## Supplementary Table 1  Analysis of intra-observer reliability

<table>
<thead>
<tr>
<th></th>
<th>Evaluation 1</th>
<th>Evaluation 2</th>
<th>Difference</th>
<th>p-value</th>
<th>Correlation coefficient (p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DD score</td>
<td>2.7 (± 1.0)</td>
<td>2.8 (± 1.0)</td>
<td>-0.12 (± 0.33)</td>
<td>0.083</td>
<td>0.886 (&lt;0.001)</td>
</tr>
<tr>
<td>EP score</td>
<td>1.8 (± 1.7)</td>
<td>1.9 (± 1.7)</td>
<td>-0.10 (± 0.30)</td>
<td><strong>0.025</strong></td>
<td>0.890 (&lt;0.001)</td>
</tr>
<tr>
<td>FD score</td>
<td>1.6 (± 0.9)</td>
<td>1.6 (± 0.9)</td>
<td>-0.04 (± 0.28)</td>
<td>0.317</td>
<td>0.965 (&lt;0.001)</td>
</tr>
<tr>
<td>FS score</td>
<td>0.4 (± 0.7)</td>
<td>0.5 (± 0.8)</td>
<td>-0.10 (± 0.42)</td>
<td>0.096</td>
<td>0.790 (&lt;0.001)</td>
</tr>
</tbody>
</table>

1 Mean (± SD)  
2 Statistical significance of difference between evaluation 1 and evaluation 2, Wilcoxon signed-rank test  
3 Spearman correlation coefficient  
DD score – Disc degeneration score, EP score – Endplate score, FD score – Facet degeneration score, FS score – Foraminal stenosis score
### Supplementary Table 2  Analysis of inter-observer reliability

<table>
<thead>
<tr>
<th></th>
<th>Evaluation 1</th>
<th>Evaluation 2</th>
<th>Difference</th>
<th>p-value</th>
<th>Correlation coefficient (p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DD score</td>
<td>2.7 (± 1.0)</td>
<td>3.0 (± 0.7)</td>
<td>-0.28 (± 0.61)</td>
<td><strong>0.035</strong></td>
<td>0.673 (&lt;0.001)</td>
</tr>
<tr>
<td>EP score</td>
<td>1.8 (± 1.7)</td>
<td>2.0 (± 1.5)</td>
<td>-0.16 (± 0.58)</td>
<td>0.059</td>
<td>0.732 (&lt;0.001)</td>
</tr>
<tr>
<td>FD score</td>
<td>1.6 (± 0.9)</td>
<td>1.0 (± 0.7)</td>
<td>0.62 (± 0.60)</td>
<td>&lt;0.001</td>
<td>0.771 (&lt;0.001)</td>
</tr>
<tr>
<td>FS score</td>
<td>0.4 (± 0.7)</td>
<td>0.8 (± 0.9)</td>
<td>-0.40 (± 0.53)</td>
<td>&lt;0.001</td>
<td>0.747 (&lt;0.001)</td>
</tr>
</tbody>
</table>

1 Mean (± SD)
2 Statistical significance of difference between observer 1 and observer 2, Wilcoxon signed-rank test
3 Spearman correlation coefficient
DD score – Disc degeneration score, EP score – Endplate score, FD score – Facet degeneration score, FS score – Foraminal stenosis score
**Supplemental Figure 1** Intra-observer reliability evaluated for DD score using Bland-Altman plots
Supplemental Figure 2  Intra-observer reliability evaluated for EP score using Bland-Altman plots
Supplemental Figure 3 Intra-observer reliability evaluated for FD score using Bland-Altman plots.

Bland-Altman plot
intra-observer reliability for FD score

Supplemental Figure 3 Intra-observer reliability evaluated for FD score using Bland-Altman plots.
Supplemental Figure 4 Intra-observer reliability evaluated for FS score using Bland-Altman plots
Supplemental Figure 5 Inter-observer reliability evaluated for DD score using Bland-Altman plots
Supplemental Figure 6 Inter-observer reliability evaluated for EP score using Bland-Altman plots
Supplemental Figure 7 Inter-observer reliability evaluated for FD score using Bland-Altman plots

Bland-Altman plot
inter-observer reliability for FD score

average of two evaluations

+ 1.96*SD

MEAN

- 1.96*SD

difference between two evaluations
**Supplemental Figure 8** Inter-observer reliability evaluated for FS score using Bland-Altman plots

Bland-Altman plot
inter-observer reliability for FS score

The figure shows a Bland-Altman plot for inter-observer reliability in evaluating FS score. The plot displays the difference between two evaluations against the average of the two evaluations. The mean difference is indicated by the horizontal line, and the upper and lower limits of agreement are shown by the dashed lines, which are at ±1.96 times the standard deviation (SD) of the differences.