



# Evaluating the Reliability of MRI-Derived Biomechanically-Relevant Measures

Menekse Salar Barim<sup>1</sup>(✉), Richard F. Seseck<sup>2</sup>, M. Fehmi Capanoglu<sup>2</sup>, Sean Gallagher<sup>2</sup>, Mark C. Schall Jr.<sup>2</sup>, and Gerard A. Davis<sup>2</sup>

<sup>1</sup> Oak Ridge Institute for Science and Education (ORISE) Research Fellow, Cincinnati, OH 45202, USA

mzs0053@auburn.edu

<sup>2</sup> Auburn University, Auburn, AL 36830, USA

**Abstract.** Human geometric dimensions have been estimated and approximated in several ways, most recently using Magnetic Resonance Imaging (MRI) techniques. The reliability of MRI-based measurement of structures has been shown to be relatively high. However, a limitation of reliability evaluations is that they often only compare assessments of the same MRI image (e.g., “slice” of the back); differences are only a function of analyst dexterity (in tracing or measuring the structures). Ideally, a reliability test should compare estimates of biomechanical structures using different scans analyzed by different analysts. This presents a “worst case” scenario and provides a robust test of the process’s repeatability. Existing databases of vertebral and intervertebral dimensions tend to be limited with respect to measures of repeatability/reliability with relatively narrow study populations and/or parameters recorded. The objectives of this study were (1) to provide a more accurate data set of lumbar spinal characteristics from 144 Magnetic Resonance Imaging (MRI) scans which were reviewed and measured using the Osirix software program and (2) to assess inter- and intra-rater reliability of the MRI process itself. Reliability for the entire process was evaluated using the aforementioned worst-case scenario of comparing two distinct scans of the same subject with different researchers performing each MRI scan and different researchers performing measurements of the various aspects of vertebral and intervertebral disc dimensions. Geometric dimensions were consistent with measurements obtained in previous MRI-based studies. As expected, larger discrepancies were observed in the “worst case” scenarios (scanners and analysts both different). However, worst case variation was relatively low with 3.6% average absolute difference for anterior endplate measurements, for example, as compared to 2.6% average absolute difference for analysts re-rating their own scans after 1 month. The process for obtaining MRI-derived biomechanical measures appears to be robust.

**Keywords:** Lumbar vertebrae · Endplate morphology  
MRI scan/rescan reliability

## 1 Introduction

Human geometric dimensions have been estimated and approximated in several ways, most recently using Magnetic Resonance Imaging (MRI) techniques. The reliability of MRI-based measurement of structures has been shown to be high (e.g.,  $\geq .90$  ICC) [1, 2]. Low back pain (LBP) is one of the most prevalent and costly health problems experienced by industry [3]. Direct measurements of the spine in multiple planes can provide valuable information about the human vertebrae, particularly for improving subject specific biomechanical models. Research efforts have been made to measure the geometry of the low back using medical imaging techniques and been reported frequently. However, a comprehensive review of the reliability and veracity of the methods themselves has not been studied at the level presented herein. Specifically, a comparison of different scans by operators and reviewed/measured by different analysts has not been conducted on substantive sample size.

Magnetic resonance imaging (MRI) is increasingly used to assess patient lumbar spinal health. MRI has important benefits for imaging the musculoskeletal system [4–6], which provide better visualization of anatomic and pathologic structures, including cartilage, bones and ligaments [4, 7–9]. Morphometric analysis helps to determine the relationships of vertebrae with the anatomical dimensions of low back structures. These morphometric measurements have been questioned by reviewers, specifically, the reliability of the MRI data collection process used here and used previously [10, 11]. While regression relationships to predict low back parameters have been presented, the veracity of their measurement methods has not been adequately studied.

In order to evaluate how precisely these data are collected, a comprehensive scan-rescan study was conducted. Scan-rescan variability is very important because poor reliability of the measurement method itself could call into question the usefulness and accuracy of the regression results.

## 2 Methodology

### 2.1 Study Sample Size

MRI scans of the lumbar intervertebral segments (L2/S1) of thirty-six (36) subjects (20 males and 16 females) who were 19 years of age or greater were scanned on a 3T scanner using standardized T2 weighted protocol. Subject demographics (age, gender, height and weight) were obtained. The average age was 23.7 ( $\pm 3.1$ ) years for males and the average mean was 25.4 ( $\pm 4.7$ ) years for females. Institutional Review Board (IRB) approval and subject informed consent was obtained prior to data collection. MRI data were obtained using a dedicated abdominal coil (Fig. 1). Subjects were placed in a lying position (supine posture) on the scanner, foot support was provided, and they were instructed to keep their body stable (no motion during MRI scans to minimize artifacts).



**Fig. 1.** MRI procedure

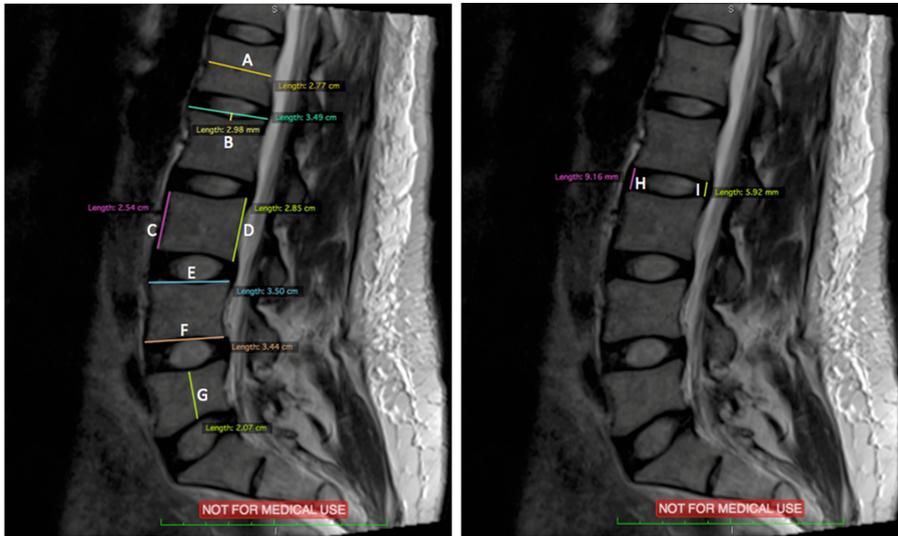
## 2.2 Measurements

MRIs were performed on a 3T unit (Siemens Verio open-bore, Auburn University Research Park, Alabama) using a dedicated abdomen coil. The protocol included the following sequences: Axial Continuous T2-weighted, Sagittal Continuous T2-weighted, and Axial Multi group T2-weighted images with the following parameters; T2-weighted spin-echo (TR 3440 ms; TE 41 ms). All MR images were obtained at a 3-mm slice thickness with 385 FoV read and 100% FoV phase.

Two level-3 MRI certified analysts were provided scans and performed measurements in random order to determine the intra and inter-rater reliability of MRI parameters. In total, 15 parameters were measured. These parameters were measured for the L2/L3, L3/L4, L4/L5, and L5/S1 lumbar regions. Parameters are as follows: Anterior Vertebrae Height (AVH), Posterior Vertebrae Height (PVH), Vertebral Height Index (VHI), Average Height Index (AHI), Sagittal Vertebrae Body Width (SVBW), Sagittal Vertebrae Body Height (SVBH), Height/Weight Index (HWI), Superior Vertebrae Body Length (SVBL) and Inferior Vertebrae Body Length, Length Index (LI) are shown in Fig. 2.

## 2.3 Repeatability of Measurements

In order to assess the reliability and the repeatability of measurements, two different observers measured all parameters three times with at least one month between repeated measurements of the same scan. Data from two observers and six sets of measurements were compared. In the lumbar MRI scans, there are 50 different slices, which can be chosen to perform measurements. In order to test the reliability, specific image



**Fig. 2.** Sagittal MRI scan with measurements. (A) Sagittal Vertebrae Body Width, (B) Concavity Height, (C) Anterior Vertebral Height, (D) Posterior Vertebral Height, (E) Superior Vertebral Body Length, (F) Inferior Vertebral Body Height, (G) Sagittal Vertebral Body Height, (H) Anterior IVD Height, (I) Posterior IVD Height

slices were not selected prior to measurements. Each observer chose the slice they thought was most appropriate for the measurement in question. The results show that, on average, the same slice was selected 61% of the time and observations were within one slice (3 mm) 90% of the time. This is across all conditions including analysts looking at the same scans and worst-case comparisons of different scans and different analysts. The highest levels of absolute agreement occurred when subjects reanalyzed the same scans. The lowest absolute agreement occurred with different analysts, regardless of scan. Analysts were within 1 slice of each other 78%-85% of the time for these three comparisons and in no case did analysts differ by more than two slices (6 mm).

### 3 Results

Mean values, standard deviations and range of data for the lower lumbar spine, which were obtained from the MRI measurements are reported. Scheffe tests were done choosing four different scenarios, which are Inter-rater reliability (same scan different observer/analysts), Intra-Best X (observer X measures her/his own scan two different times), Intra-Best Y (observer Y measures her/his own scan two different times) and Worst-Case (different scans observed by different analysts). According to the results, the most different measurements were observed in Worst-Case, which was predicted before the study (Tables 1 and 2).

**Table 1.** Vertebral body dimensions of lumbar region (L2, L3, L4, L5, S1)

Dimensions	Sex	L2	L3	L4	L5	S1
Anterior Vertebrae Height (AVH)	M	2.66 ± 0.17	2.77 ± 0.19	2.83 ± 0.19	2.89 ± 0.19	3.09 ± 0.19
		(2.09–3.14)	(2.32–3.25)	(2.4–3.42)	(2.29–3.2)	(2.59–3.57)
		2.46 ± 0.18	2.6 ± 0.17	2.62 ± 0.19	2.67 ± 0.19	2.98 ± 0.23
	F	(1.96–2.82)	(2.20–2.91)	(2.05–3.05)	(2.3–3.15)	(2.39–3.37)
		2.57 ± 0.21	2.69 ± 0.20	2.74 ± 0.22	2.8 ± 0.22	3.04 ± 0.22
		(1.96–3.14)	(2.20–3.25)	(2.05–3.42)	(2.29–3.2)	(2.39–3.57)
Posterior Vertebrae Height (PVH)	M	2.78 ± 0.19	2.84 ± 0.18	2.73 ± 0.16	2.52 ± 0.19	2.47 ± 0.26
		(2.33–3.18)	(2.3–3.2)	(2.27–3.03)	(2.02–2.83)	(1.55–3.14)
		2.57 ± 1.94	2.6 ± 0.17	2.48 ± 0.17	2.26 ± 0.18	2.29 ± 0.21
	F	(2.02–2.99)	(2.05–2.93)	(2.13–3)	(1.85–2.65)	(1.81–3.26)
		2.69 ± 0.22	2.73 ± 0.21	2.62 ± 0.21	2.40 ± 0.22	2.39 ± 0.26
		(2.02–3.18)	(2.05–3.2)	(2.13–3.03)	(1.85–2.83)	(1.55–3.26)
Sagittal Vertebrae Body Width (SVBW)	M	2.90 ± 0.23	3.08 ± 0.22	3.12 ± 0.25	2.98 ± 0.24	2.26 ± 0.26
		(2.29–3.41)	(2.67–3.98)	(2.57–3.74)	(2.32–3.75)	(1.75–2.96)
		2.57 ± 0.25	2.74 ± 0.25	2.79 ± 0.25	2.69 ± 0.26	1.89 ± 0.21
	F	(1.78–3.14)	(2–3.18)	(2.14–3.18)	(2.1–3.32)	(1.43–2.4)
		2.75 ± 0.28	2.93 ± 0.29	2.98 ± 0.3	2.86 ± 0.29	2.09 ± 0.3
		(1.78–3.41)	(2–3.98)	(2.14–3.74)	(2.1–3.75)	(1.43–2.96)
Sagittal Vertebrae Body Height (SVBH)	M	2.32 ± 0.23	2.37 ± 0.27	2.41 ± 0.24	2.34 ± 0.25	2.55 ± 0.25
		(1.71–2.71)	(1.72–3.35)	(1.77–2.88)	(1.65–2.76)	(1.87–3.13)
		2.24 ± 0.17	2.27 ± 0.16	2.77 ± 0.16	2.15 ± 0.19	2.44 ± 0.21
	F	(1.85–2.71)	(1.96–2.73)	(1.83–2.66)	(1.63–2.49)	(1.94–2.8)
		2.28 ± 0.21	2.33 ± 0.23	2.34 ± 0.22	2.26 ± 0.25	2.50 ± 0.24
		(1.71–2.71)	(1.72–3.35)	(1.77–2.88)	(1.63–2.76)	(1.87–3.13)
Superior Vertebrae Body Length (SVBL)	M	3.08 ± 0.20	3.22 ± 0.19	3.27 ± 0.22	3.28 ± 0.22	3.05 ± 0.24
		(2.38–3.58)	(2.63–3.88)	(2.76–3.84)	(2.83–4)	(2.31–4.25)
		2.77 ± 0.25	2.91 ± 0.23	2.97 ± 0.23	2.99 ± 0.22	2.76 ± 0.24
	F	(1.93–3.3)	(2.18–3.36)	(2.42–3.33)	(2.51–3.37)	(2.31–3.17)
		2.94 ± 0.27	3.08 ± 0.26	3.13 ± 0.27	3.15 ± 0.26	2.92 ± 0.28
		(1.93–3.58)	(2.18–3.88)	(2.42–3.84)	(2.51–4)	(2.31–4.25)
Inferior Vertebrae Body Length (IVBL)	M	3.14 ± 0.19	3.22 ± 0.22	3.31 ± 0.22	3.14 ± 0.24	1.94 ± 0.38
		(2.57–3.71)	(2.26–3.76)	(2.92–3.88)	(2.36–4.05)	(1.14–2.99)
		2.81 ± 0.25	2.91 ± 0.24	2.99 ± 0.24	2.89 ± 0.28	1.52 ± 0.23
	F	(1.91–3.3)	(2.21–3.29)	(2.27–3.35)	(2.33–3.44)	(1.03–2.08)
		2.99 ± 0.27	3.08 ± 0.28	3.16 ± 0.28	3.03 ± 0.29	1.76 ± 0.38
		(1.91–3.71)	(2.21–3.76)	(2.27–3.88)	(2.33–4.05)	(1.03–2.99)

**Table 2.** Scheffe's test results for vertebrae dimensions

Dimensions	Contrast coefficient	L2	L3	L4	L5	S1
Anterior Vertebral Height (AVH)	Inter rater	0.7305	0.9963	0.9904	0.9903	0.9944
	Inter Best X	0.9618	0.9974	1.0000	0.9995	0.9943
	Inter Best Y	0.9998	0.7578	0.9607	0.9129	0.9949
	Worst Case	0.8759	0.5398	0.2904	0.1672	0.0830
Posterior Vertebral Height (PVH)	Inter rater	0.0000	0.0000	0.1028	0.0766	0.0496
	Inter Best X	0.9865	0.9943	0.8590	0.9999	1.0000
	Inter Best Y	0.8429	0.5571	0.9992	0.8121	0.9745
	Worst Case	0.0000	0.0000	0.0426	0.0001	0.0127
Sagittal Vertebrae Body Width (SVBW)	Inter rater	0.7059	0.0000	0.0002	0.0973	0.3597
	Inter Best X	0.9997	0.9736	1.0000	0.9912	1.0000
	Inter Best Y	0.9910	0.9601	1.0000	1.0000	1.0000
	Worst Case	0.0463	0.0074	0.0003	0.7799	0.1002
Sagittal Vertebrae Body Height (SVBH)	Inter rater	0.0006	n/a	0.1873	0.9884	0.7720
	Inter Best X	0.9822	n/a	1.0000	1.0000	0.9992
	Inter Best Y	1.0000	n/a	1.0000	1.0000	1.0000
	Worst Case	0.0266	n/a	0.7488	0.9996	0.9863
Superior Vertebrae Body Length (SVBL)	Inter rater	0.0216	0.0000	0.0002	0.0002	0.9887
	Inter Best X	1.0000	0.9836	1.0000	1.0000	1.0000
	Inter Best Y	0.9881	1.0000	0.9994	1.0000	0.9993
	Worst Case	0.0034	0.0000	0.0000	0.0000	0.0321
Inferior Vertebrae Body Length (SVBL)	Inter rater	0.0220	0.0071	0.1949	0.3384	n/a
	Inter Best X	0.9992	0.9998	0.9995	0.9927	n/a
	Inter Best Y	1.0000	1.0000	1.0000	0.9997	n/a
	Worst Case	0.0000	0.0001	0.0143	0.0396	n/a

## 4 Discussion

Lumbar vertebrae measurements have been performed by a number of studies [12–20]. In all of these studies, the accuracy of the measurement techniques were not reported. Only sample sizes were reported; no measures of repeatability were included. In the present study, the age range was not broad enough to draw conclusions regarding age differences in lumbar spine measurements. However, the number of subjects was sufficient to explore the repeatability of the measurement process itself and to provide accurate information regarding geometric dimensions of vertebrae.

MRI shows great promise for improving biomechanical modelling of the lumbar spine using subject specific information. This study demonstrates the high degree of repeatability associated with MRI as a means of measuring biomechanically relevant structures.

## 5 Conclusion

The scan matters! There are differences based on the scan taken. Average absolute differences were greatest when different scans were compared. For example, the average absolute difference expected between measures of the same scan for the L2 Anterior Vertebrae Height was 3% (max observed 11%) while the average absolute difference expected for worst case comparisons of the L2 Anterior Vertebrae Height was 4.5% (max observed 20%). This study demonstrates that MRI derived measures are consistent both within observers and across different observers and different scans. Regressions derived from such MRI data sets and used to predict structure sizes should therefore be impacted much more by actual individual differences between subjects than by the measurement error of analysts.

**Acknowledgement.** This publication was partially supported by Grant #2T420H008436 from NIOSH. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention. Use of trade names is for identification only and does not constitute endorsement by the Public Health Service or by the U.S. Department of Health and Human Service.

## References

1. Tang R, Gungor C, Sesek RF, Foreman KB, Gallagher S, Davis GA (2016) Morphometry of the lower lumbar intervertebral discs and endplates: comparative analyses of new MRI data with previous findings. *Eur Spine J* 25(12):4116–4131
2. Gungor C, Tang R, Sesek RF, Foreman KB, Gallagher S, Davis GA (2015) Morphological investigation of low back erector spinae muscle: historical data populations. *Int J Ind Ergon* 49:108–115
3. Zhou SH, McCarthy ID, McGregor AH, Coombs RRH, Hughes SPF (2000) Geometrical dimensions of the lower lumbar vertebrae – analysis of data from digitized CT images. *Eur Spine J* 9(3):242–248
4. Schibany N, Ba-Ssalamah A, Marlovits S, Mlynarik V, Nobauer-Huhmann IM, Striessnig G, Shodjai-Baghini M, Heinze G, Trattnig S (2005) Impact of high field (3.0 T) magnetic resonance imaging on diagnosis of osteochondral defects in the ankle joint. *Eur J Radiol* 55:283–288
5. Barr C, Bauer JS, Malfair D, Ma B, Henning TD, Steinbach L, Link TM (2007) MR imaging of the ankle at 3 Tesla and 1.5 Tesla: protocol optimization and application to cartilage, ligament and tendon pathology in cadaver specimens. *Eur Radiol* 17(6):1518–1528
6. Shapiro MD (2006) MR imaging of the spine at 3T. *Magn Reson Imaging Clin N Am* 14:97–108
7. Link TM, Sell CA, Masi JN et al (2006) 3.0 vs 1.5 T MRI in the detection of focal cartilage pathology: ROC analysis in an experimental model. *Osteoarthr Cartil* 14:63–70
8. Phan CM, Matsuura M, Bauer JS et al (2006) Trabecular bone structure of the calcaneus: comparison of MR imaging at 3.0 and 1.5 T with micro-CT as the standard of reference. *Radiology* 239:488–496
9. Zhao J, Krug R, Xu D, Lu Y, Link TM (2009) MRI of the spine: image quality and normal-neoplastic bone marrow contrast at 3T versus 1.5 T. *Am J Roentgenol* 192(4):873–880

10. Tang R (2013) Morphometric analysis of the human lower lumbar intervertebral discs and vertebral endplates: experimental approach and regression models. Doctoral Dissertation, Auburn University, Auburn, AL
11. Gungor C (2013) Prediction of the erector spinae muscle lever arm distance for biomechanical models. Doctoral Dissertation, Auburn University, Auburn, AL
12. Berry JL, Moran JM, Berg WS et al (1987) A morphometric study of human lumbar and selected thoracic vertebrae. *Spine* 12:362–367
13. Einstein S (1983) Lumbar canal morphometry for computerized tomography in spinal stenosis. *Spine* 8(2):187–191
14. Fang D, Cheung K, Ruan D, Chan F (1994) Computed tomographic osteometry of the Asian lumbar spine. *J Spinal Disord* 7:307–316
15. Gilad I, Nissan M (1985) Sagittal evaluation of elemental geometrical dimensions of human vertebrae. *J Anat* 143:115–120
16. Gilad I, Nissan M (1985) Sagittal radiographic measurements of the cervical and lumbar vertebrae in normal adults. *Br J Radiol* 58:1031–1034
17. Nissan M, Gilad I (1984) The cervical and lumbar vertebrae - an anthropometric model. *Eng Med* 13:111–114
18. Larsen JL, Smith D (1980) Vertebral body size in lumbar spinal canal stenosis. *Acta Radiol Diagn* 21:785–788
19. Postacchini F, Ripani M, Carpano S (1983) Morphometry of the lumbar vertebrae – an anatomic study in two Caucasoid ethnic groups. *Clin Orthop* 172:296–303
20. Van Schaik JJ, Verbiest H, Van Schaik FD (1985) Morphometry of lower lumbar vertebrae as seen on CT scans: newly recognized characteristics. *AJR Am J Roentgenol* 145(2):327–335