Accepted Manuscript

Validity and reliability of computerized measurement of lumbar intervertebral disc height and volume from magnetic resonance images

Ales Neubert, Jurgen Fripp, Craig Engstrom, Yaniv Gal, Stuart Crozier, Michael Kingsley

PII: S1529-9430(14)00565-8

DOI: 10.1016/j.spinee.2014.05.023

Reference: SPINEE 55901

To appear in: *The Spine Journal*

Received Date: 21 November 2013

Revised Date: 6 March 2014

Accepted Date: 20 May 2014

Please cite this article as: Neubert A, Fripp J, Engstrom C, Gal Y, Crozier S, Kingsley M, Validity and reliability of computerized measurement of lumbar intervertebral disc height and volume from magnetic resonance images, *The Spine Journal* (2014), doi: 10.1016/j.spinee.2014.05.023.

Validity and reliability of computerized measurement of lumbar intervertebral disc height and volume from magnetic resonance images

A. Neubert[1,2], J. Fripp[1], C. Engstrom[3], Y. Gal[2], S. Crozier[2] and M. Kingsley[4]

[1] The Australian E-Health Research Centre, CSIRO Computational Informatics, Brisbane, Australia

[2] School of Information Technology & Electrical Engineering, University of Queensland, Brisbane, Australia

[3] School of Human Movement Studies, University of Queensland, Brisbane, Australia

[4] Exercise Physiology, La Trobe Rural Health School, La Trobe University, Victoria, Australia

Keywords:
intervertebral disc, magnetic resonance imaging, computerized measurement, morphometry

Corresponding author:

Michael Kingsley, PhD.
Exercise Physiology, La Trobe Rural Health School, La Trobe University, Victoria 3550, Australia

Email: M.Kingsley@latrobe.edu.au
Phone: +61 3 5444 7589
Fax: +61 3 5444 7977

Validity and reliability of computerized measurement of lumbar intervertebral disc height and volume from magnetic resonance images

Abstract

Background context

Magnetic resonance (MR) examinations of morphological characteristics of intervertebral discs (IVDs) have been used extensively for biomechanical studies and clinical investigations of the lumbar spine. Traditionally, the morphological measurements have been performed using time- and expertise-intensive manual segmentation techniques not well suited for analyses of large-scale studies.

Purpose

The purpose of this study is to introduce and validate a semi-automated method for measuring IVD height and mean sagittal area (and volume) from MR images to determine if it can replace the manual assessment and enable analyses of large MR cohorts.

Study Design/Setting

This study compares semi-automated and manual measurements and assesses their reliability and agreement using data from repeated MR examinations.

Methods

Seven healthy asymptomatic males underwent 1.5T MR examinations of the lumbar spine involving sagittal T2-weighted fast spin echo images obtained at baseline, pre-exercise and post-exercise conditions. Measures of the mean height and the mean sagittal area of lumbar IVDs (L1/L2 – L4/L5) were compared for two segmentation approaches: 1) a conventional manual method (10-15 minutes

to process one IVD); and 2) a specifically developed semi-automated method (requiring only a few mouse clicks to process each subject).

This research was supported under Australian Research Council's Linkage Projects funding scheme XXX (total amount of AUD $540k during 2010 – 2013). No conflicts of interest were reported by the authors of this manuscript.

Results

Both methods showed strong test-retest reproducibility evaluated on baseline and pre-exercise examinations with strong intra-class correlations for the semi-automated and manual methods for mean IVD height (ICC = 0.99, 0.98) and mean IVD area (ICC = 0.98, 0.99), respectively. A bias (average deviation) of 0.38 mm (4.1%, 95% confidence interval: 0.18 to 0.59 mm) was observed between the manual and semi-automated method for the IVD height, while there was no statistically significant difference for the mean IVD area (0.1 ± 3.5%). The semi-automated and manual methods both detected significant exercise-induced changes in IVD height (0.20 mm, 0.28 mm) and mean IVD area (5.7 mm^2, 8.3 mm^2), respectively.

Conclusions

The presented semi-automated method provides an alternative to time- and expertise-intensive manual procedures for analysis of larger cross-sectional, interventional and longitudinal MR studies for morphometric analyses of lumbar IVDs.

1 Introduction

Radiological examinations of the morphological characteristics of lumbar intervertebral discs (IVDs) such as height and volume have been used extensively for biomechanical studies and clinical investigations of the human spine [1,2]. Shape and volume of the IVDs influence the load-carrying capacity of the spinal column, and morphological abnormalities such as IVD space narrowing and thinning have been associated with acute or chronic disabilities of the lumbar spine [3]. Magnetic resonance (MR) imaging offers an ideal non-invasive modality to study the morphology of IVDs with excellent soft tissue visualization without exposing the participant to ionizing radiation [4]. However, the etiology of lower back pain and the relation to MR imaging findings is poorly understood [5–7]. Efforts in the search for imaging biomarkers that can be correlated to clinical symptoms have focused on biochemical MR [8,9], diffusion MR [10], sodium MR [11], and on assessment of morphological IVD features from structural MR imaging [12]. Morphometric measures derived from MR images include IVD height (anterior, middle, posterior) [1], width (inferior, middle, superior) [13], concavity of the vertebral endplate [14,15] and IVD volume [2]. These measures have been used in the assessment of spine pathologies, including disc degeneration [1,16,17], disc herniation [18] and changes associated with osteoporosis [19,20], or to quantify biophysical effects on IVDs, such as ageing [1,21], lifestyle factors [22] and acute exercise bouts [23,24]. In addition to diagnostic testing or interventional studies, morphometric features have the potential to benefit surgical planning [18] and facilitate the design of IVD prosthesis [25]. Quantification of weight-bearing and impact activities on the IVDs can help to understand the effects of exercise training interventions and to design safe exercise prescriptions [23,24].

Traditionally, morphological measurements on IVDs in MR studies have been performed by specifically trained analysts using manual segmentation techniques [1,2,20]. These manual approaches are time- and expertise-demanding tasks, often requiring numerous hours of labor for

image segmentation and considerable supervised training of analysts from experienced radiologist to help reduce notable inter-rater (subjective) variation [26]. Computerized segmentation approaches have the potential to systematically deliver reproducible measurements to accelerate data analysis and processing in large research or clinical investigations. Several (semi-)automated approaches for two-dimensional (2D) segmentation of IVDs from MR images have been proposed in the literature [27–31]. Morphometric data for the IVDs are subsequently extracted from the segmented IVDs using additional computer algorithms, which for measurements such as the 'average' IVD height is a nontrivial task. Computer driven automations for measuring IVD morphology have been developed for computed tomography (CT), where the high-intensity contrast between the cortical bone of the vertebrae and the IVDs allows well-defined reconstruction of the vertebrae and vertebral endplates. The IVD height can then be estimated from the distances between prominent points of the vertebral endplates [25] and the volume estimated from the measured IVD height and the axial surface of the vertebral endplates [32]. Recently, Tan et al. [33] have presented a semi-automated method for measuring average IVD heights from CT with high accuracy (95% confidence intervals (CI) between -1 and 1 mm) and reproducibility (intra-class correlation coefficient (ICC) > 0.98) requiring the manual input of one point per vertebra. Currently, very few computerized methods for morphometric analysis of IVDs from MR images have been presented in the literature. Boos et al. [34] and Roberts et al. [21] presented computerized measurements of IVD height from mid-sagittal MR slices but their approaches did not incorporate volumetric analyses. Moreover, the approach of Boos et al. [34] involved considerable user interaction and the method of Roberts et al. [21] was not quantitatively validated against manual analyses.

In this work, a semi-automated three-dimensional (3D) analysis algorithm was used to calculate the average height and sagittal area of lumbar IVDs from T2-weighted MR images for comparison against manual measurements. The method is based on our previously presented automated

segmentation algorithm [35] that extracts 3D IVD volumes from MR scans. The segmentation method was previously quantitatively validated on a dataset of 14 asymptomatic volunteers [35] and successfully applied to a further dataset of 11 symptomatic patients with lower back pain (presenting 43/64 abnormal lumbar IVDs) [36]. In this paper, further procedures to obtain computerized morphological measurements of the IVDs are presented and validated against manual segmentation data. The aims of this study were: (i) to assess the reproducibility of IVD average height and area measurements for the semi-automated and manual methods (test-retest); (ii) to evaluate the measurement agreement between both methods (validity); and (iii) to compare MR measures of IVD morphology from both methods following an acute exercise bout.

2 Methods

2.1 Experimental design and imaging dataset

Seven male volunteers (18–23 years of age, 69.4 ± 5.1 kg, 174 ± 2 cm) with no history of chronic back pain, injury or associated musculoskeletal disease gave written consent to participate in the study, which had ethics approval from XXX. Participants attended three MR scanning sessions on two days. On both days, the participants reported directly to the clinic within 30 minutes of rising from bed, after at least 10 hours of bed rest, and having undertaken minimal ambulatory activity. Participants were instructed to refrain from exercise for 24 hours and from consuming food for a minimum of 10 hours before examination. A baseline MR examination was performed on day 1 and two MR examinations on day 2. The first MR examination on day 2 was used to obtain pre-exercise data for evaluation of measuring agreement with baseline day 1 data using repeated measurements. The second (post-exercise) MR examination followed 30 minutes of treadmill running at moderate (70% of $HR_{reserve}$) intensity and was used to evaluate changes in mean IVD height and area

determined with the semi-automated and manual methods. Further details about the experimental protocol can be found in XXX [24].

The MR examinations were performed on a 1.5T high-definition 16-channel system (GE Medical Systems, Waukesha, WI). Sagittal T2-weighted fast spin echo images (TR 2200-3000ms, TE 110ms, slice thickness 3mm, slice spacing 3.2mm, image matrix 352 × 320, pixel spacing 0.5469 × 0.5469mm, 4 excitations, 11-13 sagittal slices) were acquired from the participants as they lay in a supine position. The individual MR examinations lasted between 7 and 10 minutes.

2.2 Measurements

2.2.1 Manual digitization

The mean IVD height and area per sagittal slice of the individual L1/L2 – L4/L5 IVDs were computed from manual segmentation procedures performed by a single operator, with no prior digitization experience, after training and familiarization with processing the MR data. The manual segmentation approach involved digitizing a minimum of 7 points along the visible superior and inferior vertebral endplate surfaces defining the interface between IVDs and adjacent vertebral bodies in the sagittal slices (Figure 1A). Intraobserver reliability was determined using 10 repeated measurements of a randomly selected example for all IVD locations. Repeated measurements were completed on separate days with the operator blinded to previous measurements. The ICC and standard error of measurement for IVD height and volume were 0.99 (95% CI: 0.98 to 1.00) and 0.027mm (95% CI: 0.023 to 0.030 mm) and 0.99 (95% CI: 0.99 to 1.00) and 0.07 mm^3 (95% CI: 0.06 to 0.08 mm^3), respectively. Adjacent digitized points were linearly interpolated and the vertical distances between the segmented endplates were computed at 1 mm intervals. Additional details on the experimental methods including intra-observer reliability can be found in XXX [24] and in Figure 1A. Mean IVD height for an individual disc was determined as the average of all the

endplate distances (vertical heights) from the sagittal slices processed for data analyses. A standardized approach using the central 7 sagittal images for each IVD was used in the present analyses to enable complete coverage of the nucleus pulposus, which has an essential role for distributing hydraulic pressure during loading. Mean IVD area across the same 7 central MR slices was used as a volumeric estimate of the IVD mid-substance surrounding the nucleus pulposus.

2.2.2 Semi-automated method

An automated algorithm for 3D segmentation of IVDs and vertebral bodies modified from XXX [35] was used to segment the individual lumbar IVDs and vertebral bodies from the MR images. In the pre-processing stage, a customized intensity adjustment method based on the N4 bias field correction algorithm [37] was initially applied to the region of the lumbar spinal column. Image acquisition noise was reduced by anisotropic diffusion (15 iterations, time step 0.01, conductance 1.0) and image signal intensity histograms were normalized to the signal intensity histogram of an atlas image. Subsequently, images were reformatted using B-Spline interpolation to an isotropic resolution of 0.5469 mm.

The segmentation algorithm is based on an active shape model approach [38] defining models of anatomical variability. These models have to be located at initial referent positions in the MR image before they are deformed to fit previously trained signal intensity information. In the present study, a manual initialization was performed by the user through a single mouse click in the mid-sagittal portion of each lumbar vertebral body in the baseline scans. The input points were used to approximate the lumbar spine curvature and pre-cursor IVD models were automatically positioned mid-way between the points marking the vertebral bodies and oriented to follow the estimated curve. A rigid registration algorithm [39] and propagation of results from the baseline segmentations were used for automatic initialization on images from the pre-exercise and post-exercise MR examinations.

The automated segmentation algorithm provides 3D masks of individual IVDs and vertebral bodies (Figure 1B). These masks fully cover the visualized IVD volumes, including sections exceeding the (parallel) anterior or posterior margins of adjacent vertebral bodies not included in the manual digitization protocol. Therefore, a constrained (bone-bounded) IVD region was defined to enable a direct comparison to manually extracted measures. This was completed by dilating the IVD segmentation mask using a ball structuring element with 1 mm diameter, and by finding the intersection with the segmentation masks of adjacent vertebral bodies (Figure 1B). The anterior and posterior boundary points of these intersections were linearly connected (dotted lines in Figure 1B,C,D) between the inferior and superior borders of vertebral bodies, defining a constrained IVD region used for computation of IVD height and area.

Automatic measurement of an objects' height (or thickness) involves creating a unique association between points on its inner and outer boundary, and some definition of the distance between them. For calculations of IVD heights in this work, we applied the method of Laplacian thickness presented by Jones et al. [40] which is used widely in medical image analysis [41,42]. The inner and outer boundaries are assigned potentials ψ_0 and ψ_1 ($\psi_0 \neq \psi_1$) for which the Laplace's equation $\Delta\psi=0$ is solved. This system defines a partial derivative equation with boundary conditions that allows a unique solution. The thickness at any point is calculated from the solution function as a length of the unique streamline (path of the gradient) connecting boundaries and passing through this point (see Jones et al. [40] for further details).

The Laplacian thickness in the present study was computed in 3D over the 'extended' IVD regions (Figure 1C), defining smooth boundary conditions for Laplace's equation and avoiding any discontinuities at the anterior and posterior IVD borders. The mean IVD height and area were computed by averaging over voxels belonging to the 'non-extended' (bone-bounded) IVD region in the 7 original central sagittal slices used for manual digitization.

2.3 Statistical analysis

Test-retest reliability of the semi-automated and manual methods were assessed on baseline and pre-exercise conditions using the ICC, 95% CI and limits of agreement (LOA) in Bland-Altman plots. Concurrent agreement between the semi-automated and manual methods was determined using pooled repeated measurements according to Bland and Altman [43].

A three-way repeated measures (RM) ANOVA was performed for comparisons between the measurement methods (semi-automated and manual), timing conditions (baseline, pre-exercise, post-exercise) and IVD levels (L1/L2 – L4/L5) on IVD height and area. Significant main effects for the measurement method and timing condition were further investigated *post-hoc* using pairwise comparisons with Bonferroni adjustment. Two-way RM ANOVAs were performed to determine the main effect of IVD level and timing condition for both measuring methods independently.

In all ANOVA tests, an *a priori* significance level of $p < 0.05$ was adopted for rejection of the null hypothesis and the effect size was reported using partial eta-squared (η_p^2). Statistical analyses were carried out using IBM SPSS Statistics 22.0 (SPSS; Chicago, IL).

3 Results

Prior to quantitative analyses, the quality of the semi-automated segmentations was visually assessed to exclude data from failed segmentations. The segmentation was considered unsatisfactory, if a patently discrepancy from the imaged IVD volume was identified within the first two seconds of the visual inspection. Overall, three IVDs (1× L2/L3, 2× L4/L5) were removed from the analysis. The reported mean values and LOA were computed on the remaining 25 IVDs. In all ANOVA tests, the missing values for the three IVDs were replaced using trend analysis to maximize the degrees of freedom. These estimates were computed from the group mean of the other IVDs at the

same segmental level multiplied by a scaling factor for the particular subject. The scaling factor was computed from values at the adjacent superior IVD level as the ratio of the value for the particular subject, computed over the mean measurement at this superior level.

3.1 Reproducibility

Tables 1 and 2 present values for the mean IVD height and area. The semi-automated and manual methods both demonstrated very strong test-retest reproducibility for baseline and pre-exercise height (ICC = 0.99, 0.98) and area (ICC = 0.98, 0.99) measurements. Bland-Altman plots of reproducibility for each method are shown in Figure 2. The 95% LOA for the mean IVD height were 0.93 ± 3.99% (semi-automated) and 1.64 ± 5.55% (manual). The 95% LOA for the mean IVD area were 0.78 ± 7.40% (semi-automated) and 2.34 ± 6.05% (manual) method.

3.2 Concurrent agreement

The interaction between method, timing condition and IVD level factors was not statistically significant ($F_{6,36}$ = 0.68, p = 0.101). There was a significant difference in mean IVD height measured by each method ($F_{1,6}$ = 20.4, p = 0.004, $\eta_p^2 = 0.773$), timing ($F_{2,12}$ = 17.8, p < 0.001, $\eta_p^2 = 0.841$) and IVD levels ($F_{3,18}$ = 31.0, p < 0.001, $\eta_p^2 = 0.930$). The statistically significant bias between the measurement methods was 0.38 mm (95% CI: 0.18 to 0.59 mm, 4.1 ± 2.2%). The bias is further illustrated by the pooled 95% RM LOA on baseline and pre-exercise data (-0.42 to 1.16 mm), and the Bland-Altman plots in Figure 3A,B. Two-way RM ANOVA for the semi-automated measurements revealed a significant main effect of timing ($F_{2,12}$ = 15.5, p < 0.001, $\eta_p^2 = 0.720$) and IVD level ($F_{3,18}$ = 25.3, p < 0.001, $\eta_p^2 = 0.808$). Baseline and pre-exercise mean IVD heights measured by the semi-automated method were not statistically different (95% CI: -0.07 to 0.23 mm, p = 0.365) and there was a significant reduction in mean IVD height after exercise (95% CI: 0.04 to 0.35 mm, p = 0.010). Similarly, there was a significant main effect of timing ($F_{2,12}$ = 12.4, p = 0.001, $\eta_p^2 = 0.673$) and IVD level ($F_{3,18}$ = 31.9, p < 0.001, $\eta_p^2 = 0.842$) on the mean IVD height measured by the manual method.

No statistically significant difference was found between baseline and pre-exercise conditions (95% CI: -0.09 to 0.40 mm, $p = 0.229$), and the reduction in the mean IVD height after exercise was statistically significant (95% CI: 0.06 to 0.45 mm, $p = 0.009$).

For the mean IVD area, there was no significant difference between the methods ($F_{1,6} = 0.003$, $p = 0.956$), while the main effects of timing ($F_{2,12} = 13.8$, $p = 0.001$, $\eta_p^2 = 0.697$) and IVD level ($F_{3,18} = 23.8$, $p < 0.001$, $\eta_p^2 = 0.798$) were statistically significant. The interaction between measuring method, timing condition and IVD level were not statistically significant ($F_{6,36} = 0.886$, $p = 0.515$). The 95% CI between both methods was -8.9 to 9.3 mm^2 (0.1 ± 3.5%) and the pooled 95% RM LOA were -27 to 30 mm^2 (there was no systematic bias). The Bland-Altman plots are presented in Figure 3C,D. Two-way RM ANOVA for mean IVD area measured by the semi-automated method showed a significant main effects of timing ($F_{2,12} = 4.4$, $p = 0.037$, $\eta_p^2 = 0.423$) and IVD level ($F_{3,18} = 16.7$, $p < 0.001$, $\eta_p^2 = 0.736$). Baseline and pre-exercise mean IVD area measured by the semi-automated method were not statistically different (95% CI: -8.3 to 14.2 mm^2, $p > 0.999$) and IVD area was significantly reduced after exercise (95% CI: -0.3 to 10.7 mm^2, $p = 0.030$). For the manual method, significant main effects existed for timing ($F_{2,12} = 16.1$, $p < 0.001$, $\eta_p^2 = 0.728$) and IVD level ($F_{3,18} = 27.4$, $p < 0.001$, $\eta_p^2 = 0.821$). No significant difference was found between baseline and pre-exercise conditions (95% CI: -1.5 to 13.0 mm, $p = 0.118$), and the reduction in mean IVD area after exercise was statistically significant (95% CI: 2.5 to 13.9 mm, $p = 0.005$).

3.3 Computational time

The estimated average time needed for the manual digitization of a single IVD was between 10 and 15 min to obtain both the mean height and area. This estimation does not include staff training, pilot work and other preparations because variability in software, operator skill and training procedures are likely to limit generalizability.

The (semi)-automated computer calculations were performed as a single thread on an Intel i5-

3750 3.4 GHz CPU with 16GB RAM. An average time to pre-process one MR scan was 21 s, to segment one IVD 15 min and to extract the mean area and height with the Laplacian method 4 min 30 s per IVD. Several computational optimization strategies can be implemented to significantly reduce the calculation time, such as region-of-interest image cropping and parallelization in the segmentation. The manual initialization was performed in 5-8 seconds for each case, decidedly reducing the manual processing time from at least 40 minutes (10 to 15 minutes per IVD).

4 Discussion

Both the manual and semi-automated method showed strong test-retest reliability for calculating average IVD height and area. The observed range of the LOA for both mean IVD height and area obtained from the current 3 mm thick sagittal slices highlights the challenges of performing such measurement on 2D MR images (either manually or automatically). The use of higher resolution MR scans (*e.g.* 3D SPACE as used in our recent work [36]) would likely improve the reproducibility of the computerized method. The pooled 95% LOA showed good congruent agreement between methods, considering the reproducibility LOA of both approaches. There is a systematic bias between the semi-automatically and manually measured IVD heights related to the different computation techniques that would need to be considered when comparing findings based on manual approaches similar to present study. In contrast, there was no systematic bias between the methods for mean IVD area.

There have been a limited number of studies investigating computerized methods for MR measurements of IVD height and volume. The results of this study support the future use of our specifically developed methods for computation of IVD height (using Laplacian thickness) and IVD volume (using sagittal areas) for *in vivo* studies of IVD morphometry. The concurrent agreement is comparable to those reported by Tan et al. [33] on CT data, who found (using a similar manual intervention step) differences between manual and semi-automated measurements of -0.072 ± 0.65

mm, 0.36 ± 0.68 mm, and 0.012 ± 0.59 mm for anterior, middle and posterior mid-sagittal IVD height, respectively. This translates to the LOA of -1.35 to 1.20 mm, -0.97 to 1.69 mm, and -1.14 to 1.17 mm. LOA of our technique fit well within these ranges despite the fact that MR data is generally more challenging to (semi-)automatically process due to the higher resolution and well-defined boundary of the vertebral cortical bone in CT data [28,44]. Future work will be required to further improve the accuracy by identifying sources of variation for both semi-automated and manual techniques to increase the power in detecting clinically important changes in IVD morphology [1,20].

XXX [24] have previously reported that 30 min of moderate-intensity running results in a decreased mean IVD height and volume as determined with manual digitization. In the present study, both semi-automated and manual analyses showed a statistically significant decrease in the measures adopted in this work after exercise. These results provide a compelling rationale for the use of the semi-automated method in larger MR investigations into longitudinal morphometric changes, where a change of at least 0.20 mm in IVD height (as in the present study) is expected. Pfirrmann *et al.* [1] reported that the IVD height decreased by 0.43 mm with each grade of IVD degeneration (using 4 grades: normal, mild, moderate, severe). Kwok *et al.* [20] found an increase in middle IVD height of 0.85 mm between normal and osteopenia group, and 0.92 mm between normal and osteoporosis group. This comparison suggests that our technique has the requisite sensitivity to be used for quantitative analyses in clinical studies in symptomatic populations.

Using the semi-automated method reduced the manual processing time to a few seconds per subject required for manual location of the lumbar vertebral bodies and to visually assess the quality of the obtained segmentation results. Our highly automated method does not require any expert training, is not reliant on the subjective identification of IVD boundaries and reduces the processing time required by the manual method (10-15 minutes to measure a single IVD). It offers a time- and expertise-efficient, low cost solution for analyses of large MR datasets/study cohorts. It reduces the manual processing time from one hour to a few seconds per subject and does not require the

operator specialized previous knowledge or training in manual IVD segmentation. Moreover, it provides reproducible and objective results, and the capacity for advanced quantitative analysis of local morphological changes (Figure 4).

4.1 Study limitations and future work

The relatively small subject database is a limitation of the present study. Although the results are promising, the general application of the approach in terms of different experimental design, demographics, pathologies or MR acquisition protocols require further investigation. At this stage, asymptomatic participants were chosen for the current study to assess day-to-day reliability of the technique under controlled conditions without effects of pathology. Despite this, the segmentation scheme has previously been quantitatively validated using several 3D similarity metrics on 68 lumbar and 46 thoracic IVDs (including 7 degenerative IVDs) from 14 asymptomatic subjects [35]. Furthermore, the segmentation algorithm was previously used and validated in a morphological study on 11 symptomatic patients [36]. In that previous work [36], the segmentation results were evaluated by computing the Dice score similarity coefficient [45], and by comparing measures of mid-sagittal middle IVD height and width against manual references. Since the performance of the quantification method presented in the current study is heavily based on the segmentation algorithm, there is good evidence that the proposed methods will generalize well on broader population with IVD pathology. This will however need to be carefully evaluated in future work on larger cohorts. Nevertheless, the number of participants presented in the current study is comparable to other similar validation studies, presenting results on 8 [33] or 10 [34] subjects.

The application of this computerized technique on diverse and larger datasets to focus on detecting localized changes in IVD morphology due to more intensive and/or complex loading interventions represent a logical progression for future work. The computerized approach provides

volumetric segmentations that can be used to assess regional and shape differences, hence potentially allowing observation and quantification of the presence and location of subtle IVD changes. To enhance research and clinical application for investigations that utilize 2D MR images, further effort is required to yield a fully automated processing pipeline by eliminating the manual initialization step.

Orientation of individual IVDs with respect to acquisition plane, slice gap, slice thickness and image resolution influence the visualization of IVDs in MR images and are potential sources of variability that the current study did not directly account for. Future work will include investigations of effects or MR acquisition parameters on morphological measures.

5 Conclusion

This study presents a computerized method for measurement of IVD height (using Laplace's equation) and volume (using sagittal areas) from 2D MR scans of the lumbar spine. The results were compared against measures obtained by manual digitization. Strong reliability was observed for both manual and semi-automated methods and good congruent agreement between the methods was found. Both methods detected acute changes induced by moderate intensity exercise. The presented computerized method has the potential to replace time- and expertise- intensive manual procedures. Quantitative morphological data on IVD geometry can benefit larger clinical (cross-sectional, interventional or longitudinal) studies, design of IVD prosthesis and surgical planning or help studying influence of physical activity and exercise on the spine.

Acknowledgment

The authors wish to acknowledge Mr. XXX (XXX) for his assistance with the data collection, Mr. XXX (XXX) for the manual digitization and Mr. XXX (XXX) for his technical assistance. This research was supported under XXX.

References

[1] Pfirrmann CWA, 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;24:1086–94.

[2] Gocmen-Mas N, Karabekir H, Ertekin T. Evaluation of Lumbar Vertebral Body and Disc: A Stereological Morphometric Study. Int J Morphol 2010;28:841–7.

[3] Beattie PF, Meyers SP. Magnetic resonance imaging in low back pain: general principles and clinical issues. Phys Ther 1998;78:738–53.

[4] Cousins JP, Haughton VM. Magnetic resonance imaging of the spine. J Am Acad Orthop Surg 2009;17:22–30.

[5] Borenstein DG, O'Mara JWJ, Boden SD, Lauerman WC, Jacobson A, Platenberg C, et al. The value of magnetic resonance imaging of the lumbar spine to predict low-back pain in asymptomatic subjects: a seven-year follow-up study. J Bone Joint Surg Am 2001;83:1306–11.

[6] Modic MT, Ross JS. Lumbar degenerative disk disease. Radiology 2007;245:43–61.

[7] Emch TM, Modic MT. Imaging of lumbar degenerative disk disease: history and current state. Skeletal Radiol 2011;40:1175–89.

[8] Johannessen W, Auerbach JD, Wheaton AJ, Kurji A, Borthakur A, Reddy R, et al. Assessment of human disc degeneration and proteoglycan content using T1rho-weighted magnetic resonance imaging. Spine 2006;31:1253–7.

[9] Blumenkrantz G, Zuo J, Li X, Kornak J, Link TM, Majumdar S. In vivo 3.0-tesla magnetic resonance T1rho and T2 relaxation mapping in subjects with intervertebral disc degeneration and clinical symptoms. Magn Reson Med 2010;63:1193–200.

[10] Beattie PF, Morgan PS, Peters D. Diffusion-weighted magnetic resonance imaging of normal and degenerative lumbar intervertebral discs: a new method to potentially quantify the physiologic effect of physical therapy intervention. J Orthop Sports Phys Ther 2008;38:42–9.

[11] Noebauer-Huhmann I-M, Juras V, Pfirrmann CWA, Szomolanyi P, Zbýň Š, Messner A, et al. Sodium MR imaging of the lumbar intervertebral disk at 7 T: correlation with T2 mapping and modified Pfirrmann score at 3T-preliminary results. Radiology 2012;265:555–64.

[12] Mayerhoefer ME, Stelzeneder D, Bachbauer W, Welsch GH, Mamisch TC, Szczypinski P, et al. Quantitative analysis of lumbar intervertebral disc abnormalities at 3.0 Tesla: value of T2 texture features and geometric parameters. NMR Biomed 2012;25:866–72.

[13] Beattie EE, Yoder JH, Moon SM, Vresilovic EJ, Elliott DM, Wright AC. Quantification of intervertebral disc cartilaginous endplate morphology using MRI. 2012 38th Annu Northeast Bioeng Conf 2012:103–4.

[14] He X, Liang A, Gao W, Peng Y, Zhang L, Liang G, et al. The relationship between concave angle of vertebral endplate and lumbar intervertebral disc degeneration. Spine 2012;37:E1068–73.

[15] Wang Y, Battié MC, Videman T. A morphological study of lumbar vertebral endplates: radiographic, visual and digital measurements. Eur Spine J 2012;21:2316–23.

[16] Frobin W, Brinckmann P, Kramer M, Hartwig E. Height of lumbar discs measured from radiographs compared with degeneration and height classified from MR images. Eur Radiol 2001;11:263–9.

[17] Videman T, Battié MC, Parent E, Gibbons LE, Vainio P, Kaprio J. Progression and determinants of quantitative magnetic resonance imaging measures of lumbar disc degeneration: a five-year follow-up of adult male monozygotic twins. Spine 2008;33:1484–90.

[18] Carragee EJ, Kim DH. A Prospective Analysis of Magnetic Resonance Imaging Findings in Patients With Sciatica and Lumbar Disc Herniation : Correlation of Outcomes With Disc Fragment and Canal Morphology. Spine 1997;22:1650–60.

[19] Yang Z, Griffith JF, Leung PC, Lee R. Effect of Osteoporosis on Morphology and Mobility of the Lumbar Spine. Spine (Phila Pa 1976) 2009;34:115–21.

[20] Kwok AWL, Wang Y-XJ, Griffith JF, Deng M, Leung JCS, Ahuja AT, et al. 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 2012;37:E1415–21.

[21] Roberts N, Gratin C, Whitehouse GH. MRI analysis of lumbar intervertebral disc height in young and older populations. J Magn Reson Imaging 1997;7:880–6.

[22] Luoma K, Vehmas T, Riihimäki H, Raininko R. Disc height and signal intensity of the nucleus pulposus on magnetic resonance imaging as indicators of lumbar disc degeneration. Spine 2001;26:680–6.

[23] Dimitriadis AT, Papagelopoulos PJ, Smith FW, Mavrogenis AF, Pope MH, Karantanas AH, et al. Intervertebral Disc Changes after 1 h of Running: a Study on Athletes. J Int Med Res 2011;39:569–79.

[24] XXX

[25] Van der Houwen EB, Baron P, Veldhuizen AG, Burgerhof JGM, van Ooijen PMA, Verkerke GJ. Geometry of the intervertebral volume and vertebral endplates of the human spine. Ann Biomed Eng 2010;38:33–40.

[26] Deeley MA, Chen A, Datteri R, Noble JH, Cmelak AJ, Donnelly EF, et al. Comparison of manual and automatic segmentation methods for brain structures in the presence of space-occupying lesions: a multi-expert study. Phys Med Biol 2011;56:4557–77.

[27] Shi R, Sun D, Qiu Z, Weiss KL. An Efficient Method for Segmentation of MRI Spine Images. IEEE Int. Conf. Complex Med. Eng., Ieee; 2007, p. 713–7.

[28] Michopoulou SK, Costaridou L, Panagiotopoulos E, Speller R, Panayiotakis G, Todd-Pokropek

A. Atlas-based segmentation of degenerated lumbar intervertebral discs from MR images of the spine. IEEE Trans Biomed Eng 2009;56:2225–31.

[29] Chevrefils C, Chériet F, Aubin C-E, Grimard G. Texture analysis for automatic segmentation of intervertebral disks of scoliotic spines from MR images. IEEE Trans Inf Technol Biomed 2009;13:608–20.

[30] Law MWK, Tay K, Leung A, J. Garvin G, Li S. Intervertebral disc segmentation in MR images using anisotropic oriented flux. Med Image Anal 2013;17:43–61.

[31] Seifert S, Wachter I, Schmelzle G, Dillmann R. A knowledge-based approach to soft tissue reconstruction of the cervical spine. IEEE Trans Med Imaging 2009;28:494–507.

[32] McGirt MJ, Eustacchio S, Varga P, Vilendecic M, Trummer M, Gorensek M, et al. A prospective cohort study of close interval computed tomography and magnetic resonance imaging after primary lumbar discectomy: factors associated with recurrent disc herniation and disc height loss. Spine 2009;34:2044–51.

[33] Tan S, Yao J, Yao L, Ward MM. High precision semiautomated computed tomography measurement of lumbar disk and vertebral heights. Med Phys 2013;40:011905.

[34] Boos N, Wallin A, Aebi M, Boesch C. A New Magnetic Resonance Imaging Analysis Method for the Measurement of Disc Height Variations. Spine 1996;21:563–70.

[35] XXX

[36] XXX

[37] Tustison NJ, Gee JC. N4ITK: Nick's N3 ITK implementation for MRI bias field correction. Insight J 2009;2009:1–8.

[38] Cootes TF, Taylor CJ, Cooper DH, Graham J. Active shape models-their training and application. Comput Vis Image Underst 1995;61:38–59.

[39] Ourselin S, Roche A, Subsol G. Reconstructing a 3D structure from serial histological sections. Image Vis Comput 2001;19:25–31.

[40] Jones SE, Buchbinder BR, Aharon I. Three-dimensional mapping of cortical thickness using Laplace's Equation. Hum Brain Mapp 2000;11:12–32.

[41] Acosta O, Bourgeat P, Zuluaga MA, Fripp J, Salvado O, Ourselin S. Automated voxel-based 3D cortical thickness measurement in a combined Lagrangian-Eulerian PDE approach using partial volume maps. Med Image Anal 2009;13:730–43.

[42] XXX

[43] Bland J, Altman D. Measuring agreement in method comparison studies. Stat Methods Med Res 1999;8:135–60.

[44] Klinder T, Ostermann J, Ehm M, Franz A, Kneser R, Lorenz C. Automated model-based

vertebra detection, identification, and segmentation in CT images. Med Image Anal 2009;13:471–82.

[45] Dice L. Measures of the Amount of Ecologic Association Between Species. Ecology 1945;26:297 – 302.

Figures

Figure 1 Illustration of manual and automatic procedures for calculation of mean IVD height. (A) Manual assessment involves delineation of the IVD boundaries by the human operator (black dots) after which a sequence of height measures are (automatically) interpolated at 1mm intervals along these edges (dotted vertical lines). (B) Semi-automatically obtained segmentation masks of vertebral bodies (yellow) and a dilated mask of an IVD (green, overlap with the vertebral bodies in red). Anterior and posterior boundary points of the intersections (black dots) in each slice define a bone bounded IVD space that is used for automatic computations. (C) The extended IVD regions defining potentials ψ_0 (inferior IVD boundary) and ψ_1 (superior IVD boundary) for Laplace's equation. (D) The final Laplacian thickness (quantified by 1mm for visualization purposes). For the colored version of the figure please refer to the online version of the article. IVD, intervertebral disc.

Figure 2 Bland-Altman plots showing the relationship between mean values and differences between baseline and pre-exercise measures of mean IVD height using the (A) semi-automated and (B) manual method, and of mean IVD area using the (C) semi-automated and (D) manual method. The dotted lines show 95% limits of agreement. IVD, intervertebral disc.

Figure 3 Bland-Altman plots showing the congruent agreement between the semi-automated and manual measurements of mean IVD height using the (A) baseline and (B) pre-exercise data, and of mean IVD area using the (C) baseline and (D) pre-exercise data. The dotted lines show 95% limits of agreement. IVD, intervertebral disc.

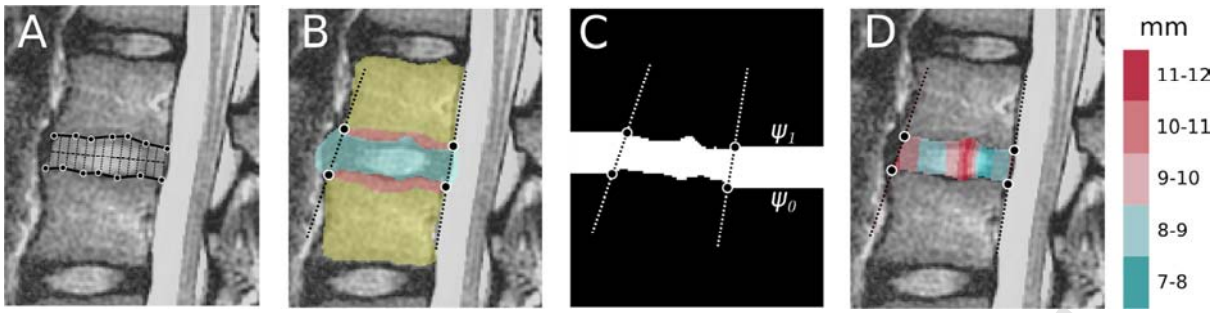
Figure 4 Example IVD (L4/L5) height at (A) baseline, (B) pre-exercise and (C) post-exercise, computed with the Laplacian method illustrating a decrease in overall IVD height across the profile of the disc following exercise. For colored version of the figure please refer to the online version of the article. s, superior; a, anterior; r, right.

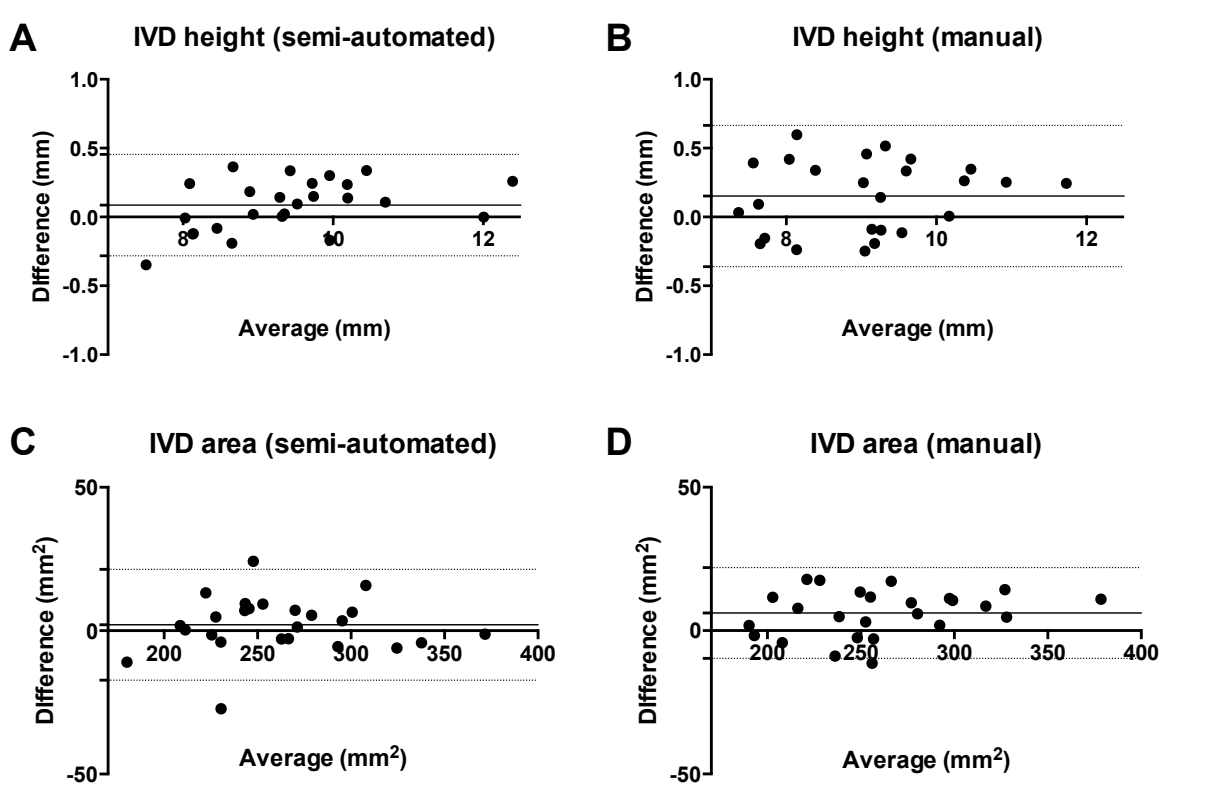
Table 1 Mean disc height (mean ± standard deviation) measured with the semi-automated and manual method.

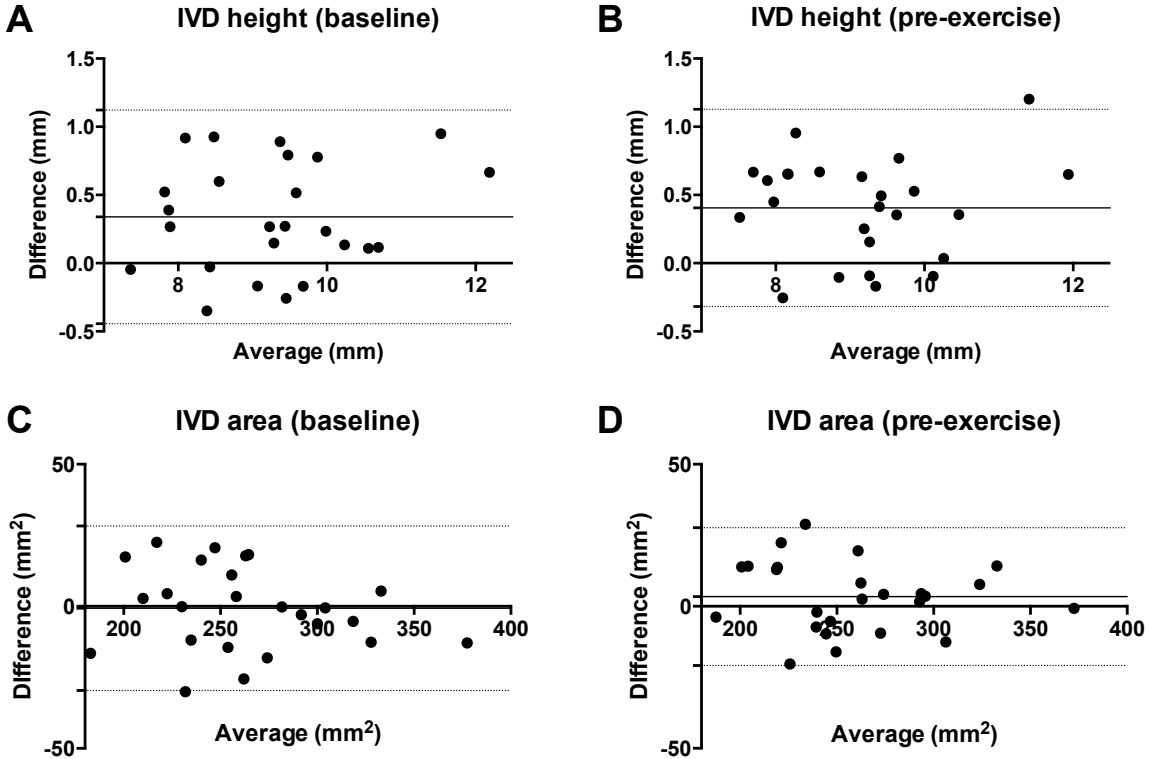
	Semi-automated (*mm*)			Manual (*mm*)		
	Baseline	**Pre-exercise**	**Post-exercise**	**Baseline**	**Pre-exercise**	**Post-exercise**
L1/L2	8.26 ± 0.56	8.32 ± 0.43	8.05 ± 0.45	7.75 ± 0.29	7.70 ± 0.31	7.38 ± 0.40
L2/L3	9.25 ± 0.74	9.14 ± 0.77	8.93 ± 0.70	9.08 ± 0.50	8.80 ± 0.64	8.65 ± 0.56
L3/L4	9.92 ± 0.69	9.74 ± 0.65	9.54 ± 0.71	9.71 ± 0.65	9.57 ± 0.52	9.23 ± 0.51
L4/L5	10.79 ± 1.40	10.66 ± 1.37	10.59 ± 1.33	10.32 ± 1.16	10.16 ± 1.08	9.87 ± 0.98

Table 2 Mean disc sagittal area (mean ± standard deviation) measured with the semi-automated and manual method.

	Semi-automated (mm^2)			Manual (mm^2)		
	Baseline	**Pre-exercise**	**Post-exercise**	**Baseline**	**Pre-exercise**	**Post-exercise**
L1/L2	226 ± 33	225 ± 27	220 ± 26	216 ± 24	211 ± 23	202 ± 20
L2/L3	267 ± 30	263 ± 34	257 ± 31	267 ± 30	259 ± 32	250 ± 30
L3/L4	275 ± 39	273 ± 36	267 ± 36	282 ± 37	276 ± 32	270 ± 33
L4/L5	294 ± 56	292 ± 61	286 ± 51	301 ± 58	295 ± 54	285 ± 49




A
IVD height (semi-automated)
Difference (mm)
Average (mm)
1.0
0.5
0.0
-0.5
-1.0
8
10
12
B
IVD height (manual)
Difference (mm)
Average (mm)
C
IVD area (semi-automated)
Difference (mm²)
Average (mm²)
50
0
-50
200
250
300
350
400
D
IVD area (manual)
Difference (mm²)
Average (mm²)


A
IVD height (baseline)
Difference (mm)
1.5
1.0
0.5
0.0
-0.5
8
10
12
Average (mm)
B
IVD height (pre-exercise)
Difference (mm)
1.5
1.0
0.5
0.0
-0.5
8
10
12
Average (mm)
C
IVD area (baseline)
Difference (mm²)
50
0
-50
200
250
300
350
400
Average (mm²)
D
IVD area (pre-exercise)
Difference (mm²)
50
0
-50
200
250
300
350
400
Average (mm²)