RESEARCH ARTICLE

JOR *Spine* OPEN ACCESS

Development of normalized quantitative measures of lumbar disc degeneration

Samuel King[1] 🟢 | John Magnussen[1] | James Elliott[2,3] | Mark Jonathan Hancock[1]

[1]Faculty of Medicine, Health and Human Sciences, Macquarie University, Sydney, Australia

[2]Faculty of Medicine and Health, The Kolling Institute, The University of Sydney, St Leonards, Australia

[3]Feinberg School of Medicine, Northwestern University, Chicago, Illinois, USA

Correspondence
Samuel King, Faculty of Medicine, Health and Human Sciences, Talavera Road, Ground Floor, Macquarie University, Sydney, NSW 2109, Australia.
Email: samuel.king3@hdr.mq.edu.au

Abstract

Background: Lumbar disc degeneration (DD) is widely regarded as a likely contributor to low back pain (LBP), but the association between DD and LBP is relatively weak. No known studies have normalized quantitative measures of DD severity relative to multiple variables such as age, height, and disc level. This study developed normalized quantitative measures (z-scores) of disc signal intensity (DSI) and disc height (DH) to rate relative severity of DD.

Methods: Raw (unnormalized) quantitative measures of DSI and DH alongside potential normalization variables were acquired from MRI scans and clinical data of 76 patients. The associations between the raw quantitative measures and potential normalization variables were investigated to develop the normalized quantitative measures (z-scores) of DSI and DH. Construct validity was assessed by comparing the normalized measures to an experienced radiologist's subjective measures of relative severity of DSI and DH loss.

Results: CSF signal intensity, age, and disc level were significantly associated with raw DSI ($R^2 = 0.06$, 0.25, and 0.09, respectively). Lumbar height and disc level were significantly associated with raw DH ($R^2 = 0.13$ and 0.31). Normalizing DSI and DH by these variables resulted in stronger relationships ($R^2 = 0.39$ and 0.37) than raw DSI and DH ($R^2 = 0.24$ and 0.31) with the radiologist's subjective measures. Normalized DSI and DH were both normally distributed ($p = 0.32$ and 0.12).

Conclusions: Construct validity and the distributions suggested that normalized quantitative measures of DSI and DH are better than existing measures of DSI and DH at rating relative DD severity. Determining whether normalized quantitative measures are more predictive of clinical outcomes is important future research.

KEYWORDS
degenerative disc disease, disc degeneration, disc height, disc signal intensity, LBP, low back pain, magnetic resonance imaging, MRI, normalization, quantitative measures

1 | INTRODUCTION

Understanding of the causes of low back pain (LBP) is limited, despite the enormous societal burden of the condition.[1] Lumbar disc

degeneration (DD) is widely thought to be a potential contributor to LBP, but its clinical relevance is debated due to the relatively weak association between DD and LBP outcomes.[2–4] There is strong evidence that DD progresses with age and is typically more severe in the lower lumbar levels, regardless of LBP.[5–8] However, DD is more common in people with LBP than those without LBP.[2] These findings suggest that raw measures of DD may be less clinically relevant because of normal aging or variability between disc levels, which should be accounted for when rating DD.

Quantitative measures of DD are a highly reliable method for rating DD compared to more common subjective ratings such as the Pfirrmann scale.[9–11] Measuring the disc signal intensity (DSI) and disc height (DH) on MRI are commonly used to quantify DD in research.[9–15] However, raw measures of DSI and DH are unlikely to accurately represent relative severity unless they are normalized by sources of variability between images or patients. For example, older people would typically be expected to have lower DSI, and shorter people would typically be expected to have lower DH. Other variables such as the disc level and image intensity may also cause variability in quantitative measures of DSI and DH. These sources of variability may be considered noise since they may reduce the clinical relevance of the raw quantitative measures.

Previous studies have adjusted raw quantitative measures of DSI and DH for a single variable, such as adjusting the DH by the patient's vertebral body height.[9,10,12–15] However, to our knowledge, previous studies have not systematically normalized (measured as a z-score) raw quantitative measures for multiple key variables that may otherwise collectively reduce their clinical relevance. If raw quantitative measures of DD are normalized to account for these important variables, this should result in improved measures of clinically relevant DD severity.

The overall aim of this study was to develop normalized quantitative measures (z-scores) of DSI and DH. Specifically, this project aimed to: (1) investigate the strength of association between raw quantitative measures of DSI and DH and various potential normalization variables (e.g., age, height, and disc level), to identify the variables to include in the normalization models; (2) investigate which normalization models produce measures of DSI and DH that are most strongly associated with subjective measures of relative severity of DSI and DH loss reported by an experienced radiologist, for the purpose of construct validity testing; and (3) investigate the normality of the distributions of the normalized quantitative measures of DSI and DH, to evaluate whether some of the suspected noise of the raw quantitative measures may have been removed after normalization.

2 | MATERIALS AND METHODS

2.1 | Participants

This study used deidentified data from 76 participants who underwent standardized lumbar MRI scans in a previous study investigating risk factors for a recurrence of LBP.[16] The participants consisted of 46 males and 30 females aged 22–75 years who had previously experienced one or more episodes of acute, nonspecific LBP lasting at least 24 h.[16] Participants were included if they had recovered from a previous episode of acute, nonspecific LBP within 3 months prior to imaging.[16] The exclusion criteria were: previous spinal surgery; contraindication to MRI; or unable to complete the primary follow-up electronically (via email or text message).[16]

2.2 | Imaging process

All images came from a 3.0 T Siemens Verio (Siemens Healthineers, Erlangen, Germany) MRI Scanner with a multichannel phased array surface coil.[16] The standardized imaging procedures for all participants included sagittal fast spin echo T1 (TR 650 ms, TE 6.3 ms) and T2 (TR 4500 ms, TE 101 ms), STIR (TR 3800 ms, TE 35 ms, TI 215 ms), and axial T2 (TR 5000 ms, TE 116 ms) scans.[16] All imaging sequences were 4-mm thick with 1-mm interslice space.[16] Sagittal sequences used a 320-mm FOV, while axial sequences used a 200-mm FOV.[16]

2.3 | Collection of the raw quantitative measures and potential normalization variables

The raw quantitative measures were defined as the unnormalized measures of DSI and DH. The potential normalization variables were defined as the variables to be investigated for normalizing the raw quantitative measures of DSI and DH. Definitions of the raw quantitative measures and potential normalization variables are provided in Table 1.

To collect the raw quantitative measures and potential normalization variables, MRI data were extracted from each of the five lumbar discs of all 76 MRI scans (380 measurements). All MRI measurements were taken using a midsagittal view by a researcher who underwent training from an experienced radiologist. Horos DICOM-viewing software (horosproject.org) was used to obtain the MRI data using the software's in-built measurement functions. The researcher started formal MRI data extraction only after demonstrating high levels of intraobserver reliability on a sample of five scans (25 disc levels), measured by the ICC.

2.4 | Collection of the radiologist's subjective measures of relative severity of DSI and DH loss

There is no gold standard rating for relative DD severity, so an experienced radiologist's subjective measures of relative severity of DSI and DH loss were used as a comparative reference for the normalized quantitative measures. The rater was a senior neuroradiologist with more than 20 years' experience in teaching hospital and academic radiology, who reports more than 2000 spine-related studies per annum. These subjective ratings of relative severity were provided for the five lumbar discs from all 76 images and called the radiologist's

TABLE 1　Definitions of the raw quantitative measures and potential normalization variables.

Variable name[a] (abbreviated)	Definition (units)	Summary[b] ($n = 380$)
Raw quantitative measures		
Disc signal intensity (DSI)	Mean signal intensity within the disc	142 (53)
Disc height (DH)	Mean of anterior, middle, and posterior disc heights (mm)	9.9 (1.8)
Potential normalization variables		
Disc level, categorical	Spinal level of disc (1 or L1/L2, 2 or L2/L3, 3 or L3/L4, 4 or L4/L5, 5 or L5/S1)	Not applicable
Sex, categorical, n (%)	Sex of participant (male, female)	46 (60.5)
Age	Age of participant (years)	45.6 (12.8)
Body height (BH)	Body-height of participant (cm)	171.4 (9.6)
Bodyweight (BW)	Bodyweight of participant (kg)	79.8 (1.9)
BMI	BMI of participant (kg/cm²)	26.9 (5.1)
CSF signal intensity (CSFSI)	Mean signal intensity of CSF region with a minimum area of 1 cm²	950 (199)
Spinal cord signal intensity (SCSI)	Mean signal intensity of spinal cord region with a minimum area of 1 cm²	345 (84)
Vertebral body height (VBH)	Mean of anterior, middle, and posterior vertebral body heights (mm)	25.4 (1.9)
Total lumbar height (LH)	Sum of means of anterior, middle, and posterior L1–L4 and L4–S1 heights (mm)	176 (11)
Radiologist's normalized measures, categorical		
Radiologist's normalized DSI	(0) Extremely low to (6) extremely high relative to (3) typical DSI[c]	2.44 (0.73)
Radiologist's normalized DH	(0) Extremely low to (6) extremely high relative to (3) typical DH[d]	2.60 (0.68)

[a]Variables are continuous, unless otherwise specified.
[b]Summaries are provided as "mean (SD)", unless otherwise specified (abbreviations: SD, standard deviation; n, population size).
[c]"Typical DSI" is relative to the surrounding intensities, age, sex, and disc level.
[d]"Typical DH" is relative to the vertebral/lumbar height, age, sex, and disc level.

normalized measures. The radiologist rated a given patient's DSI and DH on a 0–6 scale, relative to the vertebral/lumbar height (surrounding signal intensities for DSI), age, sex, and disc level (see Table 1). The radiologist was provided with a graphic of the rating scales for both DSI and DH based on the standard normal distribution, which was also supplemented by written rating scales (see Supporting Information; Figure S1 and Table S1). A seven-point scale was elected to retain reliability and accuracy of the radiologist's assessment, while

remaining large enough to sufficiently differentiate between patient groupings for comparisons to the normalized quantitative measures.

When rating DSI, the instructions given to the radiologist were: "Subjectively normalize the patient's DSI for surrounding signal intensities, age, sex, and disc level." When rating DH, the instructions given to the radiologist were: "Subjectively normalize the patient's DH for vertebral/lumbar height, age, sex, and disc level." The radiologist's reliability was measured using the Kendall's Rank Correlation Coefficient, τ. The radiologist conducted repeated ratings of the radiologist's measures from a random sample of five scans (25 disc levels), 6 weeks after conducting the initial ratings.

2.5 │ Statistical analyses

All statistical analyses were performed using R. Aim 1 was addressed using univariable and multivariable linear regression, comparing the potential normalization variables (predictors) to the raw quantitative measures (outcomes). If disc level was not the predictor variable, linear mixed models were used, setting disc level as the random effects variable to account for dependencies within participants.

Initially, univariable linear modeling was conducted. If the univariate relationship was sufficiently strong ($R^2 \geq 0.05$), then the corresponding potential normalization variable would be included in the subsequent multivariable models. The linear regression assumptions and multicollinearity between potential normalization variables were both tested. Multivariable linear modeling was then conducted using the potential normalization variables selected from univariable modeling. Multivariable models predicting raw quantitative DSI and DH were constructed using a manual forward selection process, by iteratively adding potential normalization variables to the models. The multivariable models with the strongest relationships (greatest R^2) with raw quantitative DH and DSI were then identified, and thus the final normalization variables were selected.

Normalization of the raw quantitative measures of DSI and DH was then conducted using the final normalization variables. Two methods of normalization were performed, depending on whether the normalization variables were continuous or categorical. For continuous normalization variables, each linear normalization variable term corresponding to the multivariable model was subtracted from the raw quantitative measure (subtraction method), and the z-scores were then calculated. For categorical normalization variables, the quantitative measures were decomposed into subpopulations (e.g., individual disc levels) based on all the values the normalization variable could take (subpopulation method). After calculating the z-scores within the individual subpopulations, they were recomposed into a single population.

Aim 2 was addressed using univariable linear regression, by separately comparing both the raw and normalized quantitative measures (predictors) to the radiologist's normalized measures (outcomes). To select the final normalization models for DSI and DH, the best performing regression models were evaluated after using a manual forward selection process, by iteratively normalizing the raw quantitative

measures by different combinations of normalization variables. This was assessed by measuring the change of R^2 relative to the regression models comparing the radiologist's normalized measures to the raw quantitative measures.

Finally, Aim 3 was addressed by creating the distributions of the normalized quantitative measures of DSI and DH to investigate their normality and compare to the distributions of the raw quantitative measures. Normality of the distributions was assumed if the Shapiro–Wilk test returned a *p*-value greater than 0.05.

3 | RESULTS

3.1 | Summary of the raw quantitative data and radiologist's normalized measures

Generally, the population consisted of mostly overweight, middle-aged individuals, with slightly more males than females and an average body-height of 171 cm (Table 1). Raw quantitative DH (mean of anterior, middle, and posterior DHs) was 9.9 mm on average.

3.2 | Reliability of the MRI data and radiologist's normalized measures

Reliability of the quantitative MRI data was excellent, with almost all ICC values ≥0.95, excluding the posterior DH (ICC = 0.90), and majority (8 of 15) ICC values ≥0.98. See Table S2 for all MRI data reliability results. The radiologist's normalized DSI (subjective) had a τ value of 0.70, while the radiologist's normalized DH (subjective) had a τ value of 0.79.

3.3 | Univariable comparison of the potential normalization variables (predictors) to the raw quantitative measures of DSI and DH (outcomes)

Age had the strongest statistically significant relationship with raw quantitative DSI ($R^2 = 0.25$) of all potential normalization variables (Table 2). Bodyweight, BMI, and disc level had statistically significant relationships with raw quantitative DSI ($R^2 = 0.07$, 0.09, and 0.09, respectively). CSF signal intensity (CSFSI) had a statistically significant relationship with raw quantitative DSI ($R^2 = 0.06$) that was slightly stronger than spinal cord signal intensity. Sex did not have a statistically significant relationship with raw quantitative DSI.

Disc level had the strongest statistically significant relationship with raw quantitative DH ($R^2 = 0.31$) of all potential normalization variables (Table 2). Lumbar height had a statistically significant relationship with raw quantitative DH ($R^2 = 0.13$) that was much stronger than vertebral body height and body-height ($R^2 = 0.02$). Bodyweight, BMI, and sex all had statistically significant relationships with raw quantitative DH ($R^2 = 0.05$, 0.07, and 0.05, respectively). Age did not have a statistically significant relationship with raw quantitative DH.

3.4 | Multivariable comparison of the potential normalization variables (predictors) to the raw quantitative measures of DSI and DH (outcomes)

Mean CSFSI, age, BMI, and disc level were selected as the potential normalization variables for raw quantitative DSI to be included in the multivariable linear modeling (Table 3). The best performing multivariable linear model including the mean CSFSI, age, and disc level had an R^2 of 0.43.

Lumbar height, bodyweight, and disc level were selected as the potential normalization variables for raw quantitative DH to be included in the multivariable linear modeling (Table 3). Sex was not included due to the low sample size and much weaker relationship with raw quantitative DH than disc level. BMI was not included to avoid multicollinearity with lumbar height and bodyweight. The best performing multivariable linear model including the lumbar height and disc level had an R^2 of 0.45.

3.5 | Univariable comparison of the raw and normalized quantitative measures of DSI and DH (predictors) to the radiologist's normalized measures of DSI and DH (outcomes)

Raw quantitative DSI had an R^2 of 0.31 with the radiologist's normalized DSI (Table 4). There was a 3% decrease in R^2 (0.30) when normalizing raw quantitative DSI by subtracting the linear CSFSI term (DSI − 0.07 CSFSI). There was a 39% increase in R^2 (0.43) when normalizing raw quantitative DSI by subtracting the linear age term (DSI + 2.20 age or DSI − [−2.20] age). There was a 42% increase in R^2 (0.44) when normalizing raw quantitative DSI by simultaneously subtracting both the linear CSFSI and age terms (DSI − 0.07 CSFSI + 2.20 age). When sub-populating quantitative DSI by disc level (e.g., Levels [DSI]), R^2 always decreased compared to the measures that were not sub-populated (e.g., $R^2 = 0.31$ for DSI [raw] and $R^2 = 0.26$ for Levels [DSI]). It was found that the radiologist's normalized DSI measures were not entirely independent of disc level, indicating that the radiologist had likely not completely normalized DSI by disc level. Thus, disc level was retained in the normalization model for quantitative DSI (Levels [DSI − 0.07 CSFSI + 2.20 age]), which had an R^2 of 0.37, even though this was lower than the model excluding disc level.

Raw quantitative DH had an R^2 of 0.24 with the radiologist's normalized DH (Table 4). There was no increase in R^2 (0.24) when normalizing raw quantitative DH by subtracting the linear lumbar height term (DH − 0.06 LH). Sub-populating raw quantitative DH by disc level, using Levels (DH), resulted in a 50% increase in R^2 (0.36). Normalizing raw quantitative DH by subtracting the linear lumbar height term and sub-populating by disc level, using Levels (DH − 0.06 LH), resulted in a 63% increase in R^2 (0.39). Thus, the normalization model for quantitative DH was selected to be the Levels (DH − 0.06 LH) model.

TABLE 2 Univariable linear regression results comparing the raw quantitative DSI and DH measures to the potential normalization variables.

Normalization variables		Raw quantitative DSI			Raw quantitative DH		
		R^2	β	95% CI	R^2	β	95% CI
CSFSI		0.06	0.06	0.04–0.09	Not tested		
SCSI		0.04	0.13	0.07–0.19			
VBH		Not tested			0.02	0.14	0.06–0.22
LH					0.13	0.06	0.05–0.07
BH					0.04	0.04	0.02–0.05
Age		0.25	−2.10	−2.43 to −1.77	<0.01	0.01	−0.01 to 0.02
BW		0.07	−0.76	−1.02 to −0.50	0.07	0.03	0.02–0.03
BMI		0.09	−3.10	−4.05 to −2.15	0.05	0.08	0.05–0.11
Sex	Male	<0.01	Reference		0.05	Reference	
	Female		−6	−14 to 6		−0.8	−1.1 to −0.5
Disc level	1	0.09	Reference		0.31	Reference	
	2		2	−14 to 18		1.2	0.8–1.7
	3		−4	−20 to 12		2.3	1.8–2.7
	4		−31	−47 to −15		2.8	2.3–3.3
	5		−38	−54 to −22		2.2	1.7–2.7

Abbreviations: BH, body height; BW, bodyweight; CSFSI, CSF signal intensity; DH, disc height; DSI, disc signal intensity; LH, lumbar height; R^2, coefficient of determination; SCSI, spinal cord signal intensity; VBH, vertebral body height; β, beta coefficient.

TABLE 3 Summary of the best performing multivariable linear models for raw quantitative DSI and DH.

Normalization variable		Raw quantitative DSI				
		R^2	Intercept	95% CI	β	95% CI
CSF signal intensity		0.43	185	160–210	0.07	0.05–0.10
Age					−2.20	−2.51 to −1.88
Disc level	1				Reference	
	2				2.4	−10.2 to 15.1
	3				−4.4	−17.1 to 8.2
	4				−31.1	−43.8 to −18.5
	5				−37.8	−50.5 to −25.2
		Raw quantitative DH				
Lumbar height		0.45	−2.7	−4.9 to −0.5	0.06	0.05–0.07
Disc level	1				Reference	
	2				1.2	0.8–1.7
	3				2.3	1.8–2.7
	4				2.8	2.4–3.2
	5				2.2	1.8–2.6

Abbreviations: DH, disc height; DSI, disc signal intensity; R^2, coefficient of determination; β, beta coefficient.

3.6 | Distributions of the raw and normalized quantitative measures of DSI and DH

From Shapiro–Wilk testing, the *p*-value of the raw quantitative DSI distribution was <0.001, indicating that the distribution was not normal (Figure 1). For the normalized quantitative DSI distribution, the *p*-value was 0.12, instead indicating that the distribution was normal. For the raw and normalized quantitative DH distributions, the respective *p*-values were 0.32 and 0.42, indicating that both distributions were normal.

4 | DISCUSSION

### 4.1 | Summary of key findings

The main outcome of this study was that normalized quantitative measures of DSI and DH were developed to rate relative DD severity. Raw quantitative DSI was normalized by the mean CSF signal intensity, age, and disc level; and raw quantitative DH was normalized by the total lumbar height and disc level. Construct validity was

TABLE 4 Univariable linear regression results comparing the radiologist's normalized measures to the raw and normalized quantitative measures.

Normalization method	Quantitative measure	Radiologist's normalized DSI		
		R^2 (Δ_{rel})	β	95% CI
Raw	DSI	0.31	0.76	0.65–0.88
Subtraction	DSI − 0.07 CSFSI	0.30 (−3%)	0.74	0.63–0.87
Subtraction	DSI + 2.20 age[a]	0.43 (+39%)	0.88	0.78–0.99
Subtraction	DSI − 0.07 CSFSI + 2.20 age[a]	0.44 (+42%)	0.89	0.79–0.99
Subpopulation	Levels (DSI)	0.26 (−16%)	0.69	0.58–0.81
Combination	Levels (DSI − 0.07 CSFSI)	0.25 (−19%)	0.68	0.56–0.80
Combination	Levels (DSI + 2.20 age)[a]	0.36 (+16%)	0.82	0.71–0.93
Combination	Levels (DSI − 0.07 CSFSI + 2.20 age)[a]	0.37 (+19%)	0.83	0.72–0.94
		Radiologist's normalized DH		
Raw	DH	0.24	0.76	0.67–0.86
Subtraction	DH − 0.06 LH	0.24 (0%)	0.77	0.68–0.86
Subpopulation	Levels (DH)	0.36 (+50%)	0.88	0.76–0.99
Combination	Levels (DH − 0.06 LH)	0.39 (+63%)	0.92	0.80–1.03

Abbreviations: CSFSI, CSF signal intensity; DH, disc height; DSI, disc signal intensity; LH, lumbar height; R^2, coefficient of determination; β, beta coefficient; Δ_{rel}, relative change from raw R^2.

[a]Age β coefficient from multivariable modeling was −2.20, so subtracting this term yields +2.20 age.
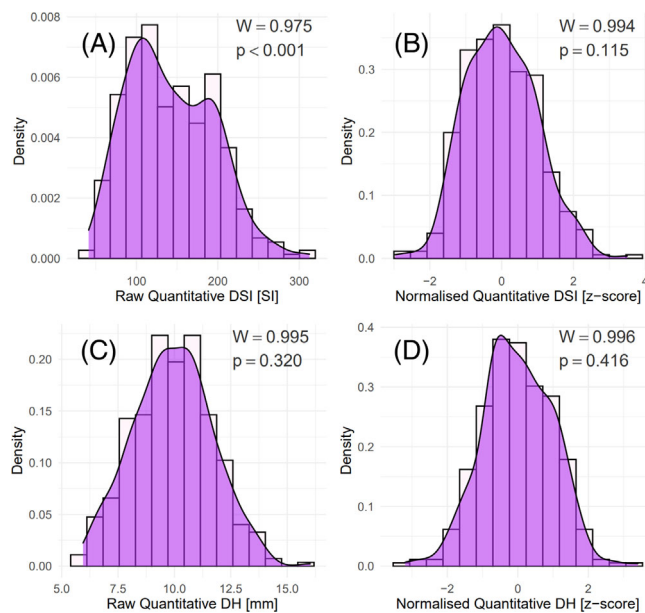
FIGURE 1 (A): Raw quantitative DSI distribution; (B): normalized quantitative DSI distribution; (C): raw quantitative DH distribution; (D): normalized quantitative DH distribution.
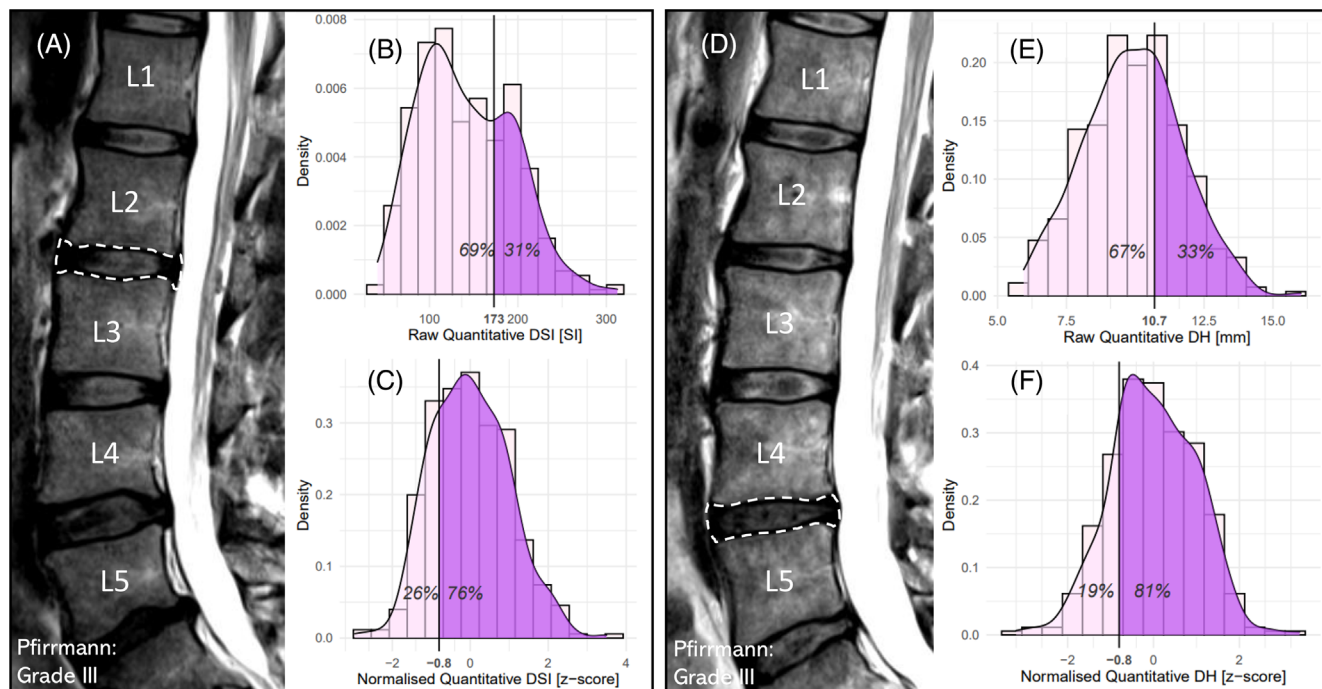
supported since the normalized quantitative measures of DSI and DH had associations that were stronger than the raw quantitative measures when compared to the radiologist's normalized measures (the reference standard used for relative DD severity). The raw quantitative DSI distribution was not normal, while the raw quantitative DH distribution was normal; however, the normalized quantitative DSI and DH distributions were both normal. The distributions suggest that some of the suspected noise of raw quantitative measures was removed after normalization. These findings suggest that normalized quantitative measures are better measures of relative severity than raw quantitative measures, but this requires further clinical investigation before definitive conclusions can be drawn.

4.2 | Implications of key findings

Since the normalized quantitative measures of DSI and DH are likely better measures of relative DD severity than the raw quantitative measures, they may be more strongly associated with clinical symptoms. Consider Figure 2, first evaluating the L2/L3 DSI (Pfirrmann grade III) of a 33-year-old male (Figure 2A) and their position in the raw (Figure 2B) and normalized (Figure 2C) quantitative DSI distributions. Raw quantitative DSI rates this person in the top 31% of people. However, normalized quantitative DSI rates this person in the bottom 26% of people. Similarly, the next part of Figure 2 evaluates the L4/L5 DH (Pfirrmann grade III) of a 55-year-old male (Figure 2D) and their position in the raw (Figure 2E) and normalized (Figure 2F) quantitative DH distributions. Raw quantitative DH rates this person in the top 33% of people, while normalized quantitative DH rates this person in the bottom 19% of people. As is clear, normalization markedly changes where some people are rated in the distributions. These examples demonstrate that the interpretation of quantitative measures of DSI and DH substantially changes after normalization. The results also show that these normalized quantitative measures (z-scores) can vary greatly for the same Pfirrmann grade.

FIGURE 2 (Left) (A): MRI scan from a 33-year-old male with raw DSI of 173 and normalized DSI (z-score) of −0.8 for the L2/L3 disc (Pfirrmann grade III); (B): their position in the raw quantitative DSI distributions; (C): their position in the normalized quantitative DSI distributions. (Right) (D): MRI scan from a 55-year-old male with raw DH of 10.7 mm and normalized DH (z-score) of −0.8 for the L4/L5 disc (Pfirrmann grade III); (B): their position in the raw quantitative DH distributions; (C): their position in the normalized quantitative DH distributions.

4.3 | Future directions

Demonstrating that normalized quantitative measures of DD are more strongly associated with important clinical outcomes is needed before they can be confidently considered better than either raw quantitative measures or subjective measures such as Pfirrmann grade.[17] There are three primary study designs toward achieving this: prognostic validation; treatment validation; and cross-sectional validation. Prognostic validation would involve a longitudinal cohort to investigate whether the normalized quantitative measures are more predictive of future LBP than unnormalized (raw) measures. Treatment validation would involve a randomized control trial investigating whether normalized quantitative measures are better than unnormalized measures at identifying patients who respond better to a specific intervention. Cross-sectional validation would involve a cross-sectional study investigating whether normalized quantitative measures are better than unnormalized measures at either identifying patients with and without LBP or differentiating between severities of LBP.

4.4 | Study limitations

A clear limitation of this study was using the radiologist's measures as the construct validity reference for the normalized quantitative measures of DSI and DH. However, the radiologist's measures have face validity and are believed to be the best available reference measures

of relative severity defined by real-world clinical practice. The key limitation of the current work is the lack of clear evidence that normalized quantitative measures are superior to existing measures at predicting LBP outcomes. As mentioned in the previous section, this is important future research.

5 | CONCLUSIONS

This study developed normalized quantitative measures of DSI and DH (z-scores) to rate relative DD severity. The findings suggest that normalized quantitative measures of DSI and DH are better measures of relative DD severity than raw quantitative measures. However, the normalized quantitative measures of DD need to be demonstrated to have strong associations with important clinical outcomes before they can be confidently considered better than existing measures for research and clinical purposes.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

ORCID

Samuel King https://orcid.org/0000-0002-8489-937X

REFERENCES

1. Hartvigsen J, Hancock MJ, Kongsted A, et al. What low back pain is and why we need to pay attention. *Lancet*. 2018;*391*(10137): 2356-2367.
2. Brinjikji W, Diehn F, Jarvik J, et al. MRI findings of disc degeneration are more prevalent in adults with low back pain than in asymptomatic controls: a systematic review and meta-analysis. *Am J Neuroradiol*. 2015;*36*(12):2394-2399.
3. Steffens D, Hancock M, Maher C, Williams C, Jensen TS, Latimer J. Does magnetic resonance imaging predict future low back pain? A systematic review. *Eur J Pain*. 2014;*18*(6):755-765.
4. Steffens D, Hancock MJ, Pereira LS, Kent PM, Latimer J, Maher CG. Do MRI findings identify patients with low back pain or sciatica who respond better to particular interventions? A systematic review. *Eur Spine J*. 2016;*25*(4):1170-1187.
5. Brinjikji W, Luetmer PH, Comstock B, et al. Systematic literature review of imaging features of spinal degeneration in asymptomatic populations. *Am J Neuroradiol*. 2015;*36*(4):811-816.
6. Pfirrmann CW, Metzdorf A, Elfering A, Hodler J, Boos N. Effect of aging and degeneration on disc volume and shape: a quantitative study in asymptomatic volunteers. *J Orthop Res*. 2006;*24*(5):1086-1094.
7. Urquhart DM, Kurniadi I, Triangto K, et al. Obesity is associated with reduced disc height in the lumbar spine but not at the lumbosacral junction. *Spine*. 2014;*39*(16):E962-E966.
8. Nagashima M, Abe H, Amaya K, et al. A method for quantifying intervertebral disc signal intensity on T2-weighted imaging. *Acta Radiol*. 2012;*53*(9):1059-1065.
9. Salamat S, Hutchings J, Kwong C, Magnussen J, Hancock MJ. The relationship between quantitative measures of disc height and disc signal intensity with Pfirrmann score of disc degeneration. *Springerplus*. 2016;*5*(1):1-7.
10. Niemeläinen R, Videman T, Dhillon S, Battié M. Quantitative measurement of intervertebral disc signal using MRI. *Clin Radiol*. 2008;*63*(3):252-255.
11. Videman T, Battié MC, Gibbons LE, Gill K. Aging changes in lumbar discs and vertebrae and their interaction: a 15-year follow-up study. *Spine J*. 2014;*14*(3):469-478.
12. Jarman JP, Arpinar VE, Baruah D, Klein AP, Maiman DJ, Muftuler LT. Intervertebral disc height loss demonstrates the threshold of major pathological changes during degeneration. *Eur Spine J*. 2015;*24*(9): 1944-1950.
13. Aissiou M, Périé D, Mac-Thiong J-M. Normalized intervertebral disc MRI signal as a biomarker of pain. *J Biomed Sci Eng*. 2013;*6*:372-380.
14. Haefeli M, Kalberer F, Saegesser D, Nerlich AG, Boos N, Paesold G. The course of macroscopic degeneration in the human lumbar intervertebral disc. *Spine*. 2006;*31*(14):1522-1531.
15. Inoue H, Ohmori K, Miyasaka K, Hosoe H. Radiographic evaluation of the lumbosacral disc height. *Skeletal Radiol*. 1999;*28*(11):638-643.
16. Hancock MJ, Maher CM, Petocz P, et al. Risk factors for a recurrence of low back pain. *Spine J*. 2015;*15*(11):2360-2368.
17. Pfirrmann CW, Metzdorf A, Zanetti M, Hodler J, Boos N. Magnetic resonance classification of lumbar intervertebral disc degeneration. *Spine*. 2001;*26*(17):1873-1878.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.