

## 1 **Radiographic evaluation of lumbar intervertebral disc height index: an intra and inter-** 2 **rater agreement and reliability study**

3 Xiaolong Chen, PhD<sup>1,2</sup> \*, Stone Sima<sup>3</sup>, Harvinder S. Sandhu, MD<sup>4</sup>, Jeff Kuan<sup>5</sup>, Ashish D. Diwan, PhD<sup>1,2</sup>

4 <sup>1</sup> Spine Labs, St. George & Sutherland Clinical School, University of New South Wales, New South Wales,  
5 Australia

6 <sup>2</sup> Spine Service, Department of Orthopaedic Surgery, St. George & Sutherland Clinical School, University of  
7 New South Wales, New South Wales, Australia

8 <sup>3</sup> 3<sup>rd</sup> Year Student of Bachelor of Medical Studies/Doctor of Medicine at University of New South Wales, NSW,  
9 Australia

10 <sup>4</sup> Spinal Surgical Service, Hospital for Special Surgery, Weill Medical College of Cornell University, New  
11 York, USA

12 <sup>5</sup> St. George MRI, Healthcare Imaging, Kirk Place, Kogarah, NSW, Australia

13 **\*Corresponding author:** Xiaolong Chen, PhD

14 Level 3, WR Pitney Building, St. George & Sutherland Clinical School

15 The University of New South Wales, Kogarah, Sydney, NSW 2217

16 Phone: + 61 414 825 498

17 E-mail address: [xiaolong.chen1@unsw.edu.au](mailto:xiaolong.chen1@unsw.edu.au)

18 **Acknowledgment:** The authors would like to thank Vivek A.S. Ramakrishna (PhD Candidate, Spine Labs in St.  
19 George & Sutherland Clinical School and School of Mechanical and Manufacturing Engineering, University of  
20 New South Wales (UNSW)) for drawing the Figure 1 in this manuscript.

21 **Conflicts of Interest:** All authors declare that they have no conflict of interest related to this work.

22 **Funding disclosure(s) statement:** This work was supported by a Research Training Program scholarship and a  
23 University Postgraduate Award from the Australian Government and The University of New South Wales  
24 (UNSW) to XLC.

25 **Ethical approval:** The study was approved by UNSW (ref no: HC180423) Human Research Ethics  
26 Committees. All procedures performed in studies involving human participants were in accordance with the

**NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.**

27 ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration  
28 and its later amendments or comparable ethical standards.

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47 **Abstract**

48 **Purpose:** To evaluate intra- and inter-rater agreement and reliability of seven reported disc height index (DHI)  
49 measurement methods on standing lateral X-ray of lumbar spine.

50 **Methods:** The adult patients who had standing lateral X-ray of lumbar spine were recruited. Seven methods  
51 were used to measure DHI of each lumbar intervertebral disc level. Bland and Altman's Limits of Agreement  
52 (LOA) with standard difference were calculated to examine intra- and inter-rater agreements between two out of  
53 seven methods for DHI. Intra-class correlations (ICC) with 95% confidence intervals were calculated to assess  
54 intra- and inter-rater reliability.

55 **Results:** The intra-rater reliability in DHI measurements for 288 participants were ICCs from 0.807 (0.794,  
56 0.812) to 0.922 (0.913, 0.946) by rater 1 (SS) and from 0.827 (0.802, 0.841) to 0.918 (0.806, 0.823) by rater 2  
57 (XC). Method 2, 3, and 5 on all segmental levels had bias (95% CI does not include zero) or/and out of the  
58 acceptable cut-off proportion (>50%). A total of 609 outliers in 9174 segmental levels' LOA range. Inter-rater  
59 reliability was good-to-excellent in all but method 2 (0.736 (0.712, 0.759)) and method 5 (0.634 (0.598, 0.667)).  
60 ICCs of related lines to good-to-excellent reliability methods was excellent in all but only indirect line in  
61 method 1 and 4 (ICCs lie in the range from 0.8 to 0.9).

62 **Conclusion:** Following structured protocol, intra- and inter-rater reliability was good-to-excellent for most DHI  
63 measurement methods on X-ray. However, in the presence of vertebral rotation, one should exercise caution in  
64 using complicated methods to define vertebral landmarks.

65 **Keywords:** Lumbar disc herniation, discectomy, disc height, disc height index, agreement, reliability.

66

67

68

69

70

## 71 **Introduction**

72 Low back pain (LBP) is the leading cause of disability worldwide with a lifetime prevalence that  
73 exceeds 90% [1]. Within the vast differential of LBP, the degeneration of intervertebral disc (IVD) is considered  
74 as a significant contributor [2]. Radiological examinations of the morphologic characteristic of lumbar IVD such  
75 as height has been found to be related to the degeneration of IVD [3]. The change of IVD height influences the  
76 load-carrying capacity of the spinal column, and morphologic abnormalities such as IVD space narrowing and  
77 thinning have been potentially associated with acute or chronic disabilities of the lumbar spine [4]. However,  
78 there is a paucity of information using different methods to estimate the disc height (DH) and its clinical  
79 significance. Therefore, an accurate and efficient measurement for IVD height is required.

80 Compared with lying supine during the magnetic resonance imaging (MRI) and computed tomography  
81 (CT) scan, the standing X-ray of lumbar spine can better present the state of IVD under load. Therefore, X-ray is  
82 considered as the most frequently used technique despite known difficulties, both in interpretation and clinical  
83 significance of findings. Clinicians often rely on their own subjective interpretation of lumbar spine radiographs,  
84 however, numerous methods for DH using X-ray published in the literature have been described as more  
85 accurate, albeit, and more time consuming [5-10]. DH can be measured as an absolute value, although this may  
86 be influenced by the magnification and position of the patient on the scan. Simple values can be used in daily  
87 practice for quick comparisons. For more in-depth studies and more accurate readings, the disc height index  
88 (DHI) has been introduced. By normalising images, variations in the size of the vertebral column and position of  
89 the patient do not affect the final measurement and allow for a reliable analysis. Many DHI measurement  
90 methods of IVD has been discussed previously in the literature [5-10]. However, this lack of consensus leads to  
91 great inter- and even intra-rater variability. A simple and reproducible method to measure DHI is required.

92 Bland and Altman's Limits of Agreement (LOA) is the most popular [11], and recommended statistical  
93 method for evaluation of agreement [12, 13]. The standard error of measurement (SEM) is similarly regarded as  
94 a suitable parameter of agreement, but is, however, sensitive to variability in the population [14]. Although  
95 recent study reported use of LOA for evaluating agreement of measurements on intervertebral disc morphology  
96 using MRI images [15], it is rarely used when evaluating agreement in the different measurements of DHI using  
97 X-ray. Therefore, we need to use LOA to evaluate intra- and inter-rater agreement and reliability of DHI using  
98 the previously reported methods [5-10].

99           The primary objective of this study is to evaluate intra- and inter-rater agreement and reliability of  
100 seven previously reported DHI measurement methods.

## 101 **Methods**

102           Ethical approval was obtained from the Human Research Ethics Committee of the University of New  
103 South Wales (NRR-HC180423) for the intra- and inter-rater agreement and reliability study using repeated  
104 measurement methods of individuals' X-rays.

### 105 *Design and Patients*

106           The study is conducted as a retrospective review of radiological images, radiology reports, and  
107 demographic data of patients over the age of 18 years who had routine standing lateral X-ray of lumbar spine  
108 from St George MRI in Sydney (Australia) from March 2017 onwards. Only patients who signed the consent  
109 form to allow use of their de-identified data for research and auditing purposes were included in the present  
110 study. The patients who had a history of spine surgery were excluded from the study.

### 111 *Measurements*

112           The standard standing lateral X-ray images of lumbar spine were assessed. The patient is naturally  
113 standing up, looking horizontally, hands resting on a vertical support, upper limbs relaxed, elbows half bent  
114 [16]. The corresponding radiology reports were read by the first author (XC). Seven methods were used to  
115 measure the DHI of each lumbar IVD level on standing later X-ray images (L1-L2, L2-L3, L3-L4, L4-L5, and  
116 L5-S1) [5-10]. The protocol and details of DHI measurement methods are presented as follows and showed in  
117 Electronic Supplementary Material 1 (ESM\_1) and Fig. 1.

118           Method 1 of DHI is expressed as a ratio of the sum of anterior and posterior IVD height to disc  
119 diameter [5]. Method 2 of DHI is expressed as a ratio of the mid-disc height to mid-vertebral body height [6].  
120 Method 3 of DHI is expressed as a ratio of the mid-disc height to disc diameter [6]. Method 4 of DHI is  
121 expressed as a ratio of the mean of anterior, middle, and posterior IVD height to the sagittal diameter of the  
122 proximal vertebral body [7]. Method 5 of DHI is expressed as a ratio of IVD height to vertebral height which  
123 cross the centre of adjacent vertebral bodies [8]. Method 6 of DHI is expressed as a ratio of the mean of anterior,  
124 middle, and posterior IVD height to the mean of proximal and distal vertebral body height [9]. Method 7 of DHI

125 is expressed as a ratio of the sum of anterior and posterior IVD height to the sum of superior and inferior disc  
126 depth [10].

127 A quadrilateral was drawn to define the vertebral corners and minimize the affection of osteophytes. A  
128 line was drawn cross the potential points of each corner which was caused by the vertebral rotation for inexact  
129 body position during the scan and the anatomy deformity (such as scoliosis, vertebral rotation, and vertebral  
130 fracture). Mid-point of the line was identified as the real vertebral corner. Direct line was draw cross the two  
131 points which were located at the vertebral body. Indirect line was drawn cross the potential points which were  
132 location at direct lines.

133 If MRI scans already performed and presented in St Georgy MRI, the images were assessed the IVD  
134 degeneration. IVD degeneration is defined as the presence of at least one of the following: nucleus pulposus  
135 degeneration, IVD bulge or IVD herniation, annular tear, Modic changes of endplate, and Schmorl's node [17-  
136 21]. Nucleus pulposus degeneration is defined as Pfirrmann grade  $\geq 3$  [22]. Participants were allocated into  
137 different groups (degeneration group and no degeneration group) based on the IVD degeneration status.

138 In order to reduce the potential bias due to difference of equipment and software, raters used Apple  
139 MacBook with integrated touchpads and the *InteleViewer<sup>TM</sup>* diagnostic imaging software for measurement.

#### 140 ***Training and Blinding of Raters***

141 Two raters conducted the measurements: one is a medical student (SS) who has no prior training in the  
142 interpretation of radiological images (Rater 1); the other is an experienced spine surgeon and back pain  
143 researcher (XC) with extensive experience in interpreting radiological images (Rater 2). Thirty participants from  
144 the final data collection period were randomly selected for training. Each rater reviewed the 30 cases  
145 independently, after which the cases were collectively reviewed, and consensus were reached on the  
146 measurement procedures. Once the raters reached an agreement on the measurement procedures, the data of  
147 these 30 cases was used to analysis the intra-rater reliability. The intra- and inter-rater agreement were tested  
148 between two out of seven measurements performed by each rater. The inter-rater reliability was tested between  
149 two raters who were purposely chosen to represent an inexperienced, and an experienced interpreter of  
150 radiological images.

151 To enhance the quality and applicability of the study, both raters were blinded in several aspects. Each  
152 rater was blinded to his own prior measurements and the findings of the other ratter. The order of participants

153 was randomly changed between the two intra-rater measurement sessions. There was a 2-week interval between  
154 the first and second measurement sessions to lessen the likelihood of recognition of participants.

### 155 *Statistical analysis*

156 Numeric variables are presented as mean  $\pm$  standard deviation (SD). Categorical variables are  
157 summarized using counts (n) and percentages (%). The intra- and inter-rater agreement between two out of  
158 seven methods for DHI were analysed using Bland and Altman's LOA. LOA is based on graphical techniques  
159 and provides a plot of mean differences (MDs) between the two methods of measurement (the bias), as well as  
160 the SD of the differences (Fig. 2). The 95% confident intervals (95% CI) of MDs were reported to describe the  
161 precision of the bias. If the 95% CI doesn't include zero, it can be assumed that there is a bias. Furthermore,  
162 LOA was presented as a proportion of mean values for each method. The proportion will be calculated as  
163 follows:  $((\text{upper LOA} + (-1 * (\text{lower LOA}))) / (\text{the mean})) * 100\%$ . Following previously published data, we  
164 consider percentages lower than 50% as an indicator of acceptable precision [15].

165 Intra-class correlation coefficient (ICC) estimates, and their 95% CI were calculated using SPSS  
166 statistical package version 24 (SPSS Inc, Chicago, IL) based on a single-rating, absolute-agreement, 2-way  
167 fixed-effects model for intra-rater reliability. Results of inter-rater reliability was evaluated with ICC based on a  
168 single-rating, consistency, 2-way random-effects model in all participants and different degeneration groups.  
169 Values of ICC less than 0.5, between 0.5 and 0.75, between 0.75 and 0.9, and greater than 0.90 are indicative of  
170 poor, moderate, good, and excellent reliability, respectively [23]. Subgroup analysis was performed based on  
171 different segmental level, the status of IVD degeneration, and different related lines (direct and indirect line).

### 172 *Factors analysis on the Bland and Altman's plot*

173 Potential factors for the data that were far above or below the LOA on the graphs were assessed and  
174 reported in a narrative form.

## 175 **Results**

176 In total, the standing lumbar X-ray from 288 participants were included in this study for evaluation of  
177 both intra- and inter-rater reliability and agreement. There were 122 females and 166 males, all aged between 19  
178 and 89 years. Of 367 lumbar levels with IVD degeneration in 278 participants who performed MRI scans (Table  
179 1).

180 ***Intra-rater reliability***

181 The intra-rater reliability for DHI of all measurement methods, using ICC, was good-to-excellent from  
182 0.807 (0.794, 0.812) to 0.922 (0.913, 0.946) by rater 1 and from 0.827 (0.802, 0.841) to 0.918 (0.806, 0.823) by  
183 rater 2, respectively (Table 2).

184 ***Inter-rater agreement***

185 ***Method 1***

186 The 95% CI of mean difference of DHI on segmental level L3-L4 ranged between -0.006 and 0.005,  
187 with LOA ranging between -0.10 and 0.10 (LOA as proportion of mean values is 33.9%). The 95% CI of mean  
188 difference of DHI on segmental level L5-S1 ranged between -0.014 and 0.002, with LOA ranging between -0.15  
189 and 0.13 (LOA as proportion of mean values is 46.7%) (Table 3 and Fig. 3A).

190 ***Method 2, 3, and 5***

191 The 95% CI of mean difference of DHI on all lumbar levels did not include zero or LOA as proportion  
192 of mean values is more than 50% (Table 3, Fig. 3B, Fig. 3C, and Fig. 3E).

193 ***Method 4***

194 The 95% CI of mean difference of DHI on segmental level L4-L5 ranged between -0.008 and 0.002,  
195 with LOA ranging between -0.06 and 0.06 (LOA as proportion of mean values is 38.7%) (Table 3 and Fig. 3D).

196 ***Method 6***

197 The 95% CI of mean difference of DHI on segmental level L3-L4 ranged between -0.003 and 0.010,  
198 with LOA ranging between -0.12 and 0.12 (LOA as proportion of mean values is 17%). The 95% CI of mean  
199 difference of DHI on segmental level L4-L5 ranged between -0.009 and 0.001, with LOA ranging between -0.08  
200 and 0.08 (LOA as proportion of mean values is 40%) (Table 3 and Fig. 3F).

201 ***Method 7***

202 The 95% CI of mean difference of DHI on segmental level L3-L4 ranged between -0.003 and 0.003,  
203 with LOA ranging between -0.04 and 0.04 (LOA as proportion of mean values is 27.6%). The 95% CI of mean  
204 difference of DHI on segmental level L5-S1 ranged between -0.006 and 0.001, with LOA ranging between -0.06  
205 and 0.06 (LOA as proportion of mean values is 40%) (Table 3 and Fig. 3G).



## 206 *Inter-rater reliability*

207 The inter-rater reliability for measurements of DHI was good-to-excellent in all but method 2 and 5  
208 (ICCs ranged from 0.634 (0.598, 0.667) to 0.984 (0.982, 0.985); method 2: 0.736 (0.712, 0.759); method 5:  
209 0.634 (0.598, 0.667)) (Table 4).

## 210 *Temporal analysis*

211 Based on different segmental level, ICCs for DHI on segment level L1-L2 was moderate in method 2,  
212 3, and 5 groups (ICC: 0.641 (0.568, 0.705), 0.718 (0.657, 0.770), 0.500 (0.409, 0.582)). ICCs for DHI on  
213 segment level L2-L3 was moderate in method 2 and 5 groups (ICC: 0.620 (0.543, 0.686), 0.726 (0.666, 0.776)).  
214 ICCs for DHI on segment level L3-L4 was moderate in method 2 and 5 groups (ICC: 0.693 (0.628, 0.749),  
215 0.728 (0.669, 0.778)) (Table 4).

216 Based on the status of IVD degeneration, ICCs of DHI on all segmental levels in degeneration group  
217 and no degeneration group have a similar range based on the classification criterion for poor, moderate, good,  
218 and excellent reliability (ESM\_2\_Table 1).

219 ICCs of related lines to good-to-excellent reliability methods was excellent in all but only indirect line  
220 in method 1 and 4 (ICCs lie in the range from 0.8 to 0.9, ESM\_2\_Table 2).

## 221 *Factors analysis on the Bland and Altman plot*

222 A total of 609 outliers in 9174 segmental levels' data includes 57 outliers in the method 1 group, 65  
223 outliers in the method 2 group, 171 outliers in the method 3 group, 182 outliers in the method 4 group, 37  
224 outliers in the method 5 group, 42 outliers in the method 6 group, and 55 outliers in the method 7 group  
225 (ESM\_2\_Table 3). The nucleus pulposus degeneration (394) and disc herniation (186) affected the raters to  
226 distinguish vertebral corners and structural boundaries.

## 227 **Discussion**

228 The reduction of IVD height is the key point in the pathological process of IVD degeneration, and the  
229 diseases of lumbar degeneration often demonstrate the reduction of IVD height in the radiographic images.  
230 Therefore, a reproducible method to measure IVD height is required. To be the best of our knowledge, this is the  
231 first study to investigate the intra- and inter-rater reliability and agreement of previously reported DHI methods  
232 to measure DH on the standing lateral lumbar X-ray images. We used a structured protocol including

233 descriptions of testing positions, standard training session of measurements on images for raters, unified  
234 measurement platform and tools, and blinding of raters [24].

235           Although the measurements on X-ray images would be affected by body posture and vertebral position  
236 of the patient on the scan and the experience of raters [6, 15, 25-27], our study still shown that intra-rater  
237 reliability was good-to-excellent for all the seven DHI assessment methods on X-ray images by both  
238 inexperienced and experienced raters. A possible explanation is the existence of division in the process of  
239 calculating the DHI, which can minimize the measurement bias by the inconsistent magnification and vertebral  
240 position on the X-ray scan. We posit the systematic training and structured protocol to conduct the measurement  
241 to be the other main cause of the good-to-excellent intra-rater reliability on DHI measurement methods on X-ray  
242 images. Therefore, the systematic training before measurement and a standard measurement process following  
243 structured protocol could provide a good-to-excellent intra-rater reliability for the DHI measurement on X-ray  
244 images.

245           Agreement is commonly used to evaluate how well the measurements produced by two raters, devices  
246 or systems agree with each other, while reliability is concerned with measurement error plus the variability  
247 between study objects and the focus is distinction between persons [15, 28]. Previously published study  
248 recommended reporting inter-rater agreement parameters via LOA, and further, when reporting reliability using  
249 ICC, they should be reported together with error estimates such as the standard error of the mean [28].  
250 Following the results of Bland and Altman's LOA, we found that the DHI measurements in method 2, 3, and 5  
251 on all segmental levels and method 1, 4, 6 and 7 on some special segmental levels had bias or/and out of the  
252 acceptable cut-off proportion. Due to different numbers of indirect lines in each method, it indicates a poor-to-  
253 moderate consensus regarding the anatomical delineation on the length measurements between the two raters.  
254 These were consistent with the status of ours' study that all indirect lines involved in method 2 and 5 and partial  
255 indirect lines involved in method 1, 3, and 4 with a poor-to-moderate agreement. Meanwhile, nucleus pulposus  
256 degeneration and disc herniation were showed to impact of the inter-rater agreement on distinguish vertebral  
257 corners and structural boundaries.

258           This study uses both LOA and reliability to express reproducibility. The inter-rater reliability was  
259 good-to-excellent in all but method 2 and 5. Although IVD degeneration can cause discs to lose height and  
260 might potentially affect the accuracy and agreement of DHI measurement [9], our findings denied the influence  
261 of IVD degeneration on the inter-rater reliability results in different measurement methods on DHI

262 (ESM\_2\_Table 1). The potential risk factors for the moderate inter-rater reliability on DHI measurement in  
263 method 2 and 5 include measurement bias of indirect lines and other bias from anatomical structure. Due to use  
264 of multiple indirect lines in method 2 and 5, the potential secondary measurement bias following the first bias  
265 by the inexact positioning of vertebral corners and indistinguishable IVD boundaries between structures during  
266 drawing the direct line might cause the moderate inter-reliability. This indicates that a complicated measurement  
267 method would cause a poor-to-moderate consensus between raters. Despite good-to-excellent inter-rater  
268 reliability on DHI measurement in method 1, 3, 4, 6, and 7, there still showed relatively poor results of inter-  
269 rater reliability on indirect lines in method 1 and 4 (ESM\_2\_Table 2). Therefore, the accurate and effective  
270 determination of vertebral corners for direct line can significantly reduce the measurement error. As for the  
271 positioning of vertebral corners, two possible interfering factors could be the presence of osteophytes and the  
272 rotation of vertebral body, hence, modifying the visual appearance of the vertebra [6, 15, 29]. While our  
273 structured protocol could minimize the influence of osteophytes on marking the corners, it can't provide a  
274 method to avoid the objective factor that leads to vertebral rotation. For instance, upper vertebral rotation by  
275 IVD no perpendicular to the projection might be the reason for moderate inter-rater reliability of DHI  
276 measurements on upper segmental level (L1-L2 and L2-L3). Meanwhile, the shorter DH of the upper IVD could  
277 induce cumulative error in the marking of vertebral corners, which was posited to be the other reason. We  
278 couldn't find studies that definitively discussed any of these factors regarding similar problems with  
279 measurement bias of indirect lines, vertebral rotation, or boundary distinction. However, we still thought that  
280 these could be the main reasons why some ICCs of inter-rater reliability were moderate. As it stands, our study  
281 potentially showed that there was a good-to-excellent intra- and inter-rater reliability and agreement on the DHI  
282 measurements in method 7 for all IVD segmental levels. For future use of these methods, specification in  
283 advance of measurements, and persistent implementation of detailed protocol for the location of projection,  
284 measurement of indirect lines, and dealing with vertebral rotation, should be conducted by all raters.

### 285 *Study limitation and future study*

286 Several methodological issues require consideration. First, the potential measurement error due to  
287 notably inexact definition of anatomic measurement points, the location of projection during the scan, definition  
288 of standard process to fix vertebral rotation, and intra- and inter-rater variation, despite a structural protocol  
289 being provided to raters. Future, a standardized protocol to assess DHI was required. Second, due to the  
290 difficulties in distinguishing the boundary of disc on X-ray, the raters can only use point-based measurement  
291 method instead of area-based method. Third, the acceptable precision of the range of LOA set at 50% following

292 previously published data would affect the results [15]. Fourth, due to the different reference values of each DHI  
293 method, the direct comparison between two out of seven measurement methods can't be done. Finally, the aim  
294 of this study was to establish reproducibility and reliability, not to report prevalence or reference values for  
295 either a general or a clinical population. Future multicenter study on the validity of different measurement  
296 methods is needed.

## 297 **Conclusion**

298 The intra-rater and most inter-rater reliability for DHI measurement was good-to-excellent for different  
299 methods following a structured protocol. However, the inter-rater reliability was moderate in some DHI  
300 measurement methods, indicating difficulties in the performance of these tests. The complicated methods (more  
301 indirect lines) and IVD degeneration (nucleus pulposus degeneration and disc herniation) potentially affected  
302 the agreement on inter-rater measurements. Future multicenter study on the validity of different measurement  
303 methods following a standardized protocol is needed.

304

305

306

307

308

309

310

311

312

313

314

315

316

317

318

319

320

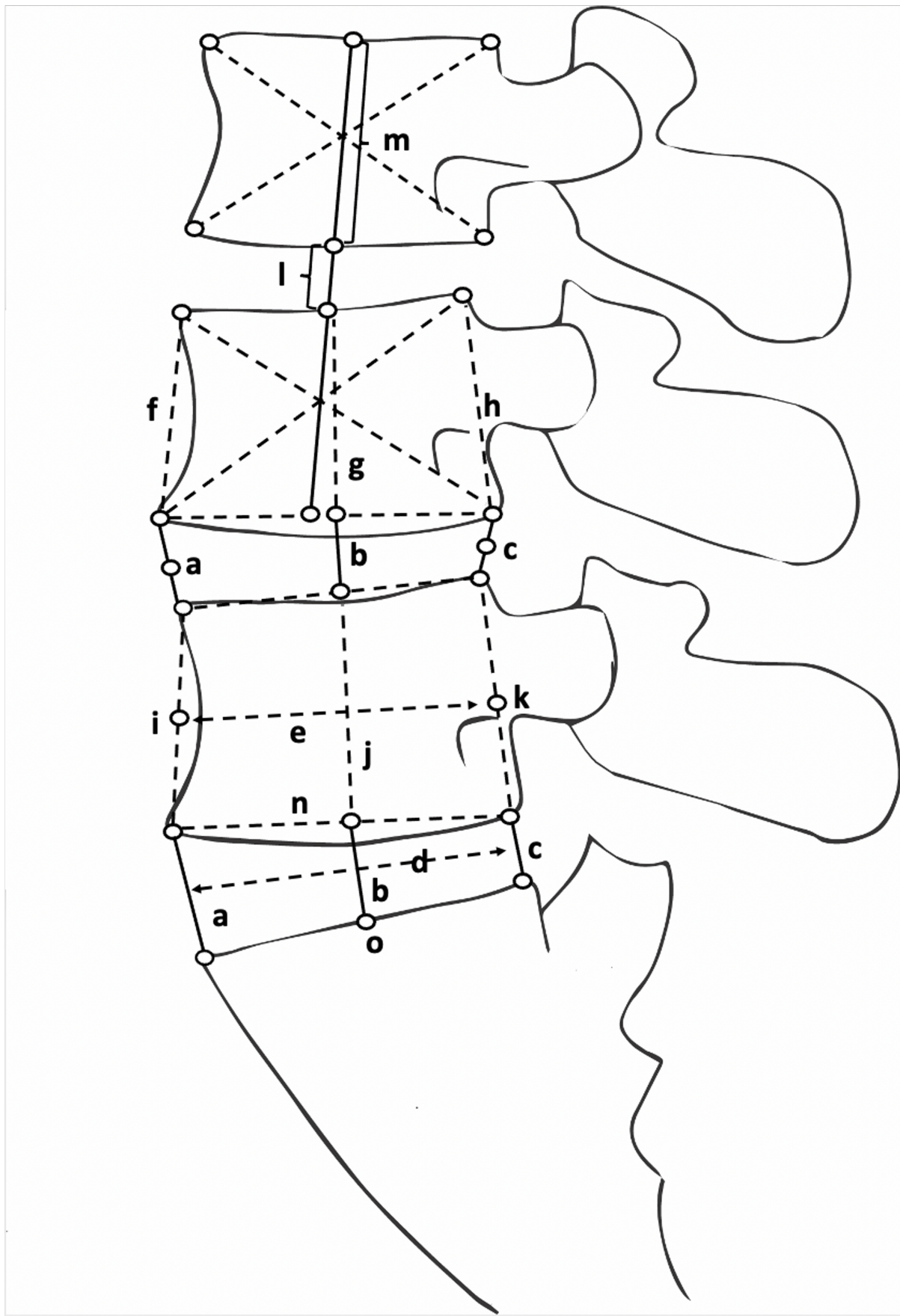
## 321 **References**

- 322 1. Hartvigsen J, Hancock MJ, Kongsted A, Louw Q, Ferreira ML, Genevay S, Hoy D, Karppinen J, Pransky G,  
323 Sieper J, Smeets RJ, Underwood M, Buchbinder R, Hartvigsen J, Cherkin D, Foster NE, Maher CG,  
324 Underwood M, van Tulder M, Anema JR, Chou R, Cohen SP, Menezes Costa L, Croft P, Ferreira M, Ferreira  
325 PH, Fritz JM, Genevay S, Gross DP, Hancock MJ, Hoy D, Karppinen J, Koes BW, Kongsted A, Louw Q, Öberg  
326 B, Peul WC, Pransky G, Schoene M, Sieper J, Smeets RJ, Turner JA, Woolf A (2018) What low back pain is  
327 and why we need to pay attention. *The Lancet* 391:2356-2367. doi: 10.1016/s0140-6736(18)30480-x
- 328 2. de Schepper EI, Damen J, van Meurs JB, Ginai AZ, Popham M, Hofman A, Koes BW, Bierma-Zeinstra SM  
329 (2010) The association between lumbar disc degeneration and low back pain: the influence of age, gender, and  
330 individual radiographic features. *Spine (Phila Pa 1976)* 35:531-536. doi: 10.1097/BRS.0b013e3181aa5b33
- 331 3. Pfirrmann CW, Metzdorf A, Elfering A, Hodler J, Boos N (2006) Effect of aging and degeneration on disc  
332 volume and shape: A quantitative study in asymptomatic volunteers. *J Orthop Res* 24:1086-1094. doi:  
333 10.1002/jor.20113
- 334 4. Beattie PF, Meyers SP (1998) Magnetic resonance imaging in low back pain: general principles and clinical  
335 issues. *Phys Ther* 78:738-753. doi: 10.1093/ptj/78.7.738
- 336 5. Farfan HF (1973) Mechanical disorders of the low back. Lea & Febiger, Philadelphia
- 337 6. Pope MH, Hanley EN, Mattern RE, Wilder DG, Frymoyer JW (1977) Measurement of intervertebral disc  
338 space height. *Spine (Philadelphia, Pa 1976)* 2:p.282-286
- 339 7. Inoue H, Ohmori K, Miyasaka K, Hosoe H (1999) Radiographic evaluation of the lumbosacral disc height.  
340 *Skeletal Radiol* 28:638-643. doi: 10.1007/s002560050566
- 341 8. **Kim KT, Park SW, Kim YB** (2009) Disc height and segmental motion as risk factors for recurrent lumbar  
342 disc herniation. *Spine (Phila Pa 1976)* 34:2674-2678. doi: 10.1097/BRS.0b013e3181b4aaac
- 343 9. Jarman JP, Arpinar VE, Baruah D, Klein AP, Maiman DJ, Muftuler LT (2015) Intervertebral disc height loss  
344 demonstrates the threshold of major pathological changes during degeneration. *Eur Spine J* 24:1944-1950. doi:  
345 10.1007/s00586-014-3564-8
- 346 10. Akeda K, Yamada T, Inoue N, Nishimura A, Sudo A (2015) Risk factors for lumbar intervertebral disc  
347 height narrowing: a population-based longitudinal study in the elderly. *BMC Musculoskelet Disord* 16:344. doi:  
348 10.1186/s12891-015-0798-5

- 349 11. Zaki R, Bulgiba A, Ismail R, Ismail NA (2012) Statistical methods used to test for agreement of medical  
350 instruments measuring continuous variables in method comparison studies: a systematic review. *PLoS One*  
351 7:e37908. doi: 10.1371/journal.pone.0037908
- 352 12. Bland JM, Altman DG (2010) Statistical methods for assessing agreement between two methods of clinical  
353 measurement. *International Journal of Nursing Studies* 47:931-936. doi: 10.1016/j.ijnurstu.2009.10.001
- 354 13. Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hrobjartsson A, Roberts C, Shoukri M, Streiner  
355 DL (2011) Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin*  
356 *Epidemiol* 64:96-106. doi: 10.1016/j.jclinepi.2010.03.002
- 357 14. Atkinson G, Nevill AM (1998) Statistical methods for assessing measurement error (reliability) in variables  
358 relevant to sports medicine. *Sports Med* 26:217-238. doi: 10.2165/00007256-199826040-00002
- 359 15. Tunset A, Kjaer P, Samir Chreiteh S, Secher Jensen T (2013) A method for quantitative measurement of  
360 lumbar intervertebral disc structures: an intra- and inter-rater agreement and reliability study. *Chiropr Man*  
361 *Therap* 21:26. doi: 10.1186/2045-709X-21-26
- 362 16. Vialle R, Levassor N, Rillardon L, Templier A, Skalli W, Guigui P (2005) Radiographic analysis of the  
363 sagittal alignment and balance of the spine in asymptomatic subjects. *J Bone Joint Surg Am* 87:260-267. doi:  
364 10.2106/JBJS.D.02043
- 365 17. Suthar P, Patel R, Mehta C, Patel N (2015) MRI evaluation of lumbar disc degenerative disease. *J Clin*  
366 *Diagn Res* 9:TC04-09. doi: 10.7860/JCDR/2015/11927.5761
- 367 18. Taher F, Essig D, Lebl DR, Hughes AP, Sama AA, Cammisa FP, Girardi FP (2012) Lumbar degenerative  
368 disc disease: current and future concepts of diagnosis and management. *Adv Orthop* 2012:970752. doi:  
369 10.1155/2012/970752
- 370 19. Urrutia J, Besa P, Campos M, Cikutovic P, Cabezon M, Molina M, Cruz JP (2016) The Pfirrmann  
371 classification of lumbar intervertebral disc degeneration: an independent inter- and intra-observer agreement  
372 assessment. *Eur Spine J* 25:2728-2733. doi: 10.1007/s00586-016-4438-z
- 373 20. Williams FM, Manek NJ, Sambrook PN, Spector TD, Macgregor AJ (2007) Schmorl's nodes: common,  
374 highly heritable, and related to lumbar disc disease. *Arthritis Rheum* 57:855-860. doi: 10.1002/art.22789
- 375 21. Zhang YH, Zhao CQ, Jiang LS, Chen XD, Dai LY (2008) Modic changes: a systematic review of the  
376 literature. *Eur Spine J* 17:1289-1299. doi: 10.1007/s00586-008-0758-y

- 377 22. Sharma A, Lancaster S, Bagade S, Hildebolt C (2014) Early pattern of degenerative changes in individual  
378 components of intervertebral discs in stressed and nonstressed segments of lumbar spine: an in vivo magnetic  
379 resonance imaging study. *Spine (Phila Pa 1976)* 39:1084-1090. doi: 10.1097/BRS.0000000000000265
- 380 23. Koo TK, Li MY (2016) A guideline of selecting and reporting intraclass correlation coefficients for  
381 reliability research. *J Chiropr Med* 15:155-163. doi: 10.1016/j.jcm.2016.02.012
- 382 24. Lucas NP, Macaskill P, Irwig L, Bogduk N (2010) The development of a quality appraisal tool for studies of  
383 diagnostic reliability (QAREL). *J Clin Epidemiol* 63:854-861. doi: 10.1016/j.jclinepi.2009.10.002
- 384 25. Neubert A, Fripp J, Engstrom C, Gal Y, Crozier S, Kingsley MI (2014) Validity and reliability of  
385 computerized measurement of lumbar intervertebral disc height and volume from magnetic resonance images.  
386 *Spine J* 14:2773-2781. doi: 10.1016/j.spinee.2014.05.023
- 387 26. Cousins JP, Haughton VM (2009) Magnetic resonance imaging of the spine. *J Am Acad Orthop Surg* 17:22-  
388 30. doi: 10.5435/00124635-200901000-00004
- 389 27. Schlager A, Ahlqvist K, Rasmussen-Barr E, Bjelland EK, Pingel R, Olsson C, Nilsson-Wikmar L,  
390 Kristiansson P (2018) Inter- and intra-rater reliability for measurement of range of motion in joints included in  
391 three hypermobility assessment methods. *BMC Musculoskelet Disord* 19:376. doi: 10.1186/s12891-018-2290-5
- 392 28. de Vet HC, Terwee CB, Knol DL, Bouter LM (2006) When to use agreement versus reliability measures. *J*  
393 *Clin Epidemiol* 59:1033-1039. doi: 10.1016/j.jclinepi.2005.10.015
- 394 29. Saraste H, Brostrom LA, Aparisi T, Axdorph G (1985) Radiographic measurement of the lumbar spine. A  
395 clinical and experimental study in man. *Spine (Phila Pa 1976)* 10:236-241. doi: 10.1097/00007632-198504000-  
396 00008
- 397
- 398
- 399
- 400
- 401
- 402
- 403
- 404
- 405
- 406

407 **Fig. 1** The details of disc height index (DHI) measurements



408



409 Note: a: The shortest distance between the anterior edges of the neighbouring endplate will be recorded as the  
410 anterior disc height; b: The mid-disc height between the upper and lower bisection points is measured at the  
411 midpoint of vertebrae; c: The shortest distance between the posterior edges of the neighbouring endplate will be  
412 recorded as the posterior disc height; d: The disc diameter will be measured between the midpoints of the lines  
413 drawn from the endpoints of the superior vertebral endplate to the inferior; e: The sagittal diameter of the  
414 vertebral body from the anterior to posterior margin will be measured at the mid-vertebral level; f-h: The  
415 proximal vertebral body height will be measured from the anterior (f), middle (g), and posterior (h) portions of  
416 each respective disc level; i-k: The distal (DV) vertebral body height will be measured from the anterior (i),  
417 middle (j), and posterior (k) portions of each respective disc level; l, m: The mid-vertebral line is the line  
418 connecting the L3 and L4 centres. The centre of the vertebral body is a crossing point of 2 diagonal lines of each  
419 vertebral body (l is intervertebral disc height, m is intervertebral height); n: superior disc depth; o: inferior disc  
420 depth. According to the classification of related lines, line a, c, f, h, i, k, n and o are defined as direct lines and  
421 line b, d, e, g, j, l, and m are defined as indirect lines.

422 Method 1:  $DHI = [(a+c)/d] * 100\%$

423 Method 2:  $DHI = (b/g) * 100\%$  or  $(b/j) * 100\%$

424 Method 3:  $DHI = (b/d) * 100\%$

425 Method 4:  $DHI = [(a+b+c)/3/e] * 100\%$

426 Method 5:  $DHI = (l/m) * 100\%$

427 Method 6:  $DHI = [2*(a+b+c)/((f+g+h)+(i+j+k))] * 100\%$

428 Method 7:  $DHI = [(a+c)/(n+o)] * 100\%$

429

430

431

432

433

434

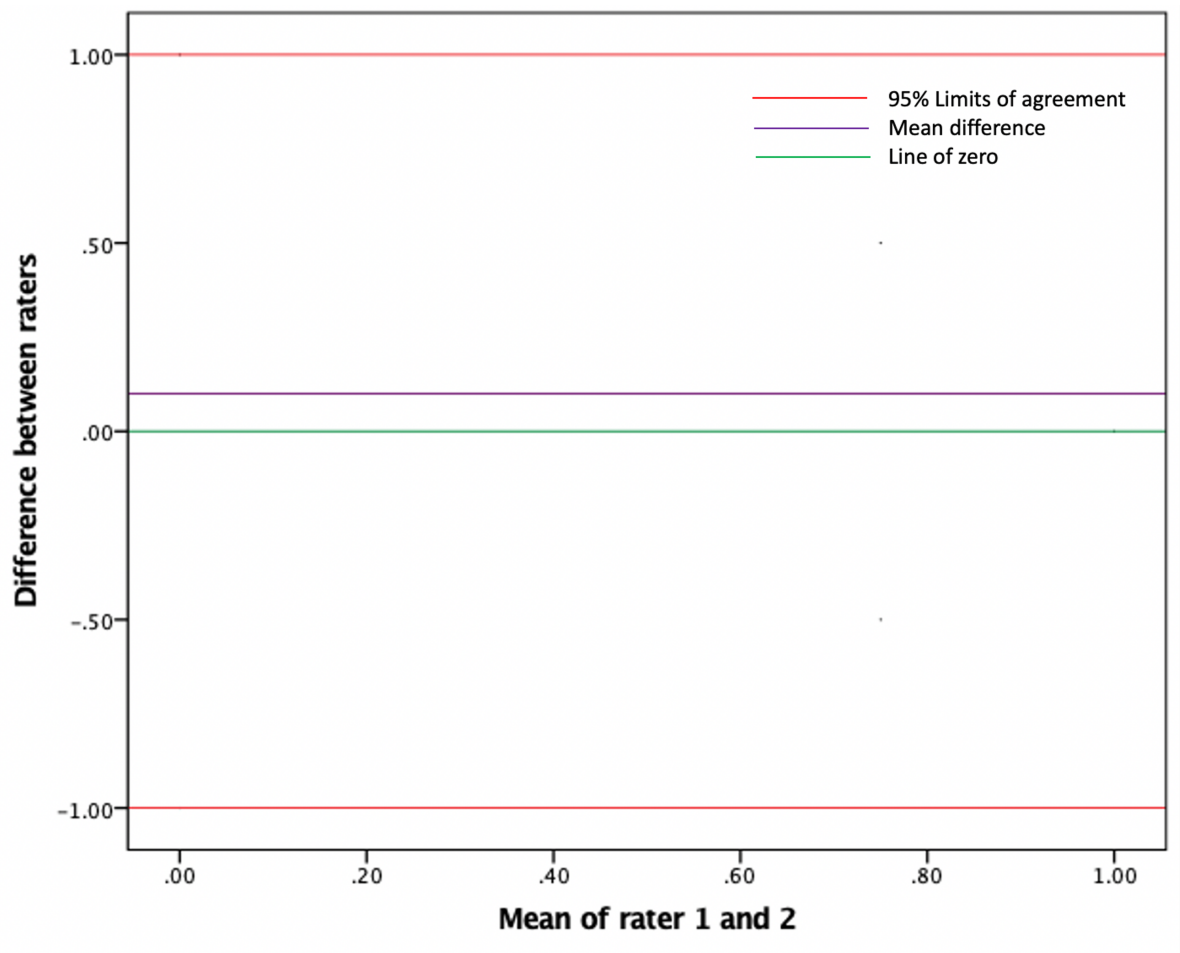
435

436

437

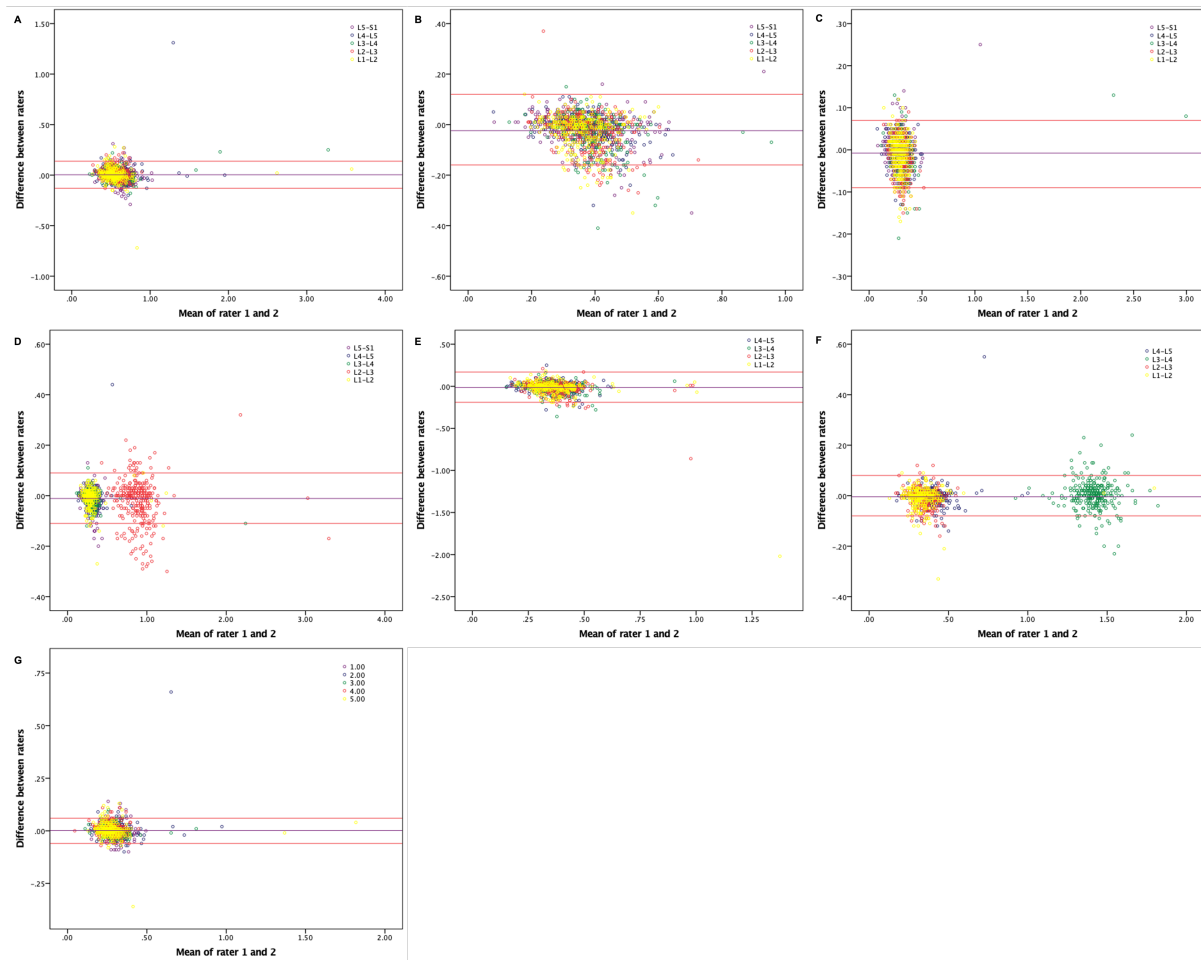
438

439 **Fig. 2** The Bland and Altman plot of Limits of Agreement (LOA) between two raters on the different  
440 measurements for disc height index (DHI). The y-axis shows the mean difference between raters'  
441 measurements, and the x-axis shows the mean value of both raters' measurements. The green line shows the  
442 range of mean difference includes zero. The purple line shows the mean difference between measurements. Red  
443 lines show the 95% LOA.



444  
445  
446  
447  
448  
449  
450  
451  
452  
453

454 **Fig. 3** The Bland and Altman plot showing the relationship between mean values and differences between rater  
455 1 and rater 2 on the measurements of DHI using two out of seven reported methods (A: method 1; B: method 2;  
456 C: method 3; D: method 4; E: method 5; F: method 6; G: method 7). Mean difference with 95% confidence  
457 intervals (CI) of the measurements between rater 1 and rater 2 was reported to describe the precision of the bias.  
458 The purple line shows the mean difference between measurements. Red lines show the 95% Limits of  
459 Agreement (LOA), between which 95% of all measurement differences are located.



460

461

462

463

464

465

466

467

468

469 **Table 1** Patient demographic and clinic-radiological information

<b>Parameter</b>	<b>Number of patients</b>
F:M	122:166
Age	47.67±16.79
Diagnosis	
Spondylolisthesis	32 (11.1%)
Disc herniation	57 (19.8%)
Spinal stenosis	174 (60.4%)
Scoliosis	11 (3.8%)
Normal	88 (30.6%)
MRI scans (number of patients)	278 (96.5%)
Intervertebral disc degeneration (number of patients)	231 (83.1%)
Lumbar levels with intervertebral disc degeneration (total)	367
L1-L2	2 (0.5%)
L2-L3	0
L3-L4	36 (9.8%)
L4-L5	160 (43.6%)
L5-S1	169 (46%)

470 F: female; M: male; MRI: Magnetic resonance imaging

471

472

473

474

475

476

477

478

479

480

481

482

483

484

485

486

487

488

489 **Table 2** Intra-rater measures' reliability results

Measurement method		N	Rater 1_ICC (95% CI)	Rater 2_ICC (95% CI)
Method 1	L1-L2	30	0.907 (0.902, 0.921)	0.917 (0.908, 0.924)
	L2-L3	30	0.867 (0.860, 0.882)	0.866 (0.854, 0.884)
	L3-L4	30	0.876 (0.861, 0.893)	0.888 (0.876, 0.893)
	L4-L5	30	0.822 (0.811, 0.843)	0.858 (0.836, 0.873)
	L5-S1	30	0.855 (0.841, 0.873)	0.878 (0.856, 0.893)
	All	150	0.875 (0.872, 0.889)	0.907 (0.902, 0.928)
Method 2	L1-L2	30	0.821 (0.817, 0.842)	0.845 (0.822, 0.864)
	L2-L3	30	0.807 (0.794, 0.812)	0.827 (0.802, 0.841)
	L3-L4	30	0.922 (0.913, 0.946)	0.912 (0.895, 0.946)
	L4-L5	30	0.823 (0.812, 0.844)	0.834 (0.828, 0.862)
	L5-S1	30	0.842 (0.817, 0.881)	0.868 (0.860, 0.890)
	All	150	0.848 (0.807, 0.855)	0.871 (0.869, 0.889)
Method 3	L1-L2	30	0.822 (0.810, 0.864)	0.842 (0.838, 0.876)
	L2-L3	30	0.843 (0.831, 0.861)	0.861 (0.849, 0.873)
	L3-L4	30	0.830 (0.815, 0.856)	0.833 (0.805, 0.856)
	L4-L5	30	0.853 (0.842, 0.883)	0.869 (0.857, 0.883)
	L5-S1	30	0.851 (0.812, 0.865)	0.863 (0.841, 0.876)
	All	150	0.850 (0.832, 0.864)	0.865 (0.844, 0.875)
Method 4	L1-L2	30	0.882 (0.962, 0.896)	0.892 (0.882, 0.898)
	L2-L3	30	0.831 (0.822, 0.848)	0.842 (0.812, 0.855)
	L3-L4	30	0.877 (0.872, 0.882)	0.879 (0.868, 0.885)
	L4-L5	30	0.852 (0.834, 0.881)	0.872 (0.865, 0.889)
	L5-S1	30	0.916 (0.901, 0.923)	0.918 (0.806, 0.823)
	All	150	0.879 (0.869, 0.912)	0.887 (0.875, 0.914)
Method 5	L1-L2	30	0.817 (0.801, 0.841)	0.841 (0.823, 0.866)
	L2-L3	30	0.854 (0.833, 0.867)	0.858 (0.843, 0.872)
	L3-L4	30	0.845 (0.840, 0.869)	0.876 (0.855, 0.883)
	L4-L5	30	0.861 (0.832, 0.877)	0.873 (0.847, 0.879)
	All	120	0.858 (0.836, 0.873)	0.871 (0.845, 0.881)
Method 6	L1-L2	30	0.878 (0.848, 0.881)	0.882 (0.878, 0.891)
	L2-L3	30	0.822 (0.817, 0.881)	0.871 (0.862, 0.889)
	L3-L4	30	0.852 (0.827, 0.881)	0.866 (0.854, 0.887)
	L4-L5	30	0.856 (0.843, 0.883)	0.884 (0.872, 0.894)
	All	120	0.859 (0.846, 0.879)	0.878 (0.873, 0.888)
Method 7	L1-L2	30	0.860 (0.842, 0.866)	0.869 (0.851, 0.878)
	L2-L3	30	0.851 (0.812, 0.865)	0.863 (0.841, 0.876)
	L3-L4	30	0.845 (0.814, 0.868)	0.871 (0.855, 0.886)
	L4-L5	30	0.855 (0.843, 0.865)	0.865 (0.844, 0.875)
	L5-S1	30	0.912 (0.878, 0.922)	0.864 (0.848, 0.872)
	All	150	0.866 (0.844, 0.916)	0.868 (0.858, 0.876)

490 N: number of levels; ICC: Intraclass correlation coefficient; 95% CI: 95% confidence interval.

491

492

493

494 **Table 3** Inter-rater measures agreement results

	Level	N	Mean	SD	Mean difference (95% CI)	95% LOA	LOA as proportion of mean values (%)
<b>Method 1</b>	L1-L2	288	0.55	0.07	0.00 (-0.008, 0.008)	-0.14, 0.14	50.9
	L2-L3	288	0.55	0.05	0.01 (0.005, 0.016) *	-0.09, 0.11	36.4
	L3-L4	288	0.59	0.05	0.00 (-0.006, 0.005)	-0.10, 0.10	33.9
	L4-L5	288	0.61	0.09	0.01 (0.001, 0.022) *	-0.17, 0.19	59.0
	L5-S1	288	0.6	0.07	-0.01 (-0.014, 0.002)	-0.15, 0.13	46.7
	All	1440	0.58	0.07	0.00 (0.002, 0.007) *	-0.14, 0.14	48.3
<b>Method 2</b>	L1-L2	288	0.36	0.07	-0.03 (-0.034, -0.019) *	-0.17, 0.11	77.8
	L2-L3	288	0.38	0.07	-0.03 (-0.039, -0.021) *	-0.17, 0.11	73.7
	L3-L4	288	0.39	0.07	-0.03 (-0.036, -0.020) *	-0.17, 0.11	71.8
	L4-L5	288	0.39	0.06	-0.02 (-0.026, -0.014) *	-0.14, 0.10	61.5
	L5-S1	288	0.37	0.06	-0.02 (-0.024, -0.009) *	-0.14, 0.10	64.9
	All	1440	0.38	0.07	-0.02 (-0.033, -0.018) *	-0.16, 0.12	73.7
<b>Method 3</b>	L1-L2	288	0.29	0.04	-0.01 (-0.015, -0.006) *	-0.09, 0.07	55.2
	L2-L3	288	0.3	0.04	-0.01 (-0.019, -0.010) *	-0.09, 0.07	53.3
	L3-L4	288	0.32	0.04	-0.01 (-0.014, -0.005) *	-0.09, 0.07	50.0
	L4-L5	288	0.3	0.03	0.00 (-0.009, -0.001) *	-0.06, 0.06	40.0
	L5-S1	288	0.27	0.04	0.00 (-0.003, 0.006)	-0.08, 0.08	59.3
	All	1440	0.3	0.04	-0.01 (-0.021, -0.009) *	-0.09, 0.07	53.3
<b>Method 4</b>	L1-L2	288	0.29	0.03	-0.01 (-0.013, -0.005) *	-0.07, 0.05	41.4
	L2-L3	288	0.9	0.09	-0.02 (-0.033, -0.013) *	-0.20, 0.16	40.0
	L3-L4	288	0.31	0.03	-0.01 (-0.012, -0.006) *	-0.07, 0.05	38.7
	L4-L5	288	0.31	0.03	0.00 (-0.008, 0.002)	-0.06, 0.06	38.7
	L5-S1	288	0.3	0.04	-0.01 (-0.016, -0.008) *	-0.09, 0.07	53.3
	All	1440	0.42	0.05	-0.01 (-0.018, -0.008) *	-0.11, 0.09	47.6
<b>Method 5</b>	L1-L2	288	0.36	0.13	-0.02 (-0.037, -0.008) *	-0.27, 0.23	138.9
	L2-L3	288	0.37	0.08	-0.02 (-0.030, -0.012) *	-0.18, 0.14	86.5
	L3-L4	288	0.37	0.06	-0.01 (-0.021, -0.078) *	-0.13, 0.11	64.9
	L4-L5	288	0.36	0.05	-0.01 (-0.012, 0.003)	-0.11, 0.09	55.6
	All	1440	0.37	0.09	-0.02 (-0.021, -0.008) *	-0.20, 0.16	97.3
<b>Method 6</b>	L1-L2	288	0.34	0.04	-0.01 (-0.012, -0.003) *	-0.10, 0.06	47.1
	L2-L3	288	0.35	0.04	-0.01 (-0.013, -0.005) *	-0.10, 0.06	45.7
	L3-L4	288	1.41	0.06	0.00 (-0.003, 0.010)	-0.12, 0.12	17.0
	L4-L5	288	0.4	0.04	0.00 (-0.009, 0.001)	-0.08, 0.08	40.0
	All	1440	0.62	0.04	-0.00 (-0.010, 0.003)	-0.08, 0.08	25.8
<b>Method 7</b>	L1-L2	288	0.27	0.04	0.00 (-0.002, 0.006)	-0.08, 0.08	59.3
	L2-L3	288	0.28	0.02	0.01 (0.003, 0.008) *	-0.03, 0.05	28.6
	L3-L4	288	0.29	0.02	0.00 (-0.003, 0.003)	-0.04, 0.04	27.6
	L4-L5	288	0.31	0.05	0.00 (0.001, 0.010) *	-0.10, 0.10	64.5
	L5-S1	288	0.3	0.03	0.00 (-0.006, 0.001)	-0.06, 0.06	40.0
	All	1440	0.29	0.03	0.00 (0.002, 0.010) *	-0.06, 0.06	41.4

495 N: number of levels; SD: standard deviation; CI: confidential intervals; LOA: Limits of Agreement

496 \* Bias was considered present if the 95% CI did not include zero.

497

498

499 **Table 4** Inter-rater measures' reliability results

Measurement	Level	N	Inter-rater_ICC (95% CI)
Method 1	L1-L2	288	0.956 (0.945, 0.965)
	L2-L3	288	0.877 (0.847, 0.901)
	L3-L4	288	0.972 (0.964, 0.977)
	L4-L5	288	0.844 (0.807, 0.874)
	L5-S1	288	0.852 (0.817, 0.881)
	All	1440	0.927 (0.919, 0.934)
Method 2	L1-L2	288	0.641 (0.568, 0.705)
	L2-L3	288	0.620 (0.543, 0.686)
	L3-L4	288	0.693 (0.628, 0.749)
	L4-L5	288	0.779 (0.729, 0.821)
	L5-S1	288	0.849 (0.813, 0.878)
	All	1440	0.736 (0.712, 0.759)
Method 3	L1-L2	288	0.718 (0.657, 0.770)
	L2-L3	288	0.781 (0.731, 0.822)
	L3-L4	288	0.980 (0.975, 0.984)
	L4-L5	288	0.866 (0.834, 0.892)
	L5-S1	288	0.888 (0.860, 0.910)
	All	1440	0.936 (0.930, 0.942)
Method 4	L1-L2	288	0.954 (0.942, 0.963)
	L2-L3	288	0.938 (0.922, 0.950)
	L3-L4	288	0.977 (0.972, 0.982)
	L4-L5	288	0.872 (0.841, 0.897)
	L5-S1	288	0.824 (0.783, 0.858)
	All	1440	0.984 (0.982, 0.985)
Method 5	L1-L2	288	0.500 (0.409, 0.582)
	L2-L3	288	0.726 (0.666, 0.776)
	L3-L4	288	0.728 (0.669, 0.778)
	L4-L5	288	0.761 (0.708, 0.805)
	All	1152	0.634 (0.598, 0.667)
Method 6	L1-L2	288	0.951 (0.938, 0.961)
	L2-L3	288	0.847 (0.811, 0.877)
	L3-L4	288	0.862 (0.829, 0.889)
	L4-L5	288	0.886 (0.859, 0.909)
	All	1152	0.995 (0.995, 0.996)
Method 7	L1-L2	288	0.959 (0.949, 0.968)
	L2-L3	288	0.891 (0.864, 0.913)
	L3-L4	288	0.927 (0.909, 0.942)
	L4-L5	288	0.840 (0.802, 0.871)
	L5-S1	288	0.867 (0.835, 0.893)
	All	1440	0.916 (0.908, 0.924)

500 N: number of levels; ICC: Intraclass correlation coefficient; 95% CI: 95% confidential intervals

501

502

503

504 **ESM\_1: The protocol of disc height index (DHI) measurement**

505 Patients will have erect, standing lateral (LR) and anteroposterior (AP) X-rays taken at various time points (per  
506 study schema). Hard copies of X-rays will be printed and will also be stored as digital images from the  
507 *InteleViewer<sup>TM</sup>* diagnostic imaging software/PACS system of the radiology provider. Care will be taken to  
508 maintain the aspect ratio of the digital images.

509 **(1) Standing Position of Patients**

510 The patient is naturally standing up, looking horizontally, hands resting on a vertical support, upper limbs relaxed,  
511 elbows half bent.

512 The projection: The projection direction is perpendicular to the third vertebral body of the lumbar spine.

513 **(2) The Quality of Standing Lateral X-ray Images of Lumbar Spine**

514 The entire lumbar spine should be visible from T12-L1 - L5-S1 (superior to include the T12-L1, inferior to include  
515 the sacrum, anterior to include the anterior border of the lumbar vertebral bodies, and posterior to include all  
516 elements of the posterior column, particularly the spinous processes). Superimposition of the greater sciatic  
517 notches, the superior articulating facets and the superior and inferior endplates. This indicates a true lateral has  
518 been achieved. Adequate image penetration and image contrast is evident by clear visualization of lumbar  
519 vertebral bodies, with both trabecular and cortical bone demonstrated.

520 **(3) Measurement Methods**

521 The contour of the vertebral body and upper and lower endplate of adjacent vertebral will be indicated by lines.

522 The vertebral corners could be located and confirmed.

523 A quadrilateral will be drawn to define the vertebral corners and minimize the affection of osteophytes.

524 A line is drawn cross the potential points of each corner which can be caused by the vertebral rotation for inexact  
525 body position during the scan and the anatomy deformity (such as scoliosis, vertebral rotation, and vertebral  
526 fracture).

527 Mid-point of the line will be identified as the real vertebral corner.

528 **(4) Potential Bias**

529 In order to reduce the potential bias due to difference of equipment and software, Apple MacBook with integrated  
530 touchpads/Computer with Microsoft Windows and the *InteleViewer<sup>TM</sup>* diagnostic imaging software/PACS system  
531 will be used for measurement.

532

533



534 ESM\_2\_ **Table 1** Inter-rater measures' reliability results of disc height index (DHI) on all segmental levels in  
535 degeneration and no degeneration group

Measurement method		Number of Levels	ICC (95% CI)
Method 1	L1-L2		
	Degeneration	2	-
	No degeneration	276	0.957 (0.946, 0.966)
	L2-L3		
	Degeneration	278	0.879 (0.849, 0.903)
	No degeneration	0	
	L3-L4		
	Degeneration	36	0.969 (0.940, 0.984)
	No degeneration	242	0.972 (0.965, 0.979)
	L4-L5		
Degeneration	160	0.941 (0.920, 0.956)	
No degeneration	118	0.888 (0.843, 0.921)	
Method 2	L5-S1		
	Degeneration	169	0.877 (0.836, 0.907)
	No degeneration	109	0.789 (0.706, 0.851)
	L1-L2		
	Degeneration	2	-
	No degeneration	276	0.659 (0.586, 0.721)
	L2-L3		
	Degeneration	278	0.629 (0.552, 0.695)
	No degeneration	0	
	L3-L4		
Degeneration	36	0.889 (0.794, 0.942)	
No degeneration	242	0.641 (0.560, 0.709)	
L4-L5			
Degeneration	160	0.805 (0.743, 0.853)	
No degeneration	118	0.795 (0.718, 0.853)	
Method 3	L5-S1		
	Degeneration	169	0.859 (0.813, 0.894)
	No degeneration	109	0.830 (0.761, 0.881)
	L1-L2		
	Degeneration	2	-
	No degeneration	276	0.729 (0.669, 0.780)
	L2-L3		
	Degeneration	278	0.789 (0.740, 0.829)
	No degeneration	0	
	L3-L4		
Degeneration	36	0.996 (0.992, 0.998)	
No degeneration	242	0.962 (0.951, 0.970)	
L4-L5			
Degeneration	160	0.871 (0.828, 0.904)	
No degeneration	118	0.852 (0.793, 0.895)	
Method 3	L5-S1		
	Degeneration	169	0.896 (0.861, 0.922)
No degeneration	109	0.875 (0.822, 0.913)	

Method 4	L1-L2		
	Degeneration	2	-
	No degeneration	276	0.956 (0.945, 0.965)
	L2-L3		
	Degeneration	278	0.942 (0.927, 0.954)
No degeneration	0		
L3-L4	Degeneration	36	0.885 (0.786, 0.940)
	No degeneration	242	0.982 (0.977, 0.986)
L4-L5	Degeneration	160	0.933 (0.910, 0.951)
	No degeneration	118	0.904 (0.864, 0.932)
L5-S1	Degeneration	169	0.867 (0.824, 0.900)
	No degeneration	109	0.766 (0.755, 0.824)
Method 5	L1-L2		
	Degeneration	2	-
	No degeneration	276	0.500 (0.409, 0.582)
	L2-L3		
	Degeneration	278	0.727 (0.666, 0.778)
No degeneration	0		
L3-L4	Degeneration	36	0.713 (0.505, 0.843)
	No degeneration	242	0.744 (0.682, 0.796)
L4-L5	Degeneration	160	0.768 (0.695, 0.824)
	No degeneration	118	0.763 (0.715, 0.795)
Method 6	L1-L2		
	Degeneration	2	-
	No degeneration	276	0.854 (0.819, 0.883)
	L2-L3		
	Degeneration	278	0.727 (0.666, 0.778)
No degeneration	0		
L3-L4	Degeneration	36	0.857 (0.737, 0.924)
	No degeneration	242	0.863 (0.827, 0.892)
L4-L5	Degeneration	160	0.941 (0.920, 0.956)
	No degeneration	118	0.930 (0.900, 0.951)
Method 7	L1-L2		
	Degeneration	2	-
	No degeneration	276	0.960 (0.950, 0.969)
	L2-L3		
	Degeneration	278	0.894 (0.867, 0.915)
No degeneration	0		
L3-L4	Degeneration	36	0.865 (0.751, 0.929)
	No degeneration	242	0.933 (0.915, 0.948)
L4-L5			

	Degeneration	160	0.944 (0.924, 0.959)
	No degeneration	118	0.811 (0.793, 0.846)
L5-S1	Degeneration	169	0.892 (0.856, 0.919)
	No degeneration	109	0.803 (0.725, 0.861)

536 ICC: Intraclass correlation coefficient; 95% CI: 95% confidence interval.

537

538

539

540

541

542

543

544

545

546

547

548

549

550

551

552

553

554

555

556

557 ESM\_2\_ **Table 2** Inter-rater measures' reliability results of related lines to each measurement method (method  
 558 could be used on all segment levels and ICCs are good-to-excellent) on disc height index (DHI) of all segmental  
 559 levels

Measurement method		Number of Levels	ICC (95% CI)
Method 1	L1-L2 Anterior disc height Posterior disc height Disc diameter	288	0.989 (0.986, 0.991)
		288	0.954 (0.943, 0.964)
		288	0.864 (0.855, 0.871)
	L2-L3 Anterior disc height Posterior disc height Disc diameter	288	0.967 (0.959, 0.974)
		288	0.954 (0.942, 0.963)
		288	0.867 (0.859, 0.874)
	L3-L4 Anterior disc height Posterior disc height Disc diameter	288	0.962 (0.952, 0.970)
		288	0.955 (0.944, 0.964)
		288	0.866 (0.858, 0.873)
	L4-L5 Anterior disc height Posterior disc height Disc diameter	288	0.958 (0.947, 0.966)
		288	0.939 (0.924, 0.952)
		288	0.963 (0.853, 0.870)
	L5S1 Anterior disc height Posterior disc height Disc diameter	288	0.956 (0.945, 0.965)
		288	0.852 (0.817, 0.881)
		288	0.859 (0.848, 0.867)
Method 3	L1-L2 Mid-disc height Disc diameter	288	0.930 (0.912, 0.944)
		288	0.964 (0.955, 0.971)
	L2-L3 Mid-disc height Disc diameter	288	0.913 (0.892, 0.931)
		288	0.967 (0.959, 0.974)
	L3-L4 Mid-disc height Disc diameter	288	0.930 (0.912, 0.944)
		288	0.966 (0.958, 0.973)
	L4-L5 Mid-disc height Disc diameter	288	0.949 (0.937, 0.960)
		288	0.963 (0.953, 0.970)
	L5S1 Mid-disc height Disc diameter	288	0.942 (0.927, 0.953)
		288	0.959 (0.948, 0.967)
Method 4	L1-L2 Anterior disc height Mid-disc height Posterior disc height Sagittal vertebral diameter	288	0.989 (0.986, 0.991)
		288	0.930 (0.912, 0.944)
		288	0.954 (0.943, 0.964)
		288	0.877 (0.870, 0.881)
	L2-L3 Anterior disc height Mid-disc height	288	0.967 (0.959, 0.974)
		288	0.913 (0.892, 0.931)

	Posterior disc height	288	0.954 (0.942, 0.963)
	Sagittal vertebral diameter	288	0.883 (0.866, 0.879)
	L3-L4		
	Anterior disc height	288	0.962 (0.952, 0.970)
	Mid-disc height	288	0.930 (0.912, 0.944)
	Posterior disc height	288	0.955 (0.944, 0.964)
	Sagittal vertebral diameter	288	0.873 (0.866, 0.879)
	L4-L5		
	Anterior disc height	288	0.958 (0.947, 0.966)
	Mid-disc height	288	0.949 (0.937, 0.960)
	Posterior disc height	288	0.939 (0.924, 0.952)
	Sagittal vertebral diameter	288	0.873 (0.866, 0.879)
	L5S1		
	Anterior disc height	288	0.956 (0.945, 0.965)
	Mid-disc height	288	0.942 (0.927, 0.953)
	Posterior disc height	288	0.852 (0.817, 0.881)
	Sagittal vertebral diameter	288	0.869 (0.861, 0.876)
Method 7	L1-L2		
	Anterior disc height	288	0.989 (0.986, 0.991)
	Posterior disc height	288	0.954 (0.943, 0.964)
	Superior disc depth	288	0.967 (0.958, 0.974)
	Inferior disc depth	288	0.963 (0.954, 0.971)
	L2-L3		
	Anterior disc height	288	0.967 (0.959, 0.974)
	Posterior disc height	288	0.954 (0.942, 0.963)
	Superior disc depth	288	0.967 (0.958, 0.974)
	Inferior disc depth	288	0.994 (0.992, 0.995)
	L3-L4		
	Anterior disc height	288	0.962 (0.952, 0.970)
	Posterior disc height	288	0.955 (0.944, 0.964)
	Superior disc depth	288	0.965 (0.956, 0.972)
	Inferior disc depth	288	0.961 (0.951, 0.969)
	L4-L5		
	Anterior disc height	288	0.958 (0.947, 0.966)
	Posterior disc height	288	0.939 (0.924, 0.952)
	Superior disc depth	288	0.964 (0.955, 0.971)
	Inferior disc depth	288	0.959 (0.948, 0.967)
	L5S1		
	Anterior disc height	288	0.956 (0.945, 0.965)
	Posterior disc height	288	0.852 (0.817, 0.881)
	Superior disc depth	288	0.965 (0.956, 0.972)
	Inferior disc depth	288	0.944 (0.930, 0.955)

560 ICC: Intraclass correlation coefficient; 95% CI: 95% confidence interval.

561 Anterior disc height: The shortest distance between the anterior edges of the neighbouring endplate will be

562 recorded as the anterior disc height; Mid-disc height: The mid-disc height between the upper and lower

563 bisection points is measured at the midpoint of vertebrae; Posterior disc height: The shortest distance between

564 the posterior edges of the neighbouring endplate will be recorded as the posterior disc height; Disc diameter:  
565 The disc diameter will be measured between the midpoints of the lines drawn from the endpoints of the superior  
566 vertebral endplate to the inferior. e: The sagittal diameter of the vertebral body from the anterior to posterior  
567 margin will be measured at the mid-vertebral level.

568

569

570

571

572

573

574

575

576

577

578

579

580

581

582

583

584

585

586

587

588

589

590

591

592

593

594

595

596

597

598

599

600

601 ESM\_2\_ **Table 3** Potential factors for the outliers (out of the 95% Limits of Agreement (LOA) on the Bland and  
602 Altman plot)

Method	Level	Number of outliers	Potential factor
Method 1	L1-L2	12	9 outliers of nucleus pulposus degeneration, 3 outliers of disc herniation
	L2-L3	19	15 outliers of nucleus pulposus degeneration, 4 outliers of disc herniation
	L3-L4	10	8 outliers of nucleus pulposus degeneration, 1 outlier of disc herniation, 1 normal
	L4-L5	2	2 outliers of nucleus pulposus degeneration
	L5-S1	14	7 outliers of nucleus pulposus degeneration, 6 outliers of disc herniation, 1 normal
	All	57	41 outliers of nucleus pulposus degeneration, 14 outliers of disc herniation, 2 normal
Method 2	L1-L2	15	9 outliers of nucleus pulposus degeneration, 5 outliers of disc herniation, 1 normal
	L2-L3	15	10 outliers of nucleus pulposus degeneration, 5 outliers of disc herniation
	L3-L4	13	10 outliers of nucleus pulposus degeneration, 3 outliers of disc herniation
	L4-L5	10	8 outliers of nucleus pulposus degeneration, 2 outliers of disc herniation
	L5-S1	12	6 outliers of nucleus pulposus degeneration, 5 outliers of disc herniation, 1 normal
	All	65	43 outliers of nucleus pulposus degeneration, 20 outliers of disc herniation, 2 normal
Method 3	L1-L2	37	21 outliers of nucleus pulposus degeneration, 15 outliers of disc herniation, 1 normal
	L2-L3	26	20 outliers of nucleus pulposus degeneration, 6 outliers of disc herniation
	L3-L4	21	16 outliers of nucleus pulposus degeneration, 3 outliers of disc herniation, 2 normal
	L4-L5	45	38 outliers of nucleus pulposus degeneration, 13 outliers of disc herniation, 4 normal
	L5-S1	42	24 outliers of nucleus pulposus degeneration, 17 outliers of disc herniation, 1 normal
	All	171	109 outliers of nucleus pulposus degeneration, 54 outliers of disc herniation, 8 normal
Method 4	L1-L2	57	34 outliers of nucleus pulposus degeneration, 17 outliers of disc herniation, 6 normal
	L2-L3	5	4 outliers of nucleus pulposus degeneration, 1 outlier of disc herniation
	L3-L4	42	31 outliers of nucleus pulposus degeneration, 7 outliers of disc herniation, 4 normal
	L4-L5	44	29 outliers of nucleus pulposus degeneration, 11 outliers of disc herniation, 4 normal
	L5-S1	34	18 outliers of nucleus pulposus degeneration, 15 outliers of disc herniation, 1 normal
	All	182	116 outliers of nucleus pulposus degeneration, 51 outliers of disc herniation, 15 normal
Method 5	L1-L2	1	1 outlier of nucleus pulposus degeneration
	L2-L3	12	6 outliers of nucleus pulposus degeneration, 6 outliers of disc herniation
	L3-L4	10	6 outliers of nucleus pulposus degeneration, 4 outliers of disc herniation

	L4-L5	14	8 outliers of nucleus pulposus degeneration, 6 outliers of disc herniation
	All	37	21 outliers of nucleus pulposus degeneration, 16 outliers of disc herniation
Method 6	L1-L2	11	5 outliers of nucleus pulposus degeneration, 6 outliers of disc herniation
	L2-L3	10	7 outliers of nucleus pulposus degeneration, 3 outliers of disc herniation
	L3-L4	13	9 outliers of nucleus pulposus degeneration, 4 outliers of disc herniation
	L4-L5	8	6 outliers of nucleus pulposus degeneration, 2 outliers of disc herniation
	All	42	27 outliers of nucleus pulposus degeneration, 15 outliers of disc herniation
Method 7	L1-L2	6	5 outliers of nucleus pulposus degeneration, 1 outlier of disc herniation
	L2-L3	17	14 outliers of nucleus pulposus degeneration, 3 outliers of disc herniation
	L3-L4	17	13 outliers of nucleus pulposus degeneration, 3 outliers of disc herniation, 1 normal
	L4-L5	1	1 outlier of nucleus pulposus degeneration
	L5-S1	14	4 outliers of nucleus pulposus degeneration, 9 outliers of disc herniation, 1 normal
	All	55	37 outliers of nucleus pulposus degeneration, 16 outliers of disc herniation, 2 normal

603