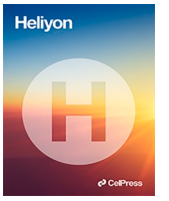
Heliyon 9 (2023) e20764

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Relation of lumbar intervertebral disc height and severity of disc degeneration based on Pfirrmann scores

Xiao-long Chen [a,*], Xiang-yu Li [a], Yu Wang [a], Shi-bao Lu [a,**]

[a] *Department of Orthopaedics, Xuanwu Hospital Capital Medical University, Xicheng District, Beijing, China*

ARTICLE INFO

Keywords:
Lumbar spine
Intervertebral disc degeneration
Disc height
Magnetic resonance image
Agreement
Reliability

ABSTRACT

Background: Disc height (DH) change is considered one of the most critical factors in assessing intervertebral disc degeneration (IVD). Pfirrmann et al. developed a scoring system for disc degeneration evaluation based on changes in DH in magnetic resonance imaging (MRI). While the relationship between DH measurements and Pfirrmann scores for disc degeneration has been explored, the validity of different DH measuring techniques or their connection with disc degeneration is yet uncertain. The present study investigates intra-rater and inter-rater agreement and reliability of different DH measurement methods on MRI and evaluates the relationship between different DH measurement methods and Pfirrmann scores of IVD degeneration, as well as between different Pfirrmann scores and clinical outcomes.

Methods: Adult patients with MRI scans of the lumbar spine were recruited. Eight DH measuring techniques were tested for intra-rater and inter-rater agreement and reliability. Bland and Altman's Limits of Agreement (LOA) was used to evaluate intra-rater and inter-rater agreements. Intra-rater and inter-rater reliability were evaluated using intra-class correlations (ICC) with 95 % confidence intervals (95 % CI). The association between DH and Pfirrmann scores was examined using one-way ANOVA.

Results: Excellent intra-rater reliability was reported for 332 participants on DH (ranging from 0.912 (0.901, 0.923) to 0.973 (0.964, 0.981) and from 0.902 (0.892, 0.915) to 0.975 (0.962, 0.985) by two independent raters). All measuring methods had high intra-rater agreement, except for methods 4 and 5. All methods had good-to-excellent of inter-rater reliability on DH (ICCs ranging from 0.812 (0.795, 0.828) to 0.995 (0.994, 0.995)) except for the posterior disc material length of method 5 (ICC 0.740 (0.718, 0.761)). Methods 1 to 6 for evaluating DH in patients with spondylolisthesis had poor inter-rater reliability. The IVD levels with grades IV and V in Pfirrmann scores had significantly lower DH than the IVD levels with grades I to III in Pfirrmann scores. IVD levels with grades IV and V in Pfirrmann scores had significantly higher VAS and ODI than IVD levels with grades I in Pfirrmann scores.

Conclusion: A good-to-excellent intra-rater and inter-rater reliability was achieved on most DH measuring methods on MRI following a standardized and structured protocol. However, small anatomical structures and different tissue borders could influence measurements. Additionally, DH can differentiate between grade IV and V Pfirrmann scores, and severe IVD degeneration (IV and V Pfirrmann) is linked to clinical outcomes.

* Corresponding author.
** Corresponding author.
E-mail addresses: chensmalldragon@163.com (X.-l. Chen), spinelu@163.com (S.-b. Lu).

https://doi.org/10.1016/j.heliyon.2023.e20764
Received 4 September 2022; Received in revised form 27 September 2023; Accepted 5 October 2023
Available online 6 October 2023

1. Introduction

Low back pain (LBP) is the leading cause of disability and lost productivity globally, posing a significant economic burden [1]. Health expenditure for LBP has been estimated to be up to $A5 billion in Australia and US$100 per year in the United States [2]. Identifying the potential aetiology of this condition and implementing the most effective treatment are urgent priorities to lessen the global burden of LBP.

Previous research has linked intervertebral disc (IVD) degeneration to LBP [3]. The hallmarks of disc degeneration include the structural alterations of the annulus fibrosis, nucleus pulposus, endplate, and subchondral bone due to the changes in fluid, proteoglycans, and collagen [4]. Radiographic changes supportive of IVD degeneration include the loss of disc height, formation of osteophytes leading to constriction of articular facets and intervertebral foramen, and changes in characteristics and location of disc material [5,6]. However, IVD degeneration has proven to be a difficult entity to investigate; its description is ambiguous, with different radiographic metrics that are difficult to classify and quantify. In theory, changes in disc height (DH) could reflect changes in the microstructure of IVD. In addition, a radiological investigation found a potential relationship between the change in lumbar DH and IVD degeneration [7]. Changes in DH could cause a redistribution of the loading on the lumbar spine because of the reduction in the function of IVD as a cushion and shock absorber. Furthermore, narrowing and thinning of IVD space have been linked to pain intensity, chronic pain, and lumbar spine impairments [8]. However, other investigations revealed that only major pathologic changes in DH may be utilized to diagnose IVD degeneration [9–11], and different techniques of assessing DH yielded different results [7,9,12–17]. Therefore, a reproducible DH measurement is required.

Intra-rater with inter-rater reliability refers to the extent to which two or more raters agree on addressing the consistency of implementation of a rating system. The statistical methods provide a score indicating the degree of homogeneity, or consensus, in the assessments given by judges. Bland and Altman's Limits of Agreement (LOA) was recently the most popular [18], recommending a statistical approach for evaluating agreement [19,20]. As a result, the application of reliability with LOA agreement is required for evaluating the accuracy and efficacy measurement for DH.

There is mounting evidence of different measuring methods based on different radiological techniques, including X-ray, computed tomography (CT), and magnetic resonance imaging (MRI) [7,9,12–17,21,22]. Clinicians frequently described and interpreted the images based on their experiences and information from literature and practice. MRI outperforms radiography and CT scans in revealing the morphologic characteristics of IVDs with good imaging of anatomical structures without ionizing radiation [22], including IVD height [7], IVD volume [23], and disc degeneration [10]. Pfirrmann et al. [10] developed a scoring system to evaluate disc degeneration based on changes in DH in MRI. Furthermore, the relationship between DH and disc degeneration classification by Pfirrmann scores has been investigated [11]. However, the scoring system does not specify the measurement methods. Previous research has shown that varying intra-rater and inter-rater reliability results in different disc height index measurement methods on X-rays can alter Pfirrmann scores [21]. While several related studies have reported potential bias on different measurement methods, research on the validity of different measurement methods on DH or the relationship with disc degeneration based on different DH measurement methods is lacking. Therefore, it is intriguing to determine whether the DH measured using different measurement methods differs in the different grades of IVD degeneration in Pfirrmann scores [10] and whether the DH change in different Pfirrmann scores varies in different clinical outcomes (e.g., LBP and disability).

In this view, the current study aims to: 1) evaluate intra-rater and inter-rater agreement and reliability of different DH measurement methods on MRI; 2) investigate the relationship between different DH measurement methods and Pfirrmann scores of IVD degeneration; and 3) assess the relationship between different Pfirrmann scores and clinical outcomes (e.g., LBP and disability).

2. Materials and methods

2.1. Participants

The Human Research Ethics Committee of Xuanwu Hospital Capital Medical University approved this study. Adult participants (age $\geq$18 years) who had routine MRI scans of the lumbar spine from September 2018 onwards (n = 332) were enrolled in this retrospective investigation. Participants who underwent lumbar spine surgery before MRI scan and/or diagnosed with scoliosis (defined as a lateral deviation of the spine from the normal plumb line with a Cobb angle $\geq$10° for the magnitude of curve in plain radiography) and/or vertebral anomaly (the deficiency presumably occurs during vertebral somite formation, including wedge vertebra, hemivertebra, fusion of the vertebrae, transitional vertebrae, and butterfly vertebra, etc.) were excluded from the study. All participants consented to use their demographic data, clinical outcomes, and radiological data for research purposes.

2.2. Clinical outcome

Two questionnaires were used to assess pain intensity by Visual Analogue Scale (VAS: 0 - no pain; 10 - worst pain imaginable) and function and disability by Oswestry disability index (ODI: a validated tool on 10-items, each item was manually rated with 5 points for six possible responses (the first statement is score 0; the last statement is score 5), giving a potential score between 0 % (normal) to 100 % (severe disability)) during the recruitment. All data were collected and stored in the Xuanwu Hospital Capital Medical University by an experienced spine surgeon (YW).

2.3. MRI acquisition

All participants had their lumbar spine scanned with a 3.0 T Trio Tim scanner (Siemens, Erlangen, Germany). MRI yielded sagittal T1-weighted, sagittal T2-weighted, and axial T2-weighted images. The field of view (FOV), repetition time (TR)/echo time (TE), matrix size, slice thickness, slice per slab, and number of excitations (NEX) during the sagittal T1-weighted scan were 310 * 310 mm, 550 ms/9.6 ms, 320 * 320, 4.0 mm, 11, and 2, respectively. The FOV, the TR/TE, matrix size, slice thickness, slice per slab, and NEX during the sagittal T2-weighted scan were 310 * 310 mm, 2700 ms/97 ms, 320 * 320, 4.0 mm, 11, and 2, respectively. The FOV, the TR/TE, matrix size, slice thickness, slice per slab, and NEX during the axial T2-weighted scan were 210 * 210 mm, 3400 ms/102 ms, 320 * 320, 4.0 mm, 15, and 2, respectively. The research team selected the mid-sagittal section of the T2-weighted slice for DH assessment. The DH was measured using an Apple MacBook with integrated touchpads and the Philips DICOM Viewer (Philips, Best, the Netherlands) to reduce potential bias.

2.4. Measurements

The DH/DHI was evaluated using a mid-sagittal T2WI slice of the lumbar spine. Each segmental level of the lumbar spine (defined

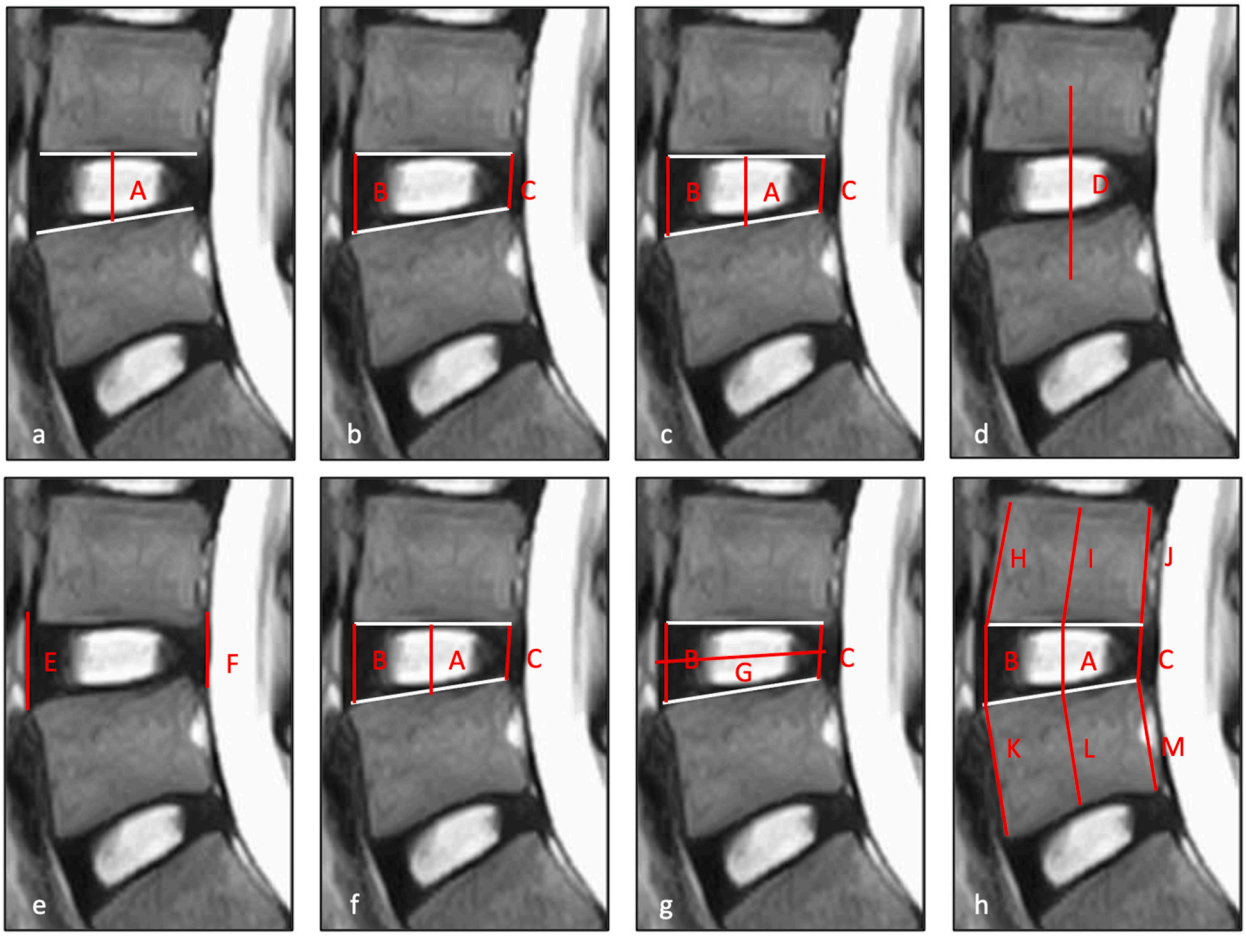


Fig. 1. The details of disc height (DH) and disc height index (DHI) measurements. (a) Method 1 of DH; (b) Method 2 of DH; (c) Method 3 of DH; (d) Method 4 of DH; (e) Method 5 of DH; (f) Method 6 of DH; (g) Method 7 of DHI; (h) Method 8 of DHI. Note: A: The mid-disc height between the upper and lower bisection points is measured at the midpoint of vertebrae; B: The shortest distance between the anterior edges of the neighbouring endplate will be recorded as the anterior disc height; C: The shortest distance between the posterior edges of the neighbouring endplate will be recorded as the posterior disc height; D: The distance between the centroids of upper and lower vertebral body; E: The distance between anterior and posterior boundaries of anterior herniated disc material; F: The distance between anterior and posterior boundaries of posterior herniated disc material; G: The disc diameter will be measured between the midpoints of the lines drawn from the endpoints of the superior vertebral endplate to the inferior; H–J: The proximal vertebral body height will be measured from the anterior (H), middle (I), and posterior (J) portions of each respective disc level; K–M: The distal vertebral body height will be measured from the anterior (K), middle (L), and posterior (M) portions of each respective disc level. According to the classification of related lines, line B, C, E, F, H, J, K, and M are defined as direct lines, and line A, D, G, I, and L are defined as indirect lines.

as IVD level from L1-L2 to L5-S1) was assessed using six DH methods [7,12,13,15–17] and two-disc height index (DHI) methods [7,9]. The protocol for measuring the DH and DHI is outlined in Supplementary Material 1 and **Figure [1(a–h)]**.

Method 1 of DH is defined as the midpoint IVD height [12]. Method 2 of DH is defined as the mean of the anterior and posterior IVD height [7]. Method 3 of DH is defined as anterior, middle, and posterior IVD height, respectively [13]. Method 4 of DH is defined as the distance between the centroids of the upper and lower vertebral body of the IVD [15]. Method 5 of DH is defined as the anterior and posterior IVD material length [16]. Method 6 of DH is defined as the mean of the anterior, middle, and posterior IVD heights [17].

Method 7 of DHI is defined as the ratio of total anterior and posterior IVD height to disc diameter [7]. Method 8 of DHI is defined as the ratio of the mean of anterior, middle, and posterior IVD height to the mean of proximal and distal vertebral body height [9].

The direct line was defined as a line that may be directly traced on the vertebral body. The indirect line was defined as the line that should be drawn as a crossing direct line(s).

The IVD degeneration was evaluated using the five categories of Pfirrmann's score [10]. Pfirrmann grade $\geq$ III was defined as disc degeneration [24]. Two raters (XC and YW) independently assessed the disc degeneration. The third rater (SL) settled the discrepancy between the two raters.

2.5. Training

Two researchers (XYL as rater 1 and YW as rater 2) performed the DH and DHI measurements. Both raters are experienced spine surgeons and back pain experts. Twenty participants were selected for measuring training. The intra- and inter-rater agreement was assessed between two out of six DH and two DHI measurements performed by each rater. The inter-rater reliability was assessed between two raters who were purposely selected to represent a novice and an expert radiological image interpretation. All measuring techniques and results were blinded to each other. There were two weeks intervals between the first and second measurement sessions to reduce potential bias.

2.6. Statistical analysis

Intra-rater and inter-rater reliability were assessed using intra-class correlation coefficient (ICC) and their 95 % confidence

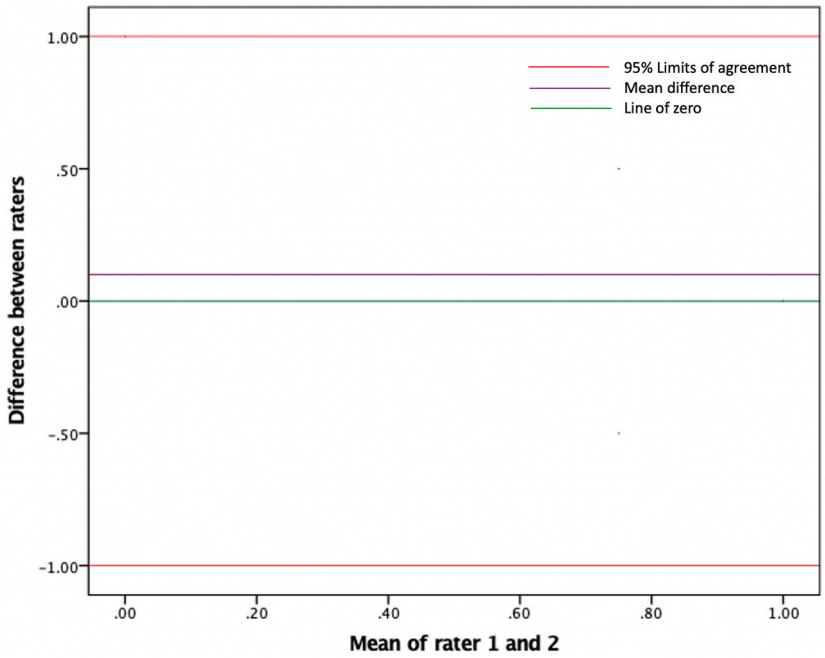


Fig. 2. The Bland and Altman plot of Limits of Agreement (LOA) between two raters on the different measurements for disc height (DH). The y-axis shows the mean difference (MD) between raters' measurements, and the x-axis shows the mean value of both raters' measurements. The LOA plot will show the relationship between mean values and differences between rater 1 and rater 2 on the measurements of DH using two out of previously reported methods. MD with 95 % confidence intervals (CI) of the measurements between rater 1 and rater 2 was reported to describe the precision of the bias. The green line shows the range of mean difference including zero. The purple line shows the MD between measurements. Red lines show the 95 % (LOA), between which 95 % of all measurement differences are located. If the 95 % CI doesn't include zero, it can be assumed that there is a bias. Furthermore, LOA was presented as a proportion of mean values for each method. The proportion will be calculated as follows: ((upper LOA +(−1*(lower LOA)))/(the mean)) *100 %. Following previously published data, we consider percentages lower than 50 % as an indicator of acceptable precision. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

intervals (95 % CI). Excellent, good, moderate, and poor reliability were defined as ICC values greater than 0.90, between 0.75 and 0.9, between 0.5 and 0.75, and less than 0.5, respectively [25]. Different diagnoses of included participants (e.g., lumbar disc herniation, lumbar spinal stenosis, and lumbar spondylolisthesis (defined as one vertebra over subjacent vertebra without an associated disruption or defect in the vertebral ring)), different lines, and different levels were employed for subgroup analysis (see Fig. 1).

Bland and Altman's Limits of Agreement (LOA) is recommended as a standard approach for assessing agreement between two measuring methods [18–20,26]. It is imperative to investigate intra-rater and inter-rater agreement on different DH measurement methods employing MRI using LOA with standard error of measurement (SEM). LOA is depicted as the mean differences (MDs) between two measuring methods with the SD of the differences (Fig. 2). The precision of bias was presented as 95 % CI of MDs. The bias is referred to as zero, not in 95 % CI. The proportion of LOA was calculated formulas follows: ((upper LOA +(−1*(lower LOA)))/(the mean)) *100 %. A proportion lower than 50 % was considered acceptable based on previously published data [16]. The variables responsible for the deviation of the results out of the LOA range have been described.

The correlations between DH/DHI and Pfirrmann scores and between Pfirrmann scores and clinical scores (VAS LBP and ODI) were investigated using one-way ANOVA. Post hoc testing with Tukey's test was performed for statistically significant ANOVA results. The strength of the relationship was assessed using linear regression. All statistical analyses were performed using SPSS statistical package version 24 (SPSS Inc, Chicago, IL). A *P* value less than 0.05 denoted statistical significance.

3. Results

There were 332 participants (142 females and 190 males; mean age 47.81 ± 16.86 years, ranging from 19 to 82 years). There are a total of 1660 lumbar spine levels in these participants. A total of 496 lumbar spine levels with IVD degeneration was found in these participants using Pfirrmann grade, including 16 IVD degeneration at the level of L1-L2, 31 at the level of L2-L3, 77 at the level of L3-L4, 187 at the level of L4-L5, and 185 at the level of L5-S1 (Table 1). The mean VAS LBP and ODI at preoperative were 4.42 ± 1.81 and 25.46 ± 10.70, respectively.

Table 1
Patient demographic and clinic-radiological information.

	Number of patients
Female: Male	142 (36.1 %): 190 (63.9 %)
Age (years)	47.81 ± 16.86
Diagnosis	
Spondylolisthesis	35 (10.5 %)
Disc herniation	68 (20.5 %)
Spinal stenosis	129 (38.9 %)
Normal	100 (30.1 %)
Intervertebral disc degeneration based on the Pfirrmann score [10]	
L1-L2	
Grade I	193 (58.1 %)
Grade II	123 (37 %)
Grade III	16 (4.9 %)
L2-L3	
Grade I	165 (49.7 %)
Grade II	136 (41 %)
Grade III	30 (9 %)
Grade IV	1 (0.3 %)
L3-L4	
Grade I	120 (36.1 %)
Grade II	135 (40.7 %)
Grade III	64 (19.3 %)
Grade IV	13 (3.9 %)
L4-L5	
Grade I	58 (17.5 %)
Grade II	87 (26.2 %)
Grade III	124 (37.3 %)
Grade IV	54 (16.3 %)
Grade V	9 (2.7 %)
L5-S1	
Grade I	55 (16.6 %)
Grade II	92 (27.7 %)
Grade III	99 (29.8 %)
Grade IV	68 (20.5 %)
Grade V	18 (5.4 %)
VAS LBP	4.42 ± 1.81
ODI	25.46 ± 10.70

VAS: Visual Analogue Scale; LBP: low back pain; ODI: Oswestry Disability Index.
Continuous data are presented as the mean ± standard deviation (SD). Dichotomous data are presented as numbers and percentages.

Table 2
Intra- and inter-rater measures' reliability results based on different measurement methods on disc height.

	Level/Diagnosis	N	Intra-rater reliability		Inter-rater_ICC (95 % CI)
			Rater 1_ICC (95 % CI)	Rater 2_ICC (95 % CI)	
Method 1	L1-L2	332			0.894 (0.870, 0.914)
	L2-L3	332			0.873 (0.844, 0.896)
	L3-L4	332			0.898 (0.875, 0.917)
	L4-L5	332			0.924 (0.907, 0.938)
	L5-S1	332			0.916 (0.896, 0.931)
	Lumbar disc herniation	340			0.900 (0.868, 0.925)
	Lumbar spinal stenosis	645			0.873 (0.851, 0.893)
	Spondylolisthesis	175			0.529 (0.448, 0.601)
	Normal	500			0.954 (0.946, 0.960)
	All	1660	0.960 (0.951, 0.972)	0.967 (0.958, 0.982)	0.902, (0.892, 0.910)
Method 2	L1-L2	332			0.971 (0.965, 0.977)
	L2-L3	332			0.953 (0.942, 0.962)
	L3-L4	332			0.940 (0.927, 0.952)
	L4-L5	332			0.926 (0.909, 0.940)
	L5-S1	332			0.929 (0.913, 0.942)
	Lumbar disc herniation	340			0.951 (0.934, 0.963)
	Lumbar spinal stenosis	645			0.977 (0.973, 0.980)
	Spondylolisthesis	175			0.556 (0.478, 0.625)
	Normal	500			0.931 (0.919, 0.942)
	All	1660	0.962 (0.947, 0.975)	0.961 (0.957, 0.971)	0.948 (0.943, 0.953)
Method 3_ Anterior disc height	L1-L2	332			0.983 (0.979, 0.986)
	L2-L3	332			0.951 (0.940, 0.960)
	L3-L4	332			0.942 (0.929, 0.953)
	L4-L5	332			0.900 (0.877, 0.919)
	L5-S1	332			0.938 (0.924, 0.950)
	Lumbar disc herniation	340			0.950 (0.933, 0.963)
	Lumbar spinal stenosis	645			0.977 (0.973, 0.980)
	Spondylolisthesis	175			0.675 (0.613, 0.729)
	Normal	500			0.930 (0.917, 0.941)
	All	1660	0.973 (0.964, 0.981)	0.975 (0.962, 0.985)	0.955 (0.950, 0.959)
Method 3_ Posterior disc height	L1-L2	332			0.946 (0.933, 0.956)
	L2-L3	332			0.933 (0.918, 0.946)
	L3-L4	332			0.934 (0.918, 0.946)
	L4-L5	332			0.929 (0.913, 0.943)
	L5-S1	332			0.873 (0.845, 0.896)
	Lumbar disc herniation	340			0.939 (0.918, 0.954)
	Lumbar spinal stenosis	645			0.971 (0.966, 0.975)
	Spondylolisthesis	175			0.479 (0.465, 0.487)
	Normal	500			0.923 (0.909, 0.935)
	All	1660	0.957 (0.942, 0.968)	0.955 (0.946, 0.962)	0.928 (0.921, 0.934)
Method 4	L1-L2	332			0.890 (0.865, 0.910)
	L2-L3	332			0.946 (0.934, 0.956)
	L3-L4	332			0.903 (0.880, 0.921)
	L4-L5	332			0.900 (0.877, 0.919)
	Lumbar disc herniation	272			0.836 (0.779, 0.880)
	Lumbar spinal stenosis	516			0.975 (0.970, 0.979)
	Spondylolisthesis	140			0.445 (0.345, 0.536)
	Normal	400			0.931 (0.898, 0.959)
	All	1328	0.921(0.913, 0.925)	0.930 (0.925, 0.938)	0.917 (0.908, 0.925)
Method 5_ Anterior disc material length	L1-L2	332			0.848 (0.814, 0.875)
	L2-L3	332			0.760 (0.711, 0.802)
	L3-L4	332			0.861 (0.830, 0.886)
	L4-L5	332			0.671 (0.607, 0.726)
	L5-S1	332			0.845 (0.811, 0.873)
	Lumbar disc herniation	340			0.800 (0.739, 0.847)
	Lumbar spinal stenosis	645			0.872 (0.852, 0.889)
	Spondylolisthesis	175			0.384 (0.283, 0.479)
	Normal	500			0.803 (0.769, 0.832)
	All	1660	0.928 (0.916, 0.935)	0.921 (0.914, 0.931)	0.812 (0.795, 0.828)
Method 5_ Posterior disc material length	L1-L2	332			0.802 (0.794, 0.820)
	L2-L3	332			0.786 (0.771, 0.802)
	L3-L4	332			0.808 (0.767, 0.842)
	L4-L5	332			0.427 (0.418, 0.440)
	L5-S1	332			0.873 (0.845, 0.897)
	Lumbar disc herniation	340			0.842 (0.793, 0.880)
	Lumbar spinal stenosis	645			0.744 (0.708, 0.777)
	Spondylolisthesis	175			0.387 (0.293, 0.474)

(*continued on next page*)

Table 2 (*continued*)

	Level/Diagnosis	N	Intra-rater reliability		Inter-rater_ICC (95 % CI)
			Rater 1_ICC (95 % CI)	Rater 2_ICC (95 % CI)	
	Normal	500			0.740 (0.698, 0.777)
	All	1660	0.912 (0.901, 0.923)	0.902 (0.892, 0.915)	0.740 (0.718, 0.761)
Method 6	L1-L2	332			0.956 (0.946, 0.964)
	L2-L3	332			0.935 (0.920, 0.948)
	L3-L4	332			0.931 (0.915, 0.944)
	L4-L5	332			0.932 (0.917, 0.945)
	L5-S1	332			0.929 (0.912, 0.942)
	Lumbar disc herniation	340			0.944 (0.926, 0.958)
	Lumbar spinal stenosis	645			0.972 (0.967, 0.976)
	Spondylolisthesis	175			0.560 (0.483, 0.629)
	Normal	500			0.918 (0.903, 0.931)
	All	1660	0.945 (0.936, 0.957)	0.927 (0.924, 0.933)	0.938 (0.932, 0.944)
Method 7	L1-L2	332			0.948 (0.936, 0.958)
	L2-L3	332			0.856 (0.825, 0.883)
	L3-L4	332			0.968 (0.960, 0.974)
	L4-L5	332			0.706 (0.647, 0.756)
	L5-S1	332			0.827 (0.790, 0.858)
	Lumbar disc herniation	340			0.912 (0.883, 0.934)
	Lumbar spinal stenosis	645			0.962 (0.956, 0.968)
	Spondylolisthesis	175			0.753 (0.688, 0.810)
	Normal	500			0.801 (0.767, 0.830)
	All	1660	0.944 (0.936, 0.951)	0.938 (0.929, 0.946)	0.897 (0.887, 0.906)
Method 8	L1-L2	332			0.938 (0.923, 0.949)
	L2-L3	332			0.824 (0.786, 0.856)
	L3-L4	332			0.846 (0.812, 0.874)
	L4-L5	332			0.810 (0.770, 0.844)
	Lumbar disc herniation	272			0.993 (0.990, 0.995)
	Lumbar spinal stenosis	516			0.997 (0.996, 0.997)
	Spondylolisthesis	140			0.993 (0.992, 0.995)
	Normal	400			0.993 (0.992, 0.995)
	All	1328	0.932 (0.922, 0.939)	0.922 (0.917, 0.931)	0.995 (0.994, 0.995)

N: number of levels; ICC: Intraclass correlation coefficient; 95 % CI: 95 % confidence intervals.
Intra-rater and inter-rater reliability were assessed by intra-class correlation coefficient (ICC) and their 95 % confidence intervals (95 % CI). The excellent, good, moderate, and poor reliability were presented as the values of ICC greater than 0.90, between 0.75 and 0.9, between 0.5 and 0.75, and less than 0.5, respectively.

3.1. Intra-rater reliability

Intra-rater reliability on different measurement methods was excellent, ranging from 0.912 (0.901, 0.923) to 0.973 (0.964, 0.981) and from 0.902 (0.892, 0.915) to 0.975 (0.962, 0.985) by two independent raters, respectively (Table 2).

3.2. Intra-rater agreement

All measurements had an excellent intra-rater agreement, except methods 4 and 5, which had bias (95 % CI of MD does not include zero) and unsatisfactory precisions (proportion of mean values $\geq$ 50 %) (Table 3).

3.3. Inter-rater reliability

Inter-rater reliability was good-to-excellent in all cases (ICCs ranging from 0.812 (0.795, 0.828) to 0.995 (0.994, 0.995)), except for the posterior disc material length of method 5 (ICC 0.740 (0.718, 0.761)) (Table 2).

Subgroup analysis was performed based on the different segmental levels. A good-to-excellent inter-rater reliability was found in all measurements at all lumbar spine levels. However, the ICCs on using method 5 for posterior disc material length and anterior disc material length at the level L4-L5 were poor (ICC: 0.427 (0.418, 0.440)) and moderate (0.671 (0.607, 0.726)) (Table 2), respectively.

Different diagnoses showed poor ICCs of DH in patients with spondylolisthesis in method 1 to method 6 (ranging from 0.384 (0.283, 0.479) to 0.675 (0.613, 0.729)), moderate in method 7 (0.753 (0.688, 0.810)), and excellent in method 8 (0.993 (0.992, 0.995)) (Table 2).

3.4. Relationship between DH (anterior, middle, and posterior DH and DHI) and Pfirrmann score

The ANOVA test revealed no significant association between DH measures (anterior, middle, and posterior DH and DHI) and grades I to III in Pfirrmann scores. However, IVD levels with grades IV and V in Pfirrmann scores had significantly lower DH than IVD levels with grades I to III in Pfirrmann scores (Table 4).

Table 3
Inter-rater measures' agreement results based on different measurement methods on disc height.

	Level	N	SD	Mean difference (95 % CI)	95 % LOA	LOA as proportion of mean values (%)
Method 1	L1-L2	332	5.82	−1.59 (−1.89, −1.31) *	−7.18, 3.80	40.9
	L2-L3	332	5.52	−1.79 (−2.13, −1.48) *	−7.43, 3.85	35
	L3-L4	332	5.82	−1.60 (−1.88, −1.30) *	−6.89, 3.69	36
	L4-L5	332	5.95	−1.40 (−1.66, −1.14) *	−6.03, 3.23	24
	L5-S1	332	2.36	−1.12 (−1.38, −0.87) *	−5.75, 3.51	46
Method 2	L1-L2	332	1.83	−0.84 (−1.05, −0.64) *	−4.42, 2.75	35
	L2-L3	332	1.69	−0.77 (−0.96, −0.06) *	−4.08, 2.54	38
	L3-L4	332	1.96	−1.04 (−1.26, −0.84) *	−4.88, 2.80	40
	L4-L5	332	2.10	−0.86 (−1.10, 0.65)	−4.98, 3,26	46
	L5-S1	332	2.29	−1.10 (−1.34, −0.86) *	−5.59, 3.39	44.5
Method 3_ADH	L1-L2	332	2.08	−0.97 (−1.22, −0.76) *	−5.05, 3.11	37
	L2-L3	332	2.17	−1.00 (−1.25, −0.79) *	−5.25, 3.25	48
	L3-L4	332	2.41	−1.23 (−1.51, −0.99) *	−5.95, 3.49	45
	L4-L5	332	2.76	−1.15 (−1.45, −0.87) *	−6.56, 4.26	35
	L5-S1	332	2.93	−1.16 (−1.48, −0.83) *	−6.90, 4.58	33
Method 3_PDH	L1-L2	332	2.27	−0.70 (−0.95, −0.49) *	−5.15, 3.75	63
	L2-L3	332	1.67	−0.55 (−0.72, −0.37) *	−3.82, 2.72	65
	L3-L4	332	1.92	−0.84 (−1.06, −0.65) *	−4.60, 2.92	72
	L4-L5	332	2.29	−0.58 (−0.82, −0.33) *	−5.07, 3.91	75
	L5-S1	332	2.34	−1.06 (−1.33, −0.81) *	−5.65, 3.53	72
Method 4	L1-L2	332	2.99	−1.37 (−1.73, −1.05) *	−7.23, 4.49	73
	L2-L3	332	2.88	−1.52 (−1.54, −1.22) *	−7.16, 4.12	80
	L3-L4	332	2.53	−1.27 (−1.55, −1.02) *	−6.23, 3.69	67
	L4-L5	332	2.76	−1.14 (−1.46, −0.87) *	−6.55, 4.27	66
Method 5_ADML	L1-L2	332	3.77	−1.70 (−2.08, −1.26) *	−9.09, 5.69	67
	L2-L3	332	4.68	−1.17 (−1.68, −0.68) *	−10.34, 8.00	78
	L3-L4	332	3.83	−1.72 (−2.14, −1.31) *	−9.23, 5.79	80
	L4-L5	332	5.31	−1.63 (−2.30, −1.09) *	−11.10, 8.78	77
	L5-S1	332	4.83	−1.04 (−1.57, 0.51)	−10.51, 8.43	66
Method 5_PDML	L1-L2	332	2.76	−0.13 (−0.44, −0.18) *	−5.53, 5.28	84
	L2-L3	332	0.77	−0.28 (−0.36, −0.21) *	−1.79, 1.23	86.5
	L3-L4	332	3.15	−1.27 (−1.58, −0.92) *	−7.44, 4.90	64.9
	L4-L5	332	6.01	−1.62 (−2.25, −0.91) *	−13.40, 10.16	84
	L5-S1	332	2.13	−0.49 (−0.72, −0.26) *	−4.66, 3.68	89
Method 6	L1-L2	332	1.98	−1.09 (−1.30, −0.87) *	−4.97, 2.79	46.1
	L2-L3	332	1.98	−1.11 (−1.32, −0.91) *	−0.10, 0.06	45.7
	L3-L4	332	2.10	−1.22 (−1.46, −1.01) *	−5.34, 2.90	35
	L4-L5	332	2.02	−1.04 (−1.27, −0.84) *	−5.00, 2.92	40
	L5-S1	332	2.21	−1.11 (−1.34, −0.87) *	−5.44, 3.22	45
Method 7	L1-L2	332	0.22	0.00 (0.00, 0.01) *	−0.08, 0.08	49.3
	L2-L3	332	0.10	0.01 (0.00, 0.01) *	−0.03, 0.05	28.6
	L3-L4	332	0.20	0.00 (0.00, 0.01) *	−0.04, 0.04	27.6
	L4-L5	332	0.14	0.00 (0.00, 0.01) *	−0.10, 0.10	44.5
	L5-S1	332	0.07	0.00 (−0.01, 0.01)	−0.14, 0.14	40
Method 8	L1-L2	332	0.12	0.00 (0.00, 0.01) *	−0.08, 0.08	44.5
	L2-L3	332	0.06	0.01 (0.00, 0.01) *	−0.03, 0.05	28.6
	L3-L4	332	0.10	0.00 (0.00, 0.01) *	−0.04, 0.04	27.6
	L4-L5	332	0.08	0.00 (0.00, 0.01) *	−0.10, 0.10	44.5

N: number of levels; SD: standard deviation; CI: confidence intervals; LOA: Limits of Agreement; ADH: anterior disc height; PDH: posterior disc height; ADML: anterior disc material length; PDML: posterior disc material length; * Bias was considered present if the 95 % CI did not include zero. The proportion lower than 50 % was regarded as an acceptable precision.

3.5. Relationship between Pfirrmann score and clinical outcomes

The post hoc analysis revealed no link between clinical outcomes (VAS LBP and ODI) and grades II and III in Pfirrmann scores. However, IVD levels with grades IV and V in Pfirrmann scores had significantly higher VAS LBP than IVD levels with grades I to III in Pfirrmann scores. The IVD levels with grades IV and V in Pfirrmann scores had significantly higher ODI than the IVD levels with grades I in Pfirrmann scores. The scores of VAS LBP and ODI exhibited a moderate to strong negative correlation with the DH in the IVD levels with grade IV (VAS LBP: $r = -0.311$, $P = 0.025$; ODI: $r = -0.245$, $P = 0.086$) and in the IVD levels with grade V (VAS LBP: $r = -0.537$, $P = 0.000$; ODI: $r = -0.512$, $P = 0.025$), respectively (Table 5).

4. Discussion

All DH/DHI measuring methods demonstrated good-to-excellent intra-rater and inter-rater reliability except the measurement on disc material. There was a significant link between the reduction of DH in different measuring methods (anterior, middle, and posterior

Table 4
Relationships between different measurement methods on disc height in all disc levels and the different grades of intervertebral disc degeneration using Pfirrmann score.

	Disc Height in Grade I (N = 591) (mm)		Disc Height in Grade II (N = 573) (mm)		Disc Height in Grade III (N = 333) (mm)		Disc Height in Grade IV (N = 136) (mm)		Disc Height in Grade V (N = 27) (mm)		P value_Post Hoc ($p < 0.05$)
	Rater 1	Rater 2	Rater 1	Rater 2	Rater 1	Rater 2	Rater 1	Rater 2	Rater 1	Rater 2	
ADH	14.77 (7.75)	15.73 (7.60)	15.27 (10.17)	15.95 (10)	12.68 (6.32)	14.44 (6.25)	9.62 (6.74)	11.86 (5.28)	14.05 (8.38)	10.71 (6.81)	Pfirrmann IV and V compared to all other Pfirrmann levels
MDH	12.76 (6.42)	14.33 (6.14)	12.50 (6.14)	13.77 (5.93)	10.65 (4.98)	12.36 (5.03)	8.38 (3.98)	10.17 (4.25)	6.71 (2.12)	7.71 (2.86)	
PDH	9.34 (6.11)	10.12 (6.18)	8.86 (5.28)	9.33 (5.14)	7.68 (4.08)	8.59 (4.25)	6.44 (3.96)	7.69 (4.26)	13.52 (2.60)	9.91 (3.12)	
Method 7	0.59 (0.13)	0.58 (0.14)	0.59 (0.22)	0.58 (0.20)	0.58 (0.13)	0.58 (0.13)	0.51 (0.14)	0.51 (0.15)	0.53 (0.20)	0.52 (0.20)	
Method 8	0.51 (0.38)	0.51 (0.38)	0.56 (0.43)	0.56 (0.42)	0.79 (0.54)	0.79 (0.53)	0.79 (0.52)	0.79 (0.53)	0.73 (0.52)	0.73 (0.53)	

N: number of levels; ADH: anterior disc height; MDH: middle disc height; PDH: posterior disc height; All values are mean (Standard Deviation).

IVD height and DHI) and the severity of IVD degeneration (Pfirrmann grade IV and V). Clinical outcomes (VAS, LBP, and ODI) are linked with the severity of IVD degeneration based on Pfirrmann scores.

The DH changes are considered the major radiological alteration associated with pathological changes in the IVD degeneration process and lumbar degenerative disorders. Therefore, an efficient, accurate, and reproducible DH measuring method is warranted. The present work first evaluated the intra-rater and inter-rater reliability and agreement on previously published measurement methods for lumbar DH using MRI and then assessed the relationship between different DH measuring methods and Pfirrmann scores of disc degeneration. A structured protocol was employed to direct the patients to the appropriate testing postures and train the raters based on a standard training session to reduce potential bias [27].

Previous research has shown that the patient's body posture and the vertebral position during scanning, as well as the rater's expertise, are potential factors influencing MRI readings [16,22,28–30]. A structured protocol for scanning and measuring might decrease measurement bias. The present study found good-to-excellent intra-rater reliability for all DH and DHI measuring methods on MRI following our measurement protocol. Several concerns have been explored as potential causes of the outcome. First, systematic and standard training sessions and protocols are the most important variables for reducing measurement bias. Second, the division in the DH and DHI measurement procedure could lessen the affection caused by vertebral position inconsistent magnification. Therefore, a systematic training session and a standardized and structured protocol might result in good outcomes in intra-rater reliability on different measurement methods.

The agreement is used to estimate how close the results of repeated measurements provided by two raters are by evaluating the measurement error and variability in repeated measurements [16,31]. Bland and Altman's LOA with error estimates were widely suggested as a standard statistical tool for reporting intra- and inter-rater agreement [31]. We demonstrated that the DH measurements in methods 4 and 5 were biased or/and out of the acceptable cut-off proportion using Bland and Altman's LOA. A plausible reason is a measuring bias by indirect lines and small anatomical structures (e.g., the length of anterior disc material and the length of posterior disc material) on MRI in methods 4 and 5, which concur with previous findings [11,16].

The reproducible DH and DHI measurement method(s) on MRI was first evaluated using reliability and Bland and Altman's LOA in this investigation. A good-to-excellent inter-rater reliability result was found in all measurment methods for measuring the DH and DHI at the lumbar spine but method 5. Meanwhile, spondylolisthesis significantly impacted the inter-rater agreement in identifying structural boundaries and vertebral corners. The key explanation for the poor inter-rater reliability of DH measurement in method 5 is measurement bias from anatomical structures.

Disc degenerative alterations in imaging studies, including the intensity of nucleus pulposus, disc herniation, and decreased disc height are probable causes of LBP in the population [32]. However, there is no systematic investigation of the reliability and agreement on the different measuring methods for disc height loss(e.g., structural integrity and morphological changes) and potential correlations on the effect on clinical outcomes. To our knowledge, this study is the first to use the Pfirrmann classification to assess the relationship between different DH measuring methods and IVD degeneration. The present study also established a link between the severity of IVD degeneration and clinical outcomes. DH alterations occur during the major pathologic changes of the lumbar spine. Because of the limited sample size, most previously published research contradicts our results by finding no association between DH and different grades in Pfirrmann scores [9,33]. A previous study also reported that low DH is associated with grades IV and V in Pfirrmann scores [11].

4.1. Study limitation and future study

Several methodological concerns must be addressed. First, potential measurement error exists, including varied scanning locations, training processes, measurement processes, diurnal variations, and effect of raters' activities within its estimates of intra- and inter-rater reliability. A structured and standardized protocol for scanning and measuring is warranted in the future. Second, the

Table 5
Relationship between the disc height in different grade of intervertebral disc degeneration using Pfirrmann score and clinical outcomes.

	Disc Height in Grade I (N = 591) (mm)					Disc Height in Grade II (N = 573) (mm)					Disc Height in Grade III (N = 333) (mm)					Disc Height in Grade IV (N = 136) (mm)					Disc Height in Grade V (N = 27) (mm)					P value_Post Hoc ($p < 0.05$)
	L1-L2	L2-L3	L3-L4	L4-L5	L5-S1	L1-L2	L2-L3	L3-L4	L4-L5	L5-S1	L1-L2	L2-L3	L3-L4	L4-L5	L5-S1	L1-L2	L2-L3	L3-L4	L4-L5	L5-S1	L1-L2	L2-L3	L3-L4	L4-L5	L5-S1	
VAS LBP	3.82 (1.73)	3.89 (1.83)	4.11 (1.86)	4.01 (1.75)	4.04 (1.92)	4.25 (1.66)	4.38 (1.76)	4.57 (1.79)	4.69 (1.93)	4.65 (1.91)	4.3 (1.82)	4.29 (1.78)	4.20 (1.75)	4.42 (1.80)	4.49 (1.69)	–	4.49 (1.93)	4.65 (1.91)	4.53 (1.72)	4.52 (1.80)	–	–	–	4.67 (1.32)-	4.67 (1.53)	Pfirrmann IV and V compared to Pfirrmann I
	r = 0.038 (P = 0.729)					r = −0.041 (P = 0.711)					r = −0.065 (P = 0.584)					r = −0.311 (P = 0.025)					r = −0.537 (P = 0.000)					
ODI	21.16 (10.11)	21.06 (10.98)	22.88 (11.69)	22.09 (10.75)	22.37 (11.54)	24.34 (11.01)	24.56 (10.34)	26.27 (9.98)	26.30 (11.65)	27.21 (11.14)	24.34 (10.44)	25.28 (10.23)	24.63 (10.27)	26.42 (10.03)	24.31 (9.63)	–	25.76 (10.35)	26.58 (11.01)	23.96 (10.73)	26.93 (10.41)	–	–	–	28.44 (7.73)-	26.76 (10.74)	Pfirrmann V compared to Pfirrmann I
	r = 0.041 (P = 0.716)					r = −0.056 (P = 0.664)					r = −0.61 (P = 0.593)					r = −0.245 (P = 0.086)					R = −0.512 (P = 0.000)					

N: number of levels; VAS: Visual Analogue Scale; LBP: low back pain; ODI: Oswestry Disability Index; mm: millimetre; All values are mean (Standard Deviation).

different definitions for acceptable precision of the LOA range influenced the final results. Third, the subgroup analysis based on age and sex was missing due to the small sample size in each subgroup, potentially influencing the link between DH and IVD degeneration. Finally, a direct comparison of DHI ratio values using previously published measurement methods is required.

5. Conclusion

A good-to-excellent intra- and inter-rater reliability on DH could be achieved in most measurement methods using MRI following a structured protocol. However, caution should be exercised while measuring indirect lines and small anatomical structures on MRI and designating anatomical landmarks. Spondylolisthesis influences inter-rater agreement on different measuring methods. The DH measurements differ solely from grades IV and V in the Pfirrmann classification. Severe IVD degeneration is related to pain and disability.

Funding disclosure(s) statement

No funding.

Ethics statement

This study was approved by the Human Research Ethics Committee of Xuanwu Hospital Capital Medical University (KS2022151-1). The patients/participants provided their written informed consent to participate in this study.

Data availability statement

The data that support the findings of this study are available from the corresponding author, Xiaolong Chen, upon reasonable request.

CRediT authorship contribution statement

Xiao-long Chen: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Writing – original draft, Writing – review & editing. **Xiang-yu Li:** Data curation, Formal analysis, Investigation, Methodology, Writing – review & editing. **Yu Wang:** Data curation, Formal analysis, Investigation, Resources, Software, Writing – review & editing. **Shi-bao Lu:** Conceptualization, Resources, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2023.e20764.

References

[1] J. Hartvigsen, M.J. Hancock, A. Kongsted, Q. Louw, M.L. Ferreira, S. Genevay, D. Hoy, J. Karppinen, G. Pransky, J. Sieper, R.J. Smeets, M. Underwood, R. Buchbinder, J. Hartvigsen, D. Cherkin, N.E. Foster, C.G. Maher, M. Underwood, M. van Tulder, J.R. Anema, R. Chou, S.P. Cohen, L. Menezes Costa, P. Croft, M. Ferreira, P.H. Ferreira, J.M. Fritz, S. Genevay, D.P. Gross, M.J. Hancock, D. Hoy, J. Karppinen, B.W. Koes, A. Kongsted, Q. Louw, B. Öberg, W.C. Peul, G. Pransky, M. Schoene, J. Sieper, R.J. Smeets, J.A. Turner, A. Woolf, What low back pain is and why we need to pay attention, Lancet 391 (2018) 2356–2367, https://doi.org/10.1016/s0140-6736(18)30480-x.

[2] B.I. Martin, R.A. Deyo, S.K. Mirza, J.A. Turner, B.A. Comstock, W. Hollingworth, S.D. Sullivan, Expenditures and health status among adults with back and neck problems, JAMA 299 (2008) 656–664, https://doi.org/10.1001/jama.299.6.656.

[3] E.I. de Schepper, J. Damen, J.B. van Meurs, A.Z. Ginai, M. Popham, A. Hofman, B.W. Koes, S.M. Bierma-Zeinstra, The association between lumbar disc degeneration and low back pain: the influence of age, gender, and individual radiographic features, Spine 35 (2010) 531–536, https://doi.org/10.1097/BRS.0b013e3181aa5b33.

[4] J.P. Urban, S. Roberts, Degeneration of the intervertebral disc, Arthritis Res. Ther. 5 (2003) 120–130, https://doi.org/10.1186/ar629.

[5] R.C. da Costa, S. De Decker, M.J. Lewis, H. Volk, C. Canine Spinal Cord Injury, Diagnostic imaging in intervertebral disc disease, Front. Vet. Sci. 7 (2020), 588338, https://doi.org/10.3389/fvets.2020.588338.

[6] V. Haughton, Imaging intervertebral disc degeneration, J Bone Joint Surg Am 88 (Suppl 2) (2006) 15–20, https://doi.org/10.2106/JBJS.F.00010.

[7] C.W. Pfirrmann, A. Metzdorf, A. Elfering, J. Hodler, N. Boos, Effect of aging and degeneration on disc volume and shape: a quantitative study in asymptomatic volunteers, J. Orthop. Res. 24 (2006) 1086–1094, https://doi.org/10.1002/jor.20113.

[8] P.F. Beattie, S.P. Meyers, Magnetic resonance imaging in low back pain: general principles and clinical issues, Phys. Ther. 78 (1998) 738–753, https://doi.org/10.1093/ptj/78.7.738.

[9] J.P. Jarman, V.E. Arpinar, D. Baruah, A.P. Klein, D.J. Maiman, L.T. Muftuler, Intervertebral disc height loss demonstrates the threshold of major pathological changes during degeneration, Eur. Spine J. 24 (2015) 1944–1950, https://doi.org/10.1007/s00586-014-3564-8.
[10] C.W. Pfirrmann, A. Metzdorf, M. Zanetti, J. Hodler, N. Boos, Magnetic resonance classification of lumbar intervertebral disc degeneration, Spine 26 (2001) 1873–1878, https://doi.org/10.1097/00007632-200109010-00011.
[11] S. Salamat, J. Hutchings, C. Kwong, J. Magnussen, M.J. Hancock, The relationship between quantitative measures of disc height and disc signal intensity with Pfirrmann score of disc degeneration, SpringerPlus 5 (2016) 829, https://doi.org/10.1186/s40064-016-2542-5.
[12] M.T. Al-Hadidi, D.H. Badran, A.M. Al-Hadidi, J.H. Abu-Ghaida, Magnetic resonance imaging of normal lumbar intervertebral discs, Saudi Med. J. 22 (2001) 1013–1018.
[13] A.W. Kwok, Y.X. Wang, J.F. Griffith, M. Deng, J.C. Leung, A.T. Ahuja, P.C. Leung, Morphological changes of lumbar vertebral bodies and intervertebral discs associated with decrease in bone mineral density of the spine: a cross-sectional study in elderly subjects, Spine 37 (2012) E1415–E1421, https://doi.org/10.1097/BRS.0b013e31826f561e.
[14] M.I. Kingsley, L.A. D'Silva, C. Jennings, B. Humphries, V.J. Dalbo, A.T. Scanlan, Moderate-intensity running causes intervertebral disc compression in young adults, Med. Sci. Sports Exerc. 44 (2012) 2199–2204, https://doi.org/10.1249/MSS.0b013e318260dbc1.
[15] N. Boos, A. Wallin, M. Aebi, C. Boesch, A new magnetic resonance imaging analysis method for the measurement of disc height variations, Spine 21 (1996) 563–570, https://doi.org/10.1097/00007632-199603010-00006.
[16] A. Tunset, P. Kjaer, S. Samir Chreiteh, T. Secher Jensen, A method for quantitative measurement of lumbar intervertebral disc structures: an intra- and inter-rater agreement and reliability study, Chiropr Man Therap 21 (2013) 26, https://doi.org/10.1186/2045-709X-21-26.
[17] D.H.K. Chow, E.M.K. Yuen, L. Xiao, M.C.P. Leung, Mechanical effects of traction on lumbar intervertebral discs: a magnetic resonance imaging study, Musculoskelet Sci Pract 29 (2017) 78–83, https://doi.org/10.1016/j.msksp.2017.03.007.
[18] R. Zaki, A. Bulgiba, R. Ismail, N.A. Ismail, Statistical methods used to test for agreement of medical instruments measuring continuous variables in method comparison studies: a systematic review, PLoS One 7 (2012), e37908, https://doi.org/10.1371/journal.pone.0037908.
[19] J.M. Bland, D.G. Altman, Statistical methods for assessing agreement between two methods of clinical measurement, Int. J. Nurs. Stud. 47 (2010) 931–936, https://doi.org/10.1016/j.ijnurstu.2009.10.001.
[20] J. Kottner, L. Audige, S. Brorson, A. Donner, B.J. Gajewski, A. Hrobjartsson, C. Roberts, M. Shoukri, D.L. Streiner, Guidelines for reporting reliability and agreement studies (GRRAS) were proposed, J. Clin. Epidemiol. 64 (2011) 96–106, https://doi.org/10.1016/j.jclinepi.2010.03.002.
[21] X. Chen, S. Sima, H.S. Sandhu, J. Kuan, A.D. Diwan, Radiographic evaluation of lumbar intervertebral disc height index: an intra and inter-rater agreement and reliability study, J. Clin. Neurosci. 103 (2022) 153–162, https://doi.org/10.1016/j.jocn.2022.07.018.
[22] J.P. Cousins, V.M. Haughton, Magnetic resonance imaging of the spine, J. Am. Acad. Orthop. Surg. 17 (2009) 22–30, https://doi.org/10.5435/00124635-200901000-00004.
[23] H.S. Karabekir, N. Gocmen-Mas, M. Edizer, T. Ertekin, C. Yazici, D. Atamturk, Lumbar vertebra morphometry and stereological assesment of intervertebral space volumetry: a methodological study, Ann. Anat. 193 (2011) 231–236, https://doi.org/10.1016/j.aanat.2011.01.011.
[24] A. Sharma, S. Lancaster, S. Bagade, C. Hildebolt, Early pattern of degenerative changes in individual components of intervertebral discs in stressed and nonstressed segments of lumbar spine: an in vivo magnetic resonance imaging study, Spine 39 (2014) 1084–1090, https://doi.org/10.1097/BRS.0000000000000265.
[25] T.K. Koo, M.Y. Li, A guideline of selecting and reporting intraclass correlation coefficients for reliability research, J Chiropr Med 15 (2016) 155–163, https://doi.org/10.1016/j.jcm.2016.02.012.
[26] G. Atkinson, A.M. Nevill, Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine, Sports Med. 26 (1998) 217–238, https://doi.org/10.2165/00007256-199826040-00002.
[27] N.P. Lucas, P. Macaskill, L. Irwig, N. Bogduk, The development of a quality appraisal tool for studies of diagnostic reliability (QAREL), J. Clin. Epidemiol. 63 (2010) 854–861, https://doi.org/10.1016/j.jclinepi.2009.10.002.
[28] M.H. Pope, E.N. Hanley, R.E. Matteri, D.G. Wilder, J.W. Frymoyer, Measurement of Intervertebral Disc Space Height, vol. 2, Spine, 1977, pp. 282–286 (Philadelphia, Pa 1976).
[29] A. Neubert, J. Fripp, C. Engstrom, Y. Gal, S. Crozier, M.I. Kingsley, Validity and reliability of computerized measurement of lumbar intervertebral disc height and volume from magnetic resonance images, Spine J. 14 (2014) 2773–2781, https://doi.org/10.1016/j.spinee.2014.05.023.
[30] A. Schlager, K. Ahlqvist, E. Rasmussen-Barr, E.K. Bjelland, R. Pingel, C. Olsson, L. Nilsson-Wikmar, P. Kristiansson, Inter- and intra-rater reliability for measurement of range of motion in joints included in three hypermobility assessment methods, BMC Musculoskelet Disord 19 (2018) 376, https://doi.org/10.1186/s12891-018-2290-5.
[31] H.C. de Vet, C.B. Terwee, D.L. Knol, L.M. Bouter, When to use agreement versus reliability measures, J. Clin. Epidemiol. 59 (2006) 1033–1039, https://doi.org/10.1016/j.jclinepi.2005.10.015.
[32] D. Chou, D. Samartzis, C. Bellabarba, A. Patel, K.D. Luk, J.M. Kisser, A.C. Skelly, Degenerative magnetic resonance imaging changes in patients with chronic low back pain: a systematic review, Spine 36 (2011) S43–S53, https://doi.org/10.1097/BRS.0b013e31822ef700.
[33] A.J. Teichtahl, D.M. Urquhart, Y. Wang, A.E. Wluka, S. Heritier, F.M. Cicuttini, A Dose-response relationship between severity of disc degeneration and intervertebral disc height in the lumbosacral spine, Arthritis Res. Ther. 17 (2015) 297, https://doi.org/10.1186/s13075-015-0820-1.