Changes in the text: *Manuscript- Results-Radiographic findings:* Eight patients demonstrated adjacent segment degeneration, but no patient had symptomatic ASD requiring reoperation.
: Manuscript - Discussion – Limitations: This series conducted dynamic radiographic analysis at final follow-up, ASD is best assessed with MRI[24]. This restricts the applicability of our ASD findings, and we suggest future studies utilize MRI at follow-up.

Comment 4: The overall rate of morbidity is 13/60 patients - 22% is reported honestly but deserves a mention in the body of the results section
Reply 4: We have added the overall rate of morbidity in 13/37 patients -22% into the results section.
Changes in the text: *Complications and reoperations:* The overall rate of complications was 22% (13/37) in this series.

Reviewer B

Comment 1: This is a retrospective study to evaluate the clinical and radiographic outcomes of patients undergoing LTDR with M6-L. Some comments are shown below.
Abstract: in the result section, 16 patients underwent standalone LTDR and 42 patients with a hybrid procedure. However, there are 60 patients in this study. How about the remaining 2 patients?
Reply 1: We have amended the *Abstract-Results* to include the 2 patients who had a 3-level procedure.

We have amended the *Manuscript-Results* to include the 2 patients who had a 3-level procedure. Also shown in Table 1.

Changes in the text: *Abstract-Results:* Sixteen (27%) underwent standalone LTDR, 42 (70%) a hybrid procedure, and 2 (3%) a 3-level procedure.

Manuscript-Results: Most patients (n=42, 70%) had a hybrid procedure, 16 (27%) underwent standalone LTDR, and 2 (3%) a 3-level procedure. Twenty-three (38%) patients were lost to follow-up.

Comment 2: # Statistical analysis: for numerical data, the authors need to perform a normal distribution first. Data should be displayed as mean with standard deviation when normally distributed and as median with IQR or range when not normally distributed. The present results show inconsistencies including SD, range, and 95% CI. For categorical data, they should be illustrated as number with percentage.
Reply 2: Thank you for this expert correction. We have amended Table 1 to fit your suggestions with 6-month follow-up being normally distributed showing mean, median, IQR and SD. Not normally distributed data including follow-up time in years was displayed as median and IQR. Categorical data is illustrated as a n,%.
Changes in the text: Please see revised Table 1.

Comment 3: # There are too many tables in this study, and some tables can be combined. For example, table 2 and table 3, table 4 and table 5, and table 6 and table 7. The same goes for figures.
Reply 3: Thank you. Based on your expert advice, we have reduced the number of Tables from 8 to now only 3 Tables. We have combined the old Tables 2 and 3 into a new Table 2. Also the old Tables 4 and 5 were deleted as the data is described in the text. The old Table 6 is now Table 3. The old Table 7 and Table 8 were deleted as the data is described in the text.

We have deleted the old Figures 2 (VAS back), 3 (VAS leg), 4 (ODI) and 5 (SF-12). All this data is in the new Table 2. We have deleted Figure 6 (NASS PSI scores) as described in the text
Changes in the text: See Table section with new Tables 1,2 and 3. See Figure section with new Figures 1,2,3,4 and 5.

Results-Patient Reported Outcomes: Paragraph 2 - The patients available for additional follow-up were very satisfied with the operation. Using the NASS PSI, 31 patients (83.8%) reported a score of 1 indicating that the procedure met their expectations. Four patients indicated a score of two and two individual patients recorded a score of 3 and 4.

Discussion paragraph 6 - We found lumbar arthroplasty with M6-L resulted in high patient satisfaction consistent with mid- to long-term results with the Prodisc-

L[17]. The majority (94.6%) of our patients responded as NASS PSI of 1 or 2, indicating they would undergo the same operation again for the same result. Two patients responded with a higher NASS score which was more attributable to the access procedure. One patient who reported a NASS PSI of 3 suffered a deep wound infection and another patient who reported a score of 4 suffers from leg pain and substance abuse.

Comment 4: In addition, line 122 indicated that the VAS showed "noticeable improvements …". If the authors say there is a clear improvement, the results should be tested; however, the tables or figures did not show any p-values. These data are best analyzed by a professional statistician.
Reply 4: We have amended *Results-Patient Reported Outcomes*- editing to fit your recommendations after review again by our statistician.
Changes in the text: Preoperative mean VAS back, VAS leg, ODI, and SF-12 (physical / mental) showed improvements postoperatively at 6 weeks and again at 6 months postoperative. PROMs showed statistically significant improvements ($p<0.05$) from baseline to last follow-up (*Table 2*).

Comment 5:# For the radiographic outcomes, 50% of patients were lost to follow-up. Though the authors say the rate is acceptable, there still existed some bias need to be considered. How is the difference between follow-up and unfollowed patients, such as age, gender, DDD severity at the beginning?
Reply 5: There was no significant difference in PROMs between patients who had final follow-up x-rays and those who declined x-rays.

There was no difference in age, sex and PROMs between followed-up and unfollowed patients.

Changes in the text: *Results-Radiographic findings*- Thirty (50%) patients consented for a final radiographic follow-up. There was no difference in PROMs at 6-weeks and 6-months between patients that obtained final follow-up x-rays and those who did not (all $p>0.05$).

Results-Patient Reported Outcomes - There was no difference in age, sex and PROMs between followed-up (n=37) and unfollowed patients.

Reviewer C

There are multiple major issues with this study that prevent it from being acceptable for publication

Comment 1: Loss of follow up of almost 40% of the initial cohort.
Reply 1: We acknowledge this loss of follow-up but our study is real and honest in clinical practice. In *Discussion* paragraph 8 we state - We had a radiographic loss to follow-up of 50%. Patients declined dynamic radiographs secondary to COVID-19 restrictions, busy lifestyle, and risks of radiation exposure. These patients reported a positive improvement in PROMs with no significant difference observed between patients who obtained radiographs, and those who did not.
Changes in the text: *Discussion-Limitations:* This study's extended follow-up period showed a radiographic loss to follow-up of 50% and final PROMs loss to follow-up of 38%. We found no difference in PROMs between patients that consented to final follow-up x-rays and those who declined x-rays. Ideally further studies with a larger sample size should be considered. However, research suggests that a loss to follow-up of up to 60% can be acceptable[25].

Comment 2: The data presented seems inconsistent and tailored to paint a forced favorable picture on the M6-L.

Reply 2: We wanted to present an honest long-term follow-up of our patients with the M6-L device with outcomes and complications. We report in *Discussion* paragraph 5 that- "The M6-L demonstrated long-term effectiveness and durability. We found no device failures and no evidence of the presence of mid-term wear induced osteolysis. There is only one reported case of M6-L device failure in the literature[14]. No patient developed ASD that required surgical intervention".

Also, we report in *Discussion* paragraph 6 that – "We found lumbar arthroplasty with M6-L resulted in high patient satisfaction consistent with mid- to long-term results with the Prodisc-L[17]. The majority (94.6%) of our patients responded as NASS PSI of 1 or 2, indicating they would undergo the same operation again for the same result".

But we have deleted the last sentence in the Discussion-Limitations that our study "still supports the M6-L as an effective long-term LDTR device".

Changes in the text: *Discussion-Limitations:* last sentence deleted: ~~Therefore, whilst the loss to follow-up in our study is not ideal, it still supports the M6-L as an effective long-term LDTR device.~~

Comment 3: The study population is very heterogeneous with patients having TDR + ALIF and TDR alone being compared head to head, these groups should be looked at separately as the addition of an ALIF will significantly alter the biomechanical behavior of the adjacent levels, making the conclusion of the current calculation confusing to interpret.

Reply 3: We agree and have added a third sentence to paragraph 1 in *Discussion-Limitations.*

Changes in the text: We note the heterogeneity of this study and that inclusion of patients with hybrid constructs may limit the generalizability of our study. However, including these patients is important given the concern for long-term follow-up for device related failures.
